

Spatially Varying Autoregressive Models for Prediction of New HIV Diagnoses

Lyndsay Shand, Bo Li, Trevor Park,¹ Dolores Albarracín²

Abstract

In demand of predicting new HIV diagnosis rates based on publicly available HIV data that is abundant in space but has few points in time, we propose a class of spatially varying autoregressive (SVAR) models compounded with conditional autoregressive (CAR) spatial correlation structures. We then propose to use the copula approach and a flexible CAR formulation to model the dependence between adjacent counties. These models allow for spatial and temporal correlation as well as space-time interactions and are naturally suitable for predicting HIV cases and other spatio-temporal disease data that feature a similar data structure. We apply the proposed models to HIV data over Florida, California and New England states and compare them to a range of linear mixed models that have been recently popular for modeling spatio-temporal disease data. The results show that for such data our proposed models outperform the others in terms of prediction.

Keywords: Bayesian hierarchical models, conditional autoregressive models, copula, spatio-temporal data

Short title: Spatio-Temporal Models for HIV Prediction

¹Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA. E-mail: *lshand2;libo;thp2@illinois.edu*.

²Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA. E-mail: *dalbarra@illinois.edu*.

1 Introduction

Human immunodeficiency virus (HIV) infections are life-changing events to those inflicted and if left untreated can lead to acquired immunodeficiency disease (AIDS). The Centers for Disease Control and Prevention (CDC) has reported that although nationally the number of newly diagnosed cases of HIV has declined by 19% in the last decade, progress has been uneven. For example, some demographic groups, such as African Americans, and some geographic regions, such as the south of the US, have shown slower declines, if any, exhibiting consistently high HIV rates ³. Yet other regions that were not traditionally affected have seen dramatic outbreaks, such as Scott county, Indiana. Regional prediction of disease is central to orchestrating appropriate public health responses. The National HIV/AIDS Strategy ⁴ identifies a key goal of intensifying efforts in the communities with the greatest concentration of HIV cases. Developing models to predict future diagnoses should allow health departments to intervene before the surge in new diagnoses occurs. Although admittedly imperfect, earlier intervention offers the possibility of reaching people living with HIV sooner, and of improving health and decreasing infectiousness through timely treatment.

In this article, we will focus on the prediction of new HIV diagnosis rates at the county level using publicly available data. It is well-known that the presence of infection in a region is partly influenced by the social and economic demographics and the prevalence of other sexually transmitted diseases (STD) in its population. Known demographic variables linked to health disparities include but are not limited to education, income, health-care access, sexual orientation, and ethnicity. However, these covariates alone are insufficient to explain the entire variability in the HIV data. After the effects due to the covariates are removed, the spread of infection across the

³<http://www.cdc.gov/hiv>

⁴<https://www.aids.gov/federal-resources/national-hiv-aids-strategy/overview/>

U.S. still exhibits strong spatially and temporally varying patterns as values among neighboring regions and time periods tend to be similar. It is therefore necessary to have appropriate modeling techniques that incorporate both the variability due to space-time dependencies and the variability due to covariates. Another major challenge that comes with the prediction of HIV lies in the rarity of the disease, leading to few to no incidents in many regions. This sparsity strongly encourages prediction methods to take advantage of the spatial and temporal dependency structures so that the statistical inference at one location can borrow strength from neighboring regions in both space and time. All aforementioned challenges motivate us to investigate efficient statistical models with the goal of predicting new HIV diagnoses.

Incidences of disease are most popularly modeled as Poisson or binomial random variables with mean functions dependent on relative risk, expressed as the ratio of observed risk to expected risk. The expected risk is calculated by standardizing the observed number of people at risk in each combination of region and period by age, gender or other categorization. This standardization can be performed either externally or internally, depending on whether the information retrieved for standardization is from another data source or from within the given data. The relative risk is then modeled through a link function, say, log or logit link function. Modeling of the transformed relative risk at region i and period t , $\eta_{i,t}$, is the main goal of disease modeling. This goal is usually achieved through an additive linear function which is essentially the sum of linear effects from certain covariates and spatial/temporal/spatio-temporal random effects. The spatial or temporal random effects may be viewed as surrogates for unobserved regional or time-changing covariates.

Spatial dependence for aggregated disease incidence is most commonly modeled through the conditional autoregressive (CAR) model. The intrinsic CAR (ICAR) model proposed by Besag et al. (1991) has been widely used but it imposes restrictions

on the correlation structure and is mainly appropriate for data with strong spatial correlation. To make the model more flexible, Besag et al. (1991) proposed to combine the ICAR model with an additional set of independent random effects to allow for a range of correlation strengths. Cressie (1993) and Stern and Cressie (2000) proposed to have a single set of random effects that still enables varying strengths of spatial correlation to be captured. Leroux et al. (1999) modified Cressie (1993) to make it more theoretically appealing by providing the explicit joint distribution of the spatial observations. Later this model was used in MacNab (2003) to analyze intraventricular hemorrhage incidence rates. Lee (2011) showed that among the above mentioned models, Leroux et al. (1999) is preferable because it consistently produces the best model fit across a large range of possible spatial correlation strengths.

Spatio-temporal disease models vary widely in their fashion of incorporating temporal variability and its interactions with spatial variability. Bernardinelli et al. (1995) and Sun et al. (2000) captured the spatially varying temporal trend using a linear function of time with region-specific intercepts and slopes. Assunção et al. (2001) incorporated an additional quadratic term in time to allow for curved trends. Alternatively, MacNab and Dean (2001, 2002) used B-splines to model the temporal trend independently for each region such that the models share information in time but not across space. Waller et al. (1997) and Xia and Carlin (1998) on the contrary built independent spatial models for each time period. Although such a model is highly flexible, it can quickly become high dimensional by treating each time point separately. To make the temporal evolution less restrictive and to allow observations to share information in both space and time, Knorr-Held (2000) proposed a generalized linear mixed model framework with spatial and temporal main effects, modeled as an ICAR and random walk respectively, as well as a possible space-time interaction term, expressed as the Kronecker product of the main effects. Lagazio et al. (2003) and

Schmid and Held (2004) extended this model by incorporating temporally varying covariates, creating age-period-cohort models. Later, Nobre et al. (2005) proposed a dynamic generalized linear mixed model by allowing both the fixed and random effects to evolve over time. Martínez-Beneito et al. (2008) proposed a multivariate autoregressive model that uses a single, spatially-invariant parameter to estimate the disease’s changing rate over time and used Besag et al. (1991) to capture the spatial correlation in the innovations of the autoregressive model.

More recently, nonparametric methods have also become engaged in modeling the dependencies in health and disease data. Cai et al. (2014) built a semiparametric model with spatio-temporally varying regression coefficients which are further decomposed into fixed, spatially varying and temporally varying components. While the temporally varying components are modeled through a dynamic model, the spatially varying components are modeled via a nonparametric Dirichlet process. Motivated by Knorr-Held (2000), Bauer et al. (2016) decomposed disease risk into purely spatial and purely temporal components and a space-time interaction term. The space-time interaction term is modeled via tensor product splines instead of as a product of the main effects, thus reducing the complexity of the model. Ugarte et al. (2012) applied a similar additive structure to prostate cancer mortality data and modeled the random effects using P-splines. Ruiz-Medina et al. (2014) proposed another nonparametric approach to disease modeling by treating the risk of breast cancer mortality as functional data.

While the majority of the aforementioned statistical spatio-temporal models mainly focus on capturing the underlying pattern of disease risk and/or the association between the disease and environmental variables, our goal here is to make predictions for disease rates. We propose a new class of spatially varying autoregressive (SVAR) models for this purpose, and compare them to six competing models

based on the popular additive modeling framework summarized in Knorr-Held (2000) and additionally two spatially invariant autoregressive models. Our SVAR models allow the temporal correlation to be location specific while still remaining parsimonious so that the models are governed by only a few parameters. This is particularly important for the HIV data presented here, because the data only has at most six observations in time for each county, making the individual estimation of an autoregressive model for each county unaffordable. We will evaluate the prediction capabilities of our models versus the additive models and spatially invariant models by applying them to the HIV data over Florida, California and New England states.

This paper is organized as follows: Section 2 describes the data, Section 3 introduces our SVAR models and presents the competing models, and Section 4 presents the model comparison and prediction results. Finally, Section 5 provides a brief discussion on the implications and possible extensions of the approaches explored. Additional plots and all derivations are deferred to the Supplement.

2 Data

Annual new HIV diagnosis data from 2008 to 2014 at the county level across the United States is available at *AIDSVu.org*. HIV rates are reported as the number of cases per 100,000 people for a given county. In order to protect the privacy of individuals, HIV rates in a county may be suppressed in any of the following situations: (1) a county has very few cases (< 5) or has a small population size (< 100), (2) the state health department requested not to release its data to AIDSVu due to re-release agreements with the CDC, and (3) there are no counties in the state such as in Alaska, District of Columbia and Puerto Rico. Due to the rareness of the disease as well as the confidentiality constraints, only 25% of all possible county-time observa-

tions across the United States have new diagnoses available in the given time frame. Figure 1 shows the sparsity of new diagnosis rates across the US in 2012. In this map, negative rates indicate missing values. Hot spots seen on this map agree with reports made by the CDC stating that the highest rates of diagnoses are in the South, the West and the Northeast. For this reason, we focus on three concentrated areas of the US: Florida with 67 counties and 75% of county-time observations available, California with 58 counties and 59% of county-time observations available, and the group of seven New England states: Connecticut, Delaware, Maryland, Massachusetts, New Jersey, New York, and Pennsylvania with 199 counties and 74% county-time observations available collectively.

General demographic annual summaries by county such as age, gender, sex at birth, race, house-hold information, and population for the years 2007-2014 were obtained from the United States Census Bureau. Other social and economic variables such as education, marital status, income, and same-sex couple households are also available from the American Community Survey run by the U.S. Census Bureau. Since annual demographic data for all variables is only available for highly populated counties, we use 3-year or 5-year averaged data to replace the missing annual or 3-year averaged variables respectively. Prevalence of other STDs such as chlamydia and syphilis are also available from *healthindicators.gov*. Although there are over 100 possible covariates from all demographic and prevalence data described above, past research suggests that only a small fraction of social, economic, and disease variables are significantly related to the spread of HIV. However, as not all variables currently available have been previously investigated, we performed our own preliminary analysis for initial variable selection. We considered variables observed in the previous year as covariates and using a step-wise regression procedure performed on all US data available, we selected 28 significant variables. The relatively more significant co-

variates include race, education level, activity level, chlamydia, and number of same sex households, which are known to be related to the epidemic nature of the HIV disease. The effects of these variables on HIV have also been shown in previous HIV studies (e.g., Balaji et al., 2013 and Albarracín et al., 2010).

3 Model development and comparison

Our goal is to make county level one year ahead predictions of new HIV diagnosis rates based on the observed county level new HIV diagnoses and selected covariates. The new HIV diagnoses is only observed from 2008-2014. For such a short time series, a linear or quadratic temporal trend model may impose too strong an assumption. We therefore choose an autoregressive model with order 1 (AR(1)) structure to model the temporal correlation of each county. However, an obvious challenge with this data set is that it is statistically unreliable to fit individual AR(1) models for each county due to so few observed time points. Moreover, the evolution rates of neighboring counties are likely similar so independent AR(1) models may fail to account for the anticipated similarity among the autoregressive coefficients, ρ , in nearby or adjacent counties.

In the exploratory data analysis we fit an AR(1) model for each county independently and examined the spatial correlation of the estimated ρ s. Figure 2 shows the independent ρ estimates by county (only for counties with no more than 3 missing years) and Table 1 reports both the Moran's I and Geary's C testing results. Both tests provides strong evidence of spatial correlation for New England, weak evidence for California, and no evidence for Florida, yet Figure 2 shows similar ρ estimates among neighboring counties for all three regions. Compared to New England, the testing results for Florida and California could be more influenced by the pattern of

missing counties due to the smaller number of counties in these two states.

To accommodate the characteristics of the data and also ensure the model inference to be feasible, we propose to jointly model the individual evolution rates in each county by modeling the county specific ρ as a spatially dependent random process using the CAR model developed by Leroux et al. (1999). The choice of this CAR model follows the argument in the Introduction. This SVAR modeling strategy has several advantages: first, it allows for flexible county-specific autoregressive coefficients; second, it makes the estimation of an AR(1) model for each county reliable by borrowing strength from neighbors since the spatial locations will essentially act as “replicates” in the estimation; and third, it dramatically reduces the rank of the model, easing the potential for overfitting.

3.1 Model formulation

Let $Y_{i,t}$ denote the new HIV diagnosis rate for county $i = 1, \dots, n$ at year $t = 1, \dots, T$. In general, disease modeling first approximates the distribution of raw counts using a Poisson or binomial process with the mean defined as a function of the standardized relative risk, and then focuses on modeling the relative risk strategically using linear mixed models. These models can be put in the framework of generalized linear mixed models (Breslow and Clayton, 1993). However, if the counts are sufficiently large then some type of transformation can be used to attain the normality of the transformed data (Waller et al., 1997). A transformation can be advantageous over Poisson or binomial approximation because the Gaussian model for the transformed data can allow for direct modeling of overdispersion in the data. Due to the sensitivity of the data, rates below five are suppressed, ensuring that the majority of data are sufficiently large and normality may be achieved by a transformation of $Y_{i,t}$. For our data, we found that a log transformation works well in transforming data to

approximate Gaussian distribution, see Figures 1-4 in the Supplement. We therefore first take a log transformation of $Y_{i,t}$ and then remove its global mean to obtain a zero-mean Gaussian random process $Z_{i,t}$.

We then focus on modeling $Z_{i,t}$ using Bayesian hierarchical models (BHMs). We model $Z_{i,t}$ in the first level of the BHM as below:

$$Z_{i,t} = \eta_{i,t}(\boldsymbol{\beta}, \boldsymbol{\theta}) + \epsilon_{i,t}, \quad (3.1)$$

where $\eta_{i,t}$ is a spatio-temporal random process governed by parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. An important feature of aggregated disease modeling is that the variance of $Y_{i,t}$ depends on its corresponding population size, i.e., $n_{i,t}$ of county i at time t . To achieve such variance specification, we specify $\boldsymbol{\epsilon} \sim N(0, \sigma^2 \mathbf{Q})$ in model (3.1), where \mathbf{Q} is an $nT \times nT$ diagonal matrix with the $(it)^{th}$ diagonal entry as $q_{it} = c/(n_{i,t}Y_{i,t})$ for a constant $c = 100,000$. The derivation for this particular form of Q can be seen in the Supplement Section 2. At the second level, we model $\eta_{i,t}$ using a SVAR model. The basic form of our SVAR model is

$$\eta_{i,t}(\boldsymbol{\beta}, \boldsymbol{\theta}) = X_{i,t-1}^T \boldsymbol{\beta} + \psi_{i,t} \rho_i (Z_{i,t-1} - X_{i,t-2}^T \boldsymbol{\beta}), \quad (3.2)$$

where $X_{i,t-k}^T \boldsymbol{\beta}$ denotes the linear effects of the previous year's covariates at county i on $Z_{i,t-k+1}$ for $k = 1, 2$, and $\rho_i \in (-1, 1)$, $i = 1, \dots, n$, are spatially varying AR(1) coefficients. The coefficient $\psi_{i,t} = \sqrt{q_{it}/q_{i(t-1)}}$ is added to ensure that ρ_i measures the correlation between two random components, because the variance of $Z_{i,t}$ is proportional to q_{it} .

Depending on the dependency structure in data, we may add additional random effects to Model (3.2). For example, if the spatial correlation in ρ_i is not sufficient to capture the dependency in $Z_{i,t}$, we can add a spatial random effect term ϕ_i to the model:

$$\eta_{i,t}(\boldsymbol{\beta}, \boldsymbol{\theta}) = X_{i,t-1}^T \boldsymbol{\beta} + \psi_{i,t} \rho_i (Z_{i,t-1} - X_{i,t-2}^T \boldsymbol{\beta}) + \phi_i, \quad (3.3)$$

or additionally a spatio-temporal interaction random effect $\delta_{i,t}$:

$$\eta_{i,t}(\boldsymbol{\beta}, \boldsymbol{\theta}) = X_{i,t-1}^T \boldsymbol{\beta} + \psi_{i,t} \rho_i (Z_{i,t-1} - X_{i,t-2}^T \boldsymbol{\beta}) + \phi_i + \delta_{i,t}.$$

The spatially varying autoregressive models capture the space-time correlation, also called spillover effects between neighbors, while the random effects can account for spatial or spatiotemporal variability from either correlated or uncorrelated unknown factors. Note that the major difference between Model (3.3) and Martínez-Beneito et al. (2008) is that we allow ρ_i to vary in space. Moreover, our spatial random effects $\boldsymbol{\phi} = (\phi_1, \dots, \phi_n)^T$ will be governed by the CAR model from Leroux et al. (1999) rather than Besag et al. (1991) to capture a broader range of spatial correlation strength. For all models considered throughout this paper, coefficients in $\boldsymbol{\beta}$ are given independent normal priors.

3.2 Prior specification for ρ_i

The prior specification for ρ_i is challenging because the typical Gaussian CAR prior is not applicable due to the truncated range of ρ_i . We therefore propose to model the prior for ρ_i using a copula approach which enables modeling the marginal distribution separately from the dependency structure. By this strategy we are able to model the dependency of ρ_i using a Gaussian random field while still maintaining the flexibility of choosing an appropriate marginal distribution to respect the truncated nature of ρ_i . We first define the joint cumulative distribution function (CDF) of ρ_1, \dots, ρ_n through a Gaussian copula model:

$$F(\rho_1, \dots, \rho_n) = C_{\Omega}(U(\rho_1), \dots, U(\rho_n)), \quad (3.4)$$

where $U(\cdot)$ represents the CDF of the marginal distribution of ρ_i , and $C_{\Omega}(\cdot)$ is a Gaussian copula with covariance matrix Ω defined by:

$$C_{\Omega}(u_1, \dots, u_n) = \boldsymbol{\Phi}_{\Omega}(\boldsymbol{\Phi}^{-1}(u_1), \dots, \boldsymbol{\Phi}^{-1}(u_n)), \quad (3.5)$$

where $\Phi_{\Omega}(\cdot)$ is the joint CDF of a multivariate Gaussian distribution with mean zero and covariance matrix Ω , and $\Phi^{-1}(\cdot)$ is the inverse CDF of a standard normal random variable.

Here we assume $U(\cdot)$ to be uniform(-1,1) for simplicity. This choice allows us to simplify (3.4) into

$$F(\rho_1, \dots, \rho_n) = C_{\Omega} \left(\frac{\rho_1 + 1}{2}, \dots, \frac{\rho_n + 1}{2} \right).$$

With (3.5), we further have

$$F(\rho_1, \dots, \rho_n) = \Phi_{\Omega} \left(\Phi^{-1}\left(\frac{\rho_1 + 1}{2}\right), \dots, \Phi^{-1}\left(\frac{\rho_n + 1}{2}\right) \right). \quad (3.6)$$

Then the joint prior of ρ_1, \dots, ρ_n is the density function corresponding to the joint CDF in (3.6). Given that the density of the Gaussian copula in (3.5) is

$$c_{\Omega}(u_1, \dots, u_n) = \frac{\phi_{\Omega}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n))}{\prod_{i=1}^n \phi(\Phi^{-1}(u_i))},$$

where $\phi_{\Omega}(\cdot)$ is the density of $\Phi_{\Omega}(\cdot)$ and $\phi(\cdot)$ is the density of $\Phi(\cdot)$, we obtain the prior for ρ_1, \dots, ρ_n as:

$$\pi(\rho_1, \dots, \rho_n) = \frac{\phi_{\Omega} \left(\Phi^{-1}\left(\frac{\rho_1+1}{2}\right), \dots, \Phi^{-1}\left(\frac{\rho_n+1}{2}\right) \right)}{\prod_{i=1}^n \frac{1}{2} \phi \left(\Phi^{-1}\left(\frac{\rho_i+1}{2}\right) \right)}. \quad (3.7)$$

In practice, we may replace 1 with 0.9999 and 2 with 1.9998 in (3.7) to avoid the singularity issue at the boundary of ρ_i 's domain. The marginal distribution for ρ_i can be assumed to be any other distribution defined on (-1,1), for example, a truncated normal distribution. However, in such cases a more complex form of the prior $\pi(\cdot)$ will be expected which may significantly increase the computational burden. Given no a priori information for ρ_i the uniform distribution seems to be a reasonable choice.

Examples of using copula-based methods in a geostatistical setting can be seen in Bárdossy (2006) and Kazianka and Pilz (2010). We choose the covariance matrix Ω

in the Gaussian copula model to follow the structure of the Leroux et al. (1999) CAR model with variance τ_ρ^2 and spatial correlation parameter λ_ρ :

$$\Omega = \tau_\rho^2[(1 - \lambda_\rho)I + \lambda_\rho R]^{-1}, \quad (3.8)$$

where R denotes a neighborhood matrix with the i^{th} diagonal element as the total number of neighbors for county i while the $(i, j)^{th}$ off-diagonal element being -1 if counties i and j share a border and 0 otherwise. In Model (3.8), a smaller/larger λ_ρ indicates weaker/stronger spatial correlation with $\lambda_\rho = 0$ corresponding to complete spatial independence and $\lambda_\rho = 1$ to the ICAR model.

Note Model (3.2) can also be extended to include a second order autoregressive term if necessary, although this extension is not appropriate for our data due to the very few time points available. With a spatially varying AR(2) model, we will have both the first and second order coefficients, $\boldsymbol{\rho}_1 = \{\rho_{11}, \dots, \rho_{1n}\}$ and $\boldsymbol{\rho}_2 = \{\rho_{21}, \dots, \rho_{2n}\}$, at n spatial locations. Taking into account the stationarity constraints on each pair of (ρ_{1k}, ρ_{2k}) for $k = 1, \dots, n$, i.e., $|\rho_{2k}| < 1$ and $|\rho_{1k}| < 1 - \rho_{2k}$, we can model the prior for $(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2)$ jointly using Gaussian copulas, $C_{\Omega_1}(\cdot)$ and $C_{\Omega_2}(\cdot)$, which depend on covariance matrices Ω_1 and Ω_2 respectively. We assume marginals $\rho_{2k} \sim \text{uniform}(-1, 1)$ and $\rho_{1k}|\rho_{2k} \sim \text{uniform}(\rho_{2k} - 1, 1 - \rho_{2k})$ for simplicity. The joint prior can then be specified as

$$\pi(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) = \pi(\boldsymbol{\rho}_1|\boldsymbol{\rho}_2)\pi(\boldsymbol{\rho}_2),$$

where by (3.7),

$$\pi(\boldsymbol{\rho}_2) = \frac{\phi_{\Omega_2}(\Phi^{-1}(\frac{\rho_{21}+1}{2}), \dots, \Phi^{-1}(\frac{\rho_{2n}+1}{2}))}{\prod_{i=1}^n \frac{1}{2}\phi(\Phi^{-1}(\frac{\rho_{2i}+1}{2}))},$$

and

$$\pi(\boldsymbol{\rho}_1|\boldsymbol{\rho}_2) \sim \frac{\phi_{\Omega_1}(\Phi^{-1}(\frac{\rho_{11}+1-\rho_{21}}{2-2\rho_{21}}), \dots, \Phi^{-1}(\frac{\rho_{1n}+1-\rho_{2n}}{2-2\rho_{2n}}))}{\prod_{i=1}^n \frac{1}{2}\phi(\Phi^{-1}(\frac{\rho_{1i}+1-\rho_{2i}}{2-2\rho_{2i}}))}.$$

3.3 Other competing models

Most models developed for spatio-temporal disease aim at describing the disease pattern and their dependency structure rather than at prediction. For example, the spatio-temporal models in Knorr-Held (2000) as well as the models reviewed in López-Quílez and Muñoz (2009) are all proposed as generalized linear mixed models for the purpose of disease modeling. These models mainly differ from our models in the construction of $\eta_{i,t}$ specified in (3.2). Below lists eight different forms of $\eta_{i,t}$ that will be compared to our SVAR models:

1. $X_{i,t-1}^T \boldsymbol{\beta}$,
2. $X_{i,t-1}^T \boldsymbol{\beta} + \phi_i$,
3. $X_{i,t-1}^T \boldsymbol{\beta} + \alpha_t$,
4. $X_{i,t-1}^T \boldsymbol{\beta} + \delta_{i,t}$,
5. $X_{i,t-1}^T \boldsymbol{\beta} + \alpha_t + \phi_i$,
6. $X_{i,t-1}^T \boldsymbol{\beta} + \alpha_t + \phi_i + \delta_{i,t}$,
7. $X_{i,t-1}^T \boldsymbol{\beta} + \psi_{i,t} \rho (Z_{i,t-1} - X_{i,t-2}^T \boldsymbol{\beta})$,
8. $X_{i,t-1}^T \boldsymbol{\beta} + \psi_{i,t} \rho (Z_{i,t-1} - X_{i,t-2}^T \boldsymbol{\beta}) + \phi_i$,

where $X_{i,t-k}^T \boldsymbol{\beta}$, $k = 1, 2$ is the same as that in Model (3.2), ϕ_i denotes the spatial random effect of county i , α_t the temporal random effect of year t , and $\delta_{i,t}$ the space-time interaction effect. Models 2 through 6 are constructed by adapting Knorr-Held (2000)'s framework to our prediction priority and to make them more comparable to our proposed models. Specifically, we replace its temporal random walk prior with an AR(1) structure and its spatial effects prior following Besag et al. (1991) with that of Leroux et al. (1999). These modifications add more flexibility to the models as the strength of the temporal and spatial correlations in the data can be directly

inferred by the correlation parameters in α_t and ϕ_i , respectively. Models 7 and 8 are specified as invariant autoregressive models across spatial locations, resembling those in Martínez-Beneito et al. (2008).

We specify the priors for $\boldsymbol{\phi} = (\phi_1, \dots, \phi_n)^\top$, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_T)^\top$ and $\boldsymbol{\delta} = (\delta_{11}, \dots, \delta_{nT})^\top$ as the following:

$$\boldsymbol{\phi} \sim N(0, \Sigma_\phi), \quad \boldsymbol{\alpha} \sim N(0, \Sigma_\alpha), \quad \boldsymbol{\delta} \sim N(0, \Sigma_\delta),$$

where Σ_ϕ follows the structure of the Leroux et al. (1999) CAR model:

$$\Sigma_\phi = \tau_\phi^2 [(1 - \lambda_\phi)I + \lambda_\phi R]^{-1},$$

where τ_ϕ^2 denotes the effect variance and $\lambda_\phi \in [0, 1]$ the strength of the spatial correlation. The covariance matrix $\Sigma_\alpha = \tau_\alpha^2 \rho_\alpha^D$ is assumed to follow an AR(1) structure with variance parameter τ_α^2 , temporal correlation parameter ρ_α and $T \times T$ temporal distance matrix D . We specify Σ_δ as the Kronecker product of AR(1) temporal correlation and Leroux et al. (1999) spatial correlation structures:

$$\Sigma_\delta = \tau_\delta^2 \rho_\delta^D \otimes [(1 - \lambda_\delta)I + \lambda_\delta R]^{-1},$$

where τ_δ^2 is the variance parameter, ρ_δ and λ_δ denote the temporal and spatial correlation parameters respectively. This indicates a separable space-time covariance structure. A small caveat for Models 4 and 6 is that if ρ_δ and λ_δ are both zero, then $\delta_{i,t}$ is indistinguishable from $\epsilon_{i,t}$. Thus, in implementing Models 4 and 6, estimates of ρ_δ and λ_δ should be carefully monitored.

3.4 Hyperpriors

The temporal and spatial correlation structures of the random effects are defined through hyperparameters which will be estimated via MCMC sampling. The variance parameters, σ^2 , τ_α^2 , τ_ϕ^2 , τ_δ^2 , and τ_ρ^2 are all given a semi-conjugate inverse gamma

hyperprior $IG(a, b)$ and can be sampled via the Gibbs sampler. The correlation parameters λ_ϕ , λ_δ , and λ_ρ are given uniform(0,1) hyperpriors while ρ_α is given a uniform(-1,1) hyperpriors and are each sampled via the Metropolis-Hastings algorithm. For models that include the spatio-temporal random effects, $\delta_{i,t}$, such as Models 4 and 6, a stronger prior for the temporal correlation, ρ_δ , has to be specified to attain convergence of the posterior samples. Specifically, a uniform(0, 0.9) prior is used for ρ_δ as opposed to the uniform(-1, 1) used for ρ_α . This is possibly due to the short time length of the data which makes it challenging to estimate the space-time covariance structure.

Following Bernardinelli et al. (1995), we use more informative hyperpriors for the variance parameters, which will help with the MCMC convergence in the case of sparse data. We performed small experiments to evaluate the sensitivity of the model fitting to the choice of hyperprior parameters using all three datasets studied here. We indeed found the results to be insensitive to the choice of hyperprior parameters for the New England dataset, the largest dataset among the three, whereas the more informative hyperpriors yield slightly better prediction results than the less informative ones for the smaller datasets of Florida and California. Details on the sensitivity analysis can be found in the Supplement Section 4. Bernardinelli et al. (1995) also demonstrated the same conclusion on the sensitivity to hyperpriors, and more discussion on the choice of hyperprior parameters is in Bernardinelli et al. (1995) and Best et al. (1999).

3.5 Prediction

Given the fitted model, it is straightforward to make predictions for the following year. The prediction, $\widehat{Z}_{i,t+1}$, is obtained by sampling from the posterior predictive distribution through forward sampling. For example, with Model (3.3) we have for

each iteration of the MCMC algorithm,

$$\widehat{Z}_{i,t+1} = X_{i,t}^T \widehat{\boldsymbol{\beta}} + \widehat{\rho}_i (Z_{i,t} - X_{i,t-1}^T \widehat{\boldsymbol{\beta}}) + \widehat{\phi}_i + \widehat{\epsilon}_{i,t+1}, \quad \widehat{\epsilon}_{i,t+1} \sim N(0, \widehat{\sigma}^2 q_{it}).$$

Note that for year $t + 1$ we do not expect to know $q_{i(t+1)}$ so we assume $q_{i(t+1)} = q_{it}$ and hence $\psi_{i,t+1} = 1$, due to the relatively slow change over time of q_{it} at each county. This is verified in Figure 5 of the Supplement. For Models 3, 5, and 6, which include temporal and spatio-temporal random effects, we first use forward sampling to sample $\widehat{\alpha}_{t+1}$ and $\widehat{\delta}_{i,t+1}$ before computing $\widehat{Z}_{i,t+1}$. For instance, predictions for Model 6 have the form:

$$\widehat{Z}_{i,t+1} = X_{i,t}^T \widehat{\boldsymbol{\beta}} + \widehat{\alpha}_{t+1} + \widehat{\phi}_i + \widehat{\delta}_{i,t+1} + \widehat{\epsilon}_{i,t+1},$$

where $\widehat{\epsilon}_{i,t+1} \sim N(0, \widehat{\sigma}^2 q_{i(t+1)})$, $\widehat{\alpha}_{t+1} = \widehat{\rho}_\alpha \widehat{\alpha}_t + \widehat{\epsilon}_{t+1}^\alpha$ where $\widehat{\epsilon}_{t+1}^\alpha \stackrel{iid}{\sim} N(0, \widehat{\tau}_\alpha^2 (1 - \widehat{\rho}_\alpha^2))$, and $\widehat{\delta}_{i,t+1} = \widehat{\rho}_\delta \delta_{i,t} + \widehat{\epsilon}_{i,t+1}^\delta$ where $(\widehat{\epsilon}_{1,t+1}^\delta, \dots, \widehat{\epsilon}_{n,t+1}^\delta)^\top \sim N(0, \widehat{\tau}_\delta^2 (1 - \widehat{\rho}_\delta^2) [(1 - \widehat{\lambda}_\delta) I_n + \widehat{\lambda}_\delta R]^{-1})$. Finally, posterior predictions in the original scale, $\widehat{Y}_{i,t+1}$, are obtained by adding back the detrended global mean of $\widehat{Z}_{i,t+1}$ and then taking the exponential.

4 Prediction of HIV data

For each HIV dataset of Florida, California and New England states as described in Section 2, we first compare Models (3.2) and (3.3) to all eight models in Section 3.3 by withholding the new HIV diagnosis rates in 2014 as testing data and using the rates from 2008-2013 as training data. We then predict the new HIV diagnosis rates in 2015 using the rates observed from 2008-2014 and based on our best performing models.

4.1 Model assessment measures

We employ the mean squared prediction error (MSPE), the Gneiting and Raftery (2007) adaptation of the predictive model choice criterion (PMCC) of Gelfand and

Ghosh (1998), and the continuous rank probability score (CRPS) (Gneiting and Raftery, 2007) to evaluate the prediction performance. Let \mathbf{y}_{obs} be the 2008-2013 observed HIV, let y_k , $k = 1, \dots, p$ be the observed county level HIV in 2014 where $p = 47, 33$ and 112 for Florida, California and New England states, respectively, and let Y_k , $k = 1, \dots, p$ denote the random variable of the HIV in 2014. We have

$$MSPE = \frac{1}{p} \sum_{k=1}^p (\hat{y}_k - y_k)^2,$$

$$PMCC = \sum_{k=1}^p \frac{(\hat{y}_k - y_k)^2}{Var(Y_k | \mathbf{y}_{obs}, \hat{\theta})} + \sum_{k=1}^p \log[Var(Y_k | \mathbf{y}_{obs}, \hat{\theta})],$$

and

$$CRPS = \frac{1}{p} \sum_{k=1}^p (E_F |Y_k - y_k| - \frac{1}{2} E_F |Y_k - Y'_k|),$$

where Y_k and Y'_k in CRPS are independent replicates sampled from the posterior predictive distribution F . The latter two measures directly take advantage of the posterior predictive distribution resulting from MCMC sampling. The PMCC is a balanced loss function composed of two parts; the first part is interpreted as a measure of the goodness of fit and the second as a predictive variance penalty term. CRPS assesses how concentrated the predictive distribution is around the observed values. As opposed to MSPE and PMCC measures, the CRPS evaluates the models based on the entire predictive distribution rather than only its first two moments. Models which minimize these measures are preferred. In addition to these three measures, we also compare the empirical coverage probability (ECP) at a 95% nominal level of all models. ECP evaluates the uncertainty quantification of the posterior predictive distribution, and is expected to be close to the nominal level if the uncertainty is correctly quantified. Finally, a baseline one-step ahead prediction in time series is to use the observation of the previous time point as the prediction. We therefore compare the performance of all predictions to this baseline prediction.

4.2 Model comparison

Table 2 reports the predictions for the 2014 new HIV diagnosis rates using all 10 models over the three regions. All predictions are obtained by taking the posterior median of the sampling chain of $\widehat{Y}_{i,t+1}$, to limit the influence of big values resulted from taking exponentiation of $\widehat{Z}_{i,t+1}$. Model (3.2) stands out in terms of CRPS and PMCC among the 10 models for all three datasets. In terms of MSPE, we again see that (3.2) is selected as the best model for California and New England. However, Model (3.2) yields a large MSPE for Florida. For this region, Model (3.3) offers the smallest MSPE and the second smallest PMCC and CRPS which nevertheless are comparable to their smallest respective values. Moreover, all ECPs provided by Models (3.2) and (3.3) are comparable to the nominal level.

Overall, Model (3.2) appears to provide the best prediction for California and New England while Model (3.3) provides the best prediction for Florida. A more careful investigation shows that the Florida data itself exhibits spatial correlation although no strong correlation is present in ρ_i , therefore modeling the spatial correlation or spillover effect only through ρ_i as in Model (3.2) is likely insufficient. In such case, Model (3.3) which includes spatial random effects to capture the spatial variability arising from unknown factors is more coherent to the data. Figure 3 shows the posterior means of ρ_i from Models (3.2) for each region. It exhibits a somewhat similar pattern but with different values as in Figure 2. Figure 4 shows the predicted HIV new diagnosis rates in 2014 based on our best performing statistical models, i.e., Model (3.2) for California and New England and Model (3.3) for Florida, paired with their corresponding observations. Each pair of observations and predictions is seen in high agreement.

Although the model comparison varies slightly for different regions, some consistent patterns emerge. Comparing the results of Model 1 and Model 3 we can see

that for these three datasets the temporal random effects have little effect to improve the data modeling and thus prediction. This implies that the annual average of adjusted HIV new diagnosis rates by taking covariates into account contributes very little to explain the data variation. This is verified by the small magnitude of the posterior estimates of $\alpha_1, \dots, \alpha_T$ that is shown in the Supplement Figure 6. The autocorrelation of adjusted HIV at each individual county is however nonnegligible. In model 8, the estimates for ρ are around 0.23, 0.18, and 0.17 for Florida, California, and New England, respectively. The estimates for ρ_i in Models (3.2) and (3.3) are also significantly different from zero at many counties, see Figure 3. Including the autocorrelation, even assumed fixed as in Model 7, helps to improve the prediction compared to Models 1 and 3.

In contrast to temporal random effects, the spatial random effects seem to help greatly with the prediction performance. This is reflected by the relatively good performance of Models 2, 5, 6 and 8 all of which exclusively include the spatial random effects. This suggests that spatial correlation in the data should be taken into account. Furthermore, the fashion in which the spatial dependence is introduced within the model also clearly makes a difference on prediction performance. Based on Table 2, modeling the spatial correlation through autoregressive coefficients, as in our SVAR Models, offer the best predictions. The significant advantage of allowing for spatially varying autoregressive coefficients is directly shown by comparing the SVAR models to Models 7 and 8 which have invariant autoregressive coefficients. The spatio-temporal random effects in Models 4 and 6 seem to offer no help for modeling these data sets. Indeed the inclusion of this random effect can even hurt the prediction power by comparing Model 5 and Model 6. This could be due to a possible overparameterization of Model 6 for these datasets.

To further illustrate the role of the spatially varying autoregressive function in

the prediction of HIV, Figure 5 shows a breakdown of the contribution from each of the individual terms, $X\beta$ and $\rho_i(Z_{i,t-1} - X_{i,t-2}\beta)$, in Model (3.2) by plotting their respective posterior means based on the California data. California was chosen for illustration due to the small number of counties it has which provides a better visual. The highly varying pattern of the component attributed to $\rho_i(Z_{i,t-1} - X_{i,t-2}\beta)$ shows that the spatially varying autoregressive function makes an important contribution in capturing the variability of the HIV data. If the autoregressive term were insignificant we would expect to see a straight line of zeroes or a curve that is very close to the zero line. Figure 5 shows that the curve of $\rho_i(Z_{i,t-1} - X_{i,t-2}\beta)$ is clearly non-negligible and is indeed helpful to adjust the effect of $X\beta$ in many counties, such as counties 6031 and 6065.

4.3 Exchangeable prior for ρ_i

Other covariance structure in priors of ρ_i may further improve the performance of our SVAR models depending on the data properties. As suggested by a reviewer, we also implemented an exchangeable prior for ρ_i , i.e., all diagonal entries of the neighborhood matrix R used in Leroux et al. (1999) are set to be $n - 1$ while all off-diagonal entries are -1 . This corresponds to a simpler neighborhood structure compared to the neighborhood matrix defined in Section 3.2, as the exchangeable prior considers all counties to be neighbors.

The prediction results using the exchangeable prior are reported in the last two rows of Table 2. Although neither of those two rows uniformly outperforms the other models, we indeed observe some advantages of using an exchangeable prior. Some notable advantages can be seen in the improved performance of (3.3) and (3.2) for Florida and California respectively in terms of MSPE and CRPS when using exchangeable priors. Further investigation of the results indicates that in the case of

moderately sized data with relatively low evidence of spatial correlation among neighbors, as with Florida and California datasets, the simpler covariance structure of an exchangeable prior reduces the volatility of the ρ_i estimates and improves predictions that largely overestimate the observations. Figure 3 and Figure 4 compare the estimates of ρ_i and predicted new HIV diagnosis rates in 2014 using exchangeable priors to those using the classic neighborhood structure defined in Section 3.2, respectively. The exchangeable prior obviously makes the estimates of ρ_i smoother, particularly for Florida and New England, although the two sets of predictions are hardly distinguishable in visualization except for a few counties.

4.4 Prediction of 2015 new HIV diagnoses

Due to the sluggish procedure of the public reporting of new HIV diagnosis data, the 2014 HIV data was not released until July 2016. Thus it will take at least until July 2017 to access the 2015 HIV data. We make predictions for the 2015 new HIV diagnoses using Model (3.3)^e for Florida, (3.2)^e for California and (3.2) for New England based on the the annual HIV data collected from 2008 to 2014 and the available covariate information between 2007 to 2014. The website *healthindicators.gov* where we retrieve many of the covariates has paused their data collection service in June 2016 due to budget cuts, having only collected data through 2013, and it is unclear when they would resume the service. This leads to incomplete covariate information for 2014. We used 2014 variables whenever they are available and an average of previous years data to replace the missing 2014 variables. The influence of the missing variables should be minimal. On the one hand, we find that the covariates are relatively stable over time and on the other hand, only four variables of 2014, two pertaining to weather, excessive drinking and inactivity, are missing from our selected set of 28 variables. Our predictions can provide a timely surrogate for the true data, and thus

are crucial for identifying HIV hotspots and making decisions on prevention, testing and treatment efforts.

Figure 6 shows the predicted 2015 new HIV diagnosis rates and their corresponding 95% credible interval widths in log scale. Counties with no previous observed data, e.g, Alpine County, CA which additionally has little to no observed data in its 5 neighboring counties as well, exhibit large credible widths. Figure 6 shows that the counties that are expected to have the highest rates, $Y_{i,t}$, in 2015 include: Union, FL (140.1), Baltimore City, MD (78.2), Hamilton, FL (76.1), Prince Georges, MD (63.0), Essex, NJ (61.4), Bronx, NY (59.1), Washington, FL (58.8), Broward, FL (57.7), Miami-Dade, FL (55.7), New York, NY (55.3), and San Francisco, CA (50.7). The numbers in parentheses are our predicted new HIV diagnosis rates for that county. We find that these counties tend to have a higher percentages of African Americans, people in the age range of 25-44, and people living in same sex households than other counties. It is also seen that Florida in general has higher predicted new diagnosis rates than the other two regions, which is an indication of severe HIV infection in the South.

5 Discussion

Spatial correlation arises often because observations of geographic proximity tend to be similar, and it usually decays as observations become further apart. With aggregated count data (areal data) in epidemiology, spatial correlation is most often present when regions of interest share borders which enables certain type of communications between neighbors. Moreover, Besag (1974) highlighted some important environmental causes that are inherently spatially structured and thus can induce spatial structuring in the response as well. Ignoring spatial autocorrelation in the

analysis may lead to erroneous conclusion (e.g., Kühn, 2007) and the importance of incorporating spatial autocorrelation has been realized in many disciplines such as ecology and epidemiology (Bahn et al., 2006; Auchincloss et al., 2012). In particular, Auchincloss et al. (2012) pointed out that considering spatial correlation is likely to improve specificity of exposure and disease relationships, reduce measurement error, and deepen our understanding of the relationships between location and health. Besides the methods reviewed and described in this paper, Banerjee et al. (2015) and references therein provide a rich source for modeling spatial disease.

To make predictions for spatially aggregated data over a fixed number of counties, it is natural to investigate the characteristics of each individual time series, so having an individual autoregressive model for each county immediately follows. However, due to the spatial correlation of time series at adjacent counties and the potential necessity to borrow strength from neighbors in model fitting when each series is short in length, it is beneficial to jointly model the parameters of each time series. For these reasons, we proposed our SVAR models, which are shown to be useful for prediction and can be advantageous over the linear mixed models that have been popular for spatio-temporal disease data modeling. In our models, the space-time interaction is incorporated in a unique fashion by allowing for spatially varying autoregressive coefficients. This procedure is more convenient than including a nonseparable spatio-temporally correlated random effect in the model as the nonseparable space-time covariance function is often difficult to model and to manipulate in computation.

The feature of our SVAR models that allows the temporal autocorrelation coefficients to vary spatially while still remaining parsimonious is particularly important for the HIV datasets analyzed in this article, which have very few observed time points at each county and thus cannot afford a separate model fitting for each county. We compared our models to eight competing models and found that our SVAR models

perform the best when jointly viewing their MSPE, PMCC, CRPS, and ECP based on new HIV diagnoses rates in Florida, California and New England states. In the geostatistics setting, Nobre et al. (2011) proposed a spatially varying autoregressive model of order p to model space-time continuous processes of sea surface temperatures, where a kernel convolution of bounded variables, particularly beta random variables, are proposed as priors for the bounded process of autoregression coefficients. Our copula model for the prior of autoregressive coefficients is more computationally efficient than the kernel convolution and can be generalized to any non-Gaussian prior.

The form of the neighborhood matrix R in the CAR model is determined by the underlying lattice structure of the data. Given no a priori knowledge on the underlying structure of the data, the simple 0-1 adjacency rule for neighborhood matrix as implemented in this paper is commonly used (Wall, 2004). Whereas, in particular cases where the data consists of highly irregular lattices, Clayton and Bernardinelli (1992) and Stern and Cressie (2000) suggested a weighted scheme for elements of R . These weights can be further determined based on the distance between areas or the similarities between key covariates (Earnest et al., 2007).

There are many ways to expand our SVAR models to make them more flexible and better accommodate different spatio-temporal processes. For example, the AR(1) model can be replaced by a higher order autoregressive model where coefficients at each order are allowed to vary spatially. We take a log transformation to attain the Gaussian properties of our data. However, we are aware that such transformation may not be appropriate for other data sets. Our model can easily be adapted to accommodate other types of transformation such as the square root transformation, Freeman-Tukey transformation of the form $\sqrt{x} + \sqrt{x+1}$ (Waller et al., 1997), or other frameworks such as specifying $Y_{i,t}$ to follow a Poisson or binomial distribution.

In the latter case, we model $Y_{i,t} \sim \text{Poisson}(E_{i,t}r_{i,t})$ or $Y_{i,t} \sim \text{binomial}(n_{i,t}, p_{i,t})$, where $E_{i,t}$ and $r_{i,t}$ are the expected and relative rates respectively and $n_{i,t}$ and $p_{i,t}$ are the population at risk and the risk rate for the i^{th} county at time t . We then model $\log(r_{i,t})$ or $\text{logit}(p_{i,t})$ in a similar fashion as $Z_{i,t}$ in this paper.

Finally, we focus on developing an effective statistical modeling framework for HIV prediction based on the observed data. The autoregressive nature of our models makes it challenging to perform prediction for counties with suppressed data in previous years. Although it is well-known that the missing data not at random can cause bias in the parameter estimation (Little and Rubin, 2014), its effect on predictions made using our modeling framework is not very clear. A study of how to appropriately incorporate information from the missing data to better inform the prediction is currently underway.

Acknowledgements

The authors thank the editor, the associate editor, and the referees for constructive suggestions that have improved the content and presentation of this article. The authors also thank Sophie Lohmann and Alex Morales for helpful discussions on this work. This research was partially supported by NIH grant R56 and NSF grants AGS-1602845.

References

- Albarracín, D., M. B. Tannenbaum, L. R. Glasman, and A. J. Rothman (2010). Modeling structural, dyadic, and individual factors: the inclusion and exclusion model of hiv related behavior. *AIDS and Behaviors* 14(S2), 239–249.

- Assunção, R. M., I. A. Reis, and C. D. Oliveira (2001). Diffusion and prediction of leishmaniasis in a large metropolitan area in Brazil with a Bayesian space-time model. *Statistics in Medicine* 20(15), 2319–2335.
- Auchincloss, A. H., S. Y. Gebreab, C. Mair, and A. V. D. Roux (2012). A review of spatial methods in epidemiology 2000-2010. *Annual review of public health* 33, 107–122.
- Bahn, V., R. J. O’Connor, and W. B. Krohn (2006). Importance of spatial autocorrelation in modeling bird distributions at a continental scale. *Ecography* 29, 835–844.
- Balaji, A. B., K. E. Bowles, B. C. Le, and G. Paz-Bailey and A M Oster (2013). High HIV incidence and prevalence and associated factors among young MSM. *AIDS* 27(2), 269–278.
- Banerjee, S., B. P. Carlin, and A. E. Gelfand (2015). Bayesian inference. In F. Buena, V. Isham, N. Keiding, T. Louis, R. L. Smith, and H. Tong (Eds.), *Hierarchical modeling and analysis for spatial data, 2nd edition*, pp. 100–107. CRC Press.
- Bárdossy, A. (2006). Copula-based geostatistical models for groundwater quality parameters. *Water Resources Research* 42(11). W11416.
- Bauer, C., J. Wakefield, H. Rue, S. Self, Z. Feng, and Y. Wang (2016). Bayesian penalized spline models for the analysis of spatio-temporal count data. *Statistics in Medicine* 35(11), 1848–1865.
- Bernardinelli, L., D. Clayton, and C. Montomoli (1995). Bayesian estimates of disease maps: how important are priors? *Statistics in Medicine* 14, 2411–2431.

- Bernardinelli, L., D. G. Clayton, C. Pascutto, C. Montomoli, M. Ghislandi, and M. Songini (1995). Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine* 14(21-22), 2433–2443.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B* 36(2), 192–236.
- Besag, J., J. York, and A. Mollié (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 43(1), 1–59.
- Best, N. G., A. A. Richard, A. Thomas, L. A. Waller, and E. Conlon (1999). Bayesian disease models for spatially correlated disease and exposure data. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Spatial Modeling*. Oxford University Press.
- Breslow, N. E. and D. G. Clayton (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 88(421), 9–25.
- Cai, B., A. B. Lawson, M. M. Hossain, J. Choi, R. S. Kirby, and J. Liu (2014). Bayesian semiparametric model with spatially-temporally varying coefficients selection. *Statistics in Medicine* 32(21), 3670–3685.
- Clayton, D. and L. Bernardinelli (1992). *Bayesian methods for mapping disease risk*, pp. 205–220.
- Cressie, N. (1993). *Statistics for Spatial Data*. USA: Wiley Series in Probability and Mathematical Statistics.
- Earnest, A., G. Morgan, K. Mengersen, L. Ryan, R. Summerhayes, and J. Beard (2007). Evaluating the effect of neighbourhood weight matrices on smoothing prop-

- erties of conditional autoregressive (car) models. *International Journal of Health Geographics* 6, 54.
- Gelfand, A. E. and S. K. Ghosh (1998). Model choice: a minimum posterior predictive loss approach. *Biometrika* 85(1), 1–11.
- Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association* 97(458), 590–600.
- Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* 102(477), 359–378.
- Kazianka, H. and J. Pilz (2010). Copula-based geostatistical modeling of continuous and discrete data including covariates. *Stochastic Environmental Research and Risk Assessment* 24, 661–673.
- Knorr-Held, L. (2000). Bayesian modeling of inseparable space-time variation in disease risk. *Statistics in Medicine* 19, 2555–2567.
- Kühn, I. (2007). Incorporating spatial autocorrelation may invert observed patterns. *Diversity and Distributions* 13, 66–69.
- Lagazio, C., A. Biggeri, and E. Dreassi (2003). Age-period-cohort models and disease mapping. *Environmetrics* 14(5), 475–490.
- Lee, D. (2011). A comparison of conditional autoregressive models used in bayesian disease mapping. *Spatial and Spatio-temporal Epidemiology* 2, 79–89.
- Leroux, B. G., X. Lei, and N. Breslow (1999). Estimation of disease rates in small areas: a new mixed model for spatial dependence. In M. Halloran and D. Berry (Eds.), *Statistical models in epidemiology, the environment and clinical trials*. New York: Springer-Verlag.

- Little, R. J. A. and D. B. Rubin (2014). *Statistical analysis with missing data*. Hoboken, NJ, USA: John Wiley & Sons, Inc.
- López-Quílez, A. and F. Muñoz (2009). Review of spatio-temporal models for disease mapping. Technical report, Departament d'Estadística, Universitat de València.
- MacNab, Y. C. (2003). Hierarchical bayesian modeling of spatially correlated health service outcome and utilization rates. *Biometrics* 59, 305–316.
- MacNab, Y. C. and C. B. Dean (2001). Autoregressive spatial smoothing and temporal spline smoothing for mapping rates. *Biometrics* 57(3), 949–956.
- MacNab, Y. C. and C. B. Dean (2002). Spatio-temporal modelling of rates for the construction of disease maps. *Statistics in Medicine* 21(3), 347–358.
- Martínez-Beneito, M. A., A. López-Quílez, and P. Botella-Rocamora (2008). An autoregressive approach to spatio-temporal disease mapping. *Statistics in Medicine* 27(15), 2874–2889.
- Mercer, L. D., J. Wakefield, A. Pantazis, A. M. Lutambi, H. Mansanja, and S. Clark (2015). Space-time smoothing of complex survey data: small area estimation for child mortality. *Annals of Applied Statistics* 9(4), 1889–1905.
- Nobre, A., B. Sansó, and A. Schmidt (2011). Spatially varying autoregressive processes. *Technometrics* 53, 310–321.
- Nobre, A. A., a M Schmidt, and H. F. Lopes (2005). Spatio-temporal models for mapping the incidence of malaria in par . *Environmetrics* 16(3), 291–304.
- Ruiz-Medina, M. D., R. M. Espejo, M. D. Ugarte, and A. F. Militino (2014). Functional time series analysis of spatio-temporal epidemiological data. *Stochastic Environmental Research and Risk Assessment* 28(4), 943–954.

- Schmid, V. and L. Held (2004). Bayesian extrapolation of space-time trends in cancer registry data. *Biometrics* 60(4), 1034–1042.
- Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B* 64(4), 583–639.
- Stern, H. S. and N. Cressie (2000). Posterior predictive model checks for disease mapping models. *Statistics in Medicine* 19, 2377–2397.
- Sun, D., R. K. Tsutakawa, H. Kim, and Z. He (2000). Spatio-temporal interaction with disease mapping. *Statistics in Medicine* 19(15), 2015–2035.
- Ugarte, M. D., T. Goicoa, J. Etxeberria, and A. F. Militino (2012). A p-spline anova type model in space-time disease mapping. *Stochastic Environmental Research and Risk Assessment* 26, 835–845.
- Wall, M. M. (2004). A close look at the spatial structure implied by the car and sar models. *Journal of the American Statistical Association* 121, 311–324.
- Waller, L. A., B. P. Carlin, H. Xia, and A. E. Gelfand (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association* 92, 607–617.
- Xia, H. and B. P. Carlin (1998). Spatio-temporal models with errors in covariates: mapping ohio lung cancer mortality. *Statistics in Medicine* 17(18), 2025–2043.

Table 1: Test Statistics and p-values for Moran’s I and Geary’s C on testing the null hypothesis of no spatial correlation among ρ_i versus the one-sided alternative hypothesis of positive spatial correlation.

	Florida		California		New England	
	Statistic	p-value	Statistic	p-value	Statistic	p-value
Moran’s I	0.0343	0.3216	0.1072	0.1389	0.2598	0.0003
Geary’s C	0.9610	0.3758	0.8098	0.0665	0.7384	0.0005

Table 2: Prediction results of 2014 new HIV diagnosis rates for Florida, California, and New England.

Model	Florida				California				New England			
	MSPE	PMCC	CRPS	ECP	MSPE	PMCC	CRPS	ECP	MSPE	PMCC	CRPS	ECP
(3.2)	75.30	221.8	3.905	0.9574	7.231	94.84	1.456	0.9697	45.88	393.7	2.557	0.9196
(3.3)	56.89	223.0	3.906	0.9574	9.715	98.11	1.621	0.9697	66.43	399.3	2.949	0.9107
1	78.26	242.9	4.387	0.9787	10.42	113.9	1.843	0.9697	67.47	442.7	3.041	0.9821
2	58.04	242.0	4.152	1.0000	8.397	138.5	2.092	1.0000	61.41	416.3	2.751	0.9643
3	80.29	245.1	4.367	0.9362	10.28	106.4	1.748	0.9394	65.11	442.2	2.983	0.9018
4	74.91	275.2	4.650	1.0000	9.457	125.6	1.919	0.9697	64.09	455.9	2.958	0.9821
5	60.62	231.7	3.990	0.9787	7.951	116.2	1.801	1.0000	55.10	385.3	2.605	0.9018
6	67.73	240.5	4.223	1.0000	8.031	107.2	1.685	0.9697	63.31	411.1	2.722	0.929
7	76.86	230.8	4.106	0.9362	9.297	107.8	1.704	0.9697	58.65	436.5	2.838	0.9821
8	65.67	226.8	3.922	0.9574	8.289	107.9	1.646	0.9697	57.78	398.4	2.717	0.9286
Y_{t-1}^*	61.51	–	–	–	10.03	–	–	–	62.58	–	–	–
(3.2) ^e	77.91	224.2	3.988	0.9574	<i>6.875</i>	96.14	<i>1.446</i>	0.9697	49.01	398.7	2.595	0.9286
(3.3) ^e	<i>53.72</i>	221.9	<i>3.843</i>	0.9362	7.244	<i>93.30</i>	1.453	0.970	55.14	<i>387.1</i>	2.754	0.9107

Y_{t-1}^* indicates the baseline prediction for each county, and ^e indicates that an exchangeable prior was used for ρ_i . The numbers in bold text are the smallest numbers of their respective columns excluding the last two rows. The numbers in italic in the last two rows are the smallest numbers of their respective columns.

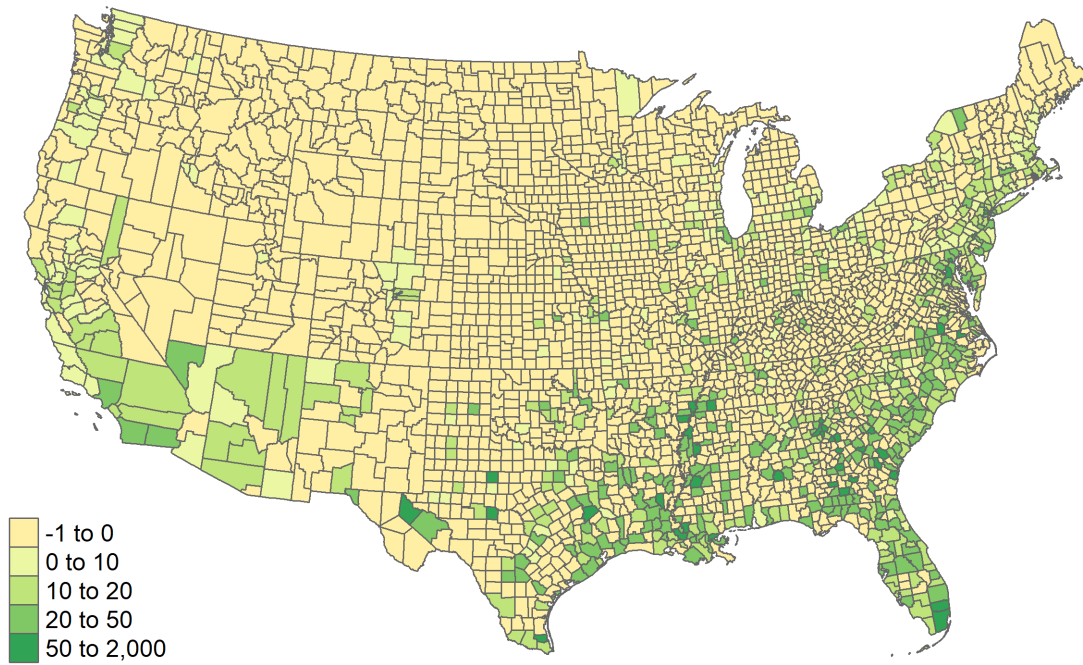


Figure 1: 2012 new HIV diagnosis rates in cases per 100,000 across the United States.

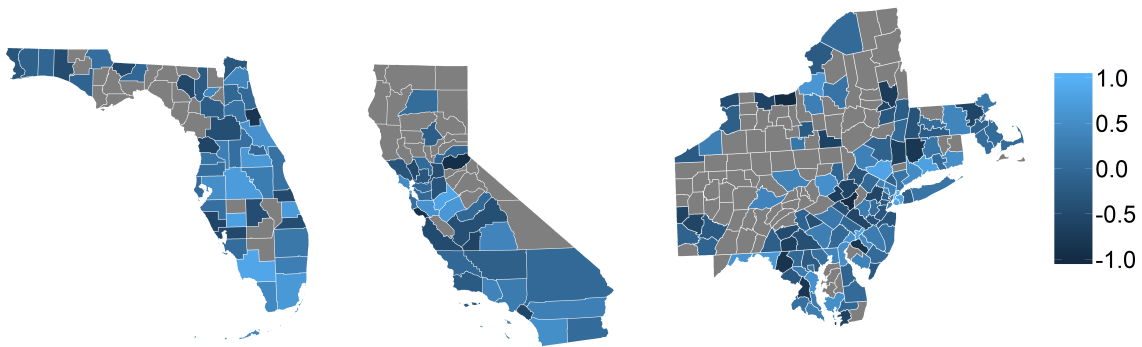


Figure 2: Spatial maps of independent ρ estimates using maximum likelihood estimation for counties in Florida, California and New England.

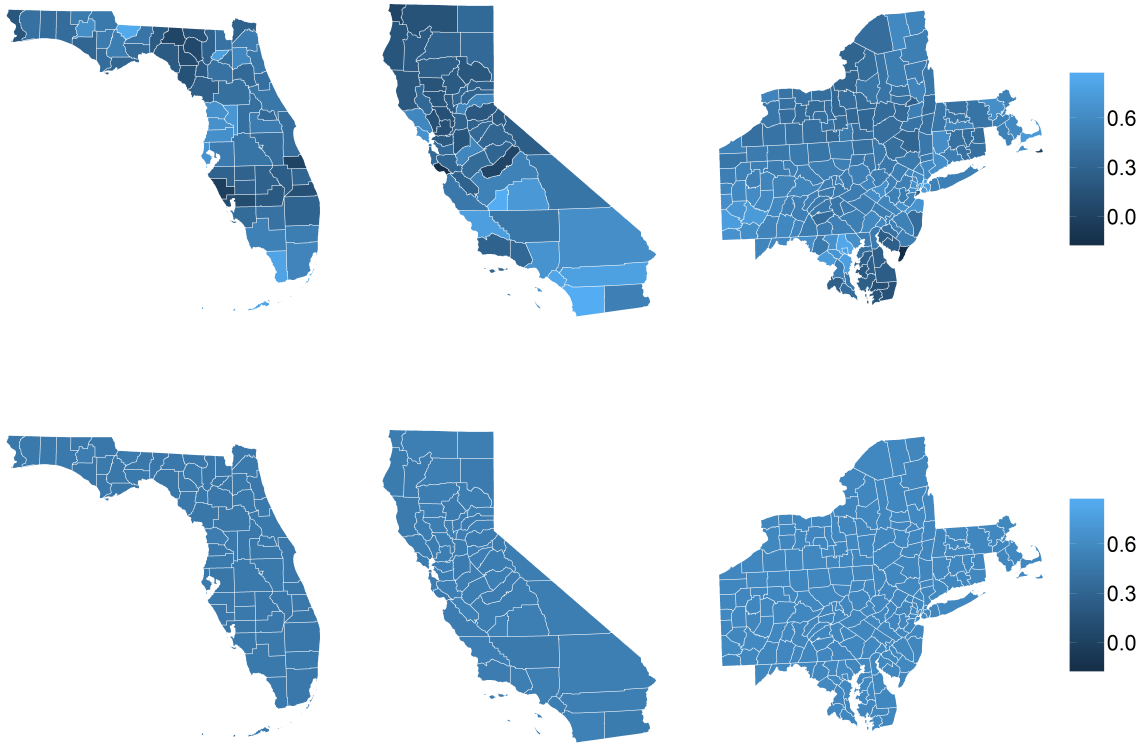


Figure 3: Spatial maps of posterior means of ρ_i from Model (3.2) (top row) and (3.2)^e (bottom row) for Florida, California and New England.

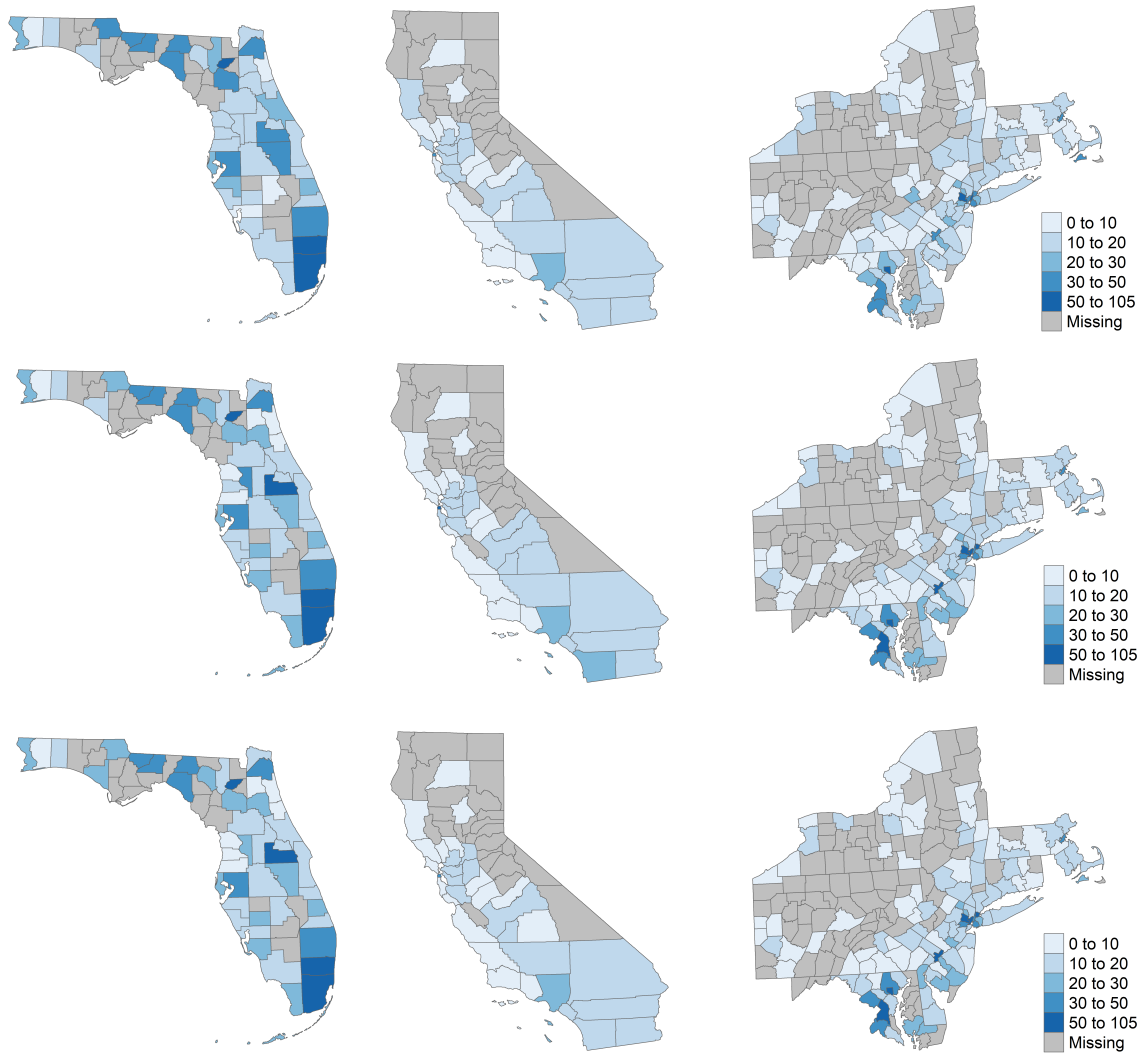


Figure 4: Observed (top row) and predicted (middle and bottom rows) new HIV diagnosis rates in 2014 for Florida, California and New England. Middle row is obtained using priors with neighborhood structure and bottom row is obtained using exchangeable priors.

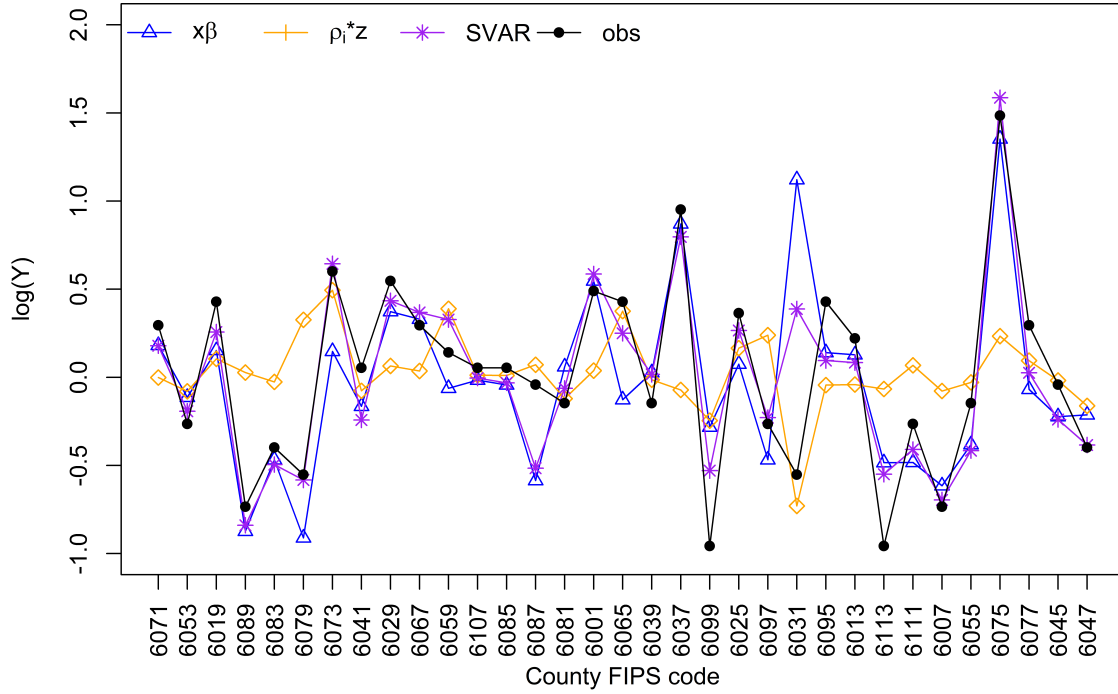


Figure 5: Break down of contribution from each term in Model (3.2) to the prediction of 2014 new HIV diagnosis rates in California, where ρ_i^*Z indicates the contribution from $\rho_i(Z_{i,t-1} - X_{i,t-2}\beta)$ and SVAR indicates the overall model prediction.

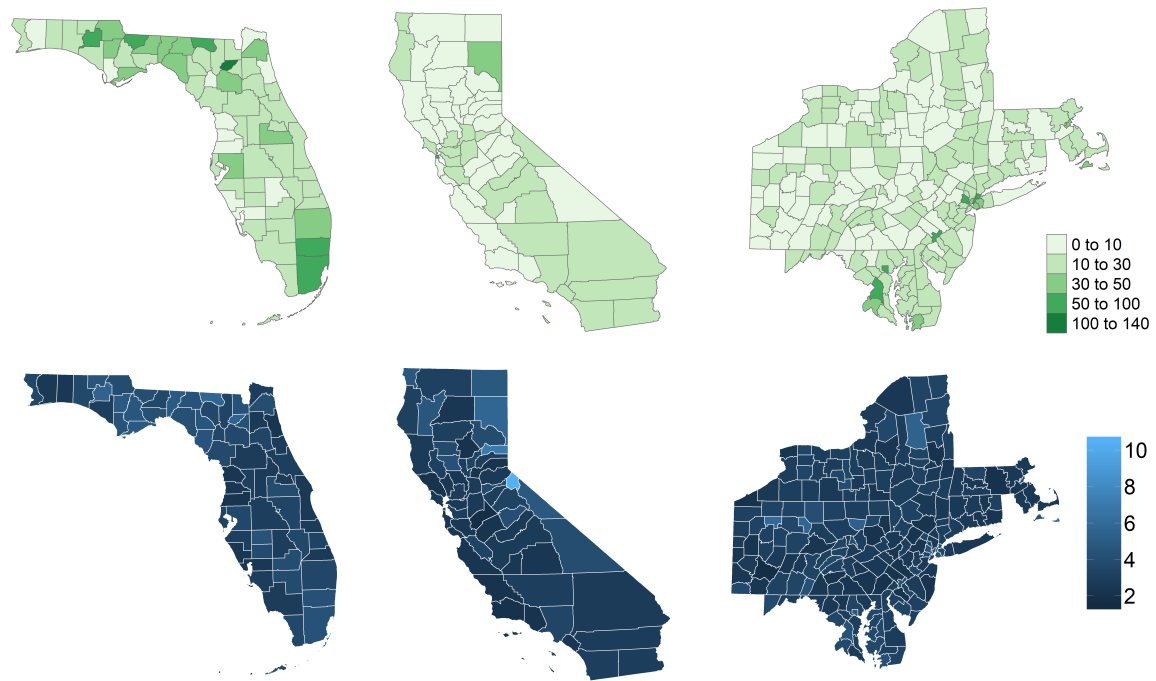


Figure 6: Predicted 2015 new HIV diagnosis rates (top row) and their corresponding 95% credible interval widths in log scale (bottom row) for Florida, California and New England using $(3.3)^e$, $(3.2)^e$ and (3.2) for each state respectively.