Global environmental drivers of influenza

Ethan R. Deyle[a], M. Cyrus Maher[b,c], Ryan D. Hernandez[d], Sanjay Basu[e], and George Sugihara[a,1]

[a]Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093; [b]Human Longevity Inc., Mountain View, CA 94041; [c]Department of Epidemiology and Translational Science, University of California, San Francisco, CA 94158; [d]Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA 94158; and [e]Stanford Prevention Research Center, Stanford University School of Medicine, Stanford, CA 94305

Edited by Alan Hastings, University of California, Davis, CA, and approved September 16, 2016 (received for review May 13, 2016)

In temperate countries, influenza outbreaks are well correlated to seasonal changes in temperature and absolute humidity. However, tropical countries have much weaker annual climate cycles, and outbreaks show less seasonality and are more difficult to explain with environmental correlations. Here, we use convergent cross mapping, a robust test for causality that does not require correlation, to test alternative hypotheses about the global environmental drivers of influenza outbreaks from country-level epidemic time series. By moving beyond correlation, we show that despite the apparent differences in outbreak patterns between temperate and tropical countries, absolute humidity and, to a lesser extent, temperature drive influenza outbreaks globally. We also find a hypothesized U-shaped relationship between absolute humidity and influenza that is predicted by theory and experiment, but hitherto has not been documented at the population level. The balance between positive and negative effects of absolute humidity appears to be mediated by temperature, and the analysis reveals a key threshold around 75 °F. The results indicate a unified explanation for environmental drivers of influenza that applies globally.

epidemiology | empirical dynamic modeling | nonlinear dynamics | state-dependence | physical–biological coupling

A diverse group of drivers and mechanisms has been put forward to explain the wintertime occurrence of seasonal influenza outbreaks. Laboratory experiments show that relative humidity controls droplet size and aerosol transmission rates (1). Experiments with mammalian models showed that viral shedding by hosts increases at low temperature (2). Strong laboratory evidence has emerged that absolute humidity has a controlling effect on airborne influenza transmission (3).

Nevertheless, questions remain as to how these potential causal agents are expressed at the population level as epidemic control variables. At the population level, environmental factors covary, multiple mechanisms can coact, and infection dynamics are influenced by many other important processes (4), such as human crowding, rapid viral evolution, and international travel patterns. Perhaps not surprisingly, statistical analyses of population level data have produced contradictory results. Although correlations between influenza incidence and both temperature and absolute humidity are easy to find in temperate countries (5) and individual US states (6), such associations are weak or altogether absent in data from tropical countries (5).

Here we use new methods appropriate for disease dynamics to identify the causal drivers of influenza acting at the population level. Using time series data across countries and latitudes, we find that absolute humidity drives influenza across latitudes, and that this effect is modulated by temperature. At low temperatures, absolute humidity negatively affects influenza incidence (drier conditions improve survival of the influenza virus when it is cold), but at high temperatures, absolute humidity positively affects influenza (wetter conditions improve survival of the influenza virus when it is warm). This population-level finding supports the conclusions of theoretical work on viral envelope stability (7) and sorts out disagreements in laboratory studies (8).

Our analysis is divided into three parts. First, we show how dynamic resonance can explain the success of linear statistical methods in detecting causal effects in temperate latitudes and how mirage correlations are symptoms of their failure in the tropics. This motivates taking an empirical dynamic modeling (EDM) approach (9). Second, we use the EDM method of detecting causality, convergent cross-mapping (CCM), to show how absolute humidity and temperature combine to produce a unified explanation for influenza outbreaks that applies across latitudes. Finally, we probe the mechanistic effects of each variable, showing that absolute humidity has the most direct effect on influenza, and that this effect is modulated nonlinearly by temperature.

Results

Correlation and Seasonality. It is well known that correlative approaches can fail to provide an accurate picture of cause and effect in a dynamic system, and this is especially true in nonlinear systems where interdependence between variables is complex. Such systems are known to produce mirage correlations that appear, disappear, and even reverse sign over time (9). Persistent correlations tend to only occur in specific circumstances; most notably, when there is synchrony between driver and response variables (the effect of the driving variable is strong enough that the response becomes enslaved to the driver). This is key, because basic host–pathogen dynamics are known to exhibit dynamical resonance when strongly forced by periodic drivers (10). Dynamical resonance causes the intrinsic nonlinear epidemiological dynamics to become synchronized (phase-locked) to the simple cyclic motion of the environmental driver. However, when drivers do not induce synchrony, the underlying nonlinear dynamics can cause the statistical relationship between driver and response to become very complex.

Indeed, the same simple epidemiological SIRS model of ref. 10 (a basic epidemiological model consisting of "susceptible-infected-recovered-susceptible") illustrates how the identical

Significance

Patterns of influenza outbreak are different in the tropics than in temperate regions. Although considerable experimental progress has been made in identifying climate-related drivers of influenza, the apparent latitudinal differences in outbreak patterns raise basic questions as to how potential environmental variables combine and act across the globe. Adopting an empirical dynamic modeling framework, we clarify that absolute humidity drives influenza outbreaks across latitudes, find that the effect of absolute humidity on influenza is U-shaped, and show that this U-shaped pattern is mediated by temperature. These findings offer a unifying synthesis that explains why experiments and analyses disagree on this relationship.

Author contributions: E.R.D., M.C.M., R.D.H., S.B., and G.S. designed research; E.R.D. and G.S. performed research; E.R.D. and M.C.M. analyzed data; M.C.M., R.D.H., and S.B. contributed to interpretation of results; and E.R.D. and G.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

See Commentary on page 12899.

[1]To whom correspondence should be addressed. Email: gsugihara@ucsd.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1607747113/-/DCSupplemental.

mechanistic effect of a driver can produce very different behavior when only the periodicity is changed (Fig. 1). Simulating the model with a strongly seasonal environmental driver produces consistent seasonal outbreaks that correlate very strongly to the driver (Fig. 1*A*, where seasonality is quantified by the Pearson's correlation between the raw values and day of year mean). However, this changes when the driver is replaced with another signal that has the same magnitude, but with much weaker seasonality (low correlation between raw values and day of year mean). The weakened seasonality results not only in much less seasonality in influenza but also much lower correlation between influenza infections and climate (Fig. 1*B*). The mechanisms have not changed, only the spectrum (periodicity) of the driver.

Moreover, not all regions of the world have strong climate seasonality. Although the seasonal cycle in a temperate country such as Germany can explain more than 90% of the variance in absolute humidity across multiple decades, in tropical countries such as Singapore, it explains less than 30%. Indeed, there is a strong correspondence between seasonality of climate and seasonality of influenza. Fig. 2 shows the seasonality in absolute humidity (Fig. 2*A*) and influenza (Fig. 2*B*) across countries. The countries with the least seasonality in environment (yellow shades) also have the least seasonal influenza (Spearman's ρ = 0.73). This is most pronounced in the tropics, where it is important to note that correlations between influenza and environment are also most elusive.

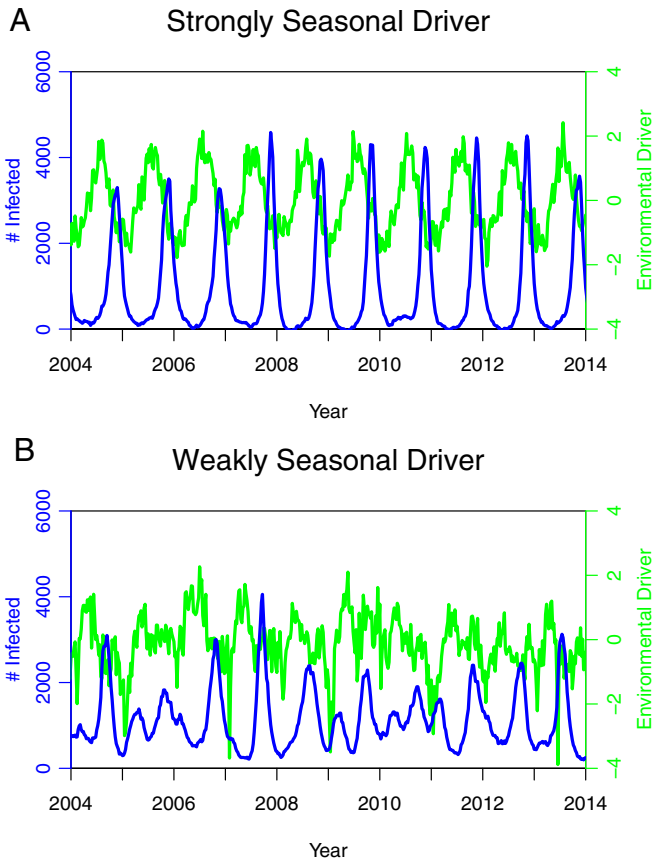

Fig. 1. Stochastic SIRS model with strongly and weakly seasonal drivers. (*A*) A strongly periodic environment induces synchrony through dynamical resonance, causing peaks in infection to correlate with seasonal lows in the seasonal environment. (*B*) If the same SIRS model is driven by a seasonal signal with the same variance, but much weaker seasonality, there is no dynamic resonance, infection peaks show much weaker seasonality, and correlation between infection and environment is much lower.

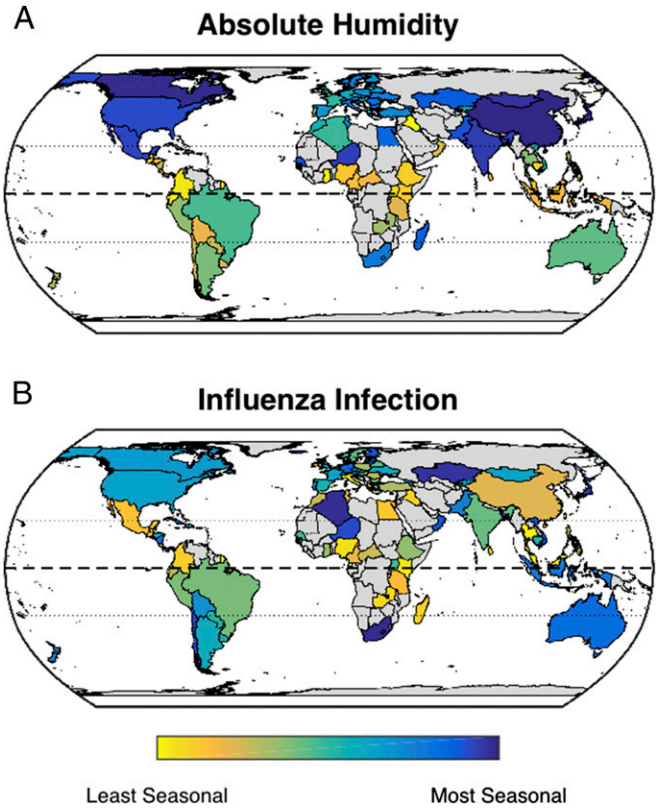

Fig. 2. Correspondence between seasonality of environment and seasonality of influenza infection. Countries are colored from the least seasonal to most seasonal for absolute humidity (*A*) and influenza infection (*B*). The Spearman correlation between the two is high: ρ = 0.73.

These results illustrate a crucial point: lack of seasonality in influenza and lack of correlation between influenza and climate do not indicate a lack of environmental forcing. Indeed, a poor correlation between climate and influenza is to be expected with a nonlinear relationship where seasonality is weak; however, where seasonality is strong, dynamical resonance can emerge to produce the appearance of a linear relationship (correlation). To examine these ideas critically and to develop a more unified understanding of how climate affects influenza globally requires an approach that can cope with the nonlinear interactions inherent in the system.

Cross-mapping. Convergent cross-mapping is a method for detecting causality in nonlinear dynamic systems (see the following playlist of three brief animations (adapted from the supplement of ref. 12): https://www.youtube.com/watch?v=fevurdpiRYg&list=PL-SSmlAMhY3bnogGTe2tf7hpWpl508pZZ).It looks for signatures of the cause in the time series of the affected variable. The idea is that if a variable, *X*, can predict the current or previous state of another variable, *Y*, then *Y* causally influenced *X*. For example, if sardine abundance can be used to recover sea surface temperature, then temperature had a causal effect on sardines. We examine cross-map prediction between influenza and the purported environmental drivers: absolute humidity, temperature, relative humidity, and precipitation. The mutual seasonality of influenza and these environmental variables makes it especially important to distinguish causal interactions from spurious correlation. Thus, we compare the cross-map prediction measured for the observational environmental time series with the null expectation obtained from cross-mapping with surrogate time series having the same seasonal cycle as the actual driver, but with randomized

anomalies. Causal forcing is established when cross-map prediction is significantly better for the real environmental driver than it is for the null surrogates (11).

Fig. 3 shows box-and-whisker plots of the null distributions for cross-map skill (ρ_{CCM}) ordered according to distance from the equator (absolute latitude). The measured CCM skill values are plotted on top as red circles that are filled if the value is significantly different from the null distribution ($P \leq 0.05$), and open otherwise. Absolute humidity (AH) and temperature (T) both show significant forcing in countries across latitude. The results have very high metasignificance (Fisher's method): $P < 1.6 \times 10^{-6}$ for AH and $P < 2.9 \times 10^{-3}$ for T. There is even higher metasignificant evidence for forcing across latitudes by relative humidity: $P < 1.8 \times 10^{-14}$. However, paired Wilcox tests (nonparametric generalization of Student's t test) in the prediction skill confirm that the cross-map effect with respect to RH is weaker on average than with either AH or T ($P < 0.02$), but that the strength of effects of AH and T are not significantly different from one another. Finally, there is some evidence that precipitation may be a causal variable in a few countries (global metasignificance, $P < 7.3 \times 10^{-3}$; tropics only, $P < 0.023$).

If all other things are equal, the magnitude of cross-map skill (ρ_{CCM}) can indicate the strength of causal effect. On this basis, one might conclude that because the absolute level of cross-map skill is lower in the tropics, the causal effect of climate (absolute humidity or temperature) is weaker. However, this would be incorrect, as "all things equal" does not apply across latitudes because of the differences in seasonality, and hence differences in the baseline predictability in these systems. This dichotomy is illustrated with the model shown in Fig. 1. Both model realizations have the same ultimate magnitude of the effect of AH on flu.

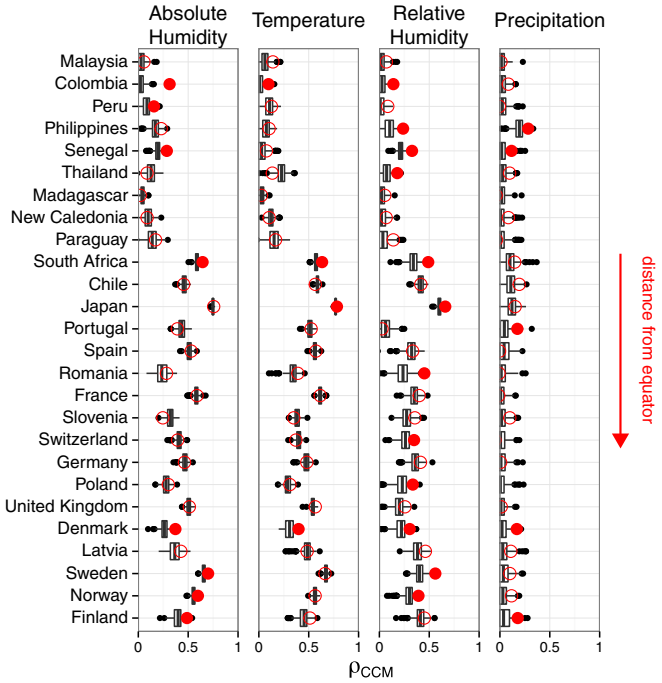

Fig. 3. Detecting cross-map causality beyond shared seasonality of environmental drivers on influenza. Red circles show the unlagged cross-map skill (ρ_{CCM}) for observed influenza predicting purported seasonal drivers: absolute humidity, temperature, relative humidity, and precipitation. Together with this, box-and-whisker plots show the null distributions for ρ_{CCM} expected from random surrogate time series that shares the same seasonality as the true environmental driver. Countries are ordered according to distance from the equator (absolute latitude). Filled circles indicate that the measured ρ_{CCM} is significantly better than the null expectation ($P \leq 0.05$).

In Fig. 1*A*, however, the driver has much stronger seasonality, and is therefore more predictable. The seasonal cycle, being identical from year to year, is trivial to predict; meaning the stronger the effect of the seasonal cycle, the easier the driver is to predict. Thus, the absolute level of ρ_{CCM} is substantially higher in Fig. 1*A* than Fig. 1*B*, even though the strength of causal effect is likely unchanged. Therefore, in this case, the skill of CCM (ρ_{CCM}) should not be used as a relative measure of causal strength when comparing CCM across latitudes, as the climate time series in tropical and temperate countries do not have the same basic levels of predictability.

Multivariate EDM. To examine whether AH or T has a stronger causal effect on influenza according to previous work (12), we use a multivariate EDM approach that looks for improvement in forecasting when a suspected causal variable is included. In brief, if a multivariate empirical dynamic model containing a potential driving variable Y produces better forecasts of X than without, then Y causally influenced X. The results are summarized in Fig. 4. The results show strong evidence that AH and T are drivers of influenza, as including either variable leads to improved forecast skill ($P < 0.001$ globally). However, in the tropics (where AH and T generally have weaker correlation), we find that including AH and T together as embedding coordinates leads to slightly more improvement over either one alone. In temperate countries, where correlations between AH and T are extremely high (generally > 0.9), these variables contain almost identical information, and hence there is less difference in forecast skill on average between embeddings with AH, T, or both. Overall, multivariate forecast improvement suggests the possibility that AH and T taken together have a nonlinear effect on influenza that is global, but that this effect is concealed in the temperate region by their strong correlation to each other.

To explore the mechanism for this interaction further, we predict the hypothetical change in influenza incidence, denoted Δflu, at historical points that would occur from small increases and decreases in the environmental driver [EDM scenario exploration (13)]. This simple device allows us not only to directly quantify the magnitude and direction of effect but also to keep track of how the effect changes through time and with varying environmental conditions. Fig. 5*A* shows country by country the magnitude of the effect of absolute humidity on influenza, $\Delta flu/\Delta AH$. Countries at high latitudes generally show a negative effect of absolute humidity on influenza, whereas low-latitude countries generally show a positive effect. This is consistent with the hypothesized U-shaped response of influenza to AH suggested by certain experiments (8).

We can examine this effect further by plotting the scenario exploration results for all the countries together. Fig. 5*B* shows that the effect of AH on influenza is negative at low AH ($\Delta flu/\Delta AH < 0$), but positive at high AH ($\Delta flu/\Delta AH > 0$), consistent with a U-shaped response of influenza survival to absolute humidity. The negative effect on incidence at low AH and positive effect at high AH appear roughly equivalent when the data are normalized to average total reported cases in a year in that country.

The same analysis for temperature (Fig. 5*C*) does not exhibit such a clear state-dependent (nonlinear) effect. Temperature changes can have a positive ($\Delta flu/\Delta T > 0$) or negative ($\Delta flu/\Delta T < 0$) effect on influenza at the same temperature. Indeed, the molecular arguments for a U-shaped effect of absolute humidity on influenza also predict that temperature should be a control on the balance between the positive and negative effects of absolute humidity (7).

With this in mind, we look at the effect of absolute humidity on influenza ($\Delta flu/\Delta AH$) as a function of temperature (Fig. 5*D*). This is perhaps the most interesting picture to emerge, as there are a number of features that corroborate and elaborate existing

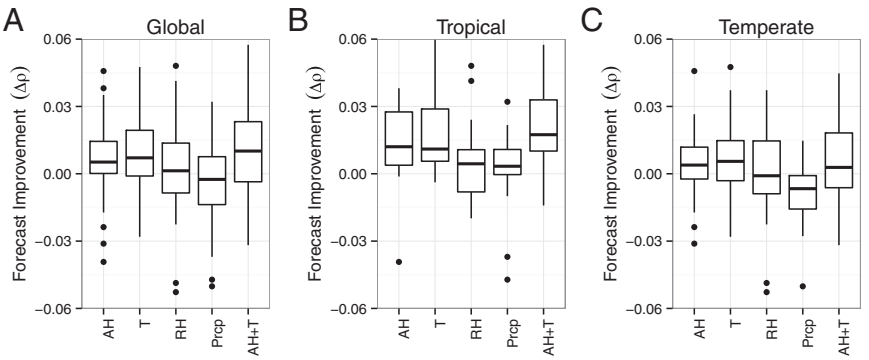

Fig. 4. Forecast improvement with multivariate EDM. Causal effect is demonstrated if EDM forecast skill (ρ) improves when a driver variable is included in the EDM model. This is quantified by $\Delta\rho = \rho$(with driver) − ρ(without driver), where ρ is the Pearson's correlation between observations and EDM predictions. Including either absolute humidity (AH) or temperature (T) leads to significant ($P < 0.001$) improvement in forecast skill both globally (*A*) and the tropics specifically (*B*). The significance is more marginal when looking solely at temperate countries (*C*) ($P < 0.07$ for AH; $P < 0.03$ for T). In the tropics, including both (AH +T) is marginally better than either AH or T alone, suggesting possible compound effects of temperature and humidity.

ideas. We observe the following: temperature has a relatively loose control on ($\Delta flu/\Delta AH$) when it is below ∼70 °F, but the effect on ($\Delta flu/\Delta AH$) is consistently negative; the positive effect of absolute humidity appears more strongly controlled by temperature and appears to be restricted to a narrow band of temperature between 75 °F and 85 °F; at the highest temperatures, the effect of absolute humidity goes to zero in exact concordance with the laboratory finding that aerosol transmission of influenza is blocked at 30 °C (86 °F) (14).

These results suggest that the balance between positive and negative effects of absolute humidity appears to shift somewhere between 70 °F and 75 °F. This is especially clear if we look back at the plot in Fig. 5*B* that shows the effect of absolute humidity on influenza across the global range of absolute humidity. Fig. 6 shows the same data, but now the points are split between two panels, based on temperature. On the left are observations where temperature was below 75 °F; on the right are observations where the temperature was between 75 °F and 85 °F. The red lines represent the 0.1 and 0.9 quantile regressions. The quantile regressions show that the measured effect of AH on influenza is almost always negative when T < 75 °F, and almost always positive when 75 °F ≤ T ≤ 85 °F.

Discussion

Prior population-level analyses have focused on explanations for influenza seasonality, using correlations. However, correlation is a limited tool for understanding causality in nonlinear systems such as those exhibiting complex host–pathogen dynamics. Building on previous work on dynamic resonance (10), we show that persistent correlations between influenza and environment should only be expected where seasonality is strongest; that is, in temperate countries (as is clear from Fig. 2).

This motivates adopting an EDM framework to address the general problem of identifying external drivers of nonlinear dynamics. We are able to provide clear tests at the population level of alternative hypotheses about global climatic drivers of influenza and show that there are general rules that span temperate and tropical latitudes. We find clear evidence that temperature and

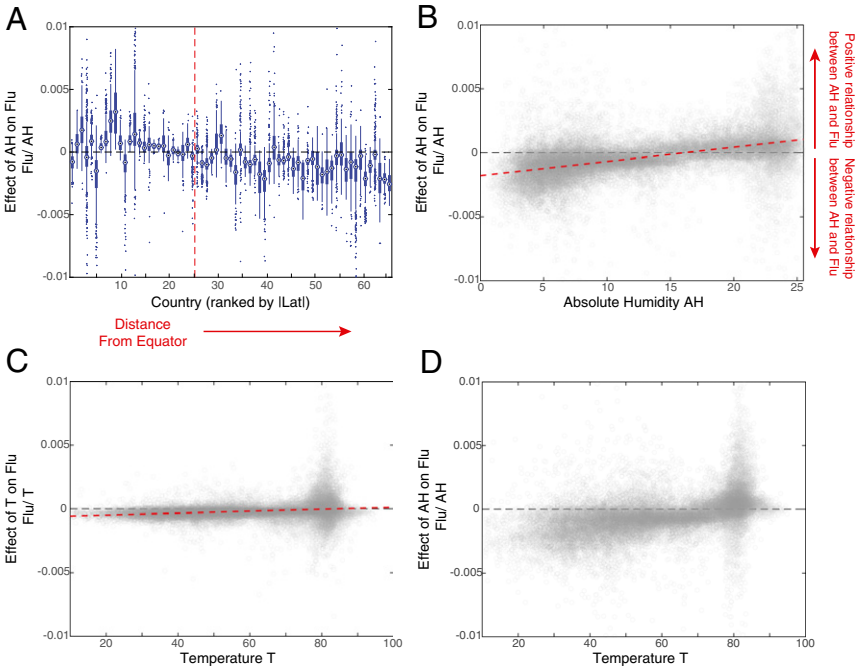

Fig. 5. Scenario exploration with multivariate EDM gives a measure of the effect of environment on influenza infection by predicting the change in influenza (Δflu) that results from a small change in absolute humidity (ΔAH) or temperature (ΔT). Analysis covers countries with at least 208 observations (4 y) and area <1.5 million mi^2. (*A*) Range of values for Δflu/ΔAH for each country across latitude, ordered by distance from equator (*SI Appendix*, Table S1). Countries closest to the equator tend to show a positive effect of AH on influenza infection, whereas countries furthest from the equator show a negative effect (red line indicates the tropical boundary). (*B*) Effect of absolute humidity on influenza (Δflu/ΔAH) as a function of AH grouped over all countries. Each point represents the estimated effect from scenario exploration for one historical point in one of the study countries. At low AH (typical of high-latitude countries), AH has a negative effect on influenza infection, whereas at high AH (typical of low-latitude countries), AH has a positive effect on influenza. (*C*) Effect of temperature on influenza (Δflu/ΔT) as a function of T. Evidence of a single global effect is much weaker, but it suggests there might be important temperature thresholds. (*D*) How the effect of absolute humidity on influenza (Δflu/ΔAH) changes as a function of T. Additional details given in *SI Appendix, Section 5*.

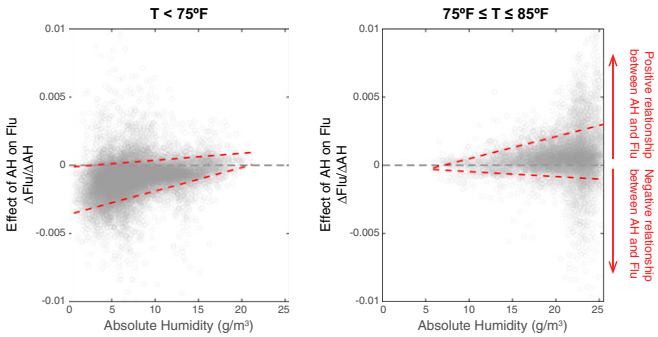

Fig. 6. Temperature thresholds in the effect of absolute humidity (ΔAH) on influenza (Δflu). The results of scenario exploration in Fig. 4*B* are replotted according to temperature. (*Left*) Values correspond to observations when T was below 75 °F. (*Right*) Values correspond to observations when T was between 75 °F and 85 °F. The red dashed lines indicate the 0.1 and 0.9 quantile regressions.

absolute humidity have more direct effects on global influenza than precipitation and relative humidity (Figs. 3 and 4), and that these environmental drivers are important at all latitudes, regardless of the degree of seasonality in the environment and influenza.

Critically, we find that AH and T interact nonlinearly (Fig. 4). That is, the relationships among temperature, humidity, and influenza are not understood by trying to study the influence of the different variables separately (e.g., by linear correlation) but, rather, by studying their mutually interdependent effects. Indeed, not only does nonlinear analysis show a U-shaped effect of absolute humidity on influenza, but also that temperature mediates this relationship (Figs. 5 and 6). This closely corroborates the theoretical molecular basis for temperature and absolute humidity effects on influenza (7). In cold environments, the viral envelope is prone to disruption; hence, drier air (low AH) promotes the spread of influenza when temperature is low. In warm environments, the viral envelope is prone to desiccation; hence, wetter air (high AH) promotes the spread of influenza when temperature is high. The EDM analysis suggests that this tradeoff between positive and negative effects happens in the neighborhood of 24 °C (75 °F).

At present, there are no modeling or laboratory results against which to compare these threshold results. Critically, the recent laboratory experiments that demonstrated a negative monotonic effect of *AH* on influenza (2, 3) only varied temperatures as high as 20 °C (68 °F), and thus could only produce the negative effect of AH that we found at lower temperatures. Our results set the stage for laboratory studies that experimentally test this threshold by varying temperature and humidity over the full range of conditions experienced globally. Augmented with further laboratory testing, these population-level results could help set the stage for public health initiatives such as placing humidifiers in schools and hospitals during cold, dry, temperate winters (as suggested by ref. 3), and in the tropics, perhaps using dehumidifiers or air conditioners set above 75 °F to dry air in public buildings.

For influenza, the interaction between transmission and environment is understood to be one of many important epidemiological processes, including strain-dependent effects, spatial dynamics within countries (15), spatial dynamics between countries (16), and antigenic drift (17). One of the advantages of EDM is that these other factors, insofar as they interrelate with the deterministic dynamics in countrywide infection, are indirectly accounted for, albeit phenomenologically, in the reconstructed attractors (18, 19). Nevertheless, the most complete understanding of influenza will come from treating all these factors integratively (4). This sets the stage for future studies that will specifically incorporate these other processes in a broader multivariate setting.

Methods

Data. Total laboratory confirmed influenza A and B cases per week were retrieved on April 2, 2014, from the World Health Organization via FluNet by country (apps.who.int/globalatlas/dataQuery/) for January 1, 1996–March 26, 2014 (Dataset S1). Ideally, we would like to analyze an index of incidence density (per capita), and thus need to account for population size and reporting rates. To account for changes in population size, we divide by linearly interpolated annual population data taken from ref. 20.

Accounting for changes in reporting rate over time is a more difficult issue to address. A typical approach is to divide weekly incidence by the total reported incidence for that country, that year. However, this masks all year-to-year differences in influenza infections, including those that arise naturally from the nonlinear intrinsic dynamics of host–pathogen dynamics and from the state-dependent effect of climate variability. Such standardization would artificially inflate the seasonal signature in influenza incidence, and hence would make the task of disentangling causality from shared seasonality harder, not easier.

However, accounting for the substantial differences in reporting rates between countries can be addressed to first order by dividing weekly incidence by the total reports per year in that country averaged over all years reported. Note that CCM is unaffected by arbitrary scaling, so this normalization only affects the comparisons between countries (Figs. 4–6).

Weekly temperature and absolute humidity data were calculated from National Oceanic and Atmospheric Administration Global Surface Summary of the Day (ftp://ftp.ncdc.noaa.gov/pub/data/gsod). A single value was calculated for each country by taking a simple average over all available stations.

Relative humidity was approximated from temperature and absolute humidity by assuming standard atmospheric pressure. Precipitation data were taken from the combined National Centers for Environmental Prediction Climate Climate Forecast System (CFSR and CFSv2) for the $2° \times 2°$ grid square corresponding to the largest urban area (Dataset S1).

In both cases, there are potential issues with spatial aggregation. Spatial aggregation has a tendency to mask nonlinear dynamics (21), yielding time series that are dominated by linear dynamics such as seasonality. For the influenza data, the only real option for a global analysis is to look at country-wide data. However, there are other reasonable options for the physical data. In some countries such as Thailand, where population is very localized (relative to the size of the country), it is entirely likely that taking a single weather station closest to the largest city will be representative of the country as a whole. However, this approach is much harder to justify for a country such as Germany, where there are multiple large urban areas distributed over the area of the country. Ultimately, taking the approach of simple averages across all of the available monitoring sites is a simple approach that does not require country-by-country expert knowledge of climate and geography. Because spatial averaging tends to enhance seasonality, this potentially makes it hard to establish significance, but (given the surrogate methods we use) is not expected to produce spurious relationships and will tend to produce conservative results.

Seasonality. The seasonal cycle is determined using a smoothing spline (smoothing parameter = 0.8) to the target variable as a function of day of year, where the spline is wrapped December 31–January 1. Unlike other methods for extracting seasonal cycles based on Fourier decomposition, this method works for both time series variables that have a mostly sinusoidal waveform (e.g., temperature) and time series variables that have very nonsinusoidal waveforms (e.g., precipitation or influenza incidence that has a hard boundary at 0). This gives the environmental variable as a function of the day of year. The Pearson's correlation between the observed value and the day of year average (i.e., the seasonal cycle) is used as a quantitative measure for "seasonality."

EDM. EDM is a quantitative framework that uses time series data to reconstruct and study the underlying attractor (9, 11–13, 19, 21–26). Dynamical systems are typically studied in terms of parametric equations; for example, a SIR model of disease outbreak. These equations can then be solved to generate the changes of the system variables through time. When viewed in state space (multivariate space in which each axis corresponds to a system variable), this becomes a trajectory that traces out the underlying attractor of the system. The attractor for an unknown system from nature can be reconstructed directly from a complete set of observational time series (illustrated in the brief animation here: https://www.youtube.com/watch?v=fevurdpiRYg), but it can also be reconstructed

from a single time series by using time-lagged coordinates (https://www.youtube.com/watch?v=QQwtrWBwxQg). The dynamic attractor is a complete representation of the system, and thus can be studied in place of parametric equations to predict and understand systems such as host–pathogen dynamics.

This general framework can be applied in a number of ways—to detect causation (9), forecast future states (22, 23), track interactions (24), and so on—using the R package "rEDM" (https://cran.r-project.org/web/packages/rEDM/index.html), which includes a technical description of the methods in a vignette (https://cran.r-project.org/web/packages/rEDM/vignettes/rEDM_tutorial.html documentation). Additional details on the exact calculations are given here. Note that EDM prediction skill in this article is always measured with Pearson's correlation (ρ) between observed and predicted values.

CCM Analysis and Seasonal Surrogates. The basic idea of convergent cross-mapping is to use prediction between variables as a test for causality. If variable Y had a causal effect on variable X, then causal information of variable Y should be present in X, and so the attractor recovered for variable X should be able to predict the states of variable Y (https://www.youtube.com/watch?v=iSttQwb-_5Y).

For basic CCM analysis, we use simplex projection, which has a single parameter to select: the embedding dimension E. As in ref. 24, we select E on the basis of the optimal prediction of cross-mapping lagged 1 wk, then measure the unlagged cross-map skill with this value of E. Some of the influenza time series have long stretches of zero-values, which will contain little or no information about the climate. As a consequence, we exclude from the prediction set all stretches of zeros at least 8 time points long (the maximum embedding dimension tested) to improve the sensitivity of the analysis.

Checking for convergence in cross-map skill (i.e., that cross-map skill improves with the amount of data used) is a general way to distinguish cross-mapping from spurious correlation (9). However, we are concerned here with a more specific problem of distinguishing driving effects from mutual seasonality. This is more directly addressed by developing a null test with surrogate time series, in the vein of ref. 11.

For a forcing variable $Z(t)$ (e.g., absolute humidity or temperature), we calculate the day of year average $\bar{Z}$ (i.e., the seasonal cycle) as earlier, and the seasonal anomaly as the difference between the observed value $Z(t)$ and the day of year average for that day, $\tilde{Z}(t) = Z(t) - \bar{Z}(t)$. We then randomly shuffle (permute) the time indices of the seasonal anomalies. Adding the shuffled anomalies back to the season average gives a surrogate time series Z^* that has the same seasonal average as Z, but with random anomalies. If Z is in fact a driver of influenza, then influenza will be sensitive not only to the seasonal component of Z but also to the anomalies. Thus, influenza should better predict the real time series Z than the surrogate Z^*. In practice, we repeat the shuffling procedure 500 times to produce an ensemble of surrogates.

Note that in determining the metasignificance of the CCM tests, we apply Fisher's method, which relies on the assumption that the different P values are independent. However, both influenza epidemics and climate have spatial dependencies between countries. Because there is no practical way to determine the covariance between P values in this experiment, Fisher's method is the best guide, but it is important to note that the metasignificance may be somewhat anticonservative.

Multivariate EDM: Forecast Improvement. When driver variables are stochastic (e.g., seasonal anomalies of climactic variables), multivariate forecast improvement can be used as a test for causality (12). A stochastic variable is considered causal if explicitly including it as a coordinate in the state space leads to improved nearest-neighbor forecasts. In the case of seasonal influenza, however, we cannot necessarily treat the environmental time series as stochastic variables. This means that information about the drivers is already contained in the univariate embedding (9, 25). Thus, we modify the method of ref. 12 as follows.

We determine the optimal univariate embedding dimension, E^*, for each influenza time series, following ref. 26. A univariate embedding with dimension $E < E^*$ will be "under embedded"; that is, it will not contain full information about the system state and dynamics. In this case, incorporating information about a driver in a multivariate embedding will generally lead to an increase in forecast skill. Thus, we calculate the improvement in forecast skill, using simplex projection of the univariate embedding with $E = E^*-1$ and the same embedding, but with the candidate environmental variable(s) included as a coordinate. As with CCM, we exclude all stretches of zeros at least eight time points long (maximum tested E) from the prediction set to improve the sensitivity of the analysis. Significant improvement is calculated using a one-sample Wilcox test (nonparametric Student's t test), and significant differences in the improvement between variables is calculated using a paired two-sample Wilcox test.

Multivariate EDM: Scenario Exploration. Multivariate scenario exploration is a way to empirically assess the effect of a small change in a physical driver (e.g., absolute humidity) on influenza incidence. We predict the effect of a small increase in absolute humidity or temperature on influenza 2 wk later to understand the sensitivity of influenza outbreaks to the environment. For each historical time point, t, we predict influenza with a small increase ($+\Delta Z/2$) and a small decrease ($-\Delta Z/2$) in historically measured driver $Z(t)$. The difference in predicted influenza is $\Delta\text{flu} = \text{flu}_{t+1}[Z = Z(t) + \Delta Z/2] - \text{flu}_{t+2}[Z = Z(t) - \Delta Z/2]$, and the ratio of $\Delta flu/\Delta Z$ quantifies the sensitivity of influenza infection to the driver Z at time t. We use $\Delta Z = 0.2$ g/m^3 and 0.5 °F for absolute humidity and temperature, respectively. These values correspond to ~5% of the SD of these variables across all of the countries analyzed. Forecasts were performed using S-maps (22), with $E = 7$ and $\theta = 0.9$.

1. Xie X, Li Y, Chwang ATY, Ho PL, Seto WH (2007) How far droplets can move in indoor environments–Revisiting the Wells evaporation-falling curve. *Indoor Air* 17(3):211–225.
2. Lowen AC, Mubareka S, Steel J, Palese P (2007) Influenza virus transmission is dependent on relative humidity and temperature. *PLoS Pathog* 3(10):1470–1476.
3. Shaman J, Kohn M (2009) Absolute humidity modulates influenza survival, transmission, and seasonality. *Proc Natl Acad Sci USA* 106(9):3243–3248.
4. Lofgren E, Fefferman NH, Naumov YN, Gorski J, Naumova EN (2007) Influenza seasonality: Underlying causes and modeling theories. *J Virol* 81(11):5429–5436.
5. Tamerius JD, et al. (2013) Environmental predictors of seasonal influenza epidemics across temperate and tropical climates. *PLoS Pathog* 9(3):e1003194.
6. Shaman J, Pitzer VE, Viboud C, Grenfell BT, Lipsitch M (2010) Absolute humidity and the seasonal onset of influenza in the continental United States. *PLoS Biol* 8(2): e1000316.
7. Minhaz Ud-Dean SM (2010) Structural explanation for the effect of humidity on persistence of airborne virus: Seasonality of influenza. *J Theor Biol* 264(3):822–829.
8. Schaffer FL, Soergel ME, Straube DC (1976) Survival of airborne influenza virus: Effects of propagating host, relative humidity, and composition of spray fluids. *Arch Virol* 51(4):263–273.
9. Sugihara G, et al. (2012) Detecting causality in complex ecosystems. *Science* 338(6106): 496–500.
10. Dushoff J, Plotkin JB, Levin SA, Earn DJD (2004) Dynamical resonance can account for seasonality of influenza epidemics. *Proc Natl Acad Sci USA* 101(48):16915–16916.
11. Tsonis AA, et al. (2015) Dynamical evidence for causality between galactic cosmic rays and interannual variation in global temperature. *Proc Natl Acad Sci USA* 112(11): 3253–3256.
12. Dixon PA, Milicich MJ, Sugihara G (1999) Episodic fluctuations in larval supply. *Science* 283(5407):1528–1530.
13. Deyle ER, et al. (2013) Predicting climate effects on Pacific sardine. *Proc Natl Acad Sci USA* 110(16):6430–6435.
14. Lowen AC, Steel J, Mubareka S, Palese P (2008) High temperature (30 degrees C) blocks aerosol but not contact transmission of influenza virus. *J Virol* 82(11):5650–5652.
15. Viboud C, et al. (2006) Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* 312(5772):447–451.
16. Russell CA, et al. (2008) The global circulation of seasonal influenza A (H3N2) viruses. *Science* 320(5874):340–346.
17. Smith DJ, et al. (2004) Mapping the antigenic and genetic evolution of influenza virus. *Science* 305(5682):371–376.
18. Takens F (1981) Detecting strange attractors in turbulence. *Dynamical Systems and Turbulence, Warwick 1980*, eds Rand DA, Young L (Lector Notes in Mathematics, New York), pp 366–381.
19. Ye H, et al. (2015) Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. *Proc Natl Acad Sci USA* 112(13):E1569–E1576.
20. The World Bank (2015) *Health Nutrition and Population Statistics*. Available at data.worldbank.org/indicator/SP.POP.TOTL. Accessed April 4, 2014.
21. Sugihara G, Grenfell B, May RM (1990) Distinguishing error from chaos in ecological time series. *Philos Trans R Soc Lond B Biol Sci* 330(1257):235–251.
22. Sugihara G, May RM (1990) Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature* 344(6268):734–741.
23. Sugihara G (1994) Nonlinear forecasting for the classification of natural time series. *Philos T Roy Soc A* 348(1688):477–495.
24. Deyle ER, May RM, Munch SB, Sugihara G (2016) Tracking and forecasting ecosystem interactions in real time. *Proc Biol Sci* 283(1822):20152258.
25. Stark J, Broomhead DS, Davies ME, Huke J (2003) Delay embeddings for forced systems. II. Stochastic forcing. *J Nonlinear Sci* 13(6):519–577.
26. Glaser SM, et al. (2014) Complex dynamics may limit prediction in marine fisheries. *Fish Fish* 15(4):616–633.