

On the inadequacy of species distribution models for modelling the spread of SARS-CoV-2: response to Araújo and Naimi

Joseph D. Chipperfield¹, Blas M. Benito², Robert B. O'Hara³, Richard J. Telford⁴, Colin J. Carlson^{5,6}

1. Norwegian Institute for Nature Research, Thormøhlensgate 55, 5006 Bergen, Norway
2. Department of Ecology & Multidisciplinary Institute for Environment Studies "Ramon Margalef", University of Alicante, San Vicente del Raspeig, Alicante, Spain
3. Department of Mathematical Sciences and Centre for Biodiversity Dynamics, Norwegian University of Science and Technology, Trondheim, Norway
4. Department of Biological Sciences, University of Bergen and Bjerknes Centre for Climate Research, Post Box 7803, N-5020 Bergen, Norway
5. Department of Biology, Georgetown University, Washington, D.C., USA
6. Center for Global Health Science and Security, Georgetown University Medical Center, Washington, D.C., USA

The ongoing pandemic of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is causing significant damage to public health and economic livelihoods, and is putting significant strains on healthcare services globally. This unfolding emergency has prompted the preparation and dissemination of the article "Spread of SARS-CoV-2 Coronavirus likely to be constrained by climate" by Araújo and Naimi (2020). The authors present the results of an ensemble forecast made from a suite of species distribution models (SDMs), where they attempt to predict the suitability of the climate for the spread of SARS-CoV-2 over the coming months. They argue that climate is likely to be a primary regulator for the spread of the infection and that people in warm-temperate and cold climates are more vulnerable than those in tropical and arid climates. A central finding of their study is that the possibility of a synchronous global pandemic of SARS-CoV-2 is unlikely. Whilst we understand that the motivations behind producing such work are grounded in trying to be helpful, we demonstrate here that there are clear conceptual and methodological deficiencies with their study that render their results and conclusions invalid.

What follows is a response to the Araújo and Naimi article centered around three main criticisms:

1. Given the fact that SARS-CoV-2 has a primary infection pathway of direct contact, it is in an active spreading phase, and remains largely underreported in the Global South, it represents an inappropriate system for analysis using the SDM framework.
2. Even if we were to accept that an SDM framework would be applicable here, the methodology presented in the article strays far from best-practice guidelines for the application of SDMs.
3. The dissemination strategy of the authors failed to respect the frameworks of risks adhered to in other academic disciplines pertaining to public health, resulting in erroneous but well-publicised claims with broad policy implications before any scientific oversight could be applied.

Criticism 1: Inappropriate application of SDMs for the system

Perhaps the most foundational issue is that species distribution modeling is fundamentally inappropriate for the task the authors are attempting. The ecology, transmission pathways, and spatial epidemiology of the SARS-CoV-2 virus are incompatible with the most basic of assumptions that SDMs make.

SARS-CoV-2 is directly transmitted

Not all pathogens are appropriate for SDM methods; their biological meaning and methodological relevance increases with the level of environmental control over a pathogen or parasite's life cycle. Metazoan parasites with a free living stage, like ticks or tapeworms, have the most readily conceptualized ecological niche, while microparasites (bacteria, viruses, and prions) are a greater challenge. For a vector-borne virus, geographic limits of transmission might be set by the strength of temperature effects on a mosquito biting rate (Shapiro *et al.* 2017), or of relative humidity on tick desiccation (Rodgers *et al.* 2007). There are cases where species distribution modeling has been

helpful for influenza and for other coronaviruses, but only in accounting for where virus is detected in reservoirs or where spillover into humans occurs (Reeves *et al.* 2015; Williams *et al.* 2008; Zhu and Peterson 2014). These are processes with a much clearer set of drivers: for example, the density of intermediate livestock hosts (poultry for influenza, camels for MERS-CoV), or seasonal vegetation and precipitation drivers explaining the distribution and health of waterfowl. For SARS-CoV-2, a novel virus originating in a (presumed) single spillover event with an unknown geographic location (Andersen *et al.* 2020), these methods are again fundamentally inappropriate. Moreover, for the human-to-human outbreak of SARS-CoV-2, ecological determinism becomes less identifiable, and spatial structure of outbreaks is more determined by social contact patterns. This is especially true for human respiratory viruses: climatic influences on aerosol transmission are doubly confounded through microclimate effects, especially if most human-to-human contact occurs indoors.

Producing a model that indicates some regions might not be at risk due to climate misses the fact that, for the most part, all parts of the world are at risk from all respiratory viruses. This isn't to say climate drivers aren't relevant, but instead, that their signal is much less identifiable, and often requires specialized methods that look at incidence (case totals and severity over time and space), rather than geographic occurrence (binary presence/absence). Influenza offers a useful comparison point: temperature and relative humidity are strong drivers of seasonal influenza outbreak severity, duration, and geography (Lowen and Steel 2014), but these influences are still poorly understood and hotly contested, and exploring them has required specialized epidemiological models (Marr *et al.* 2019; Park *et al.* 2019). Drawing an inference that some regions are outside influenza risk based on these findings would be fundamentally inappropriate: influenza remains transmissible anywhere in the world, between any two people.

This doesn't mean climate-based SDMs *cannot* be run. Machine learning methods will almost always pull signals from predictors and outcomes with spatial structure, even when the predictors have no causal link to outcomes (Fourcade *et al.* 2017). In ecology, modeling the presence of any entity based only on climate will certainly lead to the conclusion that climate is a predictor of the presence of said entity, whatever the strength of the effect is. For example, a niche model could be run with identical procedures to the Araújo and Naimi paper with a geolocated database of HIV cases. Despite long-term indirect effects of climate through socioeconomic development on HIV (Burke *et al.* 2015), there is no part of the world unsuitable for HIV transmission. If a niche model were run for HIV, it could indeed perform with high accuracy, and be read by an uninformed ecologist as implying Africa has more suitable climates for HIV than the rest of the world, while the Arctic and Antarctic are too cold for sexual transmission to occur. This would be readily debunked, missing the obvious realities of why HIV prevalence is highest in Africa (weakened public health infrastructure due to colonialism and cultural factors influencing heterosexual transmission). By ascribing climate as the proximal determinant of the distribution of SARS-CoV-2, the authors have here committed a nearly identical fallacy.

The distribution of SARS-CoV-2 is not in equilibrium with climate

Putting aside for the moment the lack of adequacy of SDMs to model the spread of respiratory viruses, there are also major issues with their use to model the spread of an ongoing pandemic. SDMs rely on the key assumption that the target species is in equilibrium with climate, meaning that its distributional data must be a fair representation of the combination of climatic conditions at which the species thrives (Guisan and Zimmermann 2000). However, environmental stochasticity coupled with non-environmental factors such as dispersal limitations and/or historical or biogeographical constraints prevent species reaching an equilibrium status, even after thousands of years of steady dispersal during the climatically stable Holocene (Araújo and Pearson 2005, Svenning and Skov 2004). This is particularly true for invasive species during the introduction and colonization stages of the invasion (Theoharides and Dukes 2007; Phillips *et al.* 2008), when their observed distributions do not match their potential distributions due to lags in range filling related with colonization time lag and/or dispersal limitations. In the context of SDMs, modelling the distribution of the ongoing

SARS-CoV-2 pandemic faces the same very challenges modelling invasive species in the earlier stages of invasion does. These challenges generally involve the modelling of truncated climate envelopes that lead to underprediction of the potential distribution of the target taxon (Václavík and Meentemeyer 2012).

The assumption of equilibrium with climate must be “pondered carefully and stated explicitly”, more so when SDMs are to be transferred to a different time or space (Araújo and Peterson 2012). Araújo and Naimi argue that the distribution of SARS-CoV-2, due to its fast spread, may have reached equilibrium with climate, and therefore, SDMs are suitable to project the risk of spread in the forthcoming months. However, the ongoing changes in the distribution of SARS-CoV-2 (see figure 1) is highly suggestive of an organism that fulfills all the standard criteria of not being in distributional equilibrium with its climate.

To test the authors’ assumptions of equilibrium, we run one of the SDMs used in the ensemble forecast of Araújo and Naimi using data at different stages of the epidemic. We used a Generalized Additive Model (Wood 2017) parameterised with all available data at three different stages: the early stage of the epidemic (22nd January 2020), the period during which the Araújo and Naimi manuscript was being prepared (11th March 2020), and two days after the Araújo and Naimi manuscript was placed on a preprint server (18th March 2020). We used the same climatic covariates as the authors did from the CHELSA climatology dataset (Karger *et al.* 2017): the accumulated precipitation and the mean, minimum, and maximum temperature. Using these models we made a prediction of the probability of occurrence of the virus for each cell in the April climatology layer (figure 2).

If the distribution of SARS-CoV-2 really was in equilibrium with climate then the predicted distributions of the virus would be similar. In figure 2 we demonstrate stark differences between the predicted distributions. The difference between the early and manuscript-preparation stages is to be expected as almost two months of confirmed case data was collected between these stages. To this the authors might argue that the virus was not in equilibrium with the climate in January, hence why the April predictions are poor when only using data from this time period, but that now it has settled into an equilibrium. However, this notion is quickly disabused when we consider the continental-scale increases in predicted occurrence in north-central Asia and central North America when we use the full confirmed cases dataset updated only two-days after the authors submitted their article to the preprint server.

Public health interventions are ignored

A final conceptual problem with the application of SDMs to the spread of SARS-CoV-2 is the difference between ecological and epidemiological patterns of occurrence, and the unique nature of epidemic data. Since the origin of a novel coronavirus in Hubei Province, governments at every scale have been working to contain the outbreak. In the most mild cases, this has involved recommendations of social distancing or travel restrictions; in the case of China and Italy, entire cities have been quarantined for extended periods. As for any other human-to-human pandemic, all scientific evidence indicates the strongest determinant of the epidemic’s course will be individual behavioral choices, and the public health measures put in place to influence or encourage them; the effect of climate is almost certainly dwarfed by the mobilization of an international emergency response. The entire world is moving rapidly to contain the pandemic and crucially, strategies vary significantly across countries and within them. Treating occurrence data from countries with this range of control strategies as equal only further obscures biases that might accidentally correlate with climate, and lead to misleading conclusions.

Criticism 2: Flaws in the application of the SDMs

Even in the core fields of application of SDMs such as biogeography and macroecology, there are a number of well-documented and commonly encountered misapplications that have led to a large

variation in the quality of predictions of species' ranges (Thuiller *et al.* 2004). These concerns resulted in the authors contributing to the publication of a paper defining a number of standards that future applications of SDMs should adhere to in order to prevent poorly-grounded inferences and incorrect advice being provided to managers and policy-makers (Araújo *et al.* 2019). But assessing the SARS-CoV-2 analysis under their own assessment criteria, we find the study to be mostly 'deficient' - the lowest grade possible under their classification scheme. Here follows a list of their assessment criteria for which their study fails to meet the required standards and the rationale for such classification.

The response variable has high spatial error and is inadequately sampled

Araújo and Naimi use a response variable that is derived from the number of reported cases in the John Hopkins University Mapping 2019-nCoV portal (Dong *et al.* 2020). The Dong *et al.* (2020) dataset is designed to feed an app dashboard, not to calibrate SDMs, and therefore its geocoding doesn't need to be accurate. In fact, the dataset reports covid19 cases "at the province level in China, at the city level in the USA, Australia, and Canada; and at the country level otherwise". This implies that almost half of the training data used by Araujo and Naimi to fit their SDMs considers the climatic conditions occurring at the geographical centroids of very large and environmentally heterogeneous regions of the world. For example, all the Russian case data is associated with the climate at a single point in the middle of a forest of the Yeniseysky district of Krasnoyarsk Krai; a location far away from any centres of human population. From figure 1 we can see other similar problematic spatial errors in the mapping of confirmed cases of COVID-19 in countries such as Brazil and Greenland.

This particular issue is worsened by the arbitrary criterion adopted by Araújo and Naimi, who assumed that places with fewer than five reported cases would unlikely be representative of the distribution of SARS-CoV-2 and therefore categorised those as 'absences' for the purposes of distribution modelling. Five cases aggregated across a city is much more relevant than five cases aggregated across an entire country and, as is apparent from Figure 1, this naive classification without accounting for the spatial extent of the aggregation results in the removal of nearly all reported cases in the US and Saharan and Sub-Saharan Africa. The latter is particularly telling because it is exactly about these regions that the authors make the bold claim that, according to their model results, they should be considered 'low-risk'.

Moreover, underreporting is a documented issue in many countries, and low case counts can represent early stages of the spread of the coronavirus. Surveillance in Africa is notably incomplete, given both limited diagnostic capacity and high comorbidities with other severe respiratory infections (e.g., meningococcal meningitis in West Africa), and may register as a number of false negatives. Considering that the lack of testing during the early stages of the pandemic has been widespread (Li *et al.* 2020), especially in the Global South, this strategy of treating such data points as 'absences' is risky. While we understand the intention to not include terminal importations (imported cases that do not start outbreaks), better-suited frameworks were developed during the Zika epidemic that directly analyze terminal importations, and extrapolate how climate and epidemiological drivers determine where outbreaks can and cannot be established (Fox *et al.* 2019).

Instead, filtering occurrences with small reported outbreaks (as the authors have) was likely to produce the exact finding the authors achieved: better-funded surveillance systems in the Global North are correlated with temperature, leading to the prediction that Saharan and sub-tropical biomes might be too arid for COVID-19 transmission.

Assessment of sufficiency/adequacy of predictor variables not performed

The authors use one precipitation and three temperature related climatologies from the CHELSA dataset (Karger *et al.* 2017). Even if we put aside, for reasons stated earlier in this paper, that these climate variables are likely to be very distal predictors at best, there are still significant problems with

mismatched spatial scaling between the response and predictor variables. The CHELSA dataset is applied at the relatively fine-scale 30 arc second resolution (resulting in grids of approximately 1km resolution at the equator). However, given the inconsistent nature of the response variable, it is unclear how the authors reconcile these scale issues and harmonise the predictor and response datasets. For those countries represented in the John Hopkins data by a single centroid coordinate (such as Brazil and Russia), is there any attempt to address the fact that cases from these countries can cover a broad range of climatologies? The manuscript is silent about what spatial averaging of the predictor variables was performed, if any.

Additionally, when predictor variables are highly correlated it becomes difficult to assess statistically which of those is the primary driver of changes in the response variable (Dormann *et al.* 2013; De Marco Júnior and Corrêa Nóbrega 2018). This causes two main problems: firstly, this widens the uncertainty associated with the estimation of the parameters and, secondly, this can cause bias and uncertainty of predictions into areas where the correlation structure between the predictor variables breaks down (Engler and Rödder, 2012). Here, Araújo and Naimi use three temperature-related predictor climatologies (mean, minimum and maximum) that, unsurprisingly, are highly correlated (the April temperature-based climatologies have a *minimum* Spearman's correlation value of 0.995). Beyond the regularisation routines employed in a small subset of the models in the ensemble (MAXENT provides internal LASSO regularisation, Phillips *et al.* 2017), there is no attempt by the authors to assess the degree to which such collinearity biases their predictions or to address such collinearity through employing consistent fitting mechanisms or predictor variable pre-processing. None of the recommendations of addressing collinearity concerns in the Dormann *et al.* 2013 review have been applied.

No checking of model assumptions or justification of complexity

The authors test the performance of the models through measures of predictive success (area under the curve and the true skill statistic; Fielding and Bell 1997) calculated on subsets of the dataset. These performance criteria give relative estimates of the predictive performance of the models contained in the ensemble but do not adequately test for situations where the model is under- or overspecified- (i.e. where important predictor variables are absent or where too many extraneous predictors are present respectively). There is no application of methods that give appropriate weightings to penalise for model complexity (such as the information criteria approaches described in Burnham and Anderson 2002) and there is no attempt to perform model checking through analysis of residuals (such as through the deployment of residual checking software such as that in the DHARMA package, Hartig 2020) which may diagnose cases of missing predictor variables. Whilst some of the models employed in the ensemble (such as MAXENT) do have some internal regularisation routines, once these are incorporated into the ensemble it is unclear how their parsimony enforcement is diluted with other models that do no parsimony weighting as part of the fitting process. In either case, regularisation does not help if the model is under-specified.

Criticism 3: Poor dissemination strategy

In addition to the conceptual and technical points mentioned earlier in this response, we believe that it is also very important to address a number of problematic aspects to the way in which the announcement of this research was made available to the public. On 13th March 2020 the official twitter account of the Ecography journal published a tweet showing an animated GIF of the predicted distribution of 'spreading risk' for COVID-19 in different months of the year. At this point in time there was no preprint available to make an accurate assessment of the provenance of the predictions of spreading risk, the submitted version of the article was still undergoing peer-review, and it wasn't even clear which journal was handling the submission. Although the situation could be ascertained by following the link in the tweet to the authors' blog (and this was eventually clarified in response to the original tweet), the tweet itself did not make it clear that the work was unpublished and, to the casual

observer, the tweet looked like the journal promoting an accepted research paper due to be published in an upcoming issue.

A preprint version of the article did not appear on medRxiv until 16th March 2020. After some criticism of the unusual situation of the official twitter account being used to promote the yet-unpublished work of members of its editorial team, Ecography eventually deleted their original tweet. Unfortunately, in the meantime, significant damage had already been done; various media outlets had picked up the research and had already reported on the outputs contained within. The irregular dissemination strategy of the authors had directly led to the reporting of findings based on unpublished research that at the time hadn't been deposited in a preprint server, with significant implications for public health, during a global pandemic.

The authors have responded to this criticism by arguing that peer review can be a long and frustrating process and that, given the urgency of the situation, they had a moral obligation to disseminate the research outputs early. We agree that, in certain situations, there can be a case for releasing research outputs before they have been given a chance to be reviewed by the community but, in these situations, there are two clear points that differ from the case presented here:

- First, the full data and methodology need to be released publically. Long after the initial research outputs were announced there was still no access to the preprint. Only an incomplete version of the methodology was made available to the community and this led to the situation where the research outputs were being widely reported on before any scientific scrutiny could be applied.
- Secondly, in the absence of a preprint, social media was employed as the main tool of dissemination. Figures shared over social media without context or an underlying scientific article have a wide scope to be interpreted incorrectly. In particular, it was difficult even for trained ecologists to ascertain the meaning of 'spreading risk' as employed by the authors in the main figure accompanying the social media posts by following the summarised methodology contained in the blog post. There seemed little regard for how this term would be interpreted by a wider audience.

Although the stakes of ecological problems are often global and consequential, the nature of the problems addressed by ecologists and epidemiologists are fundamentally different - as are the nature of problems addressed by the majority of epidemiologists versus the specific infectious disease epidemiologists working on COVID-19. During an unprecedented crisis, most scientists will want to participate and contribute helpful efforts. However, there is a tremendous possibility for harm when ecologists produce hasty work outside their expertise. Many disease ecologists are working in collaboration with teams led by epidemiologists and public health response programs, forming consortia of skills in medicine, epidemiology, biostatistics and many others. But small teams pursuing ecological research on COVID-19 with little to no familiarity with the nuances of a zoonotic virus pandemic - an entire rich subfield of research that draws on ecology but is based far outside it - should be careful before disseminating work that has yet to be sense-checked by researchers working in outbreak response. As a best case, it may draw criticism and undermine the credibility of individual scientists, or broader disciplinary traditions like species distribution models. As a worst case, it may genuinely influence the public response to public health interventions. For example, by falsely convincing people living in certain areas that the risk they face from COVID-19 is low or that virus seasonality will render public health interventions irrelevant, we undermine social distancing efforts and other advisory policies. Social media and press coverage of claims like those proposed by Araújo and Naimi, if widely disseminated enough to change the behavior of an uninformed audience, might have the potential to quite literally put millions of people back at risk during the highest fatality pandemic of our lifetimes.

Acknowledgements

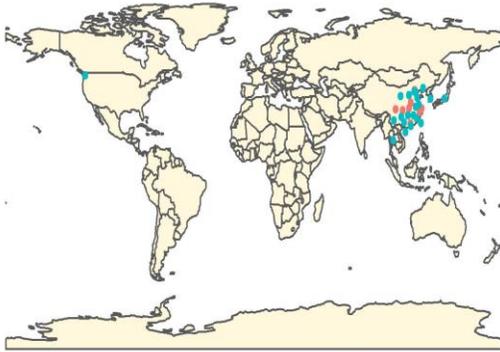
The authors would like to thank Bob Muscarella for his constructive comments on an earlier draft of the manuscript.

References

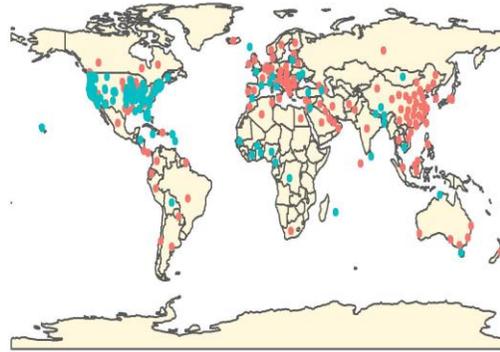
- Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., Garry, R. F. (2020) The proximal origin of SARS-CoV-2. *Nature Medicine*, <https://doi.org/10.1038/s41591-020-0820-9>
- Araújo, M. B., Anderson, R. P., Barbosa, A. M., Beale, C. M., Dormann, C. F., Early, R., Garcia, R. A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R. B., Zimmermann, N. E., Rahbek, C. (2019) Standards for distribution models in biodiversity assessments. *Science Advances*, 5, DOI: 10.1126/sciadv.aat4858
- Araújo, M. B., Naimi, B. (2020) Spread of SARS-CoV-2 coronavirus likely to be constrained by climate. *MedRxiv* <https://doi.org/10.1101/2020.03.12.20034728>
- Araújo, M. B., Pearson, R. G. (2005) Equilibrium of species' distributions with climate. *Ecography*, 28, 693-695.
- Araújo, M. B., Peterson, A. T. (2012) Uses and misuses of bioclimatic envelope models. *Ecology*, 93, 1527-1539.
- Burke, M., Gong, E., Jones, K. (2015) Income shocks and HIV in Africa. *The Economic Journal*, 125, 1157-1189.
- Burnham, K. P., Anderson, D. R. (2002) *Model selection and inference: a practical information-theoretic approach*. Springer-Verlag, New York.
- De Marco Júnior, P., Corrêa Nóbrega, C. (2018) Evaluating collinearity effects on species distribution models: an approach based on virtual species simulation. *PLoS ONE*, 13, e0202403
- Dong, E., Du, H., Gardner, L. (2020) An interactive web-based dashboard to track COVID-19 in real time. *Lancet*, [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1)
- Dormann, C. F. (2007) Promising the future? Global change projections of species distribution. *Basic and Applied Ecology*, 8, 387-397
- Dormann, C. F., Elith, J., Bacher, S., Buchman, C., Carl, G., Carré, G., García Márquez, J. R., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osbourne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., Lautenbach, S. (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36, 27-46.
- Engler, J. O., Rödder, D. (2012) Disentangling interpolation and extrapolation uncertainties in ecological niche models: a novel visualization technique for the spatial variation of predictor variable collinearity. *Biodiversity Informatics*, 8, 30-40.
- Fielding, A. H., Bell, J. F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24, 38-49.
- Fourcade, Y., Besnard, A. G., Secondi, J. (2017) Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography*, 27, 245-256.
- Fox, S. J., Bellen, S. E., Perkins, T. A., Johansson, M. A., Meyers, L. A. (2019) Downgrading disease transmission risk estimates using terminal importations. *PLoS Neglected Tropical Diseases*, 13, e0007395.
- Guisan, A., Zimmermann, N. E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, 135, 147-186.
- Hartig, F. (2020) DHARMA: Residual diagnostics for hierarchical (multi-level / mixed) regression models. R package version 0.2.7
- Karger, D.N., Conrad, O., Böhrer, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P. & Kessler, M. (2017) Climatologies at high resolution for the earth's land surface areas. *Scientific Data* 4, 170122.
- Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., Shaman, J. (2020) Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*, DOI: 10.1126/science.abb3221
- Lowen, A. C., Steel, J. (2014) Roles of humidity and temperature in shaping influenza seasonality. *Journal of Virology*, 88, 7692-7695.
- Marr, L. C., Tang, J. W., Van Mullekom, J., Lakdawala, S. S. (2019) Mechanistic insights into the effect of humidity on airborne influenza virus survival, transmission and incidence. *Royal Society Interface*, 16, 20180298.
- Park, J.-E., Son, W.-S., Ryu, Y., Choi, S. B., Kwon, O., Ahn, I. (2019) Effects of temperature, humidity, and diurnal temperature range on influenza incidence in a temperate region. *Influenza*, 14, 11-18.
- Phillips, B. J., Chipperfield, J. D., Kearney, M. R. (2008) The toad ahead: challenges of modelling the range and spread of an invasive species. *Wildlife Research*, 35, 222-234.
- Phillips, S. J., Anderson, R. P., Dudík, M., Blair, M. E. (2017) Opening the black box: an open-source release of Maxent. *Ecography*, 40, 887-893.
- Reeves, T., Samy, A. M., Peterson, A. T. (2015) MERS-CoV geography and ecology in the Middle East: analyses of reported camel exposures and a preliminary risk map. *BMC Research Notes* 8, 801.

- Rodgers, S. E., Zolnik, C. P., Matherm T. N. (2007) Duration of exposure to suboptimal atmospheric moisture affects nymphal blacklegged tick survival. *Journal of Medical Entomology*, 44, 372-375.
- Shapiro, L. L. M., Whitehead, S. A., Thomas, M. B. (2017) Quantifying the effects of temperature on mosquito and parasite traits that determine the transmission potential of human malaria. *PLoS Biology*, 15, e2003489.
- Svenning, J. -C., Skov, F. (2004) Limited filling of the potential range in European tree species. *Ecology Letters*, 7, 565-573.
- Theoharides, K. A., Dukes, J. S. (2007) Plant invasion across space and time: factors affecting nonindigenous species success during four stages of invasion. *New Phytologist*, 176, 256-273.
- Thuiller, W., Araújo, M. B., Pearson, R. G., Whittaker, R. J., Brotons, L. , Lavorel, S. (2004) Uncertainty in predictions of extinction risk. *Nature* 430, 34
- Václavík, T., Meentemeyer, R. K. (2012) Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion. *Diversity and Distributions*, 18, 73-83.
- Williams, R. A. J., Folorunso, O. F., Peterson, A. T. (2008) Predictable ecology and geography of avian influenza (H5N1) transmission in Nigeria and West Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 102, 471-479.
- Wood, S. N. (2017) *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, Boca Raton.
- Zhu, G., Peterson, A. T. (2014) Potential geographic distribution of the novel avian-origin influenza A (H7N9) virus. *PLoS One*, 9, e93390.

22nd January 2020



11th March 2020



18th March 2020

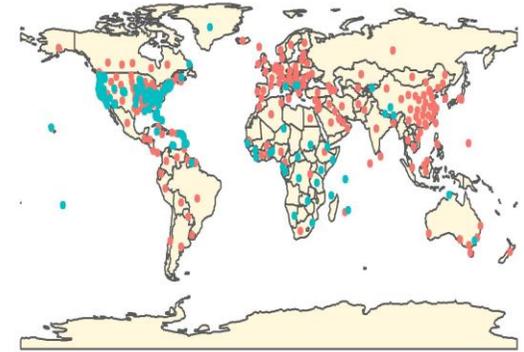
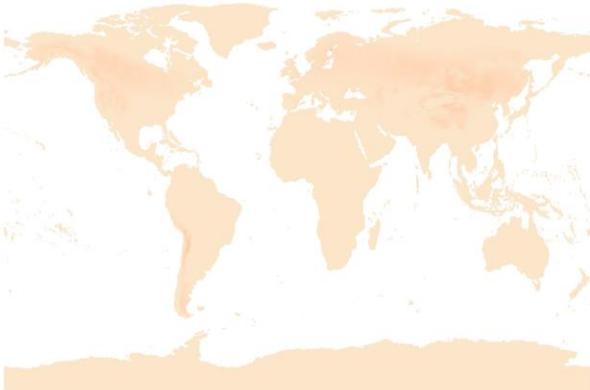
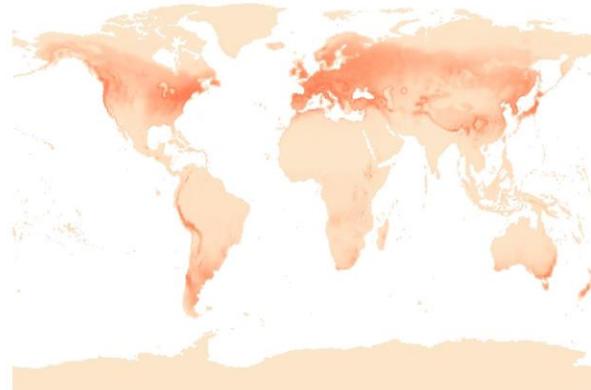


Figure 1: Confirmed cases of SARS-CoV-2 according to the Dong *et al.* (2020) dataset at different stages of the epidemic. Red points represent locations with five or more confirmed cases and have been classified as ‘presences’ in the species distribution modelling framework of Araújo and Naimi. Blue points represent locations with fewer than five cases and are therefore too few to be classified as ‘presences’ in Araújo and Naimi.

22nd January 2020



11th March 2020



18th March 2020

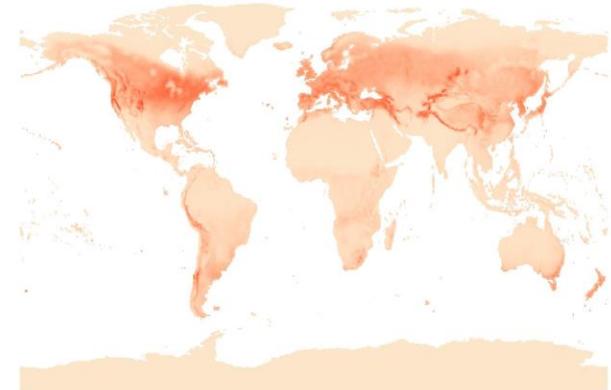


Figure 2: Prediction of occurrence probabilities for SARS-CoV-2 for April using a generalized additive model using the confirmed cases dataset of Dong *et al.* (2020) and the same classification scheme and covariates employed in Araújo and Naimi. The three panels represent models fitted using occurrence data from the beginning of the epidemic up until a particular cutoff date (shown above each panel).