World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:15, No:5, 2021

Bayesian Geostatistical Modelling of COVID-19 Datasets

I. Oloyede

Open Science Index, Mathematical and Computational Sciences Vol:15, No:5, 2021 publications.waset.org/10012054.pdf

***Abstract*—**The COVID-19 dataset is obtained by extracting weather, longitude, latitude, ISO3666, cases and death of coronavirus patients across the globe. The data were extracted for a period of eight day choosing uniform time within the specified period. Then mapping of cases and deaths with reverence to continents were obtained. Bayesian Geostastical modelling was carried out on the dataset. The study found out that countries in the tropical region suffered less deaths/attacks compared to countries in the temperate region, this is due to high temperature in the tropical region.

***Keywords*—**COVID-19, Bayesian, geostastical modelling, prior, posterior.

I. Introduction

BAYESIAN geostatistical modelling had been adopted by numerous researchers to optimize the parameters of the predictor variables of logistic (Bernoulli) and Poisson regression of geostatistical data. This is due to its robust performance over the classical approach. Reference [1] adopted Bayesian geostatistical modelling to model infection data which provided robust estimates of morbidity risk for area with scanty data. Reference [2] used Bayesian binomial geostatistical model to obtain the effects of environment and socio-economic characteristics on the Soil Transmitted Helminth (STH) infection risk.

Logarithmic score criterion was used to select the most important predictor variables from the combination of covariates of STH species specific infection [3]. Reference [4] fitted Bayesian geostatistical logistic regression models with location-specific random effects and selected the most appropriate predictor covariates. Reference [5] estimated STH infection in unobserved location using MCMC procedure cum joint Bayesian kriging.

Geographical information system had been used along with Bayesian geostatistical modelling of the STH species in Brazil. Reference [1] modelled location with log Gaussian Cox process using improper prior conjugated with likelihood considering the fact that the data obtained were informative such that the degree of information sample were under controlled when improper prior were placed on the parameters. Reference [6] proposed shared latent model of the animal species with non-informative prior which is due to the nature of the data that violate point reference spatial data. Geostatistical data of coronavirus were extracted manually from different repositories within a specified period of time for uniformity.

I. Oloyede is with the Department of Statistics, University of Ilorin, P.M.B 1515, Ilorin, Nigeria (e-mail: oloyede.i@unilorin.edu.ng).

Reference [7] pointed out that R software is an excellent tool for disease surveillance with reproducible health data analysis. Thus, summaries and maps of disease risk estimates might be presented in interactive dashboards using flexdashboard [8] and web applications using shiny [9]. Reference [10] opined that their predictive risk maps were useful for disease control managers for prioritizing control interventions and for providing a tool for more efficient surveillance-response mechanisms. They concluded that the estimates from their study could serve as benchmark for monitoring and evaluation of future efforts in an attempt to disease elimination.

The sections in the study are as follow: sequel to Section I that contains introduction to the review of literature is the Section II that examined the coronavirus pandemic, Section III looked into methodical design which emphasized the statistical framework of the study, in Section IV, the sources of data were explained. Data analysis and interpretation were set out in Section V while Section VI dealt with conclusion. This study therefore examines the effect of weather conditions on the outbreak of coronavirus (COVID-19) pandemic and the death that may result from it.

II. Coronavirus Pandemic

Reference [11] claimed that the emergency of novel coronavirus SARS-CoV-2 popularly known as COVID-19from a city called Wuhan in China recorded a total of 14,557 laboratory-confirmed cases as of February 2, 2020, this was reported in [12]. The spread of the deadly disease has been majorly attributed to transmission from person-to-person in hospital, market, schools, mosques or churches [14]. Reference [11] established that in Japan, some tourists who were in cluster came in contact with tourists in Wuhan and got infected. The deadly disease ravaged economy of virtually all the continents in 2020, WHO had no choice than to declare it as global pandemic. Deaths from the scourge hit many countries in large numbers. Despite the fact, the disease is at minimum at the last quarter of 2020, it left many countries in a state of economic recession particularly the third world countries.

Reference [11] affirmed that the Middle East Respiratory Syndrome (MERS) coronavirus or endemic human coronaviruses (HCoV) can exist on inanimate object surfaces such as metal, wood, glass or plastic for a period of 9 days. Various claims were in different quarters as it can exist in atmosphere, various precautions with respect to hygiene were given such as washing of hands regularly with soap, water and or sanitizer. People were asked to keep physical and social

Open Science Index, Mathematical and Computational Sciences Vol:15, No:5, 2021 publications.waset.org/10012054.pdf

distancing at least two meters apart. People around the world have adopted nose masking as a preventive measure to reduce the spread of COVID-19. Reference [11] concluded that effective surface disinfection will minimize and reduce the spread of the deadly COVID-19. Reference [15] reported that the coronavirus disease (COVID-19) came into limelight in December 2019 in the city of Wuhan, China which spread rapidly due to close contact amongst people, thus 70,548 infections and 1,770 deaths in mainland China and 413 infections in Japan were confirmed as at February, 16, 2020. Large indices of cases and deaths were recorded across the globe during 2020.

The World Health Organization (WHO) recommended that "to ensure that environmental cleaning and disinfection procedures are followed consistently and correctly. Thoroughly cleaning environmental surfaces with water and detergent and applying commonly used hospital-level disinfectants (such as sodium hypochlorite) are effective and sufficient procedures."[13]

Reference [15] pointed out Chloroquine, aged-long malaria drug which has high degree of efficacy in the prevention and treatment as well as anti-inflammatory agent for the curing of rheumatoid arthritis and lupus erythematosus. It was revealed that the drug has potential broad-spectrum antiviral activities which are as a result of augmenting endosomal pH that is needed for virus/cell fusion, which equally interferes with the glycosylation of cellular receptors of SARS-CoV [16].

Reference [16] revealed that chloroquine with anti-viral and anti-inflammatory properties has potent efficacy in treatment of patients with COVID-19 pneumonia. The drug is economical and less reactive with prolong history of about 70 years in the prevention and treatment of the aforementioned diseases. Due to the exigency of the need to proffer solution to the ravaging pandemic, chloroquine phosphate was recommended to treat COVID-19 associated pneumonia in larger populations.

III. Methodical Design

Let Y_i be the number of cases of COVID-19 at region C_i $\{i = 1, \dots, n\}$ under study. Let N be total number of tested individual quarantined for each COVID-19 infection, thus Y follows Poisson distribution such that $Y_i \sim (\lambda)$ where λ is np, with p the probability that COVID-19 is confirmed from quarantined or tested individual and symptom in country C_i. This study collected data and modelling it in Bayesian geostatistical paradigm. The study specified the model as $Y_i = \beta X_I + v_i + w_j + e_{ij}$ where $\beta's$ denote the coefficients of the covariates X with v_i as the location specific random effect, w_i denotes exchangeable random effect and e_{ij} denotes random error.

Let a Poisson experiment with average number of successes within a given region be λ, then it is given by:

$$P(y, \lambda) = \frac{e^{-\lambda}\lambda^y}{y!}, \lambda > 0 \; and \; y = 0,1,2,3, \dots. \quad (1)$$

where y is the actual no of successes of the outcome from the experiment, while λ denotes μ and $\hat{\sigma}^2$ of the distribution.

$$E(y) = Var(y) = \lambda \quad (2)$$

The negative likelihood function of β can be written as:

$$L = -log\left[\prod_{i=1}^{n} \frac{e^{-\lambda_i - \lambda_i^{y_i}}}{y_i!}\right] \quad (3)$$

$$= y_i log(\lambda_i) + \lambda_i + log(y_i!) \quad (4)$$

$$= y_i X'_i\beta + exp(X'_i\beta) + log(y_i!) \quad (5)$$

Then, maximum likelihood estimator for β is:

$$\hat{\beta} = \frac{\partial L}{\partial \beta} = \left(y_i + exp(X_i\beta)\right)X_i \quad (6)$$

Setting (6) equal to zero does not admit a closed form solution for β. Hence, a numerical method must be used in order to obtain the estimate.

The poison likelihood is

$$p(y|X, \beta) = \prod_{i=1}^{N} \frac{1}{y!}\mu(x_i)^{y_i} e^{-\mu(x_i)} \quad (7)$$

$$= \prod_{i=1}^{N} \frac{1}{y(y-1)!} e^{(X\beta)y_i} e^{-e^{X\beta}} \quad (8)$$

Since $p(y|X, \beta)$ is not intractable, if conjugated with normal prior, there will be need to make use of log-gamma which is approximately Gaussian when $\alpha \to$ to be large.

Let $\mu \backsim Gamma(a, b)$

$$p(\mu|a, b) = \frac{1}{\Gamma(a)b^a}\mu^{a-1}e^{-\frac{\mu}{b}} \quad (9)$$

Thus $\lambda = log\mu$, this follows log-gamma distribution, as it is established in the literature, that for large a the log-gamma density approximately Gaussian distribution. Let $b = 1$ and $a = y$ then we have

$$p(\lambda|y, b) = \frac{1}{\Gamma(y)}\mu^{a-1}e^{-\mu} = \frac{1}{(y-1)!}e^{(X\beta)(y-1)}e^{-e^{X\beta}} \approx G(X\beta|logy.y^{-1}) \quad (10)$$

Therefore we have

$$p(y|X, \beta) = \prod_{i=1}^{N} \frac{1}{y} G(X\beta|logy_i.y_i^{-1})) \quad (11)$$

$$= \frac{|\Sigma_y|^{1/2}}{(2\pi)^{N/2}} e^{-\frac{1}{2}\left\|X'\beta - \log(y)\right\|^2_{\Sigma_y}} \quad (12)$$

where $\Sigma_y = diag\left[\frac{1}{y_1}, \dots, \frac{1}{y_N}\right]$

$$p(\beta|X, y) = p(y|X, \beta) * p(\beta) \quad (13)$$

Taking log of both sides we have

Open Science Index, Mathematical and Computational Sciences Vol:15, No:5, 2021 publications.waset.org/10012054.pdf

$$\log(p(\beta|X,y)) \propto \log(p(y|X,\beta)) * \log(p(\beta) \approx -\frac{1}{2}\|X'\beta - \log(y)\|^2_{\Sigma_y} - \frac{1}{2}\|\beta\|^2_{\Sigma_p} \quad (14)$$

The terms that do not contribute to the function of β parameter are dropped, then there is need to complete the squares and expand the term, thus the posterior distribution is approximately Gaussian as

$$p(\beta|X,y) \approx G(\beta|\hat{\mu}_\beta, \hat{\Sigma}_\beta) \quad (15)$$

where

$$\hat{\mu}_\beta = \left(X\Sigma_y^{-1}X' + \Sigma_p^{-1}\right)^{-1} X\Sigma_y^{-1}\log(y) \quad (16)$$

$$\hat{\Sigma}_\beta = \left(X\Sigma_y^{-1}X' + \Sigma_p^{-1}\right)^{-1} \quad (17)$$

IV. Sources of the Data

The basic variables are cases and deaths of coronavirus, temperature, humidity, precipitation, wind, longitude and latitude, iso3666. The study is important to clear the air about Africa view relating to coronavirus as it is born from the speculation that Africa has genetic composition that will not make the virus to infect them. The dataset showed those continents that were seriously affected by the deadly virus. The researcher and policy makers can benefit from the data and maps in order to strengthen their decision making. It gives signal to the global community that they should at all-time be prepare for any eventuality.

Data on temperature, humidity, precipitation and wind were obtained from [17] in a period of eight days, specifically between the hour of 8.00 am and 11.59 am for uniformity. Data on cases and deaths due to coronavirus were obtained from [18]. Data on longitude and latitude were obtained from Google map for a period of three days mainly for 211 countries. Data on ISO3666 were obtained from [19]. All the data were matched together to arrive at the dataset. All the data were matched together to arrive at the dataset.

V. Data Analysis and Interpretation

The integrated nested Laplace approximation (INLA) [7] approach returns an object that contains the information of the fitted model including several summaries and the posterior marginal of the parameters, the linear predictors, and the fitted values. The study modelled data on coronavirus (COVID-19) cases and deaths with weather conditions in 211 countries. The objective of this study is to use cases reported to assess each country's performance and identify whether any country performs unusually well or poor.

TABLE I
Fixed Effects of Weather Data on Cases of Coronavirus: Fixed Effects

	mean	sd	0.025	0.975	t-test	p-value
Intercept	13.674	0.964	11.78	15.567	14.185	0.0001
Temperature	-0.060	0.022	-0.103	-0.017	-2.727	0.0069
Precipitation	0.006	0.009	-0.011	0.023	0.667	0.5057
Humidity	-0.052	0.010	-0.072	-0.032	-5.20	0.0001
Wind	-0.096	0.023	-0.141	-0.052	-4.174	0.0001

The column mean showed that temperature, humidity and wind have negative effects on the spread of COVID-19, though they are significant but the effects close to zero, this implies that countries with high temperature, humidity and wind suffer less deaths/attacks compared to those countries in the temperate region, this is supported by the maps shown in Figs. 2 and 3. Precipitation has positive effect but not significant. Columns 0.025quant and 0.975quant contained the lower and upper limits of 95% credible intervals of the cases and provided measures of uncertainty. The fixed effects from Table I depicted that the temperature varies *across* countries by one unit, so also *cases* decreases by *0.06* units holding other variables constant, the marginal effect of precipitation on cases is 0.006 unit while other variables are constant. The marginal effects of humidity and wind are 0.052 and 0.096 units respectively while other variables are held constant.

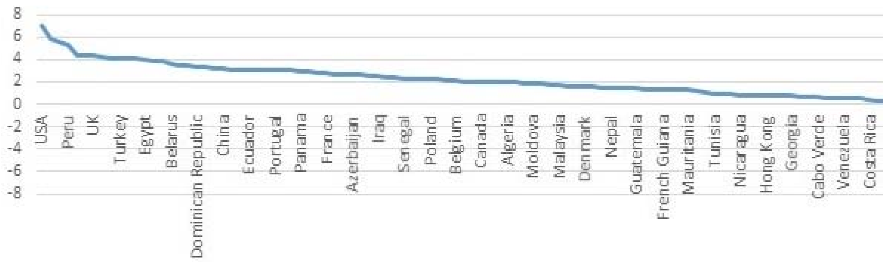

Fig. 1 Random effects of weather data on cases of coronavirus

The column for the mean showed that USA, Peru and United Kingdom were with the highest posterior means of the coronavirus pandemic, Columns for the 25% and 95% quadrant respectively contain the lower and upper limits of 95% credible intervals of the cases and provided measures of uncertainty.

VI. Conclusion

The study investigated the effect of weather condition on the outspread of coronal virus pandemic in 211 selected countries; the outcome of the study showed the evidence that the virus is global pandemic, thus the disease is peculiar to a particular continent. It is a worldwide pandemic. The study concluded that temperature, humidity and wind have negative effects on the spread of COVID-19, though they are significant but the effects were close to zero, this implies that countries in tropical region suffer less deaths/attacks compared to those countries in the temperate region.

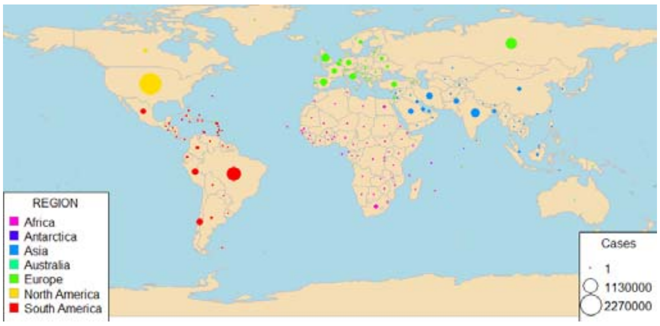

Fig. 2 Cases of coronavirus patients in seven different continents as at fifth May, 2020. The map showed the cases of COVID-19 as collected by webometric. It is sparsely distributed

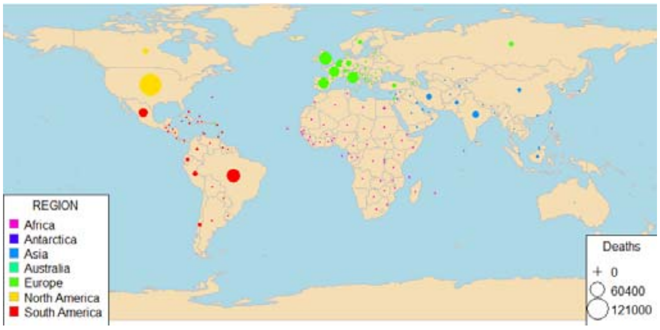

Fig. 3 Deaths of coronavirus patients in seven different continents as at fifth May, 2020. The highest death recorded in United State of America, Asia and part of South America

The age long malaria and inflammatory drug resurfaced as potential cure of coronal virus in some quarters. Most countries were experiencing recession, particularly the third world countries. This was due to prolong lockdown which inhibited production. The good side of COVID-19 is the improved hygienic conditions on the part of citizenry, people have chosen it as point of duty to maintain proper hygiene, keeping physical distancing has become the order of the day particularly in public gathering.

References

[1] D. Pati, B.J Reich and D.B Dunson, 'Bayesian geostatistical modelling with informative sampling locations', Biometrika , 98, 1, pp. 35–48 doi: 10.1093/biomet/asq067C _2011 Biometrika Trust Printed in Great Britain B.

[2] H. Rue, S. Martino and N. Chopin, 'Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations'. J.R. Stat Soc Ser B Stat Methodol 71:319-392, 2009

[3] I. M Held, M. Winton, K. Takahashi, T. Delworth, F. Zeng, and G.K. Vallis., 'Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing', J.Clim., 23, 2418– 2427, doi:10.1175/2009JCLI3466.1 , 2010.

[4] G. C. S. Ronaldo, S. Nadine, E. B. Maria, M.C. Edgar, C. Frédérique, U. Jürg, and V. Penelope, 'Spatial analysis and risk mapping of soil-transmitted helminth infections in Brazil, using Bayesian geostatistical models', Geospatial Health 8(1), pp. 97-110, 2013.

[5] P. Diggle, J. Tawn, R. A. Moyeed, 'Model-based geostatistics'. Appl. Stat. 1998;47:299–350.

[6] P. Diggle, R. Menezes and T. Su, 'Geostatistical inference under preferential sampling (with discussion)'. Appl. Statist. 59, 191–232, 2010.

[7] M. Paula, 'Small Area Disease Risk Estimation and Visualization Using R', The R Journal Vol. 10/1, 2018.

[8] J. Allaire, 'Flexdashboard: R Markdown Format for Flexible Dashboards, https://CRAN.Rproject.org/package=flexdashboard. R package version 0.5. p504/2017.

[9] Chang W., Cheng J., Allaire J., Xie Y., and McPherson J. Shiny: Web Application Framework for R, 2017. URL https://CRAN.R-project.org/package=shiny. R package version 1.0.5. p504

[10] G. C. S. Ronaldo, S. Nadine, Maria E. B., M.C. Edgar, C. Frédérique, U. Jürg and V. Penelope ,'Spatial analysis and risk mapping of soil-transmitted helminth infections in Brazil, using Bayesian geostatistical models', Geospatial Health 8(1), 2013, pp. 97-110

[11] K. Gunter, 'Potential role of inanimate surfaces for the spread of coronaviruses and their inactivation with disinfectant agents', Infection Prevention in Practice, www.elsevier.com/locate, 2020.doi: 10.1016/j.infpip.2020.100044

[12] WHO.' Infection prevention and control during health care when novel coronavirus (nCoV) infection is suspected'. Interim guidance. 25 January 2020.

[13] WHO(2020). 'Novel Coronavirus (2019-nCoV)'. Situation Report 13. WHO;2020.

[14] J.F. Chan, Yuan S, K.H. Kok, K.K. To, H. Chu, nd J. Yang, 'A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster'. Lancet 2020. https://doi.org/10.1016/s0140- 6736(20) 30154-9.

[15] G. Jianjun, T. Zhenxue and Y. Xu, 'Breakthrough: Chloroquine phosphate has shown apparent efficacy in treatment of COVID19 associated pneumonia in clinical studies', BioScience Trends Advance Publication,DOI:10.5582/bst.2029.01047,www.bioscienctrends.com/2020

[16] L. Kim, COVID-19 Had Us All Fooled, but Now We Might Have Finally Found Its Secret' Medium Post https://www.truthorfiction.com/the-covid-19-chloroquine-controversy-explained/, 2020

[17] Data on Temperature, Humidity, Precipitation and Wind were obtained from www.timeanddate.com /2020

[18] Data on Cases and Deaths due to coronavirus were obtained from https://www.worldometers.info/coronavirus/ on 4/5/2020

[19] Data on ISO3666 were obtained from https://github.com/lukes/ISO-3166-Countries-with-Regional-Codes/blob/master/all/all.csv/2020.

Open Science Index, Mathematical and Computational Sciences Vol:15, No:5, 2021 publications.waset.org/10012054.pdf