



Pan-African evolution of within- and between-country COVID-19 dynamics

Paddy Ssentongo^{a,b,1}, Claudio Fronterre^{c,1}, Andrew Geronimo^{d,1}, Steven J. Greybush^e, Pamela K. Mbabazi^f, Joseph Muvawala^f, Sarah B. Nahalamba^f, Philip O. Omadi^f, Bernard T. Opar^g, Shamim A. Sinnar^h, Yan Wang^e, Andrew J. Whalen^{a,i}, Leonhard Held^{j,2}, Christopher Jewell^{c,2}, Abraham J. B. Muwanguzi^{f,2}, Helen Greatrex^{e,2}, Michael M. Norton^{a,2}, Peter J. Diggle^{c,2}, and Steven J. Schiff^{f,a,d,k,2,3}

^aCenter for Neural Engineering, Department of Engineering Science and Mechanics, The Pennsylvania State University, University Park, PA 16802; ^bDepartment of Public Health Sciences, The Pennsylvania State University College of Medicine, Hershey, PA 17033; ^cCentre for Health Informatics, Computing and Statistics, Lancaster University, Lancaster LA1 4YW, United Kingdom; ^dDepartment of Neurosurgery, The Pennsylvania State University College of Medicine, Hershey, PA 17033; ^eDepartment of Meteorology and Atmospheric Science, The Pennsylvania State University, University Park, PA 16802; ^fNational Planning Authority, Kampala, Uganda; ^gMinistry of Health, Kampala, Uganda; ^hDepartment of Medicine, The Pennsylvania State University College of Medicine, Hershey, PA 17033; ⁱDepartment of Neurosurgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114; ^jEpidemiology, Biostatistics and Prevention Institute, University of Zurich, 8001 Zurich, Switzerland; and ^kCenter for Infectious Disease Dynamics, and Department of Physics, The Pennsylvania State University, University Park, PA 16802

Edited by Douglas S. Massey, Princeton University, Princeton, NJ, and approved May 25, 2021 (received for review December 28, 2020)

The coronavirus disease 2019 (COVID-19) pandemic is heterogeneous throughout Africa and threatening millions of lives. Surveillance and short-term modeling forecasts are critical to provide timely information for decisions on control strategies. We created a strategy that helps predict the country-level case occurrences based on cases within or external to a country throughout the entire African continent, parameterized by socioeconomic and geoeconomic variations and the lagged effects of social policy and meteorological history. We observed the effect of the Human Development Index, containment policies, testing capacity, specific humidity, temperature, and landlocked status of countries on the local within-country and external between-country transmission. One-week forecasts of case numbers from the model were driven by the quality of the reported data. Seeking equitable behavioral and social interventions, balanced with coordinated country-specific strategies in infection suppression, should be a continental priority to control the COVID-19 pandemic in Africa.

COVID-19 modeling | forecast | Africa | meteorology | Human Development Index

The ongoing coronavirus disease 2019 (COVID-19) pandemic in Africa is threatening millions of lives, a crisis compounded by the continent's unique spectrum of disease and fragile healthcare infrastructure (1). Essential to African countries' efforts to control the pandemic are effective methods to track and predict new cases and their sources in real time. Time-critical interpretation of daily case data is required to inform public health policy on mitigation strategies and resource allocation. To address this need, we developed a data-driven disease surveillance framework to track and predict country-level case incidence from internal and external sources. We chose a spatiotemporal strategy to take advantage of and combine openly available data on coronavirus epidemiology, social policy affecting human movement and public health, meteorological factors, and socioeconomic and demographic variables, seeking to inform rapid policy development.

The first COVID-19 case on the continent was reported in Egypt on February 14, 2020. By August 13, 2020, over 1 million new cases and over 20,000 deaths had been reported in all African Union (AU) Member States, according to the Africa Centres for Disease Control and Prevention (<https://africacdc.org/covid-19/>). Over 44 million cases and 190,000 deaths in Africa are projected within the first year of the pandemic (2). Although Africa has a younger age distribution that could theoretically lead to fewer symptomatic or severe infections (3), modeling predicts that the relatively low healthcare capacity in many parts of Africa, in combination with the large, intergenera-

tional households (4), could lead to infection fatality rates higher than those seen in high-income countries (1). In addition, the high prevalence of comorbidities such as HIV/AIDS is predicted to lead to increased risk of severe COVID-19 in infected individuals (5). Moreover, the coexistence of infectious diseases such as malaria (6), tuberculosis (7), dengue (8), Ebola (9), and others (10) poses additional significant medical and infrastructure challenges in controlling the COVID-19 epidemic in Africa.

Meteorological variables have been linked to the transmission of and survival from seasonal influenza (11–14), severe acute respiratory syndrome coronavirus (SARS-CoV) (15–17), and Middle East respiratory syndrome coronavirus (MERS-CoV) (18, 19). It is therefore unsurprising that there are many recent studies exploring the link between temperature, humidity, and COVID-19. All studies to date have focused on modeling, or

Significance

We created a strategy for understanding the evolution of the COVID-19 pandemic throughout the African continent. Because high-quality mobility data are challenging to obtain across Africa, the approach provides the ability to distinguish cases arising from within a country or from its neighbors. The results further show how testing capacity and social and health policy contribute to the dynamics of cases, and generate short-term prediction of the evolution of the pandemic on a country-by-country basis. This framework improves the ability to interpret and act upon real-time complex COVID-19 data from the African continent. These findings emphasize that regional efforts to coordinate country-specific strategies in transmission suppression should be a continental priority to control the COVID-19 pandemic in Africa.

Author contributions: P.S., C.F., A.G., L.H., C.J., M.M.N., and S.J.S. designed research; P.S., C.F., A.G., S.J.G., Y.W., A.J.W., H.G., M.M.N., and S.J.S. performed research; P.S., C.F., A.G., S.J.G., Y.W., H.G., M.M.N., and S.J.S. analyzed data; and P.S., C.F., A.G., S.J.G., P.K.M., J.M., S.B.N., P.O.O., B.T.O., S.A.S., A.J.W., L.H., C.J., A.J.B.M., H.G., M.M.N., P.J.D., and S.J.S. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

See [online](#) for related content such as Commentaries.

¹P.S., C.F., and A.G. contributed equally to this work.

²L.H., C.J., A.J.B.M., H.G., M.M.N., P.J.D., and S.J.S. contributed equally to this work.

³To whom correspondence may be addressed. Email: steven.j.schiff@gmail.com.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2026664118/-DCSupplemental>.

Published June 29, 2021.

on identifying a statistical link between meteorological variables against reported COVID-19 cases, without laboratory studies. A recent systematic review reports agreement among published research, with cold and dry conditions contributing to COVID-19 transmission (20). However, these results must be considered preliminary. Early studies of COVID-19 transmission focused on the emerging pandemic during the boreal spring of 2020 (March through May), when the majority of cases were found in China, the United States, and Europe. It is difficult, therefore, to extrapolate meteorological results to the very different climates found in the tropics, for example, from the subsequent outbreaks in India and Brazil. Many low- and middle-income countries, as defined by the World Bank, are located in the tropics, where many potential confounding factors could mimic a weather signal. These factors include median age, testing and health capabilities, population density, access to sanitation, and the number of new cases arriving in a country through global travel hubs (21). It is also difficult to extract seasonality from a single outbreak that has only lasted a single full year, and considerable caution has been raised regarding tropical caseload and confounding factors (22).

The human response to the pandemic can also drastically shape its timing and intensity. Where data on social distancing are sparse, government testing and stringency policies (see *Materials and Methods*) can be used as a common surrogate to compare countries' efforts to contain the spread of the virus, bolster healthcare systems, enact rigorous testing policy, and provide economic support. The Oxford Coronavirus Government Tracker (OxCGRT) standardizes these complex systems into a set of policy metrics in each of these domains (23). More strict social policies identified in the OxCGRT have been associated with reductions in human mobility (24, 25). Across 161 countries, some of these policies were significantly associated with lower per capita mortality (26), including school closing, canceling public events, and restrictions on gatherings and international travel. Likewise, others have found that strict policies are negatively associated with the growth of new cases (27–29). The relationship between policy and observed changes in social distancing, case numbers, and mortality is complicated by an unknown delay of effect. One estimate indicates a decline in growth of new cases within 1 wk of enacting strict policy, and deceleration of growth within 2 wk (29). Although the implementation of containment policies can be and has been used by many African nations (30), lockdown cannot be maintained in these countries without a worsening of severe poverty and resultant loss of life (31, 32).

We have therefore developed a COVID-19 surveillance, modeling, and prediction strategy that explores a growing spatiotemporal database on coronavirus epidemiology, meteorology, and social policy interventions. To model the spread of COVID-19 in Africa, we employ a data-driven endemic–epidemic model (33) to 1) visualize the burden of cases, including the proportion of cases arising from sources local within country and external between country, 2) describe the factors which most correlate with spread, and 3) enable short-term forecasting of new cases. This modeling framework has been used previously to fit space–time dynamics of COVID-19 in Italy (34), Germany (35), and the United Kingdom (36) and to analyze other infectious diseases (37).

Results

COVID-19 Spread and Response. As of August 13, 2020, the 55 AU Member States had reported over 1,000,000 cases and 20,000 deaths from COVID-19. The southern region had the most cases, reporting over 50% (over 560,000 cases and 11,000 deaths) of the total for the continent. North Africa carries the highest regional case fatality rate (4%) but contributes 20% of the continent's cases, with countries such as Egypt (102 cases per

100,000), Morocco (94 cases per 100,000), and Algeria (83 cases per 100,000) driving the overall numbers (Fig. 1). As more countries conduct targeted mass screening and testing, these figures continue to change. The spatial distribution of cases per 100,000 displays no clear geographical pattern (Fig. 1). South Africa, Djibouti, Equatorial Guinea, Gabon, and Egypt carry the largest burden of cases per capita, ranging from 100 to 500 per 100,000. The epidemiological curves for the African countries display varying shapes, mostly driven by the frequency and intensity of testing. For example, the epidemiological curve of South Africa is similar to those of the United Kingdom and the United States (*SI Appendix, Fig. S1*). An exception is Tanzania, which stopped reporting new cases in late April of 2020 (*SI Appendix, Fig. S1*).

Time series for case incidence and temporally varying model inputs are shown for selected countries in Fig. 1A. The full set of case incidence time series for all countries can be found in *SI Appendix, Fig. S1*. A majority of the countries imposed containment policies, including lockdowns and curfews, in early March 2020 to prevent further COVID-19 transmission within their borders. These social policy interventions remained in effect through August 2020 for most countries (*SI Appendix, Fig. S2*). Testing policies, which were restrictive at the beginning of the pandemic due to inadequate testing infrastructure, have become more open as testing is made widely available (*SI Appendix, Fig. S3*). As expected, spatiotemporal distribution of temperatures, rainfall, and specific humidity are very heterogeneous across the continent (Fig. 1 and *SI Appendix, Figs. S4–S6*). Population-weighted averages of these three meteorological variables were calculated for each country and day. This type of weighting prioritizes the human–climate interaction over the land–climate interaction (*SI Appendix, Figs. S7–S9*).

Optimal Model. The spatiotemporal dynamics of reported cases are modeled as additive components. Two epidemic sources capture infections coming from within the country and from all other countries. An endemic component includes all contributions to the reported number of cases that are not taken into account by the epidemic part. The endemic part is not driven by previous case counts but may account for factors such as seasonality, sociodemography, animal reservoirs, and population. The epidemic part of the model has an autoregressive nature, meaning that the past number of COVID-19 cases reported both within a specific country and in the rest of the continent will be used to forecast the trend of COVID-19 cases. How much the past observations contribute to the future disease count depends on two parameters, λ for the local transmission and ϕ for the external transmission, and will be estimated from the data. In particular, the impact of cases reported in the neighboring countries depends also on a set of weights that modulate the spatial connectivity of the countries in the continent (see *Materials and Methods*). These two parameters are also functions of social policy, testing availability, and meteorological and demographic factors whose association with transmission we aim to determine. The model specification reported in Eqs. 3–5 representing the endemic, within-, and between-country component of the model, respectively, is the result of a model selection procedure based on the Akaike Information Criterion (AIC) (38). A summary of model comparison and selection process is presented in *SI Appendix, Table S1*. We began with an intercept-only model (model 1) with a population offset in the endemic component of the model and country's measure of connectivity based on a power law. More complicated versions of the epidemic component were evaluated by sequentially adding weather, demographic, stringency index, and testing policy in the formula for the within- (λ) and between-country (ϕ) components of the model. We also tested whether multiple lags for cases and

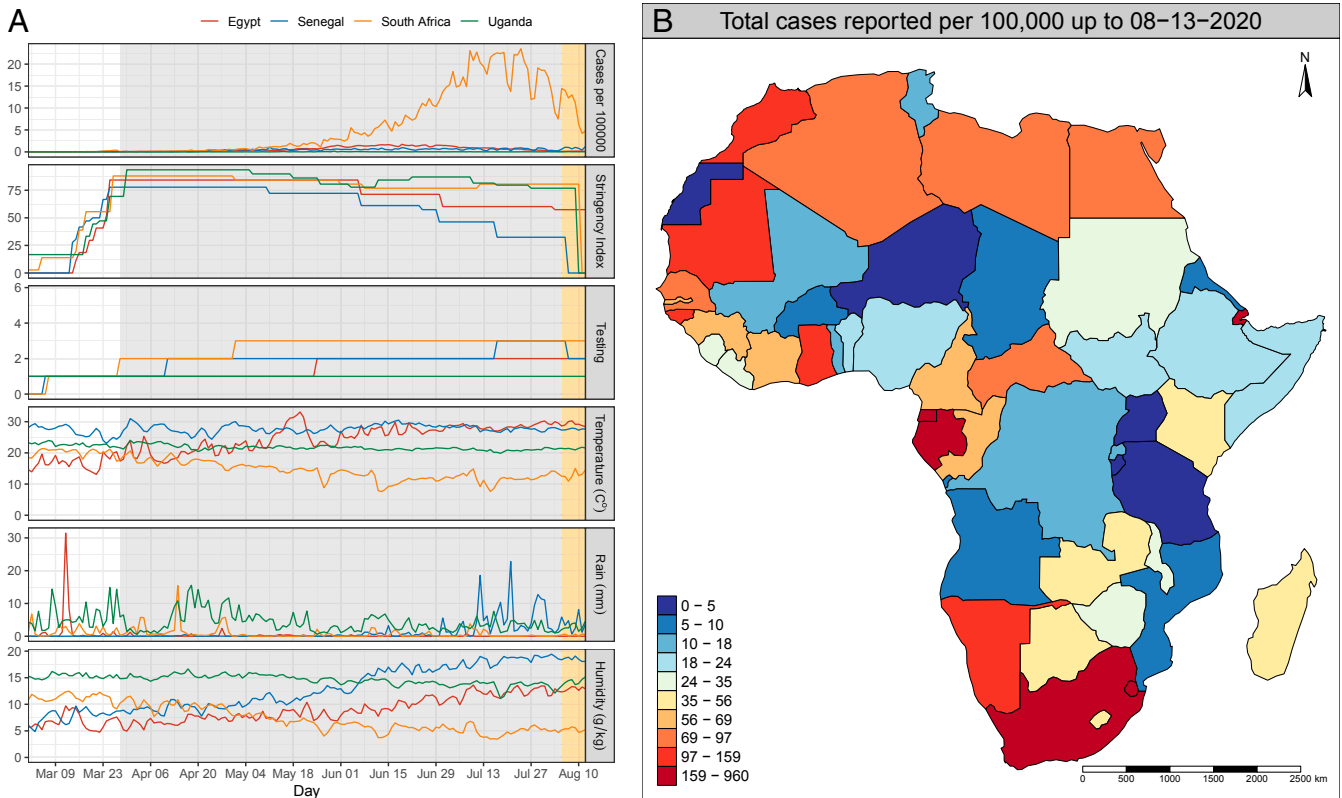


Fig. 1. Temporal distribution of reported cases, stringency index, testing policy, and weather factors. (A) Time series (February to August 2020) of (from top to bottom) daily reported cases per 100,000, stringency index, testing policy, temperature, rainfall, and specific humidity for representative countries Egypt, Senegal, South Africa, and Uganda. These time-dependent covariates were used as predictors (explanatory variables) in the best-fitting model shown in Table 1. The gray and orange shaded areas show the time window of data used to fit the model and the data held out for model validation, respectively. (B) Country-specific distribution of cumulative cases per 100,000 on August 13, 2020.

covariates better described the observed patterns: considering lags $d = 1, \dots, D$, where $D = 14$ days. Exploration of the higher-order model with Poisson and geometric lag weights revealed that, relative to the first-order model, the largest improvement in AIC ($\Delta AIC = -1,560$) was achieved with a model with $D = 7$ for all variables (SI Appendix, Fig. S10). Therefore, the model with lag 7 d for cases, testing policy, stringency index including Human Development Index (HDI), landlocked status, and population in the between-country component, and with meteorology factors, HDI, landlocked status, stringency index, and testing policy in the within-country component, yielded the lowest AIC (from 48,824 in model 1 to 48,437 in model 5 without random effects; SI Appendix, Table S1).

Due to the high spatiotemporal heterogeneity of reported cases across Africa and to better capture country-specific transmission dynamics and incidence levels not explained by observed covariates, we allowed the intercept (mean levels of λ and ϕ) in the local (4) and neighbor-driven (5) sources of infections to vary for each country as random effects. The relative risks for each explanatory variable included in this final model and the associated 95% CIs are reported in Table 1. Landlocked status, stringency index, and testing policy were significant contributing factors for the local transmission of cases. SI Appendix, Fig. S11 shows the actual contribution of the time-constant and time-varying covariates to the transmission parameters.

In addition, higher lagged mean temperature was a positive contributing factor, but higher specific humidity had a negative effect on the transmission of cases. For example, a 1 SD increase in the lagged mean temperature results in 11% higher contribution to the within-country transmission ($P = 0.023$, relative risk

[RR] 1.11, 95% CI 1.01 to 1.21). However, a 1 SD increase in the 7-d lag mean specific humidity resulted in a 14% lower contribution on the local transmission of cases ($P = 0.001$, RR 0.86, 95% CI 0.78 to 0.94). Greater accessibility of testing remained the only significant contributing factor explaining the numbers of cases from the neighboring countries. With each level increase in the openness of the testing policy from zero to four, the contribution to the transmission of cases from neighboring countries was higher by twofold ($P < 0.0001$). The overdispersion parameter decreased from the fixed effects model (1.95, 95% CI 1.87 to 2.03) to random effects model (1.70, 95% CI 1.63 to 1.78) as a sign that the random effects absorbed part of the unexplained variability between countries.

Fig. 2 shows country-specific random effects, which we used to capture unexplained contributions to transmission. A value higher (lower) than one means that a country has an average transmission rate that is higher (lower) than the rest of the continent. This may be interpreted as a country-specific propensity to generate more or fewer cases given the past number of reported infected individuals. With respect to the within-country contributions to transmission, South Africa and Djibouti are the only African countries with an effect significantly higher than one. On the other hand, the Republic of Congo is the only country with a within-country transmission rate significantly lower than the continental-level mean. With respect to the between-country contributions to transmission of cases, Benin, Cameroon, Central African Republic, Ethiopia, Gabon, Ghana, Guinea, Malawi, Republic of Congo, and Senegal had significantly higher rates of between-country transmission than the continental-level mean. Angola, Chad, Lesotho, Namibia, and

Table 1. Maximum likelihood estimates and corresponding 95% CIs for a model with a 7-d lag

Parameter	Final model		
	Relative risk	95% CI	P value
Endemic			
Intercept	11.071	(7.150, 17.142)	—
Within-country			
Intercept	0.958	(0.550, 1.669)	—
log(population)	1.036	(0.956, 1.123)	0.391
HDI	0.957	(0.819, 1.118)	0.580
Landlocked	0.674	(0.531, 0.857)	0.001
Stringency _{t-7}	1.872	(1.170, 3.000)	0.008
Testing _{t-7}	0.817	(0.729, 0.918)	0.001
Rain _{t-7}	1.045	(0.981, 1.112)	0.175
Temperature _{t-7}	1.106	(1.014, 1.206)	0.023
Humidity _{t-7}	0.856	(0.780, 0.940)	0.001
Between-country			
Intercept	0.045	(0.004, 0.481)	—
log(population)	1.428	(0.923, 2.210)	0.110
HDI	1.239	(0.584, 2.628)	0.576
Landlocked	0.844	(0.306, 2.323)	0.742
Stringency _{t-7}	1.630	(0.560, 4.749)	0.380
Testing _{t-7}	2.322	(1.676, 3.218)	<0.0001
ρ	2.186	(1.532, 2.839)	—
ψ	1.703	(1.631, 1.775)	—

For climatic variables, a 1 SD increase in climatic variables results in the shown relative risk. For stringency index, a 10% increase in stringency is associated with the increased relative risk shown. HDI and testing policy are on ordinal scales zero to three (HDI) and zero to four (testing policy). Bolded estimates are statistically significant. The spatial weight decay, ρ , reflects the strength of intercountry connectivity, and ψ , reflects the overdispersion parameter.

Tanzania had lower between-country transmission rates compared to the continental-level mean. The estimated variation of these country-specific effects in the within-country component of the model is small ($\hat{\sigma}_\lambda^2 = 0.07$) compared with their variation in the neighborhood component ($\hat{\sigma}_\phi^2 = 2.3$). Although the between-country variability of transmission resulting from cases reported outside of the country was larger, the between-country intercept (Table 1) is very small, and so the neighborhood component is, in general, a small contributor to the fit.

Contributions of Within- and Between-Country Transmission. We distinguish between endemic, within-country, and between-country contributions to the mean number of cases. Fitted values for all components according to model formulations in Eq. 1 are shown in Fig. 3, with a complete listing in *SI Appendix, Fig. S12*. The number of cases attributed to within- and between-country transmission of cases during the entire study period varied greatly. Across countries, the contribution from the endemic component was found to be minimal. Of the 46 countries analyzed, 16 of them are landlocked, and 13 (81%) of these had a substantial contribution of cases from their neighboring countries: Botswana, Burkina Faso, Burundi, Central African Republic, Ethiopia, Lesotho, Malawi, Rwanda, South Sudan, Swaziland, Uganda, Zambia, and Zimbabwe.

Short-Term Forecast. We keep the last 7 d of data out of the fitting procedure in order to use them as a forecast validation dataset. We produce 1-wk-ahead predictions and compare them with the reported data to check the quality of the model forecast. The results show that the majority of individual country

case count data are captured well within model prediction intervals (Fig. 4 and *SI Appendix, Fig. S13*). Across countries, the model predictive performance was assessed with a calibration test based on proper scoring rules as described in ref. 39. A map of *P* values for the calibration test is shown in *SI Appendix, Fig. S14*. Overall, model predictions are well calibrated, and a misalignment between forecast and observations was only detected for a few countries ($p < 0.05$, Burundi, Cameroon, Somalia, and Botswana).

Discussion

We present a COVID-19 surveillance strategy that can improve the ability of African countries to interpret the complex data available to them during the pandemic. This approach balances the simplicity and consequent robustness of an empirical model against the more complex, potentially more realistic but also more strongly assumption-driven kind of compartmental mechanistic model (40). A key feature of our approach is the ability to distinguish between case incidence arising from the local within- or neighbor-driven transmission of infection. Distinguishing within- and between-country transmission of cases allows us to identify potential strategies for social or health policy intervention. The model further enables reproducing the history of the epidemic in relationship to past policy, and producing short-term predictions of the dynamic evolution of the epidemic. Due to evolving global policy, complex weather phenomena, and changes to social behavior, these factors will change their relationship to COVID-19 prediction as the pandemic evolves.

We find that a country's testing capacity, social policy, landlocked status, temperature, and humidity are important contributing factors explaining the within- and between-country transmission of cases over the window of analysis. The availability of more testing to a wider swath of the populace is a potent contributor to reduced case transmission within country, while having the opposite effect on case transmission from neighboring countries. Testing policy, another surrogate for healthcare capability and preparedness to handle the pandemic, demonstrates this unique opposing effect on the two model components. Countries in northern and southern Africa that have relatively high HDI demonstrated comparatively higher numbers of cases per population. On the other hand, even in the face of border closures, landlocked countries depend on open borders for trade. For such countries, strict border closure measures are difficult to impose, resulting in a continual influx of cases from the neighboring countries.

The observed association of temperature and specific humidity with the case numbers, although small, points to the possible biological and behavioral responses to weather patterns, which, in turn, drive the dynamics of SARS-CoV-2 infection. Temperature and humidity are known factors in SARS-CoV, MERS-CoV, and influenza virus survival (41–43). Lower humidity has been consistently associated with a higher number of cases. Besides potentially prolonging half-life and viability of the virus, other potential mechanisms associated with low humidity include stabilization of the aerosol droplet, enhanced propagation in nasal mucosa, and impaired localized innate immunity (44). Whether the observed association is driven by the change in social behavioral patterns or the effect on the survival of SARS-CoV-2 remains to be explored (45). It is also possible that the observed contribution of meteorological factors to case transmission might be an artifact of spatial averaging and assigning one meteorological value to an entire country. It will be important to explore such associations in more detail before any policy-relevant conclusions can be drawn. Thus, at present, policy makers must focus on social-behavioral interventions such as reducing physical contact within communities and vaccination, while COVID-19 risk predictions based on climate information alone should be interpreted with caution (46).

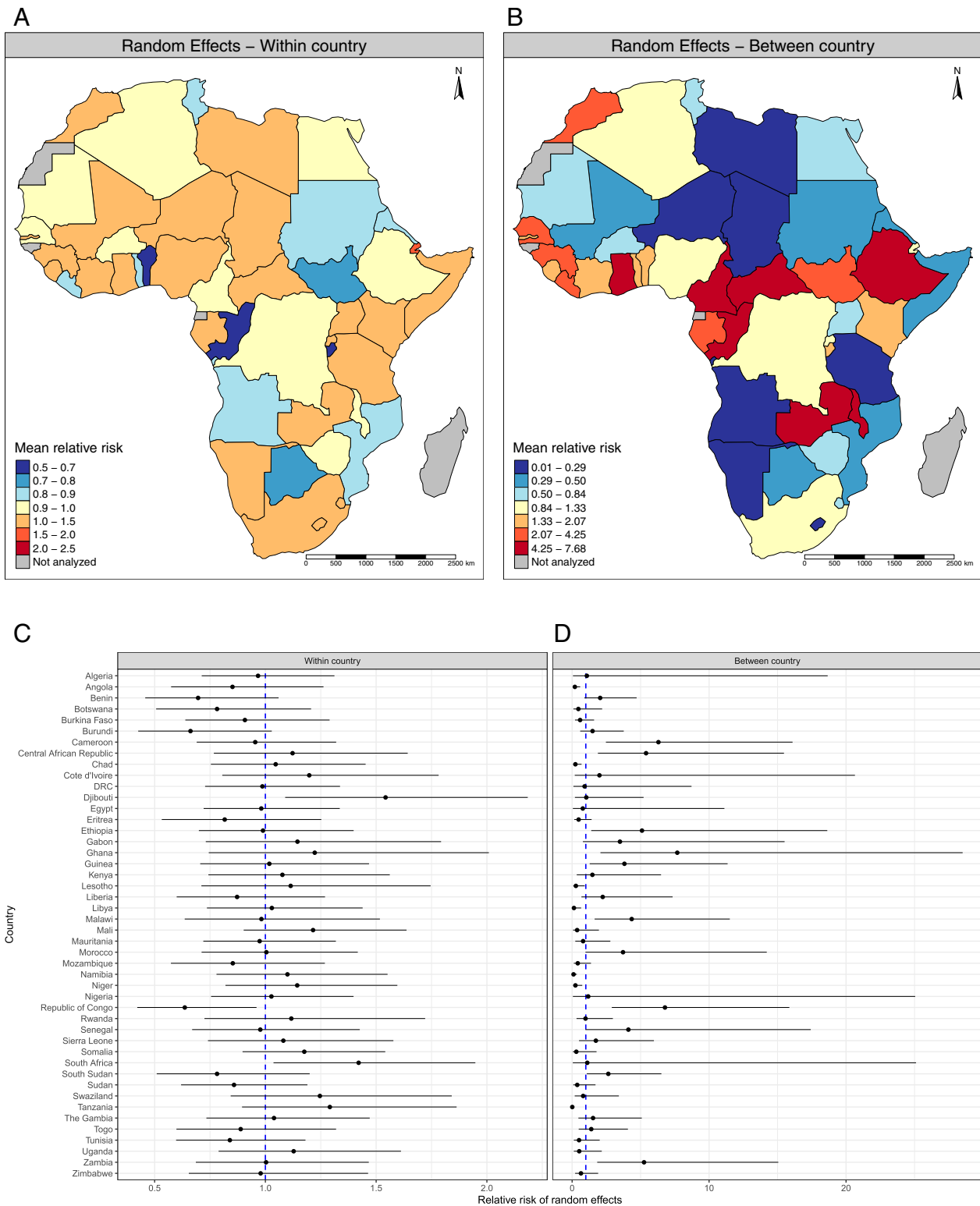


Fig. 2. Random effects for the local and neighbor components of the model. Country-specific RR and their 95% CI for (A and C) within-country and (B and D) between-country model contributions. The dashed blue line in the forest plots represents continent-level average. RR greater than one indicates higher propensity for transmission as compared to the rest of the continent.

Our infection surveillance tool adds to the public health capacity already in place on the continent to better understand transmission patterns between and within African countries. Containment and mitigation strategies to limit the spread of

the virus, including restrictions on movement, public gatherings, and schools, were implemented very early in the pandemic. In a resource-limited setting such as Africa, containment and mitigation strategies remain the most robust defense

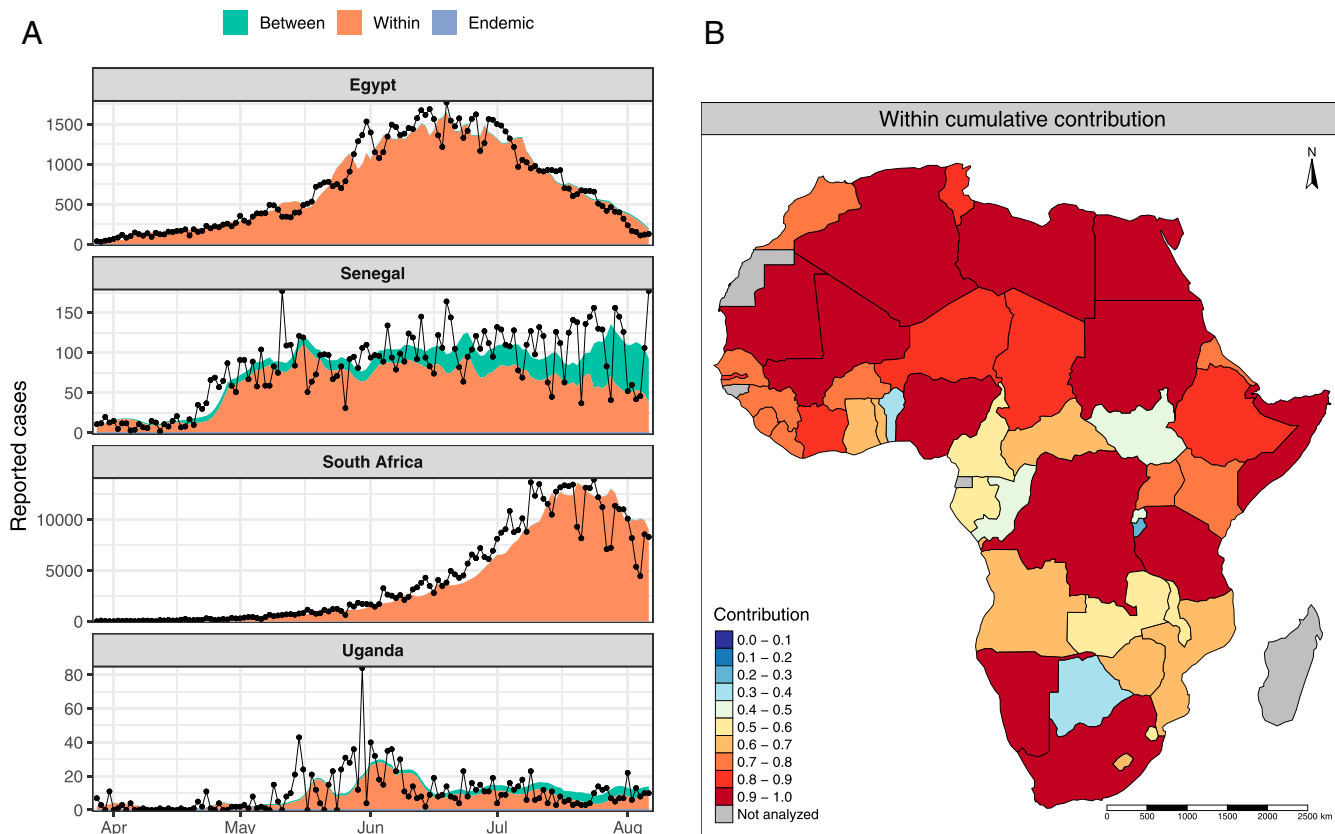


Fig. 3. Contributions of new cases from endemic, within-country, and between-country model components. (A) Black dots and connecting lines are the observations, and shaded colors are the model predictions, from the three contributing components of the model. (B) The relative contribution of new cases from within-country sources, by country. Contribution plots for all countries are given in *SI Appendix, Fig. S12*.

against high infection rates and mortality until effective vaccines are widely available. However, it is anticipated that physical distancing measures enforced to limit transmission will also restrict access to essential non-COVID-19 healthcare services, such as disruptions in the existing programs for tuberculosis, HIV/AIDS, malaria, and vaccine-preventable diseases, causing long-lasting collateral damage on the continent (32). Although between 29 million to 44 million individuals (2) in Africa were projected to become infected in the first year of the pandemic if containment measures fail, these numbers may be underestimates, since the proportion of asymptomatic infections is not well established. Since detection is biased toward clinically severe disease, the attack rate of the infection is probably substantially higher than what is reported. At the beginning of the pandemic, it was estimated that up to 86% of all infections were undocumented and were the source of 79% of the documented cases (47). Such observations explain the rapid geographic spread of the infections and challenging efforts at containment. The number of asymptomatic cases is best determined by population-based seroepidemiology data. However, due to the fragile healthcare systems of African countries, this type of disease surveillance remains limited. On the other hand, it is also plausible that the lower incidence rate of the virus in Africa is because of the investment in preparedness and response efforts toward various outbreaks on the continent (such as Ebola virus disease, Lassa fever, polio, measles, tuberculosis, and HIV) (32). This technical know-how has been swiftly adapted to COVID-19.

An additional strength of our modeling strategy is the ability to incorporate the disease-specific serial interval between sequential infections in the autoregressive model. We attempted to mimic the longer (greater than 1 d) serial interval (48, 49),

infectiousness (48, 50), and latency (48) of COVID-19 transmission, by extending the observational interval of the infectious process to several days. The Poisson autoregressive weighting method used in our modeling strategy also captures an initial increase in infectiousness and may thus be more appropriate for longer serial intervals or daily data. In their recent work, Bracher and Held (51) show that moving beyond 1-d lags to higher-order time lags improves predictive performance of these endemic-epidemic models. For our optimization scheme, we tested lags up to 14 d, and found that a lag at 7 d provided the best model fit. Short-term predictions enable the monitoring of case incidence trends but are limited by high levels of uncertainty. This is the result of the nonnegligible overdispersion detected in the data and due to the several sources of unmodeled spatial and temporal heterogeneity across the continent.

Limitations. A number of assumptions in our analysis reflect practical limitations. We used aggregated data at the country level to ensure continental coverage, but fitting this model at different spatial scales would require reformulation of the neighborhood structure and the model parameters. It is difficult to obtain accurate mobility data within and between countries throughout Africa, necessitating an indirect estimation of contact probabilities. Although our use of higher-order neighborhood (beyond sharing a border) contact patterns led to improved model fit compared to an assumption of first-order neighbors (bordering countries) only, more-accurate mobility data would improve these estimations. Because ground truth data on within-country vs. imported cases are not available, we cannot validate case origins in our model fits. Other assumptions related to the testing and stringency policies are coarse approximations of

Reported cases ● ○ Predicted mean — retrospective forecast 50% – 95% CI — retrospective forecast

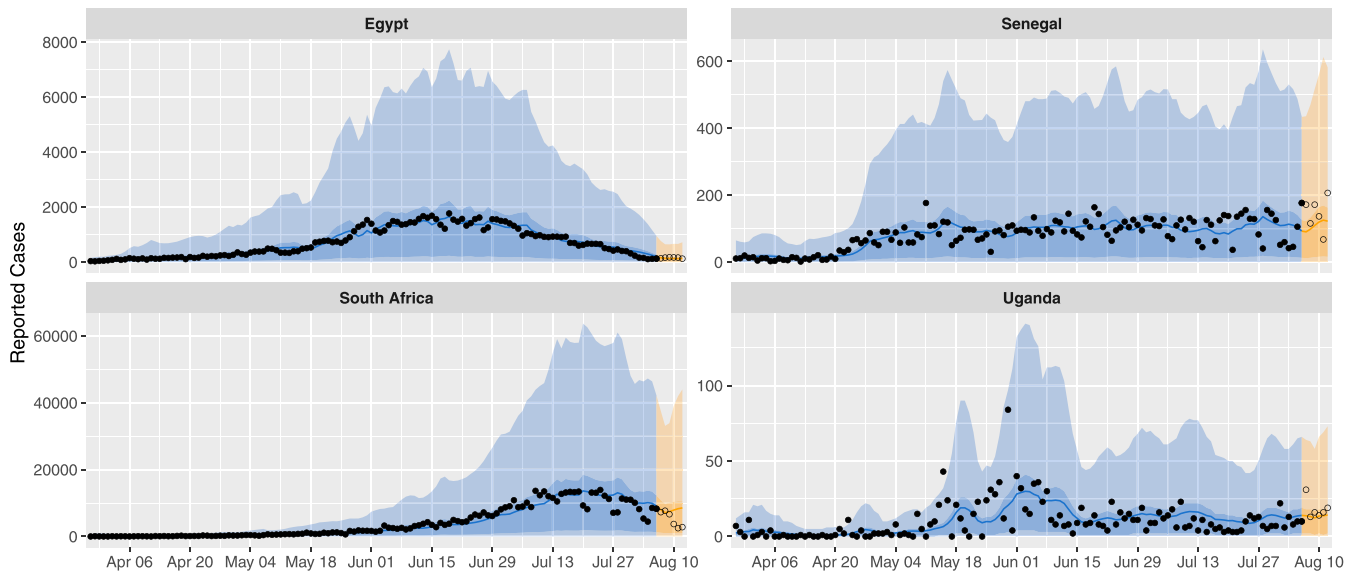


Fig. 4. Retrospective and forecast model fit for selected countries. Retrospective model fit (blue shade) and forecast (orange shade). The 50% and 95% CIs are represented by dark and light colors, respectively. Filled circles in the blue shade are observed cases, and the solid blue line is the predicted retrospective mean. Open circles in the orange shaded area represent cases during model forecast, and the solid orange line is the predicted forecast case mean. The majority of individual country case count data are captured well within model prediction intervals. For the full list of countries, see *SI Appendix, Fig. S13*.

governmental response to the surveillance and control of disease transmission. The absence of quantifiable tests per capita limits all such approaches. We excluded Equatorial Guinea, Guinea-Bissau, and Western Sahara, due to missing stringency index, and excluded the six island nations (Madagascar, Comoros, Mauritius, Seychelles, Cape Verde, São Tomé, and Príncipe), due to the lack of connectivity to mainland Africa which prevented model convergence.

SARS-CoV-2 is often carried by otherwise healthy-appearing individuals who unknowingly transmit the pathogen. In the present analysis, we could not disentangle asymptomatic and symptomatic disease. Underreporting can introduce artifacts in the autocorrelation structure and may confound the estimation of lag weights of the underlying serial interval distribution (52).

Additionally, we assumed that model coefficients were constant over time. This represents a trade-off of sufficient data required for stable model fitting versus the need to embrace the nonstationarity of the evolving pandemic. The empirical model approach we employed depends only on case data and therefore does not change as infectivity changes with viral variants, reinfection, and vaccination. We further ignored seasonal variation, implying that the interaction with weather is the same in summer and winter, or rainy and dry seasons closer to the equator. As the pandemic extends beyond the first year, introducing additional covariates to the endemic term, such as seasonal oscillations or animal reservoirs, may improve model fit.

This modeling strategy is limited by the quality of the data and the lack of nonlinear dynamics in the model. As effective vaccines become more widely available within Africa, modification of this model framework, and validation against observations and predictions, will become important to consider. Our modeling strategy enables us to focus on any window of time during an epidemic to examine the contributing properties. As the African pandemic and policies evolve, and as the variants of the virus introduce changes in infectivity, this model should be refit for later time periods to establish the changing relationships that best account for the epidemic dynamics within and between

countries. Our archived open-source code enables others to refit for any past or future time window of interest.

Conclusions. We present a pan-African COVID-19 surveillance tool to track and perform short-term forecast of COVID-19 cases and to quantify between- and within-country sources. Our analyses give insight into the sociodemographic, geodemographic, testing, mitigation/containment, and meteorological factors that influence the spread of the SARS-CoV-2 infection at the national scale. Although our strategy can be used for short-term predictions of cases, its accuracy is dependent upon the quality of testing and reported data. In settings with fragile health systems, coupled with the vulnerability of lower-HDI economies, the capacity to effectively track the pandemic is especially challenging. Such challenges point to the potential advantages in regional efforts to coordinate resources to test and report cases. Seeking equitable behavioral and social interventions, balanced with coordinated country-specific strategies in infection suppression, should be a continental priority to control the COVID-19 pandemic in Africa.

Materials and Methods

Overview. Our analyses included 46 countries of mainland Africa. We do not provide estimates for Equatorial Guinea, Guinea-Bissau, and Western Sahara, due to the missing data on stringency index, or the six island nations (Madagascar, Comoros, Mauritius, Seychelles, Cape Verde, São Tomé, and Príncipe), due to the lack of spatial connectivity. Modeling the spread of COVID-19 over the African continent poses challenges, given the extensive cultural, political, and environmental heterogeneity between countries. Indeed, this heterogeneity results in substantial variability of reported case counts across countries. It is this variability in case counts that motivates our choice of a relatively simple data-driven autoregressive modeling approach. Such a modeling approach focuses on the interaction of cases reported in time and space without hidden variables to be estimated.

Meteorology/Weather Factors. The seasonality of influenza transmission has been associated with cycles of temperature, rainfall, and specific humidity, although, in different regions of the world, transmission may peak during

the “cold-dry” season (temperate climates) or during “humid-rainy” season (tropical climates) (53).

We estimated the influence of meteorological factors on the transmission dynamics of COVID-19 in Africa. Real-time, daily, in situ synoptic weather observations are sparse across much of Africa. Therefore, daily, 10-km spatial resolution mean temperature, rainfall, and specific humidity data were obtained from UK Met-Office numerical weather prediction model output (54, 55). These data are extracted from the early time steps of the model following data assimilation, to more closely approximate an observational dataset. This approach also has the advantage that future studies have access to the same coherent dataset at a global scale for applications outside of continental Africa. The weather product that generates these data closely approximates an observational dataset at locations that have dense observation coverage, whereas, in observation-sparse areas, the dataset relies more heavily upon the numerical weather prediction model (a physics-based rather than statistical model). The meteorological dataset contained no missing data.

A population density-weighted spatial average (SI Appendix, Figs. S7–S9) was then applied for each day and country using the R package “exacttract” (56). Population density was obtained from the Gridded Population of the World version 4 (GPWv4) from the Centre for International Earth Science Information Network (57). Weighting climate variables by population gives a closer approximation to the weather conditions faced by humans living in that country compared to an unweighted average over total land area. For example, the country of Algeria, in which much of the population resides along the coast, demonstrates a cooler, wetter, and more humid climate when weighting by population (SI Appendix, Figs. S7E, S8E, and S9E).

Stringency Index and Testing Policy. To include an aggregate measure of countries’ social policies, the stringency index sourced from the OxCGRT dataset was used. This composite measure reflects government policies related to school and workplace closures, restrictions on public gatherings, events, public transportation, limitations of local and global travel, stay-at-home orders, and public education campaigns (23). A full description of these variables is provided in SI Appendix, Table S2. The stringency index is calculated from these categorical variables using a weighted average, with a range of 0 to 100 indicating weak to strict stringency measures, respectively. A time-dependent metric of testing policy was also extracted from this dataset. Ranging from zero to four, this categorical metric increases with more open and comprehensive testing policy.

HDI, Demography, United Nations Geographic Regions, and Coastline Access. In our modeling strategy, we incorporate key socioeconomic and sociodemographic epidemiological data, including HDI, population, United Nations geographic regions, and coastline access (SI Appendix, Fig. S15). HDI represents the national data on key aspects of development, namely, education, economy, and health (58). The HDI is the geometric mean of normalized indices for each of the three dimensions. The education dimension is measured by average years of schooling for adults aged 25 y and more and expected years of schooling for children of school-entering age. The economy dimension is measured by gross national income per capita, and the health dimension is assessed by life expectancy at birth. In Africa, the majority of the countries fall in the low-HDI category (SI Appendix, Fig. S15). The northern part of Africa and South Africa have a considerably higher HDI compared to the rest of the continent. Country-specific median age was correlated with HDI (Pearson’s correlation coefficient $R = 0.71$, $P < 0.0001$; SI Appendix, Fig. S16); therefore, we excluded this covariate from the model. We include in the model the 2020 population obtained from the Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat (59). The categorization of sub-Saharan and northern Africa was based on the United Nations geoscheme for Africa (60). This regional factor captures the human genetics (61), environment and climate (62), and sociocultural and sociodemographic variations of the African population (63). Finally, lack of direct access to the coastline may influence the flow of infections from neighboring countries, as border trade remains an essential operation. For example, Uganda introduced border closures and tighter preventive measures on truck drivers’ movements during the epidemic; despite this, a substantial number of new infections have been imported from truck drivers crossing the border for trade (64). Such cross-border commerce remains a crucial part of the supply chain for landlocked African countries such as Uganda and Rwanda.

Model Formulation. We chose a class of multivariate time series models for case count data introduced by Held et al. (65), and further extended by Bracher and Held (51) with the addition of higher-order distributed lags.

Conditional on past observations $Y_{i,t-d}$, $i = 1, \dots, N$, and $d = 1, \dots, D$, new COVID-19 cases Y_{it} from country i at time t are assumed to follow a negative binomial distribution with mean μ_{it} and overdispersion parameter ψ as

$$[Y_{it} | Y_{t-1}, \dots, Y_{t-D}] \sim \text{NegBin}(\mu_{it}, \psi).$$

The conditional variance is $\mu_{it} + \psi \mu_{it}^2$, which demonstrates the role of the overdispersion parameter to capture variability greater than the mean. The conditional mean μ_{it} is decomposed into three additive components,

$$\mu_{it} = \epsilon_i + \lambda_{it} \sum_{d=1}^D u_d Y_{i,t-d} + \phi_{it} \sum_{d=1}^D \sum_{j \neq i} u_d w_{ji} Y_{j,t-d}, \quad [1]$$

where ϵ_i , λ_{it} , and ϕ_{it} represent three contributions to case incidence. The first term, ϵ_i , is the so-called endemic component and captures infections arising from sources other than past observed cases (e.g., contributions from areas that are not included in the neighbor set). The two other terms in [1], λ_{it} and ϕ_{it} , constitute the epidemic part of the model and modulate how infective individuals reported in the past d days both locally and from neighboring countries will contribute to the average future number of reported cases. The strength of connection between countries is described by spatial weights w_{ji} . This intercountry transmission susceptibility is defined using a power-law formulation proposed by Meyer and Held (66),

$$w_{ji} = o_{ji}^{-\rho}, \quad [2]$$

where o_{ji} is the path distance between countries j and i (with $o_{ii} = 0$, $o_{ij} = 1$ for direct neighbors i and j and so on), and ρ is a decay parameter to be estimated from the data. The path distance o_{ji} is on an ordinal scale based upon the adjacency index. The spatial weights are normalized such that $\sum_k w_{jk} = 1$ for all rows j of the weight matrix (SI Appendix, Fig. S17).

The normalized autoregressive weights u_d are shared between the local and global epidemic components, and represent the probability for a serial interval of up to D days—which is the average time in days between symptom onset in an infectious individual (or primary case) and symptoms appearing in a newly infected individual (or secondary case) when both are in close contact (67).

The parameters ϵ_i , λ_{it} , and ϕ_{it} are constrained to be nonnegative and modeled as the natural log-transformed linear combination of different country-specific covariates. The endemic component,

$$\log(\epsilon_{it}) = \alpha^{(\epsilon)} + \log(N_i), \quad [3]$$

is decomposed as a constant $\alpha^{(\epsilon)}$ specific to the baseline endemic and a term proportional to the country-level population N_i . In the epidemic part of the model, we expect new cases to also be driven by country-specific factors: Population (N_i), HDI classifications of low, medium, or high ($HDI_i = \{0, 1, 2\}$), and land-locked (LL_i) status for each country are assumed constant over the time scale of analysis. Other forces driving new cases vary over time, as a response to either policy changes or natural fluctuation in environmental or societal patterns. Time-dependent covariates include mean daily temperature ($T_{i,t-\tau}$), rainfall ($R_{i,t-\tau}$), specific humidity ($H_{i,t-\tau}$), testing policy ($X_{i,t-\tau}$), and government stringency index ($S_{i,t-\tau}$), lagged at τ days. The full set of explanatory variables that contribute to the model from both internal and external epidemic components are formalized in [4] and [5] as

$$\begin{aligned} \log(\lambda_{it}) = & \alpha_i^{(\lambda)} + \beta^{(\lambda)} \log(N_i) + \gamma^{(\lambda)} HDI_i + \delta^{(\lambda)} LL_i + \\ & \sigma^{(\lambda)} S_{i,t-\tau} + \chi^{(\lambda)} X_{i,t-\tau} + \\ & \theta^{(\lambda)} T_{i,t-\tau} + \omega^{(\lambda)} R_{i,t-\tau} + \nu^{(\lambda)} H_{i,t-\tau} \end{aligned} \quad [4]$$

and

$$\begin{aligned} \log(\phi_{it}) = & \alpha_i^{(\phi)} + \beta^{(\phi)} \log(N_i) + \gamma^{(\phi)} HDI_i + \delta^{(\phi)} LL_i + \\ & \sigma^{(\phi)} S_{i,t-\tau} + \chi^{(\phi)} X_{i,t-\tau}, \end{aligned} \quad [5]$$

where $\alpha_i^{(\lambda)} \approx N(\alpha_0^{(\lambda)}, \sigma_\lambda^2)$ and $\alpha_i^{(\phi)} \approx N(\alpha_0^{(\phi)}, \sigma_\phi^2)$ are a set of independent country-level random effects. This modeling framework is implemented in the R package “surveillance” (68). A complete table of data sources for model input is found in SI Appendix, Table S3.

Model Fitting. We selected models based on AIC (69) if random effects were not present (*SI Appendix, Table S1*). To compare models that included random effects, we used proper scoring rules for count data (70). Scoring rules are functions $S(P, y)$ that evaluate the accuracy of a predictive distribution P against an outcome y that was observed. We chose the model with the lowest AIC or with the lowest logarithmic score computed as minus the logarithm of the predictive distribution evaluated at the observed count. We began with the first-order autoregressive modeling ($D = 1$ in [1]) of daily COVID-19 incidence using intercept-only model population offset and country connectivity. In a mechanistic interpretation of such a first-order model, the time between the appearance of symptoms in successive generations is assumed to be fixed to the observation interval at which the data are collected, here as 1 d (52).

After the estimation and illustration of this basic model, we expand the model by sequentially adding the following additional covariates: country-specific HDI, population both within country and in neighboring countries, meteorology factors, stringency index, testing policy, landlocked status, and random effects to more fully account for unobserved heterogeneity of the cases. Social policies and meteorological data were included in the model, testing for fit at different lags (for example, $T_{i,t-\tau}$, $\tau \in 0, 7, 14$ d).

1. P. G. T. Walker *et al.*, The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. *Science* **369**, 413–422 (2020).
2. J. W. Cabore *et al.*, The potential effects of widespread community transmission of SARS-CoV-2 infection in the World Health Organization African region: A predictive model. *BMJ Glob. Health* **5**, e002647 (2020).
3. B. Z. Diop, M. Ngom, C. P. Biyong, J. N. P. Biyong, The relatively young and rural population may limit the spread and severity of COVID-19 in Africa: A modelling study. *BMJ Glob. Health* **5**, e002699 (2020).
4. M. Senghore, M. K. Savi, B. Gnanon, W. P. Hanage, I. N. Okeke, Leveraging Africa's preparedness towards the next phase of the COVID-19 pandemic. *Lancet Glob. Health* **8**, e884–e885 (2020).
5. A. Clark *et al.*, Global, regional, and national estimates of the population at increased risk of severe COVID-19 due to underlying health conditions in 2020: A modelling study. *Lancet Glob. Health* **8**, e1003–e1017 (2020).
6. J. R. Gutman *et al.*, Malaria and parasitic neglected tropical diseases: Potential syndemics with COVID-19? *Am. J. Trop. Med. Hyg.* **103**, 572–577 (2020).
7. A. B. Hogan *et al.*, Potential impact of the COVID-19 pandemic on HIV, tuberculosis, and malaria in low-income and middle-income countries: A modelling study. *Lancet Glob. Health* **8**, e1132–e1141 (2020).
8. R. Ridwan, COVID-19 and dengue: A deadly duo. *Trop. Doct.* **50**, 270–272 (2020).
9. J. B. Nachega *et al.*, Responding to the challenge of the dual COVID-19 and ebola epidemics in the democratic republic of Congo—Priorities for achieving control. *Am. J. Trop. Med. Hyg.* **103**, 597–602 (2020).
10. J. Ditekemena, S. Doumbia, S. H. Ebrahim, COVID-19's final frontier: The central Africa region. *Trav. Med. Infect. Dis.* **37**, 101694 (2020).
11. A. I. Barreca, J. P. Shimshack, Absolute humidity, temperature, and influenza mortality: 30 years of county-level evidence from the United States. *Am. J. Epidemiol.* **176** (suppl..7), S114–S122 (2012).
12. J. Shaman, V. E. Pitzer, C. Viboud, B. T. Grenfell, M. Lipsitch, Absolute humidity and the seasonal onset of influenza in the continental United States. *PLoS Biol.* **8**, e1000316 (2010).
13. T. Žuk, F. Rakowski, J. P. Radomski, Probabilistic model of influenza virus transmissibility at various temperature and humidity conditions. *Comput. Biol. Chem.* **33**, 339–343 (2009).
14. J. Shaman, M. Kohn, Absolute humidity modulates influenza survival, transmission, and seasonality. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 3243–3248 (2009).
15. J. Yuan *et al.*, A climatologic investigation of the SARS-COV outbreak in Beijing, China. *Am. J. Infect. Contr.* **34**, 234–236 (2006).
16. C. Yip, W. L. Chang, K. H. Yeung, I. T. S. Yu, Possible meteorological influence on the severe acute respiratory syndrome (SARS) community outbreak at amoy gardens, Hong Kong. *J. Environ. Health* **70**, 39–47 (2007).
17. P. Bi, J. Wang, J. E. H. Weather, Driving force behind the transmission of severe acute respiratory syndrome in China? *Intern. Med. J.* **37**, 550–554 (2007).
18. E. G. Gardner *et al.*, A case-crossover analysis of the impact of weather on primary cases of Middle East respiratory syndrome. *BMC Infect. Dis.* **19**, 1–10 (2019).
19. A. Altamimi, A. E. Ahmed, Climate factors and incidence of Middle East respiratory syndrome coronavirus. *J. Infect. Pub. Health* **13**, 704–708 (2020).
20. P. Mecenás, R. Bastos, A. C. R. Vallinoto, D. Normando, Effects of temperature and humidity on the spread of COVID-19: A systematic review. *PLoS One* **15**, e0238339 (2020).
21. W. T. Siow, M. F. Liew, B. R. Shrestha, F. Muchtar, K. C. See, Managing COVID-19 in resource-limited settings: Critical care considerations. *Crit. Care* **24**, 167 (2020).
22. A. Meyer, R. Sadler, C. Faverjon, A. R. Cameron, M. Bannister-Tyrrell, Evidence that higher temperatures are associated with a marginally lower incidence of COVID-19 cases. *Front. Pub. Health* **8**, 367 (2020).
23. T. Hale *et al.*, A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nat. Hum. Behav.* **5**, 529–538 (2021).

Model Predictions. As in previous work by Held and Meyer (71), we use plug-in forecasts: forecast from the fitted model without carrying forward the uncertainty in the parameter estimates. We assess both the model fit and 1-wk-ahead forecast of the higher-order autoregressive model with the logarithmic score. The smaller the score, the better the predictive quality (72, 73). Mean scores were generated for each country's forecast, by averaging the log-score obtained for each day of the validation week.

Data Availability. All code and data are available, to both replicate the results in this paper and enable users to examine past and future time windows of interest, in *SI Appendix* and posted online at GitHub, <https://github.com/Schiff-Lab/COVID19-HHH4-Africa>.

ACKNOWLEDGMENTS. We thank the Ugandan Ministry of Health, and the National Planning Authority for providing Uganda testing data, and R. Challan and M. Ferrari for helpful discussions. We also thank the team from the UK Met Office Informatics lab for their advice and support. This work was supported by a US NIH Director's Transformative Award 1R01AI145057 (S.J.S.). M.M.N. was supported, in part, by the Covid Virus Seed Fund from the Institute for Computational and Data Sciences and the Huck Institute at The Pennsylvania State University.

24. A. H. M. B. Hussain, Strict policy interventions increase social distancing during COVID-19 pandemic: A difference-in-differences analysis. ResearchGate [Preprint] (2020). https://www.researchgate.net/publication/341193128.Strict_policy_interventions.increase.social.distancing.during.Covid-19.pandemic.A.difference-in-differences.analysis (Accessed 12 June 2021).
25. J. R. Orellana, C. A. Martelo, School closure, mobility and COVID-19: International evidence. https://harris.uchicago.edu/files/orellanamartelo_internationalmobility.pdf. Accessed 12 June 2021.
26. C. T. Leffler *et al.*, Association of country-wide coronavirus mortality with demographics, testing, lockdowns, and public wearing of masks. *Am. J. Trop. Med. Hyg.* **103**, 2400–2411 (2020).
27. A. Moustakas, Ranking the explanatory power of factors associated with worldwide new COVID-19 cases. arXiv [Preprint] (2020). <https://arxiv.org/abs/2006.00971> (Accessed 12 June 2021).
28. G. Castex, E. Dechter, M. Lorca, COVID-19: Cross-country heterogeneity in effectiveness of non-pharmaceutical interventions. *Cent. for Econ. Policy Res. COVID Econ.* **14**, 175–199 (2020).
29. Y. T. Utsunomiya *et al.*, Growth rate and acceleration analysis of the COVID-19 pandemic reveals the effect of public health measures in real time. *Front. Med.* **7**, 247 (2020).
30. B. Gaye *et al.*, Socio-demographic and epidemiological consideration of Africa's COVID-19 response: What is the possible pandemic course? *Nat. Med.* **26**, 996–999 (2020).
31. T. Divala, R. M. Burke, L. Ndeketa, E. L. Corbett, P. MacPherson, Africa faces difficult choices in responding to COVID-19. *Lancet* **395**, 1611 (2020).
32. M. M. Loembé *et al.*, COVID-19 in Africa: The spread and response. *Nat. Med.* **26**, 999–1003 (2020).
33. S. Meyer, L. Held, M. Höhle, Spatio-temporal analysis of epidemic phenomena using the R package surveillance. *J. Stat. Softw.* **77**, 1–55 (2017).
34. M. M. Dickson, G. Espa, D. Giuliani, F. Santi, L. Savadori, Assessing the effect of containment measures on the spatio-temporal dynamic of COVID-19 in Italy. *Nonlinear Dyn.* **8**, 1–14, 2020.
35. M. Bekker-Nielsen Dunbar, L. Held, Endemic-epidemic framework used in covid-19 modelling (Discussion on the paper by nunes, caetano, antunes and dias). *Revstat. Stat. J.* **18**, 565–574 (2020).
36. C. Fronterre *et al.*, COVID-19 in England: Spatial patterns and regional outbreaks. medRxiv [Preprint] (2020). <https://doi.org/10.1101/2020.05.15.20102715> (Accessed 12 June 2021).
37. C. Bauer *et al.*, Stratified space–time infectious disease modelling, with an application to hand, foot and mouth disease in China. *J. Roy. Stat. Soc. C* **67**, 1379–1398 (2018).
38. H. Akaike, A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**, 716–723 (1974).
39. W. Wei, L. Held, Calibration tests for count data. *Test* **23**, 787–805 (2014).
40. M. Paul, L. Held, A. M. Toschke, Multivariate modelling of infectious disease surveillance data. *Stat. Med.* **27**, 6250–6267 (2008).
41. K.-H. Chan, J. S. M. Peiris, S. Y. Lam, L. L. M. Poon, K. Y. Yuen, W. H. Seto, The effects of temperature and relative humidity on the viability of the SARS coronavirus. *Adv. Virol.* **2011**, 1–7 (2011).
42. L. M. Casanova, S. Jeon, W. A. Rutala, D. J. Weber, M. D. Sobsey, Effects of air temperature and relative humidity on coronavirus survival on surfaces. *Appl. Environ. Microbiol.* **76**, 2712–2717 (2010).
43. W. Yang, L. C. Marr, Dynamics of airborne influenza viruses indoors and dependence on humidity. *PLoS One* **6**, e21481 (2011).
44. E. Kudo *et al.*, Low ambient humidity impairs barrier function and innate resistance against influenza infection. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 10905–10910 (2019).
45. W. Wei *et al.*, Regional ambient temperature is associated with human personality. *Nat. Hum. Behav.* **1**, 890–895 (2017).

46. C. J. Carlson, A. C. R. Gomez, S. Bansal, S. J. Ryan, Misconceptions about weather and seasonality must not misguide COVID-19 response. *Nat. Commun.* **11**, 4312 (2020).
47. R. Li *et al.*, Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-COV-2). *Science* **368**, 489–493 (2020).
48. X. He *et al.*, Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675 (2020).
49. S. T. Ali *et al.*, Serial interval of SARS-COV-2 was shortened over time by nonpharmaceutical interventions. *Science* **369**, 1106–1109 (2020).
50. F. Zhou *et al.*, Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study. *Lancet* **395**, 1054–1062 (2020).
51. J. Bracher, L. Held, Endemic-epidemic models with discrete-time serial interval distributions for infectious disease prediction. *Int. J. Forecast.*, <https://doi.org/10.1016/j.ijforecast.2020.07.002> (2020).
52. J. Bracher, L. Held, A marginal moment matching approach for fitting endemic-epidemic models to underreported disease surveillance counts. *Biometrics*, <https://doi.org/10.1111/biom.13371> (2020).
53. J. D. Tamerius *et al.*, Environmental predictors of seasonal influenza epidemics across temperate and tropical climates. *PLoS Pathog.* **9**, e1003194 (2013).
54. UK Met Office, Met Office COVID-19 response dataset. <https://metdata.blob.core.windows.net/covid19-response-non-commercial/README.html>. Accessed 12 June 2021.
55. D. Walters *et al.*, The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations. *Geosci. Model Dev. (GMD)* **12**, 1909–1963 (2019).
56. D. Baston, Fast extraction from raster datasets using polygons. Version 0.6.1. <https://isciences.gitlab.io/exactextract/>. Accessed 12 June 2021.
57. Center for International Earth Science Information Network, *C Documentation for the Gridded Population of the World, Version 4 (GPWv4), Revision 11 Data Sets* (NASA Socioeconomic Data and Applications Center, Palisades, NY, 2020).
58. United Nations Development Programme, Human development reports database. hdr.undp.org/en/data. Accessed 12 June 2021.
59. United Nations, World population prospects 2019. <https://population.un.org/wpp/>. Accessed 12 June 2021.
60. United Nations, Statistics division. <https://unstats.un.org/unsd/methodology/m49/>. Accessed 12 June 2021.
61. A. Sanchez-Mazas, E. S. Poloni, Genetic diversity in Africa. *eLS*, <https://doi.org/10.1002/9780470015902.a0020800> (2008).
62. N. Thomas, S. Nigam, Twentieth-century climate change over Africa: Seasonal hydroclimate trends and Sahara Desert expansion. *J. Clim.* **31**, 3349–3370 (2018).
63. W. Brass, *Demography of Tropical Africa* (Princeton University Press, 2015).
64. E. Nakkazi, Obstacles to COVID-19 control in East Africa. *Lancet Infect. Dis.* **20**, 660 (2020).
65. L. Held, M. Höhle, M. Hofmann, A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Stat. Model. Int. J.* **5**, 187–199 (2005).
66. S. Meyer, L. Held, Power-law models for infectious disease spread. *Ann. Appl. Stat.* **8**, 1612–1639 (2014).
67. H. Nishiura, N. M. Linton, A. R. Akhmetzhanov, Serial interval of novel coronavirus (COVID-19) infections. *Int. J. Infect. Dis.* **93**, 284–286 (2020).
68. S. Meyer, L. Held, M. Höhle, Spatio-temporal analysis of epidemic phenomena using the R package surveillance. *J. Stat. Softw. Art.* **77**, 1–55 (2017).
69. S. Portet, A primer on model selection using the Akaike Information Criterion. *Infect. Dis. Model.* **5**, 111–128 (2020).
70. C. Czado, T. Gneiting, L. Held, Predictive model assessment for count data. *Biometrics* **65**, 1254–1261 (2009).
71. L. Held, S. Meyer, “Forecasting based on surveillance data” in *Handbook of Infectious Disease Data Analysis*, L. Held, N. Hens, P. O’Neill, J. Wallinga, Eds. (Chapman & Hall, London, UK, 2019).
72. C. Czado, T. Gneiting, L. Held, Predictive model assessment for count data. *Biometrics* **65**, 1254–1261 (2009).
73. M. Paul, L. Held, Predictive assessment of a non-linear random effects model for multivariate time series of infectious disease counts. *Stat. Med.* **30**, 1118–1136 (2011).