Assessment of Temperature Variables in Modeling Global Mapping and Distribution of COVID-19: A key factor in identification of risk region

Gabriel Salako

Department of Environmental Management and Toxicology, Kwara State University, Malete, Nigeria

Abstract*: Since the outbreak of COVID-19 in Wuhan, China, in December 2019, several works have been done and published on the role of environmental factors, especially climate in general and temperature in particular on the spread of the virus, some of which are contradictory. It has also been observed that most mapping has been overgeneralized without identifying the core infection areas. This work creatively uses distribution models to map the spread and infectivity of COVID-19 using biologically relevant temperature variables. We built ensemble COVID-19 global distribution models by fitting the selected temperature variables with over 650,000 occurrence data of COVID -19 across the globe; the ensemble models combined three algorithms: Maximum Entropy (Maxent), Generalized Linear Model (GLM), and Random Forest (RF) and was implemented in R package "SSDM" using a simple average of each SDM and display in Arc Map. Results show that the mean temperature of the coldest quarter (0-15°c) and annual mean temperature (10-22°c) are the main drivers of the virus's spread. These thresholds defined the level of risk to COVID-19 and were scales between 0 -1, with 0 being low or no risk and one the highest risk. Analysis of the regions at risk by the proportion of areas shows that Western Europe, United States, and mainland China had the most elevated regions under very high risk. At the same time, Africa, except for South Africa and Maghreb nations, was relatively at low risk. Despite few data obtained from Canada, the model predicted a high-risk zone in the eastern provinces of Toronto and Quebec and south of British Columbia. Overgeneralization in mapping was resolved in this work as a high-risk cluster was conspicuously highlighted even in an area of presumably low risk.

***Keywords*: COVID-19, environmental factors, temperature variables, distribution models, risk**

I. INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was first reported in Wuhan, China, in December 2019 which was later officially named by the World Health Organization (WHO) as coronavirus disease 2019 (COVID -19) and declared as public health emergency of international concern (PHEIC), since then more than 20 million cases have been recorded globally as at August 2020 [18]. Although environmental factors had long been identified in epidemiology as one of the three primary factors affecting the transmission of communicable disease [8], there have been controversies on the possible role of environmental factors, especially temperature, on the spread of the virus. Several works have been done and published on environmental factors, especially climate in general and temperature in particular on the spread of the virus. [15]Investigated the relationship between average temperature and transmission of COVID-19 in Sao Paulo city in Brazil and confirmed a negative linear relationship between temperature and COVID-19 transmission, a 1°c rise in temperature results in a reduced confirmed case COVID-19. [1] Their paper concluded that average temperature, minimum temperature, and air quality are highly correlated to the COVID-19 pandemic in New York, United States [16] Examined the relationship between air temperature and SARS-COV and reported that the infection was 18 times higher in colder temperature than in higher temperature. On the contrary, [17] in their study in China found an increase of 4.9% as mean temperature rise with the same margin. Although a linear relationship between COVID 19 and temperatures had been established by numbers of authors[13, 5, 1] yet there is no consensus on the nature and pattern of this relationship as various reports have drawn contradictory conclusions, and this could be as a result of disparity in data collection techniques and nature [16].

Traditionally species distribution model (SDM), otherwise known as habitat suitability, has been widely used by ecologists, environmental scientists, and biogeographers to assess the potential distribution range of species either for conservation of threatened and endangered species or for the control of invasive species [7,14]. It has equally been used in the distributional pattern of diseases and vector-carrying diseases [14]. Given data on environmental variables and occurrences of species or organisms, SDM has excellent capacity to predict the possible suitable areas such a species will likely occur. However, to the best of our knowledge, no global mapping of the response of COVID-19 to temperature variables using distribution models has been made. Also, spatial analysis on the distribution of COVID-19 by countries has been overgeneralized without identifying the core or cluster of infection.

We, therefore, creatively used this in principle to predict the possible suitable areas for the spread and distribution of COVID-19 across the globe using relevant temperature variables and the coronavirus disease 2019 (COVID-19) occurrence data obtained from the WHO dashboard [18]. This aims to map and identify vulnerable countries or regions across the globe to novel coronavirus disease 2019 (COVID-19) as a precondition for control and preparedness.

II. MATERIALS AND METHODS

The study included 252 countries and territories across the globe covering from longitudes -180 °W to180 E° W and latitude 90°N -90°W, spanning different climatic zones from cold temperate, temperate in Eurasia in northern Hemisphere to hot humid tropical and subtropical in Africa and South America in Southern hemispheres.

2.1 COVID-19 Occurrence data

COVID-19 occurrence data was mined from World Health Organization (WHO), Health mapper dashboard in June 2020. https://healthmap.org/COVID-19/A total of 650,000 point occurrence data was curated after removing data duplication and multiple entries from the downloaded data across the globe. However, fewer records were obtained for Africa, India, and Brazil at this work. Therefore, we suggested that cautions should be applied in the interpretation of the results in these areas.

2.2 Selection of Temperature Variables

To remove bias or preference for any variables, all temperature-related variables at the global level were considered for this analysis and were obtained from the Worldclim database (www.worldclim.org) at a spatial resolution of 30 arcs seconds equivalent to 1km2. (Table 1). Multicollinearity has been reported to be a significant issue that could result in model overfitting and large predictive errors [10,3]; however, to remove uncertainty caused by multicollinearity among these variables, we performed variable inflation factor (VIF) analysis using vifstep in R package, variables with VIF values more significant than ten was set as the threshold. We finally selected four variables which are ≤ 10, namely temperature variables: Bio1 = Annual mean temperature, Bio6 = Min Temperature of Coldest Month, Bio10 = Mean Temperature of Warmest Quarter, Bio11 = Mean Temperature of Coldest Quarter which satisfies the minimum vif values threshold set at ≤ 5 [2, 11].

Table 1 Temperature variables

S/No	Temperature Variables	Measurement	Source
1	Bio1 = Annual mean temperature	Degree Celsius	www.worldclim.org
2	Bio2 = Mean diurnal range (max temp – min temp) (monthly average	Degree Celsius	www.worldclim.org
3	Bio4 = Temperature Seasonality (Coefficient of Variation)	Dimensionless	www.worldclim.org
4	Bio5 = Max Temperature of Warmest Month	Degree Celsius	www.worldclim.org
5	Bio6 = Min Temperature of Coldest Month	Degree Celsius	www.worldclim.org
6	, Bio7 = Temperature Annual Range (Bio5-Bio6)	Degree Celsius	www.worldclim.org
7	Bio8 = Mean Temperature of Wettest Quarter	Degree Celsius	www.worldclim.org
8	Bio9 = Mean Temperature of Driest Quarter	Degree Celsius	www.worldclim.org
9	Bio10 = Mean Temperature of Warmest Quarter	Degree Celsius	www.worldclim.org
10	Bio11 = Mean Temperature of Coldest Quarter	Degree Celsius	www.worldclim.org

2.3 Distribution Modeling and Analytical Procedures

Ensemble Modeling has been used to address the uncertainty associated with using single algorithms in predictive modeling and has produced better results [6, 10]. We built the ensemble COVID-19 global distribution models befitting the selected temperature variables with 65,000 occurrence data of COVID-19; the ensemble models combined three algorithms: Maximum Entropy (Maxent), Generalized Linear Model (GLM), and Random Forest (RF) and was implemented in R package "SSDM" using a simple average of each SDM and display in Arc Map. The area's suitability for COVID-19 infection was scaled from 0-1, with 0 being the lowest and 1 being the highest. We used the same scale to measure the COVID-19 risk level and the proportion of landscape under the risk. Model performance was assessed using Area under the Curve (AUC) of the Receiver Operating Characteristic (ROC) and Cohen's KAPPA values, the closer the AUC values (0 ≤ 1) to 1, the more the accuracy of the model to predict the suitability of COVID 19 incidence [12]. We generated the relative contribution of the selected temperature variables in % from the ensemble model.

In contrast, the response curve was generated explicitly from the Maxent algorithm to explain the relationship between the responses of COVID-19 to the variables. We ran correlation analysis to determine the pattern of connection between our model output and the incidence of COVID-19. Zonal histogram tool was used un Arc Map to calculate the proportion of land area under different risk classes to COVID-19.

III. RESULTS

3.1 Projected distribution of COVID-19 globally

With the AUC and Cohen’s Kappa scores of 0.91 and 0.87, respectively, our model was considered very good in its prediction accuracy. Furthermore, the general analysis showed that the cool region of Western Europe, the U.S. (notably East and West coast, Midwest states), and China were predicted to have high incidence or suitable areas for the spread of COVID-19. In contrast, Africa's hot and humid region with higher temperatures (except South Africa and the Maghreb) had low or mild predicted distribution (Fig. 1). Also, the correlation between the result of our model, distribution/infectivity of COVID-19, and the incidence data showed a positive high correlation R=0.93 (Fig.2).

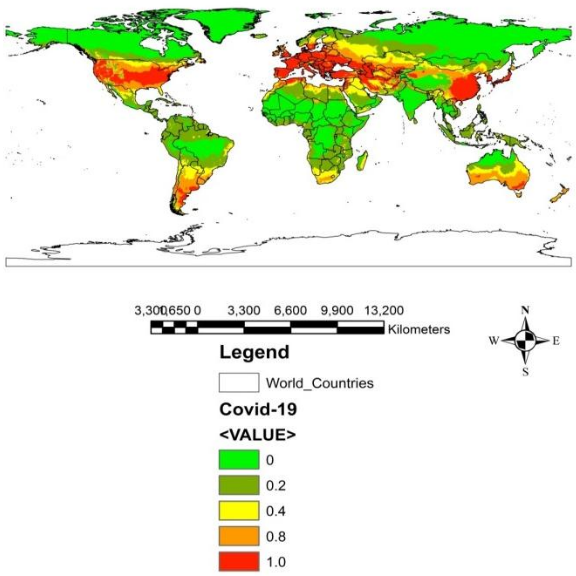

Fig. 1 Distribution of COVID-19 across the globe

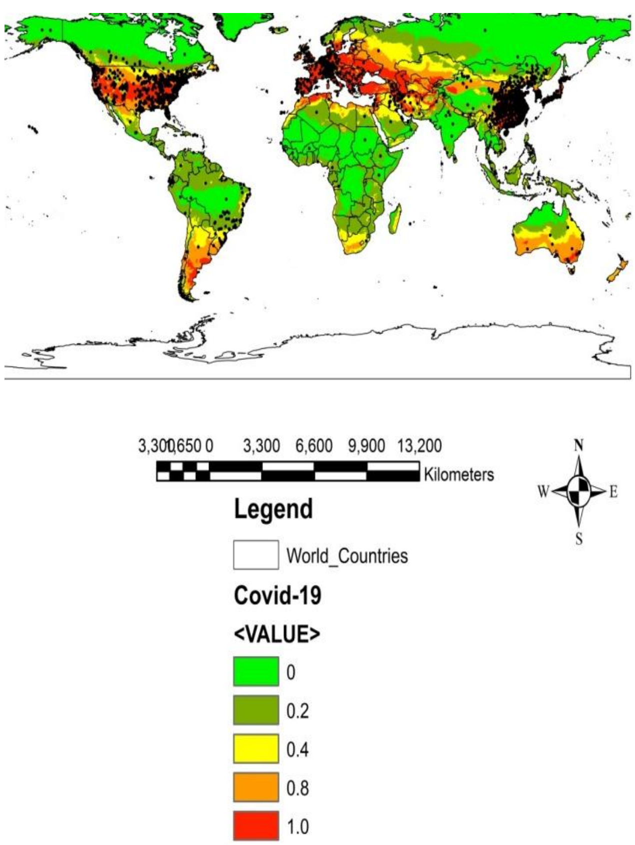

Fig. 2 COVID -19 occurrence data overlay on the projected distribution

Available data in August 2020 showed that the Maghreb nation of Algeria had a confirmed case of over 41,000, active case of 11,000 and 1446 deaths (http://COVID19.sante.gov/carte/), making it one of the hardest-hit country in Afro Arab region.

The relative contributions of the selected four biologically relevant temperature variables showed that Bio11 = Mean Temperature of Coldest Quarter and, Bio1 = Annual mean temperature contributed 45.4 and 43.8 respectively, with both contributed over 88% of the total variation. Others include a minimum temperature of the coldest month and a mean temperature of the warmest quarter with a total contribution of about 12% (Table 2). To better understand the ecological niche of COVID-19 across the globe, we examined the response curve; this explains the quantitative relationship between the environmental variables and the suitability of a species or organisms to a particular environment. One would have expected the minimum temperature of the coldest month to have contributed significantly to the spread of COVID 19, and this variable contributed only 6%. This, therefore, showed that the temperature of a single month alone could not have been enough to structure the spread of the virus but the average values of consistent, suitable temperature of at least three months in a year.

Table 2. Selected temperature variables contribution by percentage

S/No	Variable	Percent Contibution
1	Mean temperture of coldest quarter	45.4
2	Annual mean tempearature	43.8
3	Minimum tempearture of coldest month	6
4	Mean temperature of warmest month	4.8

3.2 Response of COVID-19 to temperature variables

Analysis of the response curve revealed that the mean temperature of the coldest quarter, which is the essential temperature variable, significantly affect the distribution of COVID 19 in the study; COVID -19 begins to rise at -10°c peak at 0-8° c and begin to decline when the values reach 10°c (Fig.3). Likewise, an annual mean temperature of 10-22°c was predicted to be favorable for the spread of the virus (Fig. 4).SARS virus has been found to have remained infectious for five days at a temperature between 20-220c, beyond which the infectivity reduced. [5]

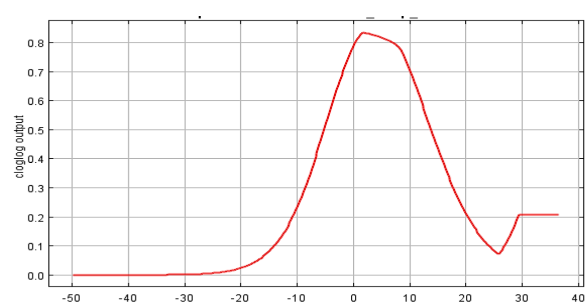

Fig. 3 Response of COVID -19 to mean temperature of coldest quarter; the curve shows COVID-19 peak between 0.0 -10 °c and correspond to the legend

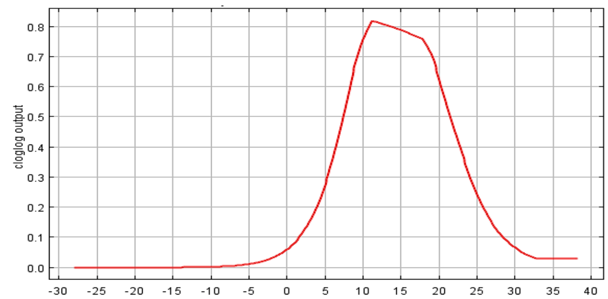

Fig. 4 Response of COVID-19 to annual mean temperature; the curve shows COVID -19 peak between 10 - 22°c as correspond to the legend

3.3 Risk Zones by proportion of landscape

The global analysis by regions at risk of COVID -19 as projected showed that Europe had about 90% of its landscape at high and very high risk of the virus notably among them are: Spain, France, Italy, and United Kingdom (UK) (Fig 5) however, the northern part of UK was shown to be relatively at less high. In Asia, the risk area is between 30-40%, except for China which had almost 50% of its total area under high risk. 50-60% of U.S. are at high and very high risk of the virus bifurcating as midline from the east coast (New York) to the west coast (California). Despite few data obtained from Canada, the model predicted a high-risk zone in the eastern provinces of Toronto and Quebec and south of British Columbia. Africa was at the lowest risk with less than 20%, except in South Africa and the Maghreb, where up to 50% of the areas are under high risk (Fig.5).

Table 3 showed the proportion of COVID-19 risk in selected countries. It should be noted that although a low percentage was calculated in some regions with high incidence case of COVID-19, such as Brazil and Algeria, this is because the more significant portion of population concentrations was confined to certain parts of the regions, and this was demarcated as a cluster of high risk in the map.

Table 3 Proportion of land area of selected countries by continents under different classes of COVID-19risk

S/No	COVID-19 Risk level	U.S.	%	China	%	Italy	%	Spain	%	Brazil	%	Algeria	%
1	Very low	129760	20	124981	22	185	1	0	0	167656	41	30814	25
2	Low	45998	7	87524	15	578	3	43	0.15	191490	47	67872	55
3	Moderate	74692	13	69549	13	574	2.9	109	0.35	36268	8	7881	6
4	High	162647	25	112623	22	849	5	3936	12.60	11793	3.9	3565	3
5	Very high	230113	35	152845	28	16831	89	26756	86.90	50	0.1	12842	11
6	Total	643210	100	547522	100	18997	100	30844	100	407212	100	122974	100

Italy	%	Spain	%	Brazil% Algeria%	%	Algeria	%
185	1	0	0	167656 41 30814 25	41	30814	25
578	3	43	0.15	191490 47 67872 55	47	67872	55
574	2.9	109	0.35	36268 8 7881 6	8	7881	6
849	5	3936	12.60	11793 3.9 3565 3	3.9	3565	3
16831	89	26756	86.90	50 0.1 12642 11	0.1	12842	11
18997	100	30844	100	407212 100 122974 100	100	122974	100

- Value shown is in pixel × 16 .5 km^2

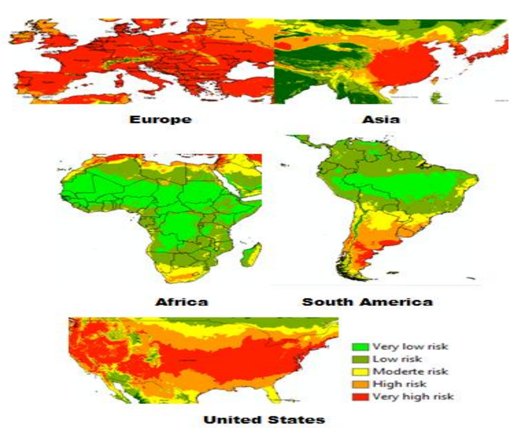

Fig. 5 COVID-19 global risk zones by regions

The Maghreb nations, in particular, could be as high as 80% as most of the concentration of population is along the coast, which invariably coincides with high-risk areas. Although Latin America as a continent was projected to be relatively at low risk compared to Europe and the U.S., a cluster of high-risk inn Argentina and southern Brazil include Sao Paolo, could have been responsible for the high incidence rate in these areas.

IV. DISCUSSIONS

Climate, especially temperature and humidity, played an essential role in the stability and survival of viruses outside the body system; its ability to remain in the environment determines its capacity to infect people and instead become epidemic on a regional scale or pandemic at the global level. Cold winter and low humidity support the survival of the virus as it persists for a longer time in the air due to its smaller sizes

[16]. The human immune system has also been related to weather and climate, as cold winter was associated with the low supply of vitamin D required by the bodies to fight infections. Coronavirus shares similar characteristics with other enveloped viruses; oily coated viruses susceptible to heat and seasonal. Severe Acute Respiratory Syndrome (SARS), another coronavirus outbreak in 2003, has been found to survive and active in cool and drier climates [5]. Their study on the survival rate of SARS under different temperature and humidity conditions reported that the Survival rates of the HCoV 229E were about 3% if exposed to the temperature of 20°c and humidity of 80% in 24 hours. [13] 2020 on the effect of temperature on the infectivity of COVID-19 in Japan found that a colder temperature of -2.1c in Hokkaido had the reported cases of 26.7 per 1, 000,000 despite its relatively lower population with other cities as compared to I case in warmer temperature in Fukushima. These findings agreed with our results that at an annual mean temperature of 20°c and above, it makes the environment unsuitable for the survival and spread of a virus-like COVID 19. Also, [15] in their study of COVID 19 relationship with temperature in Sao Paulo affirmed that each 1°c rise in temperature result in a decrease in confirmed cases of the virus but concluded that the decline could not be guaranteed when the temperature becomes warmer above 25.8c this supported our model findings that at a temperature above 26 the graph of COVID 19 flatten. Although we had fewer data for Brazil and other Latin Africa to run the model, yet the result of the model, as confirmed by the current COVID 19 global distribution, shows that significant incidence cases would be found in the south and western corner of the continent; areas considered to fall under sub-tropical or temperate climate. For example, Sao Paulo, the largest city in Brazil, was projected to be one of the epicenters in the region as it falls within the values of the two essential temperature variables: mean temperature of the coldest quarter of 0-15 c and annual mean temperature of 10- 22c where COVID-19 is predicted to be high, and this was further confirmed by the recent data from WHO [18]. This work did not specify which months does any of the region experiences these temperature variables and therefore suggested further results could be done on this on a regional basis.

V. CONCLUSIONS:

Several pieces of evidence have shown that temperature is an essential factor in the spread and infectivity of enveloped viruses in general and, to a greater extent, the novel coronavirus 2019 (COVID-19). However, this study further identifies the two biologically relevant temperature variables, which possibly account for over 80% of the spread: mean temperature of the coldest quarter and annual mean temperature. The threshold for these variables was analyzed to identify regions that are at risk of COVID -19 infection. Areas with a mean temperature of three coldest months in a year within the range of 0-15°c and the annual average temperature of 10-20°c are at high risk. These set thresholds explain some level of diversity in the spread of the virus, especially in a region with multiple climatic conditions such as Brazil and even Africa with tropical, temperate, and Mediterranean's Maghreb, but also set boundaries COVID-19 could be active. Although relative humidity is another essential climatic variable identified by other researchers, the temperature is directly related. It significantly influences relative humidity as any change in the former produces corresponding changes in the latter. This work has also tried to solve the problem of overgeneralization in COVID -19 spread mapping at the global level by identifying the cluster or hotspot areas hitherto in which an entire country or region, regardless of the hotspot or cluster of incidence, is assigned a uniform scale which may pose a severe challenge in identifying the hard-hit area.

However, it should be noted that challenges of the testing protocol may also account for variation in reported or infected cases even in the same climate or temperature zone as other factors, include the level of immunity, previous ailment, and degree of social contact play a significant role in the spread. This study focuses on the survival, stability, and infectivity of the virus under different temperature variables in the environment. It should not be confused with the virus's capacity to cause illness when exposed to the human body.

REFERENCES

[1] Bashir MF, Ma B, Bilal, Komal B, Bashir MA, Tan D. 2020. Correlation between climate indicators and COVID-19 pandemic in NewYork,USA.SciTotalEnviron.728:138835.doi:0.1016/j.scitotenv.2020.

[2] Craney, T.A., &Surles, J.G. 2002. Model-Dependent Variance Inflation Factor Cutoff Values. Quality Engineering 14: 391–403.

[3] Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B.,

[4] Skidmore, A.K., Zurell, D., &Lautenbach, S. 2012. Collinearity: a review of methods to deal with it and a simulation study evaluating theirperformance.Ecography.doi:10.1111/j.1600-87.2012.07348.x

[5] Geller Chloe, VarbanovMihayl, and DuvalE. Raphael2012. Human Coronaviruses: Insights into Environmental Resistance and Its Influence on the Development of New Antiseptic Strategies.Viruses 2012, 4: 3044-3068; doi:10.3390/v4113044

[6] Gritti, E. S., Duputie, A., Massol, F., &Chuine, I. (2013). Estimating consensus and associated uncertainty between inherently different species distribution models. *Me thods in Ecology and Evolution*, *4*: 442–452.

[7] Kumar Sunil and Stohlgren Thomas J. (2009) Maxent modeling for predicting suitable habitat for threatened and endangered tree *Canacomyricamonticola*in New CaledoniaJournal of Ecology and Natural Environment Vol. 1(4): 094-098, Available online at http://www.academicjournals.org/JENE

[8] Lowen, Anice C and Steel John 2014. Roles of Humidity and Temperature in Shaping Influenza Seasonality Journal of Virology Volume 88 Number 1:. 7692–7695

[9] Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K., &Thuiller, W.2009. Evaluation of consensus methods in redictive species distribution modelling. *Diversity and Distributions, 15*: 59–69.

[10] Mason, R.L., Gunst, R.F., & Hess, J.L. 2003. Statistical design and analysis of experiments: with applications to engineering and science. John Wiley & Sons.

[11] Nneji LM, Salako G, Oladipo SO, Abiodun B. OnadekoBabatunde E. AdedejiOlatundeOmotosoAdiahaAlda A. UgwumbaAdeniyi C. Adeola.2019. Species Distribution Modelling predicts habitat suitability and reduction of suitable habitat under future climatic scenario for *Sclerophrysperreti*: A critically endangered Nigerian endemictoad.*AfrJEcol*.2019;00:111.https//doi.org/10.1111/aje.127 13

[12] Salako, G., Chandalin, B., Aliyu, M. B., & Sawyerr, H. 2015. Modeling the suitability index of selected conifers on Mambilla Plateau Taraba State, Nigeria: Implication on planted forest. *International Journal of Agroforestry Remote Sensing and GIS*, *1*(1), 1–9

[13] Ujiie M, Tsuzuki S Ohmagari O (2020) Effect of temperature on the infectivity of COVID-19. Internatinal Journal of Infectious diseases doi.org/10.1016/j.ijid.2020.04.068

[14] UrbaniFabrizia, D'alessandro Paola, Biondi Maurizio 2017. Using Maximum Entropy Modeling (MaxEnt) to predict future trends in the distribution of high altitude endemic insects in response to climate change Bulletin of Insectology70 (2)

[15] Prata ND, Rodrigues W Bermejo P.H 2020. Temperature significantly changes COVID-19 transmission in (sub)tropical cities of Brazil Science of The Total EnvironmentVolume 729, 10 August 2020, 13886

[16] Yuan Shu, Jiang Si-Cong and Li Zi-Lin 2020. Do Humidity and Temperature Impactthe Spread of the Novel Coronavirus? Frontiers in Public health :1-4 doi: 10.3389/fpubh.2020.00240

[17] Xie J, Zhu Y2020. Association between ambient temperature and COVID-19 infection in 122 cities from China. Sci Total Environ. 724:138201. doi: 10.1016/j.scitotenv.2020.138201

[18] WHO: Coronavirus Disease 2019 (COVID-19) situation report – 23. Geneva, Switzerland: World Health Organization; 2020