Title

Long-lead predictions of eastern United States hot days from Pacific sea surface temperatures

Journal

Nature Geoscience, 9(5)

ISSN

1752-0894

Authors

McKinnon, KA
Rhines, A
Tingley, MP
et al.

Publication Date

2016-05-01

DOI

10.1038/ngeo2687

nature geoscience

ARTICLES

PUBLISHED ONLINE: 28 MARCH 2016 | DOI: 10.1038/NGEO2687

Long-lead predictions of eastern United States hot days from Pacific sea surface temperatures

K. A. McKinnon[1]*, A. Rhines[2], M. P. Tingley[3] and P. Huybers[4]

Seasonal forecast models exhibit only modest skill in predicting extreme summer temperatures across the eastern US. Anomalies in sea surface temperature and monthly-resolution rainfall have, however, been correlated with hot days in the US, and seasonal persistence of these anomalies suggests potential for long-lead predictability. Here we present a clustering analysis of daily maximum summer temperatures from US weather stations between 1982–2015 and identify a region spanning most of the eastern US where hot weather events tend to occur synchronously. We then show that an evolving pattern of sea surface temperature anomalies, termed the Pacific Extreme Pattern, provides for skillful prediction of hot weather within this region as much as 50 days in advance. Skill is demonstrated using out-of-sample predictions between 1950 and 2015. Rainfall deficits over the eastern US are also associated with the occurrence of the Pacific Extreme Pattern and are demonstrated to offer complementary skill in predicting high temperatures. The Pacific Extreme Pattern appears to provide a cohesive framework for improving seasonal prediction of summer precipitation deficits and high temperature anomalies in the eastern US.

The twenty-first century has already featured a number of costly[1] and ill-predicted[2] heat waves with negative impacts on crop production[3] and human health[4]. Climate change has the potential to increase the frequency and magnitude of heat waves[5], underscoring the importance of skillful seasonal and subseasonal forecasts[6]. Present seasonal forecast models have modest skill in predicting summer temperature across the eastern US (ref. 7) but tend to underestimate the probability of extremes at lead times longer than a week[8]. A complementary approach is to take advantage of certain mid-latitude conditions that may precede hot weather. Monthly-resolution precipitation deficits have been correlated with high temperatures in the following month[9] and anomalous mid-latitude sea surface temperature (SST) patterns have been highlighted as preceding heatwave episodes[10–13]. In the following we further explore and quantify the degree to which precipitation and sea surface temperatures provide for skillful prediction of high summer temperatures.

Defining hot weather

Our focus is on hindcasting the occurrence of high temperature anomalies during the 60 hottest days of summer (hereafter simply summer) recorded in the Global Historical Climatology Network[14]. We identify regions that tend to experience hot days synchronously through a clustering algorithm that groups stations together on the basis of simultaneously recording temperature anomalies above the 95th percentile during summer (see Methods). Clustering stations into five regions results in interpretable climate zones: the maritime west coast, the arid interior west, the semiarid and subtropical southwest, tropical Florida, and the humid and continental eastern US (Supplementary Fig. 1a). We focus on the eastern US cluster, comprising 1,613 stations, because it contains major population centres and primary regions of agriculture, both of which tend to be disproportionally affected by heat events[3,15]. Other levels of clustering are possible (Supplementary Fig. 1), and our choice represents a balance between specificity of events and inclusivity of data.

To obtain a single quantity representative of both the magnitude and spatial extent of high temperatures for each summer day, we calculate the spatial 95th percentile of temperature anomalies across the eastern US cluster, referred to as T95 (Data set 1). Figure 1 shows the values of T95 for the summer of 2012, which had a maximum T95 value of 10.4 °C on 29 June, meaning that 5% of the eastern US experienced temperatures at least 10.4 °C warmer than their climatological summer average. The average value and standard deviation of T95 between 1982 and 2015 are 4.4 °C and 2.1 °C, respectively, and we define hot days as those exceeding one standard deviation above this mean, or 6.5 °C. By definition, there are 321 hot days out of the 2,040 summer days (Fig. 1b and Supplementary Fig. 2). Results presented in the main text are based on predicting the occurrence of hot days, but similar results are obtained when predicting only the start date of consecutively occurring hot days (see Methods and Supplementary Fig. 3).

Precipitation deficits predict hot days

Previous work identified a significant correlation between hot summer days and antecedent precipitation deficits in parts of the US on monthly timescales[9]. The relationship between precipitation and temperature can be understood through local energy balance considerations, whereby low precipitation results in decreased soil moisture and reduces latent heat fluxes in favour of sensible heat fluxes. We first extend this work by resolving the relationship at daily timescales and assessing prediction skill as a function of lead time.

Precipitation variability is quantified using the standardized precipitation index (SPI; ref. 16) with a running integration period of 30 days (see Methods). Hot days are predicted based on eastern US SPI falling below a threshold, which is set to the average SPI for the following example. Lead times are counted following the last day over which precipitation is averaged. For predictions at a

[1]National Center for Atmospheric Research, Boulder, Colorado 80305, USA. [2]Department of Atmospheric Sciences, University of Washington, Seattle, Washington 98195, USA. [3]Department of Meteorology, Pennsylvania State University, Pennsylvania 16802, USA. [4]Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138, USA. *e-mail: mckinnon@ucar.edu

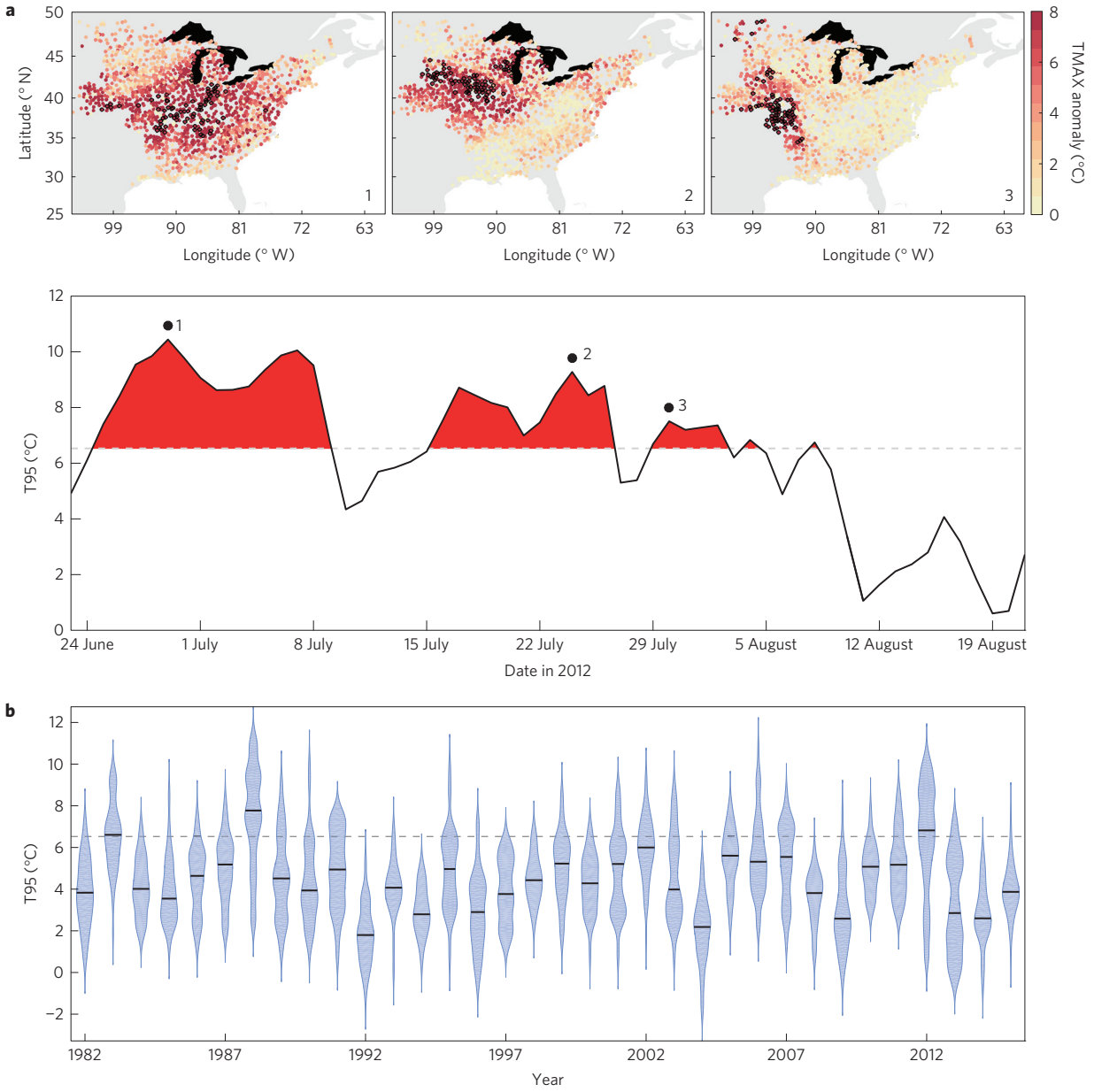

Figure 1 | Hot days and heat events are defined from station measurements of daily maximum summer temperature. a, The lower panel shows the time series of T95 for 2012, which featured three heat events, identified with filled circles as events 1, 2 and 3. Average daily maximum temperature (TMAX) anomalies at individual weather stations during the hottest day of each of these heat events are shown in the upper panels. Stations comprising the hottest 5% of the domain are outlined in black, and the colour bar is saturated at 8 °C. Hot days (indicated by red shading in the time series) are defined as T95 exceeding one standard deviation above the mean, where mean and standard deviation are calculated from all summers between 1982–2015. **b**, The distribution of T95 for each year of the study (1982–2015). Distributions are estimated using a kernel density smoother with a bandwidth of 0.5. The median value for each distribution is shown with a black horizontal line, and the dashed line is at one standard deviation above the mean of T95.

30-day lead time, below-average SPI values correspond to hot days occurring with a true positive rate of 71% and a false positive rate of 46%. These rates can be translated into an odds ratio, defined as [TPR(1 − FPR)] / [FPR(1 − TPR)], equal to 2.8, or almost a three-fold increase in the probability of a hot day (Supplementary Fig 4a).

The choice of threshold for a prediction can be varied depending on tolerance for true and false positives, with more negative SPI thresholds decreasing both the true and false positive rate. This trade-off is formalized by relative operating characteristic (ROC) curves[17] that represent the relationship between true and false positive rates as a function of threshold (Fig. 2). ROC curves are assigned a score by integrating the area under the curve. These scores are used as the skill metric throughout the analysis because they are appropriate for assessing both binary and unusual events, whereas many conventional skill scores tend to zero for rare events[18]. A model that always has the same true and false positive rate has a ROC score of 0.5, and is no more useful than a coin flip. A relevant baseline for comparison is a seasonal prediction model[19] that yielded ROC scores for extreme temperature events between 0.50–0.52 when applied to various regions across the globe.

At zero lead time, SPI-based predictions of hot days have a ROC score of 0.73, consistent with the expected relationship between dry soils and sensible heating. ROC scores decrease with lead time and are no longer significant (ROC score < 0.6) by a lead time of 45 days. Significance is estimated by bootstrapping and is presented at the 0.05 level throughout the paper (see Methods).

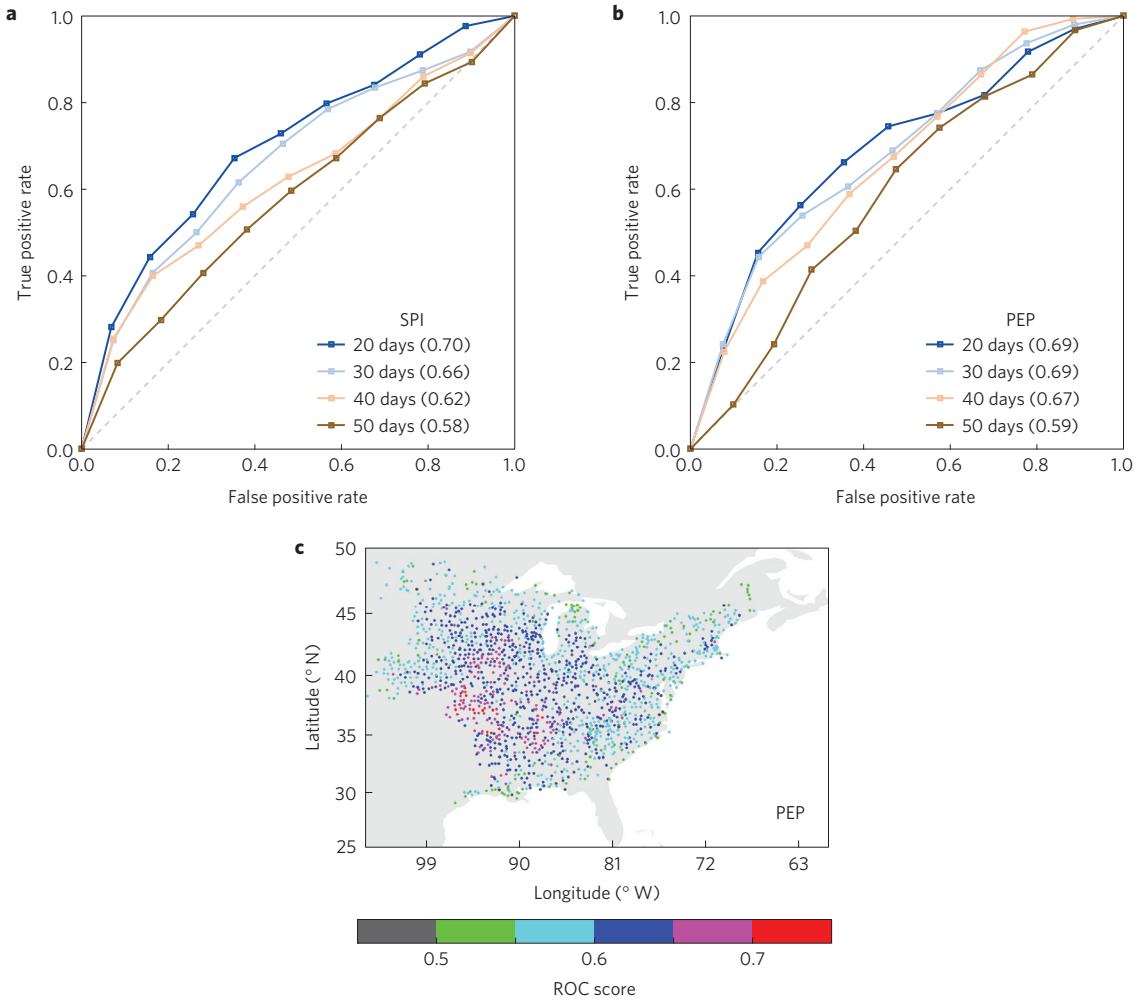

Figure 2 | Precipitation deficits and the Pacific Extreme Pattern skillfully predict hot days at long-lead times. a, Relative operating characteristic (ROC) curves for predictions of hot days at lead times of 20, 30, 40 and 50 days using the negative standardized precipitation index (SPI) with an integration time of 30 days. ROC scores for each lead time are shown in parentheses in the legend. **b**, As in **a**, but using the Pacific Extreme Pattern (PEP) as a predictor. Thresholds used to calculate true and false positive rates are indicated from the 0th (upper right squares) to the 100th (lower left squares) percentile of negative SPI (**a**) or PEP (**b**) in steps of 10%. ROC scores greater than or equal to 0.6 indicate significant (p-value < 0.05) skill. **c**, ROC scores for predicting hot days at individual stations using PEP. Hot days at individual stations are defined as having a temperature anomaly exceeding one standard deviation above their summer average.

The Pacific Extreme Pattern (PEP) predicts hot days

We next focus on the relationship between oceanic boundary conditions and the probability of hot days. To identify potential SST precursors to hot days, we create averages of daily SST anomalies with respect to hot days during the summers of 1982–2015 for a range of lead times. The average SST anomaly pattern synchronous with hot days shows coherent spatial structures in the mid-latitude oceans (Fig. 3), with a zonal tripole pattern in the mid-latitude Pacific composed of anomalies having magnitudes greater than 0.5 °C. Importantly for prediction, the mid-latitude anomalies can be traced back in time, with significant precursor anomalies evident at lead times out to 50 days in the domain between 20–50° N and 145–230° E (green box in Fig. 3). We call the evolving pattern of SST anomalies that precedes hot days the Pacific Extreme Pattern, or PEP_τ, where the subscript indicates the lead time, defined as the number of days prior to a hot day. Although PEP represents an average, this evolving structure can also be seen for individual events (Supplementary Figs 5 and 6).

To quantitatively predict hot days, we compute the spatial covariance between PEP_τ and SST anomalies in the same domain for every combination of τ and day to produce $P_{\tau,t}$. Data from a given year are excluded when calculating PEP for that year (Supplementary Fig. 7), so the relationship between hot days and P is out of sample. Excluding only the predicted year from the analysis is sufficient because P shows no interannual correlation, consistent with previous results showing that mid-latitude summer SST anomalies rarely persist for more than six months[20]. Predictions based on P have significant skill (ROC score ≥ 0.6) for lead times up to 50 days. Skill peaks at a lead time of four days, with a ROC score of 0.72, and generally decreases with increasing lead time (Fig. 2b).

Predictions can also be resolved at the station level (Fig. 2c and Supplementary Fig 8), where a hot day is defined as a temperature anomaly greater than one standard deviation above the mean of summer anomalies for that station. Extreme events at stations in the interior of the domain are better predicted than those on the edges, probably owing to a combination of coastal effects and typical locations of blocking highs (Fig. 3h). Although skill at individual stations is generally lower than that for the region as a whole, the model retains significant skill (ROC score ≥ 0.6) at 74%, 73%, 52% and 14% of stations for lead times of 20, 30, 40 and 50 days, respectively.

Finally, we conduct an out-of-sample examination of skill using pre-1982 data from HadSST3 (ref. 21). Analysis is conducted from

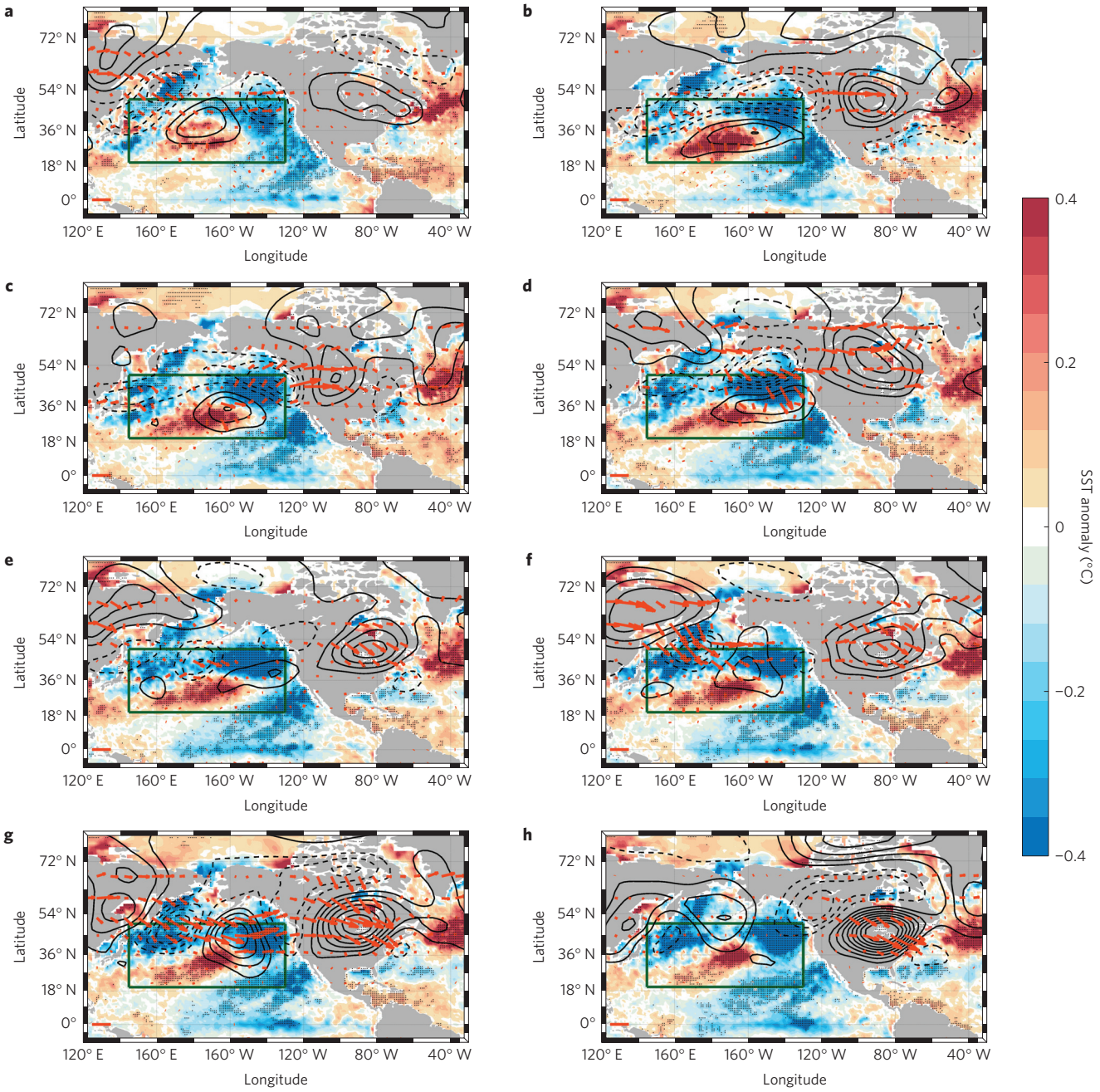

Figure 3 | SST anomalies, geopotential height anomalies and wave activity fluxes associated with eastern US hot days. a–h, Anomalies in SST (colours) and 300 mb height (contours) and wave activity flux[31] (red arrows, 5 m^2 s^{-2} scale bar in lower left of each panel) for lead times of 50 (**a**), 40 (**b**), 30 (**c**), 20 (**d**), 15 (**e**), 10 (**f**), 5 (**g**) and 0 (**h**) days. PEP is defined inside the green box. SST anomalies that are significant at the 0.05 level are stippled. Note that the colour bar for SST anomalies is saturated at ±0.4 °C. Height contours are in 10 m increments, with negative contours dashed and the zero contour suppressed. See Supplementary Methods for a movie containing the same fields at each lead time between 50 and −10 days.

1950 to 1981, coinciding with an earlier interval for which there is still good data coverage in the Pacific (Supplementary Fig. 9). Owing to the temporal resolution of the HadSST3 data set, predictions can be made only on monthly timescales, and we make two-month lead time predictions because PEP changes rapidly as a function of day for τ less than 15 days. The two-month lead time prediction using HadSST3 data has a ROC score of 0.59 (Supplementary Fig. 10), which is significant at the 0.05 level because the null distribution of ROC scores is narrower when using monthly data. Owing to concerns about data availability, we do not extend our analysis to before 1950, and therefore do not assess if our methods could have predicted the extreme heat experienced in the US during the 1930s. Prior work, however, indicates that SST anomalies in the springs of 1934 and 1936 were of opposite sign to those identified here[13]. The prolonged drought and high temperature associated with the Dust Bowl years probably involve mechanisms different from those associated with the synoptic-scale temperature events focused on here[22].

Comparison to other predictors

North Atlantic SSTs have previously been correlated with hot days in the US on seasonal timescales[13], and we find significant North Atlantic SST anomalies in our composites across a range of lead times (Fig. 3). We perform the same analysis applied to the PEP

domain, but using the region identified in ref. 16 (37.5–52.5° N, 302.5–317.5° E) to test if Atlantic anomalies are also predictive of hot days, rather than simply correlated with them. ROC scores for the Atlantic-based predictions are insignificant (ROC < 0.6) for all lead times between 0 and 50 days, averaging 0.53 (Supplementary Fig. 11a,c), leading us to focus exclusively on Pacific SST anomalies.

It is also relevant to consider if more established modes of variability permit for predictability of eastern US hot days. To compare to these standard modes, $P_{\tau,t}$ is averaged for lead times of 15–45 days across summer days to obtain a single value for each year. The resulting values explain 52% of the interannual variance (p-value < 0.01) of the summer median of T95. In contrast, none of the monthly values for January through August of the indices associated with the Pacific North America pattern, the North Atlantic Oscillation, the Pacific Decadal Oscillation, the Northern Annular Mode, and the El Niño–Southern Oscillation explain more than 9% of the interannual T95 variance, and all correlations are insignificant (p-value > 0.05).

The lack of skillful predictions based on tropical indices is consistent with minimal tropical influence on the Northern Hemisphere mid-latitudes during summer[23]. Whereas a causal relationship between tropical Pacific sea surface temperature anomalies and Northern Hemisphere weather is well established for winter conditions[24,25], the lack of a similar relationship for summer[26,27] can be understood from seasonal differences in both the magnitude of tropical anomalies and the mean atmospheric circulation[28,29]. It follows that daily-resolution tropical SSTs also confer no significant predictive skill for eastern US hot days (Supplementary Fig. 11b,d).

Links between PEP, circulation, and precipitation

The co-evolution of Pacific SST anomalies and atmospheric circulation provides insight into the origin of PEP's predictive skill. At a lead time of 50 days, PEP is characterized by an enhancement of the background climatological gradient in SST in the Pacific (Fig. 3a). Positive (negative) atmosphere-to-ocean heat fluxes are co-located with positive (negative) SST anomalies, suggesting the role of atmosphere–ocean heat fluxes in the intensification of SST anomalies over the next 40 days (Fig. 3a–f and Supplementary Fig. 12). Such intensification of SST anomalies may result from the integrated effects of essentially random atmospheric perturbations[30].

A pattern of high pressure is observed downstream from the anomalously warm SSTs in the centre of the PEP domain, both near the surface and aloft (Fig. 3 and Supplementary Fig. 13). Atmospheric simulations indicate that such barotropic pressure anomalies in the central Pacific have a large influence on atmospheric circulation in the US during summer[29], and we find a corresponding tendency towards high pressure and low precipitation over the eastern US. For example, SPI in the eastern US is more than 0.8 standard deviations below average when P_τ averaged across lead times of 15–45 days is above one standard deviation (Supplementary Fig. 14). These results suggest that PEP derives part of its predictive skill through a relationship with eastern US precipitation anomalies.

PEP changes rapidly, beginning at lead times of roughly 15 days, with the establishment of a low-pressure pattern upstream from the central Pacific high. The resulting cyclone–anticyclone pair is associated with southerly wind anomalies and heat fluxes that precede the transformation of the Pacific anomaly pattern from a meridional dipole to a zonal tripole (Supplementary Figs 12 and 15). This tripole is in quadrature with atmospheric circulation anomalies and is associated with an eastward-propagating wave activity flux[31] that converges over the eastern US (Fig. 3). Northward travel of wave activity across the Rocky Mountains suggests that orography may influence the phase of the disturbance. In the final few days before a hot day, the high over the eastern US grows rapidly (Fig. 3g,h), where associated stationary dry and clear conditions increase the radiative heating of the surface.

A similar mid-latitude atmospheric circulation pattern was shown to predict US heat waves up to 15 days in advance in an atmospheric simulation and reanalysis, although results were statistically significant only when assessing a very long model simulation[32]. We suggest that our focus on a clustered region of the US aids in identifying statistically significant precursors to hot weather.

Hindcasting the summer of 2012

To illustrate specific predictions that could be made using PEP, we present a case study for the summer of 2012 at a 40-day lead time. See Supplementary Figs 17–50 for the values of P used to make predictions for all years in 1982–2015, and the corresponding T95.

The extraordinarily hot summer of 2012 was not well predicted: on 17 May, the Climate Prediction Center's seasonal forecast was for normal June–August temperatures for the northeast and midwest, and a 33–40% chance of above normal temperatures for the southeast[33]. Instead, the summer featured three major heat events, beginning on 25 June, 16 July and 29 July (Fig. 1). Increased likelihood of these events was indicated by PEP as early as 15 May, when the value of $P_{\tau=40}$ exceeded the 80th percentile, indicating just over a threefold increase in the probability of hot days around 24 June. PEP intensified over time and, fifteen days later, $P_{\tau=40}$ indicated an almost 3.5 times increase in the odds of hot days around 9 July (Supplementary Fig. 4b). Eastern US temperatures decreased to slightly below average after 9 August. Although this shift was not predicted at 40-day lead times, $P_{\tau=20}$ was negative and decreasing by 21 July, predicting an end to the hot conditions (Supplementary Fig. 47).

We also examine predictions that could have been made using $SPI_{\tau=30}$ for 2012. A 30-day lead time is chosen such that skill is comparable to predictions using $PEP_{\tau=40}$ (Fig. 2). Although the early spring of 2012 was very dry, precipitation deficits were largely erased by a strong Nor'easter that hit the eastern US in mid-April, leading to positive values of SPI through 15 May. But from 25 May to 5 June, SPI was in the lower 20th percentile, associated with a 3.5-fold increase in the probability of hot days from 24 June to 5 July. SPI then decreased further, dropping into the lower 10th percentile by 16 June, associated with over a fourfold increase in the probability of hot days for 6–15 July (Supplementary Fig. 4a). Below-average temperatures at the end of the summer would not have been predicted by SPI, which remained negative until September.

The results presented here link two previously identified precursors to hot days—precipitation deficits[9] and an anomalous atmospheric wave train[32]—with the occurrence of PEP. Predictions are cross-validated, at daily resolution, and assessed using a skill metric that accounts for true and false positives, making the results applicable to real-time predictions. The identification of predictive skill at a seven-week lead time is an important advance over current seasonal forecast models that tend to underpredict the probability of extremes[8], although full comparison between our results and seasonal forecast models should also account for differences in forecast update times and prediction horizons. Long-lead predictions such as those presented here could be applied to reducing the impacts of extreme heat through advance warning[6], although there remain numerous avenues for further work, including combining PEP- and SPI-based predictions in an optimal manner. More broadly, few studies have explored the influence of the mid-latitude ocean on the atmosphere during the summer, and it would be useful to better determine whether the ocean forces, feed backs on, or simply acts as a passive recorder of atmospheric anomalies in the months preceding hot weather. Better determining the role of boundary conditions on the summer atmospheric circulation should allow for further increases in predictive skill.

Methods

Methods and any associated references are available in the online version of the paper.

Received 7 December 2015; accepted 24 February 2016; published online 28 March 2016

References

1. Smith, A. B. & Katz, R. W. US billion-dollar weather and climate disasters: data sources, trends, accuracy and biases. *Nat. Hazards* **67,** 387–410 (2013).
2. Dole, R. *et al*. Was there a basis for anticipating the 2010 Russian heat wave? *Geophys. Res. Lett.* **38,** L06702 (2011).
3. Battisti, D. S. & Naylor, R. L. Historical warnings of future food insecurity with unprecedented seasonal heat. *Science* **323,** 240–244 (2009).
4. Kovats, R. S. & Hajat, S. Heat stress and public health: a critical review. *Annu. Rev. Public Health* **29,** 41–55 (2008).
5. Meehl, G. A. & Tebaldi, C. More intense, more frequent, and longer lasting heat waves in the 21st century. *Science* **305,** 994–997 (2004).
6. Ebi, K. L. & Schmier, J. K. A stitch in time: improving public health early warning systems for extreme weather events. *Epidemiol. Rev.* **27,** 115–121 (2005).
7. Pepler, A. S., Díaz, L. B., Prodhomme, C., Doblas-Reyes, F. J. & Kumar, A. The ability of a multi-model seasonal forecasting ensemble to forecast the frequency of warm, cold and wet extremes. *Weath. Clim. Extremes* **9,** 68–77 (2015).
8. Luo, L. & Zhang, Y. Did we see the 2011 summer heat wave coming? *Geophys. Res. Lett.* **39,** L09708 (2012).
9. Mueller, B. & Seneviratne, S. I. Hot days induced by precipitation deficits at the global scale. *Proc. Natl Acad. Sci. USA* **109,** 12398–12403 (2012).
10. Namias, J. Anatomy of Great Plains protracted heat waves (especially the 1980 US summer drought). *Mon. Weath. Rev.* **110,** 824–838 (1982).
11. Namias, J. Spring and summer 1988 drought over the contiguous United States—Causes and prediction. *J. Clim.* **4,** 54–65 (1991).
12. Lyon, B. & Dole, R. M. A diagnostic comparison of the 1980 and 1988 US summer heat wave-droughts. *J. Clim.* **8,** 1658–1675 (1995).
13. Donat, M. G. *et al*. Extraordinary heat during the 1930s US Dust Bowl and associated large-scale conditions. *Clim. Dynam.* **46,** 413–426 (2016).
14. Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E. & Houston, T. G. An overview of the global historical climatology network-daily database. *J. Atmos. Ocean. Technol.* **29,** 897–910 (2012).
15. Smoyer, K. E., Rainham, D. G. & Hewko, J. N. Heat-stress-related mortality in five cities in Southern Ontario: 1980–1996. *Int. J. Biometeorol.* **44,** 190–197 (2000).
16. McKee, T. B., Doesken, N. J. & Kleist, J. The relationship of drought frequency and duration to time scales. In *Proc. 8th Conf. Appl. Climatol.* Vol. 17**,** 179–183 (American Meteorological Society Boston, 1993).
17. Peterson, W. W., Birdsall, T. G. & Fox, W. C. The theory of signal detectability. *Trans. IRE Prof. Group Inf. Theory* **4,** 171–212 (1954).
18. Stephenson, D., Casati, B., Ferro, C. & Wilson, C. The extreme dependency score: a non-vanishing measure for forecasts of rare events. *Meteorol. Appl.* **15,** 41–50 (2008).
19. Barnston, A. G. & Mason, S. J. Evaluation of IRI's seasonal climate forecasts for the extreme 15% tails. *Weath. Forecast.* **26,** 545–554 (2011).
20. Deser, C., Alexander, M. A. & Timlin, M. S. Understanding the persistence of sea surface temperature anomalies in midlatitudes. *J. Clim.* **16,** 57–72 (2003).
21. Kennedy, J., Rayner, N., Smith, R., Parker, D. & Saunby, M. Reassessing biases and other uncertainties in sea surface temperature observations measured *in situ* since 1850: 1. Measurement and sampling uncertainties. *J. Geophys. Res.* **116** (2011).
22. Cook, B. I., Miller, R. L. & Seager, R. Amplification of the North American Dust Bowl drought through human-induced land degradation. *Proc. Natl Acad. Sci. USA* **106,** 4997–5001 (2009).
23. Yun, K.-S., Kim, S.-Y., Ha, K.-J. & Watanabe, M. Effects of subseasonal basic state changes on Rossby wave propagation during northern summer. *J. Geophys. Res.* **116,** D24102 (2011).
24. Alexander, M. A. *et al*. The atmospheric bridge: the influence of ENSO teleconnections on air-sea interaction over the global oceans. *J. Clim.* **15,** 2205–2231 (2002).
25. Higgins, R., Kim, H. & Unger, D. Long-lead seasonal temperature and precipitation prediction using tropical Pacific SST consolidation forecasts. *J. Clim.* **17,** 3398–3414 (2004).
26. Rajagopalan, B., Cook, E., Lall, U. & Ray, B. K. Spatiotemporal variability of ENSO and SST teleconnections to summer drought over the United States during the twentieth century. *J. Clim.* **13,** 4244–4255 (2000).
27. Yang, X. & DelSole, T. Systematic comparison of ENSO teleconnection patterns between models and observations. *J. Clim.* **25,** 425–446 (2012).
28. Webster, P. J. & Holton, J. R. Cross-equatorial response to middle-latitude forcing in a zonally varying basic state. *J. Atmos. Sci.* **39,** 722–733 (1982).
29. Newman, M. & Sardeshmukh, P. D. The impact of the annual cycle on the North Pacific/North American response to remote low-frequency forcing. *J. Atmos. Sci.* **55,** 1336–1353 (1998).
30. Frankignoul, C. & Hasselmann, K. Stochastic climate models, Part II: application to sea-surface temperature anomalies and thermocline variability. *Tellus* **29,** 289–305 (1977).
31. Takaya, K. & Nakamura, H. A formulation of a wave-activity flux for stationary Rossby waves on a zonally varying basic flow. *Geophys. Res. Lett.* **24,** 2985–2988 (1997).
32. Teng, H., Branstator, G., Wang, H., Meehl, G. A. & Washington, W. M. Probability of US heat waves affected by a subseasonal planetary wave pattern. *Nature Geosci.* **6,** 1056–1061 (2013).
33. CPC Monthly & Seasonal Forecast Archive (National Weather Service Climate Prediction Center, accessed 3 December 2015); http://www.cpc.ncep.noaa.gov/products/archives/long_lead/llarc.ind.php

Acknowledgements

The authors acknowledge funding from the NSF GRFP, NASA NESSF, NCAR ASP, and NSF grant 1304309, and thank B. Farrell, C. Wunsch, C. Bitz, C. Deser, D. Schrag, D. Battisti, J. Mitrovica, M. Cane, P. Hassanzadeh and Z. Kuang for their valuable comments.

Author contributions

The authors contributed equally in designing the study and contributing analysis tools, K.A.M. and A.R. analysed data, and K.A.M. led the writing.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to K.A.M.

Competing financial interests

The authors declare no competing financial interests.

Methods

Data. Summer temperature observations are daily maxima from the Global Historical Climatology Network-Daily database (GHCND; ref. 14). We rely on weather station observations of near-surface daily maximum air temperature[14], as opposed to reanalyses that infer near-surface conditions[34] without necessarily ingesting station data[35].

The analysis is confined to the 60 warmest days of summer based on the average climatology across US weather stations (24 June–22 August on a non-leap year, and 23 June–21 August on a leap year). These climatologically warmest days of summer are when further warming may be expected to have the greatest implications for health and crops. Records from individual weather stations are included only if they have at least 80% coverage during June, July and August for at least 80% of the years considered in the analysis. There are 1613 stations in the eastern US that fulfil these requirements for the 1982–2015 period, and 1092 for the longer 1950–2015 period. Changing the data coverage requirement to focus specifically on the 60 hottest days of summer has essentially no effect on the results, because >99.5% of the stations that pass the coverage requirement for June, July and August also pass the requirement for peak summer.

Daily SST data are from the NOAA OI SST2 data set[36], which is available at daily resolution beginning in September 1981. SST data are re-gridded from 1/4° to 1° spatial resolution. Daily geopotential height and wind fields beginning in 1979 are from the NCEP-DOE II reanalysis[37] at 2.5° resolution. Atmosphere–ocean fluxes from 1985 to 2009 are from the OAFlux project[38] at 1° resolution. Monthly SST data are from the 5° resolution HadSST3 data set[21], which uses only *in-situ* measurements from ships and buoys. All analyses are in terms of anomalies that are calculated by removing the first three annual harmonics, and then removing a linear trend.

Clustering and area weighting. The region of study is defined by grouping stations together using hierarchical agglomerative clustering that minimizes the average intra-cluster Jaccard distance[39] between stations. The Jaccard distance is modified to permit the use of missing data, so that the number of positive matches is normalized only by the number of possible comparisons between the two time series and is, therefore, symmetric and non-negative. The comparison variable is an indicator function of whether or not a station has a temperature above the 95th percentile of its peak summer climatology on the same day as another station. The median fraction of available data across pairs of stations is 0.91. The algorithm incorporates no geographic information, but produces spatially distinct groups of stations (Supplementary Fig. 1). We use hierarchical agglomerative clustering rather than the k-means method because of our choice of distance metric, and because k-means clustering tends to produce circular clusters of similar size, which are not necessarily representative of the underlying data set[40].

Station areas are calculated using a spherical Voronoi tessellation[41] that determines the area around a given station such that all points in the area are closer to that station than any other. Areas are truncated at the coast, but with a 4 km buffer zone extending outside the high-resolution USGS coastline database in order to ensure coastal stations are not excluded from the analysis domain. A similar buffer is applied to the interior of the Great Lakes. Smaller bodies of water are treated as contributing to the land surface fraction. Station areas are capped at the 98th percentile of their distribution to prevent undue influence of stations that are on the edge of the domain.

Precipitation data and the standardized precipitation index. Precipitation data is from the Climate Prediction Center gauge-based gridded data set, and anomalies are calculated for the region co-located with the stations in the eastern US cluster. Precipitation deficits are quantified using the standardized precipitation index (SPI; ref. 16), which serves the dual purpose of integrating precipitation anomalies over a pre-specified time period, and mapping the distribution of integrated anomalies to a normal distribution. We calculate the SPI using integration periods of 30, 60, 90 and 180 days, but focus on the 30-day integration period in the main text because it is the most skillful predictor of hot days. Results using other integration periods are shown in Supplementary Fig. 16.

Hot days and heat events. Consecutive hot days are grouped into heat events, defined to occur if there are two or more hot days in a row (Fig. 1a). Heat events that are less than three days apart from each other are grouped into a single event, leading to a total of 64 heat events observed during the study period. Heat events are associated with their starting dates. Skill is slightly greater for predicting hot days than heat events, perhaps because the most intense heat events during the study period tend to be both long and well predicted, and are thus counted multiply as successful predictions of hot days.

True and false positive rates. The true positive rate is defined as the number of correct predictions of a hot day (true positives) normalized by the total number of hot days. The false positive rate is defined as the number of incorrect predictions of a hot day (false positives) normalized by the total number of non-hot days.

Wave activity fluxes. Wave activity fluxes (WAFs) are calculated following the methods of Takaya and Nakamura[31], which generalizes Plumb fluxes[42] to allow for a zonally variable mean state. We use the summer (JJA) average fields for the mean state, with no further smoothing, and calculate the WAFs for the anomaly composites at each lead time. The z300 composites (Fig. 3) indicate a Rossby wave train that is largely stationary, as required for the WAF calculation.

Pacific Extreme Pattern and predictions. The SST pattern associated with eastern US hot weather is termed the Pacific Extreme Pattern (PEP) and is calculated by compositing daily SST anomalies with respect to hot days at a range of lead times. The 'Extreme' in PEP refers to the association with extreme heat in the eastern US, not that the SST anomalies are necessarily extreme themselves. The similarity between an observed SST anomaly and PEP is used as a predictor for hot days. Quantitatively, we calculate the spatial covariance between SST anomalies and PEP_τ for each day and lead time, $P^*_{\tau,t} = \mathrm{COV}(\mathrm{PEP}_\tau, \mathrm{SST}'_t)$. $P^*_{\tau,t}$ is then normalized to have unit variance for each lead time, yielding the normalized index relied on for prediction, $P_{\tau,t}$. Note that both subscripts are generally omitted throughout the text. To allow for the expected variability in the evolution of a given atmospheric circulation pattern, all predictions for lead times greater than 30 days are assessed in terms of predicting a hot day within a 7-day window centred on the day of prediction.

Significance assessment. All significance estimates are based upon creating a null distribution of the quantity of interest using a block bootstrap with a block size of one year. Specifically, the time series of the predictand (either hot days or the starting date of a heat event) is shuffled 10,000 times, preserving the intra-seasonal ordering, to produce a surrogate predictand time series. Then, the original analysis is performed, but using the surrogate predictand rather than the actual one. An SST anomaly is designated as significant if its magnitude is greater than 95% of the surrogate-based SST anomalies. Similarly, a ROC score that is greater than or equal to 0.6 is found to be significant for predictions of daily-resolved hot days using either *P* or SPI. Owing to the different temporal characteristics of the predictand, ROC scores for predicting the start date of heat events are significant when they are greater than 0.56. ROC scores for monthly-resolved predictions using HadSST3 are significant when they are greater than 0.58.

Data and code availability. All data used in the analysis are publicly available. NCEP-DOE II reanalysis fields for geopotential height and 10-m zonal and meridional winds are available at http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis2.html. Daily SST data from NOAA OI SST V2 are available at ftp://ftp.cdc.noaa.gov/Datasets/noaa.oisst.v2.highres. Monthly SST data from HadSST3, version 3.1.1.0 are available at http://www.metoffice.gov.uk/hadobs/hadsst3/data/download.html. Ocean–atmosphere heat flux data from the WHOI OAFlux Project are available at http://oaflux.whoi.edu/heatflux.html. Precipitation data from the CPC Unified Gauge-Based Analysis of Daily Precipitation are available at http://www.esrl.noaa.gov/psd/data/gridded/data.unified.daily.conus.html. Daily temperature data are from the Global Historical Climatology Network-Daily database, available from http://www1.ncdc.noaa.gov/pub/data/ghcn/daily.

Monthly values for the Pacific North America pattern, North Atlantic Oscillation, and El Niño–Southern Oscillation indices are from the Climate Prediction Center, and are available at http://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/norm.pna.monthly.b5001.current.ascii, http://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/norm.nao.monthly.b5001.current.ascii, and http://www.cpc.ncep.noaa.gov/data/indices/sstoi.indices, respectively. Monthly values for the Pacific Decadal Oscillation and Northern Annular Mode indices are from the Joint Institute for the Study of the Atmosphere and Ocean, and are available at http://jisao.washington.edu/pdo/PDO.latest and http://research.jisao.washington.edu/analyses0302/#data, respectively.

See https://github.com/karenamckinnon/PEP.git for code and formatted data that allow for reproduction of the results depicted in the main text figures.

References

34. Hofstra, N., New, M. & McSweeney, C. The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data. *Clim. Dynam.* **35,** 841–858 (2010).
35. Simmons, A. *et al*. Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP/NCAR analyses of surface air temperature. *J. Geophys. Res.* **109,** D24115 (2004).
36. Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C. & Wang, W. An improved *in situ* and satellite SST analysis for climate. *J. Clim.* **15,** 1609–1625 (2002).
37. Kanamitsu, M. *et al*. NCEP–DOE AMIP-II Reanalysis (R-2). *Bull. Am. Meteorol. Soc.* **83,** 1631–1643 (2002).

38. Jin, X., Yu, L. & Weller, R. A. *Multidecade Global Flux Datasets from the Objectively Analyzed Air-sea Fluxes (OAFlux) Project: Latent and sensible heat fluxes, ocean evaporation, and related surface meteorological variables* OAFlux Project Technical Report OA-2008-01 (Woods Hole Oceanographic Institution, 2008).
39. Jaccard, P. The distribution of the flora in the alpine zone. *New Phytol.* **11,** 37–50 (1912).
40. Jain, A. K. Data clustering: 50 years beyond K-means. *Pattern Recognit. Lett.* **31,** 651–666 (2010).
41. Renka, R. J. Algorithm 772: STRIPACK: Delaunay triangulation and Voronoi diagram on the surface of a sphere. *ACM Trans. Math. Softw.* **23,** 416–434 (1997).
42. Plumb, R. Eddy fluxes of conserved quantities by small-amplitude waves. *J. Atmos. Sci.* **36,** 1699–1704 (1979).