Article

A Comparison of Bayesian Spatial Models for HIV Mapping in South Africa

Kassahun Abere Ayalew [1],*, Samuel Manda [1,2,3] and Bo Cai [4]

[1] School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg 3209, South Africa; Samuel.Manda@mrc.ac.za
[2] Biostatistics Unit, South African Medical Research Council, Pretoria 0001, South Africa
[3] Department of Statistics, University of Pretoria, Pretoria 0028, South Africa
[4] Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SC 29208, USA; BOCAI@mailbox.sc.edu
* Correspondence: kassahun.abere@gmail.com

Abstract: Despite making significant progress in tackling its HIV epidemic, South Africa, with 7.7 million people living with HIV, still has the biggest HIV epidemic in the world. The Government, in collaboration with developmental partners and agencies, has been strengthening its responses to the HIV epidemic to better target the delivery of HIV care, treatment strategies and prevention services. Population-based household HIV surveys have, over time, contributed to the country's efforts in monitoring and understanding the magnitude and heterogeneity of the HIV epidemic. Local-level monitoring of progress made against HIV and AIDS is increasingly needed for decision making. Previous studies have provided evidence of substantial subnational variation in the HIV epidemic. Using HIV prevalence data from the 2016 South African Demographic and Health Survey, we compare three spatial smoothing models, namely, the intrinsically conditionally autoregressive normal, Laplace and skew-t (ICAR-normal, ICAR-Laplace and ICAR-skew-t) in the estimation of the HIV prevalence across 52 districts in South Africa. The parameters of the resulting models are estimated using Bayesian approaches. The skewness parameter for the ICAR-skew-t model was not statistically significant, suggesting the absence of skewness in the HIV prevalence data. Based on the deviance information criterion (DIC) model selection, the ICAR-normal and ICAR-Laplace had DIC values of 291.3 and 315, respectively, which were lower than that of the ICAR-skewed t (348.1). However, based on the model adequacy criterion using the conditional predictive ordinates (CPO), the ICAR-skew-t distribution had the lowest CPO value. Thus, the ICAR-skew-t was the best spatial smoothing model for the estimation of HIV prevalence in our study.

Keywords: Bayesian; disease mapping; skew-t distribution; ICAR-normal; ICAR-Laplace; spatial random effects; spatial model

1. Introduction

Governments in sub-Saharan Africa (SSA), in collaboration with non-governmental organizations and private sectors, design national strategic plans and policies, allocate resources and implement programs in the fight against the HIV/AIDS epidemic [1,2]. Such efforts are designed to reduce HIV-related infection, morbidity and mortality. As well as understanding the level of the HIV epidemic at the national level, most governments in the region have implemented a decentralized approach to governance and service provision. Thus the need for reliable local (district)-level HIV statistics to support decision making regarding the delivery of HIV care, treatment and prevention services [3,4]. Most of the countries in SSA rely on data obtained from national HIV surveys for monitoring the level of the HIV epidemic and subsequence responses. However, the national HIV surveys are mostly empowered to produce reliable HIV estimates at national and provincial level. Crude HIV estimates at small area level could be exaggeratedly estimated due to

Int. J. Environ. Res. Public Health **2021**, *18*, 11215

2 of 10

small numbers, resulting in unstable variances [5–7]. Consequently, HIV prevention and treatment programs tailored to small areas could be based on unreliable evidence [8].

As a result, modelling approaches are used for generating local-level estimates from survey data that are originally meant to provide reliable estimates at national and provincial levels [9,10]. The most used approach has been using spatial smoothing models where spatial components are incorporated in the model as random effects. The spatial models produce reliable disease rates with improved accuracy for small areas with few sparse observations by incorporating information from local, spatially contiguous areas. The structured random effect in spatial models represents clustering of diseases over geographical areas, unobserved environmental or frailty factors which are spatially correlated but are not included as covariates in a model [11–13]. Structured spatial random effects (which consider the local effects) are mostly modelled using the intrinsic conditional autoregressive normal (ICAR-normal) model (Besag et al [13], Carlin and Banerjee [14]). The ICAR-normal model offers greater flexibility for modelling the spatial correlation than the linear mixed effects model, with only a global random effect. However, a normal spatial distribution on the structured spatial effect could be restrictive, as there could be a possibility that the normality assumption could be misspecified [15]. Misspecification of the distribution of the random effects may result in estimates of diseases rates that are biased [16,17]. The usual approach is to transform the data to normality, for example by performing a logarithm of the rates. However, if there was an appropriate theoretical model, transformation could be avoided, as it is difficult to interpret results from transformed data. In addition, the transformation could result in the loss of information [17].

A few approaches have been proposed to reduce the impact of a normal distribution assumption for spatial random components. For example, Lunn et al [18] and Manda [19] proposed a double exponential and a mixture of ICAR-normal and ICAR-double exponential, respectively, to better capture possible wider tails for the spatial random effects. Kim and Mallick [20] and Azzalini and Capitanio [21] considered a skew-normal spatial model for point referenced data. However, the structured spatial skewed random fields suffer identifiability problems (since the skewness parameter may be unknown) [22] and must be determined uniquely [23]. To solve this identifiability problems, Zhang and El-Shaarawi [24] defined a skewed stationary Gaussian process for spatial random effect based on the work by Azzalini and Capitanio [21]. In addition, Allard and Naveau [25] and Zareifard and Jafari Khaledi [26] introduced a skew-normal spatial random field based on Domınguez-Molina et al [27] and Palacios and Steel [28], respectively, for point referenced data. Other skewed spatial distributions are the skew-normal by Rantini et al [29] and Fernández and Steel [30].

Our aim, in this study, is to model the district-level HIV prevalence in South Africa using spatial smoothing methods. There is ample evidence of substantial small area variation in the distribution of HIV prevalence in Sub-Saharan Africa [31,32]. Similarly evidence has also been found in South Africa by Kim et al [33] and Gutreuter et al [34]. The distribution of the district HIV prevalence could be skewed and non-normal. Thus, we estimated the spatial distribution of the HIV prevalence among the districts in South Africa using the ICAR-normal [13], ICAR skew-t distribution (Nathoo and Ghosh [35]) and ICAR-Laplace [18] using the 2016 South African Demographic and Health Survey data. The next section presents the description of the spatial models used and the HIV data. Section 3 contains the results obtained from fitting the models to the data. We discuss the results in Section 4 and conclude in Section 5.

2. Methods and Data Source

2.1. Skew-t Spatial Random Effects Distribution

Let Y_i be the number of HIV positive individuals out of a sample of size n_i in district i $(i = 1, \ldots, 52)$. Both Y_i and n_i are adjusted to account for the survey design to become the effective number of HIV cases, Y_i^*, and the effective sample size, n_i^* [35–38].

Int. J. Environ. Res. Public Health **2021**, 18, 11215

3 of 10

A three-stage Bayesian hierarchical spatial smoothing model for a binary HIV outcome uses a binomial distribution at stage one as

$$Y_i^* | p_i \sim Binomial(n_i^*, p_i), \ i = 1, \ldots, 52$$

where p_i is the proportion (prevalence) of HIV in district i and is modelled at the second stage by a logit link function using a set of district-level predictor variables, X_i, and both unstructured and spatially structured random effects, as introduced by Besag et al. (1991).

$$\log\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + X_i\beta + u_i + v_i$$

where β_0 is the intercept; β is a vector of regression coefficients for predictor variable in X_i; u_i is the unstructured random component and it is assumed to follow a normal distribution, $u_i \sim N(0, \sigma_u^2)$; v_i is the structured spatial random component for district i.

The structured spatial random effects could be modelled using an intrinsic conditional autoregressive normal (ICAR-normal) prior (Besag et al [13], Knorr-Held and Best [12] and Carlin and Banerjee [14]) as

$$v_i | v_{-i} \sim ICARN\left(\mu_v, \sigma_v^2\right) = N\left(\frac{\Sigma_{j\sim i}v}{m_i}, \frac{\sigma_v^2}{m_i}\right)$$

where m_i is the number of neighbours of district i. Lunn et al [18] suggested an alternative model based on a Laplace/double exponential distribution (ICAR-Laplace), which is given as $u_i \sim ICARL(\mu_u, \sigma_u^2)$.

However, in situations where the distribution of HIV prevalence data could be non-normal and asymmetric, alternative spatial smoothing models that are robust and flexible could fit the data better. As a result, Nathoo and Ghosh [35] suggested the skew-t (ICAR-skew-t) spatial smoothing model, defined as

$$v_i | v_{-i} \sim ST_v\left(\frac{\Sigma_{j\sim i}v_j}{m_i}, \frac{\sigma_v^2}{m_i}, \delta_v\right)$$

For easy implementation in most Bayesian statistical software, Sahu et al [39] presented a suitable representation of skew-t distribution with k degrees of freedom. Suppose $y \sim skew - t \ (k)$, then it could be expressed as $y = \eta^{-\frac{1}{2}}(\Delta|X_0| + X)$, where $X_0 \sim N(0,1)$, $X \sim N(\mu, \sigma^2)$, Δ is the skewness parameter and $\eta \sim gamma\left(\frac{k}{2}, \frac{k}{2}\right)$. The hierarchical set-up of this stochastic representation can be given as $Y/w \sim N\left(\mu + \Delta w, \frac{\Sigma}{\eta}\right)$, where $|X_0| = w \sim N(0, I_k)I(w > 0)$. Thus, the ICAR-skew-t for the structured spatial random effect can be expressed as

$$v_i \sim N\left(\frac{\Sigma_{j\sim i}s_j}{m_i} + \delta_v w_i, \frac{\sigma_s^2}{\eta * m_i}\right)$$

where $w_i \sim N(0, I)I(w_i > 0)$, $s_{i/S_{-i}} \sim N\left(\frac{\Sigma_{j\sim i}s_j}{m_i}, \frac{\sigma_s^2}{m_i}\right)$ and σ_s^2 and δ_v are the variance of s_i and the skewness parameter, respectively. The hierarchical representation of the ICAR-skew-t model is shown in the Appendix A.

2.2. Methods for Comparing Competing Models

In this study, we used the deviance information criterion (DIC) and conditional predictive ordinates (CPO) for comparing models. The deviance information criterion was developed by Spiegelhalter et al [40] as a method used for comparing models in a Bayesian framework. It is a measure of a model's goodness of fit or adequacy adjusted for a measure

Int. J. Environ. Res. Public Health **2021**, *18*, 11215

4 of 10

of model complexity measured as effective number of parameters. Let θ and $y = y_1, \ldots, y_1$ be the model parameter and data respectively, then DIC is expressed as

$$DIC = \overline{D} + p_D = 2\overline{D} - D(\overline{\theta})$$

where $\overline{D} = E_{\theta/y}[D(\theta)] = E_{\theta/y}[-2\log p(y/\theta)]$ and is the posterior mean deviance that measures the goodness of fit or adequacy $p_D = \overline{D} - D(\overline{\theta}) = E_{\theta/y}[D(\theta)] - D(E_{\theta/y}[\theta]) = E_{\theta/y}[-2\log p(y/\theta)] - [-2\log p(y/\overline{\theta}(y)]$ is a measure of the effective number of parameters and measures model complexity; larger values of p_D suggests higher complexity of the model. It is also defined as the difference between the posterior mean of the deviance and the deviance at the posterior means of the parameters of interest; in other words, it is considered as the expected excess of the true residuals over the estimated residuals in the data conditional on the parameter θ [16]. Let $\theta^1, \ldots, \theta^k$ be parameter estimates from a converged Markov chain, then $\overline{D}$ is estimated as $\frac{1}{k} \sum_1^k D(\theta^k)$ and $D(\overline{\theta}) = D(\frac{1}{k} \sum_1^k \theta^k)$.

The CPO is a leave-one-out cross validation approach that measures the posterior probability of observing y_i when the model is fitted to all data excluding y_i and it measures the predictive ability of the fitted model. Let $Y = Y_1, Y_2, \ldots, Y_n$ be the $nX1$ data vector and Y_{-i} be the data vector without y_i. Then, the conditional predictive ordinate for observation y_i is given as

$$CPO_i = f(y_i/\mathbf{y}_{-i}) = \int f(y_i/\theta) P(\theta/y_{-i}) d\theta = E_{\theta/y}\left[\frac{1}{f(y_i/\theta)}\right]$$

where θ is the parameter vector, y_i is the ith observation and y_{-i} is the observed data set except y_i. Thus, one can estimate the value of the inverse of CPO_i by averaging the inverse probability function evaluated at y_i for each θ^k produced from the posterior density. The CPO_i values could be easily determined from the standard MCMC output which is given as

$$CPO_i = \left[\frac{1}{k} \sum_{k=1}^{K} \frac{1}{f(y_i/\theta^k)}\right]^{-1}$$

which is the harmonic mean of the probability density function evaluated at y_i for each θ^k, where K is the number of iterations. For discrete data, the comparison of CPO_i with the relative frequency determined from data without y_i (y_{-i}) enables the assessment of the predictive capacity of the fitted model to the data. In order to compare two or more competing models, the overall CPO values of each model are assessed, given as $CPO = \prod_i CPO_i$; A model with higher CPO value suggests better predictive performance than the other models; hence, this model is preferred over other models. Mostly, the CPO value is close to zero, thus the negative of the sum of the log of the CPO_i is used as indicated by Cai et al [41] and is given by $LS_{cv} = -\sum_{i=1}^{k} logCPO_i$. Thus, a model with the lowest LS_{cv} value is the best model in terms of its predictive capacity.

2.3. Implementation

The model parameters were determined using a Bayesian estimation approach via Markov Chain Monte Carlo (MCMC) as implemented in OpenBUGS [42]. The prior distributions for the regression coefficients and the unstructured random component were the same for all the three models. The prior distribution for the intercept was $\beta_0 \sim$ uniform on $(-\infty, \infty)$ and the prior for the regression coefficients was $\beta_q \sim N(0, 0.00001)$, where $q = 1, 2, 3, 4$; the variance parameters σ_u^2 and σ_v^2 were given as inverse gamma prior distributions with shape and scale parameters set at 20 and 2000, respectively. The skewness parameters for ICAR-skew-t were assigned $\delta_v \sim N(0, 0.01)$ prior. We conducted a sensitivity analysis to determine the impact of the hyper-parameters of the priors on the outcome variable; for this, we chose the most commonly used hyper-parameters, such as $IG(1000, 1000)$, $IG(10, 10)$, $IG(1, 10)$ and $IG(2, 2000)$. Since prior distributions with larger variances are considered in the model, the estimates from this analysis are expected

to be relatively robust. Moran's I test was conducted on the model residuals to determine the presence of spatial correlation [43]. We ran 100,000 iterations for each model to make inferences. We determined the number of initial iterations that needed to be discarded by assessing the history plots of each model and for each parameter. Similarly, we also investigated the autocorrelation plots of each model and each parameter to determine the selection intervals to avoid correlation problems in the generated chains.

2.4. Data

The data analyzed were obtained from the 2016 South African Demographic and Health Survey (SADHS 2016). The SADHS 2016 was conducted for evaluating the country's health programs by monitoring key milestones such as mortality, fertility, maternal and child health, nutrition, HIV, gender-based violence, etc. The data for measuring these indicators were collected by asking respondents relevant sociodemographic and behavioral characteristic questions and by collecting biological specimens. The SADHS 2016 survey employed a multistage stratified cluster sampling design to select households and/or respondents for the sample. All women between the age of 15 and 49 and men between the ages of 15 and 59 were included in the survey. Interview data were collected from a total of 8514 women and 3618 men and 6912 individuals were tested for HIV seropositivity. More information about SADHS 2016 can be obtained from the full study report [44].

The observed district-level HIV prevalence was computed by taking the survey design into account. The effective sample sizes in each district was determined by dividing the observed number of sample size at each district by the design effect [36]; the effective number of HIV cases is thus the product of effective sample size and the weighted prevalence. The number of HIV tests conducted in the survey by district varied substantially, with a sample size of between 8 tests and 455 tests, with a median sample size of 111 tests. There were some districts with zero count of HIV positive individuals in the sample. For this, we assigned them the average of the simulated data from a normal distribution with mean value equal to the average of the log of prevalence in the neighboring districts and variance as the variance of the log of the prevalence p_i calculated from all the neighboring districts divided by the number of neighbors, shown in Figure 1b; the map in Figure 1a shows the raw data not adjusted for zero positive cases. A skewness test was conducted on the prevalence, with and without adjusting for zero HIV prevalence, but no significant skewness was found.
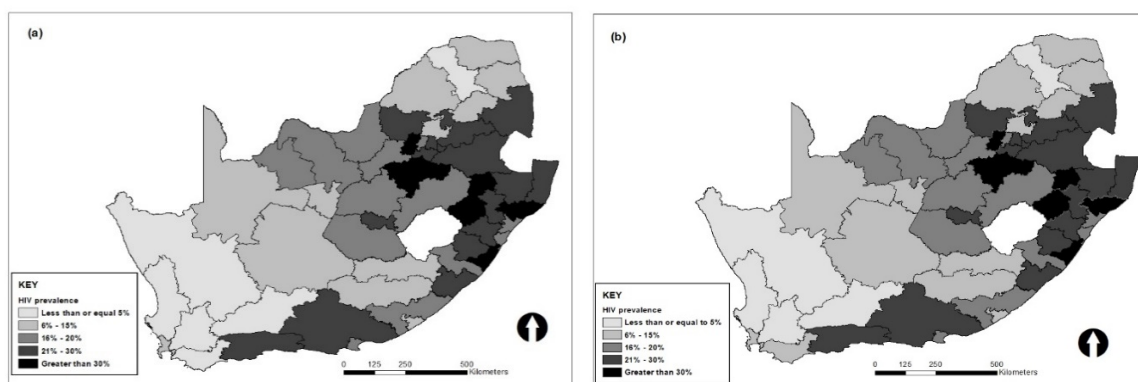
Figure 1. Map of HIV prevalence by district in South Africa before (**a**) and after (**b**) adjusting the data for zero positive tests in some districts.

The covariates included in the models are the multidimensional poverty index constructed using the 2016 community survey data [45], HIV prevalence among pregnant women obtained from the 2017 National Antenatal Sentinel Survey report [46], population density and male condom distribution coverage obtained from the 2017 district health barometer report [47]. Previous studies indicate that these factors are associated with HIV prevalence ecologically as well as individually [3,48].

3. Results

The skewness parameters for ICAR-skew-t were not significant, perhaps suggesting that the spatial component is lighter tailed (see Table 1). The model with the lowest LS_{cv} and DIC values was deemed to be the best model in its predictive performance and goodness of fit, respectively. Thus, as can be seen in Table 1, the model with the lowest LS_{cv} (170.5) is the ICAR-skew-t model, followed by the ICAR-normal model (LS_{cv}= 172.4). The ICAR-normal model and the ICAR-Laplace model have lowest (291.3) and second lowest (315) DIC values, respectively. The difference in the DIC values between these models is more than five, suggesting that there is substantial difference between the two models in terms of goodness of fit to the data, according Spiegelhalter et al [40]; however, a study by De la Cruz and Branco [49] indicated that DIC is not appropriate for such type of complex models. Thus, based on the LS_{cv} values, the ICAR-skew-t model was the best in terms of its predicative capacity as compared to the other two models used in this study.

Table 1. Comparison of the fitted models using DIC and CPO.

Covariates	ICAR-Normal	ICAR-Laplace	ICAR-Skew-t
Intercept	2.473 (−3.288, −1.65)	−2.542 (−3.321, −1.743)	−2.538 (−3.625, −1.469)
Population density	−0.0001 (−0.0003, 0.0002)	−0.0001 (−0.0003, 0.0002)	0.0001 (−0.0003, 0.0002)
Male condom distribution	−0.0070 (−0.0183, 0.0069)	−0.0064 (−0.0178, 0.0039)	−0.0069 (−0.0177, 0.0032)
Multidimensional poverty index	0.81056 (−2.826, 4.7939)	0.593 (−3.139, 4.357)	0.8934 (−2.915, 4.71)
ANC HIV prevalence	3.778 (1.673, 5.7058)	3.974 (2.074, 5.897)	3.831 (1.7, 5.931)
σ_v^2	0.0061 (0.0006, 0.6596)	0.0059 (0.0006, 0.9225)	0.0088 (0.0009, 0.4719)
σ_u^2	0.0066 (0.0007, 0.2281)	0.0106 (0.0011, 0.2434)	0.0031 (0.0004, 0.1688)
δ_u			0.05 (−0.6, 0.62)
DIC	291.3	315	348.1
LS_{cv}	172.4	174	170.5

As a sensitivity analysis, we ran the analysis using different sets of hyper-parameters for priors of the precision parameters. Thus, the mean difference in the values of the outcome variables at different choices of hyper-parameter values was observed at the third digit after the decimal point, which suggests the absence of a significant impact on the outcome variable. The Moran's I test statistic was significant (*p*-value = 0.000001), suggesting that residuals were spatially clustered. As shown in Table 1, district-level ANC prevalence is the strong predictor of district-level HIV prevalence determined from the 2016 SADHS data, whereas the other covariates were not statistically significant.

Figure 2e, shows the prevalence of HIV by district in South Africa estimated using the ICAR-skew-t spatial model (best model). According to the estimates from this model, most of the districts with high levels of HIV prevalence are located in southeastern parts of the country, while low levels of HIV prevalence are in the southwestern parts. This pattern is the same for all the maps produced using estimates from different models with or without covariates. Maps (a), (c) and (e) are estimates of the ICAR-normal, ICAR-Laplace and skew-t models with covariates, respectively; the spatial pattern of HIV prevalence is the same for these models, except the estimate from the ICAR-normal model for one district in the northwestern part. Maps (b), (d) and (f) are estimates of the ICAR-normal, ICAR-Laplace and skew-t models without covariates and the pattern of HIV prevalence by district is the same for the estimates determined using these models. One notable difference for the pattern of estimates with and without covariates for the models is that the level of HIV prevalence is lower for estimates with covariates than those without covariates in two districts in the western part.
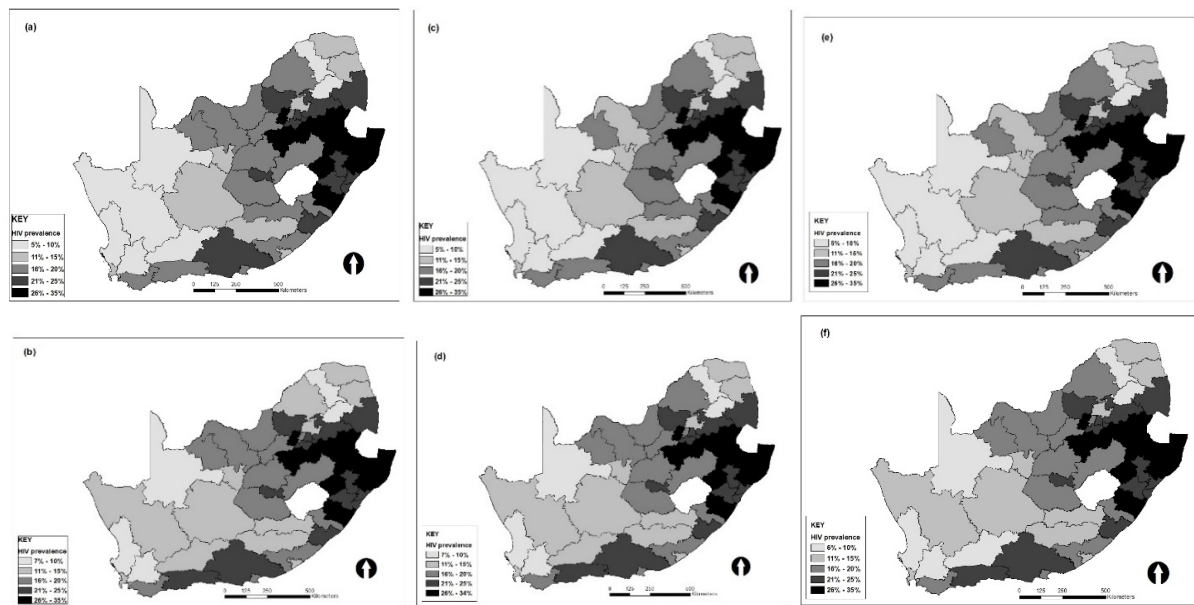
Figure 2. Estimated HIV prevalence by district in South Africa with covariates (first row **a**,**c**,**e**) and without covariates (second row **b**,**d**,**f**).

4. Discussion

HIV is a leading cause of disease burden in sub-Saharan Africa. In the era of decentralized approach to governance and service provision, designing effective HIV intervention programs and monitoring strategies at local administrative levels requires reliable estimates of local variation in HIV burden. Our study compared three spatial smoothing models, namely, the intrinsically conditionally autoregressive normal, Laplace and skew-t (ICAR-normal, ICAR-Laplace and ICAR-skew-t) in the estimation of the HIV prevalence across 52 districts in South Africa. It analyzed HIV prevalence data from the 2016 South African Demographic and Health Survey. The models were fitted using the Markov Chain Monte Carlo method in OpenBUGS, a freely available Bayesian statistical package. We found that the ICAR-skew-t distribution was the best spatial smoothing model for the estimation of HIV prevalence in our study.

We found that the districts with high levels of HIV prevalence were in the southeastern parts of the country, while low levels of HIV prevalence corresponded to the southwestern parts. Our findings are similar to those by Gutreuter et al [34] and Woldesenbet et al [46]. The estimates of HIV prevalence by district in South Africa could help governmental and non-governmental originations, as well as the private sector, to know the level of the epidemics at lower administrative level, thus prioritizing and plan appropriate public health programs tailored to each community and evaluating the combined impact of national and local public health programs.

A major weakness of our study could be that there were no HIV data in some of the sparsely populated districts; hence, we simulated data from neighboring districts to estimate prevalence of HIV in such districts; thus, the estimates for these districts may not be reliable and should be interpreted with caution. In addition, a limited number of predictors was included in the model; hence, some important predictors of district-level HIV prevalence might be missing.

5. Conclusions

In conclusion, alternative spatial distributions to ICAR-normal should be considered for modeling spatial disease outcomes. The spatial random effects could be skewed or non-normal and misspecification of the distribution of random effects could lead to estimates that are biased. This could lead to implications in the estimation of disease burden, adversely impacting policy derivations. In our study, we found that the intrinsic

Int. J. Environ. Res. Public Health **2021**, *18*, 11215

8 of 10

conditional autoregressive skew-t (ICAR-skew-t) model was the best in predicting district-level HIV prevalence compared to the ICAR-normal and ICAR-Laplace spatial models based on an analysis of the 2016 South African Demographic and Health Survey (2016 SADHS) data. District antennal clinic HIV prevalence was the most influential predictor of the district-level 2016 SADH HIV prevalence.

Appendix A

Hierarchical representation of the disease mapping model presented in Section 2.1, assuming the spatial random components follows skew-t distribution is given as follows

$$Y = Y_1, Y_2, \ldots, Y_n$$

be a one-dimensional random variable with binomial distribution

$$logit\ (p_i) = \beta_0 + X_i\beta + u_i + v_i$$

$$u_i \sim N\left(0,\ \sigma_u^2\right)$$

$$v_i / S_i, \sigma_v^2, \delta_v, w_i, \sim N\left(\frac{\Sigma_{j\sim i}u_j}{m_i} + \delta_u w_i,\ \frac{\sigma_s^2}{\eta * m_i}\right)$$

$$s_i / S_{-i} \sim N\left(\frac{\Sigma_{j\sim i}s_j}{m_i},\ \frac{\sigma_s^2}{m_i}\right)$$

$$w_i \sim N(0, I)I(w_i > 0)$$

$$\eta \sim gamma\left(\frac{k}{2}, \frac{k}{2}\right)$$

$$\beta_i \sim N(\beta_0, \Lambda),\ \mathrm{i} = 0, 1, 2, \ldots, \mathrm{k}$$

where *k* is the number of covariates

$$\sigma_v^2 \sim IG(\Omega, v)$$

$$\delta_u \sim N(0,\ \Gamma)$$

$$\sigma_s^2 \sim IG(\Omega, u)$$

$$k \sim Exp(k_0)I(k > 2)$$

where p_i is the weighted prevalence corresponding to Y_i $i = 1, 2, \ldots, 52$, σ_u^2 and σ_v^2 are variance of the spatial and the heterogeneous random component, $I(w_i > 0)$ is an indicator function, IG is inverse gamma and Exp is exponential.

Based on the likelihood distribution and the above prior specifications the posterior distribution of all the parameters assuming conditional independence between the response variable and the hyper parameters is given as

$$p\left(\mu, \beta,\ u, v, \sigma_u^2, \sigma_v^2, \delta_u, w, k, \eta, s/y\right) \propto L\left(y/\beta,\ u, v, \sigma_s^2, \sigma_v^2, \delta_u, w, s\right) P\left(\beta,\ u, v, \sigma_s^2, \sigma_v^2, \delta_u, w, k, \eta\right)$$
$$= \prod_i p(y_i/\mu_i) \prod_j (p(\beta_j/\Lambda)p(\Lambda)) p(u/\sigma_s^2)p(\sigma_s^2)p(v/\sigma_v^2)p(\sigma_v^2)p(s/\sigma_s^2)p(w)p(\delta_u)p(k)p(\eta)$$

References

1. UNAIDS. *2016–2021 Strategy on the Fast-Track to end AIDS*; UNAIDS: Geneva, Switzerland, 2015. Available online: https://www.unaids.org/sites/default/files/media_asset/20151027_UNAIDS_PCB37_15_18_EN_rev1.pdf (accessed on 1 August 2021).
2. PEPFAR. *PEPFAR 2021 Country and Regional Operational Plan (COP/ROP) Guidance for all PEPFAR Countries*; PEPFAR: Washington, WA, USA, 2021. Available online: https://www.state.gov/wp-content/uploads/2020/12/PEPFAR-COP21-Guidance-Final.pdf (accessed on 1 August 2021).
3. Manda, S.; Masenyetse, L.; Cai, B.; Meyer, R. Mapping HIV prevalence using population and antenatal sentinel-based HIV surveys: A multi-stage approach. *Popul. Health Metrics.* **2015**, *13*, 22. [CrossRef]
4. Larmarange, J. Evaluation of geospatial methods to generate subnational HIV prevalence estimates for local level planning. *AIDS* **2016**, *30*, 1467–1474.
5. Tanser, F.; Bärnighausen, T.; Cooke, G.; Newell, M.-L. Localized spatial clustering of HIV infections in a widely disseminated rural South African epidemic. *Int. J. Epidemiol.* **2009**, *38*, 1008–1016. [CrossRef] [PubMed]
6. Niragire, F.; Achia, T.; Lyambabaje, A.; Ntaganira, J. Bayesian Mapping of HIV Infection among Women of Reproductive Age in Rwanda. *PLoS ONE* **2015**, *10*, e0119944. [CrossRef] [PubMed]
7. Chimoyi, L.A.; Musenge, E. Spatial analysis of factors associated with HIV infection among young people in Uganda, 2011. *BMC Public Health* **2014**, *14*, 555. [CrossRef] [PubMed]
8. Houlihan, C.F.; Mutevedzi, P.C.; Lessells, R.J.; Cooke, G.S.; Tanser, F.C.; Newell, M.-L. The tuberculosis challenge in a rural South African HIV programme. *BMC Infect. Dis.* **2010**, *10*, 23–29. [CrossRef] [PubMed]
9. Johnson, G.D. Small area mapping of prostate cancer incidence in New York State (USA) using fully Bayesian hierarchical modelling. *Int. J. Health Geogr.* **2004**, *3*, 29. [CrossRef]
10. Leyland, A.H.; Langford, I.H.; Rasbash, J.; Goldstein, H. Multivariate spatial models for event data. *Stat. Med.* **2000**, *19*, 2469–2478. [CrossRef]
11. Lawson, A.B.; Browne, W.J.; Rodeiro, C.L.V. *Diease Mapping with WinBUGS and MLwiN*; Wiley & Sons: Chichester, UK, 2003.
12. Knorr-Held, L.; Best, N.G. A shared component model for detecting joint and selective clustering of two diseases. *J. R. Stat. Soc. Ser. A Stat. Soc.* **2001**, *164*, 73–85. [CrossRef]
13. Besag, J.; York, J.; Mollié, A. Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Stat. Math.* **1991**, *43*, 1–20. [CrossRef]
14. Carlin, B.; Banerjee, S. Hierarchical Multivariate CAR Models for Spatio-Temporally Correlated Survival Data. *Bayesian Stat.* **2003**, *7*, 45–63.
15. Arellano-Valle, R.; Bolfarine, H.; Lachos, V. Bayesian Inference for Skew-normal Linear Mixed Models. *J. Appl. Stat.* **2007**, *34*, 663–682. [CrossRef]
16. Ghosh, P.; Branco, M.D.; Chakraborty, H. Bivariate random effect model using skew-normal distribution with application to HIV-RNA. *Stat. Med.* **2007**, *26*, 1255–1267. [CrossRef]
17. Verbeke, G.; Lesaffre, E. A Linear Mixed-Effects Model with Heterogeneity in the Random-Effects Population. *J. Am. Stat. Assoc.* **1996**, *91*, 217–221. [CrossRef]
18. Lunn, D.; Jackson, C.; Best, N.; Thomas, A.; Spiegelhalter, D. *The BUGS Book: A Practical Introduction to Bayesian Analysis*; CRC: Boca Raton, FL, USA, 2012.
19. Manda, S.O.M. Macro Determinants of Geographical Variation in Childhood Survival in South Africa Using Flexible Spatial Mixture Models. In *Demographic Methods and Population Analysis*; Kandala, N.-B., Ghilagaber, G., Eds.; Springer: Dordrecht, The Netherlands, 2014.
20. Kim, H.-M.; Mallick, B.K. A Bayesian prediction using the skew Gaussian distribution. *J. Stat. Plan. Inference* **2004**, *120*, 85–101. [CrossRef]
21. Azzalini, A.; Capitanio, A. Statistical applications of the multivariate skew normal distribution. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1999**, *61*, 579–602. [CrossRef]
22. Genton, M.; Zhang, H. Identifiability problems in some non-Gaussian spatial random fields. *Chil. J. Stat.* **2012**, *3*, 171–179.
23. Gelfand, A.E.; Sahu, S.K. Identifiability, Improper Priors, and Gibbs Sampling for Generalized Linear Models. *J. Am. Stat. Assoc.* **1999**, *94*, 247–253. [CrossRef]
24. Zhang, H.; El-Shaarawi, A. On spatial ske—Gaussian processes and applications. *Environmetrics* **2009**, *21*, 33–47.
25. Allard, D.; Naveau, P. A New Spatial Skew-Normal Random Field Model. *Commun. Stat. Theory Methods* **2007**, *36*, 1821–1834. [CrossRef]

26. Zareifard, H.; Khaledi, M.J. Non-Gaussian modeling of spatial data using scale mixing of a unified skew Gaussian process. *J. Multivar. Anal.* **2013**, *114*, 16–28. [CrossRef]

27. Domınguez-Molina, J.; González-Farıas, G.; Gupta, A. *The Multivariate Closed Skew Normal Distribution*; Technical Report; Department of Mathematics and Statistics, Bowling Green State University: Bowling Green, OH, USA, 2003.

28. Palacios, M.B.; Steel, M.F.J. Non-Gaussian Bayesian Geostatistical Modeling. *J. Am. Stat. Assoc.* **2006**, *101*, 604–618. [CrossRef]

29. Rantini, D.; Iriawan, N.; Irhamah, I. Fernandez–Steel Skew Normal Conditional Autoregressive (FSSN CAR) Model in Stan for Spatial Data. *Symmetry* **2021**, *13*, 545. [CrossRef]

30. Fernández, C.; Steel, M.F.J. On Bayesian Modeling of Fat Tails and Skewness. *J. Am. Stat. Assoc.* **1998**, *93*, 359–371.

31. Dwyer-Lindgren, L.; Cork, M.A.; Sligar, A.; Steuben, K.M.; Wilson, K.F.; Provost, N.R.; Mayala, B.K.; Vander Heide, J.D.; Collison, M.L.; Hall, J.B.; et al. Mapping HIV prevalence in sub-Saharan Africa between 2000 and 2017. *Nat. Cell Biol.* **2019**, *570*, 189–193. [CrossRef]

32. Cuadros, D.F.; Abu-Raddad, L.J. Spatial variability in HIV prevalence declines in several countries in sub-Saharan Africa. *Health Place* **2014**, *28*, 45–49. [CrossRef] [PubMed]

33. Kim, H.; Tanser, F.; Tomita, A.; Vandormael, A.; Cuadros, D.F. Beyond HIV prevalence: Identifying people living with HIV within underserved areas in South Africa. *BMJ Glob. Health* **2021**, *6*, e004089. [CrossRef]

34. Gutreuter, S.; Igumbor, E.; Wabiri, N.; Desai, M.; Durand, L. Improving estimates of district HIV prevalence and burden in South Africa using small area estimation techniques. *PLoS ONE* **2019**, *14*, e0212445. [CrossRef]

35. Nathoo, F.S.; Ghosh, P. Skew-elliptical spatial random effect modeling for areal data with application to mapping health utilization rates. *Stat. Med.* **2013**, *32*, 290–306. [CrossRef]

36. Kish, L. Methods for Design Effects. *J. Off. Stat.* **1995**, *11*, 55–77.

37. Chen, C.; Wakefield, J.; Lumely, T. The use of sampling weights in Bayesian hierarchical models for small area estimation. *Spat. Spatio-Temporal Epidemiol.* **2014**, *11*, 33–43. [CrossRef]

38. Vandendijck, Y.; Faes, C.; Kirby, R.; Lawson, A.; Hens, N. Model-based inference for small area estimation with sampling weights. *Spat. Stat.* **2016**, *18*, 455–473. [CrossRef] [PubMed]

39. Sahu, S.K.; Dey, D.K.; Branco, M.D. A new class of multivariate skew distributions with applications to bayesian regression models. *Can. J. Stat.* **2003**, *31*, 129–150. [CrossRef]

40. Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P.; van der Linde, A. Bayesian measures of model complexity and fit (with discussion). *J. R. Stat. Soc. Ser. B* **2002**, *64*, 583–639. [CrossRef]

41. Cai, B.; Lawson, A.B.; Hossain, M.; Choi, J.; Kirby, R.S.; Liu, J. Bayesian semiparametric model with spatially-temporally varying coefficients selection. *Stat. Med.* **2013**, *32*, 3670–3685. [CrossRef]

42. Thomas, A.; Best, N.; Lunn, D. WinBUGS User Manual: Version 1.4. 2001. Available online: https://www.mrc-bsu.cam.ac.uk/wp-content/uploads/manual14.pdf (accessed on 1 August 2021).

43. Moran, P.A.P. Notes on Continuous Stochastic Phenomena. *Biometrika* **1950**, *37*, 17–23. [CrossRef] [PubMed]

44. National Department of Health. South Africa Demographic and Health Survey 2016. Available online: https://dhsprogram.com/pubs/pdf/FR337/FR337.pdf (accessed on 1 August 2021).

45. Fransman, T.; Yu, D. Multidimensional poverty in South Africa in 2001–2016. *Dev. S. Afr.* **2019**, *36*, 50–79. [CrossRef]

46. Woldesenbet, S.A.; Kufa, T.; Lombard, C.; Manda, S.; Ayalew, K.; Cheyip, M.; Puren, A. *The 2017 National Antenatal Sentinel HIV Survey Key Findings*; National Institute of Communicable Disease: Pretoria, South Africa, 2019.

47. Massyn, N.; Padarath, A.; Peer, N.; Day, C. *District Health Barometer 2016/17*; Health System Trust: Durban, South Africa, 2017.

48. Van Schalkwyk, C.; Dorrington, R.E.; Seatlhodi, T.; Velasquez, C.; Feizzadeh, A.; Johnson, L.F. Modelling of HIV prevention and treatment progress in five South African metropolitan districts. *Sci. Rep.* **2021**, *11*, 5652. [CrossRef]

49. De la Cruz, R.; Branco, M.D. Bayesian analysis for nonlinear regression model under skewed errors, with application in growth curves. *Biom. J.* **2009**, *51*, 588–609. [CrossRef]