Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2020;6(2):81-92

Original Article

The Effect of Weather Parameters on Covid 19 Endemic: A Global Perspective

Marina Roshini Sooriyarachchi*

Department of Statistics, University of Colombo, Sri Lanka.

ARTICLE INFO

Received 06.05.2020
Revised 23.05.2020
Accepted 11.06.2020
Published 20.06.2020

Key words:
Covid 19;
Weather;
Generalized linear mixed models (GLMM's);
Negative binomial;
Cluster

ABSTRACT

Introduction: According to the Oxford Medical Dictionary, Corona virus is the largest known viral RNA genome and causes devastating epizootics in livestock and poultry. Human corona viruses cause upper respiratory tract infections and severe acute respiratory syndrome (SARS). The initiative for this study was the extreme life threatening nature of this virus and the global pandemic it has caused. The responses were taken to be the number of deaths, number of recoveries and the number sick with the disease at a particular point in time, globally and the explanatory variables were climate variables.
Method: This is a survey type of study as the data has been extracted over a short period of time and the sampling method adopted is post cluster sampling. Simple descriptive statistics, clustering and generalized linear mixed models have been used for modelling.
Results: There was a strong regional effect of over three which was highly significant for every Covid 19 response. The air quality and temperature interaction and the air quality and humidity interaction were associated with the count of death at 0.0298 and 0.0027 levels of significance respectively. The count recovered was strongly associated with the temperature and humidity interaction and air quality at significance levels of 0.0002 and <0.0001 respectively. The Count at risk was strongly associated with the temperature, wind speeds and air quality three way interaction and this was significant at 0.0005 level.
Discussion: All four weather parameters effected one or more of the Covid 19 responses. The plots of Student residuals versus fitted values showed well-fitting models. The results of this research is useful in planning health care and allocating resources according to the region and the climate during a particular period.

Introduction

Covid 19 is a viral life-threatening infectious disease which has within a short time reached endemic proportions. This disease started in late December 2019 in Wuhan, China and by now has spread to every part of the globe. Hundreds of thousands of cases and many deaths have occurred on a global scale. In man this disease effects the upper respiratory track system and takes the form of acute pneumonia. The symptoms are coughing, fever, sneezing and difficulty in breathing. Past data has shown that this effects the older people more. After starting in China (the first patient with this disease was reported to the WHO from China) it quickly spread in huge proportions to Western Europe and USA mainly effecting Italy, France, UK, Spain, Germany and North America. Though there has not been much information of the effects of climate on Covid-19 a milder viral life-threatening infectious disease, Dengue has shown

* .roshini@stat.cmb.ac.lk

Please cite this article in press as Sooriyarachchi M R. The Effect of Weather Parameters on Covid 19 Endemic: A Global Perspective. *2020;6(2): 81-92*

to be highly related to the climate. Its vector which is mosquitoes, breeding is closely related to the climate [1], [2], [3]. Thus, however slight the evidence was of a close relationship between climate and Covid 19 it was interesting and timely to pursue a likely relationship.

The data for the study came from 2 websites, namely, a real time visualizer in the form of a world map providing for each country the number of deaths, the number recovered and the number having the disease [4] and the website for the climate of each country came from an online map provided by Ventusky [5] where from various site reports the parameters, Temperature, Relative Humidity, Wind speed and Air Quality was extracted. There were vague reports on world news on the television that high Temperature kills the virus while it is also negatively affected by humidity. One vital piece of information by studying the data was that countries with high temperature and humidity seemed to have few cases, deaths and more recoveries compared to other combinations of the two climate parameters. The logic behind taking wind speed was with the idea of the virus being spread quicker with the wind. Finally, the air quality was thought to generally affect respiratory disease.

Processing Study Data

The data was extracted from the Covid 19 database [4] by rotating the globe with the cursor and stopping at a country and double clicking on the country. This gives a pop-up menu with the number of deaths, number of recoveries and the number of cases at the particular point of time for the country checked. As many countries as possible were taken, however some countries were not shown clearly. A representative sample of 115 countries in all regions of the globe were selected. The data were stored in an EXEL CSV file. The first column represented the country which was grouped in to the region and was stored in the second column. The next three columns consisted of the three Covid 19 parameters. Then the weather website Ventusky [5] was used to extract the four weather parameters and these four parameters were put in the next four columns of the data file. A few countries had missing values and these observations were deleted. After the cleaning of the data there were 9 variables and 115 countries. Occasionally internet was used to get data which was not clearly shown by the weather website. The main difficulty was that the climate parameters varies with the location within the country. As much as possible the average climate was used for a country and when that was not available the climate of the capital was used instead. Once the data was stored it was exported to SAS 9.4.

Theory and Methodology

The data was collected from the websites on the 26th and 27th March, 2020, indicating that this was a survey type of a study. Though this was a short period of time there was adequate data to study the relationship of interest, as several countries in the entire world was used and there were more than 100 observations. Again, from the various dengue studies in which the author has been involved it was found that the disease was dependent on the geographical region [6]. Thus, from this evidence and also as it was clearly evident from the data that the region (mainly continent) was highly affective and also obviously the climate depended on the region the region was taken as a cluster variable. Thus the sampling was post cluster sampling [7]

Initially, descriptive statistics were used to get a simple idea of the data. These involved summaries and graphs. Then the data was clustered to visualize groups of countries showing similar patterns. The next and more advanced stage was modelling. In modelling the three Covid 19 parameters were univariately modelled as there did not seem to be much correlation between the three responses taken together. In depth scrutiny of the modelling procedure showed that the region was grouped in to 6 categories, namely, America, Western Europe, Eastern Europe, Asia, Africa and Oceania. As each of the responses were more similar within clusters than between clusters, the correlation within clusters had to be adjusted for.

For this purpose a Generalized Linear Mixed Model with the cluster / group effect taken to be region was used. Though the Poisson distribution was the first choice for the count data, Negative Binomial distribution was used instead because of over-dispersion.

Finally, the models were used to determine how the weather parameters were related to the Covid 19 outcomes. Here the Covid 19 outcomes are the dependent variables and the weather parameters are the independent variables. The effects were then quantified using model parameters. The entire analysis was done using SAS 9.4.

Simple Analysis

Some descriptive statistics such as counts, means, medians standard deviations, maximums and minimums for the entire data set were extracted. At the modeling stage, when continuous climate data were taken as explanatory variables these did not show much significant association with the Covid 19 responses therefore, the climate data had to be grouped into categories. This was done by calculating quantiles of the weather data and grouping each climate variable in to four categories. It was shown that this was much more efficient than a grouping based on random selection of intervals for the explanatory variables. Correlations between the response variables were looked at to decide on the type of models to be used. The correlation between all three responses put together was non-significant though two pairs, namely deaths and risk and deaths and recovered were significant but the third pair that is recovered and risk was not significant indicating that univariate models would be better than joint models. The weather parameters were correlated but the correlations were not too high. There seemed to be vague correlations between the Covid 19 parameters and the weather parameters which needed to be further explored using advanced methods. A graphical technique, namely, three dimensional graphs were used to visualize the data.

Clustering

The multivariate technique of clustering [8] was used to group similar countries with respect to the disease. The method of average clustering was used and the countries were grouped in to 8 categories. The variables that were used to group the countries were the disease parameters and the four climate variables. A tree diagram [8] was used to visualize the countries within each group.

Advanced Analysis

Generalized Linear Mixed Models (GLMMs) [9], [10] were used to model the count data on the 3 disease responses. As there were no significant correlations between all 3 parameters when taken together, univariate models were used. As this was count data it was not normally distributed. The alternatives were the Poisson and Negative Binomial distributions. On examining the data closely it was found that the Poisson distribution was inadequate as there was over-dispersion. The Negative Binomial gave a good fit to the data, providing sensible conclusions. The explanatory variables were the four weather variables and the cluster variable was region. All the variables were in the form of factors. Both forward selection and backward elimination were used to select important variables for the models. If the main effect model was inadequate then important 2 factor interactions were added.

As there was a clustering affect Generalized Linear Mixed Models (GLMMs) was used instead of Generalized Linear Models (GLMs). Here the cluster variable was taken as region as the correlation of disease responses between countries of the same region was found to be much higher than the correlation between countries of different regions. The method of estimation was Maximum Likelihood with the Laplace method of approximation. The log of the population size of the country was used as an offset. This is because the Covid 19 parameters depend on the population size of the country. The log was used to help model convergence.

Finally using the results from the model the factors affecting the Covid-19 parameters were identified. The estimates obtained were quantified to determine the direction and magnitude of the weather parameters.

Model Representation

$$\text{Log (Covid 19 Response)}_{hijkl} = \beta_{oh} + \beta_i^T + \beta_j^H + \beta_k^W + \beta_l^A \quad \ldots \quad \ldots\ldots\ldots\ldots \quad (1)$$

Here i, j, k and l are such that i corresponds to the ith level of temperature, j corresponds to the jth level of humidity, k corresponds to kth level of wind speed and l corresponds to the lth level of air quality. Here β_{oh} consists of a fixed part β_o and a random part u_{oh} that is $\beta_o + u_{oh}$ and $u_{oh} \sim$ Normal $(0, \sigma_u^2)$. The Covid 19 Responses are all assumed to have Negative Binomial distributions as the Poisson distribution is inadequate due to over-dispersion. Here the β's are the effects which are assumed to be fixed. The link for the Negative Binomial Distribution is the Log link.

Parameter Interpretation

By taking the exponential of the β's the effect on the Covid 19 Response of the rate of a level of the weather parameter with respect to its base value can be estimated. This can be represented by

Rate of the level i of the Temperature weather parameter for example is obtained with respect to its base level = Exp (β_i^T)

This is obtained by :

$$\text{Log (Estimate of the Covid 19 Response for the ith level of Temperature)} = \beta_{oh} + \beta_i^T \quad (2)$$

$$\text{Log (Estimate of the Covid 19 Response for the Base level of Temperature)} = \beta_{oh} \quad (3)$$

$$\text{Log (Estimate of the } i^{th} \text{ level of Temperature on the Covid 19 Response / Estimate of the base level of Temperature on the Covid 19 Response)} = \beta_{oh} + \beta_i^T - \beta_{oh} = \beta_i^T \quad (4)$$

$$\text{Rate of the level i of the Temperature Weather parameter with respect to its base level} = \text{Exp}(\beta_i^T) \quad (5)$$

Here u_{0h} is the random effect corresponding to the h^{th} Region.

What has been done up to now

The many reports that have emerged in the Epidemiology Unit in Sri Lanka [11] and the WHO [12] mainly report descriptive statistics. I believe this is the first attempt to do an advanced statistical analysis. In other areas differential equations, mathematical modelling and numerical analysis, physical models, biological and zoological studies, computer related studies and most importantly medical studies have been carried out. However, the author has not seen a study related to the weather. This study uses the power of computing to perform efficient statistical analysis.

Results

Descriptive Statistics

Table1 gives the number of countries in each region

Table 1. Frequency table of Regions

Region	Number of Countries
1.Western Europe	15
2. Eastern Europe	16
3.America (South+North)	19, 2
4.Africa	33
5.Oceania	3
6.Asia	27

Table 1 indicates that most of the countries are from Africa (33) and then from Asia (27) indicating the developing world. The lowest number of countries are from Oceania (3). Western Europe (15), Eastern Europe (16) and South America (19) are represented by a moderately large number of countries. There are only 2 areas from North America. These make up the 115 countries. There are 6 regions.

Table 2 represents some summary statistics in the form of means, quantiles standard deviations, maximums and minimums for each Covid 19 parameter for all the regions for each of the weather parameters.

Table 2. Summary Statistics

Variable	Mean	Q1	Q2	Q3	SD	Max	Min
deaths	202.026	0	3	16	950.07	8259.00	0
recovered	1054.58	0	6	33	7120.29	74588.00	0
No. at risk	3011.30	22	190	943	10355.92	67731.00	1.00
temperature	15.041	2.23	17.22	26.11	13.525	59.00	-11.67
humidity	67.15	50	70	90	23.99	100.00	10.00
Wind speed	9.015	3.22	4.83	11.27	11.163	88.55	0
Air quality	21.46	8	18	26	25.197	205.00	0

In Table 2 the quantiles are calculated only for the weather variables in order to categorize these explanatory data in to groups. The means show that the deaths are relatively low compared to the number recovered and having the disease (risk). The mean of air quality shows that on average countries seem to be having only moderate air quality. Humidity on the other hand is quite high on average. The standard deviation is much higher than the average for Covid-19 parameters. This indicates that the status of Covid 19 varies hugely between countries. The minimum shows that deaths and recovered from Covid 19 are zero for some countries. This is useful for the modelling stage as it shows an excess of zeros indicating that either the negative binomial or the zero inflated Poisson regression is better for modelling than the Poisson distribution. As the negative binomial is simpler and easier to interpret, this was used.

Figure 1 (a) and Figure 1 (b) indicates three dimensional graphs of temperature and humidity versus two of the more significant Covid 19 parameters, deaths and risk.

Figure 1 (a) indicates that low temperature and low humidity has resulted in a high number of deaths. Figure 1 (b) indicates that low temperature and high humidity has resulted in a large number of people currently having the disease.

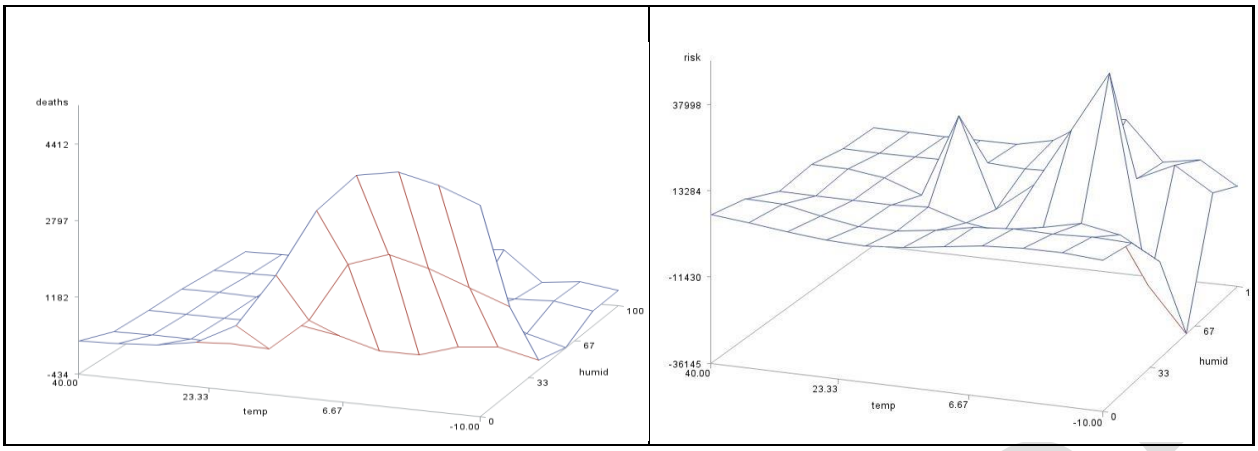

Figure 1. a. Three dimensional plot of deaths, temperature and humidity.b. Three dimensional plot of risk, temperature and humidity

Clustering of Countries

The average clustering method with 8 clusters was used and the Covid 19 responses and also the weather explanatory variables were used to group the data into similar clusters. Eight clusters were taken as this seemed to give a sensible grouping. A tree diagram was used to visualize the countries within each group. Table 3 gives the countries within each group. Most countries are within group 1 and these represent the countries having a mild form of Covid 19. These are mainly the countries in Africa, Asia and South America. This consists almost entirely of the developing world. The other groups have stronger forms of the disease while the worst forms of the disease are in Italy, USA and China and these countries are all on their own in groups 6, 7 and 8. This table also represents the 115 countries in the study.

Table 3. Countries within groups (G)

G1	G2	G3	G4	G5	G6	G7	G8
Central African Republic (CAR)/ Chad/Angola/ Guinea/ Mali/ Mauritania/ Libya/Papua New Guinea/ Congo/ Paraguay/ Nicaragua/ Somalia/ Myanmar/ Syria/El Salvador/ Tanzania/ Sudan/ Zimbabwe/ Guyana/ Gabon/ Lao/ Ethiopia/ Mongolia/ Mozambique/ Suriname/ Madagascar/ Togo/ Tunisia/ Ukraine/ Uganda/Niger/Namibia/French Guinea/ Guatemala/ Bolivia/ Senegal/ Sri Lanka/ Afghanistan/ Cameroon/ Zambia/ Bosnia/ Cambodia/ Venezuela/ Ghana/ Malta/ Mexico/ Peru/ Slovakia/ Uruguay/Kyrgyzstan/Coted'Ivoire/Kenya/Bulgaria/Morocco/Honduras/Nigeria/Costa Rica/ Burkina Faso/	Canada/ Brazil /Austria Belgium/ Korea/ Portugal /Turkey /Norway /Australia /Sweden	Switzerland UK	France Iran	Germany Spain	Italy	USA	China

Vietnam/ Oman/ Colombia/ Bangladesh/ Uzbekistan/ Azerbaijan/ Kazakhstan/ Croatia/ Armenia/ Ecuador/ Poland/Finland/South Africa/Pakistan/Algeria/New Zealand /Hungary/ Argentina/ Serbia/ Greece/ Russian Republic/ Belarus/ India/ Philippines/ Lithuania/ Romania/ Thailand/ Panama/ Saudi Arabia/ Indonesia/ Iceland/ Slovenia/ Iraq/ Czech/ Malaysia/ Luxemburg /Japan							

Modelling Data

According to the methods explained in section 3, GLMM's were fitted to each Covid 19 response separately. The cluster variable was taken to be region and the explanatory variables were taken to be the 4 weather parameters. The technique used to fit each model was to first fit the model with all two factor interactions. Then backward elimination was used to drop the terms that were not significant at the 5% level of significance. If there were convergence problems in the two factor interactions these were dropped. If the all 2 factor model did not fit the three factor interactions were added.

Model 1

The response variable was taken to be death from Covid 19. The model was fitted to all 115 countries. The selected model was

$$\text{Log}(\text{deaths}_{hijl}) = (\text{Random Intercept})_{oh} + \text{Temp}_{i} + \text{Humid}_{j} + \text{AirQ}_{l} + (\text{Temp}*\text{AirQ})_{il} + +(\text{Humid}*\text{AirQ})_{jl} \quad (6)$$

Where i corresponds to the level of temperature and i=1,..4 , j corresponds to the levels of humidity j=1,…4, l corresponds to the levels of air quality l=1,..4, l corresponds to the levels of wind speed l=1,..4 and h corresponds to the cluster h=1,2,..8,. The Generalized Pearson Chi Square Statistic for model 1 is 81.85 on 115 degrees of freedom giving p-value of 0.992.

Checking the Goodness of fit of model 1

In order to check the GOF of the model the student residuals from the model are plotted against its fitted values. The 95% confidence interval is also included. This is given in Figure 2.

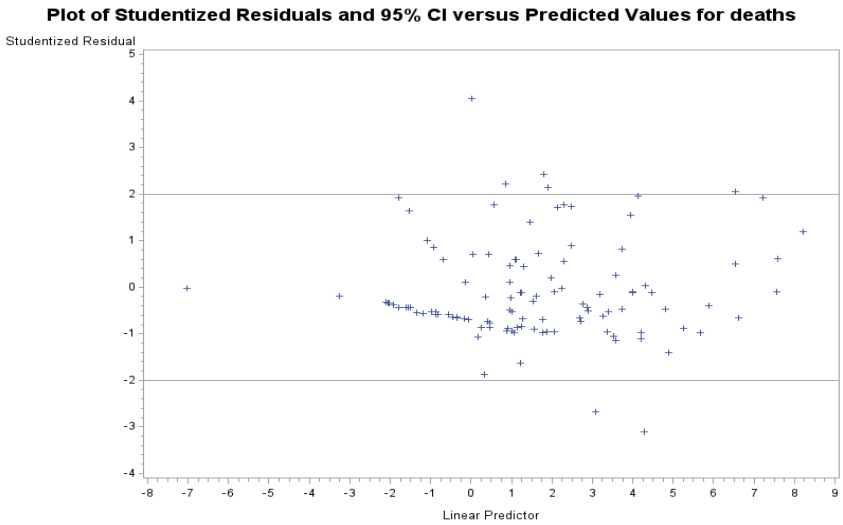

Figure 2. Plot of Student Residuals Versus Predicted values for the Death model

Figure 2 indicates that other than for 6 residuals all the other residuals are within the 95% confidence band. Here 5 out of 100 residuals are allowed outside the confidence intervals. As we have 115 observations then 6 values are allowed outside the band. This clearly shows that model 1 fits well.

Interpreting Parameters from model 1

Initially the direction of the effects will be explained. Then the magnitude will be expressed. Mainly when the air quality is good for mild to moderate temperatures the deaths are less. The rate of decrease of deaths for good air quality and mild temperature in the range 0-10 centigrade compared to good air quality and high temperature greater than 25.56 degrees centigrade is 0.021with a 0.0169 p-value.

The rate of decrease of deaths for good air quality and moderate temperature in the range 10-25.56 centigrade compared to good air quality and high temperature greater than 25.56 degrees centigrade is 0.018 with a p-value of 0.0011.

When the Temperature is mild between 0 – 10 degrees centigrade and the air quality is moderate (9-28) then the deaths are low. The rate of decrease of deaths compared to high temperature and moderate air quality is 0.0038 with a p-value of 0.0063. When the temperature is moderate (>=10-25.56) and the air quality is also moderate the deaths are less. The rate of decrease compared to the high temperature and same degree of air quality is 0.035 with a p-value of 0.0198. When air quality is good and humidity is low (<=60) and when air quality is good and humidity is moderate (70-90) then deaths are low compared to good air quality and extreme humidity (>90). The rates of the low death categories with respect to the base category is 0.019 with a p-value of 0.0379 and with p-values of 0.039 at 0.0294 respectively.

When air quality is moderate and humidity is low the deaths are less compared to moderate air quality and high humidity. The rate is 0.017 with a p-value of 0.0022.

The covariance parameter estimate of region is 3.9159 and its standard error is 2.7199. This indicates that there is a positive cluster effect.

Model 2

The model for the number recovered is of the form

$$\text{Log (recovered}_{hijl}) = (\text{Random Intercept})_{oh} + \text{Temp}_{i} + \text{Humid}_{j} + \text{AirQ}_{l} + (\text{Temp*Humid})_{ij} \quad (7)$$

The Generalized Pearson Chi Square Statistic for model 2 is 93.19 on 115 degrees of freedom giving p-value of 0.932.

In order to check the GOF of the model for the recovered the student residuals from the model2 are plotted against its fitted values. The 95% confidence interval is also included. This is given in Figure 3.

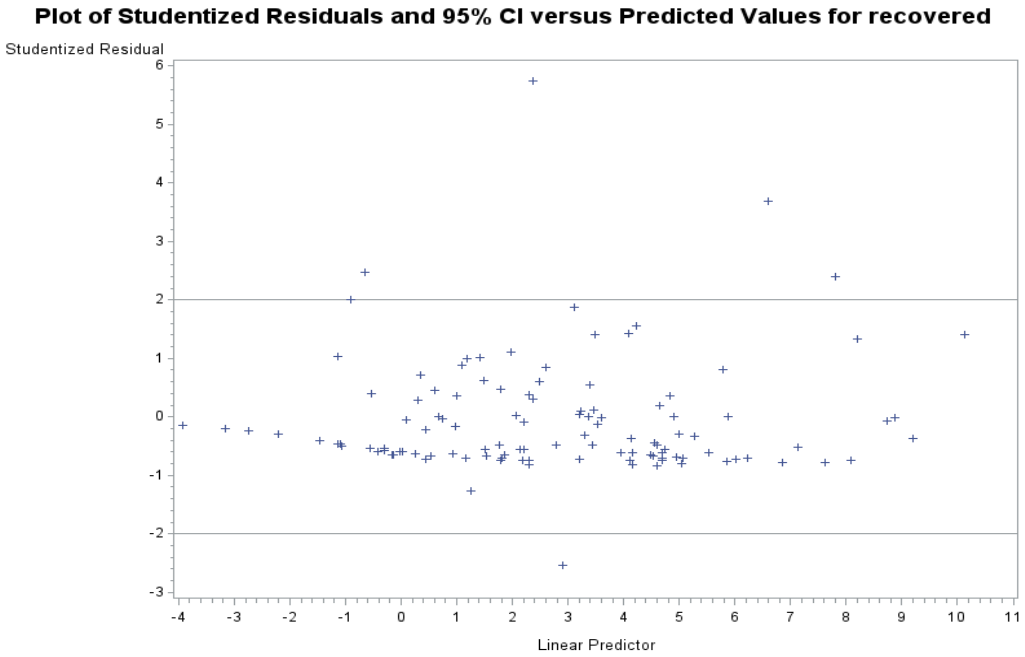

Figure 3. Student Residuals versus Predicted values of the Recovered model

Figure 3 indicates that other than for 5 residuals all the other residuals are within the 95% confidence band. Here 5 out of 100 residuals are allowed outside the confidence intervals. As we have 115 observations the 6 values are allowed outside the band. This clearly shows that model 2 fits well.

Interpretation of results from Model 2

When the air quality is good, moderate and fair the rate of recoveries is much greater than when the air quality is poor. The rates are respectively, 6.2, 11.61 and 28.78 with p-values of 0.0032, 0.0002 and <0.0001 respectively.

When the temperature is low and the humidity is low, mild to moderate the rate of recoveries is much higher than when the temperature is low and humidity is high. These values are respectively, 755.14, 1342.65 and 2107.38 78 with p-values of 0.0028, 0.0029 and 0.0007 respectively.

When the temperature is mild and the humidity is low, mild to moderate the rate of recoveries is much higher than when the temperature is mild and humidity is high. These values are respectively, 2953.95, 3383.95 and 202.78 with p-values of 0.0001, 0.0001 and 0.0127 respectively.

When the temperature is moderate and the humidity is low, mild to moderate the rate of recoveries is much higher than when the temperature is moderate and humidity is high. These values are respectively, 1033.18, 283.52 and 709.95 with p-values of <0.0001, 0.0108 and 0.0002 respectively.

.The covariance parameter estimate of region is 3.1058 with standard error 2.0355. This shows somewhat lesser region covariance than for deaths. However, the covariance seems to be significant.

Model 3

The model for the number having the disease is of the form

$$\text{Log (disease}_{hijkl}) = (\text{Random Intercept})_{oh} + \text{Temp}_{i} + \text{Humid}_{j} + \text{AirQ}_{l} + \text{WindS}_{k} + (\text{Temp*AirQ})_{il} + (\text{Temp*WindS})_{ik} + (\text{Humid*AirQ})_{jl} + (\text{Humid*WindS})_{jk} + (\text{AirQ*WindS})_{kl} + (\text{Temp*WindS*AirQ})_{ikl} \quad (8)$$

The Generalized Pearson Chi Square Statistic for model 1 is 79.57 on 115 degrees of freedom giving p-value of 0.995.

In order to check the GOF of the model3 for the number with disease student residuals from the model3 are plotted against its fitted values. The 95% confidence interval is also included. This is given in Figure 4.

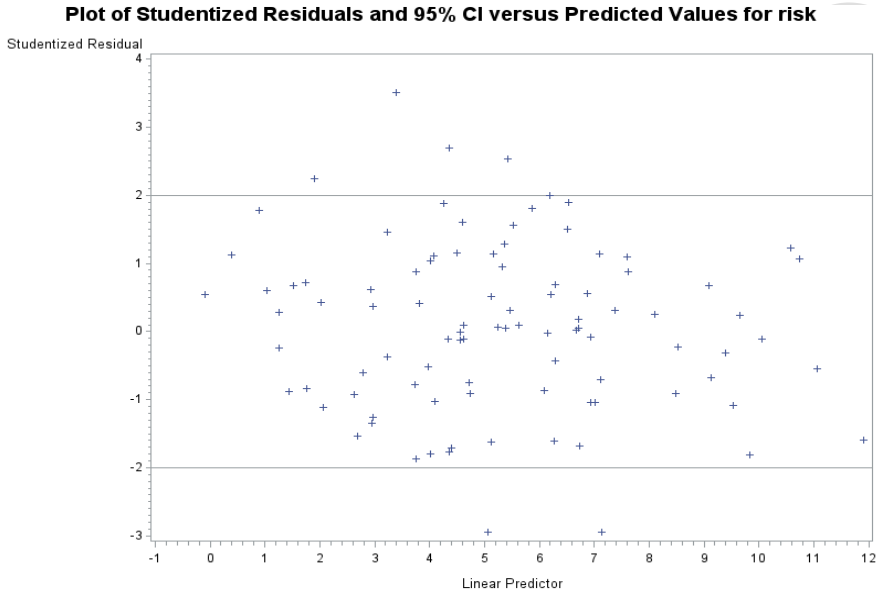

Figure 4. Plot of Student residuals versus predicted values for the disease model

Figure 4 indicates that other than for 6 residuals all the other residuals are within the 95% confidence band. Here 5 out of 100 residuals are allowed outside the confidence intervals. As we have 115 observations the 6 values are allowed outside the band. This clearly shows that model 3 fits well.

Interpretation of results from Model 3

When the temperature is mild, winds are low and the air quality is moderate the rate of disease is lower than when temperature and winds are same and the air quality is poor. The rate being 0.00282 with a p-value of 0.0038.

. When temperature is mild, winds are mild and the air quality is mild the disease rate is higher than when the temperature and winds are the same but air quality is fair. The rate being 12,625.8 with a p-value of 0.0016 which is a little unrealistic. When the temperature is high, the wind speed is low and the air quality is good the disease rate is higher than when the first 2 variables are the same and the air quality is fair however, the significance of this result is low. The rate being 83.12 with a p-value of 0.0498.

When the humidity is low and the air quality is good the rate of disease is lower than when the humidity is low and the air quality is poor. The rate being 0.048 with a p-value of 0.0235. When the humidity is mild and the air quality is good the rate of disease is lower than when the humidity is mild and the air quality is poor. The rate being 0.000018 with a p-value of <0.0001.

When the humidity is low and the air quality is moderate the rate of disease is lower than when the humidity is low and the air quality is poor. The rate being 0.00105 with a p-value of 0.0211. When the humidity is moderate and the air quality is good the rate of disease is lower than when the humidity is moderate and the air quality is poor. The rate being 0.046 with a p-value of 0.0012.

When the humidity is low and the wind speed is low the rate of disease is lower than when humidity is low and the wind speed is high. The rate being 0.049 with a p-value of 0.0116. When the humidity is low and the wind speed is moderate the rate of disease is lower than when humidity is low and the wind speed is high. The rate being 0.07 with a p-value of 0.0006. When the humidity is mild and the wind speed is low the rate of disease is lower than when humidity is mild and the wind speed is high. The rate being 0.0000074 with a p-value of 0.0009. When the humidity is low and the wind speed is mild the rate of disease is lower than when humidity is moderate and the wind speed is high. The rate being 0.000112 with a p-value of 0.0216. When the humidity is low and the wind speed is moderate the rate of disease is lower than when humidity is low and the wind speed is high. The rate being 0.00119 with a p-value of < 0.0001.

The covariance parameter estimate of region is 4.6926 with standard error 3.207. This shows the highest region covariance than for disease and the covariance seems to be significant.

Conclusions and Discussion

Summary of the Study

Most viral life threatening infectious diseases such as Dengue, Leptospirosis (rat fever), Japanese Encephalitis to name a few occur within clusters and are closely connected to the weather patterns within the cluster [13] [14]. This was the initial idea behind studying the effect of weather parameters on the responses of Covid 19. Initial descriptive statistics gave an idea of the data and the quantiles were useful for categorizing the weather variables. The continuous explanatory variables gave no relationships with the responses however, when the explanatory variables were categorized the relations were much clearer.

The correlation between all three responses put together was non-significant though two pairs, namely deaths and risk and deaths and recovered were significant the third pair that is recovered and risk was not. Therefore three univariate models were preferred over a 3 dimensional joint model. Glancing at the data it was very clear that for each response the values were more similar within the region than between the regions. Therefore, Generalized Linear Mixed Models (GLMM) were used with the cluster effect being region for modelling the responses.

From the Covid 19 visualizer data for as many categories of the United Nations were obtained. This resulted in 115 countries. The Ventusky website was used to extract the weather parameters.

As the data was collected on two days was a survey type study with post cluster sampling.

Important Conclusions

The region cluster effect was very efficient especially for deaths and diseased of Covid 19. It was less efficient for the recoveries.

Most effects of the weather on the Covid 19 parameters were complicated and took the form of two factor interactions and in the case of disease there was a single three factor interaction as well.

The main effects for air quality and wind speeds were somewhat clear with good air quality / low wind speeds resulting in favorable results. The main effects for temperature and humidity were not so clear but usually mild to moderate temperature and low to moderate humidity gave favorable results.

Overall Covid 19 showed to be closely affected by the weather parameters, particularly temperature, humidity and air quality.

The clustering of the countries with respect to Covid 19 showed that a majority of the countries with milder form of Covid 19 were together. However, the worst three countries with respect to Covid 19 status namely, Italy, USA and China formed their own clusters.

In the clustering it is clearly shown that a majority of the African, Asian, South American and Eastern European countries were clustered together and these had milder forms of Covid 19. Western Europe, USA, China, Korea and Iran showed more serious forms of Covid 19.

Limitations of the study

The first limitation was in extracting the data for the study. The Covid 19 visualizer did not show all the countries of the UN clearly. Also the Ventusky weather website sometimes showed the weather parameters on an average whilst it sometimes showed this city wise and sometimes for the capital. Also the data set was collected within two days. It would have been better to do a repeated measures analysis by collecting many observations for each country on several days. However, longitudinal generalized linear mixed modelling is far more complex and this was taken to be a preliminary study.

Though the authors sort to predict the responses from the models this was not possible as obviously there were far more unidentified explanatory variables that would affect the response. The unavailability of other covariates was a problem but the models showed good fit with only the climate parameters used.

Further work

If all the countries of the UN could be obtained and more weather parameters could be selected, a larger scale more comprehensive study could be carried out. Also a longitudinal study can be attempted.

Some regions like North America had only two sources namely, USA and Canada. This was too little to draw conclusions about this region. Thus more areas should be selected within North America.

References

1.Choi, Y., Tang, C. S., McIver, L., Hashizume, M., Chan, V., Abeyasinghe, R. R., Iddings, S., & Huy, R. (2016). Effects of weather factors on dengue fever incidence and implications for interventions in Cambodia. BMC public health, 16, 241. https://doi.org/10.1186/s12889-016-2923-2

2.Ehelepola, N.D.B., Ariyaratne, K., Buddhadasa, W.M.N.P. et al. A study of the correlation between dengue and weather in Kandy City, Sri Lanka (2003 -2012) and lessons learned. Infect Dis Poverty 4, 42 (2015). https://doi.org/10.1186/s40249-015-0075-8

3.Vishnampettai G. amachandran, Priyamvada Roy, Shukla Das, Narendra Singh Mogha, Ajay Kumar Bansal (2016). Empirical model for estimating dengue incidence using temperature, rainfall, and relative humidity: a 19-year retrospective analysis in East Delhi. *Epidemiol Health*. 2016;38:e2016052 Published online November 27, 2016 DOI: https://doi.org/10.4178/epih.e2016052

4.https://www.covidvisualizer.com (Retrieved on 26, 27 March, 2020)

5.https://www.ventusky.com/ (Retrieved on 26, 27 March, 2020)

5.Wimarsha Jayanetti, Roshini Sooriyarachchi (2015). A multilevel study of dengue Epidemiology in Sri Lanka: modeling survival of dengue patients. International Journal of Mosquito Research 2015; 2 (2): 94-101.

6.https://shodhganga.inflibnet.ac.in/bitstream/10603/112402/8/08_chapter%203.pdf (Retrieved on 15th April, 2020)

7.R.A. Johnson and D.W. Wichern Applied Multivariate Statistical Analysis (2008) Prentice and Hall, New Jersey.

8.J. Jiang (2007). Linear and Generalized Linear Mixed Models and their Applications. Springer.

9.C.E. McCulloch and S.R. Searle (2001). Generalized, Linear and Mixed Models, John Wiley and Sons.

10.http://www.epid.gov.lk/web/index.php?option=com_content&view=article&id=225&Itemid=487&lang=en (Retrieved on 10th April, 2020).

11.https://www.who.int/health-topics/coronavirus (Retrieved on 10th April, 2020)

12.H. Lin, L.Yang, Q. Liu, L. Tian (2011). Time series analysis of Japanese encephalitis and weather in Linyi City, China. *International Journal of Public Health* 57(2):289-96.

13.S.M. Fernando and M.R. Sooriyarachchi (2018), Bivariate Negative Binomial Modelling of Epidemiological Data. *Open Science Journal of Statistics and Applications* 5(3): 47-57.