

# A High-resolution Global-scale Model for COVID-19 Infection Rate

GIANPAOLO CORO and PASQUALE BOVE, Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo" – CNR, Pisa, Italy

Several models have correlated COVID-19 spread with specific climatic, geophysical, and air pollution conditions, and early models had predicted the lowering of infection cases in Summer 2020. These approaches have been criticized for their coarse assumptions and because they could produce biases if used without considering dynamic factors such as human mobility and interaction. However, human mobility and interaction models alone have not been able to suggest more innovative recommendations than simple social distancing and lockdown, and would definitely need to include information about the base environmental suitability of a World area to COVID-19 spread. This scenario would benefit from a global-scale high-resolution environmental model that could be coupled with dynamic models for large-scale and regional analyses. This article presents a 0.1° *high-resolution global-scale* probability map of low and high infection-rates of COVID-19 that uses annual-average surface air temperature, precipitation, and CO<sub>2</sub> as environmental parameters, and Italian provinces as training locations. A risk index calculated on this map correctly identifies 87% of the World countries that reported high infection rates in 2020 and 80% of the low and high infection-rate countries overall. Our model is meant to be used as an additional factor in other models for monthly weather and human mobility. It estimates the base environmental inertia that a geographical place opposes to COVID-19 when mobility restrictions are not in place and can support how much the monthly weather favors or penalizes infection increase. Its high resolution and extent make it consistently usable in global and regional-scale analyses, also thanks to the availability of our results as FAIR data and software as an Open Science-oriented Web service.

CCS Concepts: • **Computing methodologies** → **Modeling and simulation**;

Additional Key Words and Phrases: Ecological niche modelling, Coronavirus, SARS-CoV-2, COVID-19, maximum entropy

## ACM Reference format:

Gianpaolo Coro and Pasquale Bove. 2022. A High-resolution Global-scale Model for COVID-19 Infection Rate. *ACM Trans. Spatial Algorithms Syst.* 8, 3, Article 20 (January 2022), 24 pages.  
<https://doi.org/10.1145/3494531>

## 1 INTRODUCTION

A correlation between the decrease of COVID-19 cases and the summer weather in most Northern Hemisphere countries was observed in 2020 but had been previously predicted by several

Gianpaolo Coro and Pasquale Bove contributed equally to this research.

The authors acknowledge the European Open Science Cloud COVID-19 Fast Track Funding for supporting this research. Authors' address: G. Coro and P. Bove, Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo" - CNR, Pisa, Italy, Via Moruzzi 1, Pisa, Tuscany, Italy, 56124; emails: {coro, pasquale.bove}@isti.cnr.it.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2022 Association for Computing Machinery.

2374-0353/2022/01-ART20 \$15.00

<https://doi.org/10.1145/3494531>

models that had used climatic information to explore this correlation. Most of these models were large-scale spatio-temporal models that identified climatic similarities between areas affected by COVID-19 [4, 29, 43, 128, 135]. These approaches have been criticized (especially before Summer 2020), because of their approximations and coarse resolutions and assumptions, in favor of dynamic systems based on human mobility and interactions that constantly monitor and forecast infection's spread [18, 65, 69, 127, 129]. However, these dynamic systems have shown difficulties at producing effective and realizable recommendations (i.e., different from simple social distancing and lockdown) to prevent and understand infection spread [7, 77, 80, 145]. Indeed, the currently implemented mobility restrictions are still very coarse, and the only effective strategy currently pursued by World governments are vaccination campaigns. However, governments rely very much on the correlation between climatic conditions (especially during Summer) and infection spread to manage restrictions [52], but unfortunately several governments misinterpreted this correlation in 2020 to relax restrictions too much. Today, it has been widely demonstrated that climatic conditions and air pollution are essential drivers of the COVID-19 pandemic, especially if coupled with high human mobility and interaction [27, 58, 62, 72, 100, 133, 146]. Correlative models of COVID-19 with annual and monthly climatic conditions provide information about the environment that supports virus persistence, independently of social dynamics. This information is crucial when coupled with vaccination and economic strategies. Indeed, it can indicate when weather is a "good ally" (even at a monthly resolution) to fight COVID-19 and thus when restrictions can be relaxed, with consequent economic and social benefits. Otherwise, it can indicate when average weather conditions may favor the virus spread and thus acceleration on vaccination campaign and anti-COVID therapy is required.

Among the proposed climatic correlative models, **ecological niche models (ENMs)** have been used for the COVID-19 pandemic [4, 29, 113] and other viruses [16, 48, 89, 121, 122]. An ENM estimates the suitability of the environmental conditions of a geographic area to a species' presence and subsistence. It models a species-specific correlation function between the species' occurrence records in its native habitat and specific environmental parameters. This correlation is the species' niche, formally defined as the space within a hyper-volume in the environmental parameter vector space associated with the species' environmental preference that fosters the species' persistence in an area. An ENM can estimate the species' ecological niche in its native habitat (native niche) or across other geographical areas to discover new potentially suitable areas for the species (potential niche). Applying this concept to viruses allows discovering the environmental conditions that may foster the proliferation of infection cases in a population. In the case of a pandemic like COVID-19, the native niche likely corresponds to the potential niche, because the virus can adapt to many environmental conditions and is fostered by global human mobility [70, 87, 117]. The global spread of COVID-19 has highlighted the high adaptability of SARS-CoV-2 to a wide range of environments and has indicated that populated regions exist with lower infection increment rates than others but with equally high social activity [29]. Since human mobility and interactions in these areas are high, the reasons for the lower infection increment rates might be found in their environmental conditions. Indeed, studies that used ENMs indicated interesting climatic similarities between global regions that reported similar infection rates in 2020 [4, 113].

In 2020, we conducted a preliminary experiment with an ENM to estimate COVID-19 infection rate as *disease spreading speed*, i.e., the derivative of the infection cases in the increment period [29]. Our model was developed at an early stage of the pandemic in March 2020 and produced a global risk map at a half-degree spatial resolution. The model foresaw high-risk geographical areas that could favor infection spread because of their particular geophysical (surface air temperature, precipitation, and elevation) and human-related characteristics (CO<sub>2</sub> and population density). The model did not include social interactions but advised that the estimated distribution represented

the *inertia* that an area would oppose to the virus' spread independently of human social activities. Indeed, the distribution was complementary to the one modelled by human-dynamic models, and could be coupled with these models as a multiplicative factor in one overall risk function [100, 119]. Our model discovered strong suitability to high infection rate in locations that had manifested a quick infection spread, e.g., the Hubei province of China, Europe, and the western United States. Overall, it correctly predicted  $\sim 77\%$  of the countries that had reported outbreaks. This large-scale experiment clearly indicated specific parameter ranges that were most correlated with high-infection rate urbanized areas, i.e., moderate-high  $\text{CO}_2 \sim 0.03 \text{ g C m}^{-2} \text{ day}^{-1}$ , average air temperature  $\sim 11.8^\circ\text{C}$ , and moderate precipitation  $\sim 0.3 \text{ } 10^{-4} \text{ kg m}^{-2} \text{ s}^{-1}$ . Furthermore, this correlation was independent of population density, especially after 200 people per  $\text{km}^2$ .

However, the model could not explain the high infection rate of regions such as Brazil, Ecuador, and Peru. In these regions, social conditions and people interaction in overcrowded cities seemed to be more critical than climatic conditions. Nevertheless, risk assessment for these countries was biased by the coarse  $0.5^\circ$  resolution ( $\sim 50 \text{ km}$ ) of the model that included both populated and uninhabited areas within one probability cell. This bias was higher in those countries where vast uninhabited regions separate urbanized areas. Furthermore, a  $0.5^\circ$  spatial resolution was hardly combinable with city-scale social models that presented substantial variations within a  $\sim 50 \text{ km}$  range.

For this reason, this article presents an augmented version of the ENM model of Reference [29]—at an increased  $0.1^\circ$  spatial resolution—that predicts COVID-19 infection rate based on  $\text{CO}_2$ , temperature, and precipitation data. The presented experiment required a different data preparation phase, the increase of the ENM spatial resolution, and a re-assessment of an infection-increase *risk index* based on the ENM output. Our new higher-resolution model (i) produces much more information than the previous model, (ii) reveals new infection patterns, and (iii) indicates a high risk for several countries missed by the previous model. Furthermore, the availability of global COVID-19 data for the entire 2020 allowed us to conduct a more complete performance evaluation. Our new model is also more combinable with regional-scale social models of COVID-19 spread.

This article is organized as follows: Section 2 reports an overview on correlative climatic approaches for the COVID-19 pandemic. Section 3 describes our data collection and preparation phases and the modelling approach; Section 4 reports the model's performance at predicting global high-infection-rate zones based on 2020 global reports; Section 5 discusses the results and draws the conclusions.

## 2 OVERVIEW

Several experiments have modelled the complex dependency between COVID-19 spread and environmental variables to predict future outbreaks and analyze the current situation. Statistical models have highlighted a general dependency of COVID-19 spread on high temperature and humidity [61, 73, 82, 142] and have indicated the role of weather, season, and air pollution in the variation of COVID-19 cases [2, 13, 93, 132]. Other studies have modelled the dependency of COVID-19 spread on human mobility and highlighted the strong negative effects that undocumented cases and the lack of tracking data have on these models [76, 85]. Some works have tried to compensate for this issue through social-network-text mining, with limited predictive results, and have highlighted the large gap between official reports and actual cases (e.g., in China) [49, 64, 75, 147]. Internet data have also been used to understand the effect of social distancing on COVID-19 spread [57]. Finally, predictive models based on COVID-19 case reports and clinical information have been used to evaluate the effectiveness of lockdown and social distancing and predict infection waves [3, 21, 22, 26, 28, 112, 126].

Among the environmental models, ENMs have extensively proved to be effective to estimate viruses' potential spatial outreach and the environmental data that may foster disease

Table 1. Summary of All Data Involved in Our Experiment, with their Primary Sources

Data	Primary Source
Infection per Italian Province	Italian Civil Protection Department
World Infections	John Hopkins University
Average Annual Surface Air Temperature	EnviDat
Average Annual Precipitation	EnviDat
Carbon Dioxide	Copernicus Atmosphere Monitoring Service

Data access and post-processing is described in the article.

transmission [42, 44, 88, 97, 140, 149]. Building an ENM requires (i) identifying relevant environmental data to the species' presence, (ii) retrieving reliable species observation records, (iii) modelling the correlation between the species' presence and environmental data, and (iv) projecting this correlation onto a geographical space. There is a vast choice of ENM modelling techniques [99], some of which also require information on unsuitable habitat for the modelled species (absence locations) [33]. Having environmental information associated with absence records—either expert-provided or automatically estimated—is usually beneficial for parameter selection and model performance [23, 34, 99, 102]. An ENM can be built either by explicitly modelling the species' environmental preferences and physiological limits (mechanistic models) or by automatically inferring these through a machine learning or statistical model (correlative models) [99]. Correlative ENMs are the most used to predict disease outbreaks over long periods and, despite their categorical reference to *correlation*, estimate a complex function defined on environmental parameters that represents the probability of species' potential habitat suitability [48, 89, 121, 122]. However, shorter-period correlative ENMs have been able to indirectly track viruses by predicting their principal vectors' niches [55, 79, 86, 118, 125, 130, 139, 141]. Short-period ENMs—especially those based on generalized linear models—have been used also to predict COVID-19 outbreaks in large geographical areas based on meteorological conditions [1, 4, 19, 29, 110]. One widely used correlative ENM is **Maximum Entropy (MaxEnt)**, which is flexible enough to work with presence or presence/absence data and also estimates the conditions that, in the case of a virus, might end in an infection outbreak [29, 34, 36, 51, 70, 87, 104, 117]. We used MaxEnt in the methodology presented in this article (Section 3.3).

### 3 MATERIAL AND METHODS

#### 3.1 Data Selection and Preparation

One of our principal aims was to make our methodology repeatable, reproducible, and reusable in compliance with the Open Science paradigm [131]. Thus, we searched for data that met the principles of findability, accessibility, interoperability, and re-usability (FAIR data), i.e., geospatial data accessible through representational standards and published on open-access repositories. Unfortunately, the data we used at a 0.5° resolution for our previous experiment were not available as FAIR data at a 0.1° resolution. Furthermore, NASA's temperature and precipitation data were no more accessible [91]. Thus, we searched for new repositories of environmental data and transformed these data into annual-averaged grid raster files in the standard NetCDF-CF format (Table 1).

Our previous MaxEnt model indicated that average surface air temperature, precipitation, and CO<sub>2</sub> accounted for ~98% of the model-produced probability distribution, and thus we used these parameters in our new model. Average surface air temperature and precipitation have demonstrated to be important predictor variables for COVID-19 spread in other experiments [53, 119], and are commonly used in viruses' ecological niche models [16, 55, 98, 139]. Precipitation is commonly used as a surrogate of humidity [9, 20, 84], which has been correlated with COVID-19



[14, 71, 83]. Several works have demonstrated a correlation between air pollution and COVID-19 spread [62, 72, 100]. CO<sub>2</sub> can be used as a proxy for air pollution, because they are correlated, and generally with human presence. Other parameters that we did not take into account in our new model (because of their lower weight in the previous model) were elevation, population density, and other greenhouse gases (i.e., methane and nitrous oxide). Indeed, although elevation has been used to model the spread of other viruses [87, 101, 139], it is less useful to predict SARS-CoV-2 spread [29]. As for population density, complex-system experiments have demonstrated that pandemics occur beyond a critical population density that depends on the virus' infectivity, recovery rate, and mortality [48, 121]. COVID-19 has been reported both in over-crowded and less inhabited cities with high infectivity and slow recovery rates. Thus, population density is likely a weak discriminant factor—over a minimal threshold that our previous model estimated as  $\sim 200$  inhabitants per km<sup>2</sup>—to state *a priori* that an area is not at risk of infection spread.

The EnviDat infrastructure hosts a FAIR repository of 0.049° resolution ( $\sim 5$  km at the equator) data for maximum/minimum surface air temperature and average precipitation [95, 96]. The variables are historical monthly estimates from 1850 to 2005 and forecasts up to 2100, under the RCP 4.5 and 8.5 scenarios. Historical data between 1990–2005 are particularly indicated for consistent averaging and use in ecological models [96]. The data are provided as uniform-valued grids in the NetCDF standard format. The variables were estimated through the **Coupled Model Intercomparison Project - Phase 5 (CMIP5)** framework [143], in particular through four global circulation models (GCMs): ACCESS1-3, CESM1-BGC, CMCC-CM, MIROC5. We processed these data to obtain one dataset for average surface air temperature and another one for average precipitation. Using each GCM as one variable of MaxEnt would have introduced too high inter-parameter dependency and consequently lowered the model's performance [51, 108]. Thus, we used the four models' ensemble for each variable and averaged information across months and years. Finally, we estimated average surface air temperature as the average of the minimum and maximum temperature annual data, which is particularly sound for ecological niche models and climate change pattern analyses [37, 39].

In summary, we obtained average surface air temperature and precipitation (Figures 1(a) and (b)) through the following operations:

- (1) Download the 1990–2005 historical data of the four GCMs;
- (2) Restrict the time range to 2000–2005 to better focus climatic variability over the Internet and technology outbreak period;
- (3) Calculate monthly averages for each GCM to obtain annual averages;
- (4) Average the four annual GCMs;
- (5) Use first-order conservative remapping [136] to project the original data onto a 0.1° resolution grid;
- (6) Average the two ensemble maximum and minimum surface air temperature datasets to obtain one average surface air temperature dataset.

Note that data projection was performed through first-order conservative remapping, which assigned to a target cell a weighted average of the values of the intersecting cells, where each weight was proportional to the intersection area. We implemented this algorithm with the NetCDF Operators [92] and the Climate Data Operators [17] software.

As for CO<sub>2</sub>, we took FAIR data from the Copernicus Atmosphere Monitoring Service [15]. These data contain monthly carbon dioxide flux information at the soil, aggregated between 1979 and 2013. This dataset includes both human and natural activity, although human activity's average contribution is much higher than the natural one. Unfortunately, no FAIR repository currently reports CO<sub>2</sub> data at a finer scale than 0.5°. This limitation is also due to the fact that it would be

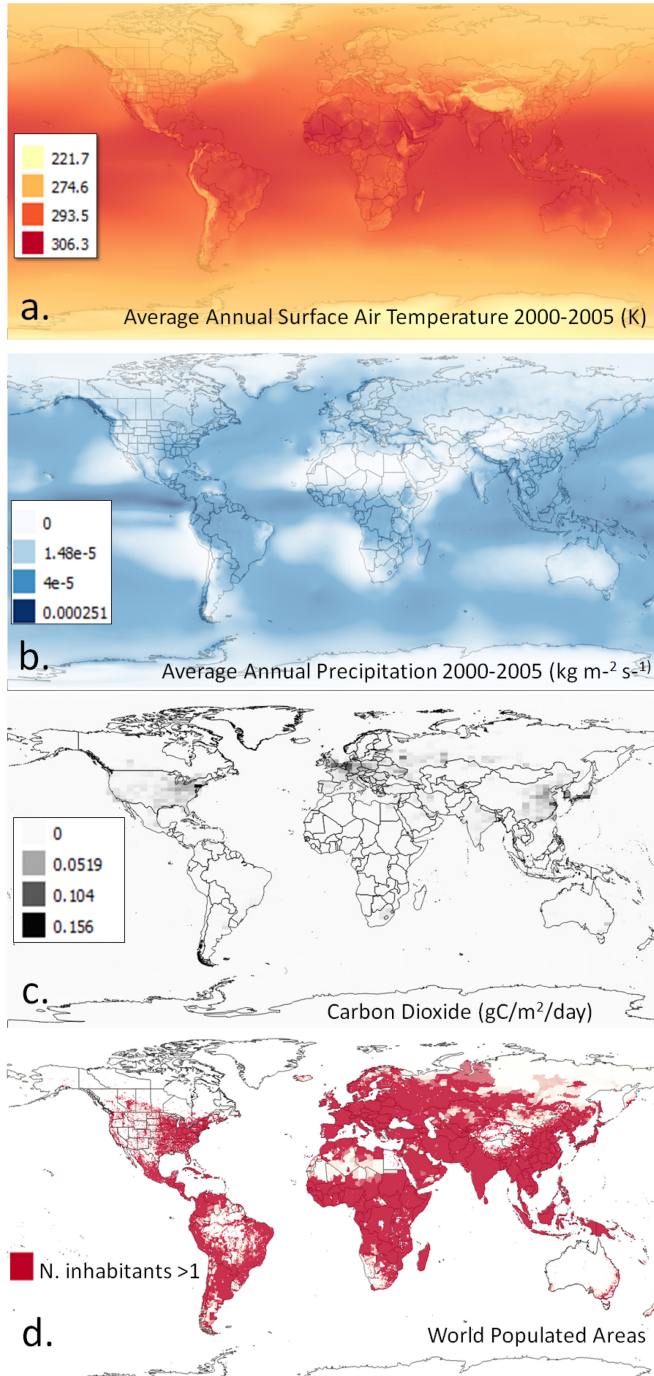


Fig. 1. Visual comparison of the environmental and human-related data used by our model: (a) average annual surface air temperature in 2000–2005, (b) average annual precipitation in 2000–2005, (c) average carbon dioxide in 1979–2013, (d) World-populated areas.

less realistic to model the concentration of this gas on smaller areas, especially for the scopes of statistical models [40]. Since carbon dioxide proved to be the most influential parameter of our previous model, we had to include it in our new model necessarily. To this aim, we processed and interpolated the CO<sub>2</sub> Copernicus data (Figure 1(c)) as follows:

- (1) Download the 1979–2013 historical data from Copernicus Atmosphere Monitoring Service;
- (2) Calculate monthly averages to obtain one annual average. This operation uses the full dataset time frame to combine values preceding the high industrialization rate of the last decades with natural CO<sub>2</sub> emission;
- (3) Down-sample the Copernicus layer to 0.1° model through first-order conservative remapping [136].

### 3.2 Observation Data

Our ENM estimated a prediction function, defined on the selected environmental parameters, for COVID-19 infection rate (here defined as the derivative of the infection cases in the increment period). In particular, we focused on infection *rate* instead of absolute infection *counts* to reduce the information bias generated by undetected cases, low cases with a high spread rate, misreports, and low numbers of serologic tests and swabs. Thus, we used the derivative of the COVID-19 case report curve and modelled the relation between environmental parameters and areas with averagely high derivative in the infection increase periods. To this aim, we needed reliable reports of infection increase that were not biased by additional factors such as mobility restrictions, vaccinations, and misreports. The best period to avoid these observational biases was March 2020, because the virus was freer to circulate, and regulations were either late or had their effect start in April. Furthermore, in March 2020, vaccination campaigns were not active yet. Finally, asymptomatic cases were poorly reported in March 2020, because swabs were only done to symptomatic patients. Consequently, the reported infection rate was proportional to the absolute severity of the infection.

Among the COVID-19 global reports of March 2020, Italian data particularly satisfied our requirements, because Italy was the first European country to be both heavily impacted by the virus and to study the pandemic, and it reported infection cases rigorously and systematically. Furthermore, a correlation between temperature, humidity, and air pollution in Italy has been highlighted throughout 2020 by both governmental and scientific investigations [68, 120, 134]. The Italian Civil Protection Department publishes, since February 2020, daily updates of COVID-19 numbers of people infected, recovered, and dead, down to a province scale [67]. March 2020 was the maximum infection period in Italy during the “first wave,” and the first lockdown started on 9 March. Symptomatic cases were basically the only cases reported during March, and the infection curve began to reduce just at the end of April (around the 23rd) [59, 66]. Thus, March 2020 in Italy was a benchmark period for our model due to increasing cases, reliable reports that reflected the severity of the epidemic, and mobility restrictions that were carried out late in the month and had a poor effect during March. In summary, we used Italian province data up to the end of March 2020 as the training observations for our ENM (Figure 2(a)). City areas with the highest derivatives of the infection cases (*high infection rate*) were selected. Environmental data associated with these areas were used as positive correlation examples.

Using only Italian observation data in our ENM could be seen as a potential source of bias, because other countries’ environmental characteristics can be very different. However, annual average environmental conditions have a large variability across Italy and represent other global conditions, from hot-humid and desert areas to cold mountain areas with perennial glaciers. Furthermore, the COVID-19 data published by the John Hopkins University for most World countries

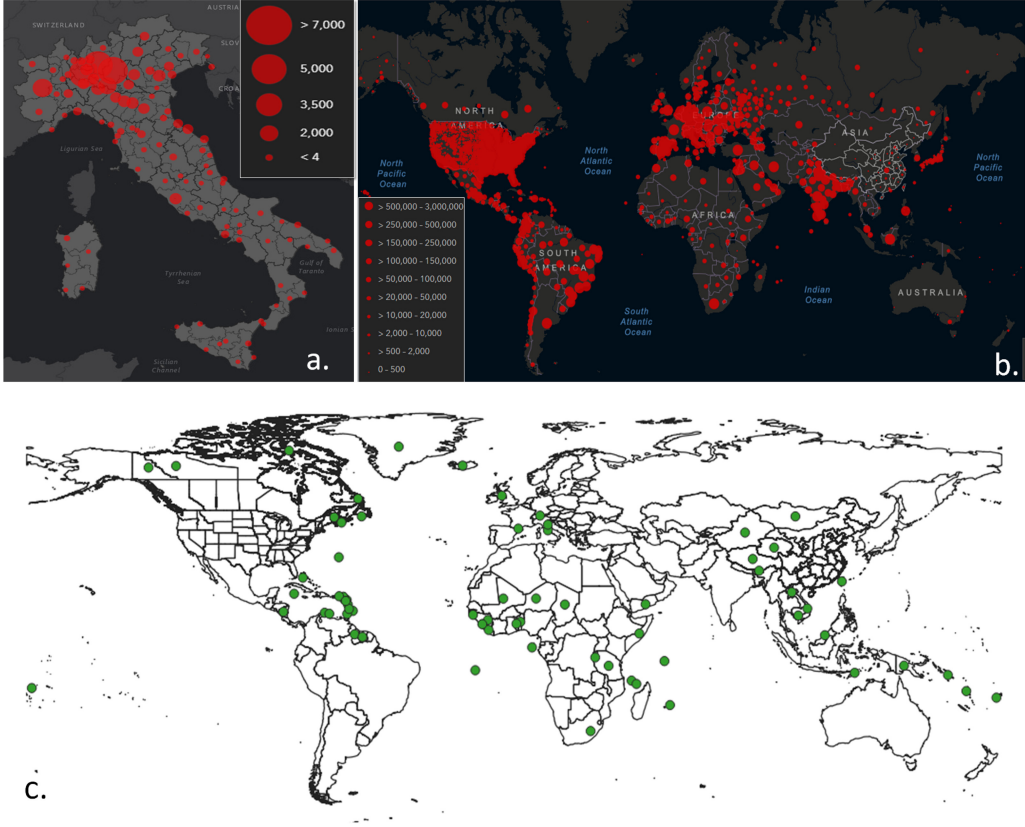


Fig. 2. Visual comparison of the infection data used by our model: (a) number of infection cases in Italian provinces on 31 March 2020, (b) cumulative infection cases of COVID-19 on 04/04/2021, (c) low-infection-rate countries calculated from infection rate data (derivatives).

[46] (Figure 2(b)) highlight the heterogeneity and poor comparability of cross-country reports. Resolution is given at a national and sometimes regional scale (e.g., China, U.S.A., and Canada), and thus it does not allow to appreciate an intra-country variability that is essential to build a  $0.1^\circ$  resolution model. Moreover, mixing cross-country data with Italian data would generate additional bias, because the other World countries use very different disease identification and monitoring approaches [115]. Thus, cross-country data are practically incomparable, especially for March 2020. Nevertheless, global data can be used to assess if a country has a high or low *risk* of fostering a high infection rate by analyzing the steepness of their infection waves. Thus, cross-country data can be used at least to assess model performance at a macroscopic level.

To identify Italian provinces and World countries with the highest infection rates, we used the following statistical analysis:

- (1) For each zone to analyze (i.e., either Italian provinces or World countries):
- (2) Take the time series of COVID-19 cases in 2020;
- (3) Apply a three-day smoothing;
- (4) Detect the maximum peak;
- (5) Select the segment of the time series between 10% of the peak value and the peak itself;

- (6) Calculate the rates of infected people in the segment, i.e., the derivatives of the COVID-19 cases' time series in the increment period;
- (7) Calculate the average infection rate;
- (8) Approximate the distribution of all infection rates as a log-normal distribution (the validity of this approximation was confirmed by a Chi-square test over Italian data);
- (9) Identify low-infection-rate zones as those with  $rate < geometric\ mean - lognormal\ standard\ deviation$
- (10) Classify the remaining zones as high-infection-rate zones.

Note that this algorithm assumes that infection rates are log-normal distributed and excludes locations with rates much lower than the distribution median (i.e., the geometric mean). This approach has demonstrated effectiveness also in other contexts [32], because it is neither too restrictive to risk producing small datasets (i.e., it does not include data above the upper confidence limit), nor too weak to include very-low infection rate zones (i.e., below the lower confidence limit). This approach was verified to be the optimal choice to balance dataset size and prediction performance. This procedure identified 54 Italian provinces and 175 World regions with a high infection rate and 53 Italian provinces and 70 World regions with a low infection rate [30] (Figure 2(c)).

### 3.3 The Ecological Niche Model

MaxEnt is a commonly used ENM [8, 34, 41, 104, 105, 107] that estimates the probability density  $\pi(\bar{x})$  of a species' presence given real-valued vectors  $\bar{x}$  of environmental parameters. During the training phase, MaxEnt learns from the environmental values in the observation locations [41, 99] and does not necessarily need information on absence locations. A high COVID-19 infection rate in a particular area can be modelled as a species presence within specific environmental conditions. Since MaxEnt is very sensitive to biased data [50], it is fundamental to provide reliable information about high-infection rates. This characteristic strengthened our choice to use Italian data of March 2020 to extract observation data. A MaxEnt model trained with a specific area's parameters at a  $0.1^\circ$  resolution can be projected at a  $0.1^\circ$  resolution on another area to produce a spatial probability distribution. Indeed, we trained a MaxEnt model at this resolution using high-infection rate Italian provinces as observation data and average annual surface air temperature, precipitation, and  $CO_2$  as environmental parameters. Finally, we projected the model at the global scale.

MaxEnt estimates  $\pi(\bar{x})$  so its entropy function  $H = - \sum \pi(\bar{x}) \ln(\pi(\bar{x}))$  is maximum on the training locations with respect to the entropy of randomly selected environmental vectors (*background points*) in the reference area. The training algorithm estimates the coefficients of a linear combination of the parameters that is used in the analytical form of  $\pi(\bar{x})$ . These coefficients represent the influence of each environmental parameter in the species' presence prediction (*percent contribution*). Furthermore, the model also estimates the *permutation importance* of each parameter in the environmental vectors. In our model,  $\pi(\bar{x})$  is the probability of a  $0.1^\circ$  cell-location to foster a high infection rate, i.e.,  $\pi(\bar{x}) = P(\text{high} - \text{infection} - \text{rate} | \bar{x})$ . This distribution can be further defined using the Bayes rule:

$$\pi(\bar{x}) = \frac{f_1(\bar{x})P(\text{high} - \text{infection} - \text{rate})}{f(\bar{x})},$$

where  $f(\bar{x})$  is a probability density over the background parameters;  $f_1(\bar{x})$  is a density function defined on the high-infection-rate locations (i.e., the training set); and  $P(\text{high} - \text{infection} - \text{rate})$  is the prior distribution of high-infection-rate zones (*prevalence*), set to 0.5, since no prior assumption can be given on infection rate. MaxEnt hypothesizes that the optimal  $f_1(\bar{x})$  is the closest function to  $f(\bar{x})$ , because without observations there would be no expectation (which is a *maximum entropy condition*). However, the model constraints  $f_1(\bar{x})$  to have its maximum on the parameters' means in



the training set locations. With this constraint, the model minimizes the Kullback-Leibler distance between  $f_1(\bar{x})$  and  $f(\bar{x})$ :

$$d(f_1(\bar{x}), f(\bar{x})) = \sum_{\bar{x}} f_1(\bar{x}) \log_2 \left( \frac{f_1(\bar{x})}{f(\bar{x})} \right),$$

which is solved by Gibbs distribution functions such as  $f_1(\bar{x}) = f(\bar{x})e^{\eta(\bar{x})}$  [104], where  $\eta(\bar{x}) = \alpha + \beta h(\bar{x})$ ;  $\alpha$  is a normalization constant that makes  $f_1(\bar{x})$  sum to 1;  $h$  is an optional transformation of  $\bar{x}$  that can model complex relationships between the environmental parameters; and  $\beta$  is the vector of *percent contributions* that indicates the quantity of information carried by each parameter [34, 36, 41]. Thus, minimizing  $\eta(\bar{x})$  consequently minimizes  $d(f_1(\bar{x}), f(\bar{x}))$ , which means that the MaxEnt training algorithm requires to solve a log-linear equation over the background and training vectors (usually done through penalized maximum likelihood [108]). MaxEnt is a more complex and adequate choice to model a species' presence than a linear or logistic regression, because it is equivalent to a Poisson regression (a generalized linear model) that appropriately models the probability of a number of events in a fixed space (i.e., a species' occurrences) [114]. However, not including information on absence locations (i.e., low infection rate areas) and dispersal constraints might reduce prediction accuracy [137, 138]. Our model mitigates this issue by using high-quality training presence information [34, 36] and reducing over-fitting issues through two advanced options offered by the used MaxEnt software [103]: Presence locations were included among background points if associated with unique combinations of environmental parameters, and *hinge* features were enabled to model a complex response to the environmental parameters and make model fitting more flexible [47, 103, 106, 148].

### 3.4 Model Availability

All results reported in this article are publicly available as FAIR data [30]. Furthermore, for our experiments, we used a MaxEnt implementation openly available as-a-service on the DataMiner cloud computing platform [6, 31]. The underlying implementation is based on the software described in Reference [109]. The DataMiner endpoint is a free-to-use Web service version of this software that automatically works with standardized input environmental data (ESRI GRID, NetCDF, GeoTIFF, etc.) that should be provided as input together with training locations in CSV format [24, 109]. Furthermore, it is a cloud computing service [38] published under the **Web Processing Service (WPS)** standard to maximize its reuse from other software that natively supports this description (e.g., References [10, 111, 144]). WPS also fosters the service re-use in other domains or for other diseases [34, 36, 40, 41]. Finally, DataMiner uses the Prov-O ontological format [74] to save computational information (input, output, parameters) and allow the repetition and reproduction of the experiments and collaborative experimentation.

### 3.5 Evaluation Metrics

We evaluated model training results using average **Area Under the Curve (AUC)**, i.e., the integral of the **Receiver Operating Characteristic (ROC)** curve. The ROC curve represents *sensitivity* vs. *1-specificity*, i.e.,  $\frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$  vs.  $1 - \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}$ . When AUC is close to 1, the model-simulated probability is significantly higher on presence locations than random locations. A threshold on the  $\pi(\bar{x})$  values can be set to transform the probability distribution into a dichotomic classification [109], where cells with values above the threshold are classified as high-infection-rate areas and as low-infection-rate areas otherwise. True positives and false negatives can be calculated based on how many actual high-infection-rate areas of the training set are captured after setting the threshold. One commonly used threshold—also adopted in this

Table 2. Results of the MaxEnt Training Phase, with Estimated Thresholds and Indication of Percent Contribution and Permutation Importance for the Used Environmental Parameters

<b>AUC</b>	0.995	
<b>Balanced omission-sensitivity threshold</b>	0.008	
<b>Peak identification threshold</b>	0.13	
<i>Parameter name</i>	<i>Percent contribution</i>	<i>Permutation importance (%)</i>
<b>Carbon Dioxide</b>	85	28.9
<b>Average Annual Surface Air Temperature</b>	8.6	57.7
<b>Average Annual Precipitation</b>	6.4	13.5

article—is the *balanced threshold*, which balances *omission rate* ( $\frac{\text{False Negatives}}{\text{True Positives} + \text{False Negatives}}$ ) and *sensitivity*. Moreover, probability *peak* areas can be identified as those locations with particularly high values, falling in the upper tail of the probability distribution across all cells.

To numerically assess the performance of our model at predicting World countries that actually reported high and low infection rates, we calculated the *risk index* of a country [29, 41] as the *average MaxEnt probability* in its populated areas (based on the data in Figure 1(d)). This index is much more accurate at a 0.1° resolution than it was at 0.5° resolution, because inhabited areas are selected with a much higher precision by the smaller cells. Countries with a *risk index* over the geometric mean of the *risk index* distribution across all countries were classified as *high-risk* countries, and as *low-risk* countries otherwise. These countries can be compared with those extracted from real data (Section 3.2). We calculated the accuracy on the prediction of countries' infection-rate as

$$\text{Accuracy} = \frac{n. \text{ of high - and low - infection - rate areas correctly identified}}{\text{overall n. of areas}}.$$

Moreover, we calculated the agreement between real and estimated low/high risk countries using Cohen's Kappa [25], which estimates accuracy with respect to chance agreement and can be interpreted using Fleiss' tables [54].

## 4 RESULTS

### 4.1 Model Training

Our MaxEnt model training phase resulted in a 0.95 AUC (Table 2), which indicates a very good learning from the training set. The *percent contributions* of the three parameters indicate that CO<sub>2</sub> has the highest weight (85%), followed by a much lower weight of average air temperature (8.6%) and precipitation (6.4%). However, average air temperature has the highest *permutation importance* (57.7%), followed by CO<sub>2</sub> (28.9%) and precipitation (13.5%). The high *permutation importance* of all parameters indicates that they all bring important information to the model, and thus that CO<sub>2</sub> alone does not account for the entire model's performance. The ranges, for each parameter, associated with the maximum probability of high infection rate can be inferred by analyzing the logistic prediction curves of the individual-parameter models [34] (Figures 3(a) and (b)). These curves represent (i) the individual-parameter logistic responses obtained after fixing the other parameter values at their averages on the training locations (marginal response curve) and (ii) the individual-parameter logistic responses obtained after excluding the other variables from the MaxEnt model. The curves agree on the probability peaks of each parameter. In particular, referring to the marginal response curves, the environmental parameter ranges that could be associated with the maximum infection rate were: 13.85 (8.85; 16.85) °C (*moderate-low*) for temperature, 0.3 (0.2; 0.4) 10<sup>-4</sup> kg m<sup>-2</sup> s<sup>-1</sup> for precipitation (*moderate*), and 0.02 (0.01; 0.05) g C m<sup>-2</sup> day<sup>-1</sup> for CO<sub>2</sub> (*moderate-high*). These ranges can be compared with those of our previous 0.5° model: CO<sub>2</sub>

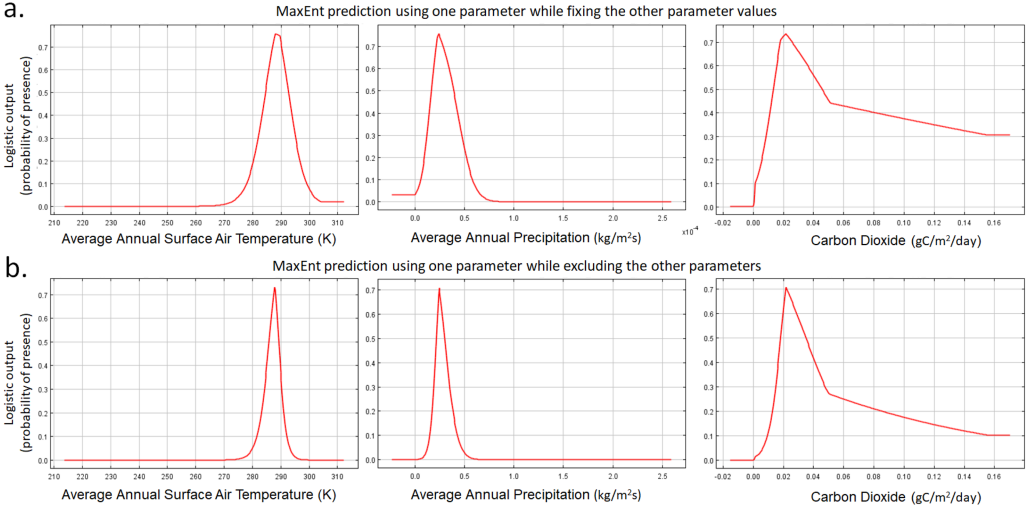


Fig. 3. Comparison between (a) the logistic prediction curves obtained by varying one environmental parameter while fixing all other parameters at their average values on the training locations and (b) the logistic prediction curves obtained by training MaxEnt on one parameter at a time, i.e., excluding the other parameters.

is slightly centered on lower values with respect to the previous  $0.03$  ( $0.01$ ;  $0.4$ )  $g\ C\ m^{-2}\ day^{-1}$  estimate; average air temperature is a bit higher than the previous  $11.8$  ( $8.0$ ;  $16.0$ )  $^{\circ}C$  estimate; and precipitation is substantially equal to the previous  $0.3$  ( $0.2$ ;  $0.45$ )  $10^{-4}\ kg\ m^{-2}\ s^{-1}$  estimate.

#### 4.2 Global-scale Projection

We projected the trained model globally at a  $0.1^{\circ}$  resolution, using the  $0.008$  estimated *balanced* threshold cut-off (Figure 4(a)). The map displays the probability that a  $0.1^{\circ}$  square on the World might foster a high infection rate. It shows probability *peaks*—over the estimated  $0.13$  *peak* threshold—in several known high-infection-rate areas, e.g., the Hubei region in China, the western and eastern coasts of the United States, and the entire Europe (Figure 4(b)). Our *risk map* is much more detailed and accurate than the previous  $0.5^{\circ}$  map and identifies additional high-infection-rate areas (Figures 5(a) and (b)). For example, South America is almost completely covered, because the bias due to the inclusion of inhabited areas within large  $0.5^{\circ}$  cells decreased when using the more accurate  $0.1^{\circ}$  cells. Other high-infection-rate regions that emerge with respect to our previous model are several countries in Asia (from Iran to Pakistan and Kazakhstan), New Zealand, and the Arabian peninsula. Also, Africa, Canada, and the United States are much better covered.

As a numerical assessment, we evaluated the accuracy of our risk map to predict real data (Table 3 and Figure 4(c)). Accuracy on low- and high-risk prediction reached  $80\%$ , with  $152$  countries (and regions) correctly identified as high-infection-rate areas (i.e., as true positives) and other  $44$  as low-infection-rate areas over a total of  $245$  comparisons. Cohen's Kappa was equal to  $0.5$ , which corresponds to *good* agreement according to Fleiss' classification. This result indicates that the model is generally accurate at predicting infection rate across World countries. We also verified that adding more World cities to the training set, e.g., Madrid, Istanbul, and Buenos Aires, did not increment this performance.

Our previous model reached  $77.25\%$  accuracy and *good* agreement ( $Kappa = 0.46$ ) on the identification of high-infection-rate countries only, but a complete accuracy calculation was not possible, because information on low-infection-rate countries was insufficient. Our new model has an  $86.9\%$

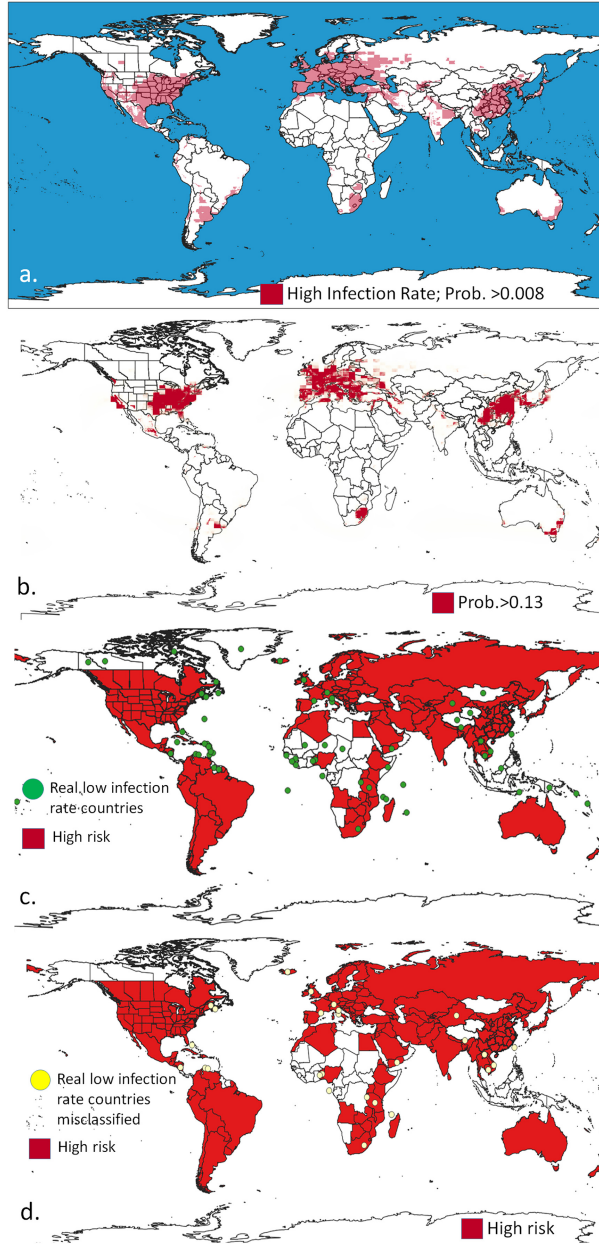


Fig. 4. Visualization of the output produced by our model: (a) distribution of high-infection-rate areas using the MaxEnt *balanced threshold* (0.008), (b) probability peak areas in the MaxEnt distribution (0.13 threshold), (c) overlap between low infection rate countries extracted from real data and our risk map, and (d) highlight of low infection rate countries not predicted by our model.

accuracy at detecting high-infection-rate countries (152 over 175, Table 3). Thus, increasing the model's spatial resolution had a highly positive effect on performance (+11.1% relative accuracy).

Most of the regions at risk that our model did not detect (Figure 4(d)) were small countries (e.g., Isle of Man, Mayotte, Bahamas, Andorra, Liechtenstein) whose statistics were sometimes

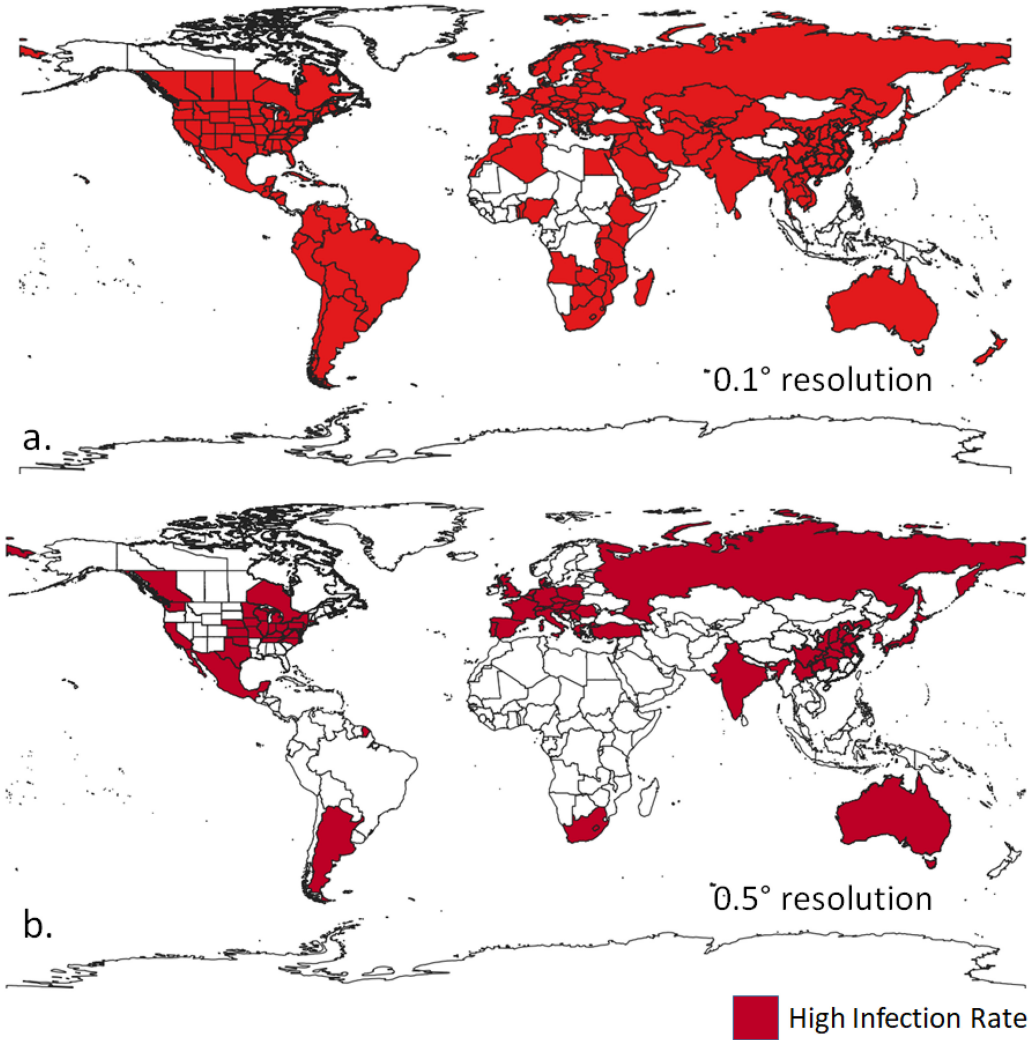


Fig. 5. Visual comparison between (a) our new high-infection-rate risk map based on a  $0.1^\circ$  resolution MaxEnt model and (b) our previous risk map based on a  $0.5^\circ$  resolution MaxEnt model.

incorporated in other countries' statistics (e.g., in Italian statistics as for Vatican State and San Marino). However, there were also large countries wrongly classified as high-infection-rate countries. Among these, Iceland is a particular case, because it experienced several infection waves (thus, it is a high-risk COVID-19 zone), but infection *rate* was never high. Other misclassified countries were Laos, Cambodia, Vietnam, Taiwan, Yemen, Costa Rica, and Xinjiang. However, there is a margin of uncertainty on the reliability of the reports from these countries [11, 60, 78, 94]. There were also 23 countries wrongly classified as low-infection-rate countries (i.e., false negatives), primarily in Indonesia and Africa. These errors were due to the low representativeness of Italian provinces for these regions [29].

### 4.3 Comparison with Other Methods

We compared our results with those of other studies to assess common and complementary aspects. This comparison was not straightforward because of heterogeneous (i) target variables (e.g.,



Table 3. Summary of Our Model's Performance on the Detection of Low and High Infection Rates in World Countries and Regions

True Positives	152
True Negatives	44
False Positives	26
False Negatives	23
Total Comparisons	245
<b>Total Accuracy</b>	<b>80%</b>
<b>Accuracy on high-infection-rate regions</b>	<b>86.9%</b>
<b>Accuracy on low-infection-rate regions</b>	<b>62.9%</b>
<b>Total Agreement - Kappa</b>	<b>0.5</b>
Kappa Interpretation (Fleiss)	Good

hospitalizations, infection cases, or infection rates), (ii) predictive variables (e.g., social, mobility-related, or environmental variables), (iii) analysis scales and resolutions (e.g., global, regional, or city scale), and (iv) modelling approaches (e.g., regression or machine learning-based models). Nevertheless, it was possible to evaluate the agreement of our model on general properties and patterns indicated by the other models. For example, some studies—using linear analyses—have indicated that high altitude may protect from COVID-19, based on the low number of cases reported in Tibet, Bolivia, and Ecuador [5]. Our results do not suggest a global dependency between low COVID-19 infection rate and high altitude. Other studies have indeed highlighted that there is no global dependency between altitude and COVID-19 spread. For example, a low fatality rate was observed in Cusco (Peru) [124], which is in a low probability zone also in our map (Figure 6(a)). However, higher rates were observed in several other high-altitude Peru regions (e.g., Cajatambo and Jauja) [124], which confutes the existence of an inter-dependency between fatality rate and altitude. This result finds confirmation in the overall low overlap between our high-infection-rate areas and the Peruvian Andes (Figure 6(a)), which demonstrates that our model can be informative also at the regional scale.

Advanced time series forecasting techniques and deep learning models have been used to predict short-term COVID-19 cases, recoveries, and deaths in high-infection-rate areas [3]. In this context, multi-seasonal signal processing techniques may outperform deep learning models, when few training data are available, because deep learning models require many examples to accurately infer and forecast the underlying structure of a time series [35, 56]. These techniques can reach a high prediction accuracy and highlight macroscopic patterns in the analyzed regions. For example, they predicted COVID-19 cases seven days ahead in Australia in 2020 (with a 94.8% accuracy) and confirmed that the coastal areas were the most impacted by the virus spread [3]. The basis of this analysis was a set of known high-infection-rate areas, most of which (78%) are also predicted by our model (Figure 6(b)). This observation indicates that our model could support these time-series analyses by suggesting the areas on which the forecasts should concentrate.

Our model can also complement the information explored by social models. For example, socioeconomic and environmental factors (e.g., wind speed, temperature, and suspended particulate matter) coupled with built-environment and high population density can foster COVID-19 infection rate, as highlighted by a recent linear correlative analysis [81]. This analysis showed a westward trend in King County (Washington, United States) in 2020. Our model agrees with this finding, since it indicates a high-infection-rate probability in the King County western area (Figure 6(c)). This observation may also indicate that our MaxEnt-modelled distribution indirectly depends on urbanization.

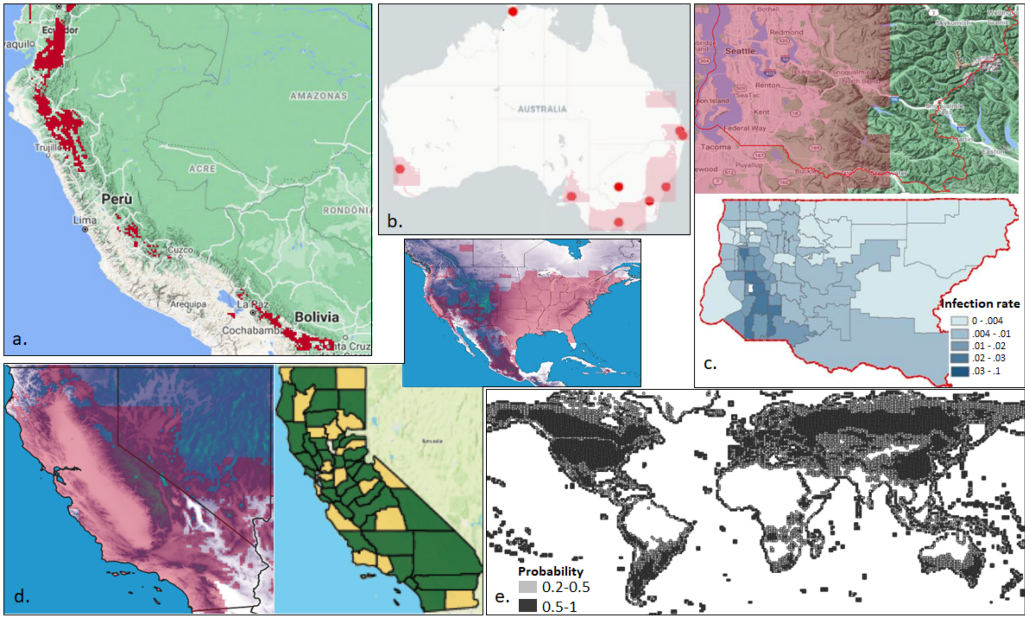


Fig. 6. Comparison between our results and other studies: (a) distribution of high-infection-rate areas (*MaxEnt value* > 0.008) in Peru (in red) that confirms poor correlation of our model with altitude [124]; (b) overlap between the 2020 high impact locations in Australia (red dots) [3] and our high-infection-rate areas (light-red zones); (c) estimated high-infection-rate areas (in red, in the upper figure) in King County (Washington, U.S.A.) vs. actual infection rate in 2020 [81]; (d) estimated high-infection-rate areas (left figure) vs. frequent-hospitalization (green and yellow areas in the right figure) and good-forecast (green areas only) counties in California [123], with a reference image reporting our estimated high-infection-rate areas in North America (upper-right figure); (e) averaged map of the global-scale model described in Reference [4] that predicts potential COVID-19 outbreak based on climatic data.

A recent deep-learning modelling approach has recently investigated the relation between air quality and COVID-19 to predict the number of hospitalisations in the United States 15 days ahead [123]. This study used air temperature, humidity, and aerosol optical depth as predictive variables. The contribution by the predictors to the forecasts depended on the study region. For example, almost all Californian counties reported a high number of cases in 2020, and the predictive environmental variables positively contributed to the forecasts in most of them (Figure 6(d), green areas). Notably, our map also reports that most California counties are at a high-infection-rate risk (Figure 6(d), left image). It also estimates a low probability for the “corner” counties (Del Norte, Modoc, Imperial, and partially in Monterey) where the deep-learning model indicates low forecast contribution by the predictive environmental variables. Our map indicates a high risk also for internal counties (e.g., Tulare, Mariposa, and San Joaquin)—that actually experienced a high-infection-rate in 2020 [46]—where the other model received low forecast contribution, likely because it used different variables (especially aerosol optical depth instead of carbon dioxide). Thus, in this case, our model may suggest effective predictive variables to include in the forecasts.

At an early stage of the pandemic, in 2020, a global-scale seasonal model predicted climate suitability for COVID-19 outbreak using an ensembled machine-learning model [4]. This approach used air temperature, monthly precipitation, surface shortwave radiation, and evapotranspiration as predictor variables and trained an ensemble of 10 different machine learning models with global

COVID-19 case reports. The target variable (infection outbreak) was slightly different from our target variable (high-infection rate). The results indicated that temperate warm and cold climates would be more favorable to COVID-19 spread, whereas tropical climates would be less favorable. However, estimation uncertainty was very high: The predicted ranges having the most significant influence on outbreaks were 9.14 (−11.43; 27.15) °C for temperature and 71.65 (3.19; 236.50) mm for precipitation. We transformed this seasonal ensemble model into an annual-average model to conduct a fair comparison with our model (which is not seasonal). To this aim, we averaged the images (since data were not available) reported in Reference [4] to obtain a coarse annual probability distribution that could highlight macroscopic probability patterns (Figure 6(e)). Very similar patterns between the two distributions emerge after this transformation (Figures 4(a), red areas, and 6(e), black areas), e.g., in southern Australian coasts, eastern United States, Mexico, southern China, Argentina, and South Africa. However, significant discrepancies exist, especially in Canada and Russia, where the annual-average model estimates extended areas with high outbreak probability. Canada and Russia actually reported low infection rates throughout 2020 and 2021 except for specific cities [46], in line with our results. The over-prediction by the ensemble model could be due to the aggregations and projections it operates over the Köppen-Geiger climate zones, which consequently highlights the higher precision of our model.

## 5 DISCUSSION AND CONCLUSIONS

This article has presented a high-resolution global map of high-infection-rate risk for COVID-19, based on average annual temperature, precipitation, and CO<sub>2</sub> data. In particular, CO<sub>2</sub> was confirmed to be the main predictive parameter, which is directly correlated with air pollution [62]. Our *risk map* was able to predict ~87% of the countries with an actual risk of high infection increase, and ~80% of both high- and low-infection-rate countries. Our model indicates that high-risk global areas are characterized by a moderate-high level of CO<sub>2</sub>, moderate-low temperatures, and a moderate level of precipitation, in agreement with other studies [72, 116].

Using Italian data to train the model against temperature, precipitation, and CO<sub>2</sub> parameters may seem a limitation, because the global data may go beyond the ranges of the predictor variables. Furthermore, infection reports are affected by human-related factors (mobility, policy, interactions) specific to the country. However, the Italian data of March 2020 represented a golden reference for our experiment. The reports indicated symptomatic cases only, with a high accuracy reflecting the severity of the pandemic, within a context of high mobility and relatively low restrictions (Section 3.2). This condition is ideal when building an ENM model for an epidemic. However, infection reports by different countries were too heterogeneous and poorly comparable and could not be used in our ENM. Finally, the large variability of environmental conditions across Italy allowed covering wide ranges of the predictor variables, because they included from hot-humid and desert areas (in the South) to cold and elevated areas (in the North).

Our model accounts for climatic factors only; thus, it is meant to be coupled with monthly climate and human-mobility and -interaction models to generate a complete view of the pandemic. Indeed, monthly climate and human dynamics are fundamental to building regional-scale models, e.g., predicting COVID-19 spread in one country. If these factors were modelled as two probability distributions ( $P_{human}$  and  $P_{weather}$ ), then our distribution  $\pi(\bar{x}) = P_{climate}$  could be an additional factor to account for the general climatic predisposition of each 0.1° cell in the region to increase infection rate. Thus, an overall score for infection spread that took into account all factors could be calculated as  $COVID\ risk = P_{human} \cdot P_{weather} \cdot P_{climate}$ . The high resolution of our model would make this multiplication useful even for regional-scale analyses. The weight of our distribution in the multiplication would be higher during lockdown periods (when mobility is low) and when monthly weather was within (or completely outside, by contrast) our predicted niche.

The produced prediction function ( $P_{climate}$ ) is more complex than the correlation assessments between COVID-19 and the single analyzed parameters that other studies have proposed [2, 13, 61, 73, 82, 132, 142]. Our function contains complementary information with respect to alternative models for COVID-19 spread prediction and can be integrated with these other models. For example, **Susceptible-Exposed-Infected-Recovered (SEIR)** models [63] have been successfully used to study the pandemic dynamic and understand the impact of environmental and social conditions on its spread [12]. SEIR models aim to predict the number of potentially infected individuals, the transmission rate, and the recovering rate by automatically inferring the model parameters from disease reports and statistics. However, these models infrequently integrate information on the driving forces that may foster infection transmission, e.g., mobility, weather, air pollution, and climate. For example, the seasonal variability of COVID-19 cases has been explored through per-season SEIR models adapted to seasonal infection reports, without explicitly considering climatic factors as seasonal factors [93]. Instead, explicitly modelling these factors would enhance SEIR models' prediction accuracy [12, 21, 26]. Our  $P_{climate}$  distribution could indeed be integrated into SEIR models as an additional factor in the "number of susceptible individuals" equation, in a similar way to what was recently proposed for environment-to-human infection modelling [90]. Otherwise, it could be a multiplication factor for the SEIR infection rate parameter ( $\beta$ ) to simulate the potential increment in the number of daily people infected by one individual due to particular climatic conditions.

Other analyses that could integrate our prediction function would be statistical analyses based on the *big data* of online searches and Google Trends. These experiments evaluate the correlation between the general interest in COVID-19 and death cases [85] but currently cannot explain the reasons behind the periodic increase of this correlation. Furthermore, our function could be coupled with risk assessments based on spatio-temporal clustering of COVID-19 cases [45] by weighting the distances of the points from the clusters by our function values. Therefore, this weight would represent the potential inertia that the infection locations potentially oppose to infection transmission due to their climatic conditions.

Our map is distributed as a FAIR dataset [30] that facilitates its integration with other models. Furthermore, software availability as an Open Science-oriented Web service [24] makes our approach reusable to build predictive models for new epidemics and COVID-19 variants.

## REFERENCES

- [1] Mohsen Ahmadi, Abbas Sharifi, Shadi Dorosti, Saeid Jafarzadeh Ghouschi, and Negar Ghanbari. 2020. Investigation of effective climatology parameters on COVID-19 outbreak in Iran. *Sci. Total Environ.* 729 (2020), 138705.
- [2] N. Al-Rousan and H. Al-Najjar. 2020. The correlation between the spread of COVID-19 infections and weather variables in 30 Chinese provinces and the impact of Chinese government mitigation plans. *Eur. Rev. Med. Pharmacol. Sci.* 24 (2020), 4565–4571.
- [3] Moutaz Alazab, Albara Awajan, Abdelwadood Mesleh, Ajith Abraham, Vansh Jatana, and Salah Alhyari. 2020. COVID-19 prediction and detection using deep learning. *Int. J. Comput. Inf. Syst. Industr. Manag. Applic.* 12 (2020), 168–181.
- [4] Miguel B. Araujo and Babak Naimi. 2020. Spread of SARS-CoV-2 Coronavirus likely to be constrained by climate. (2020). DOI: <https://doi.org/10.1101/2020.03.12.20034728>
- [5] Christian Arias-Reyes, Natalia Zubieta-DeUrioste, Liliana Poma-Machicao, Fernanda Aliaga-Raduan, Favio Carvajal-Rodriguez, Mathias Dutschmann, Edith M. Schneider-Gasser, Gustavo Zubieta-Calleja, and Jorge Soliz. 2020. Does the pathogenesis of SARS-CoV-2 virus decrease at high-altitude? *Respir. Physiol. Neurobiol.* 277 (2020), 103443.
- [6] M. Assante, L. Candela, D. Castelli, R. Cirillo, G. Coro, L. Frosini, L. Leli, F. Mangiacrapa, P. Pagano, G. Panichi et al. 2019. Enacting open science by D4Science. *Fut. Gen. Comput. Syst.* 101 (2019), 555–563.
- [7] Hamada S. Badr, Hongru Du, Maximilian Marshall, Ensheng Dong, Marietta M. Squire, and Lauren M. Gardner. 2020. Association between mobility patterns and COVID-19 transmission in the USA: A mathematical modelling study. *Lancet Infect. Dis.* 20, 11 (2020), 1247–1254.
- [8] Roger A Baldwin. 2009. Use of maximum entropy modeling in wildlife research. *Entropy* 11, 4 (2009), 854–866.



- [9] Edward B. Baskerville and Sarah Cobey. 2017. Does influenza drive absolute humidity? *Proc. Nat. Acad. Sci.* 114, 12 (2017), E2270–E2271.
- [10] Michael R. Berthold, Nicolas Cebon, Fabian Dill, Thomas R. Gabriel, Tobias Kötter, Thorsten Meinl, Peter Ohl, Kilian Thiel, and Bernd Wiswedel. 2009. KNIME-the Konstanz information miner: Version 2.0 and beyond. *ACM SIGKDD Explor. Newslett.* 11, 1 (2009), 26–31.
- [11] Emilie S. Koum Besson, Andy Norris, Abdulla S. Bin Ghouth, Terri Freemantle, Mervat Alhaffar, Yolanda Vazquez, Chris Reeve, Patrick J. Curran, and Francesco Checchi. 2021. Excess mortality during the COVID-19 pandemic: A geospatial and statistical analysis in Aden governorate, Yemen. *BMJ Global Health* 6, 3 (2021), e004564.
- [12] Hemant Bherwani, Ankit Gupta, Saima Anjum, Avneesh Anshul, and Rakesh Kumar. 2020. Exploring dependence of COVID-19 on environmental factors and spread prediction in India. *npj Clim. Atmos. Sci.* 3, 1 (2020), 1–13.
- [13] Vanessa Bianconi, Paola Bronzo, Maciej Banach, Amirhossein Sahebkar, Massimo R. Mannarino, and Matteo Pirro. 2020. Particulate matter pollution and the COVID-19 outbreak: Results from Italian regions and provinces. *Arch. Med. Sci.* 16, 5 (2020), 985.
- [14] Irina V. Biktasheva. 2020. Role of a habitat's air humidity in Covid-19 mortality. *Sci. Total Environ.* 736 (2020), 138763.
- [15] CAMS. 2019. Flux inversion reanalysis of global carbon dioxide—Fluxes and atmospheric concentrations. Retrieved from <https://atmosphere.copernicus.eu/catalogue/product/urn:x-wmo:md:int.ecmwf::copernicus:cams:prod:rean:co2:pid286>.
- [16] Colin J. Carlson, Eric R. Dougherty, and Wayne Getz. 2016. An ecological assessment of the pandemic threat of Zika virus. *PLoS Neglect. Tropic. Dis.* 10, 8 (2016), e0004968.
- [17] CDO. 2021. The Climate Data Operators software. Retrieved from [https://www.unidata.ucar.edu/software/netcdf/workshops/2012/third\\_party/CDO.html](https://www.unidata.ucar.edu/software/netcdf/workshops/2012/third_party/CDO.html).
- [18] Meng-Chun Chang, Rebecca Kahn, Yu-An Li, Cheng-Sheng Lee, Caroline O. Buckee, and Hsiao-Han Chang. 2021. Variation in human mobility and its impact on the risk of future COVID-19 outbreaks in Taiwan. *BMC Public Health* 21, 1 (2021), 1–10.
- [19] Biqing Chen, Hao Liang, Xiaomin Yuan, Yingying Hu, Miao Xu, Yating Zhao, Binfen Zhang, Fang Tian, and Xuejun Zhu. 2020. Roles of meteorological conditions in COVID-19 transmission on a worldwide scale. *medRxiv*. <https://www.medrxiv.org/content/early/2020/10/10/2020.03.16.20037168>.
- [20] Mu-Jean Chen, Chuan-Yao Lin, Yi-Ting Wu, Pei-Chih Wu, Shih-Chun Lung, and Huey-Jen Su. 2012. Effects of extreme precipitation to the distribution of infectious diseases in Taiwan, 1994–2008. *PloS One* 7, 6 (2012), e34651.
- [21] Yi-Cheng Chen, Ping-En Lu, and Cheng-Shang Chang. 2020. A time-dependent SIR model for COVID-19. DOI: <https://doi.org/10.1109/TNSE.2020.3024723>
- [22] Vinay Kumar Reddy Chimmula and Lei Zhang. 2020. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solit. Fract.* 135 (2020), 109864.
- [23] I. Chuine and E. G. Beaubien. 2008. Phenology is a major determinant of tree species range. *Ecol. Lett.* 4, 5 (2008), 500–510.
- [24] CNR. 2019. Maximum Entropy Model Web Processing Service. Retrieved from [https://services.d4science.org/group/biodiversitylab/data-miner?OperatorId=org.gcube.dataanalysis.wps.statisticalmanager.synchserver.mappedclasses.transducerers.MAX\\_ENT\\_NICHE\\_MODELLING](https://services.d4science.org/group/biodiversitylab/data-miner?OperatorId=org.gcube.dataanalysis.wps.statisticalmanager.synchserver.mappedclasses.transducerers.MAX_ENT_NICHE_MODELLING).
- [25] Jacob Cohen et al. 1960. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 1 (1960), 37–46.
- [26] Rinaldo M. Colombo, Mauro Garavello, Francesca Marcellini, and Elena Rossi. 2020. An age and space structured SIR model describing the Covid-19 pandemic. *J. Math. Industr.* 10, 1 (2020), 1–20.
- [27] Daniele Contini and Francesca Costabile. 2020. Does air pollution influence COVID-19 outbreaks? *Atmosphere* 11, 4, Article 377 (2020). <https://www.mdpi.com/2073-4433/11/4/37>.
- [28] Ian Cooper, Argha Mondal, and Chris G Antonopoulos. 2020. A SIR model assumption for the spread of COVID-19 in different communities. *Chaos, Solit. Fract.* 139 (2020), 110057.
- [29] Gianpaolo Coro. 2020. A global-scale ecological niche model to predict SARS-CoV-2 coronavirus infection rate. *Ecol. Modell.* 431 (2020), 109187.
- [30] Gianpaolo Coro. 2021. A High-Resolution Global-Scale Model for COVID-19 Infection Rate. Retrieved from <https://zenodo.org/record/4671982>.
- [31] Gianpaolo Coro, Leonardo Candela, Pasquale Pagano, Angela Italiano, and Loredana Liccardo. 2015. Parallelizing the execution of native data mining algorithms for computational biology. *Concurr. Computat.: Pract. Exper.* 27, 17 (2015), 4630–4644.
- [32] Gianpaolo Coro, Anton Ellenbroek, and Pasquale Pagano. 2021. An Open Science approach to infer fishing activity pressure on stocks and biodiversity from vessel tracking data. *Ecol. Inform.* 64 (2021), 101384.
- [33] Gianpaolo Coro, Chiara Magliozzi, Edward Vanden Berghe, Nicolas Bailly, Anton Ellenbroek, and Pasquale Pagano. 2016. Estimating absence locations of marine species from data of scientific surveys in OBIS. *Ecol. Modell.* 323 (2016), 61–76.



- [34] Gianpaolo Coro, Chiara Magliozzi, Anton Ellenbroek, and Pasquale Pagano. 2015. Improving data quality to build a robust distribution model for *Architeuthis dux*. *Ecol. Modell.* 305 (2015), 29–39.
- [35] Gianpaolo Coro, Fabio Valerio Massoli, Antonio Origlia, and Francesco Cutugno. 2021. Psycho-acoustics inspired automatic speech recognition. *Comput. Electric. Eng.* 93 (2021), 107238.
- [36] Gianpaolo Coro, Pasquale Pagano, and Anton Ellenbroek. 2013. Combining simulated expert knowledge with neural networks to produce ecological niche models for *Latimeria chalumnae*. *Ecol. Modell.* 268 (2013), 55–63.
- [37] Gianpaolo Coro, Pasquale Pagano, and Anton Ellenbroek. 2020. Detecting patterns of climate change in long-term forecasts of marine environmental parameters. *Int. J. Dig. Earth* 13, 5 (2020), 567–585. DOI: <https://doi.org/10.1080/17538947.2018.1543365>
- [38] Gianpaolo Coro, Giancarlo Panichi, Paolo Scarponi, and Pasquale Pagano. 2017. Cloud computing in a distributed e-infrastructure using the web processing service standard. *Concurr. Comput.: Pract. Exper.* 29, 18 (2017), e4219.
- [39] Gianpaolo Coro and Eugenio Trumpy. 2020. Geothermal Data NetCDF files and Models on the D4Science e-Infrastructure. Retrieved from <https://data.d4science.org/workspace-explorer-app/?folderId=Qk5YY3JMY0w1ODRqZmxtNkVQNXB6S3ROdWtkU1pYN1UrcXVSaVkyUW0vdmgyYXBYdFJ6Z0JkYWladVU0UWptVQ>.
- [40] Gianpaolo Coro and Eugenio Trumpy. 2020. Predicting geographical suitability of geothermal power plants. *Journal of Cleaner Production* 267 (2020), 121874. DOI: <https://doi.org/10.1016/j.jclepro.2020.121874>
- [41] Gianpaolo Coro, Luis Gonzalez Vilas, Chiara Magliozzi, Anton Ellenbroek, Paolo Scarponi, and Pasquale Pagano. 2018. Forecasting the ongoing invasion of *Lagocephalus scleratus* in the Mediterranean Sea. *Ecol. Modell.* 371 (2018), 37–49.
- [42] Jane Costa and A. Townsend Peterson. 2012. Ecological niche modeling as a tool for understanding distributions and interactions of vectors, hosts, and etiologic agents of Chagas disease. *Adv. Experim. Med. Biol.* 710 (2012), 59–70. DOI: [https://doi.org/10.1007/978-1-4419-5638-5\\_7](https://doi.org/10.1007/978-1-4419-5638-5_7)
- [43] M. R. Mansouri Daneshvar, M. Ebrahimi, A. Sadeghi, and A. Mahmoudzadeh. 2021. Climate effects on the COVID-19 outbreak: A comparative analysis between the UAE and Switzerland. *Model. Earth Syst. Environ.* n/a, n/a (2021), 1–14. <https://link.springer.com/article/10.1007/s40808-021-01110-x>.
- [44] Andrew J. Davison. 2007. *Overview of Classification*. Cambridge University Press, Cambridge, UK, 3–9.
- [45] M. R. Desjardins, Alexander Hohl, and E. M. Delmelle. 2020. Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: Detecting and evaluating emerging clusters. *Appl. Geog.* 118 (2020), 102202.
- [46] Ensheng Dong, Hongru Du, and Lauren Gardner. 2020. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* 20, 5 (2020), 533–534. Retrieved from [https://www.thelancet.com/journals/laninf/article/PIIS1473-3099\(20\)30120-1/fulltext](https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30120-1/fulltext).
- [47] Miroslav Dudík, Steven Phillips, and Robert E. Schapire. 2005. Correcting sample selection bias in maximum entropy density estimation. *Adv. Neural Inf. Process. Syst.* 18 (2005), 323–330.
- [48] David J. D. Earn, Pejman Rohani, Benjamin M. Bolker, and Bryan T. Grenfell. 2000. A simple model for complex dynamical transitions in epidemics. *Science* 287, 5453 (2000), 667–670.
- [49] Maria Effenberger, Andreas Kronbichler, Jae Il Shin, Gert Mayer, Herbert Tilg, and Paul Perco. 2020. Association of the COVID-19 pandemic with internet search volumes: A Google Trends™ analysis. *Int. J. Infect. Dis.* 95 (2020), 192–197.
- [50] Jane Elith and John R. Leathwick. 2009. Species distribution models: Ecological explanation and prediction across space and time. *Ann. Rev. Ecol. Evol. Syst.* 40, 1 (2009), 677–697.
- [51] Jane Elith, Steven J. Phillips, Trevor Hastie, Miroslav Dudík, Yung En Chee, and Colin J. Yates. 2011. A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* 17, 1 (Jan. 2011), 43–57.
- [52] European Commission. 2021. Coronavirus: Commission proposes a Digital Green Certificate. Retrieved from [https://ec.europa.eu/commission/presscorner/detail/en/IP\\_21\\_1181](https://ec.europa.eu/commission/presscorner/detail/en/IP_21_1181).
- [53] Gentile Francesco Ficetola and Diego Rubolini. 2021. Containment measures limit environmental effects on COVID-19 early outbreak dynamics. *Sci. Total Environ.* 761 (2021), 144432. DOI: <https://doi.org/10.1016/j.scitotenv.2020.144432>
- [54] Joseph L. Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychol. Bull.* 76, 5 (1971), 378.
- [55] Trevon L. Fuller, Marius Gilbert, Vincent Martin, Julien Cappelle, Parvize Hosseini, Kevin Y. Njabo, Soad Abdel Aziz, Xiangming Xiao, Peter Daszak, and Thomas B. Smith. 2013. Predicting hotspots for influenza virus reassortment. *Emerg. Infect. Dis.* 19, 4 (2013), 581.
- [56] Daryl L. X. Fung, Calvin S. H. Hoi, Carson K. Leung, and Christine Y. Zhang. 2021. Predictive analytics of COVID-19 with neural networks. In *International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–8.
- [57] Song Gao, Jinneng Rao, Yuhao Kang, Yunlei Liang, Jake Kruse, Doerte Doeffer, Ajay K. Sethi, Juan Francisco Mandujano Reyes, Jonathan Patz, and Brian S. Yandell. 2020. Mobile phone location data reveal the effect and geographic variation of social distancing on the spread of the COVID-19 epidemic. DOI: <https://doi.org/10.1001/jamanetworkopen.2020.20485>

- [58] Marino Gatto, Enrico Bertuzzo, Lorenzo Mari, Stefano Miccoli, Luca Carraro, Renato Casagrandi, and Andrea Rinaldo. 2020. Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. *Proc. Nat. Acad. Sci.* 117, 19 (2020), 10484–10491.
- [59] GEDI. 2020. GEDI Group Visual lab—Coronavirus data and analysis. Retrieved from <https://lab.gedidigital.it/gedi-visual/2020/coronavirus-i-contagi-in-italia/>.
- [60] Elizabeth Gonzalez, Chase Harrison, Katie Hopkins, Luisa Horwitz, Paola Nagovitch, Holly Sonnelan, and Carin Zissis. 2021. The Coronavirus in Latin America. Retrieved from <https://www.as-coa.org/articles/coronavirus-latin-america>.
- [61] Sachin S. Gunthe, Basudev Swain, Satya S. Patra, and Aneesh Amte. 2020. On the global trends and spread of the COVID-19 outbreak: Preliminary assessment of the potential relation between location-specific temperature and UV index. *J. Public Health* 42, 4 (2020), 1–10.
- [62] Yang Han, Jacqueline C. K. Lam, Victor O. K. Li, Peiyang Guo, Qi Zhang, Andong Wang, Jon Crowcroft, Shanshan Wang, Jinqi Fu, Zafar Gilani et al. 2020. The effects of outdoor air pollution concentrations and lockdowns on Covid-19 infections in Wuhan and other provincial capitals in China. Retrieved from <https://www.preprints.org/manuscript/202003.0364/v1>.
- [63] Shaobo He, Yuexi Peng, and Kehui Sun. 2020. SEIR modeling of the COVID-19 and its dynamics. *Nonlin. Dynam.* 101, 3 (2020), 1667–1680.
- [64] Thomas S. Higgins, Arthur W. Wu, Dhruv Sharma, Elisa A. Illing, Kolin Rubel, Jonathan Y. Ting, and Snot Force Alliance. 2020. Correlations of online search engine trends with coronavirus disease (COVID-19) incidence: Infodemiology study. *JMIR Public Health Surveill.* 6, 2 (2020), e19702.
- [65] Songhua Hu, Chenfeng Xiong, Mofeng Yang, Hannah Younes, Weiyu Luo, and Lei Zhang. 2021. A big-data driven approach to analyzing and modeling human mobility trend under non-pharmaceutical interventions during COVID-19 pandemic. *Transport. Res. Part C: Emerg. Technol.* 124 (2021), 102955.
- [66] ISS. 2021. Istituto Superiore di Sanità - COVID-19 integrated surveillance data in Italy. Retrieved from <https://www.epicentro.iss.it/en/coronavirus/sars-cov-2-dashboard>.
- [67] Italian Civil Protection Department. 2020. Interface for browsing and downloading COVID-19 data. Retrieved from <http://opendatadpc.maps.arcgis.com/apps/opsdashboard/index.html/b0c68bce2cce478eaac82fe38d4138b1>.
- [68] Italian Ministry of Health. 2020. FAQ on Covid-19. Retrieved from <http://www.salute.gov.it/portale/malattieInfettive/dettaglioFaqMalattieInfettive.jsp?lingua=italiano&id=228>.
- [69] Yuhao Kang, Song Gao, Yunlei Liang, Mingxiao Li, Jinmeng Rao, and Jake Kruse. 2020. Multiscale dynamic human mobility flow dataset in the US during the COVID-19 epidemic. *Sci. Data* 7, 1 (2020), 1–13.
- [70] Lisa K. Koch, Sarah Cunze, Antje Werblow, Judith Kochmann, Dorian D. Dörge, Heinz Mehlhorn, and Sven Klimpel. 2016. Modeling the habitat suitability for the arbovirus vector *Aedes albopictus* (Diptera: Culicidae) in Germany. *Parasitol. Res.* 115, 3 (2016), 957–964.
- [71] Sachiko Kodera, Essam A. Rashed, and Akimasa Hirata. 2020. Correlation between COVID-19 morbidity and mortality rates in Japan and local population density, temperature, and absolute humidity. *Int. J. Environ. Res. Pub. Health* 17, 15 (2020), 5477.
- [72] Piotr Andrzej Kowalski, Marcin Szwagrzyk, Jolanta Kiełpińska, Aleksander Konior, and Maciej Kusy. 2021. Numerical analysis of factors, pace and intensity of the coronavirus (COVID-19) epidemic in Poland. *Ecol. Inform.* 63 (2021), 101284. DOI: <https://doi.org/10.1016/j.ecoinf.2021.101284>
- [73] Yasuhiro Kubota, Takayuki Shiono, Buntarou Kusumoto, and Junichi Fujinuma. 2020. Multiple drivers of the COVID-19 spread: The roles of climate, international mobility, and region-specific conditions. *PloS One* 15, 9 (2020), e0239385.
- [74] Timothy Lebo, Satya Sahoo, Deborah McGuinness, Khalid Belhajjame, James Cheney, David Corsar, Daniel Garijo, Stian Soiland-Reyes, Stephan Zednik, and Jun Zhao. 2013. Prov-O: The Prov ontology. Retrieved from <https://www.w3.org/TR/prov-o/>.
- [75] Cuilian Li, Li Jia Chen, Xueyu Chen, Mingzhi Zhang, Chi Pui Pang, and Haoyu Chen. 2020. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. *Eurosurveillance* 25, 10 (2020), 2000199.
- [76] Ruiyun Li, Sen Pei, Bin Chen, Yimeng Song, Tao Zhang, Wan Yang, and Jeffrey Shaman. 2020. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* 368, 6490 (2020), 489–493.
- [77] Li-Lin Liang, Ching-Hung Tseng, Hsiu J. Ho, and Chun-Ying Wu. 2020. Covid-19 mortality is negatively associated with test number and government effectiveness. *Sci. Rep.* 10, 1 (2020), 1–7.
- [78] Yi-Ting Lien. 2021. Why China's COVID-19 Disinformation Campaign Isn't Working in Taiwan. Retrieved from <https://thediplomat.com/2020/03/why-chinas-covid-19-disinformation-campaign-isnt-working-in-taiwan/>.
- [79] Ariel Linden. 2006. Measuring diagnostic and predictive accuracy in disease management: An introduction to receiver operating characteristic (ROC) analysis. *J. Eval. Clin. Pract.* 12, 2 (2006), 132–139.

- [80] Kevin Linka, Alain Goriely, and Ellen Kuhl. 2021. Global and local mobility as a barometer for COVID-19 dynamics. *Biomech. Model. Mechanobiol.* 20, 2 (2021), 651–669.
- [81] Chao Liu, Zerun Liu, and Chenghe Guan. 2021. The impacts of the built environment on the incidence rate of COVID-19: A case study of King County, Washington. *Sustain. Cities Societ.* 74 (2021), 103144.
- [82] Wei Luo, Maimuna Majumder, Dianbo Liu, Canelle Poirier, Kenneth Mandl, Marc Lipsitch, and Mauricio Santillana. 2020. The role of absolute humidity on transmission rates of the COVID-19 outbreak. (2020). <https://dash.harvard.edu/handle/1/42639515>.
- [83] Yueling Ma, Yadong Zhao, Jiangtao Liu, Xiaotao He, Bo Wang, Shihua Fu, Jun Yan, Jingping Niu, Ji Zhou, and Bin Luo. 2020. Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China. *Sci. Total Environ.* 724 (2020), 138226.
- [84] Hirohiko Masunaga. 2012. Short-term versus climatological relationship between precipitation and tropospheric humidity. *J. Clim.* 25, 22 (2012), 7983–7990.
- [85] Amaryllis Mavragani. 2020. Tracking COVID-19 in Europe: Infodemiology approach. *JMIR Public Health Surveill.* 6, 2 (2020), e18941.
- [86] Kim A. Medley. 2010. Niche shifts during the global invasion of the Asian tiger mosquito, *Aedes albopictus* Skuse (Culicidae), revealed by reciprocal distribution models. *Global Ecol. Biogeog.* 19, 1 (2010), 122–133.
- [87] Robin H. Miller, Penny Masuoka, Terry A. Klein, Heung-Chul Kim, Todd Somer, and John Grieco. 2012. Ecological niche modeling to estimate the distribution of Japanese encephalitis virus in Asia. *PLoS Neglect. Tropic. Dis.* 6, 6 (2012), e1678.
- [88] Usha Kant Misra and Jayantee Kalita. 2010. Overview: Japanese encephalitis. *Prog. Neurobiol.* 91, 2 (2010), 108–120.
- [89] Stephen S. Morse, Jonna A. K. Mazet, Mark Woolhouse, Colin R. Parrish, Dennis Carroll, William B. Karesh, Carlos Zambrana-Torrel, W. Ian Lipkin, and Peter Daszak. 2012. Prediction and prevention of the next pandemic zoonosis. *Lancet* 380, 9857 (2012), 1956–1965.
- [90] Samuel Mwalili, Mark Kimathi, Viona Ojiambo, Duncan Gathungu, and Rachel Mbogo. 2020. SEIR model for COVID-19 dynamics incorporating the environment and social distancing. *BMC Res. Notes* 13, 1 (2020), 1–5.
- [91] NASA-NEX. 2015. Space Apps Challenge. Retrieved from [https://web.archive.org/web/20150924122046/https://nex.nasa.gov/nex/projects/1348/wiki/general\\_data\\_access\\_and\\_apis/](https://web.archive.org/web/20150924122046/https://nex.nasa.gov/nex/projects/1348/wiki/general_data_access_and_apis/).
- [92] NCO. 2021. The NetCDF Operators software. Retrieved from <http://nco.sourceforge.net/>.
- [93] Richard A. Neher, Robert Dyrda, Valentin Druelle, Emma B. Hodcroft, and Jan Albert. 2020. Potential impact of seasonal forcing on a SARS-CoV-2 pandemic. DOI: <https://doi.org/10.4414/smww.2020.20224>
- [94] Tran Nguyen. 2021. COVID-19 - What do we know about the situation in Vietnam? Retrieved from <https://towardsdatascience.com/covid-19-what-do-we-know-about-the-situation-in-vietnam-82c195163d7e>.
- [95] Dirk Nikolaus Karger, Dirk R. Schmatz, Gabriel Dettling, and Niklaus E. Zimmermann. 2020. High resolution monthly precipitation and temperature timeseries for the period 2006–2100. *Sci. Data* 7, 248 (2020), 1–10.
- [96] Dirk Nikolaus Karger, Dirk R. Schmatz, Gabriel Dettling, and Niklaus E. Zimmermann. 2021. EnviDat repository for high resolution monthly precipitation and temperature timeseries for the period 2006–2100. Retrieved from <https://doi.org/10.16904/envi.dat.124>
- [97] Ji-Eun Park, Woo-Sik Son, Yeonhee Ryu, Soo Beom Choi, Okyu Kwon, and Insung Ahn. 2020. Effects of temperature, humidity, and diurnal temperature range on influenza incidence in a temperate region. *Influenz. Other Respir. Virus.* 14, 1 (2020), 11–18.
- [98] Jonathan A. Patz. 1998. Predicting key malaria transmission factors, biting and entomological inoculation rates, using modelled soil moisture in Kenya. *Tropic. Med. Int. Health* 3, 10 (1998), 818–827.
- [99] Richard G. Pearson. 2007. Species’ distribution modeling for conservation educators and practitioners. *Synth. Amer. Mus. Nat. Hist.* 50 (2007), 54–89.
- [100] Rui Pei, Danielle Cosme, Mary E. Andrews, Bradley D. Mattan, and Emily Falk. 2020. Cultural influence on COVID-19 cognitions and growth speed: The role of cultural collectivism. DOI: <https://doi.org/10.31234/osf.io/fet6z>
- [101] A. Townsend Peterson, R. Ryan Lash, Darin S. Carroll, and Karl M. Johnson. 2006. Geographic potential for outbreaks of Marburg hemorrhagic fever. *Amer. J. Tropic. Med. Hyg.* 75, 1 (2006), 9–15.
- [102] A. Townsend Peterson and David A. Vieglais. 2001. Predicting species invasions using ecological niche modeling: New approaches from bioinformatics attack a pressing problem: A new approach to ecological niche modeling, based on new tools drawn from biodiversity informatics, is applied to the challenge of predicting potential species’ invasions. *BioScience* 51, 5 (2001), 363–371.
- [103] Steven J. Phillips, Robert P. Anderson, Miroslav Dudík, Robert E. Schapire, and Mary E. Blair. 2017. Opening the black box: An open-source release of MaxEnt. *Ecography* 40, 7 (2017), 887–893.
- [104] Steven J. Phillips, Robert P. Anderson, and Robert E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Modell.* 190, 3–4 (2006), 231–259.

- [105] S. J. Phillips and M. Dudík. 2008. Modeling of species distributions with MaxEnt: New extensions and a comprehensive evaluation. *Ecography* 31 (2008), 161–175.
- [106] Steven J. Phillips and Miroslav Dudík. 2008. Modeling of species distributions with MaxEnt: New extensions and a comprehensive evaluation. *Ecography* 31, 2 (2008), 161–175.
- [107] Steven J. Phillips, Miroslav Dudík, and Robert E. Schapire. 2004. A maximum entropy approach to species distribution modeling. In *21st International Conference on Machine Learning*. ACM, Association for Computing Machinery, New York, 83.
- [108] Steven J. Phillips and Miroslav Dudík. 2008. Modeling of species distributions with MaxEnt: New extensions and a comprehensive evaluation. *Ecography* 31, 2 (2008), 161–175.
- [109] Steven J. Phillips, Miroslav Dudík, and Robert E. Schapire. 2019. MaxEnt software for modeling species niches and distributions (Version 3.4.1). Retrieved from [http://biodiversityinformatics.amnh.org/open\\_source/maxent/](http://biodiversityinformatics.amnh.org/open_source/maxent/).
- [110] David N. Prata, Waldecy Rodrigues, and Paulo H. Bernejo. 2020. Temperature significantly changes COVID-19 transmission in (sub) tropical cities of Brazil. *Sci. Total Environ.* 729 (2020), 138862.
- [111] QGIS. 2020. Quantum GIS geographic information system. Retrieved from <https://qgis.org/en/site/>.
- [112] Chong Qi, Daniel Karlsson, Karl Sallmen, and Ramon Wyss. 2020. Model studies on the COVID-19 pandemic in Sweden. Retrieved from [arXiv:q-bio.PE/2004.01575](https://arxiv.org/abs/2004.01575).
- [113] Hongyan Ren, Lu Zhao, An Zhang, Liuyi Song, Yilan Liao, Weili Lu, and Cheng Cui. 2020. Early forecasting of the potential risk zones of COVID-19 in China's megacities. *Sci. Total Environ.* 729 (2020), 138995.
- [114] Ian W. Renner and David I. Warton. 2013. Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. *Biometrics* 69, 1 (2013), 274–281.
- [115] Reuters. 2020. Special Report: Italy and South Korea virus outbreaks reveal disparity in deaths and tactics. Retrieved from <https://www.reuters.com/article/us-health-coronavirus-response-specialre/special-report-italy-and-south-korea-virus-outbreaks-reveal-disparity-in-deaths-and-tactics-idUSKBN20Z27P>.
- [116] Mohammad M. Sajadi, Parham Habibzadeh, Augustin Vintzileos, Shervin Shokouhi, Fernando Miralles-Wilhelm, and Anthony Amoroso. 2020. Temperature and latitude analysis to predict potential spread and seasonality for COVID-19. Retrieved from [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3550308](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3550308).
- [117] Abdallah M. Samy and A. Townsend Peterson. 2016. Climate change influences on the global potential distribution of bluetongue virus. *PLOS ONE* 11 (2016), 1–13. <https://doi.org/10.1371/journal.pone.0163863>
- [118] Abdallah M. Samy, Stephanie M. Thomas, Ahmed Abd El Wahed, Kevin P. Cohoon, and A. Townsend Peterson. 2016. Mapping the global geographic potential of Zika virus spread. *Memorias do Instituto Oswaldo Cruz* 111, 9 (2016), 559–560.
- [119] Giovanni Scabbia, Antonio Sanfilippo, Annamaria Mazzoni, Dunia Bachour, Daniel Perez-Astudillo, Veronica Bermudez Benito, Etienne Wey, Mathilde Marchand-Lasserre, and Laurent Saboret. 2020. Does Climate Help Modeling COVID-19 Risk and to What Extent? DOI: <https://doi.org/10.21203/rs.3.rs-108398/v1>
- [120] Nicola Scafetta. 2020. Distribution of the SARS-CoV-2 pandemic and its monthly forecast based on seasonal climate patterns. *Int. J. Environ. Res. Public Health* 17, 10 (2020), 3493.
- [121] Marten Scheffer. 2009. *Critical Transitions in Nature and Society*. Vol. 16. Princeton University Press, Princeton, New Jersey.
- [122] Marten Scheffer and Egbert H. Van Nes. 2018. Seeing a global web of connected systems. *Science* 362, 6421 (2018), 1357–1357.
- [123] Ignacio Segovia Dominguez, Huikyo Lee, Yuzhou Chen, Michael Garay, Krzysztof M. Gorski, and Yulia R. Gel. 2021. Does air quality really impact COVID-19 clinical severity?: Coupling NASA satellite datasets with geometric deep learning. In *27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. ACM, 3540–3548.
- [124] Jose Segovia-Juarez, Jesús M. Castagnetto, and Gustavo F. Gonzales. 2020. High altitude reduces infection rate of COVID-19 but not case-fatality rate. *Respir. Physiol. Neurobiol.* 281 (2020), 103494.
- [125] Manuela Signorini, Rudi Cassini, Michele Drigo, Antonio Frangipane di Regalbano, Mario Pietrobelli, Fabrizio Montarsi, and Anna-Sofie Stensgaard. 2014. Ecological niche model of *Phlebotomus perniciosus*, the main vector of canine leishmaniasis in north-eastern Italy. *Geospat. Health* 9, 1 (Nov. 2014), 193–201. DOI: <https://doi.org/10.4081/gh.2014.16>
- [126] Ashutosh Simha, R. Venkatesha Prasad, and Sujay Narayana. 2020. A simple Stochastic SIR model for COVID 19 Infection Dynamics for Karnataka: Learning from Europe. Retrieved from [arXiv:q-bio.PE/2003.11920](https://arxiv.org/abs/2003.11920).
- [127] R. K. Sanayaima Singh, Md. Zubair Malik, and R. K. Brojen Singh. 2021. Diversity of SARS-CoV-2 isolates driven by pressure and health index. *Epidemiol. Infect.* 149 (2021), e38. DOI: <https://doi.org/10.1017/S0950268821000248>
- [128] Albertus J. Smit, Jennifer M. Fitchett, Francois A. Engelbrecht, Robert J. Scholes, Godfrey Dzhivhuho, and Neville A. Sweijd. 2020. Winter is coming: A southern hemisphere perspective of the environmental drivers of SARS-CoV-2 and the potential seasonality of COVID-19. *Int. J. Environ. Res. Public Health* 17, 16 (2020), 5634.
- [129] Ilaria Spassiani, Giovanni Sebastiani, and Giorgio Palù. 2021. Spatiotemporal analysis of COVID-19 incidence data. *Viruses* 13, 3 (2021), 463. DOI: <https://doi.org/10.3390/v13030463>

- [130] Kaoru Tachiiri, Brian Klinkenberg, Sunny Mak, and Jamil Kazmi. 2006. Predicting outbreaks: A spatial risk assessment of West Nile virus in British Columbia. *Int. J. Health Geog.* 5, 1 (2006), 21.
- [131] K. M. Tolle, D. S. W. Tansley, and A. J. G. Hey. 2011. The fourth paradigm: Data-Intensive scientific discovery [Point of View]. *Proc. IEEE* 99, 8 (2011), 1334–1337. DOI: <https://doi.org/10.1109/JPROC.2011.2155130>
- [132] Ramadhan Tosepu, Joko Gunawan, Devi Savitri Effendy, Hariati Lestari, Hartati Bahar, Pitrah Asfian et al. 2020. Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia. *Sci. Total Environ.* 725 (2020), 138436.
- [133] Marco Travaglio, Yizhou Yu, Rebeka Popovic, Liza Selley, Nuno Santos Leal, and Luis Miguel Martins. 2021. Links between air pollution and COVID-19 in England. *Environ. Pollut.* 268 (2021), 115859.
- [134] Tuscany Regional Health Agency. 2020. Recommendations for health operators. Retrieved from <https://www.ars.toscana.it/2-articoli/4276-nuovo-coronavirus-covid-19-informazioni-buone-pratiche-raccomandazioni-per-operatori-sanitari.html>.
- [135] Ploutarchos Tzampoglou and Dimitrios Loukidis. 2020. Investigation of the importance of climatic factors in COVID-19 worldwide intensity. *Int. J. Environ. Res. Public Health* 17, 21 (2020), 7730.
- [136] Schulzweida Uwe. 2021. MPI for Meteorology. Retrieved from <https://code.mpimet.mpg.de/projects/cdo/embedded/index.html>.
- [137] Tomáš Václavík and Ross K. Meentemeyer. 2009. Invasive species distribution modeling (iSDM): Are absence data and dispersal constraints needed to predict actual distributions? *Ecol. Modell.* 220, 23 (2009), 3248–3258.
- [138] Tomáš Václavík and Ross K. Meentemeyer. 2012. Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion. *Divers. Distrib.* 18, 1 (2012), 73–83.
- [139] George Valiakos, Konstantinos Papaspyropoulos, Alexios Giannakopoulos, Periklis Birtsas, Sotirios Tsiodras, Michael R. Hutchings, Vassiliki Spyrou, Danaï Pervanidou, Labrini V. Athanasiou, Nikolaos Papadopoulos et al. 2014. Use of wild bird surveillance, human case data, and GIS spatial analysis for predicting spatial distributions of West Nile virus in Greece. *PLoS One* 9, 5 (2014), e96935.
- [140] John Wahlgren. 2011. Influenza A viruses: An ecology review. *Infect. Ecol. Epidemiol.* 1, 1 (2011), 6004.
- [141] Nephi A. Walton, Mollie R. Poynton, Per H. Gesteland, Chris Maloney, Catherine Staes, and Julio C. Facelli. 2010. Predicting the start week of respiratory syncytial virus outbreaks using real time weather variables. *BMC Med. Inform. Decis. Mak.* 10, 1 (2010), 68.
- [142] Jingyuan Wang, Ke Tang, Kai Feng, Xin Li, Weifeng Lv, Kun Chen, and Fei Wang. 2020. High Temperature and High Humidity Reduce the Transmission of COVID-19. Retrieved from [arXiv:q-bio.PE/2003.05003](https://arxiv.org/abs/2003.05003).
- [143] WCRP WCRP. 2011. Coupled model intercomparison project—phase 5—CMIP5. *CLIVAR Exch. Spec. Issue (World Clim. Res. Prog., 2011)* 16 (2011), 32.
- [144] Katherine Wolstencroft, Robert Haines, Donal Fellows, Alan Williams, David Withers, Stuart Owen, Stian Soiland-Reyes, Ian Dunlop, Aleksandra Nenadic, Paul Fisher et al. 2013. The Taverna workflow suite: Designing and executing workflows of Web Services on the desktop, web or in the cloud. *Nucleic Acids Res.* 41, W1 (2013), W557–W561.
- [145] World Economic Forum. 2021. COVID-19 proves mobility systems require new solutions to meet changing needs. Retrieved from <https://www.weforum.org/agenda/2020/04/covid-19-proves-mobility-systems-require-new-solutions-to-meet-changing-needs/>.
- [146] Xiao Wu, Rachel C. Nethery, Benjamin M. Sabath, Danielle Braun, and Francesca Dominici. 2020. Exposure to air pollution and COVID-19 mortality in the United States: A nationwide cross-sectional study. DOI: <https://doi.org/10.1101/2020.04.05.20054502>
- [147] Xiaoling Yuan, Jie Xu, Sabiha Hussain, He Wang, Nan Gao, and Lanjing Zhang. 2020. Trends and prediction in daily new cases and deaths of COVID-19 in the United States: An internet search-interest based model. *Explorat. Res. Hypoth. Med.* 5, 2 (2020), 1.
- [148] A. Elizabeth Zaniewski, Anthony Lehmann, and Jacob McCovert. 2002. Predicting species spatial distributions using presence-only data: A case study of native New Zealand ferns. *Ecol. Modell.* 157, 2–3 (2002), 261–280.
- [149] Jianqiang Zhang, Kyoung-Jin Yoon, and Jeffrey J. Zimmerman. 2019. *Overview of Viruses*. John Wiley & Sons, Ltd, New York, New York, 425–437. DOI: <https://doi.org/10.1002/9781119350927.ch23>

Received April 2021; revised October 2021; accepted October 2021