

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

Gavin D. Madakumbura (✉ [gavindayanga@ucla.edu](mailto:gavindayanga@ucla.edu))

University of California – Los Angeles <https://orcid.org/0000-0002-0092-4153>

**Chad W. Thackeray**

University of California – Los Angeles

**Jesse Norris**

University of California – Los Angeles

**Naomi Goldenson**

University of California – Los Angeles

**Alex Hall**

University of California – Los Angeles

---

## Research Article

**Keywords:** Detection and attribution, Rx1day, artificial neural networks, layerwise relevance propagation

**Posted Date:** April 27th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-227967/v2>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Nature Communications on July 6th, 2021.

See the published version at <https://doi.org/10.1038/s41467-021-24262-x>.

# 1 Anthropogenic influence on extreme precipitation over global land areas seen 2 in multiple observational datasets

3 Gavin D. Madakumbura<sup>1\*</sup>, Chad W. Thackeray<sup>1</sup>, Jesse Norris<sup>1</sup>, Naomi Goldenson<sup>1</sup>  
4 and Alex Hall<sup>1</sup>

5 <sup>1</sup>Department of Atmospheric and Oceanic Sciences, University of California — Los  
6 Angeles, Los Angeles, CA, USA.

7 \*correspondence e-mail address: [gavindayanga@ucla.edu](mailto:gavindayanga@ucla.edu)

## 8 9 **Abstract**

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

23 **Keywords:** Detection and attribution, Rx1day, artificial neural networks, layerwise relevance propagation

## 25 **Introduction**

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

35 Recent studies have detected anthropogenic influence in historical changes to extreme  
36 precipitation across the domains of North America<sup>17,18</sup>, Europe<sup>18,19</sup>, Asia<sup>18-20</sup> and Northern Hemisphere  
37 land areas as a whole<sup>21</sup>. These attempts are part of a larger category of studies known as *Detection and*  
38 *Attribution* (D&A)<sup>22-24</sup>. Often, they initially extract the spatial or spatiotemporal patterns of climate-  
39 system response to anthropogenic forcing (so-called *fingerprints*) from an ensemble of global climate  
40 models (GCMs). Projection of observations onto these fingerprints allows for detection of the signal<sup>24,25</sup>.  
41 The presence of a signal that can be statistically distinguished from internal variability confirms the  
42 influence of external forcing. Thus, traditional D&A methods rely on long term observations<sup>24,26</sup>. In the  
43 case of extreme precipitation, traditional methods may be difficult to apply globally due to inordinately  
44 short records and large observational uncertainty, reflected in multiple global datasets produced with  
45 very different assumptions<sup>27-30</sup>. Another key difficulty with traditional methods is that the models  
46 produce a large spread in the extreme precipitation response to historical anthropogenic forcing<sup>31</sup>. This  
47 spread, the model uncertainty, occurs alongside large internal variability in the models’ simulations of  
48 the historical period. These two effects create significant uncertainty in the character of the “true”

49 anthropogenic signal. In past research, spread in the response has been suppressed by assuming the  
50 anthropogenic fingerprint can be derived from the ensemble-mean change in extreme precipitation<sup>32</sup>.  
51 Here we aim to take these uncertainties fully into account, by making no assumptions about how to  
52 derive the anthropogenic signal from GCM data.

53 Machine learning-based methods for the detection of anthropogenic influence (DAI) have been  
54 shown to overcome the reliance on trends<sup>33,34</sup> and are even capable of detecting the human influence  
55 from weather data on a single day<sup>35</sup>. An artificial neural network (ANN) is trained to predict a proxy of  
56 external forcing (e.g. the year of the data) based on the spatial maps of the target variable from an  
57 ensemble of GCM simulations. Under this supervised learning approach, the ANN learns the spatial  
58 patterns that best represent the external forcing from the background noise arising from the internal  
59 variability and model uncertainty<sup>33,34</sup>. Observations can then be fed to this trained ANN to assess the  
60 presence of an anthropogenic signal in observations<sup>33-35</sup>. This ANN DAI method can identify the non-  
61 linear combinations of the forced signal, internal climate variability and inter-model variability<sup>34</sup>. This  
62 method also has the advantage of being able to explicitly include internal variability and model  
63 uncertainty. It does not assume that any model or any model-derived quantity, such as the ensemble-  
64 mean of the models, is the “true” anthropogenic signal. And it uses the raw GCM data, with GCM  
65 internal variability included. In addition, novel visualization techniques also allow for the interpretability  
66 of the ANNs formerly considered as “black boxes”, making them *explainable*<sup>36,37</sup>, or interpretable in  
67 terms of physical processes or system behavior. Use of these visualization techniques alongside the ANN  
68 DAI method allows one to capture the time varying dynamic fingerprints of each input and evaluate  
69 their physical credibility<sup>34,38</sup>.

70 In this study, we apply the ANN DAI method and the ANN visualization technique known as  
71 Layerwise Relevance Propagation (LRP)<sup>39,40</sup> to global maps of annual maximum daily precipitation  
72 (Rx1day) over land. Using Coupled Model Intercomparison Project, phase 5 (CMIP5)<sup>41</sup> and phase 6

73 (CMIP6)<sup>42</sup> model ensembles, we first aim to understand how the ANN is detecting the anthropogenic  
74 signal and interpret it physically. Then we use the ANN to detect the anthropogenic influence on Rx1day  
75 in several land-only observational and reanalysis datasets. Thus, we are agnostic about which GCM is  
76 correct, and which gridded data set is a true representation of the observed record. In this way we  
77 efficiently generate multiple lines of evidence as to the presence of an anthropogenic signal in the  
78 various instantiations of the observed record.

79

## 80 **Results**

### 81 **ANN-identified fingerprints of anthropogenic influence**

82 We first discuss the ability of the ANN to predict the year of occurrence for a series of simulated  
83 annual Rx1day maps. Predictions of the simulated Rx1day year (Figure 1a,b) show that the ANN  
84 struggles during roughly the 1920-1970 period. But prediction accuracy gradually increases, noticeably  
85 starting from the late 20<sup>th</sup> century. This characteristic, a near constant predicted year followed by a  
86 positive trend, is consistent with the emergence of the anthropogenic signal from the noise of natural  
87 variability<sup>43</sup>. Compared to when this technique is applied to global-mean temperature (ref. 33) there is a  
88 lag in the emergence of the anthropogenic signal in extreme precipitation. This delay is likely due to  
89 larger internal and inter-model variability in extreme precipitation. We estimate this time of emergence  
90 (*departure year*) as the year after which the ANN prediction continuously exceeds a selected base period  
91 (1920-1949) (see ref. 33 and 43 for more details). In GCMs, predicted year departs from the base period  
92 in the 1970s, but the departures mostly occur later, with lower and upper quartiles of 1993 and 2014,  
93 respectively (Figure 1c). The ANN suggests that there is a detectable anthropogenic signal in the GCM's  
94 Rx1day during the historical period, consistent with traditional statistical methods<sup>44</sup>.

95 Figure 1d shows the importance of each grid box for the ANN to identify the anthropogenic  
96 signal (hereafter called *relevance patterns*, see methods), averaged over the period 1982-2015. Positive

97 (negative) values in the relevance patterns correspond to an increase (decrease) in the predicted year.  
98 Therefore, areas of positive relevance can be interpreted as the regions with a positive contribution to  
99 the prediction (i.e. the year) and negative values are the regions with a negative contribution. The sum  
100 of each grid cell value is equal to the predicted year (methods, supplementary text). By learning how to  
101 predict the year of the data, the ANN is able to detect the spatial patterns that best reflect the changing  
102 climate from background noise<sup>33,34</sup>. Therefore, the relevance patterns observed above can be  
103 considered as the ANN-identified fingerprints of anthropogenic influence on Rx1day (e.g. ref. 35).

104         The regions with positive relevance include the East Asian and African monsoon regions, and the  
105 North Pacific and Atlantic storm tracks (Figure 1d, Figure S2). The regions with negative relevance  
106 include arid and semi-arid subtropical zones such as Northern African and Middle Eastern deserts,  
107 Southern South Africa, Australian arid and semi-arid regions, and wet regions such as central and  
108 northwestern parts of South America. Regions with negative relevance coincide with areas where the  
109 dynamical component of the Rx1day trend (i.e. the contribution from the change in vertical velocity<sup>4</sup>) is  
110 largely negative (ref. 45, their Figure 3b). This offsets the Rx1day increase stemming from the  
111 thermodynamic contribution (i.e. the contribution from the increase in atmospheric moisture with  
112 warming<sup>3-5</sup>) and produces only a weak and inconsistent increase in Rx1day<sup>45</sup>. The uncertainty associated  
113 with the dynamical component has been identified as a major concern for D&A of precipitation<sup>46</sup>.

114         To understand the physical nature of the relevance patterns, we next assess the signal, and the  
115 noise components arising from internal variability and the model uncertainty. Negative relevance of the  
116 forced response is associated with a lower signal-to-noise ratio (S:N) than the regions with positive  
117 relevance (Figure 1e,f). The S:N is lower for both internal variability and model variability. This reflects  
118 both the higher uncertainty regarding the change in extreme precipitation projected by GCMs for a  
119 majority of global arid land regions, as well as larger internal variability in those regions.

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

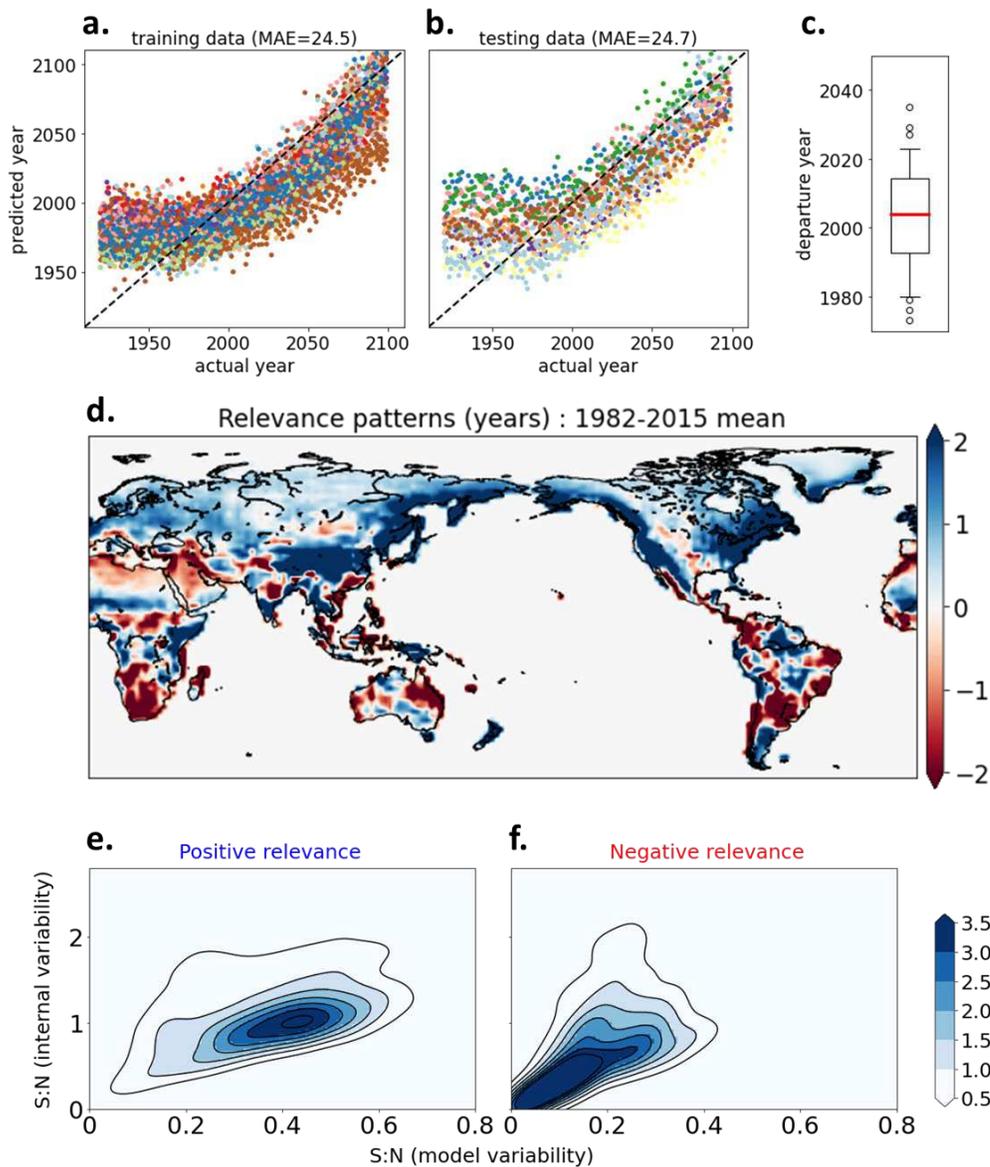
126

### 127 **Time varying fingerprints**

128 One of the main advantages of using an ANN to detect anthropogenic influence over traditional  
129 D&A methods is that time varying signals can be accounted for<sup>34,38</sup>. Changes in the signals could be due  
130 to the nonlinear evolution of the climate system or temporal and spatial variations in the forcing itself.  
131 Figure 2a shows the difference between the relevance maps for our baseline period (1920-1949) and the  
132 end of the 21<sup>st</sup> century (2070-2099). While the sum of the relevance maps derived using LRP is larger for  
133 later years in the time series (Methods), local differences can explain the redistribution of the  
134 importance with time. This ability to aggregate over regions and different samples has been identified as  
135 an advantage of using LRP to interpret deep learning models<sup>40</sup>. Notably, the relevance increases with  
136 time across Africa and Asia, which is likely to be associated in part with the enhancement of the  
137 monsoon systems<sup>47</sup>. A similar increase in relevance can be seen in North Pacific and North Atlantic land  
138 regions, possibly associated with the poleward shift of storm tracks<sup>48</sup>. South African and South  
139 American Mediterranean climate regions also show an increase in relevance, associated with subtropical  
140 drying, a robust pattern of climate change<sup>49,50</sup>. This indicates that even though dry regions have a  
141 smaller S:N compared to wet regions in terrestrial Rx1day (Fig 1d-f), some dry regions show an increase  
142 of signal and/or decrease in noise with time, yielding an increase in the relevance (Figure 2a). Among

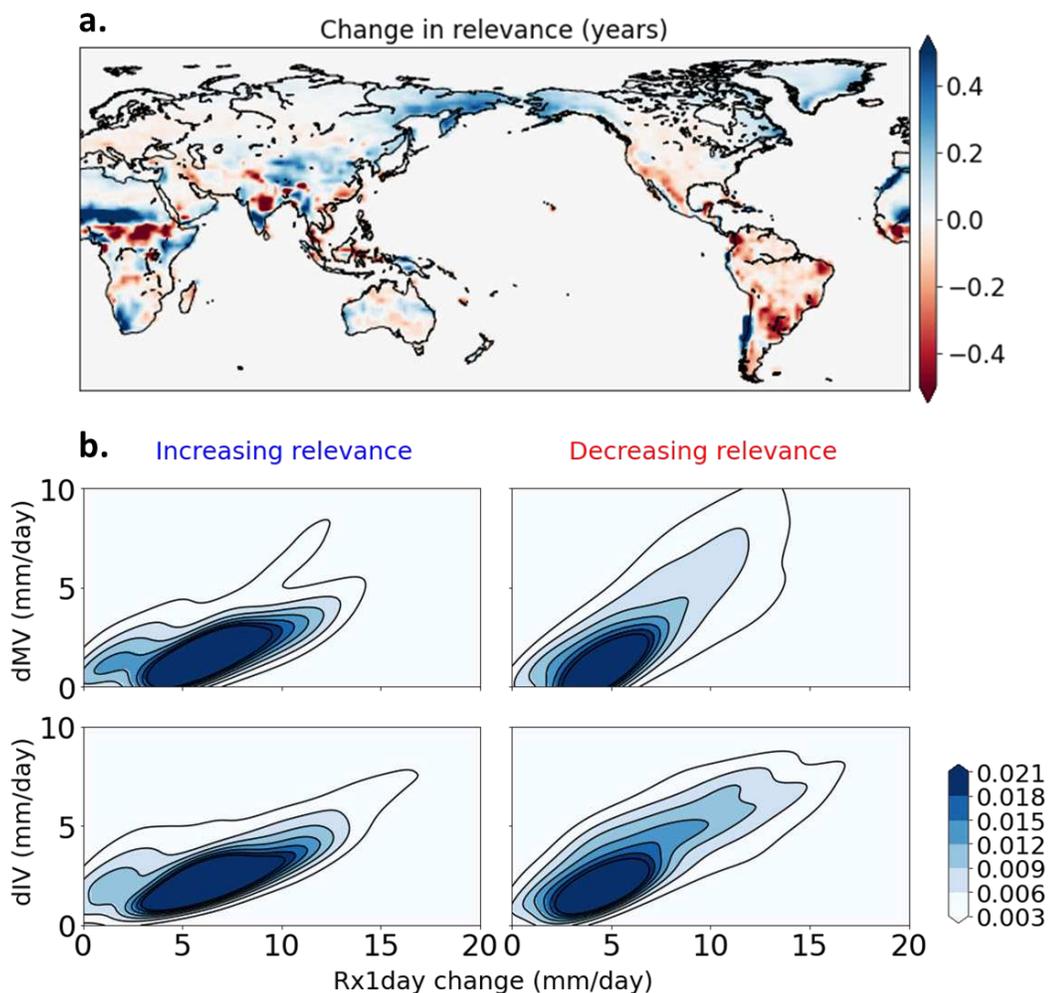
143 the regions with decreasing relevance, a majority of South America and the Western US stand out,  
144 possibly due to an increase in model uncertainty of Rx1day as the 21<sup>st</sup> Century progresses.

145 To assess the physical validity of the change in relevance determined by the ANN, we break the  
146 terrestrial Rx1day record down into its forced signal and changes in noise components between the two  
147 periods. Results show that grid cells with increasing relevance have a comparable change in Rx1day, but  
148 much less increase in both internal variability and inter-model variability compared to grid cells with  
149 decreasing relevance (Figure 4b). Therefore, the change of relevance over time is in accord with the  
150 tradeoff between increasing noise and increasing signal with time.



151  
 152  
 153 **Figure 1. Fingerprint of external forcing in simulated Rx1day learned by the ANN.**  
 154 (a,b) Actual year vs predicted year for training data derived from CMIP5 and CMIP6 GCMs (a) and testing data derived from  
 155 CMIP5 and CMIP6 GCMs (b) for a single ANN. Each GCM is represented by a different color. (c) The year of departure from the  
 156 base period, 1920-1949. Whiskers represent the 5<sup>th</sup>-95<sup>th</sup> percentiles, while blank circles represent outliers. (d) Multimodel,  
 157 ensemble-mean, layerwise-relevance-propagation-based relevance maps for Rx1day input for the period 1982-2015 from all  
 158 models. (e,f) Signal-to-noise ratio density plots for grid cells with a positive relevance (e) and negative relevance (f) in panel (d).  
 159 Signal is defined as the multi-model mean change in Rx1day between the base period 1920-1949 and 2070-2099. Noise is  
 160 defined in two ways: The first stems from internal variability and is calculated as the multimodel ensemble mean of the

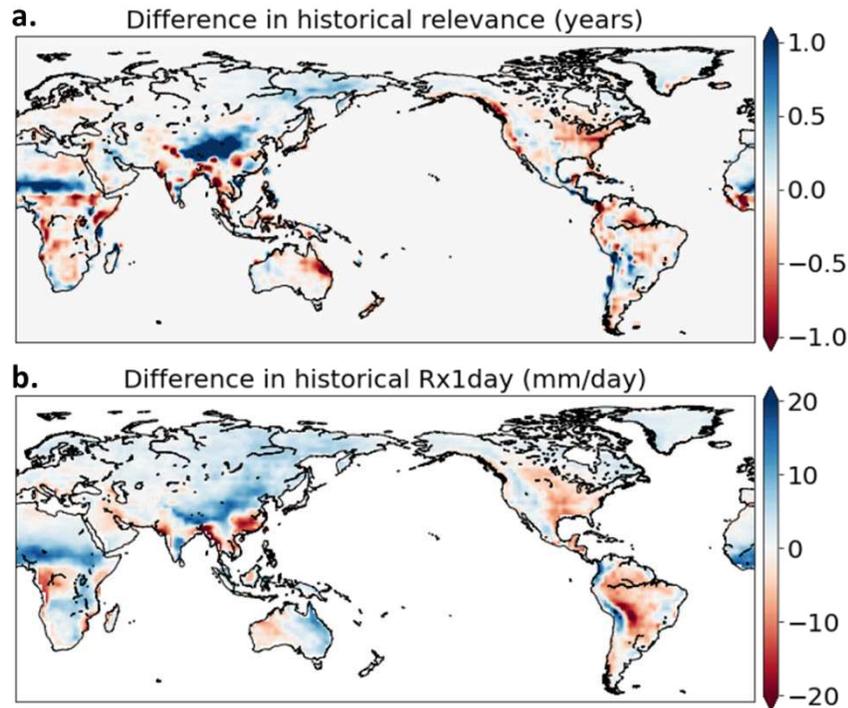
161 standard deviation in Rx1day during the base period. The second pertains to inter-model variability, and is calculated as the  
 162 inter-model standard deviation of the signal from each GCM.  
 163



164  
 165 **Figure 2. Change in the relevance patterns learned by the ANN through time.**  
 166 (a) Multimodel average change of relevance maps between 2070-2099 and 1920-1949. (b) Multimodel ensemble mean change  
 167 in Rx1day vs change in inter-model variability of Rx1day, (dMV, top panels), change in Rx1day vs change in internal variability  
 168 of Rx1day (dIV, bottom panels), between 2070-2099 and 1920-1949. Left panels show results for grid cells where relevance  
 169 increases with time in panel (a) and right panels show results for grid cells where relevance decreases. Internal variability is  
 170 calculated as the standard deviation of Rx1day timeseries and inter-model variability is calculated as the standard deviation of  
 171 mean Rx1day from all models for each time period. Prior to calculation of internal variability, the forced Rx1day trend at each  
 172 grid cell was removed by regressing onto 41-year lowess filtered annual global mean surface temperature<sup>84</sup>.  
 173

174 **Origins of the spread in the predicted year**

175           We next investigate why the ANN predicts such a large range of years depending on the data of  
176 the underlying GCM used to predict the year. This inter-model spread in the predicted year is especially  
177 pronounced before the warming signal emerges (Figure 1a,b). Here, we select four GCMs with the  
178 highest average predicted year, and four GCMs with the lowest average predicted year, during the  
179 baseline period (1920-1949). We obtain the relevance heatmaps for each year of the baseline period for  
180 these eight models and calculate the composite difference (i.e. models with high-versus-low predicted  
181 year; Fig 3a). Large positive values are seen in the African and Asian monsoon regions. The models  
182 predicting later years also have larger 20<sup>th</sup> century mean state Rx1day values in these regions (Figure  
183 3b). Thus, the GCMs that predict a later year in the baseline period have more *future-like patterns* of  
184 Rx1day in their baseline climatologies compared to other models. When projected onto the fingerprints  
185 identified by the ANN, these patterns result in a later predicted year compared to the opposite subset.  
186 This exercise suggests a potential use of ANN-based DAI methods to understand how biases in historical  
187 simulations project onto future changes<sup>31</sup>.



189

190 **Figure 3. Differences between subsets of models with high and low predicted years by the ANN during the baseline period**  
 191 **(1920-1949).** (a,b) The difference in their relevance maps (a) and Rx1day (b) between the four models with the highest mean  
 192 predicted year and the four models with lowest mean predicted year (as shown in Figure 1a-b).

193

194

### 195 **Detected anthropogenic signal in historical Rx1day records**

196 With these physical interpretations of the ANN results and relevance patterns, we use the GCM-  
 197 trained ANNs to detect whether there is a forced signal in observations. According to previous studies, a  
 198 steady global warming trend can be seen since the 1970s<sup>51</sup> and, in GCMs, the anthropogenic signal of  
 199 global-mean Rx1day has started to emerge as of the 1970s<sup>44</sup>. Therefore, according to the theoretical  
 200 basis of the response of extreme precipitation to warming<sup>3-5</sup>, one could hypothesize that GCM-simulated  
 201 and observed Rx1day should have a positive significant trend during the historical period analyzed here,

202 1982-2015. Confirming this, GCMs show a positive trend in globally-averaged Rx1day (significant at 99%  
203 in 36 out of 44 models), which cannot be explained by natural variability alone (Figure S3). In  
204 observations and reanalyses, only seven out of the eleven datasets show a significant trend ( $p < 0.01$ ) in  
205 globally-averaged Rx1day for the historical period 1982-2015, ranging from 0.02 to 0.09 mm/day/year  
206 (Table S3). Taken at face value, this large disparity in observations suggests that the observational  
207 evidence for anthropogenic influence on recent changes in extreme precipitation is weak. However,  
208 when we apply the ANN trained on Rx1day data from GCMs, to the same eleven datasets, a different  
209 story emerges.

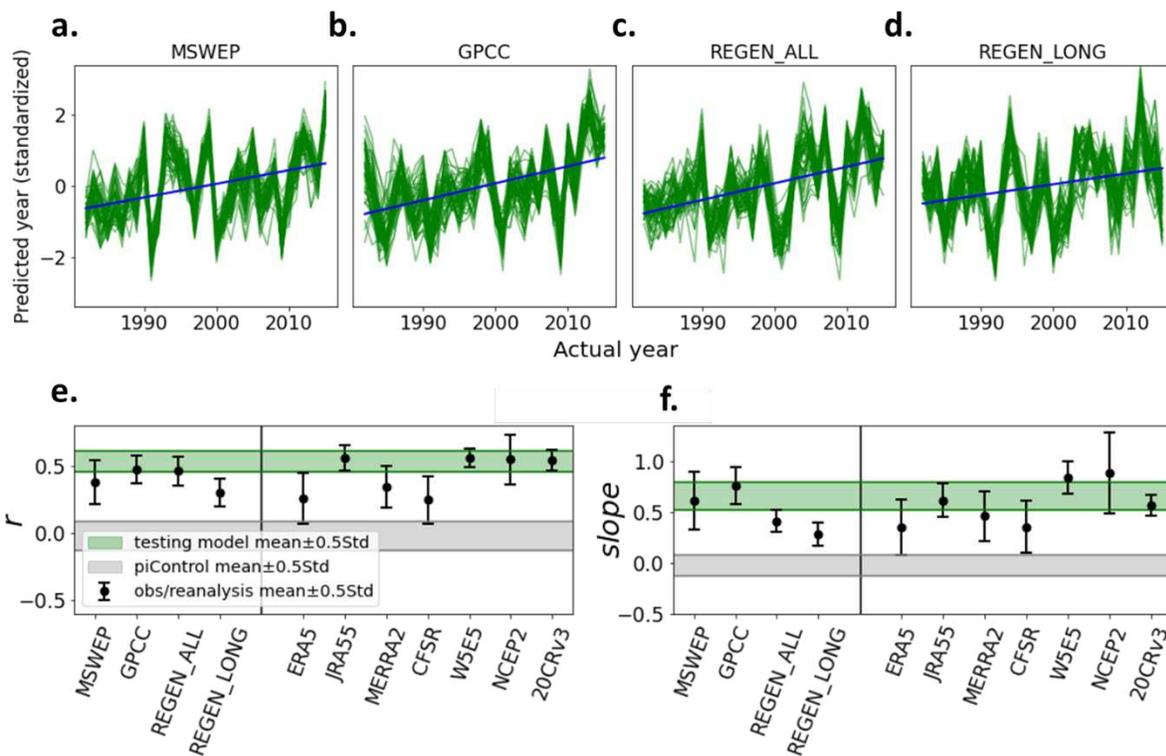
210         If an observational dataset exhibits the same forced response as the GCMs, the time series of  
211 predicted year from that dataset should have a positive correlation ( $r$ ) with the actual year and a linear  
212 regression of these two variables should produce a positive slope<sup>25,34</sup>. The metric  $r$  can be considered as  
213 an indicator of the presence of an anthropogenic signal whereas the slope is an indicator of the strength  
214 of that signal. Figure 4 shows these two metrics for observations, reanalyses and testing GCMs, from 51  
215 different ANNs trained using randomly selected training GCMs. We also calculated the two metrics ( $r$   
216 and slope) for predicted-versus-actual year given for GCM simulations with radiative forcing held  
217 constant at pre-industrial levels, which is used as a measure of natural variability (see Methods). All  
218 observations and reanalysis have positive  $r$  values (Figure 4, Figure s4), even in datasets that do not  
219 show a significant positive trend in global-average Rx1day (Table S3). This contrast is because the ANN  
220 detects a signal in the spatial distribution of Rx1day, as opposed to the global average. The  $r$  values for  
221 all observational data sets are substantially larger than those expected by natural variability (grey  
222 shaded area in Figure 4e,f). When looking at the slope, two observational datasets (MSWEP and GPCC)  
223 are in line with GCMs, along with four reanalyses (JRA55, MERRA2, W5E5 and 20CRv3). The two REGEN  
224 datasets, ERA5 and CFSR show lower slopes, whereas NCEP2 has the highest slope among the datasets  
225 considered here. In general, observational and reanalysis products show similar  $r$  values and slopes as

226 the GCMs for the same historical time period (compare the black circles and the green bands in Figure  
227 4e, f). This indicates that the observational and reanalysis products show anthropogenic influence on  
228 Rx1day that is comparable to what is shown by GCMs.

229 To estimate the statistical significance of the signal detected in the observations and reanalysis,  
230 we first estimate the noise as the standard deviation of the distribution of the slopes representing  
231 natural variability (Figure 4e,f). Then S:N is calculated in all datasets by dividing the mean slope by noise.  
232 Following two-tailed z test, S:N larger than, for instance, 1.96 corresponds to a statistical significance  
233 level 95%<sup>25,52,53</sup>. Out of the four observations, MSWEP, GPCC and REGEN\_ALL show a 95% significance  
234 and REGEN\_LONG shows a 84% significance. Among reanalyses, ERA5 and CFSR show a 90% significance  
235 while the rest show a 95% significance.

236 These results demonstrate that the absence of a significant linear trend in globally averaged  
237 Rx1day cannot be taken to mean there is no evidence of anthropogenic signal in Rx1day. This  
238 underscores the importance of exploiting the spatial pattern of the response to external forcing to  
239 extract the forced signal in observations, as opposed to the trend-based analysis<sup>35,38,54</sup>. In particular,  
240 areas with higher internal variability can act to suppress the trend in the global mean. Further evidence  
241 of the importance of spatial patterns can be seen in the fact that the average ANN-predicted values vary  
242 widely and systematically across the observational datasets (Figure S4). This is an indicator of systematic  
243 and large relative biases in the Rx1day climatologies of the various data sets (as pointed out above in the  
244 discussion of ANN applied to the GCMs, the average predicted value of the year depends on the  
245 magnitude of the Rx1day in the climatology (Figure 3, supplementary text)). Yet it is significant that the  
246 ANN *can put the years in close to the correct order*, as demonstrated by the significant correlations  
247 between actual and predicted years, even if the absolute value of the years is incorrect. This is a strong  
248 indicator that the subtle patterns and time variations of the simulated anthropogenic signal are present

249 in the observational data sets and are shared among them, despite the fact that they are systematically  
 250 biased relative to one another and likely the real world<sup>55,56</sup>.  
 251



252  
 253 **Figure 4. Metrics of the forced signal in observation-based estimates of precipitation during 1982-2015.**  
 254 (a-d) Actual year vs predicted year obtained from 51 different ANNs with different training/validation/testing sets, for four  
 255 observational datasets, MSWEP (a), GPCP (b), REGEN\_ALL (c) and REGEN\_LONG (d). Green lines show results from each ANN.  
 256 The blue line is the mean slope. Each predicted year time series is standardized in the figure for a better comparison between  
 257 datasets. (e) Correlation ( $r$ ) between the actual years and predicted years, (f) slope of the regression line between actual years  
 258 and predicted years for observational and reanalysis data (black circle with a line), and testing models (green shaded regions).  
 259 Grey shading represents a measure of natural variability derived from 220 non-overlapping 34-year segments obtained from  
 260 pre-industrial control (piControl) simulations (see Methods).

261  
 262  
 263

## 264 **Conclusions**

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

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

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

291 Several caveats of the machine learning based detection method should be noted. Compared to  
292 regression based traditional D&A methods<sup>59</sup>, the assessment of the influence of individual forcings (e.g.  
293 anthropogenic aerosols, land-use change, and natural forcings such as volcanic and solar activities) in  
294 the presented framework is challenging. We did not attempt such a breakdown in this study, and this  
295 would require methodological modifications<sup>60</sup>. Additionally, the training GCMs might be under sampling  
296 the low frequency natural variability such as Atlantic Multidecadal variability and Pacific Decadal  
297 Oscillation. This may be remedied by inflating the training dataset with paleoclimate data<sup>61</sup>. However,  
298 underestimation of the precipitation response to natural forcings such as volcanic activities and natural  
299 variability such as El Nino Southern Oscillation in GCMs could affect the results<sup>62</sup>. We also note that  
300 different ANN visualization techniques are available<sup>50,63-65</sup>, and those should be explored to understand  
301 the sensitivity of the extracted fingerprints to the ANN visualization technique. Despite these limitations,  
302 it is clear that ANN DAI methods with ANN visualization techniques are very useful and efficient in  
303 identifying the human influence on variables that are highly uncertain in GCMs, and poorly characterized  
304 in observations, such as extreme precipitation.

305

## 306 **Data and Methods**

### 307 **Data**

308 We use daily precipitation rate output from a collection of climate models participating in  
309 CMIP5 and CMIP6 (Supplementary Table 1). Data from each ensemble's historical forcing scenario is  
310 combined with future projections following a high-emissions scenario to create a time-series from 1920  
311 to 2099 for each model. Future projections from CMIP5 follow the Representative Concentration

312 Pathway 8.5 (RCP 8.5)<sup>66</sup>, while CMIP6 projections follow the Shared Socioeconomic Pathway 5–8.5 (SSP  
313 5-8.5)<sup>67</sup>. To increase our sample size, we combine both CMIP5 and CMIP6 model subsets into one  
314 ensemble, which is justifiable considering the very similar time evolution of the total anthropogenic  
315 forcing in RCP 8.5 and SSP 5-8.5 scenarios (ref. 67, their Figure 3c). We regrid all daily precipitation data  
316 to a 2° x 2° spatial grid and compute the Rx1day value for each year at each land grid point.

317 To assess the influence of natural variability, we also use pre-industrial control simulations  
318 (piControl), which are GCM simulations with radiative forcing held constant at pre-industrial levels. As  
319 the length of the piControl simulations vary between GCMs, we selected the same number of samples  
320 from a collection of 20 CMIP6 models used here (Supplementary Table 2). We extract 34-year non  
321 overlapping samples from each simulation (so as to match the length of the observational record) to  
322 represent natural variability. Each GCM provided 14 samples of this length and after removing first 3  
323 samples of each simulation to avoid climate drift<sup>42</sup>, we were left with 220 pre-industrial samples with  
324 which to assess natural variability.

325 We use four datasets of observational estimates of daily precipitation rate with global  
326 coverage: Multi-Source Weighted-Ensemble Precipitation, version 2 (MSWEP)<sup>68</sup>, Global Precipitation  
327 Climatology Centre (GPCC) version 2018<sup>69</sup>, and Rainfall Estimates on a Gridded Network (REGEN)<sup>70</sup>,  
328 including both REGEN\_ALL and REGEN\_LONG. MSWEP is a hybrid reconstruction using *in situ*, satellite  
329 and reanalysis data, whereas GPCC and the REGEN datasets are developed from ground-based  
330 measurements. REGEN\_ALL is developed by interpolating all considered station data whereas  
331 REGEN\_LONG is developed using only the stations with a data record of 40 years or longer. We further  
332 use seven widely used reanalysis products for comparison: European Centre for Medium-Range  
333 Weather Forecasts ERA5<sup>71</sup>, Japanese 55-year Reanalysis (JRA55)<sup>72</sup>, Modern-Era Retrospective analysis for  
334 Research and Applications, Version 2 (MERRA2)<sup>73</sup>, NCEP Climate Forecast System Reanalysis (CFSR)<sup>74</sup>,  
335 the bias-corrected ERA5 precipitation dataset compiled for phase 3b of the inter-sectoral impact model

336 intercomparison project (W5E5)<sup>75,76</sup>, NCEP-DOE Reanalysis 2 (NCEP2)<sup>77</sup> and NOAA-CIRES-DOE Twentieth  
337 Century Reanalysis version 3 (20CRv3)<sup>78</sup>. These observational and reanalysis datasets are selected  
338 considering the availability of full global land coverage and data for at least three decades. We selected  
339 the period 1982-2015 for observational analysis as it is the common temporal range for all datasets. All  
340 observation and reanalysis data were regridded to the same 2° x 2° spatial grid as the models, and then  
341 Rx1day was calculated at each grid point for each year.

342

### 343 **Neural network-based detection method**

344 Here we apply the method in ref. 33 (see their Figure 1a) to predict the year with which given  
345 annual Rx1day maps from GCMs are associated, a regression task. This requires the ANN to learn the  
346 signature of the forced response in simulated Rx1day. By feeding the ANN data from forced simulations,  
347 it learns to distinguish the forced signal from internal climate variability. The use of multiple GCMs helps  
348 the ANN learn the common elements of the forced response most relevant to the prediction task, a  
349 process that fully considers model uncertainty as well as internal climate variability. Input to the ANN  
350 from each model is a vectorized spatial map of Rx1day (2° x 2° spatial grid = 16200 grid values) for each  
351 year from 1920 to 2099. Our primary goal is to detect the anthropogenic signal in extreme precipitation  
352 over land (excluding Antarctica). Thus, we mask out data over the ocean at this stage, resulting in 6082  
353 land grid values. The ANN architecture consists of two hidden layers with ten nodes each. The *Rectified*  
354 *Linear Unit* activation function is used for all hidden units.

355 Approximately 60% of the models (26) are used for training the ANN, while the rest of the  
356 models are divided equally to use as validation and testing sets (9 models each). The *mean squared error*  
357 between the actual and predicted year of Rx1day is used as the loss function to be minimized during the  
358 training. For the optimizer which updates the ANN based on the gradient of the loss, we select *rmsprop*.  
359 Climate variables inherently contain spatial autocorrelation. To account for this dependence among

360 adjacent input data points, we use L2 regularization<sup>79</sup> between inputs and the first hidden layer, which  
 361 adds the sum of squared weights as a penalty term to the loss function. By iterating over L2 values of  
 362 leading order of magnitudes and inspecting the tradeoff between low prediction error and  
 363 generalizability (Figure S1), we found L2=0.001 to be a suitable value for our analysis. We trained the  
 364 model for 1000 epochs. *Early stopping* was enabled to reduce the overfitting by monitoring the  
 365 *validation loss* with a *patience value* of 50 epochs<sup>80</sup>. We repeated the training process for 51 different  
 366 training sets obtained by random combinations of GCMs, resulting in 51 different ANNs. We found that  
 367 increasing the number of hidden units or changing the other hyperparameters did not result in a  
 368 substantial increase in accuracy.

369

### 370 **Neural network interpretation using Layerwise Relevance Propagation (LRP)**

371 Assume that for a given input map,  $x$ , we get an output  $f(x)$ , in our case, the predicted year.  
 372 LRP conservatively back-propagates this value through hidden layers until it reaches the input map. This  
 373 process generates a *relevance heatmap*, indicating the areas of importance influencing the value  $f(x)$ .  
 374 The conservation property is shown in eq.1, for relevance propagation between two hidden layers  $j$  and  
 375  $k$ , 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  
 376  $d$  input features. The summation operation for each hidden layer (e.g.  $\sum_k P_k$ ) is the summation of the  
 377 relevance ( $P$ ) of all hidden units in that layer, where  $P_k$  is the relevance of a single unit in layer  $k$ . The  
 378 activation,  $a_k$  (eq. 2) is the information coming from all units in layer  $j$ , to a target unit in layer  $k$ . In eq.  
 379 2,  $a_j$  values are the individual activations of each unit in the layer  $j$ ,  $w_{jk}$  values are the weights  
 380 associated with the relationship between each unit in layer  $j$  and the target unit in layer  $k$ , and  $b_k$  is the  
 381 bias of that target unit.

382

$$383 \quad \sum_{i=1}^d P_i = \dots = \sum_j P_j = \sum_k P_k = \dots = f(x) \quad (1)$$

384

$$385 \quad \mathbf{a}_k = \text{ReLU}(\sum_j \mathbf{a}_j \mathbf{w}_{jk} + \mathbf{b}_k) \quad (2)$$

386

$$387 \quad \mathbf{P}_j = \sum_k \left( \alpha \frac{a_j w_{jk}^+}{\sum_j a_j w_{jk}^+} - \beta \frac{a_j w_{jk}^-}{\sum_j a_j w_{jk}^-} \right) \mathbf{P}_k \quad (3)$$

388

389 The relevance propagation rule from layer  $k$  to a unit in layer  $j$  is given in eq. 3. This general  
390 form is also known as the  $\alpha\beta$ -rule<sup>39,40</sup>. The components  $()^+$  and  $()^-$  indicate only positive and negative  
391 weights are being considered, respectively. The  $\alpha$  and  $\beta$  coefficients represent the relative amount of  
392 positive and negative relevance to be propagated, respectively. As shown in eq. 3, positive relevance  
393 (i.e. excitatory influence) and negative relevance (i.e. inhibitory influence) are associated with positive  
394 and negative weights, respectively. The  $\alpha$  and  $\beta$  coefficients are to be chosen with the constraints  $\alpha -$   
395  $\beta = 1$  and  $\beta \geq 0$ . The combination  $\alpha = 2$  and  $\beta = 1$  (LRP <sub>$\alpha_2\beta_1$</sub> ) have been experimentally inferred as  
396 suitable values, and have been adopted in previous research<sup>39,40,81-83</sup>. Here we adopt the LRP <sub>$\alpha_2\beta_1$</sub>  rule  
397 (Supplementary Text).

398

### 399 Data availability

400 CMIP data used are available at <https://esgf-node.llnl.gov/projects/esgf-llnl/>. Observational and  
401 reanalysis data used are available at following links : MSWEP: <http://www.gloh2o.org/>, GPCC:  
402 <https://www.dwd.de/EN/ourservices/gpcc/gpcc.html>, REGEN\_ALL:  
403 <https://doi.org/10.25914/5ca4c380b0d44>, REGEN\_LONG: <https://doi.org/10.25914/5ca4c2c6527d2>,  
404 ERA5: <https://cds.climate.copernicus.eu/>, JRA55: <https://rda.ucar.edu/datasets/ds628.0/>, MERRA2:  
405 <https://disc.gsfc.nasa.gov/>, CFSR: <https://www.ncdc.noaa.gov/data-access/>, W5E5: [https://esg.pik-](https://esg.pik-potsdam.de/)  
406 [potsdam.de/](https://esg.pik-potsdam.de/), NCEP2 and 20CRv3: <https://psl.noaa.gov/data/>.

407

408 **Code availability**

409 Neural network analysis was conducted using Python libraries TensorFlow (<https://www.tensorflow.org>)  
410 and Keras (<https://keras.io>). Neural network interpretation was carried out using the library  
411 iNNvestigate (<https://github.com/albermax/innvestigate>). Python scripts developed for the analysis and  
412 figures are available publicly at

413 <https://dx.doi.org/10.6084/m9.figshare.14479659>.

414

415 **References**

- 416 1. Handmer, J. et al. Changes in impacts of climate extremes: human systems and ecosystems. in  
417 *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* (eds  
418 Field, C. B. et al.) 231–290 (Cambridge University Press, Cambridge, UK and New York, NY, 2012).
- 419 2. Crimmins, A. et al. *The impacts of climate change on human health in the United States: A Scientific*  
420 *Assessment*. Washington, DC: US Global Change Research Program.  
421 <https://doi.org/10.7930/JOR49NQX> (2016)
- 422 3. Allen, M. R. & Ingram, W. J. Constraints on future changes in climate and the hydrologic cycle.  
423 *Nature* **419**, 224–232 (2002).
- 424 4. O’Gorman, P. A. & Schneider, T. The physical basis for increases in precipitation extremes in  
425 simulations of 21st century climate change. *Proceedings of the National Academy of Sciences of the*  
426 *United States of America* **106**, 14773–14777 (2009).
- 427 5. Trenberth, K. E., Dai, A., Rasmussen, R. M. & Parsons, D. B. The changing character of precipitation.  
428 *Bulletin of the American Meteorological Society* **84**, 1205–1218 (2003).
- 429 6. Held, I. M. & Soden, B. J. Robust response of the hydrological cycle to global warming. *Journal of*

- 430 *Climate* **19**, 5686–5699 (2006).
- 431 7. Norris, J., Chen, G. & Neelin, J. D. Thermodynamic versus dynamic controls on extreme precipitation  
432 in a warming climate from the Community Earth System Model Large Ensemble. *Journal of Climate*  
433 **32**, 1025–1045 (2019).
- 434 8. Sun, Y., Solomon, S., Dai, A. & Portmann, R.W. How often does it rain?. *Journal of Climate* **19**, 916–  
435 934 (2006).
- 436 9. Fischer, E. M., Beyerle, U. & Knutti R. Robust spatially aggregated projections of climate extremes.  
437 *Nature Climate Change* **3**, 1033–1038 (2013).
- 438 10. Kharin, V. V., Zwiers, F. W., Zhang, X. & Wehner, M. Changes in temperature and precipitation  
439 extremes in the CMIP5 ensemble, *Climatic Change* **119**, 345–357 (2013).
- 440 11. Sillmann, J., Kharin, V.V., Zwiers, F. W., Zhang, X. & Bronaugh D. Climate extremes indices in the  
441 CMIP5 multimodel ensemble: Part 2. Future climate projections. *Journal of Geophysical Research*  
442 *Atmospheres* **118**, 2473–2493 (2013).
- 443 12. Madakumbura, G. D. et al. Event-to-event intensification of the hydrologic cycle from 1.5°C to a 2°C  
444 warmer world. *Scientific Reports* **9**, 3483 (2019).
- 445 13. Swain, D. L., Langenbrunner, B., Neelin, J. D. & Hall, A. Increasing precipitation volatility in twenty-  
446 first-century California. *Nature Climate Change* **8**, 427–433 (2018).
- 447 14. Allan, R. P. & Soden, B. J. Atmospheric warming and the amplification of precipitation extremes,  
448 *Science* **321**, 1481–1484 (2008).
- 449 15. Donat, M. G., Lowry, A. L., Alexander, L. V., O’Gorman, P. A. & Maher, N. More extreme precipitation  
450 in the world's dry and wet regions, *Nature Climate Change* **6**, 508–513 (2016).
- 451 16. Donat, M. G., Angélil, O. & Ukkola, A. M. Intensification of precipitation extremes in the world's  
452 humid and water-limited regions. *Environmental Research Letters* **14**, 065003 (2019).
- 453 17. Kirchmeier-Young, M. C. & Zhang, X. Human influence has intensified extreme precipitation in North

- 454 America. *Proceedings of the National Academy of Sciences of the United States of America* **117**,  
455 13308–13313 (2020).
- 456 18. Dong, S., Sun, Y., Li, C., Zhang, X., Min, S. K. & Kim, Y. H. Attribution of extreme precipitation with  
457 updated observations and CMIP6 simulations. *Journal of Climate* **34**, 871–881 (2021).
- 458 19. Paik, S., Min, S. K., Zhang, X., Donat, M. G., King, A. D., & Sun, Q. Determining the anthropogenic  
459 greenhouse gas contribution to the observed intensification of extreme precipitation. *Geophysical*  
460 *Research Letters* **47**, e2019GL086875 (2020).
- 461 20. Dong, S., Sun, Y. & Li, C. Detection of human influence on precipitation extremes in Asia. *Journal of*  
462 *Climate* **33**, 5293–5304 (2020).
- 463 21. Min, S. K., Zhang, X. B., Zwiers, F. W. & Hegerl, G. C. Human contribution to more-intense  
464 precipitation extremes. *Nature* **470**, 378–381 (2011).
- 465 22. Bindoff, N. L. et al. Detection and attribution of climate change: From global to regional. in *Climate*  
466 *Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment*  
467 *Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. et al.) 867–952  
468 (Cambridge University Press, Cambridge, UK and New York, NY, 2013).
- 469 23. Stott, P. et al. Detection and attribution of climate change: A regional perspective. *Wiley*  
470 *Interdisciplinary Reviews: Climate Change* **1**, 192–211 (2010).
- 471 24. Hegerl, G. & Zwiers, F. Use of models in detection and attribution of climate change. *Wiley*  
472 *Interdisciplinary Reviews: Climate Change* **2**, 570–591 (2011).
- 473 25. Marvel, K. & Bonfils, C. Identifying external influences on global precipitation. *Proceedings of the*  
474 *National Academy of Sciences* **110**, 19301–19306 (2013).
- 475 26. Easterling, D. R., Kunkel, K. E., Wehner, M. F. & Sun, L. Detection and attribution of climate extremes  
476 in the observed record. *Weather and Climate Extremes* **11**, 17–27 (2016).
- 477 27. Herold, N., Behrangi, A. & Alexander, L. V. Large uncertainties in observed daily precipitation

478 extremes over land. *Journal of Geophysical Research: Atmospheres* **22**, 668–681 (2017).

479 28. Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S. & Hsu, K. L. A review of global precipitation  
480 data sets: Data sources, estimation, and intercomparisons. *Reviews of Geophysics* **56**, 79–107  
481 (2018).

482 29. Roca, R. et al. FROGS: A daily 1° × 1° gridded precipitation database of rain gauge, satellite and  
483 reanalysis products. *Earth System Science Data*, **11**, 1017–1035 (2019).

484 30. Hegerl, G. C. et al. Challenges in quantifying changes in the global water cycle. *Bulletin of the*  
485 *American Meteorological Society* **96**, 1097–1115 (2015).

486 31. O'Gorman, P. A. Sensitivity of tropical precipitation extremes to climate change. *Nature Geoscience*  
487 **5**, 697–700 (2012).

488 32. Ribes, A., Zwiers, F. W., Azais, J.-M. & Naveau, P. A new statistical approach to climate change  
489 detection and attribution. *Climate Dynamics* **48**, 367–386 (2017).

490 33. Barnes, E. A., Hurrell, J. W., Ebert-Uphoff, I., Anderson, C. & Anderson, D. Viewing forced climate  
491 patterns through an AI Lens. *Geophysical Research Letters* **46**, 13389–13398 (2019).

492 34. Barnes, E. A., Toms, B., Hurrell, J.W., Ebert-Uphoff, I., Anderson, C., & Anderson, D. Indicator  
493 patterns of forced change learned by an artificial neural network. *Journal of Advances in Modeling*  
494 *Earth Systems* **12**, e2020MS002195 (2020).

495 35. Sippel, S., Meinshausen, N., Fischer, E. M., Szekely, E. & Knutti, R. Climate change now detectable  
496 from any single day of weather at global scale. *Nature Climate Change* **10**, 35–41 (2020).

497 36. Ebert-Uphoff, I., Samarasinghe, S. & Barnes, E. Thoughtfully using artificial intelligence in Earth  
498 science. *Eos* **100**. <https://doi.org/10.1029/2019eo135235> (2019).

499 37. Toms, B. A., Barnes, E. A. & Ebert-Uphoff, I. Physically interpretable neural networks for the  
500 geosciences: applications to earth system variability. *Journal of Advances in Modeling Earth Systems*  
501 **12**, e2019MS002002 (2020).

- 502 38. Wills, R.C., Sippel, S. & Barnes, E.A. Separating forced and unforced components of climate change:  
503 The utility of pattern recognition methods in Large Ensembles and observations. *Variations* **18**, 1–10  
504 (2020).
- 505 39. Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R. & Samek, W. On pixel-wise  
506 explanations for non-linear classifier decisions by layer-wise relevance propagation. *Plos One* **10**,  
507 e0130140 (2015).
- 508 40. Montavon, G., Samek, W. & Müller, K. R. Methods for interpreting and understanding deep neural  
509 networks. *Digital Signal Processing* **73**, 1–15 (2018).
- 510 41. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bulletin*  
511 *of the American Meteorological Society* **93**, 485–498 (2012).
- 512 42. Eyring, V. et al. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6)  
513 experimental design and organization. *Geoscientific Model Development Discussion* **8**, 10539–10583  
514 (2015).
- 515 43. Mora, C. et al. The projected timing of climate departure from recent variability. *Nature* **502**, 183–  
516 187 (2013).
- 517 44. King, A. D. et al. The timing of anthropogenic emergence in simulated climate extremes.  
518 *Environmental Research Letters* **10**, 94015 (2015)
- 519 45. Pfahl, S., O’Gorman, P. A. & Fischer, E. M. Understanding the regional pattern of projected future  
520 changes in extreme precipitation. *Nature Climate Change* **7**, 423–427 (2017).
- 521 46. Shepherd, T. G. Atmospheric circulation as a source of uncertainty in climate change projections.  
522 *Nature Geoscience* **7**, 703–708 (2014).
- 523 47. Hsu, P.-C. Global monsoon in a changing climate. in *The Monsoons and Climate Change* (eds  
524 Carvalho, L. M. V. & Jones, C.) 7–24 (Springer International Publishing, Springer Cham Heidelberg,  
525 New York, Dordrecht, London, 2016).

- 526 48. Yin, J. A consistent poleward shift of the storm tracks in simulations of 21st century climate.  
527 *Geophysical Research Letters* **32**, L18701 (2005).
- 528 49. Lu, J., Vecchi, G. A. & Reichler T. Expansion of the Hadley cell under global warming. *Geophysical*  
529 *Research Letters* **34**, L06805 (2007).
- 530 50. Scheff, J. & Frierson D. M. Robust future precipitation declines in CMIP5 largely reflect the poleward  
531 expansion of model subtropical dry zones. *Geophysical Research Letters*, **39**, L18704 (2012).
- 532 51. Rahmstorf, S., Foster, G. & Cahill, N. Global temperature evolution: recent trends and some pitfalls.  
533 *Environmental Research Letters* **12**, 054001 (2017).
- 534 52. Marvel, K. et al. External influences on modeled and observed cloud trends. *Journal of Climate* **28**,  
535 4820–4840 (2015).
- 536 53. Marvel, K., Cook, B. I., Bonfils, C. J., Durack, P. J., Smerdon, J. E. & Williams, A. P. Twentieth-century  
537 hydroclimate changes consistent with human influence. *Nature* **569**, 59–65 (2019).
- 538 54. Wills, R. C. J., Battisti, D. S., Armour, K. C., Schneider, T. & Deser, C. Pattern Recognition Methods to  
539 Separate Forced Responses from Internal Variability in Climate Model Ensembles and Observations.  
540 *Journal of Climate* **33**, 8693–8719 (2020).
- 541 55. Bador, M., Alexander, L. V., Contractor, S. & Roca, R. Diverse estimates of annual maxima daily  
542 precipitation in 22 state-of-the-art quasi-global land observation datasets. *Environmental Research*  
543 *Letters* **15**, 35005 (2020).
- 544 56. Alexander, L. V., Bador, M., Roca, R., Contractor, S., Donat, M. & Nguyen, P. Intercomparison of  
545 annual precipitation indices and extremes over global land areas from in situ, space-based and  
546 reanalysis products. *Environmental Research Letters* **15**, 055002 (2020).
- 547 57. Ribes, A., Qasmi, S. & Gillett, N. P. Making climate projections conditional on historical observations.  
548 *Science Advances* **7**, eabc0671 (2021).
- 549 58. Borodina, A., Fischer, E. M. & Knutti, R. Models are likely to underestimate increase in heavy rainfall

550 in the extratropical regions with high rainfall intensity. *Geophysical Research Letters* **44**, 7401–7409  
551 (2017).

552 59. Allen, M. R. & Stott, P. A. Estimating signal amplitudes in optimal fingerprinting, Part I: Theory.  
553 *Climate Dynamics* **21**, 477–491 (2003).

554 60. Székely, E., Sippel, S., Knutti, R., Obozinski, G., & Meinshausen, N. A direct approach to detection  
555 and attribution of climate change. Preprint at <https://arxiv.org/abs/1910.03346> (2019).

556 61. Min, S. K. Human fingerprint in global weather. *Nature Climate Change* **10**, 15–16 (2020).

557 62. Zhang, X. et al. Detection of human influence on twentieth-century precipitation trends. *Nature* **448**,  
558 461–465 (2007).

559 63. Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K. & Müller, K. R. *Explainable AI: Interpreting,*  
560 *Explaining and Visualizing Deep Learning* (Springer Nature, 2019).

561 64. Gilpin, L. H. et al. Explaining explanations: an overview of interpretability of machine learning. in  
562 *2018 IEEE International Conference on Data Science and Advanced Analytics (DSAA)* 80–89 (IEEE,  
563 2018).

564 65. Linardatos, P., Papastefanopoulos, V. & Kotsiantis, S. Explainable AI: A Review of Machine Learning  
565 Interpretability Methods. *Entropy* **23**, 18 (2021).

566 66. Meinshausen, M. et al. The RCP greenhouse gas concentrations and their extensions from 1765 to  
567 2300. *Climatic Change* **109**, 213–241 (2011).

568 67. O'Neill, B. C. et al. The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6.  
569 *Geoscientific Model Development* **9**, 3461–3482 (2016).

570 68. Beck, H. E. et al. MSWEP V2 global 3-hourly 0.1° precipitation: Methodology and quantitative  
571 assessment. *Bulletin of the American Meteorological Society* **100**, 473–500 (2019).

572 69. Ziese, M, Rauthe-Schöch, A, Becker, A, Finger, P, Meyer-Christoffer, A & Schneider, U. GPCP full data  
573 daily version. 2018 at 1.0°: Daily land-surface precipitation from rain-gauges built on GTS-based and

574 historic data: Global Precipitation Climatology Centre (GPCC, <http://gpcc.dwd.de/>) at Deutscher  
575 Wetterdienst (2018).

576 70. Contractor, S. et al. Rainfall Estimates on a Gridded Network (REGEN)—A global land-based gridded  
577 dataset of daily precipitation from 1950–2013. *Hydrology and Earth System Sciences* **24**, 919–943  
578 (2020).

579 71. Hersbach, H. et al. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*  
580 **2020**, 1–51 (2020).

581 72. Kobayashi, S. et al. The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of*  
582 *the Meteorological Society of Japan* **93**, 5–48 (2015).

583 73. Gelaro, R. et al. The Modern-Era Retrospective Analysis for Research and Applications, Version 2  
584 (MERRA-2). *Journal of Climate* **30**, 5419–5454 (2017).

585 74. Saha, S. et al. The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological*  
586 *Society* **91**, 1015–1057 (2010).

587 75. Lange, S. WFDE5 over land merged with ERA5 over the ocean (W5E5). V. 1.0. GFZ Data Services.  
588 <https://doi.org/10.5880/pik.2019.023> (2019).

589 76. Cucchi, M. et al. WFDE5: bias-adjusted ERA5 reanalysis data for impact studies. *Earth System Science*  
590 *Data* **12**, 2097–2120 (2020).

591 77. Kanamitsu, M. et al. NCEP-DOE AMIP-II reanalysis (R-2). *Bulletin of the American Meteorological*  
592 *Society* **83**, 1631–1643 (2002).

593 78. Slivinski, L. C. et al. Towards a more reliable historical reanalysis: Improvements for version 3 of the  
594 Twentieth Century Reanalysis system. *Quarterly Journal of the Royal Meteorological Society* **145**,  
595 2876–2908 (2019).

596 79. Hastie, T. Ridge regularization: An essential concept in data science. *Technometrics*, **62**, 426–433  
597 (2020).

- 598 80. Géron, A. *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and*  
599 *Techniques to Build Intelligent Systems* (O'Reilly Media, 2017).
- 600 81. Böhle, M., Eitel, F., Weygandt, M. & Ritter, K. Layer-wise relevance propagation for explaining deep  
601 neural network decisions in MRI-based Alzheimer's disease classification. *Frontiers in aging*  
602 *neuroscience* **11**, 194 (2019).
- 603 82. Dobrescu, A., Valerio Giuffrida, M. & Tsiftaris, S. A. Understanding deep neural networks for  
604 regression in leaf counting. in *Proceedings of the IEEE Conference on Computer Vision and Pattern*  
605 *Recognition Workshops*, 2600–2608. Long Beach, CA, USA: IEEE (2019).
- 606 83. Grigorescu, I., Cordero-Grande, L., Edwards, A. D., Hajnal, J. V., Modat, M. & Deprez, M. Investigating  
607 Image Registration Impact on Preterm Birth Classification: An Interpretable Deep Learning  
608 Approach. in *Smart Ultrasound Imaging and Perinatal, Preterm and Paediatric Image Analysis*, 104–  
609 112, (Springer, Cham., 2019).
- 610 84. Hawkins, E. et al. Observed Emergence of the Climate Change Signal: From the Familiar to the  
611 Unknown. *Geophysical Research Letters* **47**, e2019GL086259 (2020).
- 612 85. Alber, M. et al. iNNvestigate neural networks! *Journal of Machine Learning Research* **20**, 1–8 (2019).

613

#### 614 **Acknowledgements**

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

621

622 **Author contributions**

623 GDM, CWT, JN and AH designed research. GDM developed the detection framework, conducted the  
624 analyses, and wrote the paper. All authors contributed to the interpretation of the results and edited  
625 the paper.

626

627 **Competing interests**

628 The authors declare no competing interests.

# Figures

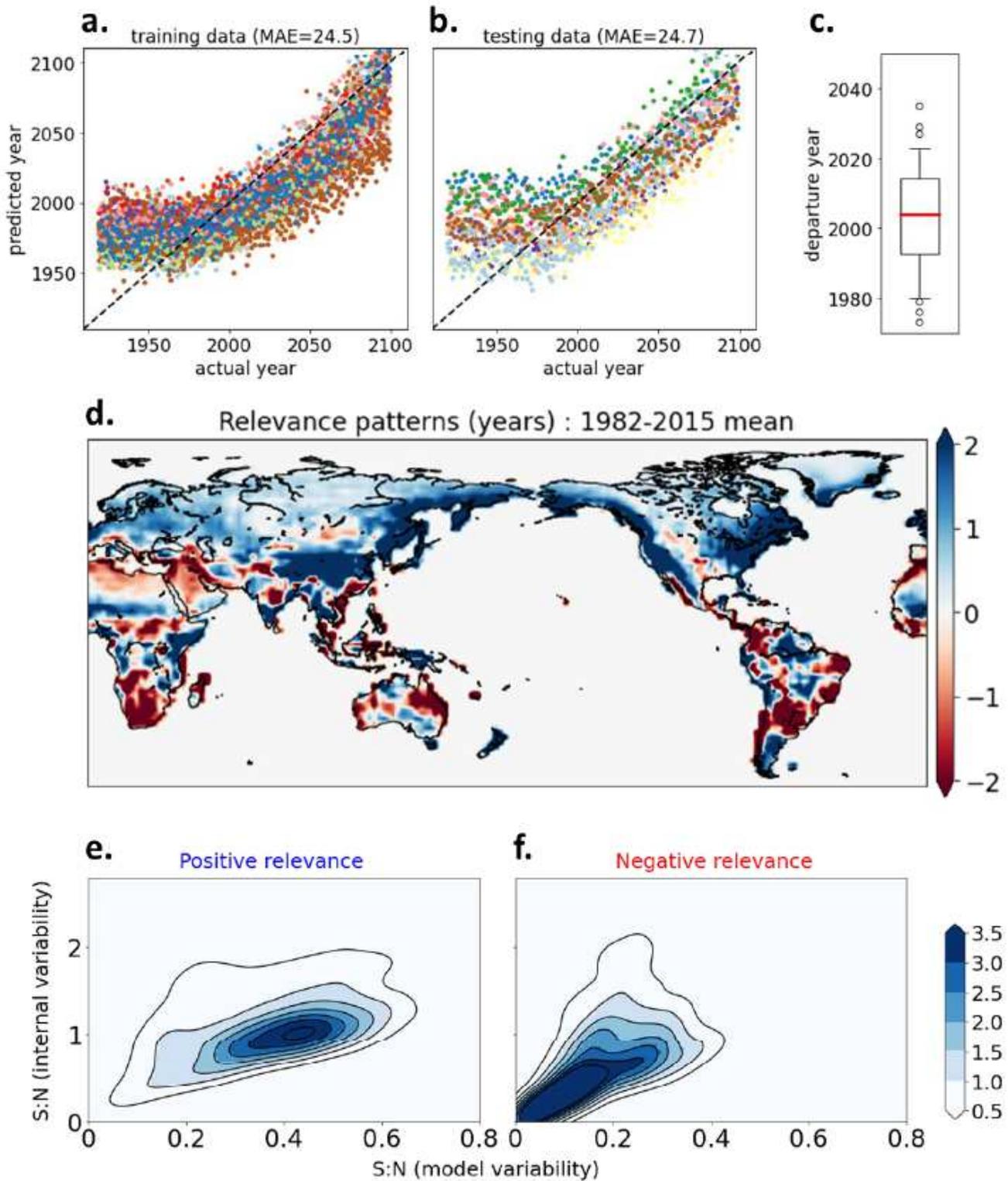
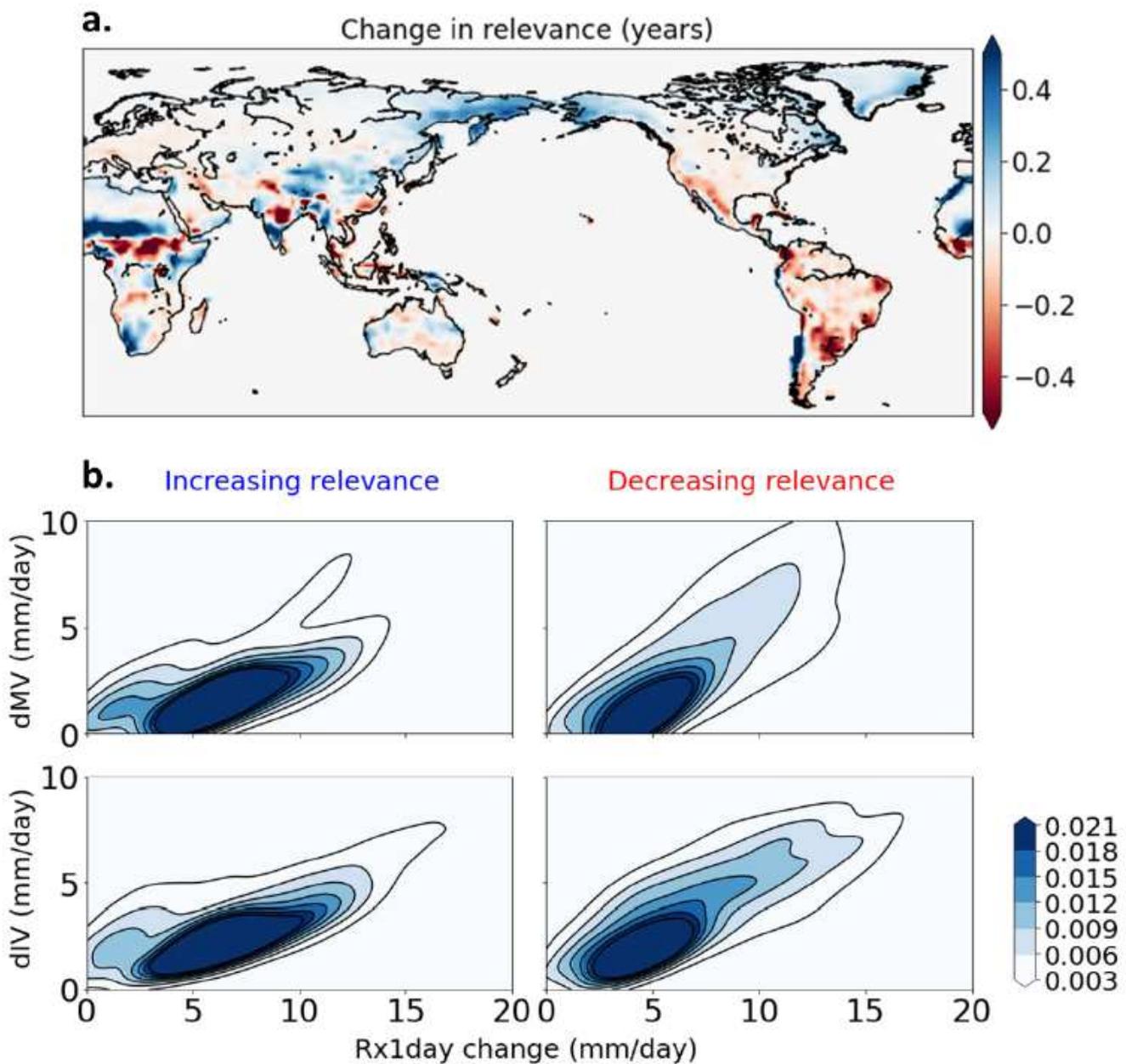


Figure 1

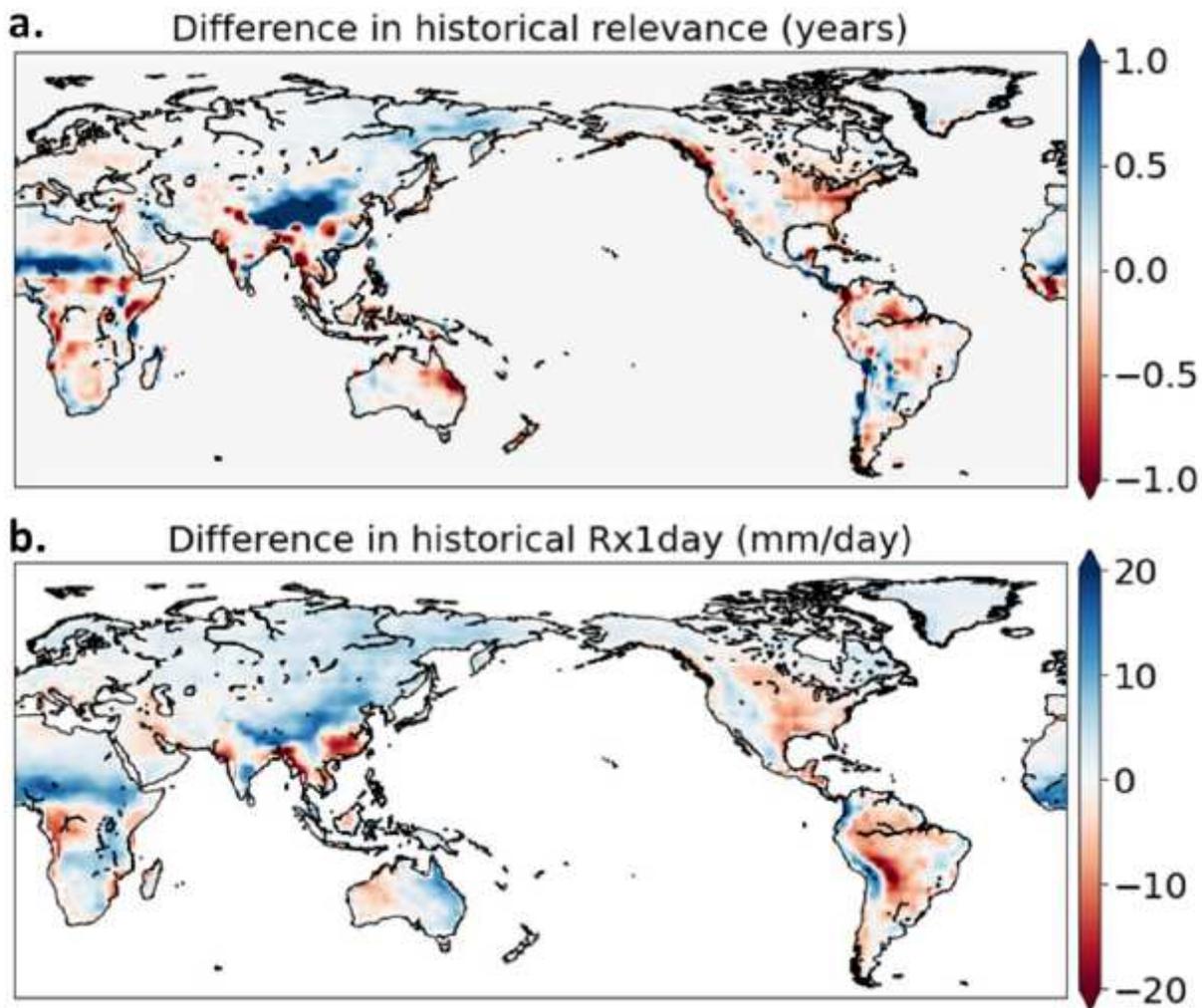
Fingerprint of external forcing in simulated Rx1day learned by the ANN. (a,b) Actual year vs predicted year for training data derived from CMIP5 and CMIP6 GCMs (a) and testing data derived from CMIP5 and CMIP6 GCMs (b) for a single ANN. Each GCM is represented by a different color. (c) The year of departure

from the base period, 1920-1949. Whiskers represent the 5th-95th percentiles, while blank circles represent outliers. (d) Multimodel, ensemble-mean, layerwise-relevance-propagation-based relevance maps for Rx1day input for the period 1982-2015 from all models. (e,f) Signal-to-noise ratio density plots for grid cells with a positive relevance (e) and negative relevance (f) in panel (d). Signal is defined as the multi-model 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 inter-model variability, and is calculated as the inter-model standard deviation of the signal from each GCM. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.



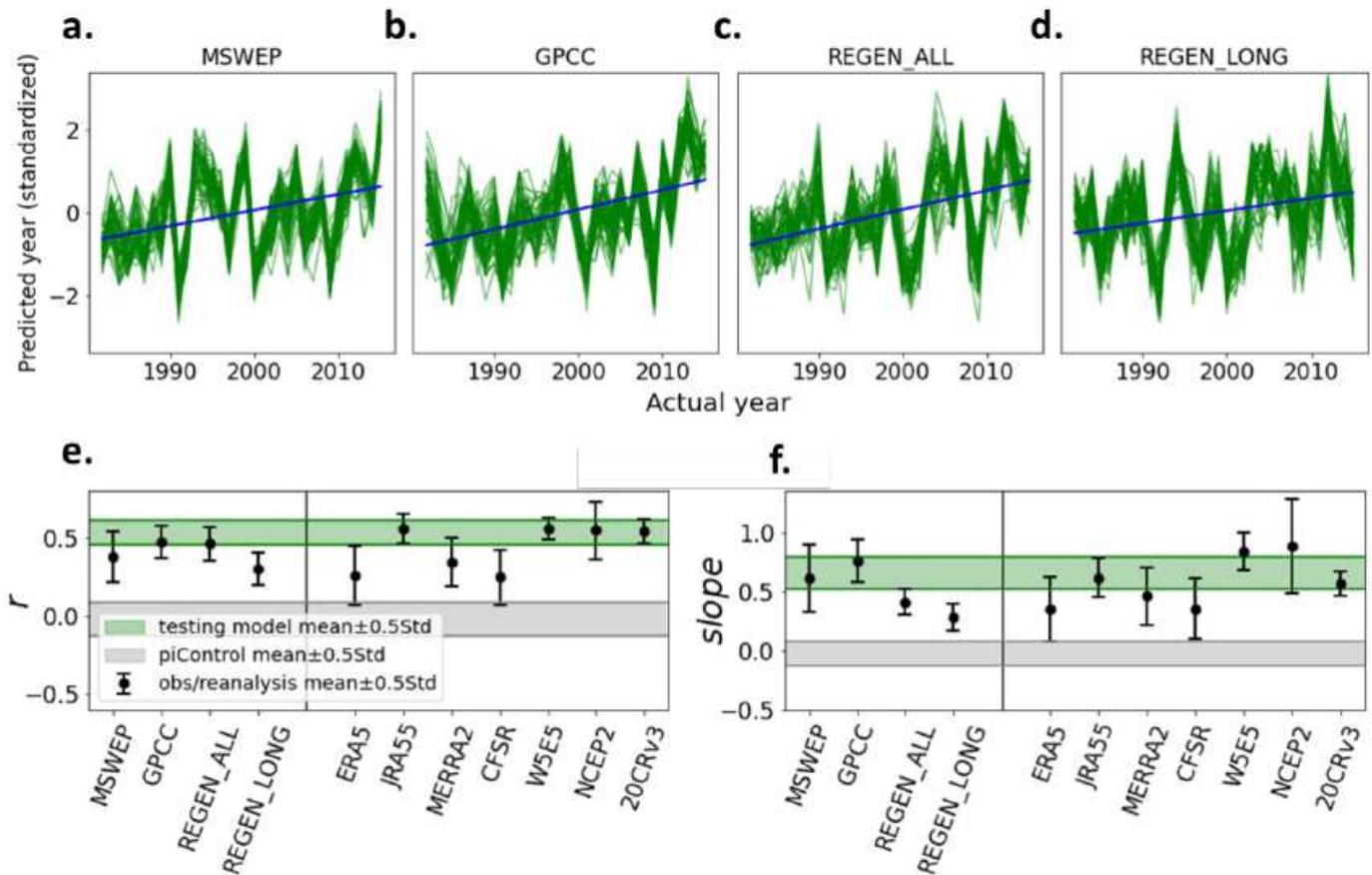
**Figure 2**

Change in the relevance patterns learned by the ANN through time. (a) Multimodel average change of relevance maps between 2070-2099 and 1920-1949. (b) Multimodel ensemble mean change in Rx1day vs change in inter-model 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. 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 timeseries and inter-model variability is calculated as the standard deviation of mean Rx1day from all models for each time period. Prior to 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>84</sup>. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.



**Figure 3**

Differences between subsets of models with high and low predicted years by the ANN during the baseline period (1920-1949). (a,b) The difference in their relevance maps (a) and Rx1day (b) between the four models with the highest mean predicted year and the four models with lowest mean predicted year (as shown in Figure 1a-b). Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.



**Figure 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 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). Grey shading represents a measure of natural variability derived from 220 non-overlapping 34-year segments obtained from pre-industrial control (piControl) simulations (see Methods).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarytext.pdf](#)