Spread of SARS-CoV-2 Coronavirus likely to be constrained by climate

Miguel B. Araújo[1,2] and Babak Naimi[3]

[1]*Department of Biogeography and Global Change, National Museum of Natural Sciences, CSIC, Calle Jose Gutierrez Abascal, 2, 28006, Madrid, Spain.*

[2]*Rui Nabeiro Biodiversity Chair, MED Institute, University of Évora, Largo dos Colegiais, 7000, Évora, Portugal.*

[3]*Department of Geosciences and Geography, University of Helsinki, 00014, PO Box 64, Helsinki, Finland*

As new cases of SARS CoV-2 (*aka* 2019-nCoV) Coronavirus are confirmed throughout the world and millions of people are being put into quarantine, few doubt the virus will reach pandemic state. Some worry it could badly hit the developing world, such as sub-Saharan Africa, potentially leading to a global human calamity. It is still early days, but using existing data we develop a large ensemble of ecological niche models that project monthly variation in climate suitability of SARS-CoV-2 Coronavirus throughout a typical climatological year. The current spread suggests a degree of climate determination with Coronavirus displaying preference for cool and dry conditions. The predecessor SARS-CoV was linked to similar climate conditions. Should the spread of SARS CoV-2 continue to follow current trends, a worst-case scenario of synchronous global pandemic is improbable. More probable is the emergence of asynchronous seasonal global outbreaks much like other respiratory diseases. People in temperate warm and cold climates are more vulnerable. Those in arid climates follow next in vulnerability, while the disease will likely marginally affect the tropics. Our projections minimize uncertainties related with spread of SARS CoV-2, providing critical information for anticipating the adequate social, economic and political responses.

Introduction

Biogeography studies the patterns and processes underlying the distribution of Life on earth. One generalization emerging from hundreds of years of natural history observations is that all organisms have a degree of environmental specialization. That is, while biomes in the planet have a range of different types of organisms[1], individual types of organisms cannot occur in every biome even when, distant apart as they might be, they converge into playing the same ecological roles within ecosystems[2]. Biogeographers and ecologists alike resort to the concept of ecological niche[3-5] to examine the relationship between the distributions of organisms and other biotic or abiotic factors controlling them. An organism is said to be within its ecological niche if death rates of the organism are lower that birth rates[6,7]. That is, an organism cannot persist beyond its ecological niche, in a sink, unless there is a regular influx of individuals from source populations. Even if organisms are regularly reaching a sink area, as one might expect with an easily dispersed pathogen, the spread and establishment of the organism will be limited by ecological constraints. Although biogeographic concepts, such as the species ecological niches, are commonly used and applied to multicellular organisms (eukaryotes), there is an increased number of studies utilizing the ecological niche concepts and associated analytical tools to investigate relationships between the distributions of unicellular organisms (prokaryotic), or viruses, and a range of environmental factors[8].

Building on the concept of ecological niche, we develop projections of monthly changes in the likelihood of SARS-CoV-2 Coronavirus outbreaks. Projections are obtained from an ensemble of 10 familiar machine learning and statistical ecological niche models[9], each with 20 copies generated with bootstrapping to account for and enable the quantification of intra-model variability to the initial conditions[10,11]. Models were trained using the distribution of all recorded SARS-CoV-2 Coronavirus cases by the 10/03/2020 with data compiled and made available to the John Hopkins University Mapping 2019-nCoV portal[12]. Regions with fewer than 5 positive cases were not included in the models. Exclusion of such sites was based on the working assumption that sites with small numbers

of positive cases are likely imported from infected regions, thus failing to provide evidence that the SARS-CoV-2 Coronavirus is being transmitted locally within its ecological niche. Predictors were temperature and precipitation values expected between January and March using 1979–2013 as reference, and with data downloaded from the high-resolution climatology database for the earth's land surface[13]. Models were then projected monthly for the rest of the year.

Results

Analysis of all positive cases of SARS-CoV-2 Coronavirus plotted against monthly temperature and precipitation values reveals that the interquartile range of average environmental temperatures associated with positive cases so far is between -4,01ºC to 15,58ºC (99% range) and -2,04ºC to 9,49ºC (95% range). For precipitation, the interquartile range ranges from 4,68 mm to 116,06 mm (99% range) and 19,75 mm to 94,43 mm (95% range). These values are estimated taking into account total numbers of positive cases, which are obviously strongly determined by contingent factors linked with the origin of the SARS-CoV-2 Coronavirus outbreak (the city of Wuhan in China) and subsequent pattern of spread. While the pattern of spread is, as it seems based on our analysis, constrained by climate, the actual numbers of positive cases are affected by non-climatic factors[14], some of which might be stochastic. Less sensitive measurements can be obtained by using presence absence of positive cases. With such an approach, the estimated interquartile ranges for temperature is -18,10ºC to 28,64ºC (99% range) and -8,81ºC to 25,65ºC (95% range). For precipitation it is 1,00 mm to 345,55 mm range (99% range) and 2,16 mm to 151,31 mm (95% range). Regardless of the approach used to quantify the climate envelope of the SARS-CoV-2 Coronavirus, we are not characterizing the exact local temperature and precipitation conditions constraining the virus spread but rather determining the type of macro-climate conditions in the places where spreading is occurring. Nevertheless, regardless of whether we calculate environmental preferences of the SARS-CoV-2 Coronavirus using the total numbers of incidences or their presence and absence, it appears the virus favors cool and dry conditions being largely absent under extremely cold and very hot and wet conditions (Figure 1).

We summarized projections by ensembles of ecological niche models by climate zones[15]. The analysis reveals that SARS-CoV-2 strives in warm temperate climates between October to May and cold temperate climates between April and September (Figure 2). Arid environments follow the temperate warm trend of seasonal probability of contracting the SARS-CoV-2 Coronavirus but with generally more moderate levels. Much of the tropics have low levels of climate suitability for spread of SARS-CoV-2 Coronavirus owing to their high temperatures and precipitation (used here as a surrogate for humidity), followed by polar climates, where conditions of extreme cold temperatures seem to be beyond the virus critical minimum tolerance values. In most of such low climate suitability areas, human populations will likely be spared from outbreaks arising from local transmissions (Figure 2).

The analysis of risk provided at the climate zone level (Figure 2), masks the sharp seasonality and the fine-grained regional variation in risk that emerges when analyzing the patterns in geographical space (Figure 3). From June to September, much of higher latitude regions of the southern hemisphere, like Argentina, Australia, Brazil, Chile, New Zealand, and Southern Africa will likely be become exposed to new outbreaks of SARS-CoV-2. Models also project highest latitude regions of the northern hemisphere to be badly hit by the Coronavirus during this period, including Canada and Russia, but also the Scandinavian countries. High elevation areas in the Andes and the Himalayas share the same prospects. Concurrently, areas that, as we speak, are of extreme concern in the northern hemisphere (chiefly Italy, Spain, France, Germany, UK, and USA) should witness a reduction in the incidence of new positive cases SARS-CoV-2 Coronavirus. Beyond September and until the end of May, conditions will be suitable for renewed outbreaks in much of warmer temperate regions of Asia, Europe and North America.

Discussion

Not all viruses are climate determined. HIV/AIDS, for example, is not affected by external environmental factors. The virus is transmitted by sexual intercourse, blood transfusions, or from mother to child during pregnancy, delivery or breastfeeding, so it never leaves the host's internal environmental conditions. In contrast, SARS-CoV-2, like other respiratory viruses, namely its predecessor SARS-CoV, involves aerial transmissions of respiratory droplets or fomites, exposing the virus to external environmental conditions.

SARS-CoV-2 Coronavirus has already set foot in most parts of the world, but virulent outbreaks with large numbers of local infections are still not global. Instead, outbreaks concentrate in the northern hemisphere, chiefly Asia, the Middle East, Central, Southern and Western Europe, and the USA. Our models support the view that the incidence of the virus will follow a seasonal pattern with outbreaks being favored by cool and dry weather, while being slowed down by extreme conditions of cold and heat as well as moist. Prevalence of respiratory disease outbreaks, such as influenza, during wintering conditions is common[16,17]. But the similarity of climate determination of SARS-CoV-2 with its predecessor SARS-CoV is noteworthy given hope that fundamental traits shared by the two Coronavirus might be conserved.

Analyses of SARS-CoV outbreaks in relation to meteorology reveal significant correlations between the incidence of positive cases and aspects of weather. For example, an initial investigation linking SARS outbreaks and temperature in Hong Kong, Guangzhou, Beijing, and Taiyuan[18], revealed significant correlations between SARS-CoV incidences and temperature seven days (the known period of incubation of SARS-CoV) before the outbreak, with environmental temperatures associated with positive cases of SARS-CoV ranging between 16ºC to 28ºC. They also found that incidence of the Coronavirus was inversely related to humidity. Another study conducted between 11 March and 22 May 2003 in Hong Kong[19] showed that SARS-CoV incidences sharply decreased as temperature increased from 15ºC to 29ºC, after which it practically disappeared. In this study,

incidences under the cooler end of the gradient were 18-fold higher than under the opposite warmer end of the gradient.

The mechanism underlying these patterns climate determination is likely linked with the ability of the virus to survive external environmental conditions prior to reaching a host. For example, a recent study examined survival of dried SARS-CoV Coronavirus on smooth surfaces and found that it would be viable for over 5 days at temperatures ranging between 11-25ºC and relative humidity of 40-50%, drastically loosing viability as temperatures and humidity increased[20]. Heat intolerance of the Corona viruses is probably related to their being covered by a lipid bilayer[21,22], which could breakdown easily as temperatures increase. Humidity in the air is also expected to affect the transmissibility of respiratory viruses. Once the pathogens have been expelled from the respiratory tract by sneezing, they literally float in the air and they do so for a longer period when the humidity is greater.

More detailed examination of SARS-CoV-2 outbreak relationships with weather events will only be possible once the spread of the virus has stabilized. The current macroecological-level analysis enables inferences that would otherwise not be possible with high-resolution data for specific case studies. That is, substitute the familiar analysis of meteorological variation at site levels matched with specific SARS-CoV-2 cases by examination of all known positive cases worldwide against an analysis of large-scale climatological variation. It is, obviously, possible that, as the virus spreads and additional climate regions witness outbreaks of positive cases, inferences made herein are altered. We are skeptical this will happen for two reasons. Firstly, there is little reason to suspect that out-of-China contaminations would have occurred only, or mainly, with trade partners in the northern hemisphere[14]. China is a big world player, having key commercial partnerships with Africa and Latin America. Yet there is not indication that meaningful local infections have taken place in these areas despite the global reporting of Coronavirus cases generally attributed to travellers coming from infected regions. Secondly, the climatic discrimination of the outbreaks is such that is seems unlikely to be a consequence of random

chance, trade preferences with China, or just the outcome of poorly developed public policies. On the contrary, the SARS-CoV-2 Coronavirus, although being still expanding, seems to have followed closely the expected (given what we know of SARS-CoV) pattern of climate suitability. This suggests the Coronavirus might have reached equilibrium with climate[23,24], which, if true, would have contributed to significantly reduce the unaccounted for data biases and uncertainties entering our ecological niche models[25].

Understanding the underlying factors involved in the successful spread of SARS-CoV-2 Coronavirus is critical to manage the timing and scale of the social, economic, and political reactions to it. While the Coronavirus is likely to spread much more widely than at present, owing to the seasonal changes of climate suitability, it is unlikely to do so with the same intensity, simultaneously. Our results will allow anticipate the timing and the magnitude of the likely public interventions to mitigate the adverse consequences of the Coronavirus on public health. Only with adequate planning will unnecessary collateral damages be imposed on individuals and the global economy.

References

1 Holt, B. G. *et al.* An Update of Wallace's Zoogeographic Regions of the World. *Science* **339**, 74-78, (2013).

2 Mendoza, M. & Araújo, M. B. Climate shapes mammal community trophic structures and humans simplify them. *Nature Communications* **10**, 5197, (2019).

3 Peterson, A. T. *et al. Ecological Niches and Geographical Distributions*. (Princeton University Press, 2011).

4 Soberón, J. & Nakamura, M. Niches and distributional areas: concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences USA* **106**, 19644-19650, (2009).

5 Chase, J. M. & Leibold, M. A. *Ecological niches - Linking classical and contemporary approaches*. (The University of Chicago Press, 2003).

6 Hutchinson, G. E. Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology* **22**, 145-159, (1957).

7 Pulliam, H. R. On the relationship between niche and distribution. *Ecology Letters* **3**, 349-361, (2000).

8 Murray, K. A., Olivero, J., Roche, B., Tiedt, S. & Guégan, J.-F. Pathogeography: leveraging the biogeography of human infectious diseases for global health management. *Ecography* **41**, 1411-1427, (2018).

9 Naimi, B. & Araújo, M. B. sdm: a reproducible and extensible R platform for species distribution modelling. *Ecography* **39**, 368-375, (2016).

10 Araújo, M. B. & New, M. Ensemble forecasting of species distributions. *Trends in Ecology and Evolution* **22**, 42-47 (2007).
11 Araújo, M. B. *et al.* Standards for distribution models in biodiversity assessments. *Science Advances* **5**, eaat4858, (2019).
12 Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*.
13 Karger, D. N. *et al.* Climatologies at high resolution for the earth's land surface areas. *Scientific Data* **4**, 170122, (2017).
14 Chinazzi, M. *et al.* The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, eaba9757, (2020).
15 Peel, M. C., Finlayson, B. L. & McMahon, T. A. Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* **11**, 1633-1644, (2007).
16 Lowen, A. C., Mubareka, S., Steel, J. & Palese, P. Influenza Virus Transmission Is Dependent on Relative Humidity and Temperature. *PLOS Pathogens* **3**, e151, (2007).
17 Tamerius, J. D. *et al.* Environmental Predictors of Seasonal Influenza Epidemics across Temperate and Tropical Climates. *PLOS Pathogens* **9**, e1003194, (2013).
18 Tan, J. *et al.* An initial investigation of the association between the SARS outbreak and weather: with the view of the environmental temperature and its variation. *Journal of Epidemiology and Community Health* **59**, 186, (2005).
19 Lin, K., Yee-Tak Fong, D., Zhu, B. & Karlberg, J. Environmental factors on the SARS epidemic: air temperature, passage of time and multiplicative effect of hospital infection. *Epidemiol Infect* **134**, 223-230, (2006).
20 Chan, K. H. *et al.* The Effects of Temperature and Relative Humidity on the Viability of the SARS Coronavirus. *Advances in Virology* **2011**, 7, (2011).
21 Raamsman, M. J. B. *et al.* Characterization of the Coronavirus Mouse Hepatitis Virus Strain A59 Small Membrane Protein E. *Journal of Virology* **74**, 2333, (2000).
22 Schoeman, D. & Fielding, B. C. Coronavirus envelope protein: current knowledge. *Virology Journal* **16**, 69, (2019).
23 Svenning, J.-C. & Skov, F. Limited filling of the potential range in European tree species. *Ecology Letters* **7**, 565-573, (2004).
24 Araújo, M. B. & Pearson, R. G. Equilibrium of species' distributions with climate. *Ecography* **28**, 693-695, (2005).
25 Araújo, M. B. & Peterson, A. T. Uses and misuses of bioclimatic envelope modeling. *Ecology* **93**, 1527-1539, (2012).
26 McCullagh, P. & Nelder, J. A. *Generalized Linear Models*. 2nd edition edn, (Chapman and Hall, 1989).
27 Hastie, T. J. & Tibshirani, R. *Generalized additive models*. (Chapman and Hall, 1990).
28 Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. *Classification and regression trees*. (Chapman and Hall, 1984).
29 Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* **29**, 1189-1232, (2001).
30 Breiman, L. Random forest. *Machine Learning* **45**, 5-32, (2001).

31 Hastie, T. & Tibshirani, R. Discriminant Analysis by Gaussian Mixtures. *Journal of the Royal Statistical Society. Series B (Methodological)* **58**, 155-176, (1996).
32 Guo, Q., Kelly, M. & Graham, C. H. Support vector machines for predicting distribution of Sudden Oak Death in California. *Ecological Modelling* **182**, 75-90, (2005).
33 Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review* **65**, 386-408, (1958).
34 Phillips, S. J., Anderson, R. P. & Schapire, R. E. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**, 231-259, (2006).
35 Friedman, J. H. Multivariate Adaptive Regression Splines. *Annals of Statistics* **19**, 1-67, (1991).
36 Hastie, T., Tibshirani, R. & Friedman, J. H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. (Springer, 2001).
37 Fielding, A. H. & Bell, J. F. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* **24**, 38-49, (1997).
38 Garcia, R. A., Burgess, N. D., Cabeza, M., Rahbek, C. & Araújo, M. B. Exploring consensus in 21st century projections of climatically suitable areas for African vertebrates. *Global Change Biology* **18**, 1253-1269, (2012).

Figures

Figure 1 | Frequency distribution of SARS-CoV-2 positive cases plotted against the world gradient of mean temperature (A) and precipitation (B) in a typical climatological series between January and March.

Figure 2 | Projected relative climate suitability for SARS-CoV-2 Coronavirus outbreaks across the 5 Köppen–Geiger climate zones of the world[15]. (A) Distribution of coarse Köppen–Geiger climate zones. (B) Monthly changes in relative climate suitability for the Coronavirus per climate zone.

Figure 3 | Projected climate suitability for SARS-CoV-2 Coronavirus outbreaks in a typical January-March (A), April-June (B), July-September (C), and October-December (D). A gif with the monthly projections is available at http://www.maraujolab.com/wp-content/uploads/2020/03/corona_risk.gif

Methods

SAR-CoV-2 Coronavirus data

We downloaded the geo-referenced coordinates of SAR-CoV-2 Coronavirus cases from the data repository (https://github.com/CSSEGISandData/COVID-19/blob/master/README.md) operated by the Johns Hopkins University Center for Systems Science and Engineering with support from ESRI Living Atlas Team and the Johns Hopkins University Applied Physics Lab. The data was downloaded on 08/03/2020.

Climate data

We downloaded temperature (mean, maximum, minimum) and precipitation (accumulated) from CHELSA (Climatologies at high resolution for the earth's land surface areas; http://chelsa-climate.org)[13]. This is a high-resolution (30 arc sec) climate data set for the earth land surface hosted by the Swiss Federal Institute for Forest, Snow and Landscape Research WSL. It provides monthly summaries covering the period starting in January 1979 until December 2013. The time series data were then aggregated monthly though averaging so to provide expected values for a typical climatological month in the recent past.

Ecological niche models

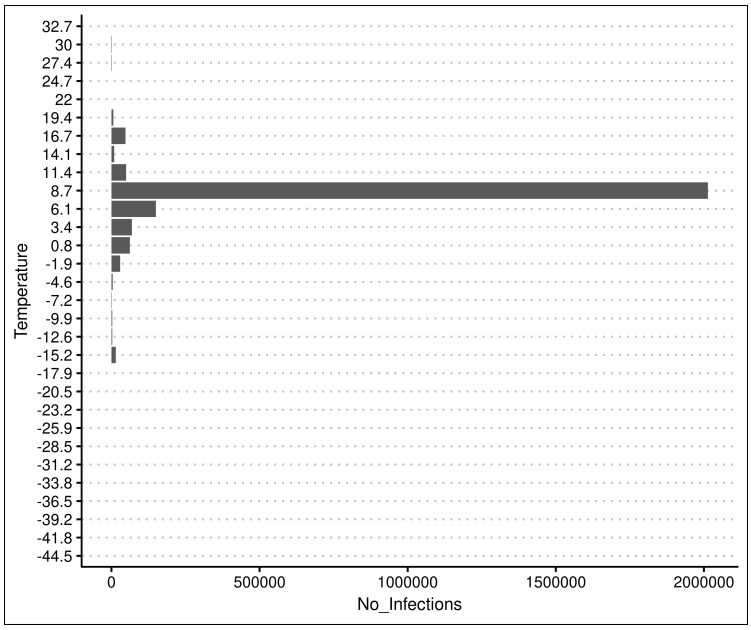
The spatial distribution of incidence SAR-CoV-2 Coronavirus records were linked to the corresponding monthly climate data. We used SDM-R platform[9] for ensemble ecological niche modeling[10] (or species distributions modeling[25]), to characterize climate conditions associated with outbreaks of SARS-CoV-2 between January and March 2020. We used 10 commonly used machine learning methods including generalized linear model (GLM)[26], generalized additive model (GAM)[27], classification and regression trees (CART)[28], boosted regression trees (BRT)[29], random forests (RF)[30], multiple discriminant analysis (MDA)[31], support vector machine (SVM)[32], multi-layer perceptron neural networks (MLP)[33], maximum entropy (Maxent)[34], and multivariate adaptive regression splines (MARS)[35]. We used a bootstrapping resampling procedure[36], with 20 replications, to generate the training and test datasets. We then fitted the ecological niche models for each replication using the training dataset and evaluated them for their performance using the test dataset. We used the area under curve (AUC) of receiver operating characteristic (ROC) plot and the true skill statistic (TSS) to measure the predictive performance of models[37]. A ROC curve plots sensitivity values (true positive fraction) on the y-axis against '1 – specificity' values (false positive fraction) for all thresholds on the x-axis. AUC is a threshold-independent metric that varies from 0 to 1 and provides a single measure of model performance. AUC values under 0.5 indicate discrimination worse than chance; a score of 0.5 implies random predictive discrimination; and a score of 1 indicates perfect discrimination. TSS is calculated as "sensitivity + specificity -1" and ranges from -1 to +1, where +1 indicates perfect agreement, a value of 0 implies agreement expected by chance, and a value of less than 0 indicates agreement worse than chance. We then used the ensemble of 200 models to calculate and project a consensus distribution of spreading risk for each month across the globe. Consensus was achieved through AUC-weighted mean across all models[38].

Acknowledgements
The authors thank and congratulate the John Hopkins University for real-time compilation and release of incidence data for SAR-CoV-2 Coronavirus, without which this investigation would not have been possible.

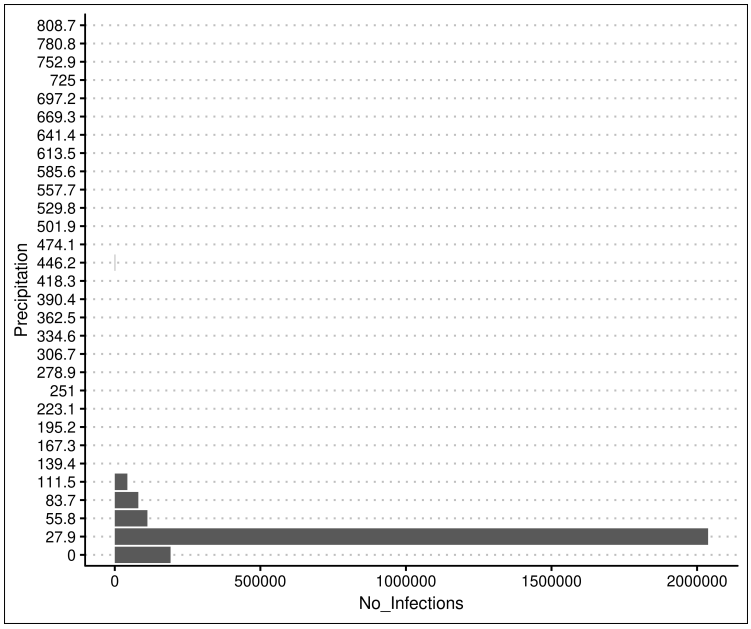
Author contributions
MBA conceived the study and wrote the manuscript. BN conducted the analysis and prepared graphical material.

Author's information
- Reprints and permissions information is available at www.nature.com/reprints
- Competing interests: None to declare.
- Correspondence and requests for materials should be addressed to maraujo@mncn.csic.es

Figures

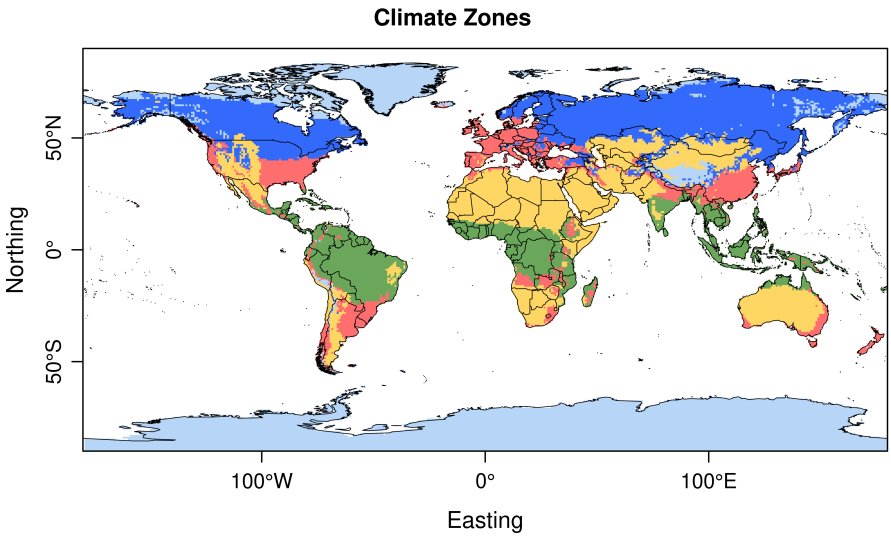


(A)

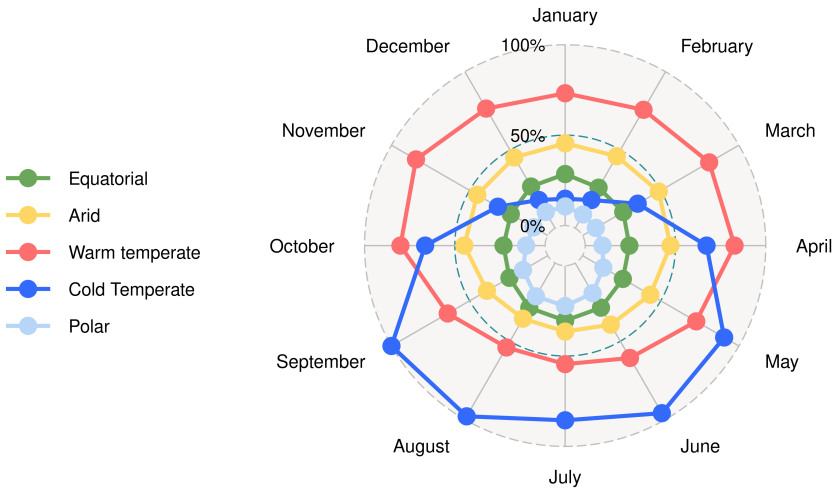


(B)

Figure 1


Climate Zones
Northing
50°N
0°
50°S
100°W
0°
100°E
Easting

(A)


Equatorial
Arid
Warm temperate
Cold Temperate
Polar
January
February
March
April
May
June
July
August
September
October
November
December
100%
50%
0%

(B)

Figure 2

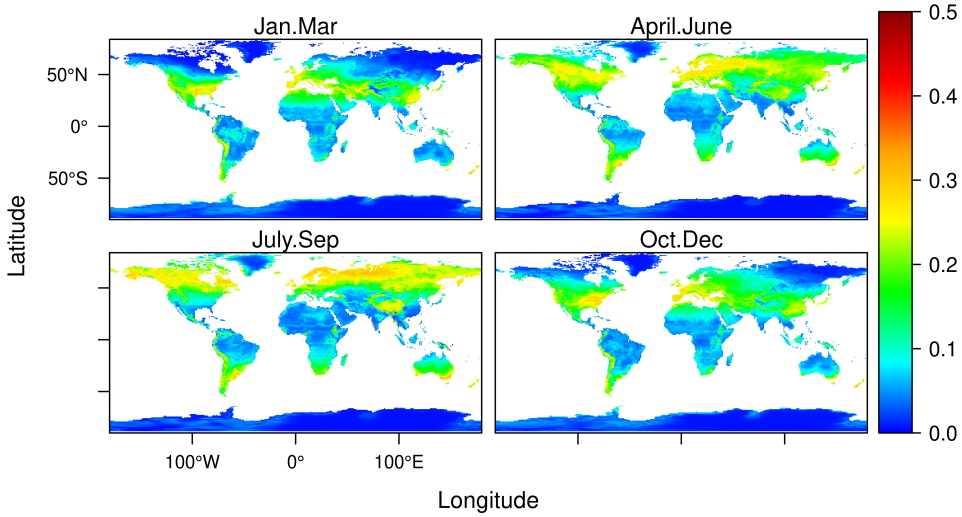

Jan.Mar
April.June
July.Sep
Oct.Dec
50°N
0°
50°S
Latitude
100°W
0°
100°E
Longitude
0.5
0.4
0.3
0.2
0.1
0.0

Figure 3