

# UCLA

## UCLA Previously Published Works

### Title

Anthropogenic influence on extreme precipitation over global land areas seen in multiple observational datasets.

### Permalink

<https://escholarship.org/uc/item/45q1g82w>

### Journal

Nature Communications, 12(1)

### Authors

Madakumbura, Gavin

Thackeray, Chad

Norris, Jesse

et al.

### Publication Date

2021-07-06

### DOI

10.1038/s41467-021-24262-x

Peer reviewed

# Anthropogenic influence on extreme precipitation over global land areas seen in multiple observational datasets

Gavin D. Madakumbura <sup>1</sup>, Chad W. Thackeray <sup>1</sup>, Jesse Norris <sup>1</sup>, Naomi Goldenson <sup>1</sup> & Alex Hall <sup>1</sup>

The intensification of extreme precipitation under anthropogenic forcing is robustly projected by global climate models, but highly challenging to detect in the observational record. Large internal variability distorts this anthropogenic signal. Models produce diverse magnitudes of precipitation response to anthropogenic forcing, largely due to differing schemes for parameterizing subgrid-scale processes. Meanwhile, multiple global observational datasets of daily precipitation exist, developed using varying techniques and inhomogeneously sampled data in space and time. Previous attempts to detect human influence on extreme precipitation have not incorporated model uncertainty, and have been limited to specific regions and observational datasets. Using machine learning methods that can account for these uncertainties and capable of identifying the time evolution of the spatial patterns, we find a physically interpretable anthropogenic signal that is detectable in all global observational datasets. Machine learning efficiently generates multiple lines of evidence supporting detection of an anthropogenic signal in global extreme precipitation.

<sup>1</sup>Department of Atmospheric and Oceanic Sciences, University of California—Los Angeles, Los Angeles, CA, USA. ✉email: [gavindayanga@ucla.edu](mailto:gavindayanga@ucla.edu)

Extreme precipitation can have devastating direct societal impacts such as flooding, soil erosion, and agricultural damage<sup>1</sup>, as well as causing indirect health risks and impacts<sup>2</sup>. Anthropogenic warming acts to intensify Earth's hydrologic cycle<sup>3</sup>. This intensification is manifested in part through increased extreme precipitation as a result of greater atmospheric moisture with warming following the Clausius–Clapeyron relationship. However, circulation changes can act to enhance or reduce this increase<sup>4–7</sup>. Future projections by climate models following climate change scenarios show a robust increase in extreme precipitation, globally and on regional scales<sup>8–11</sup>. Moreover, increased variation between wet and dry extremes is projected, which could have devastating societal impacts<sup>12,13</sup>. These changes in extreme precipitation may have already become apparent on a regional basis<sup>14–16</sup>.

Recent studies have detected anthropogenic influence in historical changes to extreme precipitation across the domains of North America<sup>17,18</sup>, Europe<sup>18,19</sup>, Asia<sup>18–20</sup>, and Northern Hemisphere land areas as a whole<sup>21</sup>. These attempts are part of a larger category of studies known as Detection and Attribution (D&A)<sup>22–24</sup>. Often, they initially extract the spatial or spatio-temporal patterns of climate-system response to anthropogenic forcing (so-called fingerprints) from an ensemble of global climate models (GCMs). Projection of observations onto these fingerprints allows for detection of the signal<sup>24,25</sup>. The presence of a signal that can be statistically distinguished from internal variability confirms the influence of external forcing. Thus, traditional D&A methods rely on long-term observations<sup>24,26</sup>. In the case of extreme precipitation, traditional methods may be difficult to apply globally due to inordinately short records and large observational uncertainty, reflected in multiple global datasets produced with very different assumptions<sup>27–30</sup>. Another key difficulty with traditional methods is that the models produce a large spread in the extreme precipitation response to historical anthropogenic forcing<sup>31</sup>. This spread, the model uncertainty, occurs alongside large internal variability in the models' simulations of the historical period. These two effects create significant uncertainty in the character of the true anthropogenic signal. In past research, spread in the response has been suppressed by assuming the anthropogenic fingerprint can be derived from the ensemble-mean change in extreme precipitation<sup>32</sup>. Here, we aim to take these uncertainties fully into account, by making no assumptions about how to derive the anthropogenic signal from GCM data.

Machine learning-based methods for the detection of anthropogenic influence (DAI) have been shown to overcome the reliance on trends<sup>33,34</sup> and are even capable of detecting the human influence from weather data on a single day<sup>35</sup>. An artificial neural network (ANN) is trained to predict a proxy of external forcing (e.g., the year of the data) based on the spatial maps of the target variable from an ensemble of GCM simulations. Under this supervised learning approach, the ANN learns the spatial patterns that best represent the external forcing from the background noise arising from the internal variability and model uncertainty<sup>33,34</sup>. Observations can then be fed to this trained ANN to assess the presence of an anthropogenic signal in observations<sup>33–35</sup>. This ANN DAI method can identify the nonlinear combinations of the forced signal, internal climate variability, and intermodel variability<sup>34</sup>. This method also has the advantage of being able to explicitly include internal variability and model uncertainty. It does not assume that any model or any model-derived quantity, such as the ensemble mean of the models, is the true anthropogenic signal. It uses the raw GCM data, with GCM internal variability included. In addition, novel visualization techniques allow for the interpretability of the ANNs formerly considered as black boxes, making them

explainable<sup>36,37</sup>, or interpretable in terms of physical processes or system behavior. The use of these visualization techniques alongside the ANN DAI method allows one to capture the time-varying dynamic fingerprints of each input and evaluate their physical credibility<sup>34,38</sup>.

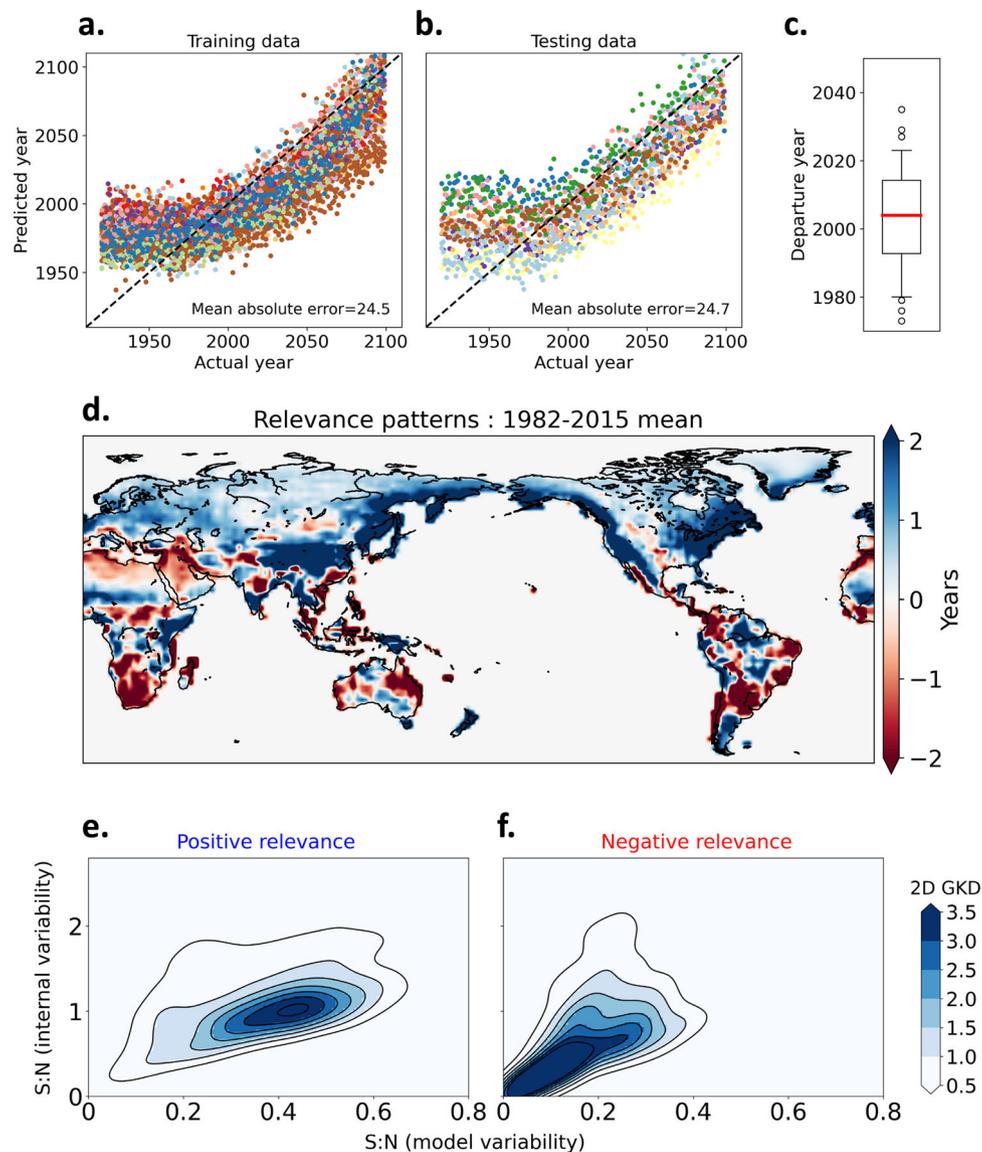
In this study, we apply the ANN DAI method and the ANN visualization technique known as Layer-wise Relevance Propagation (LRP)<sup>39,40</sup> to global maps of annual maximum daily precipitation (Rx1day) over land. Using Coupled Model Intercomparison Project, phase 5 (CMIP5)<sup>41</sup> and phase 6 (CMIP6)<sup>42</sup> model ensembles, we first aim to understand how the ANN is detecting the anthropogenic signal and interpret it physically. Then we use the ANN to detect the anthropogenic influence on Rx1day in several land-only observational and reanalysis datasets. Thus, we are agnostic about which GCM is correct, and which gridded dataset is a true representation of the observed record. In this way, we efficiently generate multiple lines of evidence as to the presence of an anthropogenic signal in the various instantiations of the observed record.

## Results

**ANN-identified fingerprints of anthropogenic influence.** We first discuss the ability of the ANN to predict the year of occurrence for a series of simulated annual Rx1day maps. Predictions of the simulated Rx1day year (Fig. 1a, b) show that the ANN struggles during roughly the 1920–1970 period. But prediction accuracy gradually increases, noticeably starting from the late twentieth century. This characteristic, a near-constant predicted year followed by a positive trend, is consistent with the emergence of the anthropogenic signal from the noise of natural variability<sup>43</sup>. Compared to when this technique is applied to global-mean temperature (ref. 33), there is a lag in the emergence of the anthropogenic signal in extreme precipitation. This delay is likely due to larger internal and intermodel variability in extreme precipitation. We estimate this time of emergence (departure year) as the year after which the ANN prediction continuously exceeds a selected base period (1920–1949)<sup>33,43</sup>. In GCMs, the predicted year departs from the base period in the 1970s, but the departures mostly occur later, with lower and upper quartiles of 1993 and 2014, respectively (Fig. 1c). The ANN suggests that there is a detectable anthropogenic signal in the GCM's Rx1day during the historical period, consistent with traditional statistical methods<sup>44</sup>.

Figure 1d shows the importance of each grid box for the ANN to identify the anthropogenic signal (hereafter called relevance patterns, see “Methods”), averaged over the period 1982–2015. Positive (negative) values in the relevance patterns correspond to an increase (decrease) in the predicted year. Therefore, areas of positive relevance can be interpreted as the regions with a positive contribution to the prediction (i.e., the year) and negative values are the regions with a negative contribution. The sum of all grid cell values is equal to the predicted year (“Methods” and Supplementary Text). By learning how to predict the year of the data, the ANN is able to detect the spatial patterns that best reflect the changing climate from background noise<sup>33,34</sup>. Therefore, the relevance patterns observed above can be considered as the ANN-identified fingerprints of anthropogenic influence on Rx1day (e.g., ref. 35).

The regions with positive relevance include the East Asian and African monsoon regions, and the North Pacific and Atlantic storm tracks (Fig. 1d and Supplementary Fig. 1). The regions with negative relevance include arid and semi-arid subtropical zones such as Northern African and Middle Eastern deserts, Southern South Africa, Australian arid and semi-arid regions, and wet regions such as central and northwestern parts of South America.



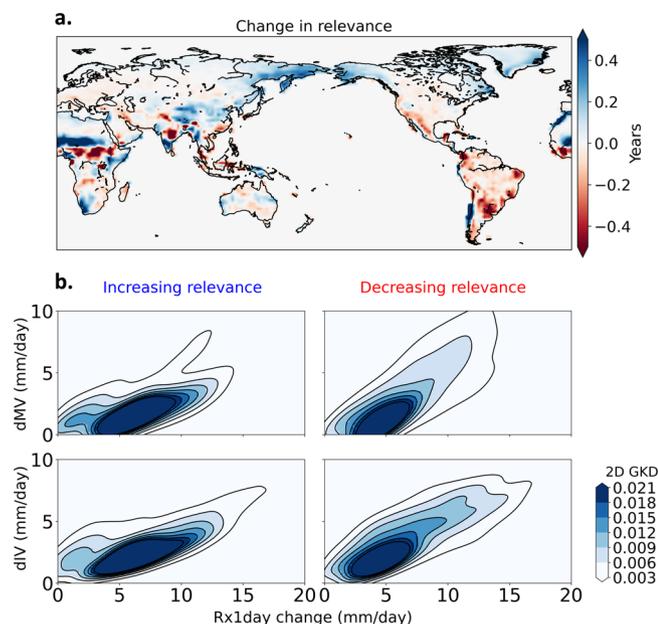
**Fig. 1 Fingerprint of external forcing in simulated Rx1day learned by the neural network.** **a, b** Actual year vs predicted year for training data derived from CMIP5 and CMIP6 global climate models (GCMs) (**a**) and testing data derived from CMIP5 and CMIP6 GCMs (**b**) for a single neural network. Each GCM is represented by a different color. **c** The year of departure from the base period, 1920–1949. Whiskers represent the 5th–95th percentiles, while blank circles represent outliers. **d** Multimodel, ensemble-mean, layer-wise-relevance-propagation-based relevance maps for annual maximum daily precipitation (Rx1day) input for the period 1982–2015 from all models. Relevance is the contribution of each grid value to the neural network’s decision (see “Methods”). **e, f** Signal-to-noise ratio (S:N): two-dimensional Gaussian kernel density (2D GKD) estimation plots for grid cells with a positive relevance (**e**) and negative relevance (**f**) in panel (**d**). Signal is defined as the multimodel mean change in Rx1day between the base period 1920–1949 and 2070–2099. Noise is defined in two ways: the first stems from internal variability and is calculated as the multimodel ensemble mean of the standard deviation in Rx1day during the base period. The second pertains to intermodel variability and is calculated as the intermodel standard deviation of the signal from each GCM.

Regions with negative relevance coincide with areas where the dynamical component of the Rx1day trend (i.e., the contribution from the change in vertical velocity<sup>4</sup>) is largely negative (ref. <sup>45</sup>, their Fig. 3b). This offsets the Rx1day increase stemming from the thermodynamic contribution (i.e., the contribution from the increase in atmospheric moisture with warming<sup>3–5</sup>) and produces only a weak and inconsistent increase in Rx1day<sup>45</sup>. The uncertainty associated with the dynamical component has been identified as a major concern for D&A of precipitation<sup>46</sup>.

To understand the physical nature of the relevance patterns, we next assess the signal, and the noise components arising from internal variability and the model uncertainty. The negative relevance of the forced response is associated with a lower signal-to-noise ratio (S:N) than the regions with positive relevance

(Fig. 1e, f). The S:N is lower for both internal variability and model variability. This reflects both the higher uncertainty regarding the change in extreme precipitation projected by GCMs for a majority of global arid land regions, as well as larger internal variability in those regions.

The ANN-based relevance patterns are consistent with the idea that previously observed long-term trends of terrestrial Rx1day are anthropogenic in origin (e.g., ref. <sup>21</sup>, their Fig. 1). Many wetland regions, such as the Asian, African, and South American monsoon regions, have experienced a robust increase in Rx1day to date<sup>15,16</sup>, whereas in arid and semi-arid subtropical zones no such trend can be seen<sup>16</sup>. The selection of regions in these previous studies (e.g., ref. <sup>16</sup>) seems to overlap with regions of higher relevance in Fig. 1d.



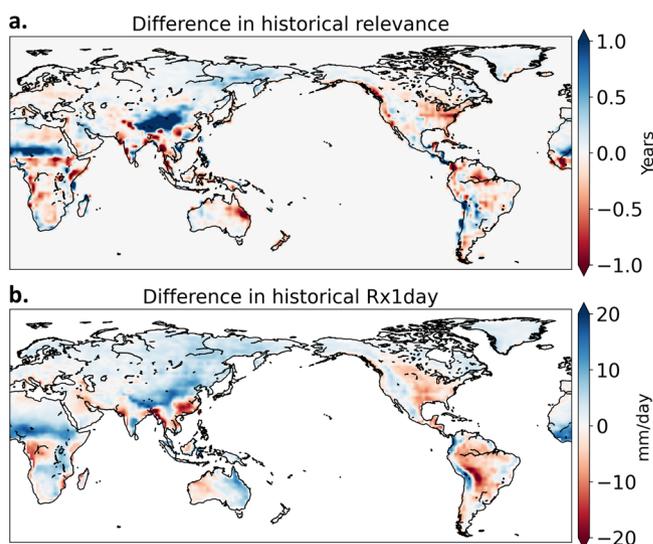
**Fig. 2** Change in the relevance patterns learned by the neural network through time. **a** Multimodel average change of relevance maps between 2070–2099 and 1920–1949. Relevance is the contribution of each grid value to the neural network’s decision, obtained using layer-wise relevance propagation (see “Methods”). **b** Multimodel ensemble-mean change in annual maximum daily precipitation (Rx1day) vs change in intermodel variability of Rx1day, (dMV, top panels), change in Rx1day vs change in internal variability of Rx1day (dIV, bottom panels), between 2070–2099 and 1920–1949. Two-dimensional Gaussian kernel density (2D GKD) estimation is shown. Left panels show results for grid cells where relevance increases with time in panel (a) and right panels show results for grid cells where relevance decreases. Internal variability is calculated as the standard deviation of Rx1day time-series and intermodel variability is calculated as the standard deviation of mean Rx1day from all models for each time period. Prior to the calculation of internal variability, the forced Rx1day trend at each grid cell was removed by regressing onto 41-year lowess filtered annual global-mean surface temperature<sup>85</sup>.

**Time-varying fingerprints.** One of the main advantages of using an ANN to detect anthropogenic influence over traditional D&A methods is that time-varying signals can be accounted for<sup>34,38</sup>. Changes in the signals could be due to the nonlinear evolution of the climate system or temporal and spatial variations in the forcing itself. Figure 2a shows the difference between the relevance maps for our baseline period (1920–1949) and the end of the twenty-first century (2070–2099). While the sum of the relevance maps derived using LRP is larger for later years in the time series (“Methods”), local differences can explain the redistribution of the importance with time. This ability to aggregate over regions and different samples has been identified as an advantage of using LRP to interpret deep-learning models<sup>40</sup>. Notably, the relevance increases with time across Africa and Asia, which is likely to be associated in part with the enhancement of the monsoon systems<sup>47</sup>. A similar increase in relevance can be seen in North Pacific and North Atlantic land regions, possibly associated with the poleward shift of storm tracks<sup>48</sup>. South African and South American Mediterranean climate regions also show an increase in relevance, associated with subtropical drying, a robust pattern of climate change<sup>49,50</sup>. This indicates that even though dry regions have a smaller S:N compared to wet regions in terrestrial Rx1day (Fig. 1d–f), some dry regions show an increase of the signal and/or decrease in noise with time, yielding an increase in the relevance (Fig. 2a). Among the regions with

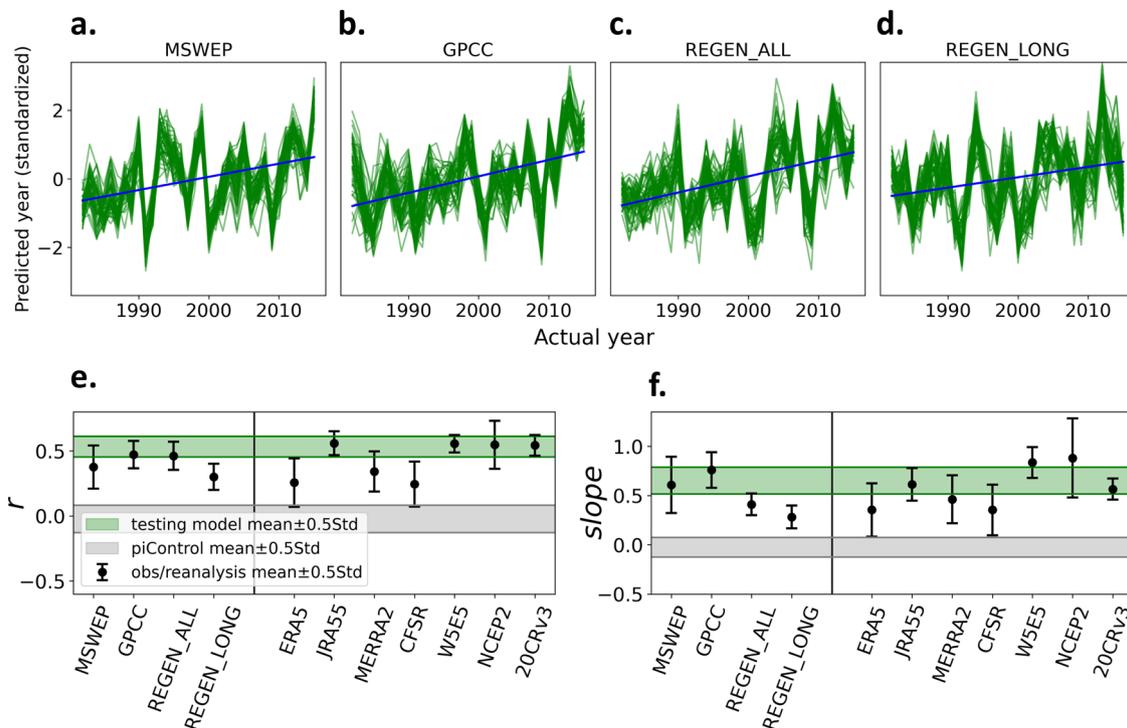
decreasing relevance, a majority of South America and the Western US stand out, possibly due to an increase in model uncertainty of Rx1day as the twenty-first century progresses.

To assess the physical validity of the change in relevance determined by the ANN, we break the terrestrial Rx1day record down into its forced signal and changes in noise components between the two periods. Results show that grid cells with increasing relevance have a comparable change in Rx1day, but much less increase in both internal variability and intermodel variability, compared to grid cells with decreasing relevance (Fig. 2b). Therefore, the change of relevance over time is in accord with the tradeoff between increasing noise and increasing signal with time.

**Origins of the spread in the predicted year.** We next investigate why the ANN predicts such a large range of years depending on the data of the underlying GCM used to predict the year. This intermodel spread in the predicted year is especially pronounced before the warming signal emerges (Fig. 1a, b). Here, we select four GCMs with the highest average predicted year, and four GCMs with the lowest average predicted year, during the baseline period (1920–1949). We obtain the relevance heatmaps for each year of the baseline period for these eight models and calculate the composite difference (i.e., models with high-versus-low predicted year; Fig. 3a). Large positive values are seen in the African and Asian monsoon regions. The models predicting later years also have larger twentieth-century mean state Rx1day values in these regions (Fig. 3b). Thus, the GCMs that predict a later year in the baseline period have more future-like patterns of Rx1day in their baseline climatologies compared to other models. When projected onto the fingerprints identified by the ANN, these patterns result in a later predicted year compared to the opposite subset. This exercise suggests a potential use of ANN-based DAI methods to understand how biases in historical simulations project onto future changes<sup>31</sup>.



**Fig. 3** Differences between subsets of models with high and low predicted years by the neural network during the baseline period (1920–1949). **a, b** The difference in their relevance maps (a) and annual maximum daily precipitation (Rx1day) (b) between the four models with the highest mean predicted year and the four models with the lowest mean predicted year (as shown in Fig. 1a, b). Relevance is the contribution of each grid value to the neural network’s decision, obtained using layer-wise relevance propagation (see “Methods”).



**Fig. 4** Metrics of the forced signal in observation-based estimates of precipitation during 1982–2015. **a–d** Actual year vs predicted year obtained from 51 different artificial neural networks (ANNs) with different training/validation/testing sets, for four observational datasets, MSWEP (**a**), GPCC (**b**), REGEN\_ALL (**c**), and REGEN\_LONG (**d**). Green lines show results from each ANN. The blue line is the mean slope. Each predicted year time series is standardized in the figure for a better comparison between datasets. **e** Correlation ( $r$ ) between the actual years and predicted years, **f** slope of the regression line between actual years and predicted years for observational and reanalysis data (black circle with a line), and testing models (green-shaded regions). Gray shading represents a measure of natural variability derived from 220 nonoverlapping 34-year segments obtained from pre-industrial control (piControl) simulations (see “Methods”). The error bar range and the range of the green and gray-shaded area show the  $\pm 0.5$  standard deviation range.

### Detected anthropogenic signal in historical Rx1day records.

With these physical interpretations of the ANN results and relevance patterns, we use the GCM-trained ANNs to detect whether there is a forced signal in observations. According to previous studies, a steady global warming trend can be seen since the 1970s<sup>51</sup> and, in GCMs, the anthropogenic signal of global-mean Rx1day has started to emerge as of the 1970s<sup>44</sup>. Therefore, according to the theoretical basis of the response of extreme precipitation to warming<sup>3–5</sup>, one could hypothesize that GCM-simulated and observed Rx1day should have a positive significant trend during the historical period analyzed here, 1982–2015. Confirming this, GCMs show a positive trend in globally averaged Rx1day (significant at 99% in 36 out of 44 models), which cannot be explained by natural variability alone (Supplementary Fig. 2). In observations and reanalyses, only seven out of the eleven datasets show a significant trend (significant at 99%) in globally averaged Rx1day for the historical period 1982–2015, ranging from 0.02 to 0.09 mm/day/year (Supplementary Table 1). Taken at face value, this large disparity in observations suggests that the observational evidence for anthropogenic influence on recent changes in extreme precipitation is weak. However, when we apply the ANN trained on Rx1day data from GCMs, to the same eleven datasets, a different story emerges.

If an observational dataset exhibits the same forced response as the GCMs, the time series of predicted year from that dataset should have a positive correlation ( $r$ ) with the actual year and linear regression of these two variables should produce a positive slope<sup>25,34</sup>. The metric  $r$  can be considered as an indicator of the presence of an anthropogenic signal whereas the slope is an indicator of the strength of that signal. Figure 4 shows these two metrics for observations, reanalyses and testing GCMs, from 51

different ANNs trained using randomly selected training GCMs. We also calculated the two metrics ( $r$  and slope) for the predicted-versus-actual year given for GCM simulations with radiative forcing held constant at pre-industrial levels, which is used as a measure of natural variability (see “Methods”). All observations and reanalysis have positive  $r$  values (Fig. 4 and Supplementary Fig. 3), even in datasets that do not show a significant positive trend in global-average Rx1day (Supplementary Table 1). This contrast is because the ANN detects a signal in the spatial distribution of Rx1day, as opposed to the global average. The  $r$  values for all observational datasets are substantially larger than those expected by natural variability (gray-shaded area in Fig. 4e, f). When looking at the slope, two observational datasets (MSWEP and GPCC) are in line with GCMs, along with four reanalyses (JRA55, MERRA2, W5E5, and 20CRV3). The two REGEN datasets, ERA5 and CFSR, show lower slopes, whereas NCEP2 has the highest slope among the datasets considered here. In general, observational and reanalysis products show similar  $r$  values and slopes as the GCMs for the same historical time period (compare the black circles and the green bands in Fig. 4e, f). This indicates that the observational and reanalysis products show anthropogenic influence on Rx1day that is comparable to what is shown by GCMs.

To estimate the statistical significance of the signal detected in the observations and reanalysis, we first estimate the noise as the standard deviation of the distribution of the slopes representing natural variability (Fig. 4e, f). Then S:N is calculated in all datasets by dividing the mean slope by noise. Following the two-tailed  $z$  test, S:N larger than, for instance, 1.96 corresponds to a statistical significance level of 95%<sup>25,52,53</sup>. Out of the four observations, MSWEP, GPCC, and REGEN\_ALL show a 95% significance and

REGEN\_LONG shows a 84% significance. Among reanalyses, ERA5 and CFSR show a 90% significance while the rest show a 95% significance.

These results demonstrate that the absence of a significant linear trend in globally averaged Rx1day cannot be taken to mean there is no evidence of an anthropogenic signal in Rx1day. This underscores the importance of exploiting the spatial pattern of the response to external forcing to extract the forced signal in observations, as opposed to the trend-based analysis<sup>35,38,54</sup>. In particular, areas with higher internal variability can act to suppress the trend in the global mean. Further evidence of the importance of spatial patterns can be seen in the fact that the average ANN-predicted values vary widely and systematically across the observational datasets (Supplementary Fig. 3). This is an indicator of systematic and large relative biases in the Rx1day climatologies of the various datasets (as pointed out above in the discussion of ANN applied to the GCMs, the average predicted value of the year depends on the magnitude of the Rx1day in the climatology (Fig. 3 and Supplementary Text)). Yet it is significant that the ANN can put the years in close to the correct order, as demonstrated by the significant correlations between actual and predicted years, even if the absolute value of the years is incorrect. This is a strong indicator that the subtle patterns and time variations of the simulated anthropogenic signal are present in the observational datasets and are shared among them, despite the fact that they are systematically biased relative to one another and likely the real world<sup>55,56</sup>.

## Discussion

Detecting anthropogenic signals in observations of extreme precipitation has been a challenging task due to the large internal variability of rare events, as well as climate model uncertainty. The limited sampling in observations adds additional uncertainty, due in part to a dataset development process that involves a variety of homogenization, extrapolation, and interpolation techniques to produce global gridded products<sup>30</sup>. Using a recently introduced ANN DAI method, we utilized the time evolution of spatial maps of Rx1day in GCMs, for historical simulations and future projections. The ANN yields fingerprints of anthropogenic signals that are physically consistent with the time evolution of the forced signal and can be distinguished from the noise arising from internal variability and substantial model uncertainty. Using this ANN DAI method, we show that the anthropogenic signal can be detected in all global terrestrial Rx1day records considered in this study. This robust detection occurs despite large systematic biases and large discrepancies in data sources and homogenization methods.

While previous trend-based D&A assessments of Rx1day have demonstrated the human influence in this variable in some regions, those studies assume the ensemble mean of the GCMs is the anthropogenic signal. This leads to questions as to whether further steps are needed to fully consider model uncertainty<sup>32,57</sup>. We made a simple attempt to examine this issue by applying the ANN DAI method to the same widely used, quality-controlled Rx1day record used in the previous trend-based D&A assessments. We applied the method twice, once using the same multimodel approach discussed elsewhere in this study, and once using a large ensemble dataset which only accounts for internal variability. Our results show that including internal variability and model uncertainty in the forced response could reduce the power of detection (Supplementary Text). Therefore, the detected signal in multiple global terrestrial Rx1day datasets in this study, with internal variability and model uncertainty being taken fully into account (Fig. 4), is a definitive affirmation of a human influence on extreme precipitation in the historical record. Note

that while all observations show this anthropogenic influence, the signal magnitude varies considerably, on par with that seen in the GCMs. This large observational uncertainty underscores a difficulty in constraining future projections of extreme precipitation with historical climate model simulations and observations<sup>31,58</sup>.

Several caveats of the machine learning-based detection method should be noted. Compared to regression-based traditional D&A methods<sup>59</sup>, the assessment of the influence of individual forcings (e.g., anthropogenic aerosols, land-use change, and natural forcings such as volcanic and solar activities) in the presented framework is challenging. We did not attempt such a breakdown in this study, and this would require methodological modifications<sup>60</sup>. In addition, the training GCMs might be undersampling the low-frequency natural variability such as Atlantic Multidecadal variability and Pacific Decadal Oscillation. This may be remedied by inflating the training dataset with paleoclimate data<sup>61</sup>. However, even with adequate sampling of natural variability in the training dataset, the underestimation of the precipitation response to natural forcings such as volcanic activities and natural variability such as El Niño Southern Oscillation in GCMs could still affect the results<sup>62</sup>. We also note that different ANN visualization techniques are available<sup>63–65</sup>, and those should be explored to understand the sensitivity of the extracted fingerprints to the ANN visualization technique. Despite these limitations, it is clear that ANN DAI methods with ANN visualization techniques are very useful and efficient in identifying the human influence on variables that are highly uncertain in GCMs, and poorly characterized in observations, such as extreme precipitation.

## Methods

**Data.** We use daily precipitation rate output from a collection of climate models participating in CMIP5 and CMIP6 (Supplementary Table 2). Data from each ensemble's historical forcing scenario are combined with future projections following a high-emissions scenario to create a time series from 1920 to 2099 for each model. Future projections from CMIP5 follow the Representative Concentration Pathway 8.5 (RCP 8.5)<sup>66</sup>, while CMIP6 projections follow the Shared Socio-economic Pathway 5–8.5 (SSP 5–8.5)<sup>67</sup>. To increase our sample size, we combine both CMIP5 and CMIP6 model subsets into one ensemble, which is justifiable considering the very similar time evolution of the total anthropogenic forcing in RCP 8.5 and SSP 5–8.5 scenarios (ref. 67, their Fig. 3c). We regrid all daily precipitation data to a  $2^\circ \times 2^\circ$  spatial grid and compute the Rx1day value for each year at each land-grid point.

To assess the influence of natural variability, we also use pre-industrial control simulations (piControl), which are GCM simulations with radiative forcing held constant at pre-industrial levels. As the length of the piControl simulations vary between GCMs, we selected the same number of samples from a collection of 20 CMIP6 models used here (Supplementary Table 3). We extract 34-year nonoverlapping samples from each simulation (so as to match the length of the observational record) to represent natural variability. Each GCM provided 14 samples of this length and after removing the first three samples of each simulation to avoid climate drift<sup>42</sup>, we were left with 220 pre-industrial samples with which to assess natural variability.

We use four datasets of observational estimates of daily precipitation rate with global coverage: Multi-Source Weighted-Ensemble Precipitation, version 2 (MSWEP)<sup>68</sup>, Global Precipitation Climatology Centre (GPCC) version 2018<sup>69</sup>, and Rainfall Estimates on a Gridded Network (REGEN)<sup>70</sup>, including both REGEN\_ALL and REGEN\_LONG. MSWEP is a hybrid reconstruction using in situ, satellite, and reanalysis data, whereas GPCC and the REGEN datasets are developed from ground-based measurements. REGEN\_ALL is developed by interpolating all considered station data, whereas REGEN\_LONG is developed using only the stations with a data record of 40 years or longer. We further use seven widely used reanalysis products for comparison: European Centre for Medium-Range Weather Forecasts ERA5<sup>71</sup>, Japanese 55-year Reanalysis (JRA55)<sup>72</sup>, Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA2)<sup>73</sup>, NCEP Climate Forecast System Reanalysis (CFSR)<sup>74</sup>, the bias-corrected ERA5 precipitation dataset compiled for phase 3b of the intersectoral impact model intercomparison project (W5E5)<sup>75,76</sup>, NCEP-DOE Reanalysis 2 (NCEP2)<sup>77</sup> and NOAA-CIRES-DOE Twentieth Century Reanalysis version 3 (20CRv3)<sup>78</sup>. These observational and reanalysis datasets are selected considering the availability of full global land coverage and data for at least three decades. We selected the period 1982–2015 for observational analysis as it is the common temporal range for all datasets. All observation and reanalysis data were

regridded to the same  $2^\circ \times 2^\circ$  spatial grid as the models, and then Rx1day was calculated at each grid point for each year.

**Neural network-based detection method.** Here, we apply the method in ref. 33 to predict the year with which given annual Rx1day maps from GCMs are associated, a regression task. This requires the ANN to learn the signature of the forced response in simulated Rx1day. By feeding the ANN data from forced simulations, it learns to distinguish the forced signal from internal climate variability. The use of multiple GCMs helps the ANN learn the common elements of the forced response most relevant to the prediction task, a process that fully considers model uncertainty as well as internal climate variability. Input to the ANN from each model is a vectorized spatial map of Rx1day ( $2^\circ \times 2^\circ$  spatial grid = 16,200 grid values) for each year from 1920 to 2099. Our primary goal is to detect the anthropogenic signal in extreme precipitation over land (excluding Antarctica). Thus, we mask out data over the ocean at this stage, resulting in 6082 land-grid values. The ANN architecture consists of two hidden layers with ten nodes each. The Rectified Linear Unit activation function is used for all hidden units.

Approximately 60% of the models (26) are used for training the ANN, while the rest of the models are divided equally to use as validation and testing sets (9 models each). The mean squared error between the actual and predicted year of Rx1day is used as the loss function to be minimized during the training. For the optimizer which updates the ANN based on the gradient of the loss, we select rmsprop. Climate variables inherently contain spatial autocorrelation. To account for this dependence among adjacent input data points, we use L2 regularization<sup>79</sup> between inputs and the first hidden layer, which adds the sum of squared weights as a penalty term to the loss function. By iterating over L2 values of leading order of magnitudes and inspecting the tradeoff between low-prediction error and generalizability (Supplementary Fig. 4), we found L2 = 0.001 to be a suitable value for our analysis. We trained the model for 1000 epochs. Early stopping was enabled to reduce the overfitting by monitoring the validation loss with a patience value of 50 epochs<sup>80</sup>. We repeated the training process for 51 different training sets obtained by random combinations of GCMs, resulting in 51 different ANNs. We found that increasing the number of hidden units or changing the other hyperparameters did not result in a substantial increase in accuracy.

#### Neural network interpretation using layer-wise relevance propagation (LRP).

Assume that for a given input map,  $\mathbf{x}$ , we get an output  $f(\mathbf{x})$ , in our case, the predicted year. LRP conservatively back-propagates this value through hidden layers until it reaches the input map. This process generates a relevance heatmap, indicating the areas of importance influencing the value  $f(\mathbf{x})$ . The conservation property is shown in Eq. 1, for relevance propagation between two hidden layers  $j$  and  $k$ , where  $k$  is the upper layer (i.e., closer to the output).  $\sum_{i=1}^d P_i$  denotes the sum of the relevance of the  $d$  input features. The summation operation for each hidden layer (e.g.,  $\sum_k P_k$ ) is the summation of the relevance ( $\mathbf{P}$ ) of all hidden units in that layer, where  $P_k$  is the relevance of a single unit in layer  $k$ . The activation,  $a_k$  (Eq. 2) is the information coming from all units in layer  $j$ , to a target unit in layer  $k$ . In Eq. 2,  $a_j$  values are the individual activations of each unit in the layer  $j$ ,  $w_{jk}$  values are the weights associated with the relationship between each unit in layer  $j$  and the target unit in layer  $k$ , and  $b_k$  is the bias of that target unit.

$$\sum_{i=1}^d P_i = \dots = \sum_j P_j = \sum_k P_k = \dots = f(\mathbf{x}) \quad (1)$$

$$a_k = \text{ReLU} \left( \sum_j a_j w_{jk} + b_k \right) \quad (2)$$

$$P_j = \sum_k \left( \alpha \frac{a_j w_{jk}^+}{\sum_j a_j w_{jk}^+} - \beta \frac{a_j w_{jk}^-}{\sum_j a_j w_{jk}^-} \right) P_k \quad (3)$$

The relevance-propagation rule from layer  $k$  to a unit in layer  $j$  is given in Eq. 3. This general form is also known as the  $\alpha\beta$ -rule<sup>39,40</sup>. The components  $()^+$  and  $()^-$  indicate only positive and negative parts are being considered, respectively. The  $\alpha$  and  $\beta$  coefficients represent the relative amount of positive and negative relevance to be propagated, respectively. As shown in Eq. 3, positive relevance (i.e., excitatory influence) and negative relevance (i.e., inhibitory influence) are associated with positive and negative weights, respectively. The  $\alpha$  and  $\beta$  coefficients are to be chosen with the constraints  $\alpha - \beta = 1$  and  $\beta \geq 0$ . The combination  $\alpha = 2$  and  $\beta = 1$  (LRP <sub>$\alpha=2\beta=1$</sub> ) has been experimentally inferred as suitable, and has been adopted in previous research<sup>39,40,81–83</sup>. Here we adopt the LRP <sub>$\alpha=2\beta=1$</sub>  rule (Supplementary Text).

#### Data availability

CMIP data used are available at <https://esgf-node.llnl.gov/projects/esgf-llnl/>. Observational and reanalysis data used are available at the following links: MSWEP: <http://www.gloh2o.org/>, GPCC: <https://www.dwd.de/EN/ourservices/gpcc/gpcc.html>, REGEN\_ALL: <https://doi.org/10.25914/5ca4c380b0d44>, REGEN\_LONG: <https://doi.org/>

[10.25914/5ca4c2c6527d2](https://doi.org/10.25914/5ca4c2c6527d2), ERA5: <https://cds.climate.copernicus.eu/>, JRA55: <https://rda.ucar.edu/datasets/ds628.0/>, MERRA2: <https://disc.gsfc.nasa.gov/>, CFSR: <https://www.ncdc.noaa.gov/data-access/>, W5E5: <https://esg.pik-potsdam.de/>, NCEP2 and 20CRv3: <https://psl.noaa.gov/data/>.

#### Code availability

Neural network analysis was conducted using Python libraries TensorFlow (<https://www.tensorflow.org>) and Keras (<https://keras.io>). Neural network interpretation was carried out using the library iNNvestigate<sup>84</sup> (<https://github.com/albermax/innvestigate>). Python scripts developed for the analysis and figures are available publicly at <https://doi.org/10.6084/m9.figshare.14479659>.

Received: 2 February 2021; Accepted: 8 June 2021;

Published online: 06 July 2021

#### References

- Handmer, J. et al. Changes in impacts of climate extremes: human systems and ecosystems. In *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* (eds Field, C. B. et al.) 231–290 (Cambridge University Press, 2012).
- Crimmins, A. et al. The impacts of climate change on human health in the United States: a scientific assessment. Washington, DC: US Global Change Research Program. <https://doi.org/10.7930/JOR49NQX> (2016).
- Allen, M. R. & Ingram, W. J. Constraints on future changes in climate and the hydrologic cycle. *Nature* **419**, 224–232 (2002).
- O’Gorman, P. A. & Schneider, T. The physical basis for increases in precipitation extremes in simulations of 21st century climate change. *Proc. Natl Acad. Sci. USA* **106**, 14773–14777 (2009).
- Trenberth, K. E., Dai, A., Rasmussen, R. M. & Parsons, D. B. The changing character of precipitation. *Bull. Am. Meteorological Soc.* **84**, 1205–1218 (2003).
- Held, I. M. & Soden, B. J. Robust response of the hydrological cycle to global warming. *J. Clim.* **19**, 5686–5699 (2006).
- Norris, J., Chen, G. & Neelin, J. D. Thermodynamic versus dynamic controls on extreme precipitation in a warming climate from the Community Earth System Model Large Ensemble. *J. Clim.* **32**, 1025–1045 (2019).
- Sun, Y., Solomon, S., Dai, A. & Portmann, R. W. How often does it rain? *J. Clim.* **19**, 916–934 (2006).
- Fischer, E. M., Beyerle, U. & Knutti, R. Robust spatially aggregated projections of climate extremes. *Nat. Clim. Change* **3**, 1033–1038 (2013).
- Kharin, V. V., Zwiers, F. W., Zhang, X. & Wehner, M. Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Clim. Change* **119**, 345–357 (2013).
- Sillmann, J., Kharin, V. V., Zwiers, F. W., Zhang, X. & Bronaugh, D. Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections. *J. Geophys. Res. Atmospheres* **118**, 2473–2493 (2013).
- Madakumbura, G. D. et al. Event-to-event intensification of the hydrologic cycle from 1.5 °C to a 2 °C warmer world. *Sci. Rep.* **9**, 3483 (2019).
- Swain, D. L., Langenbrunner, B., Neelin, J. D. & Hall, A. Increasing precipitation volatility in twenty-first-century California. *Nat. Clim. Change* **8**, 427–433 (2018).
- Allan, R. P. & Soden, B. J. Atmospheric warming and the amplification of precipitation extremes. *Science* **321**, 1481–1484 (2008).
- Donat, M. G., Lowry, A. L., Alexander, L. V., O’Gorman, P. A. & Maher, N. More extreme precipitation in the world’s dry and wet regions. *Nat. Clim. Change* **6**, 508–513 (2016).
- Donat, M. G., Angéilil, O. & Ukkola, A. M. Intensification of precipitation extremes in the world’s humid and water-limited regions. *Environ. Res. Lett.* **14**, 065003 (2019).
- Kirchmeier-Young, M. C. & Zhang, X. Human influence has intensified extreme precipitation in North America. *Proc. Natl Acad. Sci. USA* **117**, 13308–13313 (2020).
- Dong, S. et al. Attribution of extreme precipitation with updated observations and CMIP6 simulations. *J. Clim.* **34**, 871–881 (2021).
- Paik, S. et al. Determining the anthropogenic greenhouse gas contribution to the observed intensification of extreme precipitation. *Geophys. Res. Lett.* **47**, e2019GL086875 (2020).
- Dong, S., Sun, Y. & Li, C. Detection of human influence on precipitation extremes in Asia. *J. Clim.* **33**, 5293–5304 (2020).
- Min, S. K., Zhang, X. B., Zwiers, F. W. & Hegerl, G. C. Human contribution to more-intense precipitation extremes. *Nature* **470**, 378–381 (2011).
- Bindoff, N. L. et al. Detection and attribution of climate change: from global to regional. In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. et al.) 867–952 (Cambridge University Press, 2013).

23. Stott, P. et al. Detection and attribution of climate change: a regional perspective. *Wiley Interdiscip. Rev.: Clim. Change* **1**, 192–211 (2010).
24. Hegerl, G. & Zwiers, F. Use of models in detection and attribution of climate change. *Wiley Interdiscip. Rev.: Clim. Change* **2**, 570–591 (2011).
25. Marvel, K. & Bonfils, C. Identifying external influences on global precipitation. *Proc. Natl Acad. Sci. USA* **110**, 19301–19306 (2013).
26. Easterling, D. R., Kunkel, K. E., Wehner, M. F. & Sun, L. Detection and attribution of climate extremes in the observed record. *Weather Clim. Extremes* **11**, 17–27 (2016).
27. Herold, N., Behrangi, A. & Alexander, L. V. Large uncertainties in observed daily precipitation extremes over land. *J. Geophys. Res.: Atmospheres* **22**, 668–681 (2017).
28. Sun, Q. et al. A review of global precipitation data sets: data sources, estimation, and intercomparisons. *Rev. Geophysics* **56**, 79–107 (2018).
29. Roca, R. et al. FROGS: a daily 1° × 1° gridded precipitation database of rain gauge, satellite and reanalysis products. *Earth Syst. Sci. Data* **11**, 1017–1035 (2019).
30. Hegerl, G. C. et al. Challenges in quantifying changes in the global water cycle. *Bull. Am. Meteorological Soc.* **96**, 1097–1115 (2015).
31. O’Gorman, P. A. Sensitivity of tropical precipitation extremes to climate change. *Nat. Geosci.* **5**, 697–700 (2012).
32. Ribes, A., Zwiers, F. W., Azais, J.-M. & Naveau, P. A new statistical approach to climate change detection and attribution. *Clim. Dyn.* **48**, 367–386 (2017).
33. Barnes, E. A., Hurrell, J. W., Ebert-Uphoff, I., Anderson, C. & Anderson, D. Viewing forced climate patterns through an AI lens. *Geophys. Res. Lett.* **46**, 13389–13398 (2019).
34. Barnes, E. A. et al. Indicator patterns of forced change learned by an artificial neural network. *J. Adv. Modeling Earth Syst.* **12**, e2020MS002195 (2020).
35. Sippel, S., Meinshausen, N., Fischer, E. M., Székely, E. & Knutti, R. Climate change now detectable from any single day of weather at global scale. *Nat. Clim. Change* **10**, 35–41 (2020).
36. Ebert-Uphoff, I., Samarasinghe, S. & Barnes, E. Thoughtfully using artificial intelligence in Earth science. *Eos* **100**. <https://doi.org/10.1029/2019eo135235> (2019).
37. Toms, B. A., Barnes, E. A. & Ebert-Uphoff, I. Physically interpretable neural networks for the geosciences: applications to earth system variability. *J. Adv. Modeling Earth Syst.* **12**, e2019MS002002 (2020).
38. Wills, R. C., Sippel, S. & Barnes, E. A. Separating forced and unforced components of climate change: the utility of pattern recognition methods in large ensembles and observations. *Variations* **18**, 1–10 (2020).
39. Bach, S. et al. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS ONE* **10**, e0130140 (2015).
40. Montavon, G., Samek, W. & Müller, K. R. Methods for interpreting and understanding deep neural networks. *Digital Signal Process.* **73**, 1–15 (2018).
41. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorological Soc.* **93**, 485–498 (2012).
42. Eyring, V. et al. Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Dev. Discussion* **8**, 10539–10583 (2015).
43. Mora, C. et al. The projected timing of climate departure from recent variability. *Nature* **502**, 183–187 (2013).
44. King, A. D. et al. The timing of anthropogenic emergence in simulated climate extremes. *Environ. Res. Lett.* **10**, 04015 (2015).
45. Pfahl, S., O’Gorman, P. A. & Fischer, E. M. Understanding the regional pattern of projected future changes in extreme precipitation. *Nat. Clim. Change* **7**, 423–427 (2017).
46. Shepherd, T. G. Atmospheric circulation as a source of uncertainty in climate change projections. *Nat. Geosci.* **7**, 703–708 (2014).
47. Hsu, P.-C. Global monsoon in a changing climate. In *The Monsoons and Climate Change* (eds Carvalho, L. M. V. & Jones, C.) 7–24 (Springer International Publishing, Springer Cham Heidelberg, 2016).
48. Yin, J. A consistent poleward shift of the storm tracks in simulations of 21st century climate. *Geophys. Res. Lett.* **32**, L18701 (2005).
49. Lu, J., Vecchi, G. A. & Reichler, T. Expansion of the Hadley cell under global warming. *Geophys. Res. Lett.* **34**, L06805 (2007).
50. Scheff, J. & Frierson, D. M. Robust future precipitation declines in CMIP5 largely reflect the poleward expansion of model subtropical dry zones. *Geophys. Res. Lett.* **39**, L18704 (2012).
51. Rahmstorf, S., Foster, G. & Cahill, N. Global temperature evolution: recent trends and some pitfalls. *Environ. Res. Lett.* **12**, 054001 (2017).
52. Marvel, K. et al. External influences on modeled and observed cloud trends. *J. Clim.* **28**, 4820–4840 (2015).
53. Marvel, K. et al. Twentieth-century hydroclimate changes consistent with human influence. *Nature* **569**, 59–65 (2019).
54. Wills, R. C. J., Battisti, D. S., Armour, K. C., Schneider, T. & Deser, C. Pattern recognition methods to separate forced responses from internal variability in climate model ensembles and observations. *J. Clim.* **33**, 8693–8719 (2020).
55. Bador, M., Alexander, L. V., Contractor, S. & Roca, R. Diverse estimates of annual maxima daily precipitation in 22 state-of-the-art quasi-global land observation datasets. *Environ. Res. Lett.* **15**, 35005 (2020).
56. Alexander, L. V. et al. Intercomparison of annual precipitation indices and extremes over global land areas from in situ, space-based and reanalysis products. *Environ. Res. Lett.* **15**, 055002 (2020).
57. Ribes, A., Qasmi, S. & Gillett, N. P. Making climate projections conditional on historical observations. *Sci. Adv.* **7**, eabc0671 (2021).
58. Borodina, A., Fischer, E. M. & Knutti, R. Models are likely to underestimate increase in heavy rainfall in the extratropical regions with high rainfall intensity. *Geophys. Res. Lett.* **44**, 7401–7409 (2017).
59. Allen, M. R. & Stott, P. A. Estimating signal amplitudes in optimal fingerprinting. Part I: theory. *Clim. Dyn.* **21**, 477–491 (2003).
60. Székely, 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* (eds Brajard, J. et al.) 119–124 (NCAR, 2019).
61. Min, S. K. Human fingerprint in global weather. *Nat. Clim. Change* **10**, 15–16 (2020).
62. Zhang, X. et al. Detection of human influence on twentieth-century precipitation trends. *Nature* **448**, 461–465 (2007).
63. Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K. & Müller, K. R. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning* (Springer Nature, 2019).
64. Gilpin, L. H. et al. Explaining explanations: an overview of interpretability of machine learning. In *2018 IEEE 5th International Conference on Data Science and Advanced Analytics* (eds Bonchi, F. et al.) 80–89 (IEEE, 2018).
65. Linardatos, P., Papastefanopoulos, V. & Kotsiantis, S. Explainable AI: a review of machine learning interpretability methods. *Entropy* **23**, 18 (2021).
66. Meinshausen, M. et al. The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Clim. Change* **109**, 213–241 (2011).
67. O’Neill, B. C. et al. The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geoscientific Model Dev.* **9**, 3461–3482 (2016).
68. Beck, H. E. et al. MSWEP V2 global 3-hourly 0.1° precipitation: methodology and quantitative assessment. *Bull. Am. Meteorological Soc.* **100**, 473–500 (2019).
69. Ziese, M. et al. GPCC full data daily version.2018 at 1.0°: Daily land-surface precipitation from rain-gauges built on GTS-based and historic data. [https://doi.org/10.5676/DWD\\_GPCC/FD\\_D\\_V2018\\_100](https://doi.org/10.5676/DWD_GPCC/FD_D_V2018_100) (2018).
70. Contractor, S. et al. Rainfall Estimates on a Gridded Network (REGEN)—a global land-based gridded dataset of daily precipitation from 1950–2013. *Hydrol. Earth Syst. Sci.* **24**, 919–943 (2020).
71. Hersbach, H. et al. The ERA5 global reanalysis. *Q. J. R. Meteorological Soc.* **2020**, 1–51 (2020).
72. Kobayashi, S. et al. The JRA-55 reanalysis: general specifications and basic characteristics. *J. Meteorological Soc. Jpn.* **93**, 5–48 (2015).
73. Gelaro, R. et al. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *J. Clim.* **30**, 5419–5454 (2017).
74. Saha, S. et al. The NCEP climate forecast system reanalysis. *Bull. Am. Meteorological Soc.* **91**, 1015–1057 (2010).
75. Lange, S. WFDE5 over land merged with ERA5 over the ocean (W5E5). V. 1.0. GFZ Data Services. <https://doi.org/10.5880/pik.2019.023> (2019).
76. Cucchi, M. et al. WFDE5: bias-adjusted ERA5 reanalysis data for impact studies. *Earth Syst. Sci. Data* **12**, 2097–2120 (2020).
77. Kanamitsu, M. et al. NCEP-DOE AMIP-II reanalysis (R-2). *Bull. Am. Meteorological Soc.* **83**, 1631–1643 (2002).
78. Slivinski, L. C. et al. Towards a more reliable historical reanalysis: improvements for version 3 of the twentieth century reanalysis system. *Q. J. R. Meteorological Soc.* **145**, 2876–2908 (2019).
79. Hastie, T. Ridge regularization: an essential concept in data science. *Technometrics* **62**, 426–433 (2020).
80. Géron, A. *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (O’Reilly Media, 2017).
81. Böhle, M., Eitel, F., Weygandt, M. & Ritter, K. Layer-wise relevance propagation for explaining deep neural network decisions in MRI-based Alzheimer’s disease classification. *Front. Aging Neurosci.* **11**, 194 (2019).
82. Dobrescu, A., Valerio Giuffrida, M. & Tsafaris, S. A. Understanding deep neural networks for regression in leaf counting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2600–2608. <https://doi.org/10.1109/CVPRW.2019.00316> (IEEE, 2019).
83. Grigorescu, I. et al. Investigating image registration impact on preterm birth classification: an interpretable deep learning approach. In *Smart Ultrasound Imaging and Perinatal, Preterm and Paediatric Image Analysis* (eds Wang, Q. et al.) 104–112 (Springer, Cham., 2019).
84. Alber, M. et al. iNNvestigate neural networks! *J. Mach. Learn. Res.* **20**, 1–8 (2019).

85. Hawkins, E. et al. Observed emergence of the climate change signal: from the familiar to the unknown. *Geophys. Res. Lett.* **47**, e2019GL086259 (2020).

### Acknowledgements

We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups for producing and making available their model output. We also thank the Earth System Grid Federation (ESGF) for archiving the data and providing access, and various funding agencies who support CMIP and ESGF. We acknowledge support from the Regional and Global Model Analysis Program for the Office of Science of the U.S. Department of Energy through the Program for Climate Model Diagnosis and Intercomparison.

### Author contributions

G.D.M., C.W.T., J.N. and A.H. designed research. G.D.M. developed the detection framework, conducted the analyses, interpreted the results, and wrote the manuscript. C.W.T., J.N., N.G., and A.H. contributed to the interpretation of the results and edited the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-24262-x>.

**Correspondence** and requests for materials should be addressed to G.D.M.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021