nature climate change

LETTERS

https://doi.org/10.1038/s41558-019-0666-7

Climate change now detectable from any single day of weather at global scale

Sebastian Sippel[1,2,3]*, Nicolai Meinshausen[2], Erich M. Fischer[1], Enikő Székely[4] and Reto Knutti[1]

For generations, climate scientists have educated the public that 'weather is not climate', and climate change has been framed as the change in the distribution of weather that slowly emerges from large variability over decades[1–7]. However, weather when considered globally is now in uncharted territory. Here we show that on the basis of a single day of globally observed temperature and moisture, we detect the fingerprint of externally driven climate change, and conclude that Earth as a whole is warming. Our detection approach invokes statistical learning and climate model simulations to encapsulate the relationship between spatial patterns of daily temperature and humidity, and key climate change metrics such as annual global mean temperature or Earth's energy imbalance. Observations are projected onto this relationship to detect climate change. The fingerprint of climate change is detected from any single day in the observed global record since early 2012, and since 1999 on the basis of a year of data. Detection is robust even when ignoring the long-term global warming trend. This complements traditional climate change detection, but also opens broader perspectives for the communication of regional weather events, modifying the climate change narrative: while changes in weather locally are emerging over decades, global climate change is now detected instantaneously.

"Climate is what you expect, weather is what you get"[8]. This proverbial distinction between weather and climate has served for a long time in public and media discourse as a starting point for scientists to distinguish day-to-day weather variability against the long-term forced response of climate change. The 'weather versus climate paradigm' is therefore typically invoked to explain phenomena that may be perceived as contradictory, such as cold winter days in a warming climate. However, Earth's climate is in fact variable on a broad spectrum of timescales[9]. Variability on weather timescales integrates to produce internal variability on 'climate' timescales[10]. It is therefore well established that, locally or regionally, irreducible uncertainties related to internal climate variability might mask forced climate changes in temperature or hydro-meteorological variables even for decades to come[5,6].

Detection and attribution (D&A) studies aim to identify externally forced signals in the observed climate record[7] such as in surface air temperature[4], humidity[11,12] or the seasonal cycle of temperature[13]. The confidence in the detection of a key climate change metric, that is the 40-yr trend in annual mean tropospheric temperature, is very high and has exceeded a 5σ detection threshold recently[14]. Nonetheless, it is often argued that long-term trends are needed to detect a signal. Here we test whether externally forced climate change can be detected on shorter timescales, or even an arbitrary day in the observed record, if detection is based on a global spatial pattern.

We start with a simple example to illustrate the difference in warming experienced locally and globally (Fig. 1). The past decade (2009–2018) has been on average 0.7 °C warmer than an earlier period (1951–1980). Locally, deseasonalized daily temperature anomalies fluctuate due to internal weather-related variability with a magnitude of up to 30 °C (Fig. 1a,c), depending on region and season. This substantial variation implies that despite an overall warmer climate, cold anomalies or even cold records can still occur and are to be expected[15]. However, at the global scale, weather-related variations tend to cancel out, and variations in deseasonalized global daily temperature anomalies are substantially reduced (Fig. 1b,d). Consequently, mean warming of about 0.7 °C corresponds to a change of about 2.0σ globally, but only to around 0.25σ locally. This difference is consistent with the occurrence of global heat records in recent years despite large variability regionally[16]. Overall, the example anticipates that 'global weather', even in the form of a naive average, carries important although not the optimal information about climate change.

Formal D&A of externally forced signals in Earth's observed climate relies on spatial patterns, so-called 'fingerprints', that encapsulate expected physical climate responses to external forcing[7]. We extend an established fingerprint detection method[11,13,17] by incorporating a regression method that improves the separation of signal and noise based on daily data (see Methods and Extended Data Fig. 1). First, we define two key climate change metrics, annual global mean temperature (AGMT) and a decadal average of Earth's energy imbalance (EEI), that serve as target variables for detection. AGMT characterizes climate warming from a surface perspective, used in policy assessments[18,19], while EEI characterizes the state of climate change from a more physically motivated energy balance perspective[20,21]. Second, we train regularized linear regression models, using ridge regression, that predict each of the two targets from daily spatial patterns of surface temperature and/or humidity for each month and based on model simulations in the Coupled Model Intercomparison Project Phase 5 (CMIP5) multi-model archive driven by external forcing. This step yields fingerprints as maps of regression coefficients that encapsulate the relationship between global patterns of daily weather and each target metric while reducing 'noise' in regions with large internal variability or model disagreement. Finally, and similarly to established detection methods[11,13,17], we project observations onto the model-derived fingerprints to obtain a prediction of each target metric. We then assess whether externally forced climate change is detected by testing against the null hypothesis that the predicted test statistic is indistinguishable from natural variability.

Overall, we estimate AGMT from any given day with a root mean squared error (RMSE) of 0.09 °C (Pearson correlation > 0.99) across

[1]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland. [2]Seminar for Statistics, ETH Zurich, Zurich, Switzerland. [3]Norwegian Institute of Bioeconomy Research, Ås, Norway. [4]Swiss Data Science Center, ETH Zurich and EPFL, Lausanne, Switzerland. *e-mail: sebastian.sippel@env.ethz.ch

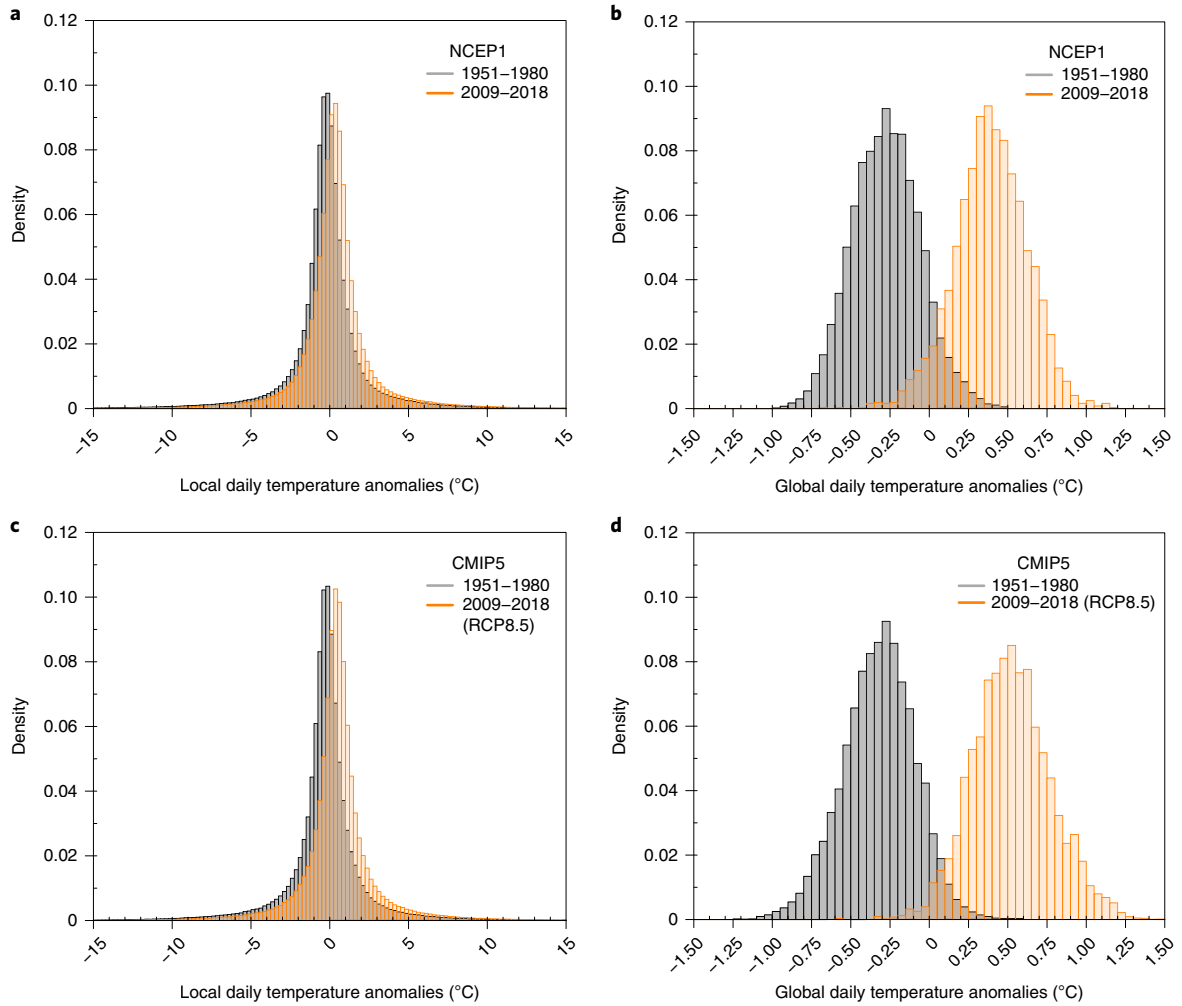


Fig. 1 | Warming of daily temperatures experienced at the local and global scale. a–**d**, The distribution of local (**a**,**c**) and global (**b**,**d**) daily temperatures in the National Centers for Environmental Prediction (NCEP) 1 reanalysis dataset (**a**,**b**) and in the CMIP5 multi-model archive (**c**,**d**). The local histograms are area-weighted on the basis of temperature anomalies from all grid cells and all seasons relative to their respective 1979–2005 mean seasonal cycle.

the CMIP5 archive on models not used for training (Extended Data Fig. 2). The RMSE is about 44% lower than for the naive daily global average used above, and the error is comparable to one year in advance forecasts based on dynamical models[22]. It is important to highlight that model structural uncertainties imply differences in each model's representation of the forced response and internal variability (fingerprint uncertainty analyses in Supplementary Text 2). However, the fingerprints discussed here are extracted via cross-validation to extrapolate well to other models. This yields fingerprints that are robust across the set of CMIP5 models because regions of model disagreement receive small regression coefficients.

To understand the spatial signature of externally forced climate change in daily data, we discuss AGMT fingerprints and their seasonal cycle, which reveals key climatological features (Fig. 2a,b). First, regression coefficients are positive throughout the globe, indicating that the fingerprints pick up a global warming signal. Second, coefficients are larger over the oceans than over land, and larger over the tropics than over mid- or high latitudes. These features illustrate that signal-to-noise ratios of forced climate change in daily data are higher over the oceans than over land, and higher in the tropics than over mid- or high-latitudes, consistent with our physical understanding of forced changes on longer timescales[23,24]. Polar regions do not receive much weight in the fingerprints despite very fast warming rates[25], because of large internal variability. Further, interannual variability due to the El Niño/Southern Oscillation in the central and eastern tropical Pacific that is different and uninformative for long-term changes leaves a distinct imprint on the fingerprints in the form of low values of regression coefficients. In boreal summer, the fingerprints' regression coefficients extend further north than in boreal winter (Fig. 2b), consistent with a reduced Equator-to-pole temperature gradient in summer and, consequently, reduced internal temperature variability in summer in the Northern Hemisphere.

Given the strong signal even in a single-day average (Fig. 1), the results so far mainly reflect the signal of long-term mean warming. However, D&A can be assessed in the absence of global mean changes, leading to confidence in detection based on spatial pattern similarity alone[13]. Hence, we subtract global mean temperature, separately for each day, and extract a second set of fingerprints. The prediction error for daily global temperature increases substantially (RMSE = 0.40 °C), and therefore lends low statistical power for detection. However, climate change manifests itself not only in surface air temperature change but also through changes in total atmospheric heat content[26], where latent heat in the form of atmospheric moisture plays an important role. Therefore, we add specific humidity (that is, the moisture content of a given air parcel (restricted to land values due to coverage)) as a predictor. The joint temperature

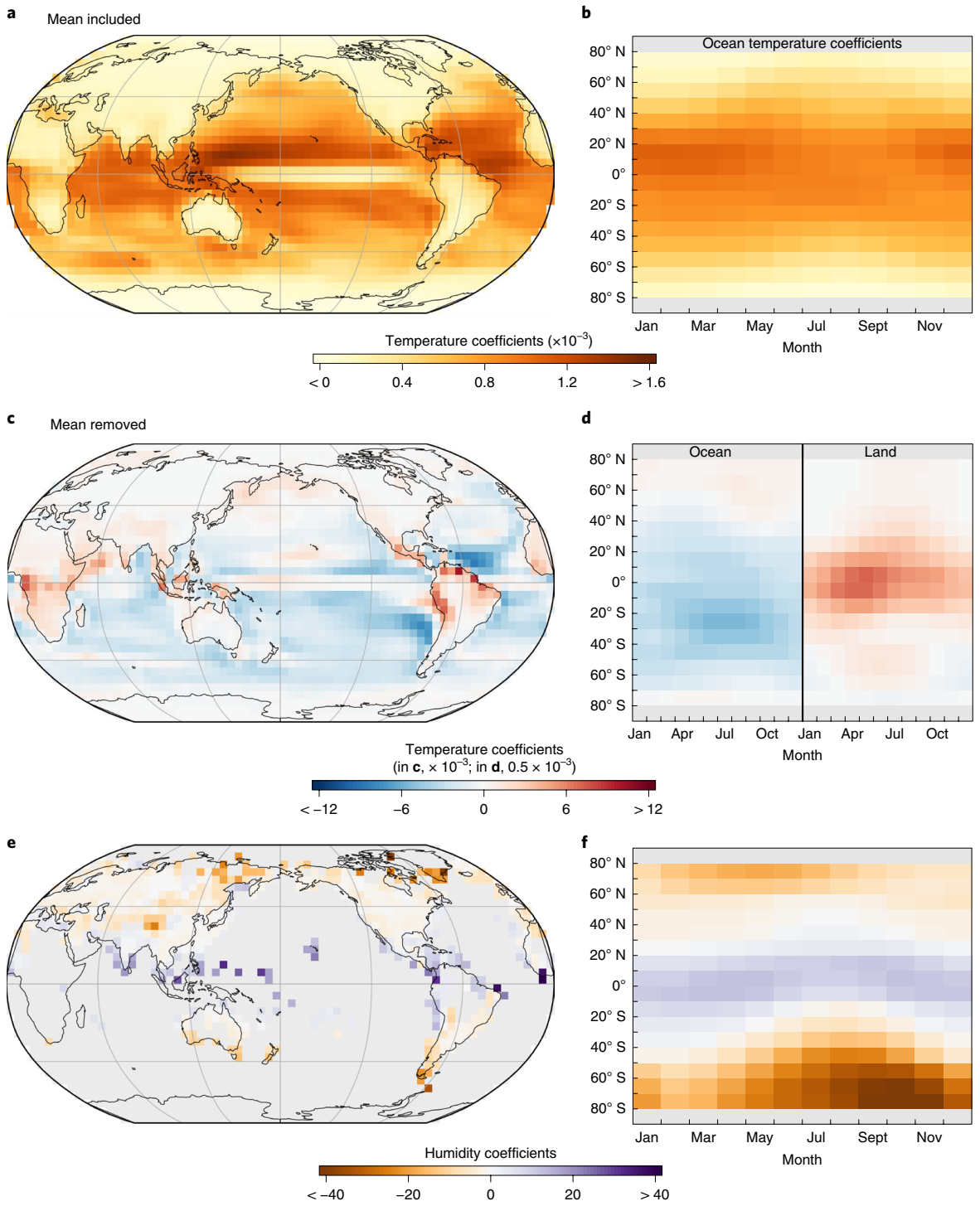


Fig. 2 | The fingerprint of external forcing and its seasonal variation. a,**b**, The annual average fingerprint of external forcing as a map of regression coefficients extracted from the spatial pattern of daily temperatures to predict AGMT, and its seasonal variation along latitudinal bands for ocean grid cells (**b**). **c**–**f**, The annual average fingerprint of external forcing extracted from daily temperatures (**c**) and daily land humidity (**e**) (with gaps in observations in grey) with the global mean signal removed, and its seasonal variation along latitudinal bands for the mean-removed fingerprint of temperature over ocean/land grid cells (**d**), and land humidity (**f**). All fingerprints are extracted from the CMIP5 multi-model archive.

and land humidity fingerprint (both mean-removed) reduces the prediction error of AGMT (RMSE = 0.22 °C). The mean-removed temperature fingerprint relies on a land–sea warming contrast at low and mid-latitudes, and fast warming at high latitudes in summer, both consistent with the physical understanding[23,24] (Fig. 2c,d). Contrasting regression coefficients are particularly confined to geographically neighbouring land and sea regions that experience low but correlated internal variability. The mean-removed humidity fingerprint relies on an Equator-to-pole gradient with positive coefficients over the tropics (corresponding to moistening trends that exceed global mean humidity changes) and negative coefficients at high latitudes (Fig. 2e,f; joint temperature and humidity

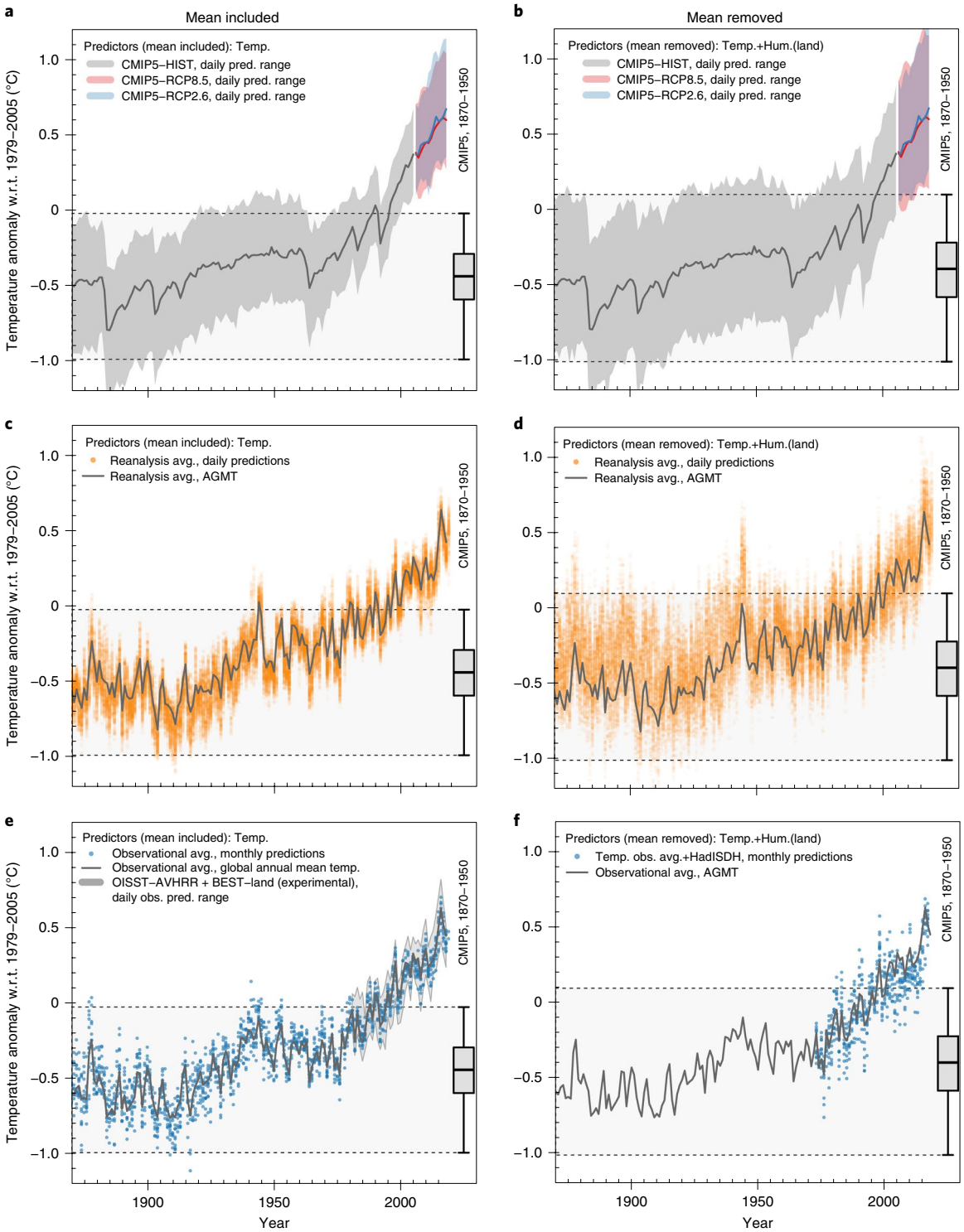


Fig. 3 | AGMT estimates from models, reanalyses and observations. a–f, Time series of AGMT (bold lines) for the CMIP5 multi-model mean (**a**,**b**), reanalysis datasets (**c**,**d**) and observations (**e**,**f**). Estimates of AGMT based on each single day (**a**–**e**, in **e** for daily observations that are experimental and shown for illustration alone, see Methods) and single months (**e**,**f**) are shown. All estimates are derived on the basis of the spatial pattern of daily/monthly temperatures with the mean signal included (**a**,**c**,**e**); and based on the spatial pattern of daily/monthly temperature and humidity with the mean signal removed (**b**,**d**,**f**). The boxplots show the 2.5th to 97.5th percentile of the distribution of the AGMT test statistic for the 1870–1950 historical period in CMIP5 with each model's contribution weighted equally. avg., average; obs., observed; pred., predicted.

fingerprints shown in Supplementary Text 2). Therefore, expected specific humidity increases are larger in warmer background climates than in cooler climates[27], consistent with the Clausius–Clapeyron relationship suggesting an approximately linear percentage increase in atmospheric moisture with warming. Seasonal variations in the humidity fingerprint indicate a northward shift of tropical regression coefficients in boreal summer, along with reduced absolute values of negative coefficients at high latitudes consistent with seasonal

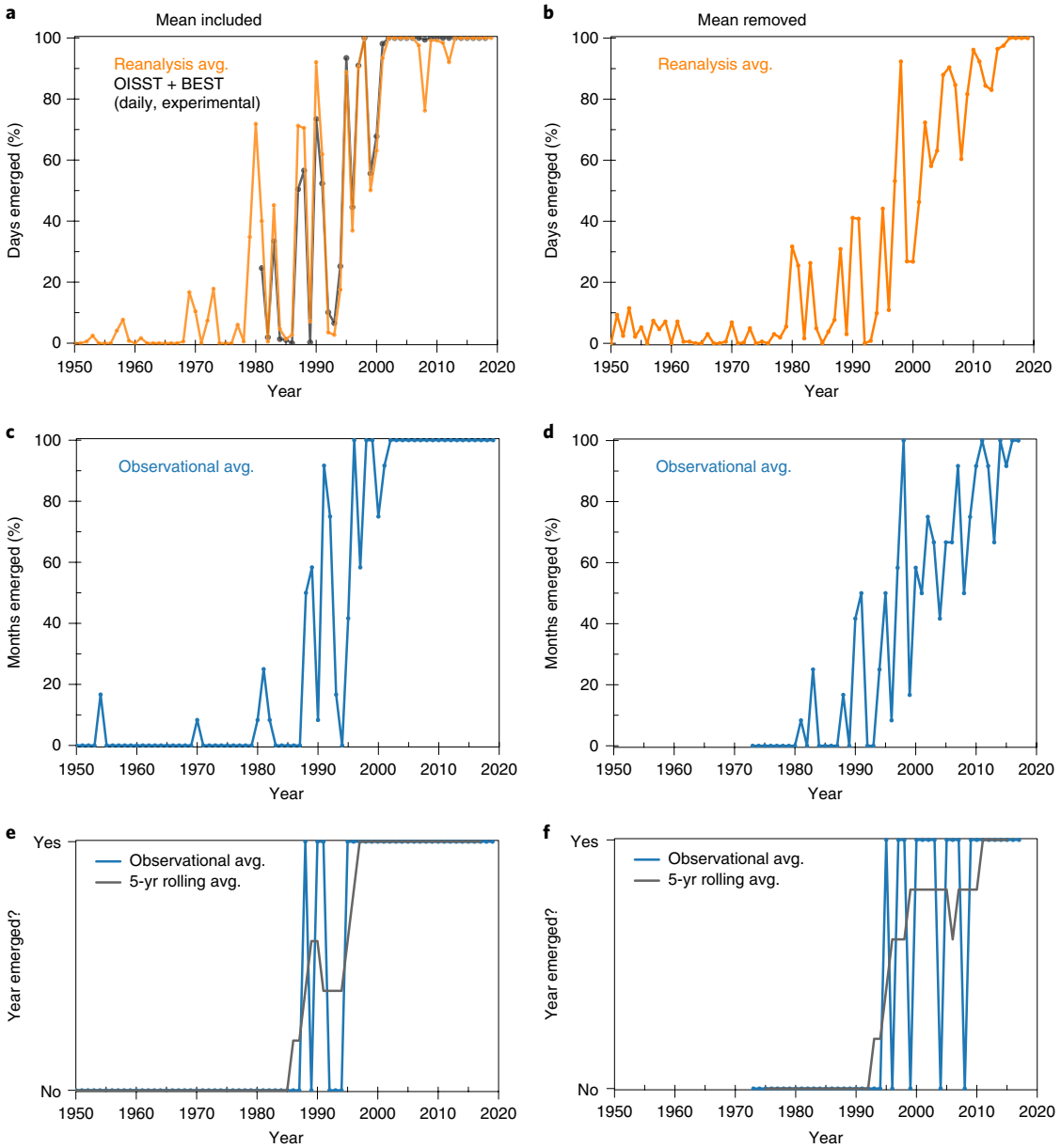


Fig. 4 | Emergence of externally forced climate change in 'global weather' from the noise of natural variability. a–d, The fraction of days in reanalysis datasets (**a**,**b**), daily observations (**a**, experimental and illustrative alone) and months in observations (**c**,**d**) that have emerged from daily natural variability (that is, exceeding the 97.5th percentile in the CMIP5 1870–1950 distribution). The estimates are based on the spatial pattern of temperature with the mean signal included (**a**,**c**,**e**) and the joint spatial pattern of temperature and land humidity with the mean signal removed (**b**,**d**,**f**). **e**,**f**, The emergence of observations based on individual years.

variations in surface climate. We conclude that physically consistent forced climate signals can be picked up from the spatial pattern of daily surface temperature and humidity even if the global mean signal is removed.

We assess whether climate change is detected at short timescales by projecting in situ observations, reanalyses output and CMIP5 model simulations onto the fingerprints. This step yields a time series of the AGMT test statistic for each day in a given year (Fig. 3) that is assessed against a proxy of natural climate variability. Here we use the 2.5th–97.5th percentile range of the test statistic distribution in CMIP5 models in 1870–1950. The proxy is conservative because of forced early twentieth-century warming[28] and exceeds the 'extremely likely' 95% level in Intergovernmental Panel on Climate Change terminology[29].

CMIP5 models predict that forced climate change can be detected at a daily basis from the early 2000s onwards, where the range of daily AGMT estimates emerges from natural variability (Fig. 3a,b). The reanalyses average indicates that since the end of 2001 the majority of days in any given year lies outside natural variability irrespective of whether the global mean trend signal has been removed or not (Fig. 3; 97.7% ('mean included') and 83.6% ('mean removed') of days detected individually from 2001 up to August 2019). Since late March 2012, every single day is detected individually (Fig. 4a, 'mean included'), and all but 3.5% of days in the 'mean removed' case (Fig. 4b). In observations, every month individually is outside natural variability since 2001 with data available up to December 2018 ('mean included'). In an experimental daily observational dataset, every day would be

individually detected since about 2008 (Fig. 3e, 'mean included'). If the mean is removed, only six months (7.3% with humidity data available up to December 2017) fall within natural variability since the end of 2010 (Fig. 3f). As the test statistic varies characteristically across timescales, global climate change is probably detectable also at shorter subdaily or even instantaneous timescales (Extended Data Fig. 3). Overall, the conclusion that a forced climate change is detected with >97.5% confidence in any individual day or month since early 2012 is robust across any of the five reanalysis and observational datasets.

EEI is a second, and arguably more fundamental, key climate change metric that determines how the energy content of the climate system is changing[20,21]. Across the CMIP5 multi-model archive, given a historical forcing trajectory, reasonable decadal EEI estimates can be determined from a single day's spatial pattern of temperature and humidity ($RMSE_{Temp., mean\text{-}incl.} = 0.18\,W\,m^{-2}$, $RMSE_{Temp.+Hum(land), mean\text{-}rem.} = 0.21\,W\,m^{-2}$). Compared to AGMT, EEI fingerprints rely on a slightly narrower latitudinal band of regression coefficients over tropical oceans. This is probably related to optimally capturing the decadal EEI signal while minimizing the influence of interannual or short-term variations. The detection analysis reveals that decadal EEI estimates in models, reanalyses and observations in fact all show recent emergence from natural variability similar to AGMT-based detection (Supplementary Text 4). Although the exact baseline of EEI estimates is uncertain (Methods), EEI estimates of around $+0.53\,W\,m^{-2}$ on average from monthly observations over the last decade since 2009 (spanning $+0.38$ to $+0.71\,W\,m^{-2}$ overall; up from $+10\,W\,m^{-2}$ on average in 1900–1950) are just slightly lower than direct observation-based EEI estimates[20]. Our results therefore indicate that the Earth has taken up energy continuously over the last decade, which will thus continue to drive climate change in the near future.

We have shown, using fingerprints of key climate change metrics, that observations, reanalyses and model simulations agree that climate change is detected now from global weather in a single year, month or even a single day. This result is robust also if detection is based only on the spatial pattern of temperature and land humidity with any global mean signal removed. The aforementioned is a remarkable physical consistency check between models and observations, and allows daily detection even if, for example, the global mean long-term trend is judged to be unreliable due to changes in observing systems. Daily global weather has therefore emerged from the background of natural variability, and the total energy content of the planet keeps increasing. As in other detection studies[13], this result relies on the assumption that climate models do not severely underestimate low-frequency variability, but our results are robust to various combinations of CMIP5 models used to construct the reference distribution (Supplementary Text 3).

Broader implications for climate science include, first, that weather on global scales carries important climate information that may be exploited for short-term D&A. This may complement the study of long-term climate trends, which may be masked locally by internal climate variability[5]. Second, statistical learning offers opportunities to optimally combine multivariate spatiotemporal information to maximize a desired climate signal against internal variability or other factors[30–32]. Third, our approach opens a broader perspective and context for the communication of regional weather events against the backdrop of a warming climate. This allows one to overcome a disconnect between formal fingerprinting studies that rely on observed long-term trends, and event attribution (that is, the assessment of externally forced changes in the magnitude or frequency of short-term (extreme) weather events that interpret model simulations probabilistically[33]). Short-term detection may help to formally assess the role of external forcing on global or sub-global 'weather' up to instantaneous timescales.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41558-019-0666-7.

Received: 6 July 2019; Accepted: 19 November 2019;
Published online: 2 January 2020

References

1. Hasselmann, K. F. On the signal-to-noise problem in atmospheric response studies. In *Joint Conference of Royal Meteorological Society, American Meteorological Society, Deutsche Meteorologische Gesellschaft and the Royal Society* (ed. Shaw, D. B.) 251–259 (Royal Meteorological Society, 1979).
2. Wigley, T. & Jones, P. Detecting CO_2-induced climatic change. *Nature* **292**, 205–208 (1981).
3. Lorenz, E. N. Chaos, spontaneous climatic variations and detection of the greenhouse effect. *Dev. Atmos. Sci.* **19**, 445–453 (1991).
4. Hegerl, G. C. et al. Detecting greenhouse-gas-induced climate change with an optimal fingerprint method. *J. Clim.* **9**, 2281–2306 (1996).
5. Deser, C., Knutti, R., Solomon, S. & Phillips, A. S. Communication of the role of natural variability in future North American climate. *Nat. Clim. Change* **2**, 775–779 (2012).
6. Hawkins, E. & Sutton, R. Time of emergence of climate signals. *Geophys. Res. Lett.* **39**, L01702 (2012).
7. Bindoff, N. L. et al. in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. et al.) Ch. 10 (IPCC, Cambridge Univ. Press, 2013).
8. Lorenz, E. N. *Climate is what you expect* http://eaps4.mit.edu/research/Lorenz/Climate_expect.pdf (1995).
9. Huybers, P. & Curry, W. Links between annual, Milankovitch and continuum temperature variability. *Nature* **441**, 329–332 (2006).
10. Schneider, E. & Kinter, J. An examination of internally generated variability in long climate simulations. *Clim. Dynam.* **10**, 181–204 (1994).
11. Santer, B. D. et al. Identification of human-induced changes in atmospheric moisture content. *Proc. Natl Acad. Sci. USA* **104**, 15248–15253 (2007).
12. Willett, K. M., Gillett, N. P., Jones, P. D. & Thorne, P. W. Attribution of observed surface humidity changes to human influence. *Nature* **449**, 710–712 (2007).
13. Santer, B. D. et al. Human influence on the seasonal cycle of tropospheric temperature. *Science* **361**, eaas8806 (2018).
14. Santer, B. D. et al. Celebrating the anniversary of three key events in climate change science. *Nat. Clim. Change* **9**, 180–182 (2019).
15. Cattiaux, J. et al. Winter 2010 in Europe: a cold extreme in a warming climate. *Geophys. Res. Lett.* **37**, GL044613 (2010).
16. Fischer, E. M., Beyerle, U. & Knutti, R. Robust spatially aggregated projections of climate extremes. *Nat. Clim. Change* **3**, 1033–1038 (2013).
17. Marvel, K. & Bonfils, C. Identifying external influences on global precipitation. *Proc. Natl Acad. Sci. USA* **110**, 19301–19306 (2013).
18. Meinshausen, M. et al. Greenhouse-gas emission targets for limiting global warming to 2 °C. *Nature* **458**, 1158–1162 (2009).
19. IPCC. *Special Report on Global Warming of 1.5 °C* (eds Masson-Delmotte, V. et al.) (WMO, 2018).
20. Trenberth, K. E., Fasullo, J. T. & Balmaseda, M. A. Earth's energy imbalance. *J. Clim.* **27**, 3129–3144 (2014).
21. Von Schuckmann, K. et al. An imperative to monitor Earth's energy imbalance. *Nat. Clim. Change* **6**, 138–144 (2016).
22. Folland, C. K. et al. High predictive skill of global surface temperature a year ahead. *Geophys. Res. Lett.* **40**, 761–767 (2013).
23. Sutton, R. T., Dong, B. & Gregory, J. M. Land/sea warming ratio in response to climate change: IPCC AR4 model results and comparison with observations. *Geophys. Res. Lett.* **34**, GL028164 (2007).
24. Mahlstein, I., Knutti, R., Solomon, S. & Portmann, R. W. Early onset of significant local warming in low latitude countries. *Environ. Res. Lett.* **6**, 034009 (2011).
25. Bekryaev, R. V., Polyakov, I. V. & Alexeev, V. A. Role of polar amplification in long-term surface air temperature variations and modern arctic warming. *J. Clim.* **23**, 3888–3906 (2010).
26. Pielke, R. A., Davey, C. & Morgan, J. Assessing 'global warming' with surface heat content. *Eos* **85**, 210–211 (2004).
27. Dunn, R. J., Willett, K. M., Ciavarella, A. & Stott, P. A. Comparison of land surface humidity between observations and CMIP5 models. *Earth Syst. Dynam.* **8**, 719–747 (2017).
28. Haustein, K. et al. A limited role for unforced internal variability in 20th century warming. *J. Clim.* **32**, 4893–4917 (2019).

29. Mastrandrea, M. D. et al. *Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties* (IPCC, 2010).
30. Barnes, E., Anderson, C. & Ebert-Uphoff, I. An AI approach to determining the time of emergence of climate change. In *Proc. 8th International Workshop on Climate Informatics: CI 2018* (eds Chen, C. et al.) 19–22 (NCAR, 2018).
31. Reichstein, M. et al. Deep learning and process understanding for data-driven Earth system science. *Nature* **566**, 195–204 (2019).
32. Sippel, S. et al. Uncovering the forced climate response from a single ensemble member using statistical learning. *J. Clim.* **32**, 5677–5699 (2019).
33. Otto, F. E. Attribution of weather and climate events. *Annu. Rev. Environ. Resour.* **42**, 627–646 (2017).

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

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Climate change detection method. D&A studies use fingerprints of external forcing and a formal comparison between model simulations and observations to assess whether the influence of external forcing can be detected in Earth's climate[7]. We first briefly describe a standard and widely used detection method[11,13,17,34] and then an extension of the method that incorporates statistical learning techniques based on regularized linear regression for daily climate change detection.
The established method involves three key steps:

First, a fingerprint is extracted from model simulations that encapsulates the spatial (or, for example, vertical) pattern of the physical response to external forcing. For instance, previous D&A studies[13,34] calculate the fingerprint as the leading empirical orthogonal function[35] of the time-evolving multi-model average pattern of forced simulations. In a second step, observations are projected onto the fingerprint to assess whether there is increasing congruency between the observations and the expected response to external forcing (which yields a signal time series, denoted $S(t)$ in previous studies[11,13,17,34]). In other words, this step measures the spatial covariance between the expected physical response to external forcing (determined purely from model simulations), and the observations. Third, statistical inference is based on a scalar signal-to-noise ratio, where it is tested whether the signal $S(t)$, summarized as a trend of a given length L, exceeds that of the noise of L-length trends in long control simulations that feature only internal or natural variability (the null hypothesis). The gist of this detection approach is that a high-dimensional testing problem is reduced to a one-dimensional test statistic. The extraction of the fingerprint captures the expected signal in response to external forcing but disregards (through averaging) whether internal variability ('climate noise') in a given region is high or low.

Here, we incorporate a regularized linear regression model that separates signal and noise in the extraction of the fingerprint (see Extended Data Fig. 1 for an illustration). We use two key climate change metrics as target variables, AGMT (in the year y that corresponds to the day i at which climate change is to be detected) and EEI (as a decadal average in the years before y). These two metrics are key indicators of climate change (AGMT[18,19]; EEI[20,21]) and serve here as target variables (that is, the one-dimensional test statistic) for climate change detection at the daily timescale. The method is conceptually similar to a machine learning approach to fingerprint extraction used recently to determine emergence times of climate change based on yearly data[30]. In a first step, the fingerprint of external forcing is extracted from forced model simulations such that the p-dimensional spatial pattern of daily temperature or humidity is related linearly to one of the target metrics (denoted here, $\mathbf{Y}_{\mathrm{mod}}$) in a regression setting (step (1) in Extended Data Fig. 1):

$$\mathbf{Y}_{\mathrm{mod}} = X_{\mathrm{mod}}\hat{\boldsymbol{\gamma}} + \boldsymbol{\varepsilon} \tag{1}$$

We store the model simulated spatial patterns of daily temperature and/or humidity in an $n \times p$ matrix X_{mod}, where n and p denote the number of training samples (that is, individual days) and the number of spatial predictors, respectively. The vector $\mathbf{Y}_{\mathrm{mod}}$ is of length n and denotes the defined target metric for each time step. The estimated fingerprint $\hat{\boldsymbol{\gamma}}$ is then a p-dimensional vector of spatial regression coefficients. Hence, the fingerprint extraction step maps the high-dimensional spatial patterns of temperature and/or humidity into a one-dimensional test statistic via $\hat{\boldsymbol{\gamma}}$, and the fingerprint therefore encapsulates the physically expected response to external forcing. This step is analogous to the standard detection method described above[11,13,17,34]: instead of extracting empirical orthogonal functions from the time-evolving average across multiple models, we use regularized linear regression in equation (1). The regression problem can be solved, and thus the respective spatial fingerprints extracted, with ridge regression that optimizes the climate change signal while giving lower weight to regions with large internal climate variability at daily timescales. In a second step, we project the observations (X_{obs}, which are fully independent from $\hat{\boldsymbol{\gamma}}$) and model simulations of natural variability (denoted X_{mod^*}, which are not used for fingerprint extraction, see the next paragraph) onto $\hat{\boldsymbol{\gamma}}$ to obtain an estimate of the target climate change metric for each individual time step that will be used as a test statistic (step (2) in Extended Data Fig. 1).

$$\hat{\mathbf{Y}}_{\mathrm{obs}} = X_{\mathrm{obs}}\hat{\boldsymbol{\gamma}} \tag{2}$$

$$\hat{\mathbf{Y}}_{\mathrm{mod}^*} = X_{\mathrm{mod}^*}\hat{\boldsymbol{\gamma}} \tag{3}$$

The fact that model simulations (X_{mod}) used for fingerprint extraction and model simulations (X_{mod^*}) used for estimating the control distribution of the test statistic under natural variability are kept separate is important, because otherwise a bias may result as the training and application of the statistical model would not be independent. In fact, we use the CMIP5 multi-model archive[36] for fingerprint extraction and to estimate the control distribution of the test statistic under natural variability, and we keep entire models (X_{mod} versus X_{mod^*}) separate for these two tasks in a cross-validation setting (similar to an iterative 'perfect-model' approach, specified in detail in the next paragraph). In a final third step (step (3) in Extended Data Fig. 1), we assess whether external forcing can be detected in the observations by testing, separately for any single day, against the distribution of the test statistic under natural variability ($\mathcal{P}[\hat{\mathbf{Y}}_{\mathrm{mod}^*}]$), using AGMT or EEI. The second and third steps of the outlined approach are again analogous to the standard detection method described above[11,13,17,34].

The advantages of this approach include, first, that the spatial pattern response to external forcing is encapsulated in the fingerprint, but regions with large internal variability or where different climate models disagree with each other receive less weight in the fingerprint. In other words, ridge regression optimizes the climate signal against 'climate noise' (internal climate variability) at the daily timescale; a step that is consistent with signal-to-noise optimization in traditional D&A studies[1,4]. Second, estimates of the two key climate change metrics used here (AGMT and EEI), derived from observations and control simulations, serve the purpose of a test statistic, but retain their climatological interpretation. Third, the detection approach in a supervised regression setting (equation (1)) allows one to evaluate the expected prediction performance within an idealised model approach using classical error metrics such as RMSE (see main text).

Statistical learning technique. The linear model employed here to extract the fingerprints in equation (1) is required to solve a high-dimensional linear regression problem. Traditional methods may fail in a high-dimensional setting because regression coefficients are not well constrained[37]. Ridge regression, a statistical learning technique developed for such situations[37,38], guards against overfitting by imposing a penalty on the complexity of the model via shrinkage of regression coefficients. Shrinkage of regression coefficients is based on the sum of squared regression coefficients (known as L2 regularization) and a ridge regression parameter λ that determines the absolute amount of shrinkage. This yields small but non-zero regression coefficients, and regression coefficients that are relatively smoothly distributed among correlated predictors[38]. As climate variables are typically correlated in space, this leads to regression coefficients that tend to be smooth in space (see, for example, Fig. 2) but without imposing any smoothness explicitly. However, the approach presented here is not necessarily confined to ridge regression for fingerprint extraction: full-blown machine learning techniques can be used for tasks analogous to this[30], or constraints based on system understanding to target individual climate forcing agents specifically could be readily implemented[39].

Data pre-processing and fingerprint extraction. Data processing and subsequent model training is set up as follows.

We extract model simulations from the CMIP5 multi-model archive[36] for training. Simulations are extracted for 24 different models from 13 modelling centres that provide daily surface temperature and specific humidity (see Supplementary Table 1). The model data consist of in total 14,100 individual model years available for fingerprint extraction, where three days per month are chosen for training: the 5th, 15th and 25th day of each month (corresponding to 507,600 days). The 14,100 model years are made up of 45 simulations using the historical (1870–2005; that is, in total 6,120 model years) and Representative Concentration Pathway (RCP)8.5 scenario (2006–2100, 4,815 model years), and 39 simulations with the RCP2.6 scenario (2006–2100, 3,705 model years). We regrid all daily model data to a common, regular 5°×5° spatial grid (that is, corresponding to $p = 2{,}592$ grid cells or spatial predictors). Next, we subtract the seasonal cycle from each day i, separately for each model and for each grid cell p, using a 31-d rolling mean seasonal cycle centred on the respective day. The 31-d rolling mean seasonal cycle is estimated from the 1979–2005 reference period. The 1979–2005 reference period is chosen to maximize the overlap between model simulations with historical forcing (that end in 2005) and observations (some of which start only in 1979, see next paragraph), and to ensure an identical processing and hence comparability between model simulations and observations. The target metrics are obtained similarly for each model. AGMT denotes the anomaly of the annual mean spatial average of surface temperatures (in the year that corresponds to day i), and is estimated separately for each model relative to its 1979–2005 average. The second target metric, EEI, denotes the energy imbalance at the top of the atmosphere[20], estimated as a decadal average in the decade before the year that corresponds to day i. Net top-of-atmosphere radiation is sensitive to drift in CMIP5 models[40,41], but the drift is approximately constant in each model[40,41]. Therefore, we implement a standard mean drift correction of net top-of-atmosphere radiation (corresponding to a linear drift in Earth's energy content[40]) by estimating the mean drift in long-term control simulations in each model. Subsequently, we subtract the mean drift from each model's EEI estimates.

Next, we train the regularized linear model to extract the respective fingerprints. For this purpose, an individual statistical model (that is, fingerprint) is trained for each month, because the expected physical response to external forcing changes with the seasonal cycle[13] (for example, Fig. 2). To increase the sample size for training of each individual month, samples from the previous and subsequent month are included in the training step. We implement a standard cross-validation scheme to determine the ridge regression parameter (λ) and to extract the final fingerprints. Cross-validation is standard practice in statistical learning and ensures, by splitting the raw dataset into separate partitions, that model fitting and model validation and selection are performed on different data (to avoid a biased performance evaluation). From a climate science perspective, cross-validation as

implemented here with a 'leave-one-model-out' strategy can be seen as an iterative perfect model approach. First, in an outer loop, we run 13 simulations ($k=13$) for which each individual climate model is iteratively left out as an unseen test set. Second, we determine the ridge regression parameter λ in an inner loop by $k-1$-fold cross-validation. That is, each of the 13 training simulations uses data from $k-1=12$ climate models (that is, one model left out iteratively), where each model is put in a separate fold ('leave-one-model-out cross-validation'). This step ensures that the fingerprints extrapolate well to an unseen model (that is, fingerprints that are robust across the CMIP5 multi-model archive). In cases where modelling centres provide separate variants of a particular model (Supplementary Table 1), these model variants are treated as part of the same model and hence put in the same fold. We ensure that each model used in the training step receives the same weight by subsampling the number of individual days used for training. The tuning parameter λ is then selected in the cross-validation as the most regularized model within one standard error of the minimum mean squared error on the out-of-fold data. This yields a fingerprint ($\hat{\boldsymbol{\gamma}}_k$) as a set of regression coefficients. Third, for each of the 13 simulations (that is, iteratively for each test set model not used in the inner loop), we predict the respective target metric for any given day i. These independent estimates of the target metric are used to estimate prediction errors (that is, the RMSEs discussed in the main text) and 1870–1950 predictions are used as a reference distribution of the test statistic under 'natural variability'. The choice of the 1870–1950 period for the control distribution is conservative, because some of the early-twentieth-century warming may have been externally forced[28] in addition to internal variability. Last, the final fingerprint ($\hat{\boldsymbol{\gamma}}$, shown in Fig. 2), for any given month, is obtained by averaging over the 13 separate fingerprints ($\hat{\boldsymbol{\gamma}}_k$) of the outer loop.

This set-up is repeated for both climate change target metrics (AGMT and EEI), and for different sets of predictors. The different sets of predictors include the daily spatial pattern of (1) temperature ('Temp.'); (2) temperature, where the global mean temperature of each day is removed from each grid cell ('Temp., mean removed); (3) specific humidity, using the mask of an observational dataset[42] available over land ('Hum. (land)'); (4) specific humidity over land areas with the global mean removed at each time step ('Hum. (land), mean removed'); and (5) combined mean-removed temperature and land-only humidity ('Temp. + Hum. (land), mean removed'; that is, (2) and (4) combined). Supplementary Table 2 and Supplementary Table 4 provide an overview of the prediction performance for all sets of predictions for AGMT and EEI prediction, respectively. As the number of available model runs differs by model (Supplementary Table 1), all multi-model quantities from CMIP5 shown in the paper (for example, 1850–1950 reference distribution of 'natural variability', 2.5th to 97.5th percentile range of daily CMIP5 predictions) are weighted such that each model receives equal weight. The prediction error metrics (RMSE) are evaluated separately for each model and season, and subsequently averaged. Training of statistical models was conducted in R (version 3.4.3)[43] using the 'glmnet' package (version 2.0-16)[44].

Fingerprint uncertainty analysis. A detailed fingerprint uncertainty analysis is provided in Supplementary Text 2. The analysis includes an assessment of signal-to-noise improvement provided by ridge regression over the simple 'daily global average fingerprint' example provided in the main text (Supplementary Text 2.1), an analysis of AGMT prediction performance depending on season (Supplementary Text 2.2) and an assessment of climate model uncertainty in fingerprint extraction (Supplementary Text 2.3). Last, we analyse how taking into account the spatial correlation structure in daily surface climate variables through ridge regression changes the resulting spatial patterns of regression coefficients (Supplementary Text 2.4).

Data from reanalyses and observations. To assess short-term detection beyond climate models, we project three daily reanalysis datasets, three monthly observational datasets and one daily observational dataset onto $\hat{\boldsymbol{\gamma}}$ to estimate our test statistic. As outlined above, the projection is performed separately for both target metrics (AGMT and EEI), for each set of predictors and for each day i using the fingerprint of the respective month.

The reanalyses include ERA-Interim[45], the NCEP/NCAR Reanalysis 1[46] and the 20th Century Reanalysis version 2c[47]. Spatial coverage of the reanalysis datasets is global, and with daily temporal coverage. The temporal coverage spans all datasets with a combined 168-yr period (that is, 1979–2018, 1948–2018, and 1851–2012, respectively).

In addition, we use monthly gridded temperature and specific humidity observations. Three monthly temperature datasets are available with near-global spatial coverage and in monthly temporal resolution. These datasets include: the Berkeley Earth Surface Temperatures[48] (BEST), the Cowtan and Way temperature reconstruction[49] (CW14) based on HadCRUT4[50] and the National Aeronautics and Space Administration's GISS Surface Temperature Analysis[51] (GISTEMP, version 3). All three datasets have global (CW14, 1850–2018) and near-global coverage (>99.4% in space after regridding to a 5°×5° regular grid, BEST, 1956–2018, and GISTEMP v3, 1957–2018) obtained through a statistical reconstruction to infill observational gaps[48,49,51] from station-based land temperatures blended with sea surface temperature measurements. Note that sea surface temperatures show slightly less warming than air temperatures above the sea[52] (for example, a difference of around 0.031 °C in the 2009–2013 period relative to 1961–1990[52]). This might imply very small differences compared with natural variability estimates from CMIP5 that are based on air temperatures and that the increase in the test statistic derived from blended observations is slightly too conservative. The fact that results based on observations and reanalyses are so similar suggests that the effect is small. For specific humidity, gridded observations are available only for land areas[27]. We use the Met Office HadISDH gridded global land surface humidity dataset[42] (spanning 1973–2017), which features a reasonable coverage of global land areas. We mask the dataset to all grid cells that have a coverage of at least 95% in time. This yields a land humidity dataset with a maximum of 3% gaps in space at any particular time step, which still samples all major land regions of the globe (Fig. 2e), and where in fact 519 out of 540 time steps (96.1%) have less than 1.5% gaps in space (after regridding to a 5°×5° regular grid; which yields 520 grid cells with data). This mask is used also for fingerprint extraction where 'land humidity' is included as a predictor. The small number of remaining gaps in the temperature and humidity gridded observations were filled with zeros (that is, corresponding to the monthly mean of the 1979–2005 reference period).

The processing of reanalysis and observational datasets follows exactly the processing of CMIP5 models. That is, all data are regridded to a regular 5°×5° grid and the seasonal cycle (31-day rolling mean for daily reanalysis datasets, and monthly mean seasonal cycle for monthly gridded observations), estimated from the 1979–2005 period, is subtracted from each grid cell before further analysis.

In addition to monthly observations, we construct a daily observational dataset that spans 1981–2018 by combining a daily sea surface temperature dataset (OISST-AVHRR[53]) with daily observational land data (Berkeley Earth Gridded Daily Data[48]). Daily observations should be considered as experimental and are shown only for illustration purposes, because Berkeley Earth daily land data are still in development. All details regarding dataset generation and discussion of potential caveats are described in Supplementary Text 1.

Robustness analysis of AGMT detection statements. We assess the robustness of detection statements in Supplementary Text 3. This includes a detection and emergence analysis equivalent to Figs. 3 and 4 but for all datasets individually (Supplementary Text 3.1), an assessment of the robustness of detection statements against individual CMIP5 models used to construct natural variability estimates (Supplementary Text 3.2) and an analysis of the influence of low-frequency variability on the distribution of the test statistic (Supplementary Text 3.3). In addition, we show how detection results depend on the timescale of analysis in Extended Data Fig. 3.

Detection analysis based on EEI. An in-depth analysis and discussion of EEI-based detection results is shown in Supplementary Text 4. The analysis includes a short description of the EEI target metric and its characteristics (Supplementary Text 4.1), additional description of EEI fingerprint extraction, analysis of fingerprints and EEI prediction errors (Supplementary Text 4.2), and a discussion and presentation of EEI-based detection results (Supplementary Text 4.3).

Data availability

All original CMIP5 data, reanalyses and observations used in this study are publicly available under the following URLs. CMIP5 model data: https://esgf-node.llnl.gov/projects/cmip5/; reanalysis: ERA-Interim (https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim), NCEP/NCAR Reanalysis 1 (https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html), NCEP/NCAR Reanalysis 2 (https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis2.html), Twentieth Century Reanalysis (https://www.esrl.noaa.gov/psd/data/20thC_Rean/); observations (monthly): GISTEMP temperature dataset, version 3 (https://data.giss.nasa.gov/gistemp/), Cowtan and Way (2014) temperature dataset, version 2 (https://www-users.york.ac.uk/~kdc3/papers/coverage2013/series.html), Berkeley Earth Monthly Land+Ocean temperature dataset (http://berkeleyearth.org/data/), Met Office gridded land surface humidity dataset (HadISDH), version 4.0.0.2017f (https://www.metoffice.gov.uk/hadobs/hadisdh/); observations (daily): Berkeley Earth Daily Land temperature dataset (Experimental, http://berkeleyearth.org/data/), NOAA Optimum Interpolation Sea Surface Temperature (OISST), AVHRR-Only (https://www.ncdc.noaa.gov/oisst). All intermediate and derived data from these products (extracted CMIP5 fingerprints and daily/monthly time series of the test statistic (that is, obtained by projecting CMIP5 models, reanalyses and observations individually onto the fingerprints)) are available at https://data.iac.ethz.ch/Sippel_et_al_2019_DailyDetection/.

Code availability

All computer code to reproduce the main results and all figures and Extended Data figures is available at https://data.iac.ethz.ch/Sippel_et_al_2019_DailyDetection/.

References

34. Marvel, K. et al. Twentieth-century hydroclimate changes consistent with human influence. *Nature* **569**, 59–65 (2019).

35. Von Storch, H. & Zwiers, F. W. *Statistical Analysis in Climate Research* (Cambridge Univ. Press, 2001).
36. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2012).
37. Friedman, J. H. Fast sparse regression and classification. *Int. J. Forecast.* **28**, 722–738 (2012).
38. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, 2009).
39. Szekely, E., Sippel, S., Knutti, R., Obozinski, G. & Meinshausen, N. A direct approach to detection and attribution of climate change. In *Proc. 9th International Workshop on Climate Informatics: CI 2019* (NCAR, in the press).
40. Gupta, A. S., Jourdain, N. C., Brown, J. N. & Monselesan, D. Climate drift in the CMIP5 models. *J. Clim.* **26**, 8597–8615 (2013).
41. Hobbs, W., Palmer, M. D. & Monselesan, D. An energy conservation analysis of ocean drift in the CMIP5 global coupled models. *J. Clim.* **29**, 1639–1653 (2016).
42. Willett, K. et al. HadISDH: an updateable land surface specific humidity product for climate monitoring. *Clim. Past* **9**, 657–677 (2013).
43. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2014); http://www.R-project.org/
44. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
45. Dee, D. P. et al. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**, 553–597 (2011).
46. Kalnay, E. et al. The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77**, 437–472 (1996).
47. Compo, G. P. et al. The twentieth century reanalysis project. *Q. J. R. Meteorol. Soc.* **137**, 1–28 (2011).
48. Rohde, R. et al. Berkeley Earth temperature averaging process. *Geoinformatics Geostat.* **1**, 1–13 (2013).
49. Cowtan, K. & Way, R. G. Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Q. J. R. Meteorol. Soc.* **140**, 1935–1944 (2014).
50. Morice, C. P., Kennedy, J. J., Rayner, N. A. & Jones, P. D. Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 data set. *J. Geophys. Res.* **117**, D08101 (2012).
51. Hansen, J., Ruedy, R., Sato, M. & Lo, K. Global surface temperature change. *Rev. Geophys.* **48**, RG000345 (2010).
52. Cowtan, K. et al. Robust comparison of climate models with observations using blended land air and ocean sea surface temperatures. *Geophys. Res. Lett.* **42**, 6526–6534 (2015).
53. Banzon, V., Smith, T. M., Chin, T. M., Liu, C. & Hankins, W. A long-term record of blended satellite and in situ sea-surface temperature for climate monitoring, modeling and environmental studies. *Earth Syst. Sci. Data* **8**, 165–176 (2016).

Acknowledgements

We thank A. Merrifield, I. Medhaug, G. Obozinski and H. Lange for comments, and we thank U. Beyerle and J. Sedlàček for the preparation and maintenance of CMIP5 data. We acknowledge funding received from the Swiss Data Science Centre within the project 'Data Science-informed attribution of changes in the Hydrological cycle' (DASH, ID C17-01). We thank the observers, creators, maintainers and providers of all datasets. Support for the Twentieth Century Reanalysis Project version 2c dataset is provided by the US Department of Energy, Office of Science Biological and Environmental Research (BER), and by the National Oceanic and Atmospheric Administration Climate Program Office. NCEP Reanalysis and NCEP Reanalysis 2 data were provided by the NOAA/OAR/ESRL PSD, Boulder, CO, USA. We thank the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups for producing and making available their model output. For CMIP, the US Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led the development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

Author Contributions

S.S. conceived the study with N.M. and R.K. and conducted the statistical analysis. All authors contributed to the interpretation of the results and the writing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

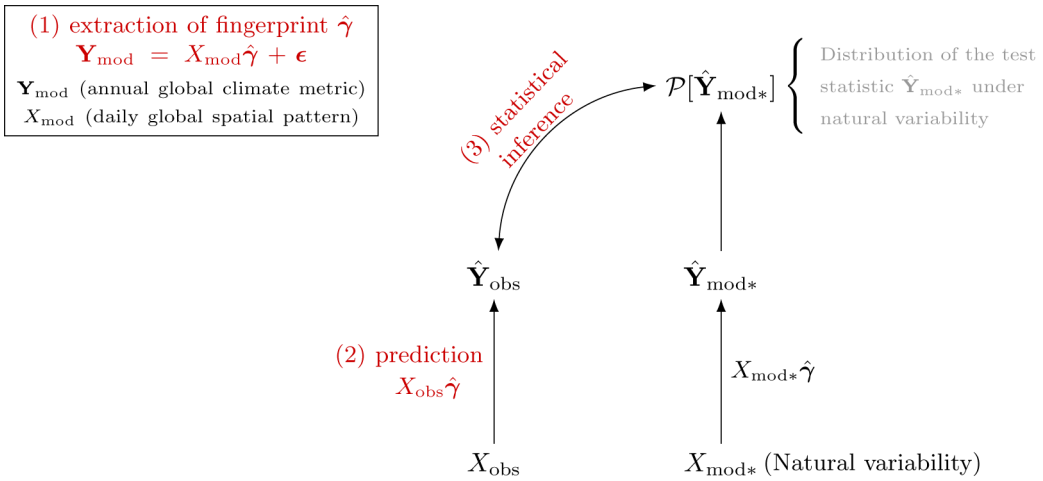
Extended data is available for this paper at https://doi.org/10.1038/s41558-019-0666-7.

Supplementary information is available for this paper at https://doi.org/10.1038/s41558-019-0666-7.

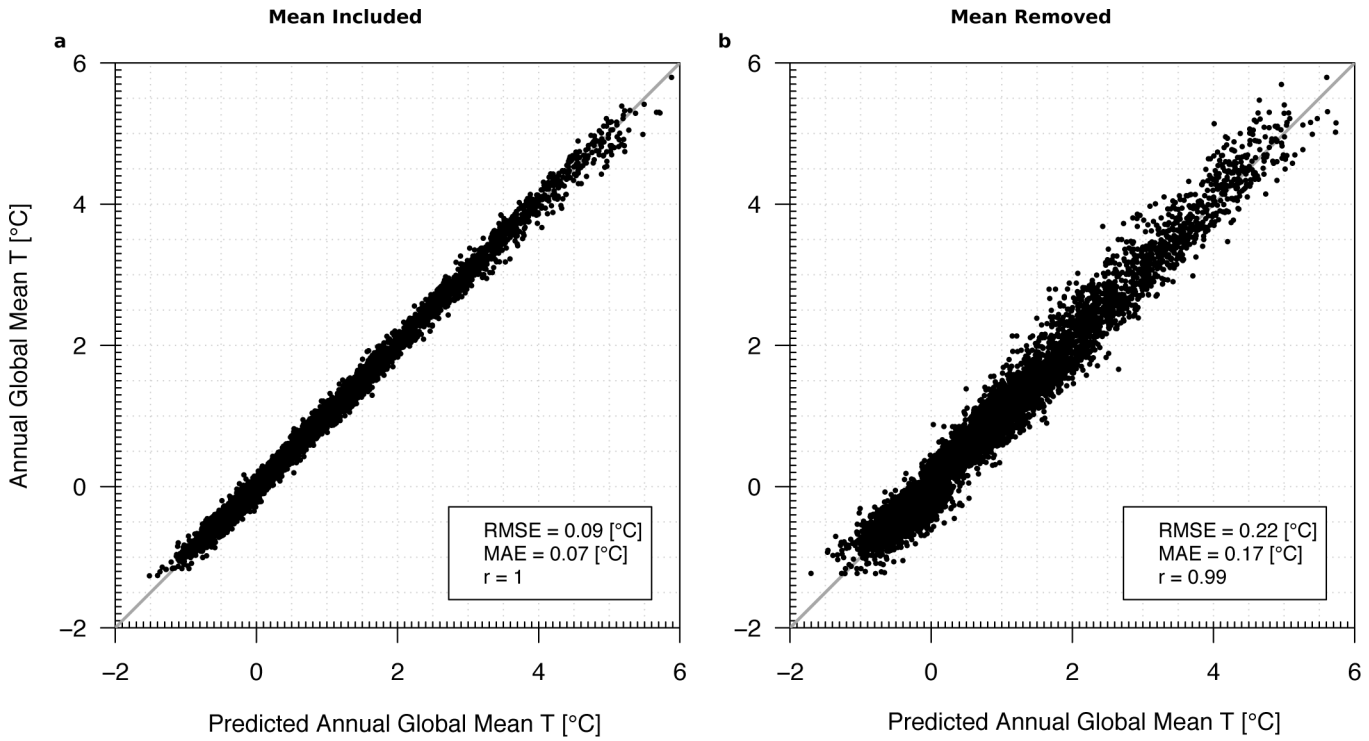
Correspondence and requests for materials should be addressed to S.S.

Peer review information *Nature Climate Change* thanks Karen McKinnon and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

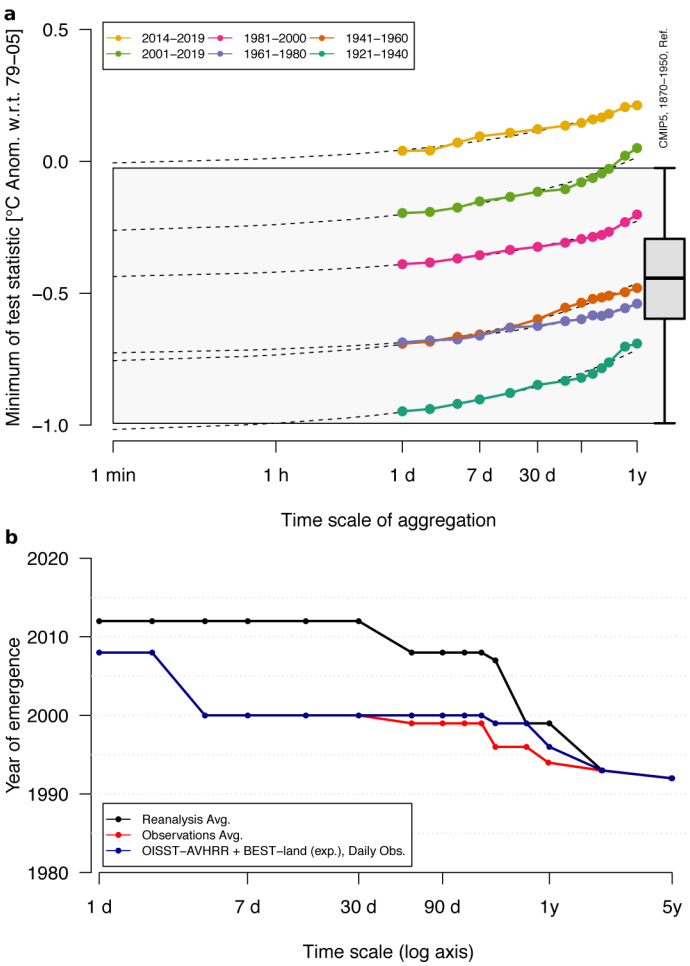
Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | Illustration of daily climate change detection method based on statistical learning. Statistical detection of externally forced climate change involves three steps. (1) A fingerprint $\hat{\boldsymbol{\gamma}}$ of external forcing on climate is extracted, using regularized linear regression, that relates model simulated spatial patterns of daily temperatures (X_{mod}) and a defined univariate target variable used as test statistic (denoted $\mathbf{Y}_{\mathrm{mod}}$ in the figure: for example annual global mean temperature, AGMT, or decadal-scale Earth's Energy Imbalance). (2) Spatial patterns of daily temperatures (X_{obs}) are projected onto the fingerprint $\hat{\boldsymbol{\gamma}}$ to predict the target variable (denoted $\hat{\mathbf{Y}}_{\mathrm{obs}}$). (3) A statistical significance test is used to infer whether external forcing on the climate system can be detected from the observed daily temperature pattern against the distribution of the test statistic under natural variability (denoted here $\mathcal{P}[\hat{\mathbf{Y}}_{\mathrm{mod}*}]$).



Extended Data Fig. 2 | Evaluation of Annual Global Mean Temperature prediction performance. Evaluation of prediction performance for the Annual Global Mean Temperature (AGMT) target metric across the multi-model CMIP5 archive from (**a**) any individual day's global temperature pattern ('Mean Included') and (**b**) the daily pattern of combined temperature and land humidity with the mean removed ('Mean Removed').



Extended Data Fig. 3 | Time scale dependence of climate change detection. (**a**) Dependence of overall minimum of the test statistic on the time scale of aggregation, shown over different time periods (coloured dots and lines) for the reanalysis time series. Dashed lines show statistical fits of a linear model in log-log space to each period and its extrapolation to sub-daily time scales. The different time scales of aggregation are obtained by successively aggregating the daily test statistics to longer time scales. The CMIP5 1870-1950 distribution of the daily test statistic is shown for comparison. (**b**) The year of emergence (that is 'detection' at any time) of global climate as a function of time scale. The figure is derived by finding, for each time scale and backwards in time from 2018, the first year in which any point does not exceed the 97.5th percentile of the CMIP5 1870-1950 reference distribution of the daily test statistic from (**a**). Over the last 20 years, climate change would have been detectable in any individual 365-day period, whereas over the last 10 years any 180-day period was detectable in reanalyses and observations. Over the last seven years, detection was possible for any individual day, and would have likely been possible even for shorter time periods. Detection in the experimental daily observational dataset (OISST+BEST) occurs slightly earlier than in daily reanalyses.