



OPEN

## Climate and COVID-19 transmission: a cross-sectional study in Africa

Ousmane Koanda<sup>1</sup>, Roland Yonaba<sup>1</sup> , Fowé Tazen<sup>1</sup>, Héla Karoui<sup>1</sup>, Mohamed Lamine Sidibé<sup>1</sup>, Babacar Lèye<sup>1</sup>, Mamadou Diop<sup>2</sup>, Harinaivo Anderson Andrianisa<sup>1</sup> & Harouna Karambiri<sup>1</sup>

The role of climate in the Coronavirus disease 2019 (COVID-19) transmission appears to be controversial, as reported in earlier studies. In Africa, the subject is poorly documented. In this study, over the period from January 1st, 2020 to September 31, 2022, the daily variations in cumulative confirmed cases of COVID-19 for each African country (54 countries) are modelled through time-series-based approaches and using meteorological factors as covariates. It is suggested from the findings that climate plays a role in COVID-19 transmission since at least one meteorological factor is found to be significant in 32 countries. In decreasing order, the most often occurring meteorological factors are dewpoint temperature, relative and absolute humidity, average temperature and solar radiation. Most of these factors show a lagged effect with confirmed cases (between 0 and 28 days). Also, some meteorological factors exhibit contrasting effects on COVID-19 transmission, resulting in both positive and negative association with cumulative cases, therefore highlighting the complex nature of the interplay between climate and COVID-19 transmission.

In December 2019, a transmissible disease emerged in the city of Wuhan (in Hubei province, China). The disease was found to be caused by a new coronavirus, the *Sars-CoV-2*<sup>1</sup>. Initially called *2019-nCoV* acute respiratory disease, later renamed *COVID-19* disease, it was declared a pandemic by the World Health Organization (WHO) on March 11, 2020, and further acknowledged as a Public Health Emergency of International Concern (PHEIC)<sup>2</sup>. The disease quickly spread throughout the world, starting in January 2020. Despite the wide array of restrictive measures largely promoted by governments worldwide<sup>3,4</sup>, the outbreak of the pandemic has been largely unprecedented: a year later, on December 31, 2020, a total of 83,638,090 cumulative confirmed cases and 1,880,710 deaths were reported globally, hence a case-fatality rate of 2.25%<sup>5</sup>.

Even though at the global scale, the outbreak of the COVID-19 disease has been exponential, the African continent emerges as the least affected in the World, according to the numbers of confirmed cases and deaths reported. The first confirmed case in Africa was reported in Egypt on February 14, 2020. On December 31, 2020, a total of 2,760,450 cumulative confirmed cases and 65,458 deaths (respectively 3.30% and 3.48% of the global count) were reported. As of September 30, 2022, the total cumulative confirmed cases in Africa rose to 12,355,709 (2.0% of the global total), while the number of deaths reached 256,764 (3.9% of the global total)<sup>5</sup>. These estimates seem low compared to those from America, Europe and Asia, which reported respectively 44.6%, 28.6% and 24.8% of the global count of confirmed cases<sup>5,6</sup>. Yet, the case-fatality rate of the disease appears to be almost twice as high in Africa (2.08% as of September 30, 2022) as compared to the global level (1.08%), indicating the higher lethality of the disease in Africa<sup>7</sup>.

A typical trait regarding the spatial patterns of the spread of the pandemic is that, at the scale of the African continent, large discrepancies are observed among countries. As of September 30, 2022, for example, South Africa alone reported 32.5% and 39.8% of the total confirmed cases and deaths (respectively) in Africa. Meanwhile, the top five hard-hit countries (*in decreasing order of severity*: South Africa, Morocco, Tunisia, Egypt and Libya) reported 60.3% and 69.7% of the total confirmed cases and deaths (respectively)<sup>5</sup>. These figures are thought to be due to specific traits of the African context, including the low rate of urbanization, the limited transport network hindering the mobility of people, and the young age of the population<sup>8,9</sup>. Also, the lower case fatality rate

<sup>1</sup>Laboratoire Eaux, Hydro-Systèmes et Agriculture (LEHSA), Institut International d'Ingénierie de l'Eau et de l'Environnement (2iE), Ouagadougou, Burkina Faso. <sup>2</sup>Laboratoire Eco-Matériaux et Habitat Durable (LEMHaD), Institut International d'Ingénierie de l'Eau et de l'Environnement (2iE), Ouagadougou, Burkina Faso. ✉email: roland.yonaba@gmail.com

in Africa of non-communicable diseases such as cancer, cardiovascular accidents and diabetes (already known as comorbidities in this pandemic) has been often put forward<sup>7,10</sup>.

Among the factors affecting the transmissibility of COVID-19, the role of climate conditions has been early investigated and is still a trending and active topic<sup>11</sup>. Meteorological conditions can influence how the virus survives and persists within a given environment, which in turn might affect the likelihood of transmission of the disease<sup>12</sup>. Overall, most of the available studies tend to report lower transmission rates at warmer temperatures but higher transmission in colder and dry environments<sup>9,13</sup>, along with exacerbating effects of humidity, higher air concentration in fine particles<sup>14</sup>, lower wind speeds<sup>15</sup> or lower solar radiation<sup>9</sup>. For example, in South Asia, Hossain et al.<sup>12</sup> identified a positive association between COVID-19 confirmed cases and minimum, maximum and average temperatures, rainfall, relative air humidity, wind speed and surface pressure. In the US, Bashir et al.<sup>16</sup> reported a positive association between COVID-19 cases and temperatures, but a negative association with rainfall, relative air humidity, wind speed and surface pressure. Zhuang et al.<sup>17</sup> reported similar findings based on evidence from 216 cities in China. Temperature and relative humidity are, on average, the factors the most frequently associated with COVID-19 cases<sup>18</sup>. Wang et al.<sup>19</sup> reported for instance that a 1 °C rise in average temperature might be responsible for a 3.1% decrease in the daily new cases of COVID-19 infections and a 1.2% decrease in related deaths. Other meteorological parameters, such as wind speed and air pollutants are also reported as likely influential on COVID-19 transmission rates<sup>12,20</sup>. Solar radiation and geographical latitude (representing the distance to the equator) have also been identified as significant in explaining COVID-19 deaths<sup>9,13</sup>. Moreover, the relationship between climate and COVID-19 transmission is thought to be complex because weather patterns in a year affect human behaviour, therefore affecting the observation of lifting of containment measures. For example, Ganslmeier et al.<sup>21</sup> reported that weather substantially affects the likelihood of social gatherings, therefore increasing the spread of the disease. Also, previous research indicates that including human behaviour and climate is critical in modelling the spread of a disease outbreak<sup>22</sup>.

Previous modelling studies attempted to analyse the spread of COVID-19 in different contexts, mainly to forecast the pandemic prevalence or assess the potential effect of meteorological conditions (as covariates) on COVID-19<sup>23</sup>. The modelling methods often used range from generalized additive models<sup>24</sup>, generalized linear models<sup>24</sup>, compartmental epidemiologic approaches based on *Susceptible, Infectious or Recovered* (SIR) models<sup>25–27</sup>, and artificial neural networks<sup>28,29</sup>. Among these methods, time series-based modelling methods such as *Autoregressive Integrated Moving Average* (ARIMA) have been largely used for COVID-19 modelling and to explore the potential effect of external regressors such as meteorological variables on COVID-19 transmission<sup>30–32</sup>.

However, despite the huge body of available literature, still, few studies addressed the case of Africa as a whole, especially regarding the specific role of meteorological conditions on COVID-19 transmission. There is still a need to assess which and to what extent climate factors influence the transmission of COVID-19<sup>33–36</sup>. The question is of importance to African countries, especially in building effective, well-informed and adapted coping strategies for the ongoing pandemic<sup>37</sup>.

In this study, we aimed to provide a continental scale assessment of the potential isolated effect of meteorological conditions on COVID-19 prevalence in the specific context of Africa. Our objectives are twofold: (i) identify the climatic factors which significantly affect COVID-19 transmission in African countries; (ii) assess the isolated level of association of these climatic factors to COVID-19 and their relative importance across African countries. In this scope, the study uses publicly available data on COVID-19 prevalence accumulated over 33 months, from January 1, 2020, to September 30, 2022, which is deemed appropriate to investigate the potential long-term effects of meteorological conditions on COVID-19 transmission.

## Results

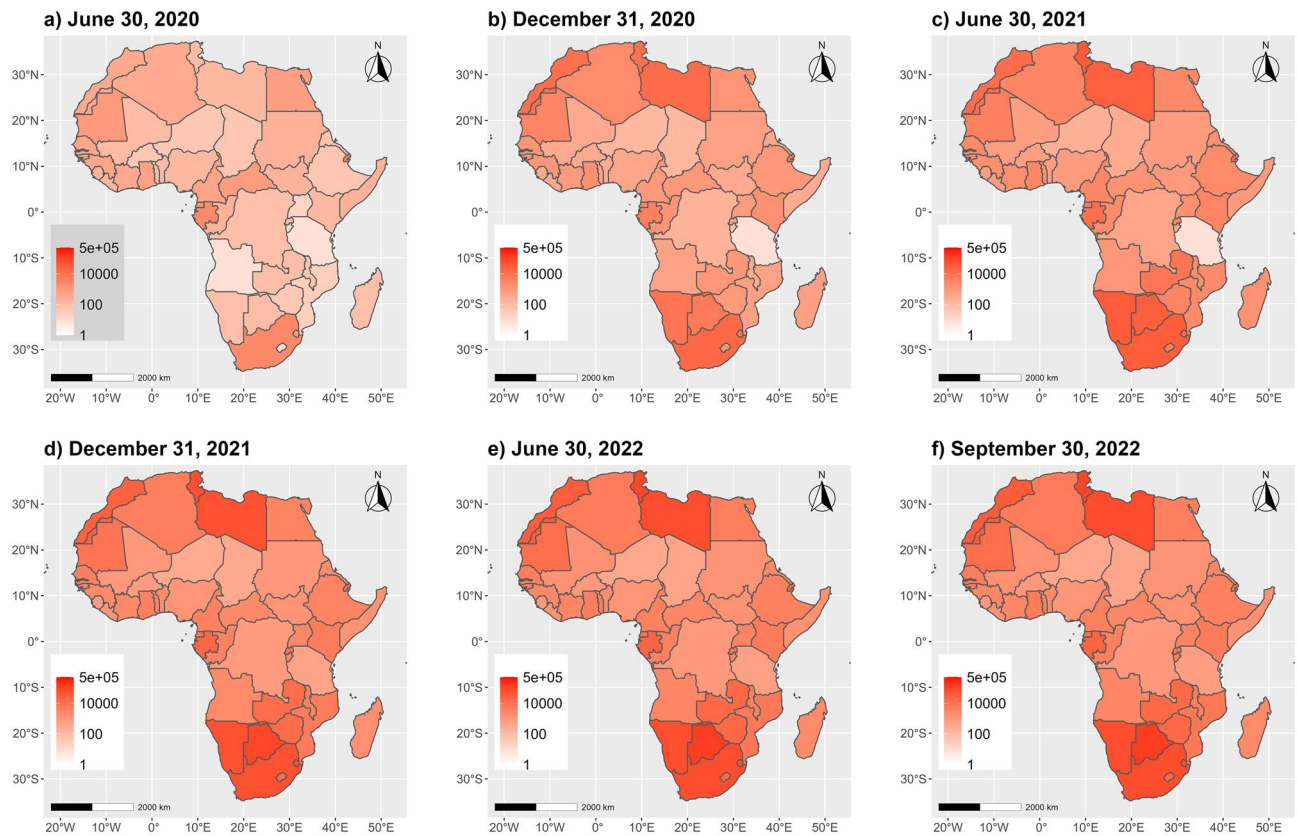
### Figures of the COVID-19 pandemic in Africa

#### *Cumulative cases and waves*

According to official reports<sup>5,6</sup>, the first confirmed case of COVID-19 was reported in Egypt (EGY) on February 14, 2020; it was shortly followed by countries mostly from Northern Africa (NAf) region, especially Algeria (DZA), Tunisia (TUN) and Morocco (MAR)<sup>9</sup>. Over time, East Africa (EAf) and Southern Africa (SAf) emerged as the regions where the incidence of the pandemic is the largest. These regions witnessed three waves of increasing magnitudes: (i) in January 2021, 500–750 daily new confirmed cases per million; (ii) in May–July 2021, 1500–1700 daily new confirmed cases per million; and (iii) in January 2022, 2000–3000 new daily confirmed cases per million. West Africa (WAf) and Northern Africa (NAf) are the second largest regions affected, with waves occurring at the same periods, although roughly similar in magnitude, with 700–900 new daily confirmed cases per million. Central Africa (CAf) is the least affected region in the continent (further details are provided as Supplementary Material, see Figs. S1, S2 and Table S1).

#### *Spatial patterns of spread of the pandemic in Africa*

The spatial patterns of evolution of the COVID-19 are presented in choropleths maps in Fig. 1. It appears that at the onset of the pandemic (January 1st to December 31, 2020), the countries located at the northern and southern poles of the continent were the most affected. This is likely attributable to the fact that such countries feature the highest rates of international trade and tourism<sup>9</sup> and are gateways to the African continent<sup>9,38,39</sup>. However, over this period and the following year, there is still no clear emerging spatial clustering or organization of the COVID-19 pandemic at the global continental level. Towards September 30, 2022, countries around the equator emerge as the least affected, while those at the northernmost and the southernmost poles of the continent appear as the highly affected ones.



**Figure 1.** Spatial and temporal evolution of COVID-19 cumulative cases (per million) in Africa. The maps were produced using the *ggmap*<sup>40</sup> and *ggplot2*<sup>41</sup> visualisation libraries under the R programming language<sup>42</sup>.

The assessment of the global spatial auto-correlation in COVID-19 incidence is presented in Table 1, which shows that until June 30, 2022, there is no significant global spatial autocorrelation at the continental level (Moran's  $I$  p-values  $> \alpha = 5\%$ ). From December 31, 2021 until September 30, 2022, a significant global spatial auto-correlation is emerging (Moran's  $I$  p-values  $< \alpha = 5\%$ ), likely because countries in the low-low cluster (countries with lower incidence surrounded by similar neighbouring countries) appear spatially located near the equator (Western, Central and East Africa regions), while highly affected countries remain altogether located at the continent poles.

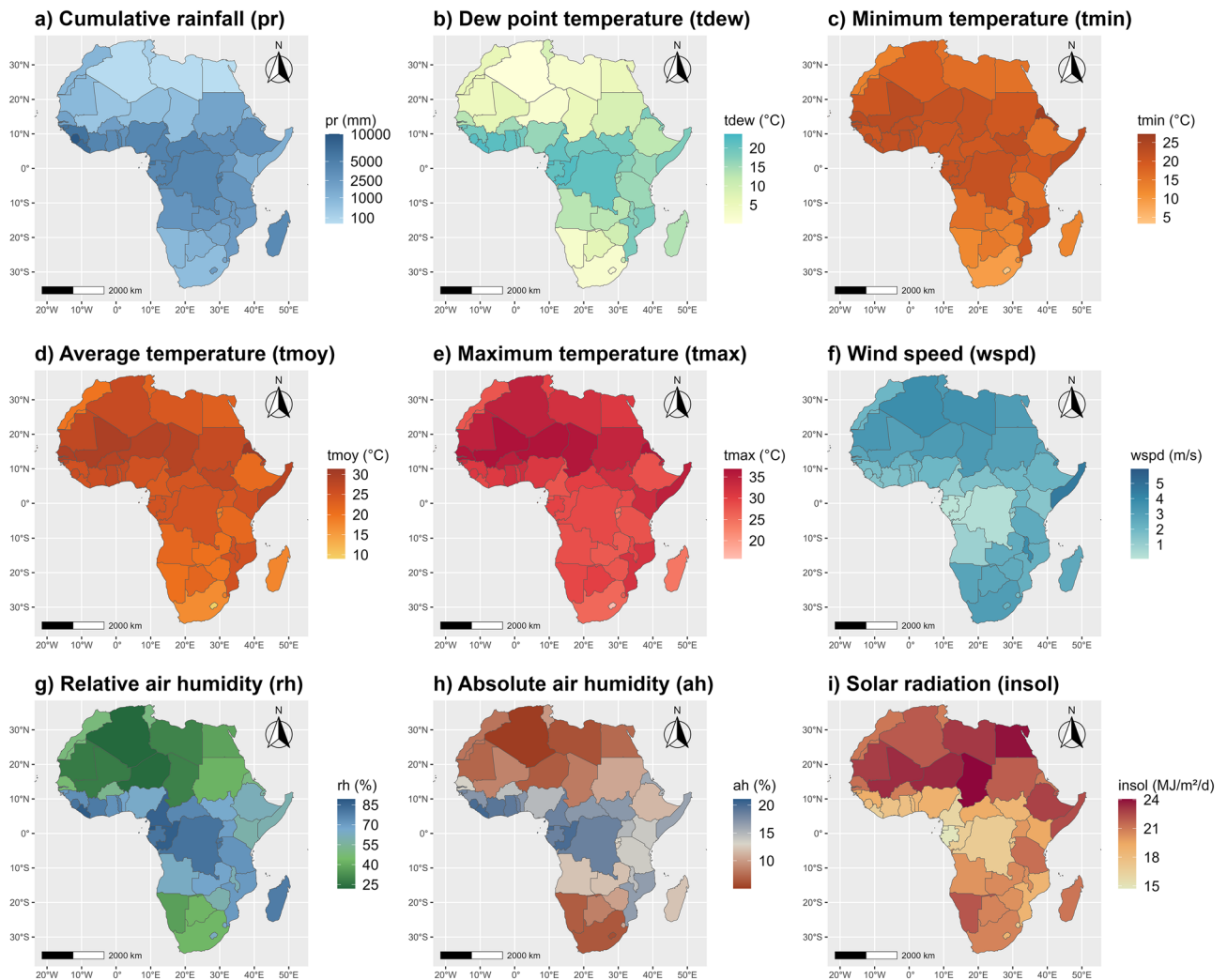
### Analysis of meteorological conditions

#### Summary description and spatial patterns of meteorological conditions

The spatial patterns of meteorological conditions in Africa are presented in Fig. 2. The driest median cumulative rainfall ( $pr$ ) is observed for the NAF region (202.3 mm), whereas it increases to 2580.7–4564.7 mm in other regions. NAF region also shows the lowest median dew point temperature ( $tdew = 5.1$  °C), which increases to 12.1 °C (SAf region) and 21.0 °C (CAf region). The highest minimum ( $tmin$ ), average ( $t moy$ ) and maximum ( $tmax$ ) median temperatures are observed in the WAF region (22.2 °C, 26.6 °C and 32.8 °C respectively), whereas the lowest values are reported in the SAf region (14.2 °C, 20.6 °C and 27.2 °C respectively). The NAF region also shows the lowest relative ( $rh$ ) and absolute ( $ah$ ) median humidity (38.7% and 7.3% respectively), but

Dates	Observed $I$	Expected $I_0$	Standard deviation	p-value
6/30/2020	-0.0123	-0.0189	0.0224	0.7684
12/31/2020	0.0235	-0.0189	0.0233	0.0693
6/30/2021	-0.0062	-0.0189	0.0165	0.4437
12/31/2021	0.0189	-0.0189	0.0186	0.0425*
6/30/2022	0.0174	-0.0189	0.0176	0.0394*
09/30/2022	0.0177	-0.0189	0.0177	0.0393*

**Table 1.** Global Moran's  $I$  for spatial auto-correlation in COVID-19 incidence.  $I$  is the computed Global Moran.  $I_0$  is the expected value of  $I$  under the null hypothesis; P-values appended with "\*" are significant at  $\alpha = 5\%$  level.



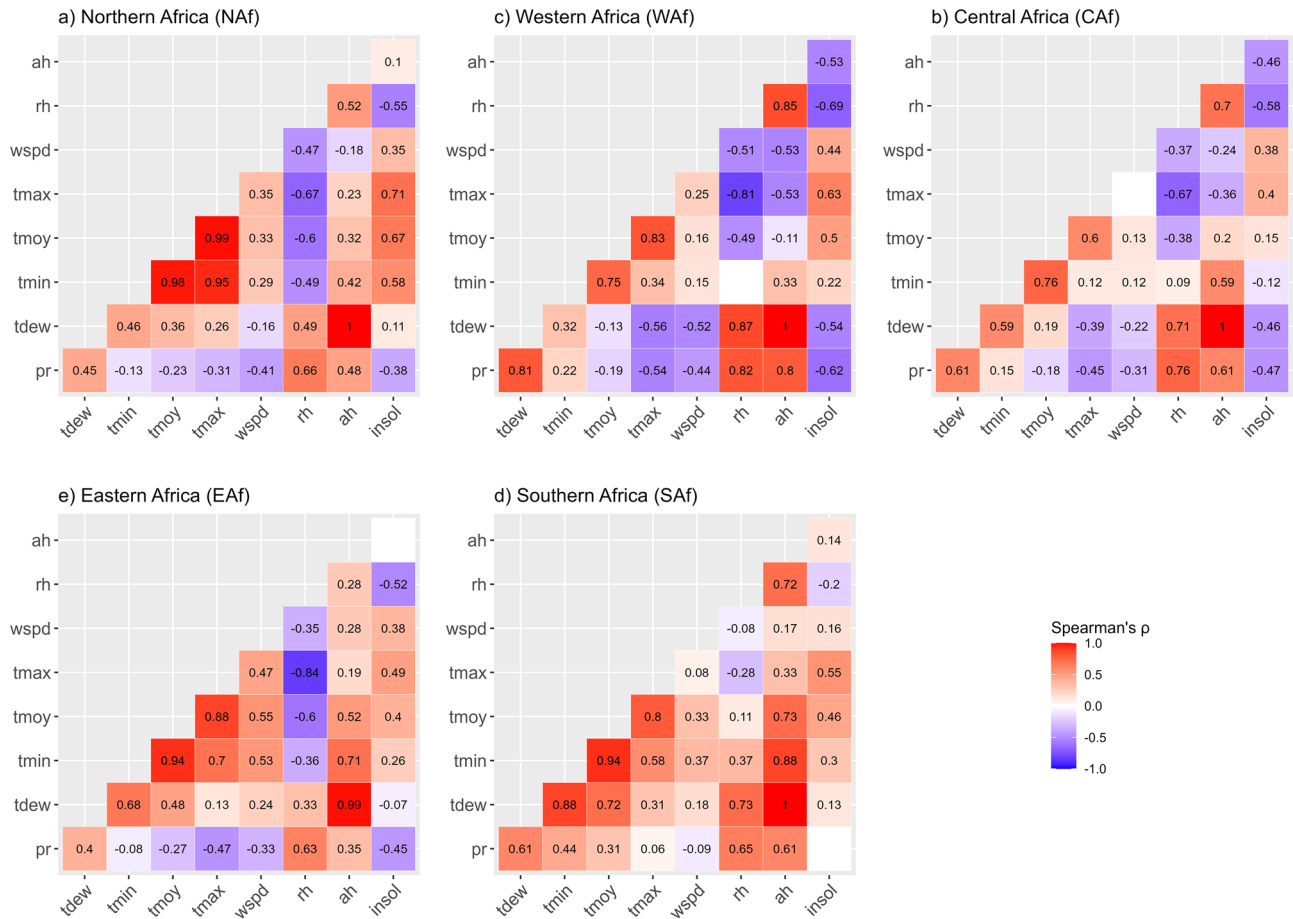
**Figure 2.** Spatial repartition of meteorological variables in Africa. The maps were produced using the *ggmap*<sup>40</sup> and *ggplot2*<sup>41</sup> visualisation libraries under the R programming language<sup>42</sup>.

the highest median wind speed (3.2 m/s). In contrast, the median relative (*rh*) and absolute (*ah*) humidity in the other regions range between 63.2% (Eaf region)–81.3% (Caf region) and 11.6% (SAf region)–18.4% (CAf region). Wind speed (*wspd*) is also generally lower in Waf (median of 2.2 m/s), Eaf (median of 2.1 m/s), SAf (median of 2.7 m/s) and CAf (median of 1.0 m/s). Finally, solar radiation (*insol*) shows the highest values in the NAf region (median of 22.0 MJ/m<sup>2</sup>/day) and the lowest values in the CAf region (median of 16.9 MJ/m<sup>2</sup>/day). The correspondence between countries and regions is given as Supplementary Material in Table S2, and further details on the statistical description of each climate variable according to each region in the continent are given as Supplementary Material in Table S3.

#### Inter-correlations between climate variables

The correlation between climate variables is presented in Fig. 3, region-wise. Dew point temperature (*tdew*), *tmax*, *tmin* and *tmoy* variables are positively associated in NAf, Eaf and SAf regions, whereas *tdew* appears to be negatively associated with *tmoy* and *tmax* in the Waf region and only with *tmax* in the CAf region. Also, *wspd* is always negatively correlated with *tdew* and *pr* in NAf, Waf and CAf regions, negatively correlated with *pr* in Eaf and SAf, but positively associated with *tdew* in Eaf and SAf regions. *ah* appears to be positively associated with all climate variables in Eaf and SAf regions, but shows negative association with *wspd* (in NAf region), *wspd*, *tmax* and *tmoy* (in Waf region), *wspd* and *tmax* (in CAf region). In all regions, however, *rh* and *ah* variables are positively associated with *tdew* and *pr*.

These inter-correlations shed light that climate variables show different forms of associations depending on the region. However, these correlations are influenced by seasonal variations (forced by the Earth's rotation around the Sun), and therefore, do not consider the unique properties of intra-annual variability of the considered climate variable over the period 2020–2022. In this regard, we complement the previous analysis with the assessment of correlations on anomalies of climate variables, estimated from climate normals (over the period 2000–2019, i.e. 20 years) and presented in Fig. 4.



**Figure 3.** Correlations between climate variables according to regions. Correlations are evaluated using the non-parametric Spearman's  $\rho$  correlation coefficient. Blank cells correspond to non-significant values (at  $\alpha = 5\%$  significance level).

The analysis of correlations between anomalies reveals similar patterns across all regions. The highest positively correlated sets are *tmoy*, *tmin* and *tmax* (especially in Naf, Waf and SAf regions), and also *rh*, *ah* and *tdew* (at similar levels in all regions). A moderately positive association is observed between *insol*, *tmoy* and *tmax* in all regions except NAF. A moderately negative association is reported between *tmax*-*tmoy* and *rh*-*ah* pairs in almost all regions, and also between *rh*-*ah* pair and *insol*. These observations reveal consistent patterns of relationship between the climate variables across the different regions in Africa, despite the seasonal patterns which can be specific to each region<sup>43</sup>.

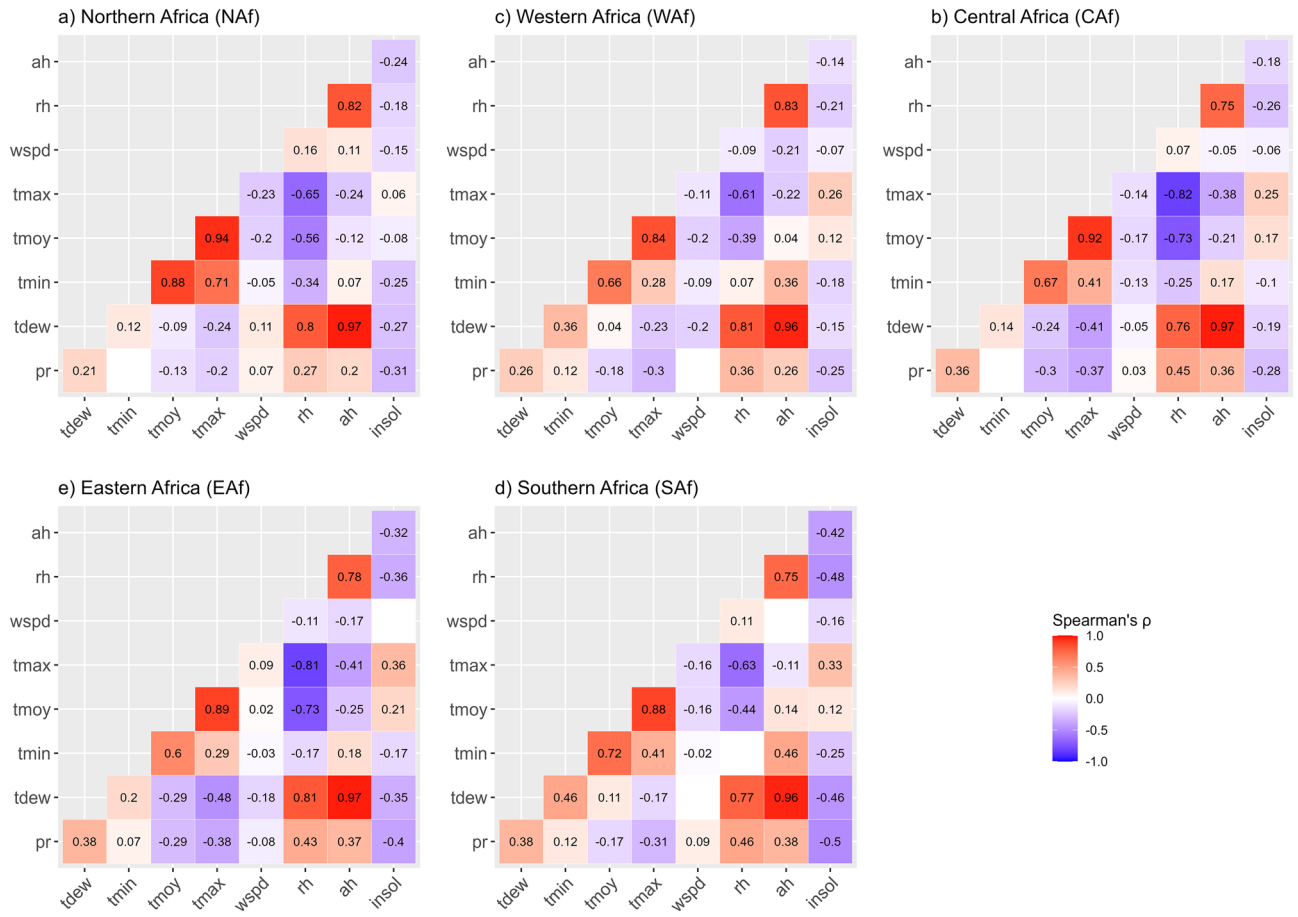
### Effect of meteorological factors on COVID-19 cumulative cases

#### Significant meteorological factors affecting COVID-19 transmission

The fitting of a *RegARIMA* model with meteorological factors as covariates on the COVID-19 time series of cumulative cases results in 156 candidate models in which the climate variable included is significant (Wald *t*-test on coefficients—*p*-value < 5%) and the residuals are similar to white noise (Ljung-Box Q test, *p*-value < 5%). However, within this total, some meteorological factors often show multiple significant lags for the same country. Most of the time, these multiple significant lags are not consecutive, but separated by lags without significance. This could be further explained by the recurrence time scale, as shown by the plots of the Autocorrelation Function (ACF) provided as a Supplementary Material (pages 10–63). There is also no apparent direct relationship between the number of significant lags and the decorrelation time scale.

To further filter the potential 'true' models, the 'duplicate' models are filtered to retain the model showing the highest association (i.e. the absolute largest Spearman's  $\rho$  correlation coefficient) with COVID-19. Therefore, a definitive total of 59 models are finally considered for further analysis. The complete list of these models is presented as a Supplementary Material (Table S4).

Overall, it appears that in 32 African countries (out of 54), at least one of the 9 meteorological factors considered in this study appears to be significant in COVID-19 cumulative cases. Interestingly, some countries were found to have the largest numbers of climate variables significantly related to COVID-19 cumulative cases. These countries include Côte d'Ivoire (CIV: *ah* *tdew*, *tmoy*, *wspd*) and Seychelles (SYC: *tdew*, *tmax*, *tmin*, *tmoy*), closely followed by Gabon (GAB), Guinea-Bissau (GNB), Liberia (LBR), South Sudan (SSD) and Togo (TGO).



**Figure 4.** Correlations between climate anomalies (2020–2022) according to regions. Anomalies are estimated from climate normals over the period 2000–2019 (20 years). Correlations are evaluated using the non-parametric Spearman's  $\rho$  correlation coefficient. Blank cells correspond to non-significant values (at  $\alpha = 5\%$  significance level).

The meteorological factors most often occurring are *tdew* (in 10 countries), followed by *rh* and *ah* (both in 9 countries). They are followed by *insol* and *tmoy* (both in 8 countries). Variables such as *pr*, *tmin* and *tmax* are both reported as significant in 4 countries and *wspd* in 3 countries.

Figure 5 provides a visual comparison of the distribution of the significant lags between climate variables and the COVID-19 cumulative cases across the 59 RegARIMA models retained while separating positive and negative association.

Some variables show a positive delayed association with COVID-19 (Fig. 5a), meaning that an increase in such variables is associated with an increase in COVID-19 cases. Such variables include *rh* and *insol* (median lags of 18.0 and 13.0 days respectively), followed by *pr* (median lag of 10.0 days), *wspd* (median lag of 7.0 days) and finally *ah* (median lag of 4.0 days).

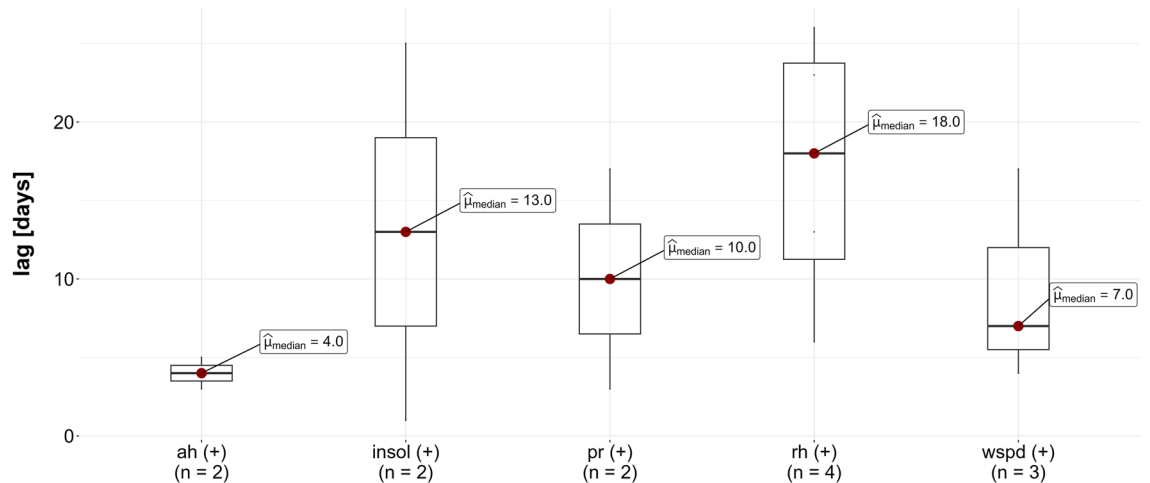
The majority of the climate variables considered in this study show a negative delayed association with COVID-19 (Fig. 5b), meaning that an increase in these variables is likely related to a decrease in COVID-19 cases. *insol* and *pr* have the highest negative association lags (medians of 25.5 and 24.5 days respectively), followed by *tmin* (median lags of 15.5 days), then *tmoy*, *ah*, *rh* and *tdew* at a similar level (median lags of 9.0, 8.0, 8.0 and 6.5 days respectively). *tmax* shows the shorter lag in terms of negative association (median of 3.5 days).

It should be noted that some variables such as *ah*, *insol*, *pr* and *rh* show both positive and negative associations, but in different countries. However, the negative associations are less dominant, in comparison with the positive ones.

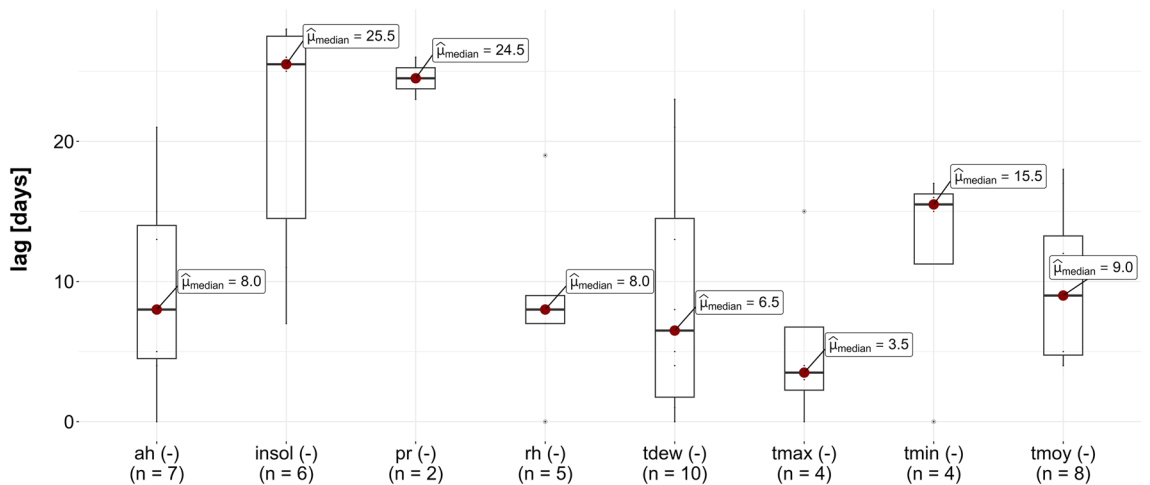
*Spatial distribution of association of climate variables to COVID-19*

The spatial distribution of countries where a positive association between a meteorological variable and COVID-19 is detected is shown in Fig. 6. The analysis shows that rainfall is positively associated with COVID-19 cases in Congo Kinshasa (COD) and Burundi (BDI). Relative humidity (*rh*) is reported significant for Gabon (GAB) in CAF, Guinea-Bissau (GNB) in Waf, but also Eswatini (SWZ) and Malawi (MWI) in Saf. Solar radiation (*insol*) is significant for Sierra Leone (SLE) and Liberia (LBR), while absolute humidity (*ah*) appears significant for Senegal (SEN) and South Africa (ZAF). Wind speed (*wspd*) is reported for Côte d'Ivoire (CIV), Kenya (KEN) and Egypt (EGY), which are mostly coastal countries.

## a) Lags for positive association



## b) Lags for negative association



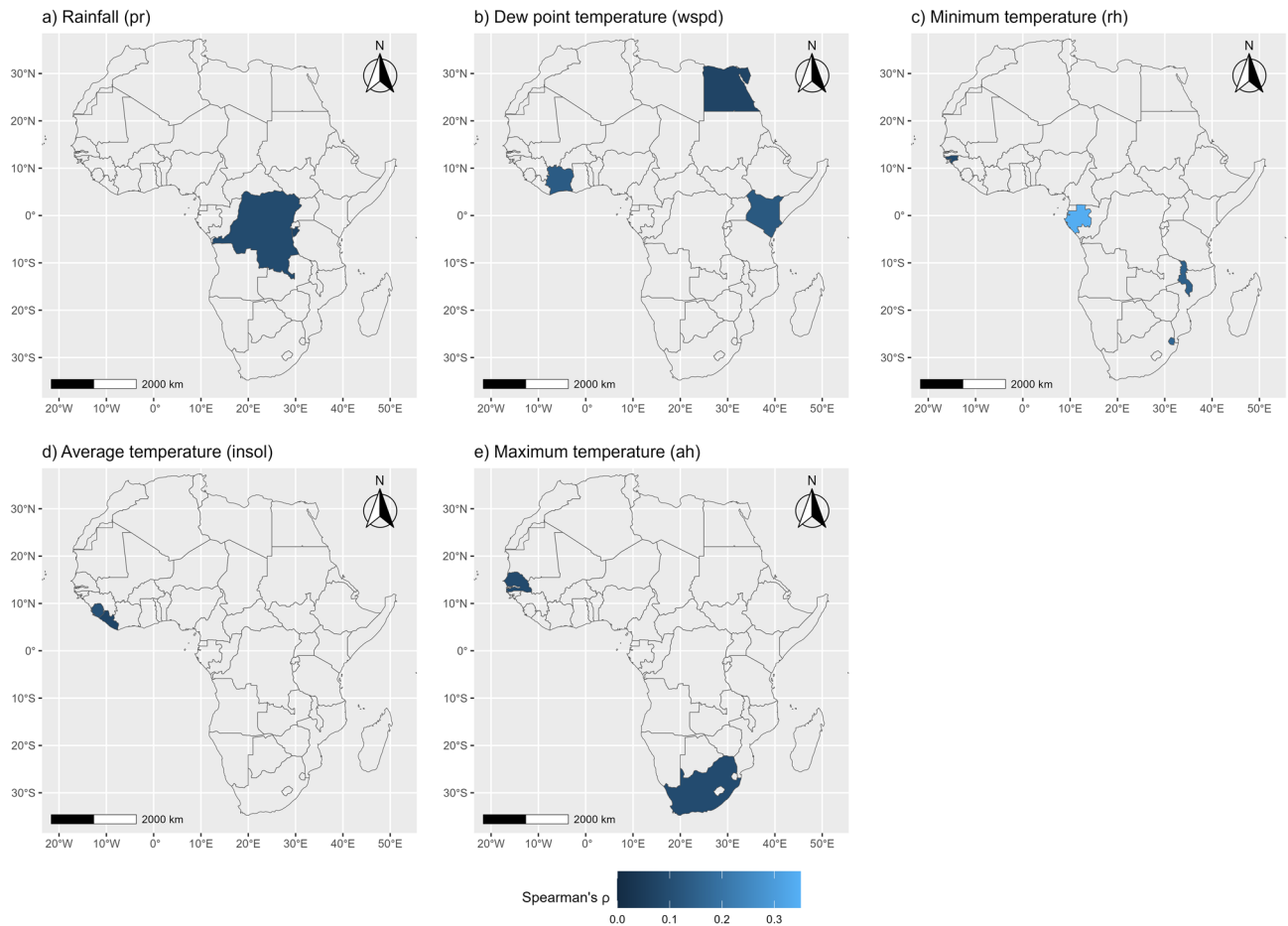
**Figure 5.** Distribution of lags according to positive and negative associations between meteorological variables and COVID-19 cumulative cases. **(a)** Distribution of lags for positive association. **(b)** Distribution of lags for negative association. The red dots show the median value over the distribution box plots, indicated in days. The number of values in each box plot is shown in parenthesis under the axis label.

The spatial distributions of countries where a negative association between meteorological variables and COVID-19 is observed are presented in Fig. 7. In this case, dew point temperature (*tdew*) is reported significant for countries mostly in Waf, Eaf and SAf (to a lesser extent). Average temperature (*tmoy*) is also negatively associated with COVID-19 in mostly Eaf countries and also the Waf region. Absolute (*ah*) and relative (*rh*) humidity are found significant mostly in Waf and Eaf, while maximum (*tmax*) and minimum (*tmin*) temperature are reported significant mostly in the Eaf region.

## Discussion

In this study, the potential effect of meteorological factors on COVID-19 transmission is evaluated in the context of African countries. The global pandemic figures show that Africa as a whole remains largely less affected by COVID-19 in comparison to Europe, America and Asia. Previous studies have suggested that this apparent resilience from Africa can be explained by the youthfulness of its population (median age of 18 years<sup>8</sup>). Also, some other factors include the low rate of urbanization, the reduced mobility rate due to the limited means of swift transportation and the low volume of international transit with other continents<sup>9,26</sup>. More recent studies even suggested the possibility of immune formation endemic to developing countries<sup>44</sup>, where interestingly the previous occurrence of diseases such as malaria, and tuberculosis have been associated with a lower prevalence of COVID-19<sup>9,45,46</sup>.

The role of meteorological conditions of COVID-19 has been investigated in previous research<sup>9,13,18,19,24,47,48</sup>. In this study, the findings regarding the specific role of each meteorological factor on COVID-19 transmission are not necessarily consistent across all the countries considered, in line with previous research<sup>12</sup>. Yet, some key signatures are emerging. First, a significant association with at least one climate variable has been detected

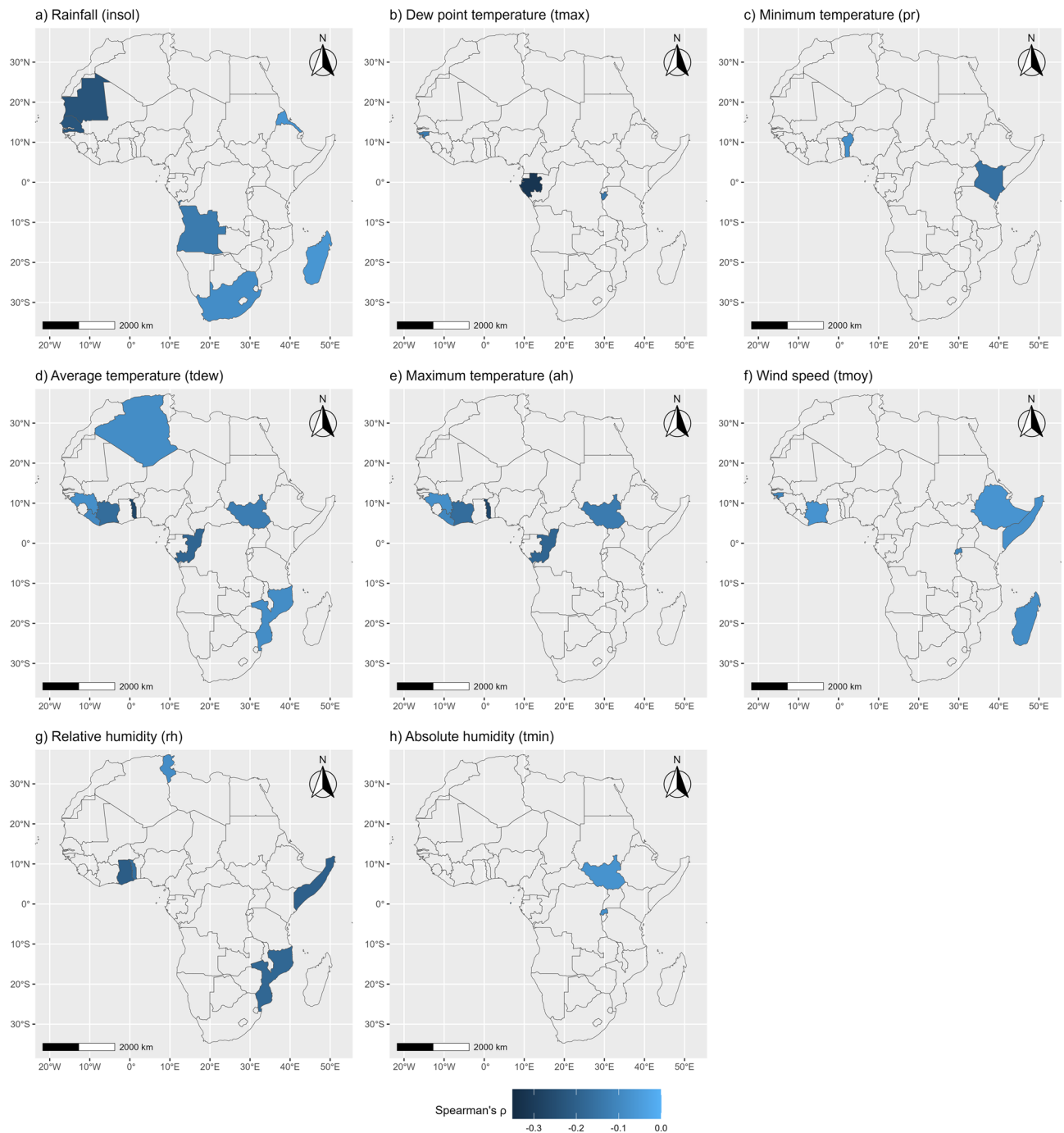


**Figure 6.** Countries showing a positive association between meteorological factors and COVID-19 cumulative cases. Correlation is estimated using the non-parametric Spearman's  $\rho$  coefficient. Only significant values (at  $\alpha = 5\%$  significance level) are presented. The maps produced using the *ggmap*<sup>40</sup> and *ggplot2*<sup>41</sup> visualisation libraries under the R programming language<sup>42</sup>.

in 32 countries (out of 54). The prominent factor appears to be dewpoint temperature (in 10 countries), for which negative associations have been observed in this study but also reported in the literature<sup>49,50</sup>. Relative and absolute humidity closely follow, (both occurring in 9 countries), in line with earlier studies<sup>12,19,51,52</sup>. To a similar extent, solar radiation and average temperature are both reported in 8 countries in this study but also in previous research, for both positive<sup>53,54</sup> and negative<sup>52</sup> associations. Finally, to a lesser extent, variables such as rainfall, minimum and maximum temperature appear as the less often occurring factors (in 4 countries)<sup>55</sup>, and wind speed (in 3 countries), as already observed in previous research<sup>3,56</sup>.

Some previous studies also highlighted the major role of average temperature and relative humidity on COVID-19 transmission, which was primarily considered an airborne disease whose transmission might be amplified in humid environments, with low temperatures, less solar radiation and higher wind speed<sup>57,58</sup>. Yet, an interesting finding highlighted in this study is the fact that some of the meteorological factors show significant positive and negative associations with COVID-19, in line with the findings of previous studies<sup>12,59</sup>. In this study, such variables include rainfall, relative and absolute humidity and solar radiation. In the case of solar radiation, previous research suggests that increased solar radiation reduces the likelihood of airborne persistence of the virus in outdoor environments, decreasing its transmission<sup>13</sup>; Yet, the positive association of high solar radiation could be explained by the fact that high temperature increases human mobility, resulting in people gathering outside, breaking lockdown measures, therefore increasing the likelihood of transmission<sup>60</sup>. As for rainfall, relative and humidity, regarding the negative association highlighted in this study, previous research suggests that it can be partly explained by the fact that most people stay home on rainy days, hence decreasing the likelihood of the spread of the disease<sup>61</sup>. Also, rainfall causes the washout of aerosols, which is known to decrease the residence time of the virus in the atmosphere<sup>12</sup>. Other studies however explain that at specific ranges of humidity, the virus can remain active for up to 4–5 days on outdoor surfaces, increasing the likelihood of transmission and resulting in positive association to COVID-19<sup>62,63</sup>. Overall, the findings in this study shed light on the complex and contrasting effect of meteorological factors on COVID-19 prevalence in Africa, but also that the climate sensitivity of the disease varies greatly in space and is not consistent across all countries. Also, it should be acknowledged that other major factors, such as socio-economic features and immunity of the populations might have a stronger influence on the spread of the pandemic<sup>9,62</sup>.





**Figure 7.** Countries showing a negative association between meteorological factors and COVID-19 cumulative cases. Correlation is estimated using the non-parametric Spearman's  $\rho$  coefficient. Only significant values (at  $\alpha = 5\%$  significance level) are presented. The maps produced using the *ggmap*<sup>40</sup> and *ggplot2*<sup>41</sup> visualisation libraries under the R programming language<sup>42</sup>.

This study highlighted major findings regarding the pandemic spread in Africa: first, the relationship between meteorological conditions and COVID-19 and their time-dependence structure have been explored. The substantial geographical differences in the prevalence have also been highlighted, especially towards the end of the study period, which resulted in a clear clustering in the severity of the disease. However, there are also some limitations which should be acknowledged. First, we should be well aware that all models are as perfect as the data used. Especially in the case of African countries, COVID-19 daily reports tend to be fraught with inaccuracies, especially when testing policies are different or not heavily promoted. Second, the analysis framework used in this study relies on time-series-based modelling, in which case time-dependency structural behaviour can sometimes result in spurious correlations, which should often be treated with caution. It should also be noted that some specific factors were not included in our analysis such as air quality due to the lack of accurate

data in the case of Africa. Air quality, especially the concentration of fine particulate matter (PM<sub>2.5</sub>) has been previously highly associated with COVID-19 in previous large-scale studies<sup>9,58</sup>. We should also indicate that this study focused on the individual effect of climate variables on COVID-19, yet the synergistic effect of all climate variables altogether was not considered, leaving out interaction effects<sup>55,64</sup>. Finally, some other shortcomings are the non-consideration of socio-economic factors, healthcare system and other development country-level indicators, and population immunity, which could explain the variation in specific sensitivities of countries to COVID-19<sup>9</sup>. However, in our understanding, such factors are less likely to explain day-to-day variations, which is the target of this study. In this regard, our findings appear to be of importance as they shed light on the contribution of climate to the spread of the pandemic. This, in turn, might help refine further modelling and forecasting attempts, sustain the preparedness of healthcare systems and foster informed management policies.

## Material and methods

### Ethical statement

The data used in this study, regarding both COVID-19, stringency index and meteorological factors, are all available from public databases. Therefore, an ethical review was not required.

### Sources of data used in this study

In this study, COVID-19 cumulative confirmed cases data were gathered for the study period covering January 1, 2020, to September 30, 2022. The data were automatically collected at the daily timestep and for each African country (54 countries in total), using the R package *COVID-19*<sup>6</sup>. The time series were treated for inconsistencies, such as decreasing values in cumulative confirmed cases from 1 day to the next (occurring for example for Benin—BEN).

The climate variables considered in this study for potential significant association with COVID-19 prevalence are cumulative rainfall (*pr*), dewpoint temperature (*tdew*), minimum (*tmin*), mean (*tmoy*) and maximum (*tmax*) temperatures, wind speed (*wspd*), relative humidity (*rh*), absolute humidity (*ah*) and solar radiation (*insol*), both at the daily timescale. The data for these variables were collected from the so-called “Modern-Era Retrospective Analysis for Research and Applications” (MERRA-2<sup>65</sup>), which is a global climate reanalysis (at the spatial resolution of 0.625° × 0.5°), accessible through the NASA POWER Data Access Viewer (<https://power.larc.nasa.gov/data-access-viewer/>). The data were automatically collected using the *nasapower* R package<sup>66</sup>. The time series were also treated for missing values using spline interpolation. The daily values of absolute humidity (*ah*) were calculated from the Clausius Clapeyron approximation, as shown in Eq. (1)<sup>9</sup>:

$$ah = \left( 6.112 \times \frac{17.67 \times tmoy}{\exp(tmoy + 243.5)} \times rh \times 2.1674 \right) / (tmoy + 273.15) \quad (1)$$

where *ah* is the absolute humidity (%), *tmoy* is the average temperature (°C), *rh* is the air relative humidity (%) and *exp* is the natural base logarithm.

Over the study period, various and different policy responses to COVID-19 have been adopted by countries to contain and mitigate the spread of the pandemic. In this study, as the analysis of the potential effect of climate variables is carried out on a cross-sectional basis, it is critical to isolate the potential effect of such different policy responses from that of the climate variables considered. Therefore, a stringency index (*si*), defined as a composite mean score of nine pandemic response measures (school closures, workplace closures, cancellation of public events, restrictions on public gatherings, closures of public transport, stay-at-home requirements, public information campaigns, restrictions on internal movements and international travel controls) is systematically considered as a potential covariate for each country’s COVID-19 confirmed cases time series. The *si* index takes a value between 0 and 100, with a higher score indicating a stricter mitigation response<sup>67</sup>. The *si* time series data were collected from Ritchie et al.<sup>68</sup>.

### Inter-relations and spatial patterns in climate variables and COVID-19 prevalence

Climate variables often exhibit medium to long-term associations depending on the local climate system. In this study, the potential relationships between the 9 climate variables (*pr*, *tdew*, *tmin*, *tmoy*, *tmax*, *wspd*, *rh*, *ah* and *insol*) are assessed through the non-parametric Spearman’s rank correlation test (at  $\alpha = 5\%$  significance level). To account for regional contrasts, the correlations were evaluated depending on regions, i.e. North Africa (NAF), West Africa (WAF), Central Africa (CAF), East Africa (EAF) and Southern Africa (SAF). Also, the average spatial distribution of these variables is presented through choropleths maps.

In addition, because of seasonal cycles, intra-annual weather variability can be typically masked, blurring the inter-relationships analysed. To address this issue, we complement the above correlation analysis with the examination of correlations between the weather variables anomalies, still using the non-parametric Spearman’s rank correlation (at  $\alpha = 5\%$  significance level). The anomalies are evaluated by subtracting the weather variables values from the day-by-day country climatological mean (or daily climate normal) evaluated over the period 2000–2019 (i.e., 20 years). The daily climate normals are calculated by averaging the daily time series for each meteorological variable for each day of the year (1, 2, ..., 365). For leap years, the date of February 29 (i.e., the 60th day of the year) was skipped to maintain the climate normals time series to a fixed length of 365 values, for ease of computation. The data were collected from MERRA-2 climate reanalysis through the use of the *nasapower* package<sup>66</sup>.

Furthermore, to evaluate the spatial clustering of the pandemic spread at the level of the continent, the spatial autocorrelation in cumulative cases (per million people) at different time points selected within the study period is assessed through Global Moran’s *I*<sup>69</sup>. Moran’s *I* statistic is a common indicator of global spatial autocorrelation,

defined as a cross-product statistic between a variable and its spatial lag, with the variable expressed in deviations from its mean value, as in Eq. (2):

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j \omega_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \quad (2)$$

where  $x_i$  and  $x_j$  is a pair of distinct observations at locations  $i, j$  (respectively),  $\bar{x}$  is the mean of the variable  $x$ ,  $\omega_{ij}$  are the elements of spatial weights matrix,  $S_0 = \sum_i \sum_j \omega_{ij}$  is the sum of all the weights and  $n$  is the number of observations. The expected value of  $I$  under the null hypothesis (no spatial autocorrelation) is given by  $I_0 = -1/(n-1)^{70}$ . The evaluation of Moran's  $I$  was carried out in this study through the *ape* R package<sup>71</sup> using the inverse Euclidean distance on all pairs of countries' centroid coordinates to compute the spatial matrix weights.

### Potential association of climate variables to COVID-19 transmission

The general workflow of the methodology used to detect the potential association between meteorological variables and COVID-19 is presented in Fig. 8. The following sections describe in detail the steps of the analysis.

#### Pre-processing COVID-19 cumulative cases time series

The time series of COVID-19 cumulative confirmed cases for each country are normalized through the *min-max* method, in the 0–1 range (Eq. 3). This operation ensures cross-comparison of the results (between countries), especially when estimating the magnitude of the effect of the climate variables analysed.

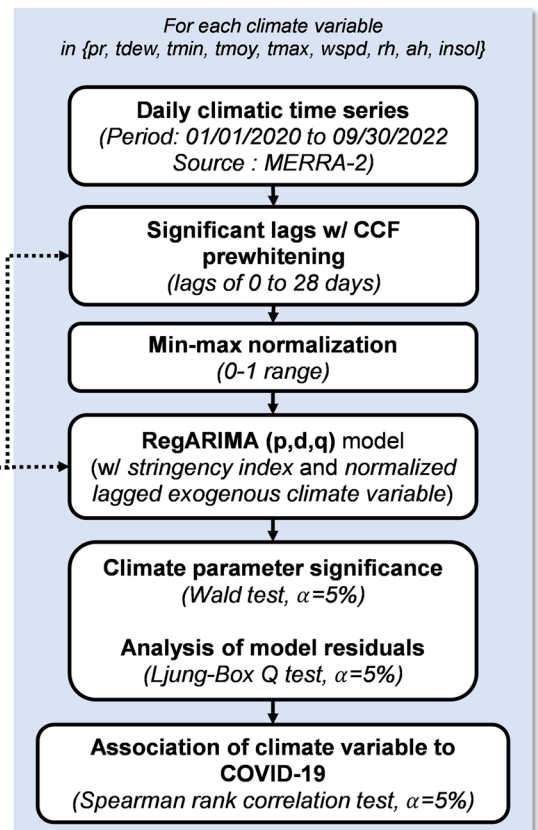
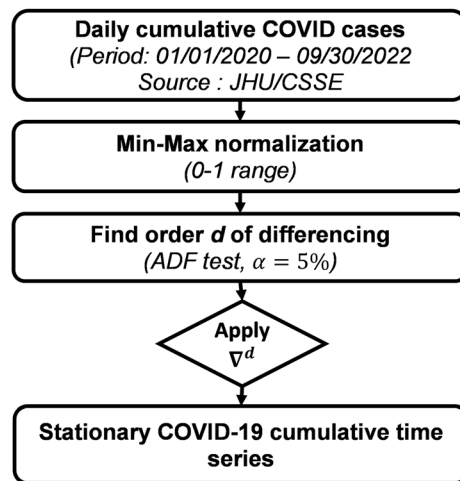
$$x = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (3)$$

where  $x$  is the normalized time series,  $X$  is the original time series,  $\min(X)$  and  $\max(X)$  are respectively the minimum and maximum values in the  $X$  time series.

#### Time series-based modelling of COVID-19 and effect of climate variables

The modelling framework used in this study is based on the *Autoregressive Integrated Moving Average* (ARIMA) model, which is a common approach used in time series modelling, resulting in the backshift notation presented in Eq. (4)<sup>72</sup>:

For each African country...



**Figure 8.** Methodology flowchart of the study.

$$(1 - B)^d Y_t = \mu + \frac{\theta(B)}{\phi(B)} \alpha_t \quad (4)$$

where  $t$  indexes time (in days),  $\mu$  is the mean term,  $B$  is the backshift operator (that is  $BY_t = Y_{t-1}$ ),  $\phi(B)$  is the autoregressive operator represented as a polynomial in the backshift operator ( $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ ),  $\theta(B)$  is the moving average operator represented as a polynomial in the backshift operator ( $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ ) and  $\alpha_t$  is the independent random error term. The ARIMA model is defined through three non-negative parameters  $p, d, q$ , respectively defined as the number of time lags, the degree of differencing and the order of the moving average model. The order of differencing  $d$  is obtained through successive applications of the Augmented Dickey-Fuller (ADF) test (at  $\alpha = 5\%$  significance level) to increasing levels of differencing applied to the original time series. However, in this study, a minimum value of  $d = 1$  is fixed to ensure that the modelled time series are typically daily new cases, instead of the raw cumulative values. The time series obtained after differencing is considered stationary. The procedure was carried out using the *ndiffs* command from the R forecast package<sup>72</sup>.

The potential effect of climate variables on COVID-19 confirmed cases is assessed through a variation of the ARIMA ( $p, d, q$ ) model, which includes exogenous explanatory variables. Previous studies indicate that the effect of climate is not necessarily immediate on COVID-19 transmission, but can typically be lagged by up to 28 days<sup>73</sup>. Therefore, in this study, for each of the 9 climate variables considered (*pr*, *tdew*, *tmin*, *tmoy*, *tmax*, *wspd*, *rh*, *ah* and *insol*), the time series data collected over the study period is initially screened to filter significant lags (between 0 to 28 days) through pre-whitening. Pre-whitening helps in evaluating at which specific lag a covariate (independent variable) affects a dependent variable<sup>74</sup>. The steps in the pre-whitening procedure include:

- i. fitting an ordinary ARIMA ( $p, d, q$ ) model to the dependent variable time series and saving the residuals (say *res<sub>Y</sub>*);
- ii. fitting the same model to the covariate time series and saving the residuals (say *res<sub>X</sub>*);
- iii. analyzing the cross-correlation function between the residuals *res<sub>X</sub>* and *res<sub>Y</sub>* and identify significant lags (using nonparametric Spearman's rank correlation at  $\alpha = 5\%$  significance level). These lags are defined as the lags for which cross-correlation exceeds the 95% confidence interval limits under the assumption that the null hypothesis ( $H_0$ : no correlation) is true. Here, to account for the length of the time series, which could deflate the reported p-value<sup>75</sup> and to provide better estimates (and confidence interval around these estimates<sup>76</sup>), the correlation was assessed through screening of the distribution of 1000 sieve random bootstrap simulations, carried out using the '*ccf\_boot*' command in the R *funtimes* package<sup>77</sup>.

Another major advantage in working with the residuals (from the pre-whitening procedure) is that such residuals are assumed to be independent, while the original meteorological variables are unlikely, since weather systems could persist for many days, therefore reducing the number of degrees of freedom of the time series. This effect is further amplified by seasonal fluctuations, which should be thoroughly addressed when assessing for statistical significance to avoid the detection of spurious correlations<sup>78,79</sup>.

Also, in this study, only strictly negative lags were retained, i.e. climate leads to COVID-19 cases, but not the other way around. The significant lags identified are applied by shifting the climate variable time series considered, which is further normalized in the 0–1 range through the *min-max* method (see Eq. 3). This ensures cross-comparison of the results, especially considering that climate variable values are defined on different ranges.

Further, a regression model with ARIMA ( $p, d, q$ ) errors is developed using simultaneously the stringency *si* index time series (also *min-max* normalized in the 0–1 range) and the considered climate covariate. The regression model obtained, hereafter referred to as *RegARIMA* ( $p, d, q$ ), is defined as in the backshift notation given in Eq. (5):

$$y_t = \beta x_t + \frac{\theta(B)}{(1 - B)^d \phi(B)} z_t \quad (5)$$

where  $y_t$  is the response variable (confirmed cases),  $\beta$  is a coefficient matrix,  $x_t$  is the covariate matrix formed by the stringency index and the climate variable considered and  $z_t$  is a white noise process. The *RegARIMA* ( $p, d, q$ ) model parameters  $p$  and  $q$  are obtained through maximum likelihood fitting using the *auto.arima* function provided with the *forecast* R package<sup>72</sup>. The Bayesian Information Criterion (BIC) is used to select the optimal model, defined as the one minimizing the BIC<sup>12,72</sup>. The definitive optimal and meaningful models are retained based on the following conditions:

- i. *the model residuals are typically similar to white noise*: this condition further ensures that the model has captured the maximum variance in the dependent variable (i.e., the COVID-19 prevalence data). This is assessed through the application of the Ljung-Box Q test (at  $\alpha = 5\%$  significance level)<sup>72,80</sup>.
- ii. *the climate variable coefficient is significant in the model*: this condition is assessed through the Wald t-test of coefficients at  $\alpha = 5\%$  significance level. This statistical test is applied with the *lmtest* R package<sup>81</sup>.
- iii. *the climate variable with its actual lag is significantly correlated to COVID-19 cumulative cases*: this condition is assessed using the non-parametric Spearman rank correlation test at  $\alpha = 5\%$  significance level. The correlation coefficient  $\rho$  is an indication of the effect size of the relationship, and assesses the association both in terms of magnitude and direction (positive or negative). A positive (negative) association implies that as the climate variable increases, the number of reported cases increases (decreases).

## Data availability

All the data used in this study are publicly available from various sources included in the manuscript.

## Code availability

The code used to analyse the data and prepare the figures is available at Github ([https://github.com/Yonaba/covid19\\_climate\\_ts\\_modelling](https://github.com/Yonaba/covid19_climate_ts_modelling)).

Received: 4 November 2022; Accepted: 26 October 2023

Published online: 31 October 2023

## References

- Holmes, E. C. *et al.* The origins of SARS-CoV-2: A critical review. *Cell* **184**, 4848–4856 (2021).
- Green, M. S. Did the hesitancy in declaring COVID-19 a pandemic reflect a need to redefine the term?. *Lancet* **395**, 1034–1035 (2020).
- d'Albis, H., Coulibaly, D., Roumagnac, A., de Carvalho Filho, E. & Bertrand, R. Quantification of the effects of climatic conditions on French hospital admissions and deaths induced by SARS-CoV-2. *Sci. Rep.* **11**, 21812 (2021).
- Miller, R. K. & Hui, I. Impact of short school closures (1–5 days) on overall academic performance of schools in California. *Sci. Rep.* **12**, 2079 (2022).
- Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020).
- Guidotti, E. & Ardia, D. COVID-19 data hub. *JOSS* **5**, 2376 (2020).
- Lawal, Y. Africa's low COVID-19 mortality rate: A paradox?. *Int. J. Infect. Dis.* **102**, 118–122 (2021).
- Lulbadda, K. T., Kobbekaduwa, D. & Guruge, M. L. The impact of temperature, population size and median age on COVID-19 (SARS-CoV-2) outbreak. *Clin. Epidemiol. Glob. Health* **9**, 231–236 (2021).
- Sidibé, M. L. *et al.* Understanding the COVID-19 pandemic prevalence in Africa through optimal feature selection and clustering: Evidence from a statistical perspective. *Environ. Dev. Sustain.* <https://doi.org/10.1007/s10668-022-02646-3> (2022).
- Randazzo, W., Cuevas-Ferrando, E., Sanjuán, R., Domingo-Calap, P. & Sánchez, G. Metropolitan wastewater analysis for COVID-19 epidemiological surveillance. *Int. J. Hyg. Environ. Health* **230**, 113621 (2020).
- Iqbal, A., Haq, W., Mahmood, T. & Raza, S. H. Effect of meteorological factors on the COVID-19 cases: A case study related to three major cities of the Kingdom of Saudi Arabia. *Environ. Sci. Pollut. Res.* **29**, 21811–21825 (2022).
- Hossain, Md. S., Ahmed, S. & Uddin, Md. J. Impact of weather on COVID-19 transmission in south Asian countries: An application of the ARIMAX model. *Sci. Total Environ.* **761**, 143315 (2021).
- Chen, S. *et al.* Climate and the spread of COVID-19. *Sci. Rep.* **11**, 9042 (2021).
- Sahoo, M. M. Significance between air pollutants, meteorological factors, and COVID-19 infections: Probable evidences in India. *Environ. Sci. Pollut. Res.* **28**, 40474–40495 (2021).
- Jefferson, T., Spencer, E., Plüddemann, A., Roberts, N. & Heneghan, C. Analysis of the transmission dynamics of COVID-19: an Open Evidence Review. <https://www.cebm.net/evidence-synthesis/transmission-dynamics-of-covid-19> (2021).
- Bashir, M. F. *et al.* Correlation between climate indicators and COVID-19 pandemic in New York, USA. *Sci. Total Environ.* **728**, 138835 (2020).
- Zhuang, Z. *et al.* Preliminary estimation of the novel coronavirus disease (COVID-19) cases in Iran: A modelling analysis based on overseas cases and air travel data. *Int. J. Infect. Dis.* **94**, 29–31 (2020).
- Kerr, G. H., Badr, H. S., Gardner, L. M., Perez-Saez, J. & Zaitchik, B. F. Associations between meteorology and COVID-19 in early studies: Inconsistencies, uncertainties, and recommendations. *One Health* **12**, 100225 (2021).
- Wang, *et al.* Impact of temperature and relative humidity on the transmission of COVID-19: A modelling study in China and the United States. *BMJ Open* **11**, e043863 (2021).
- Jüni, P. *et al.* Impact of climate and public health interventions on the COVID-19 pandemic: A prospective cohort study. *CMAJ* **192**, E566–E573 (2020).
- Ganslmeier, M., Furceri, D. & Ostry, J. D. The impact of weather on COVID-19 pandemic. *Sci. Rep.* **11**, 22027 (2021).
- Rodó, X., San-José, A., Kirchgatter, K. & López, L. Changing climate and the COVID-19 pandemic: More than just heads or tails. *Nat. Med.* **27**, 576–579 (2021).
- Briz-Redón, A. & Serrano-Aroca, Á. The effect of climate on the spread of the COVID-19 pandemic: A review of findings, and statistical and modelling techniques. *Prog. Phys. Geogr. Earth Environ.* **44**, 591–604 (2020).
- Liu, J. *et al.* Impact of meteorological factors on the COVID-19 transmission: A multi-city study in China. *Sci. Total Environ.* **726**, 138513 (2020).
- Mohamadou, Y., Halidou, A. & Kapen, P. T. A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19. *Appl. Intell.* **50**, 3913–3925 (2020).
- Zongo, P., Zorom, M., Mophou, G., Dorville, R. & Beaumont, C. A model of COVID-19 transmission to understand the effectiveness of the containment measures: Application to data from France. *Epidemiol. Infect.* **148**, e221 (2020).
- Zhang, P. *et al.* Usage of compartmental models in predicting COVID-19 outbreaks. *AAPS J.* **24**, 98 (2022).
- Torrealba-Rodriguez, O., Conde-Gutiérrez, R. A. & Hernández-Javier, A. L. Modeling and prediction of COVID-19 in Mexico applying mathematical and computational models. *Chaos Solitons Fract.* **138**, 109946 (2020).
- Conde-Gutiérrez, R. A., Colorado, D. & Hernández-Bautista, S. L. Comparison of an artificial neural network and Gompertz model for predicting the dynamics of deaths from COVID-19 in México. *Nonlinear Dyn.* **104**, 4655–4669 (2021).
- Ilie, O.-D. *et al.* Forecasting the spreading of COVID-19 across nine countries from Europe, Asia, and the American continents using the ARIMA models. *Microorganisms* **8**, 1158 (2020).
- Naing, C., Ni, H., Aung, H. H., Chan, E. & Mak, J. W. The influence of climate factors on COVID-19 transmission in Malaysia: An autoregressive integrated moving average (ARIMA) model. <https://doi.org/10.1101/2020.08.14.20175372> (2020).
- Sahai, A. K., Rath, N., Sood, V. & Singh, M. P. ARIMA modelling & forecasting of COVID-19 in top five affected countries. *Diabetes Metab. Syndr. Clin. Res. Rev.* **14**, 1419–1427 (2020).
- Adekunle, I. A., Tella, S. A., Oyesiku, K. O. & Oseni, I. O. Spatio-temporal analysis of meteorological factors in abating the spread of COVID-19 in Africa. *Heliyon* **6**, e04749 (2020).
- Martinez-Alvarez, M. *et al.* COVID-19 pandemic in west Africa. *Lancet Glob. Health* **8**, e631–e632 (2020).
- Ssentongo, P. *et al.* Tracking and predicting the African COVID-19 pandemic. <https://doi.org/10.1101/2020.11.13.20231241> (2020).
- Ai, H., Nie, R. & Wang, X. Evaluation of the effects of meteorological factors on COVID-19 prevalence by the distributed lag nonlinear model. *J. Transl. Med.* **20**, 170 (2022).
- Gwenzi, W. & Rzymiski, P. When silence goes viral, Africa sneezes! A perspective on Africa's subdued research response to COVID-19 and a call for local scientific evidence. *Environ. Res.* **194**, 110637 (2021).
- ACSS. Africa's varied COVID landscapes. *Africa Center for Strategic Studies* (2020).

39. ACSS. Mapping COVID-19 risk factors. *Africa Center for Strategic Studies* (2020).
40. Kahle, D. & Wickham, H. ggmap: Spatial Visualization with ggplot2. (2016).
41. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. (2016).
42. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2022).
43. Maidment, R. I. *et al.* The 30 year TAMSAT African rainfall climatology and time series (TARCAT) data set. *J. Geophys. Res. Atmos.* <https://doi.org/10.1002/2014JD021927> (2014).
44. Chatterjee, B., Karandikar, R. L. & Mande, S. C. *Paradoxical Case Fatality Rate dichotomy of Covid-19 among rich and poor nations points to the "hygiene hypothesis"*. <https://doi.org/10.1101/2020.07.31.20165696> (2020).
45. Iesa, M. A. M. *et al.* SARS-CoV-2 and *Plasmodium falciparum* common immunodominant regions may explain low COVID-19 incidence in the malaria-endemic belt. *New Microbes New Infect.* **38**, 100817 (2020).
46. Anjorin, A. A. *et al.* Comorbidities and the COVID-19 pandemic dynamics in Africa. *Trop. Med. Int. Health* **26**, 2–13 (2021).
47. Baker, R. E., Yang, W., Vecchi, G. A., Metcalf, C. J. E. & Grenfell, B. T. Susceptible supply limits the role of climate in the early SARS-CoV-2 pandemic. *Science* **369**, 315–319 (2020).
48. Braiman, M. Latitude dependence of the COVID-19 mortality rate—A possible relationship to vitamin D deficiency?. *SSRN Electron. J.* <https://doi.org/10.2139/ssrn.3561958> (2020).
49. Selcuk, M., Gormus, S. & Guven, M. Impact of weather parameters and population density on the COVID-19 transmission: Evidence from 81 provinces of Turkey. *Earth Syst. Environ.* **5**, 87–100 (2021).
50. Neisi, A. *et al.* Association of the corona virus (Covid-19) epidemic with environmental risk factors. *Environ. Sci. Pollut. Res.* **30**, 60314–60325 (2023).
51. Ganegoda, N. C., Wijaya, K. P., Amadi, M., Erandi, K. K. W. H. & Aldila, D. Interrelationship between daily COVID-19 cases and average temperature as well as relative humidity in Germany. *Sci. Rep.* **11**, 11302 (2021).
52. Islam, Md. M. & Noor, F. M. Correlation between COVID-19 and weather variables: A meta-analysis. *Heliyon* **8**, e10333 (2022).
53. Hachim, M. Y. *et al.* Higher temperatures, higher solar radiation, and less humidity is associated with poor clinical and laboratory outcomes in COVID-19 patients. *Front. Public Health* **9**, 618828 (2021).
54. Liu, M. *et al.* Association between temperature and COVID-19 transmission in 153 countries. *Environ. Sci. Pollut. Res.* **29**, 16017–16027 (2022).
55. Moazeni, M., Rahimi, M. & Ebrahimi, A. What are the effects of climate variables on COVID-19 pandemic? A systematic review and current update. *Adv. Biomed. Res.* **12**, 33 (2023).
56. Rendana, M. Impact of the wind conditions on COVID-19 pandemic: A new insight for direction of the spread of the virus. *Urban Clim.* **34**, 100680 (2020).
57. Barcellos, D. D. S., Fernandes, G. M. K. & de Souza, F. T. Data based model for predicting COVID-19 morbidity and mortality in metropolis. *Sci. Rep.* **11**, 24491 (2021).
58. Lim, Y. K., Kweon, O. J., Kim, H. R., Kim, T.-H. & Lee, M.-K. The impact of environmental variables on the spread of COVID-19 in the Republic of Korea. *Sci. Rep.* **11**, 5977 (2021).
59. McClymont, H. & Hu, W. Weather variability and COVID-19 transmission: A review of recent research. *IJERPH* **18**, 396 (2021).
60. Damette, O., Mathonnat, C. & Goutte, S. Meteorological factors against COVID-19 and the role of human mobility. *PLoS ONE* **16**, e0252405 (2021).
61. Shenoy, A. *et al.* God is in the rain: The impact of rainfall-induced early social distancing on COVID-19 outbreaks. *J. Health Econ.* **81**, 102575 (2022).
62. Doğan, B., Ben Jebli, M., Shahzad, K., Farooq, T. H. & Shahzad, U. Investigating the effects of meteorological parameters on COVID-19: Case study of New Jersey, United States. *Environ. Res.* **191**, 110148 (2020).
63. Lewis, D. Mounting evidence suggests coronavirus is airborne—but health advice has not caught up. *Nature* **583**, 510–513 (2020).
64. Yonaba, R. *et al.* Trends, sensitivity and estimation of daily reference evapotranspiration ET<sub>0</sub> using limited climate data: Regional focus on Burkina Faso in the West African Sahel. *Theor. Appl. Climatol.* <https://doi.org/10.1007/s00704-023-04507-z> (2023).
65. Gelaro, R. *et al.* The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *J. Clim.* **30**, 5419–5454 (2017).
66. Sparks, A. nasapower: NASA-POWER Data from R. (2021).
67. Hale, T. *et al.* A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nat. Hum. Behav.* **5**, 529–538 (2021).
68. Ritchie, H. *et al.* Coronavirus Pandemic (COVID-19). *Our World in Data* (2020).
69. Moran, P. A. P. The interpretation of statistical maps. *J. R. Stat. Soc. Ser. B Methodol.* **10**, 243–251 (1948).
70. Anselin, L. The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In *Spatial Analytical Perspectives on GIS* (eds Fischer, M. *et al.*) 111–126 (Routledge, 2019). <https://doi.org/10.1201/9780203739051-8>.
71. Paradis, E. & Schliep, K. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
72. Hyndman, R. J. & Khandakar, Y. Automatic time series forecasting: The forecast package for R. *J. Stat. Soft.* <https://doi.org/10.18637/jss.v027.i03> (2008).
73. Bañuelos Gimeno, J. *et al.* Air pollution and meteorological variables' effects on COVID-19 first and second waves in Spain. *Int. J. Environ. Sci. Technol.* <https://doi.org/10.1007/s13762-022-04190-z> (2022).
74. Razavi, S. & Vogel, R. Prewhitening of hydroclimatic time series? Implications for inferred change and variability across time scales. *J. Hydrol.* **557**, 109–115 (2018).
75. Halsey, L. G., Curran-Everett, D., Vowler, S. L. & Drummond, G. B. The fickle P value generates irreproducible results. *Nat. Methods* **12**, 179–185 (2015).
76. Wilcox, R. R. *Introduction to Robust Estimation and Hypothesis Testing* (Elsevier Science, 2011).
77. Lyubchich, V., Gel, Y. R. & Vishwakarma, S. funtimes: Functions for Time Series Analysis. (2023).
78. Bretherton, C. S., Widmann, M., Dymnikov, V. P., Wallace, J. M. & Bladé, I. The effective number of spatial degrees of freedom of a time-varying field. *J. Clim.* **12**, 1990–2009 (1999).
79. Metz, W. Optimal relationship of large-scale flow patterns and the barotropic feedback due to high-frequency eddies. *J. Atmos. Sci.* **48**, 1141–1159 (1991).
80. Ljung, G. M. & Box, G. E. P. On a measure of lack of fit in time series models. *Biometrika* **65**, 297–303 (1978).
81. Zeileis, A. & Hothorn, T. Diagnostic checking in regression relationships. *R. News* **2**, 7–10 (2002).

## Author contributions

O.K., R.Y., F.T., H.K. conceived the analysis, collected and analysed the data, prepared the figures and wrote the first draft of the paper. M.L.S., B.L., M.D., H.A.A., H.K. reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-46007-0>.

**Correspondence** and requests for materials should be addressed to R.Y.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023