

Does influenza drive absolute humidity?

Edward B. Baskerville^a and Sarah Cobey^{a,1}

Multiple lines of evidence suggest that absolute humidity and school terms affect the timing of influenza epidemics, although the contributions of these and other factors are still unresolved (1–5). In PNAS, Deyle et al. (6) applied convergent cross-mapping (CCM) to measure the impact of environmental variables on influenza activity. The authors concluded that absolute humidity and temperature are not merely correlated with influenza activity but in fact directly or indirectly cause epidemics.

The mathematical basis of CCM involves strict assumptions that may be violated in nature (7). Process noise, secular changes in parameters, and other transient dynamics can lead to situations where the underlying attractor is undersampled, causing errors in reconstruction and inference (8). There are not yet any clear statistical tests to determine when CCM has gone awry.

One simple test of the method is whether it correctly predicts that influenza does not drive absolute humidity, temperature, relative humidity, and precipitation. We used the authors' code to test these interactions for each of the 26 countries shown in figure 3 of Deyle et al. (6), with modifications, described below, to support different test assumptions. Deyle et al. (6) used surrogate time series with randomized anomalies for each of the environmental variables to construct a null distribution for significance testing. Initially, we analogously used surrogate time series only for influenza to assess whether influenza affected each of these variables. In 53 of 104 tests influenza was inferred to drive the environmental variable. When using surrogate time series for both influenza and the environmental variable influenza was inferred

to drive the environmental variable in 56 of 104 tests (Fig. 1). This general result was robust to different transformations of the time series, including

- i) using surrogates for both influenza and the environmental variable, influenza alone, or the environmental variable alone;
- ii) using simple weekly means instead of fitted splines for the seasonal surrogates;
- iii) removing all zeros from the influenza time series and restricting analyses to periods when influenza was present;
- iv) log-transforming influenza incidence; and
- v) additionally requiring that the maximum cross-correlation appear at a lag corresponding to the correct directionality in time (9).

Summaries of results from different test variants are in Table 1. More thorough testing with simulated data shows that using CCM with surrogate time series frequently leads to false positives. Other criteria for causality perform better in some instances, but none is universally reliable in a seasonally forced infectious disease system (8).

Although it may be possible to choose surrogates and transformations to obtain more favorable results, it is clear that more work is needed on the statistical basis of CCM before it can be applied out of the box. We agree with Deyle et al. (6) that weather affects the timing of influenza activity, but we do not think that analysis based on CCM supports the correct inference regarding causal direction.

The code for our analysis, along with source data and full results, is available at github.com/cobeylab/ccm-letter.

^aDepartment of Ecology and Evolution, University of Chicago, Chicago, IL 60637

Author contributions: E.B.B. and S.C. designed research; E.B.B. performed research; E.B.B. analyzed data; and E.B.B. and S.C. wrote the paper. The authors declare no conflict of interest.

¹To whom correspondence should be addressed. Email: cobey@uchicago.edu.

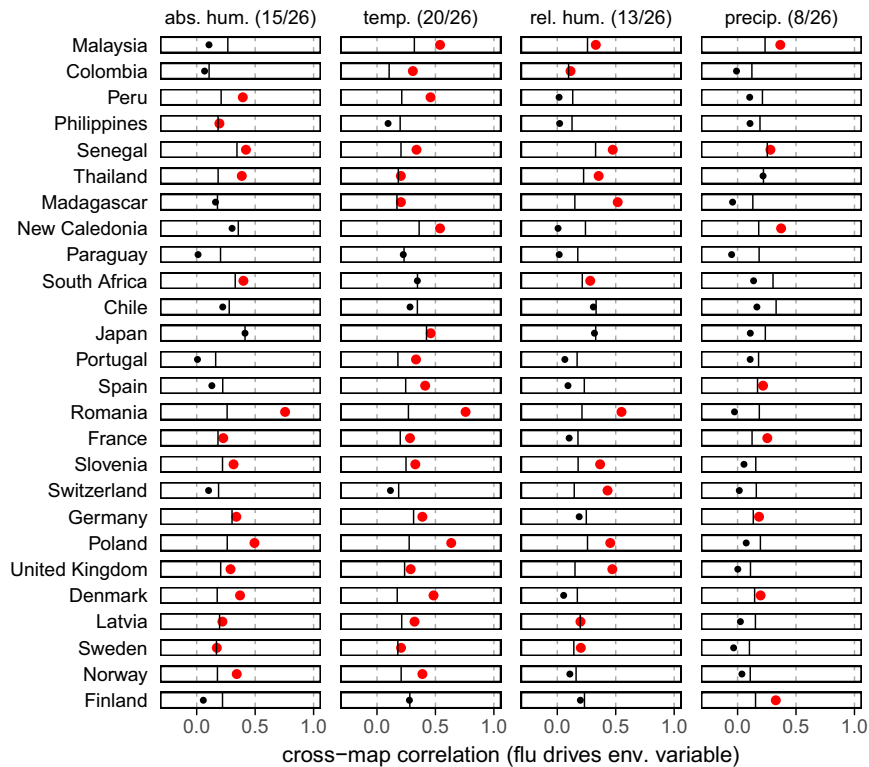


Fig. 1. CCM results testing whether influenza drives environmental variables in different countries. Circles show the cross-map correlation; significant interactions are large and red. Vertical bars show the location of the 95% quantile of the null distribution constructed from surrogate time series. Surrogates were constructed using splines for both influenza and environmental variables, and analysis was restricted to periods when influenza was present. Influenza incidence was not log-transformed.

Table 1. Summary of CCM results for whether influenza drives environmental variables, and vice versa, in different countries

Flu surrogate	Environment surrogate	Log-flu	No. significant of 104		No. significant of 104 (lag test)	
			Flu → environment	Environment → flu	Flu → environment	Environment → flu
	X		14	27	8	14
X			53	14	25	6
X	X		56	22	27	10
	X	X	17	31	11	15
X		X	55	29	25	13
X	X	X	52	34	23	15

Tests used surrogates for influenza and/or environmental variables, and either untransformed or log-transformed influenza incidence. Results are shown with and without the requirement that the maximum cross-correlation occur in the correct temporal direction. Surrogates were constructed using splines, and analysis was restricted to periods when influenza was present. Numbers are out of a total of 104 tests (26 countries, four tests each).

- 1 Bjørnstad ON, Viboud C (2016) Timing and periodicity of influenza epidemics. *Proc Natl Acad Sci USA* 113(46):12899–12901.
- 2 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.
- 3 Shaman J, Goldstein E, Lipsitch M (2010) Absolute humidity and pandemic versus epidemic influenza. *Am J Epidemiol* 173(2):127–135.
- 4 Chao DL, Halloran ME, Longini IM, Jr (2010) School opening dates predict pandemic influenza A(H1N1) outbreaks in the United States. *J Infect Dis* 202(6):877–880.
- 5 Tamerius J, Viboud C, Shaman J, Chowell G (2015) Impact of school cycles and environmental forcing on the timing of pandemic influenza activity in Mexican States, May–December 2009. *PLoS Comput Biol* 11(8):e1004337.
- 6 Deyle ER, Maher MC, Hernandez RD, Basu S, Sugihara G (2016) Global environmental drivers of influenza. *Proc Natl Acad Sci USA* 113(46):13081–13086.
- 7 Sugihara G, et al. (2012) Detecting causality in complex ecosystems. *Science* 338(6106):496–500.
- 8 Cobey S, Baskerville EB (2016) Limits to causal inference with state-space reconstruction for infectious disease. *PLoS One* 11(12):e0169050.
- 9 Ye H, Deyle ER, Gilarranz LJ, Sugihara G (2015) Distinguishing time-delayed causal interactions using convergent cross mapping. *Sci Rep* 5:14750.