

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 Global climate models produce large increases in extreme precipitation when subject  
11 to anthropogenic forcing, but detecting this human influence in observations is challenging.  
12 Large internal variability makes the signal difficult to characterize. Models produce diverse  
13 precipitation responses to anthropogenic forcing, mirroring a variety of parameterization  
14 choices for subgrid-scale processes. And observations are inhomogeneously sampled in space  
15 and time, leading to multiple global datasets, each produced with a different homogenization  
16 technique. Thus, previous attempts to detect human influence on extreme precipitation have  
17 not incorporated internal variability or model uncertainty, and have been limited to specific  
18 regions and observational datasets. Using machine learning methods, we find a physically  
19 interpretable anthropogenic signal that is detectable in all global datasets. Detection occurs  
20 even when internal variability and model uncertainty are taken into account. Machine  
21 learning efficiently generates multiple lines of evidence supporting detection of an  
22 anthropogenic signal in extreme precipitation.

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

24

## 25 Introduction

26 Extreme precipitation can have devastating direct societal impacts such as flooding, soil  
27 erosion, agricultural damages<sup>1</sup> and indirect health risks and impacts<sup>2</sup>. Anthropogenic warming acts to  
28 intensify Earth's hydrologic cycle<sup>3,4</sup>. This intensification is manifested in part through increased extreme  
29 precipitation as a result of greater atmospheric moisture with warming following the Clausius-Clapeyron  
30 relationship. However, circulation changes can act to enhance or reduce this increase<sup>4-7</sup>. If current  
31 warming trends continue, climate models project that the Earth's atmosphere overall will move towards  
32 a more intense precipitation regime<sup>8-11</sup>. Moreover, increased variation between wet and dry extremes is  
33 projected, which could have devastating societal impacts<sup>12,13</sup>. These changes in extreme precipitation  
34 may have already become apparent on a regional basis<sup>14-17</sup>.

35 Recent studies have detected anthropogenic influence in historical changes to extreme  
36 precipitation across North America<sup>18</sup> and Northern Hemisphere land areas<sup>15</sup>. These attempts are part of  
37 a larger category of studies known as *Detection and Attribution* (D&A). Often, they initially extract the  
38 spatial or spatiotemporal patterns of climate-system response to anthropogenic forcing (so-called  
39 *fingerprints*) from an ensemble of global climate models (GCMs)<sup>19,20</sup>. Projection of observations onto  
40 these fingerprints allows for signal detection as the trend of the projection<sup>21</sup>. The presence of a signal  
41 that can be statistically distinguished from internal variability confirms the influence of anthropogenic  
42 forcing. Thus, traditional D&A methods rely on long term observations<sup>22,23</sup>. In the case of extreme  
43 precipitation, traditional methods may be difficult to apply globally due to inordinately short records  
44 and large observational uncertainty, reflected in multiple global datasets produced with very different  
45 assumptions<sup>24-26</sup>. Another key difficulty with traditional methods is that the models produce a large  
46 spread in the extreme precipitation response to anthropogenic forcing<sup>27</sup>. This spread occurs alongside  
47 large internal variability in the models' simulations of the historical period. These two effects create  
48 significant uncertainty in the character of the "true" anthropogenic signal. In past research, spread in

49 the response have been suppressed by assuming the anthropogenic fingerprint can be derived from the  
50 ensemble-mean change in extreme precipitation<sup>28</sup>. Here we aim to take these uncertainties fully into  
51 account, by making no assumptions about how to derive the anthropogenic signal from GCM data.

52 A machine-learning-based method for the detection of anthropogenic influence (DAI) has been  
53 shown to overcome the reliance on trends<sup>29,30</sup> and is even capable of detecting the human influence  
54 from weather data on a single day<sup>31</sup>. An artificial neural network (ANN) is trained to predict a proxy of  
55 external forcing (e.g. the year of the data) based on the spatial maps of the target variable from an  
56 ensemble of GCM simulations. Then a forced signal can be confirmed despite the presence of internal  
57 climate variability and inter-model variability<sup>29,30</sup>. This ANN DAI method can identify the non-linear  
58 combinations of the forced signal, internal climate variability and inter-model variability<sup>30</sup>. This method  
59 also has the advantage of being able to explicitly include internal variability and model uncertainty. It  
60 does not assume that any model or any model-derived quantity, such as the ensemble-mean of the  
61 models, is the “true” anthropogenic signal. And it uses the raw GCM data, with GCM internal variability  
62 included. In addition, ANN visualization techniques also allow for the interpretability of the models  
63 formerly considered as “black boxes”, making them *explainable*<sup>32,33</sup>, or interpretable in terms of physical  
64 processes or system behavior. Use of these visualization techniques alongside the ANN DAI method  
65 allows one to capture the time varying dynamic fingerprints of each input and evaluate their physical  
66 credibility<sup>30,34</sup>.

67 In this study, we apply the ANN DAI method and the ANN visualization technique known as  
68 Layerwise Relevance Propagation (LRP)<sup>35,36</sup> to global maps of annual daily maximum precipitation  
69 (Rx1day) over land. Using Coupled Model Intercomparison Project, phase 5 (CMIP5)<sup>37</sup> and phase 6  
70 (CMIP6)<sup>38</sup> model ensembles, we first aim to understand how the ANN is detecting the anthropogenic  
71 signal and interpret it physically. Then we use the ANN to detect the anthropogenic influence on Rx1day  
72 in several land-only observational and reanalysis datasets. Thus, we are agnostic about which GCM is

73 correct, and which gridded data set is a true representation of the observed record. In this way we  
74 efficiently generate multiple lines of evidence as to the presence of an anthropogenic signal in the  
75 various instantiations of the observed record.

76

## 77 **Results**

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

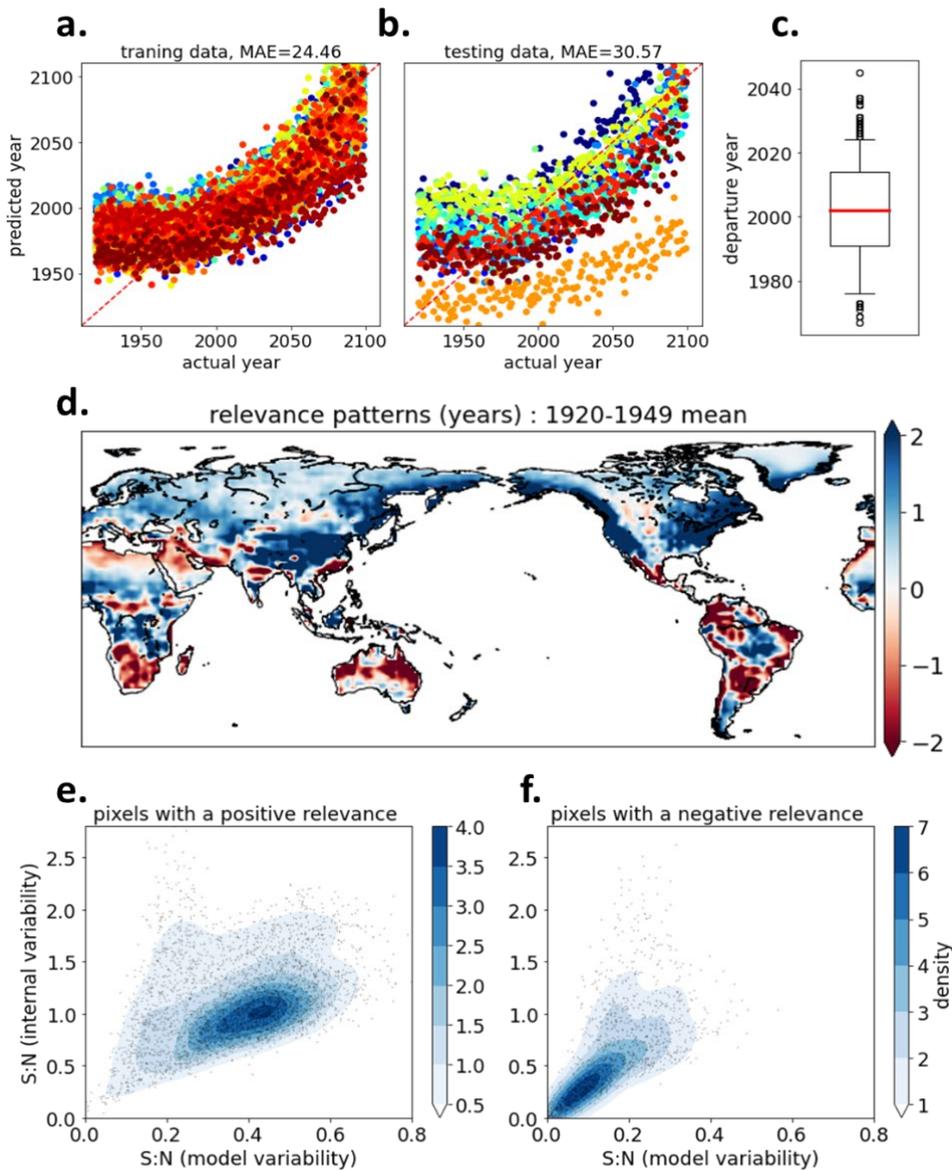
79 We first discuss the ability of the ANN to predict the year of occurrence for a series of simulated  
80 annual Rx1day maps. Predictions of the simulated Rx1day year (Figure 1a,b) show that the ANN  
81 struggles during roughly the first seven to eight decades of the analysis period. But prediction accuracy  
82 gradually increases, noticeably starting from the late 20<sup>th</sup> century. Thus, compared to when this  
83 technique is applied to global-mean temperature (ref. 29) there is a lag in the emergence of the  
84 anthropogenic signal in extreme precipitation. This delay is likely due to larger internal and inter-model  
85 variability in extreme precipitation. We estimate this time of emergence (*departure year*) as the year  
86 when the ANN prediction continuously exceeds a selected base period (1920-1949) (see ref. 29 and 39  
87 for more details). Some maps of simulated terrestrial Rx1day drawn from various populations of GCM  
88 training and testing data sets depart from the base period in the 1970s, but the departures mostly occur  
89 later, with lower and upper quartiles of 1991 and 2014, respectively (Figure 1c). According to the  
90 models, the anthropogenic signal has probably already emerged in Rx1day, consistent with traditional  
91 statistical methods<sup>40</sup>.

92 Figure 1d shows the relevance pattern identified by the ANN, averaged over the period 1920-  
93 1949. Positive (negative) values in the relevance pattern correspond to an increase (decrease) in the  
94 predicted year. Therefore, areas of positive relevance can be interpreted as the regions with an  
95 advancing tendency on the prediction (i.e. the year) and negative values are the regions with a  
96 retreating tendency. The sum of each grid cell value is equal to the predicted year (methods, Figure

97 S1d). These relevance patterns can be considered as the ANN-identified fingerprints of anthropogenic  
98 influence (e.g. ref. 31).

99         The regions with positive relevance include the East Asian and African monsoon regions, and the  
100 North Pacific and Atlantic storm tracks (Figure 1d, Figure S2). The regions with negative relevance  
101 include arid and semi-arid subtropical zones such as Northern African and Middle Eastern deserts,  
102 Southern South Africa, Australian arid and semi-arid regions, and wet regions such as central and  
103 northwestern parts of South America. Regions with negative relevance also coincide with areas  
104 exhibiting a large negative dynamical component of the Rx1day trend (ref. 41, their Figure 3b). These  
105 regions show a significant anthropogenic reduction of vertical velocities associated with Rx1day. This  
106 offsets the Rx1day increase stemming from the thermodynamic contribution, and produces only a weak  
107 and inconsistent increase in Rx1day<sup>41</sup>. The uncertainty associated with the dynamical component has  
108 been identified as a major concern for D&A of precipitation<sup>42</sup>. As suspected, negative relevance of the  
109 forced response is associated with lower signal to noise ratios than the regions with positive relevance  
110 (Figure 1e,f). The signal to noise ratios are lower for both internal variability and model variability. This  
111 reflects both the higher uncertainty regarding the change in extreme precipitation projected by GCMs  
112 for a majority of global arid land regions, as well as larger internal variability in those regions. Here we  
113 examined the time-averaged relevance patterns, but further examination of time-varying fingerprints  
114 shows that the ANN is also able to identify the time-varying nature of the signal and noise  
115 (Supplementary Text), which cannot be obtained directly by linear models<sup>34</sup>.

116         The ANN-based relevance patterns are consistent with the idea that previously observed long-  
117 term trends of terrestrial Rx1day are anthropogenic in origin (e.g. ref. 15, their Figure 1e). Many wet  
118 land regions of the world have experienced a robust increase in Rx1day to date, whereas in dry regions  
119 no such trend can be seen<sup>16,17</sup>. The selection of regions in these previous studies (e.g. ref. 16, 17) seems  
120 to overlap with the high relevance regions in Fig.1d.



122

123 **Figure 1. Fingerprint of external forcing in simulated Rx1day learned by the ANN.**

124 (a,b) Actual year vs predicted year for training data derived from CMIP5 and CMIP6 GCMs (a) and testing data derived from  
 125 CMIP5 and CMIP6 GCMs (b) for a single ANN. Each GCM is represented by a different color. (c) The year of departure from the  
 126 base period, 1920-1949, obtained from 51 different ANNs with different training/testing sets. Whiskers represent the 5<sup>th</sup>-95<sup>th</sup>  
 127 percentiles, while blank circles represent outliers. (d) Multimodel, ensemble-mean, layerwise-relevance-propagation-based  
 128 relevance maps for Rx1day input for the period 2070-2099 from all models. (e,f) Signal to noise ratio density plots for grid cells  
 129 with a positive relevance (e) and negative relevance (f) in panel (d). Signal is defined as the multi-model mean change in Rx1day  
 130 between the base period and 1920-1949. Noise is defined in two ways: The first stems from internal variability and is calculated

131 as the multimodel ensemble mean of the standard deviation in Rx1day during the base period. The second pertains to inter-  
132 model variability, and is calculated as the inter-model standard deviation of the signal from each GCM. Each dot in (e-f)  
133 corresponds to one grid cell.

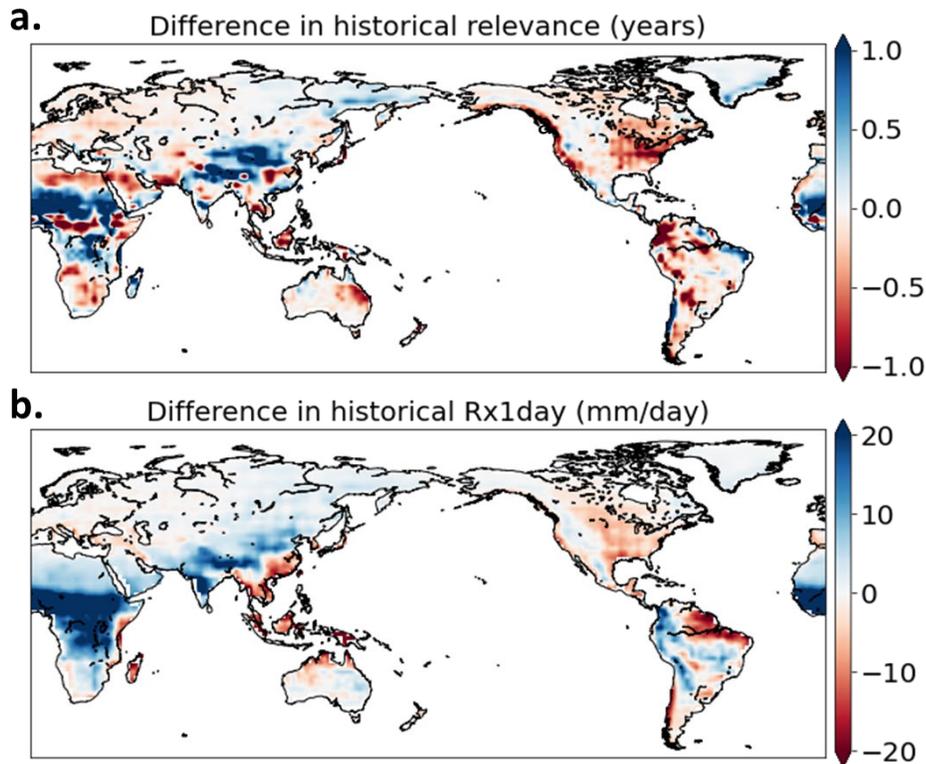
134

135

### 136 **Origins of the spread in the predicted year**

137         We next investigate why the ANN predicts such a large range of years depending on the data of  
138 the underlying GCM used to predict the year. This inter-model spread in the predicted year is especially  
139 pronounced before the warming signal emerges (Figure 1a,b). Here, we select four GCMs with the  
140 highest average predicted year, and four GCMs with the lowest average predicted year, during the  
141 baseline period (1920-1949). We obtain the relevance heatmaps for each year of the baseline period for  
142 these eight models and calculate the composite difference (i.e. high value minus low value) between  
143 two sets (Fig 2a). Large positive values are seen in the African and Asian monsoon regions. The models  
144 predicting later years also have larger 20<sup>th</sup> century mean state Rx1day values in these regions (Figure  
145 2b). Thus, the models that predict a higher value in the baseline period have *more future-like patterns* of  
146 Rx1day in their baseline climatologies compared to other models. When projected onto the fingerprints  
147 identified by the ANN, these patterns result in a later predicted year compared to the opposite subset.  
148 This exercise suggests a potential use of ANN-based DAI methods to understand how biases in historical  
149 simulations project onto future changes.

150



151

152 **Figure 2. Differences between subsets of models with high and low predicted years by the ANN during the baseline period**  
 153 **(1920-1949).** (a,b) The difference in their relevance maps (a) and Rx1day (b) between the four models with the highest mean  
 154 predicted year and the four models with lowest mean predicted year, (as shown in Figure 1a-b). The models from each subset  
 155 are determined by combining the training and testing data.

156

157

158 **Detected anthropogenic signal in historical Rx1day records**

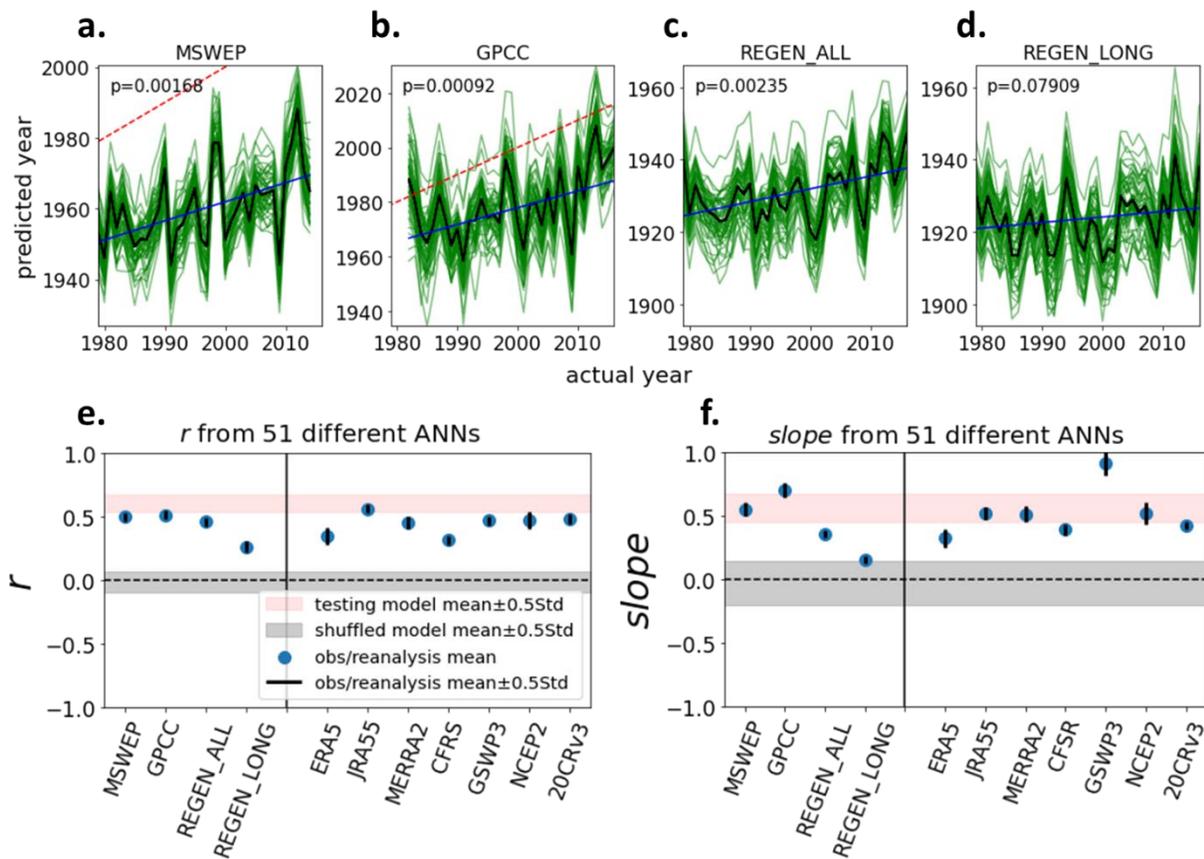
159 With these physical interpretations of the ANN results and relevance patterns, we use the GCM-  
 160 trained ANNs to detect whether there is a forced signal in observations. First, we calculate the globally-  
 161 averaged Rx1day trends in each dataset using a modified Mann-Kendall trend test<sup>43</sup>. Only seven out of  
 162 the eleven datasets show a significant trend ( $p < 0.01$ ) in globally-averaged Rx1day for the historical  
 163 period, ranging from 0.05 to 0.09 mm/day/year (Table S2). Taken at face value, this suggests that the  
 164 evidence for anthropogenic influence on recent changes in extreme precipitation is weak. However,

165 when we apply the ANN, based on Rx1day data from GCMs, to four datasets of observational  
166 precipitation estimates and seven reanalyses, a different story emerges.

167 If an observational dataset exhibits the same forced response as the GCMs, the predicted year  
168 time series from that dataset should have a statistically significant positive correlation with the actual  
169 year ( $r$ ) and a linear regression of these two variables should produce a statistically significant positive  
170 slope<sup>21,30</sup>. Figure 3 shows these two metrics for observations, reanalysis and testing GCMs, from 51  
171 random iterations of the ANN with different training/testing model sets. All observations and reanalysis  
172 have high  $r$  values (Figure 3, Figure s3), even in datasets that do not show a significant positive trend in  
173 global Rx1day record (Table S2). The  $r$  values for all observational data sets are substantially larger than  
174 those expected by chance (grey shaded area in Figure 3e,f). The slope is a measure of signal strength (or  
175 the rate of change) in the Rx1day record. Two observational datasets (MSWEP and GPCC) are in line with  
176 GCMs, along with four reanalyses (JRA55, MERRA2, NCEP2 and 20CRv3). The two REGEN datasets, ERA5  
177 and CFSR show lower slopes, whereas GSWP3 has the highest slope among the datasets considered  
178 here. In general, observational and reanalysis products show similar correlations and slopes as the GCMs  
179 for the same historical time period. (Compare the blue dots and the pink bands in Figure 3e, f.) This  
180 indicates that the observational and reanalysis products show anthropogenic influence on Rx1day that is  
181 comparable to what is expected from GCMs.

182 These results demonstrate that the absence of a significant linear trend in globally averaged  
183 Rx1day cannot be taken to mean there is no evidence of anthropogenic signal in Rx1day. This  
184 underscores the importance of exploiting the spatial pattern of the response to external forcing to  
185 extract the forced signal in observations, as opposed to the trend-based analysis<sup>31,34,44</sup>. In particular,  
186 areas of negative relevance, defined previously, can act to suppress the trend in the global mean.  
187 Further evidence of the importance of spatial patterns can be seen in the fact that the average ANN-  
188 predicted values vary widely and systematically across the observational datasets (Figure 3a-d, Figure

189 S3). This is an indicator of systematic and large relative biases in the Rx1day climatologies of the various  
 190 data sets (as pointed out above in the discussion of ANN applied to the GCMs, the average predicted  
 191 value of the year depends on the magnitude of the Rx1day in the climatology (Figure 2, supplementary  
 192 text)). Yet it is significant that the ANN *can put the years in close to the correct order*, as demonstrated  
 193 by the significant correlations between actual and predicted years, even if the absolute value of the  
 194 years is incorrect. This is a strong indicator that the subtle patterns and time variations of the simulated  
 195 anthropogenic signal are present in the observational data sets and are shared among them, despite the  
 196 fact that they are systematically biased relative to one another and likely the real world<sup>45,46</sup>.  
 197



198  
 199 **Figure 3. Metrics of the forced signal in observation-based estimates of precipitation since 1979.**

200 (a-d) Actual year vs predicted year obtained from 51 different ANNs with different training/testing sets, for four observational  
 201 datasets, MSWEP (a), GPCC (b), REGEN\_ALL (c) and REGEN\_LONG (d). Green lines show results from each ANN and the black

202 line shows the mean predicted value from all ANNs. The blue line is the best fit line of the mean predicted value. The red  
203 dashed line is the 1:1 line. (e) Correlation ( $r$ ) between the actual years and predicted years, (f) slope of the regression line  
204 between actual years and predicted years for observational and reanalysis data (blue circle with black line), and testing models  
205 (red and grey shaded regions). Grey shading illustrates the values obtained from testing models after randomly shuffling, for  
206 each iteration of training/testing sets. Slope is obtained using the Theil–Sen method<sup>47,48</sup>.

207

208

## 209 **Conclusions**

210 Detecting anthropogenic signals in observations of extreme precipitation has been a challenging  
211 task due to large internal variability of rare events, as well as climate model uncertainty. The limited  
212 sampling in observations adds additional uncertainty, due in part to a dataset development process that  
213 involves a variety of homogenization, extrapolation, and interpolation techniques to produce global  
214 gridded products. Using a recently introduced ANN DAI method which utilizes the time evolution of  
215 spatial maps of Rx1day in GCMs, subject to realistic radiative forcing, we find fingerprints of  
216 anthropogenic signals that are physically consistent with the time evolution of the forced signal. The  
217 fingerprints can be distinguished from internal variability, and emerge despite substantial model  
218 uncertainty. Using this ANN DAI method, we show that the anthropogenic signal can be detected in all  
219 global terrestrial Rx1day records considered in this study. This robust detection occurs despite large  
220 systematic biases and large discrepancies in data sources and homogenization methods.

221 While previous trend-based D&A assessments of Rx1day have demonstrated the human  
222 influence in this variable in some regions, those studies assume the ensemble-mean of the GCMs is the  
223 anthropogenic signal. This leads to questions as to whether further steps are needed to fully consider  
224 model uncertainty<sup>28</sup>. We made a simple attempt to examine this issue by applying the ANN DAI method  
225 to the same widely-used, quality-controlled Rx1day record used in the previous trend-based D&A

226 assessments. We applied the method twice, once using the same multi-model approach discussed  
227 elsewhere in this study, and once assuming the ensemble-mean time series represents the true forced  
228 response. Our results show that including internal variability and model uncertainty in the forced  
229 response reduces the power of detection (supplementary text). Therefore, the detected signal in  
230 multiple datasets in this study, with internal variability and model uncertainty being taken fully into  
231 account (Figure 3) is a definitive affirmation of a human influence on extreme precipitation in the  
232 historical record. Note that while all observations show this anthropogenic influence, the signal  
233 magnitude varies considerably, on par with that seen in the GCMs. This large observational uncertainty  
234 underscores a difficulty in constraining future projections of extreme precipitation with historical  
235 climate model simulations and observations<sup>27,49</sup>.

236 A limitation of the ANN DAI method presented here is the inability to directly quantify the  
237 detected anthropogenic influence in terms of physical units. This is especially important when  
238 comparing the influence of individual external forcings (e.g. greenhouse gases, aerosols, land use and  
239 land cover change, etc.). We also note that different ANN visualization techniques are available<sup>50</sup>, and  
240 those should be explored to understand the sensitivity of the extracted fingerprints to the ANN  
241 visualization technique. Despite these limitations, it is clear that ANN DAI methods with ANN  
242 visualization techniques are very useful and efficient in identifying the human influence on variables that  
243 are highly uncertain in GCMs, and poorly characterized in observations, such as extreme precipitation.

244

245

## 246 **References**

- 247 1. Handmer, J., *et al.* (2012), Changes in impacts of climate extremes: Human systems and ecosystems,  
248 in *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*,

- 249 edited by C. Field *et al.*, pp. 231– 290, A Special Report of Working Groups I and II of the  
250 Intergovernmental Panel on Climate Change: Managing the Risks of Extreme Events and Disasters to  
251 Advance Climate Change Adaptation, Cambridge Univ. Press, Cambridge, U. K., and New York.
- 252 2. Crimmins, A., Balbus, J., Gamble, J. L., Beard, C. B., Bell, J. E., Dodgen, D., Eisen, R. J., Fann, N.,  
253 Hawkins, M. D., Herring, S. C., Jantarasami, L., Mills, D. M., Saha, S., Sarofim, M. C., Trtanj, J., & Ziska,  
254 L. (2016). *The impacts of climate change on human health in the United States: A Scientific*  
255 *Assessment*. Washington, DC: US Global Change Research Program.  
256 <https://doi.org/10.7930/J0R49NQX>
- 257 3. Allen, M. R., & Ingram, W. J. (2002). Constraints on future changes in climate and the hydrologic  
258 cycle. *Nature*, **419**(6903), 224–232. <https://doi.org/10.1038/nature01092>
- 259 4. O'Gorman, P. A., and T. Schneider (2009), The physical basis for increases in precipitation extremes  
260 in simulations of 21st century climate change, *Proceedings of the National Academy of Sciences of*  
261 *the United States of America*, **106**(35), 14,773–14,777, doi:10.1073/pnas.0907610106.
- 262 5. Trenberth, K. E., Dai, A., Rasmussen, R. M., & Parsons, D. B. (2003). The changing character of  
263 precipitation. *Bulletin of the American Meteorological Society*, **84**(9), 1205– 1218.  
264 <https://doi.org/10.1175/BAMS-84-9-1205>
- 265 6. Held, I. M., and B. J. Soden (2006), Robust response of the hydrological cycle to global warming,  
266 *Journal of Climate*, **19**, 5686– 5699.
- 267 7. Norris, J., Chen, G., Neelin, J. D., Norris, J., Chen, G., & Neelin, J. D. (2019). Thermodynamic versus  
268 dynamic controls on extreme precipitation in a warming climate from the Community Earth System  
269 Model Large Ensemble. *Journal of Climate*, **32**, 1025–1045. <https://doi.org/10.1175/JCLI-D-18->  
270 [0302.1](https://doi.org/10.1175/JCLI-D-18-0302.1)
- 271 8. Sun, Y., S. Solomon, A. Dai, and R. W. Portmann (2006), How often does it rain?, *Journal of Climate*,  
272 **19**(6), 916– 934, doi:10.1175/JCLI3672.1

- 273 9. Fischer, E. M., U. Beyerle, and R. Knutti (2013), Robust spatially aggregated projections of climate  
274 extremes, *Nature Climate Change*, **3**(12), 1033– 1038.
- 275 10. Kharin, V. V., F. W. Zwiers, X. Zhang, and M. Wehner (2013), Changes in temperature and  
276 precipitation extremes in the CMIP5 ensemble, *Climatic Change*, **119**(2), 345– 357,  
277 doi:10.1007/s10584-013-0705-8
- 278 11. Sillmann, J., V. V. Kharin, F. W. Zwiers, X. Zhang, and D. Bronaugh (2013), Climate extremes indices in  
279 the CMIP5 multimodel ensemble: Part 2. Future climate projections, *Journal of Geophysical*  
280 *Research Atmospheres*, **118**, 2473– 2493, doi:10.1002/jgrd.50188
- 281 12. Madakumbura, G. D., Kim, H., Utsumi, N., Shiogama, H., Fischer, E. M., Seland, Ø., Scinocca, J. F.,  
282 Mitchell, D. M., Hirabayashi, Y., & Oki, T. (2019). Event-to-event intensification of the hydrologic  
283 cycle from 1.5°C to a 2°C warmer world. *Scientific Reports*, **9**(1), 3483.
- 284 13. Swain, D. L., Langenbrunner, B., Neelin, J. D., & Hall, A. (2018). Increasing precipitation volatility in  
285 twenty-first-century California. *Nature Climate Change*, **8**, 427– 433.  
286 <https://doi.org/10.1038/s41558-018-0140-y>
- 287 14. Allan, R. P., and B. J. Soden (2008). Atmospheric warming and the amplification of precipitation  
288 extremes, *Science*, **321**(5895), 1481– 1484, doi:10.1126/science.1160787
- 289 15. Min, S. K., X. B. Zhang, F. W. Zwiers, and G. C. Hegerl (2011), Human contribution to more-intense  
290 precipitation extremes, *Nature*, **470**, 378– 381, doi:10.1038/nature09763
- 291 16. Donat, M. G., A. L. Lowry, L. V. Alexander, P. A. O’Gorman, and N. Maher (2016), More extreme  
292 precipitation in the world’s dry and wet regions, *Nature Climate Change*, **6**, 508– 513.
- 293 17. Donat, M. G., Angélil, O., & Ukkola, A. M. (2019). Intensification of precipitation extremes in the  
294 world’s humid and water-limited regions. *Environmental Research Letters*, **14**, 065003.  
295 <https://doi.org/10.1088/1748-9326/ab1c8e>
- 296 18. Kirchmeier-Young, M. C., and X. Zhang (2020), Human influence has intensified extreme

- 297 precipitation in North America. *Proceedings of the National Academy of Sciences of the United*  
298 *States of America*, **117**(24), 13,308–13,313, doi:10.1073/PNAS.1921628117.
- 299 19. Stott, P., N. Gillett, G. Hegerl, D. Karoly, D. Stone, X. Zhang, and F. Zwiers (2010), Detection and  
300 attribution of climate change: A regional perspective, *Wiley Interdisciplinary Reviews: Climate*  
301 *Change*, **1**, 192– 211, doi:10.1002/wcc.34
- 302 20. Bindoff, N. L., *et al.* (2013), Detection and attribution of climate change: From global to regional, in  
303 *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth*  
304 *Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. F. Stocker *et al.*,  
305 pp. 867– 952, Cambridge Univ. Press, Cambridge, U. K., and New York.
- 306 21. Marvel, K., Cook, B. I., Bonfils, C. J., Durack, P. J., Smerdon, J. E., & Williams, A. P. (2019). Twentieth-  
307 century hydroclimate changes consistent with human influence. *Nature*, **569**(7754), 59– 65.
- 308 22. Hegerl, G., and F. Zwiers (2011), Use of models in detection and attribution of climate change, *Wiley*  
309 *Interdisciplinary Reviews: Climate Change*, **2**(4), 570– 591.
- 310 23. Easterling, D. R., Kunkel, K. E., Wehner, M. F., & Sun, L. (2016). Detection and attribution of climate  
311 extremes in the observed record. *Weather and Climate Extremes*, **11**, 17– 27.
- 312 24. Herold, N., Behrangi, A., & Alexander, L. V. (2017). Large uncertainties in observed daily precipitation  
313 extremes over land. *Journal of Geophysical Research: Atmospheres*, **22**, 668– 681.  
314 <https://doi.org/10.1002/2016JD025842>
- 315 25. Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., & Hsu, K. L. (2018). A review of global  
316 precipitation data sets: Data sources, estimation, and intercomparisons. *Reviews of Geophysics*, **56**,  
317 79– 107. <https://doi.org/10.1002/2017RG000574>
- 318 26. Roca, R., Alexander, L. V., Potter, G., Bador, M., Jucá, R., Contractor, S., Bosilovich, M. G., & Cloché,  
319 S. (2019). FROGS: A daily 1° × 1° gridded precipitation database of rain gauge, satellite and reanalysis  
320 products. *Earth System Science Data*, **11**(3), 1017–1035. <http://doi.org/10.5194/essd-11-1017-2019>

- 321 27. O'Gorman, P. A. (2012), Sensitivity of tropical precipitation extremes to climate change, *Nature*  
322 *Geoscience*, **5**(10), 697– 700, doi:10.1038/NGEO1568.
- 323 28. Ribes, A., Zwiers, F. W., Azais, J.-M., & Naveau, P. (2017). A new statistical approach to climate  
324 change detection and attribution. *Climate Dynamics*, **48**(1-2), 367– 386.
- 325 29. Barnes, E. A., Hurrell, J. W., Ebert-Uphoff, I., Anderson, C., & Anderson, D. (2019). Viewing forced  
326 climate patterns through an AI Lens. *Geophysical Research Letters*, **46**, 13,389– 13,398.  
327 <https://doi.org/10.1029/2019GL084944>
- 328 30. Barnes, E. A., Toms, B., Hurrell, J.W., Ebert-Uphoff, I., Anderson, C., & Anderson, D. (2020). Indicator  
329 patterns of forced change learned by an artificial neural network. *Journal of Advances in Modeling*  
330 *Earth Systems*, **12**, e2020MS002195. <https://doi.org/10.1029/2020MS002195>
- 331 31. Sippel, S., Meinshausen, N., Fischer, E. M., Szekely, E., & Knutti, R. (2020). Climate change now  
332 detectable from any single day of weather at global scale. *Nature Climate Change*, **10**, 35– 41.  
333 <https://doi.org/10.1038/s41558-019-0666-7>
- 334 32. Ebert-Uphoff, I., Samarasinghe, S., & Barnes, E. (2019). Thoughtfully using artificial intelligence in  
335 Earth science. *Eos*, **100**. <https://doi.org/10.1029/2019eo135235>
- 336 33. Toms, B. A., Barnes, E. A., & Ebert-Uphoff, I. (2020). Physically interpretable neural networks for the  
337 geosciences: applications to earth system variability. *Journal of Advances in Modeling Earth Systems*,  
338 **12**, e2019MS002002. <https://doi.org/10.1029/2019MS002002>
- 339 34. Wills, R.C., Sippel, S. and Barnes, E.A. (2020). Separating forced and unforced components of climate  
340 change: The utility of pattern recognition methods in Large Ensembles and observations. *Variations*,  
341 **18**(2), pp.1-10.
- 342 35. Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., & Samek, W. (2015). On pixel-wise  
343 explanations for non-linear classifier decisions by layer-wise relevance propagation. *Plos One*, **10**(7),  
344 e0130140. <https://doi.org/10.1371/journal.pone.0130140>

- 345 36. Montavon, G., Samek, W., & Müller, K. R. (2018). Methods for interpreting and understanding deep  
346 neural networks. *Digital Signal Processing*, **73**, 1– 15.
- 347 37. Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design.  
348 *Bulletin of the American Meteorological Society*, **93**(4), 485– 498. [https://doi.org/10.1175/BAMS-D-](https://doi.org/10.1175/BAMS-D-11-00094.1)  
349 11-00094.1
- 350 38. Eyring, V., S. Bony, G. A. Meehl, C. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor (2015), Overview  
351 of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and  
352 organisation, *Geoscientific Model Development Discussion*, **8**(12), 10,539– 10,583,  
353 doi:10.5194/gmdd-8-10539-2015
- 354 39. Mora, C., Frazier, A. G., Longman, R. J., Dacks, R. S., Walton, M. M., Tong, E. J., Sanchez, J. J., Kaiser,  
355 L. R., Stender, Y. O., Anderson, J. M., Ambrosino, C. M., Fernandez-Silva, I., Giuseffi, L. M., &  
356 Giambelluca, T. W. (2013). The projected timing of climate departure from recent variability. *Nature*,  
357 **502**(7470), 183–187. <https://doi.org/10.1038/nature12540>
- 358 40. King, A. D., Donat, M. G., Fischer, E. M., Hawkins, E., Alexander, L. V., Karoly, D. J., *et al.* (2015). The  
359 timing of anthropogenic emergence in simulated climate extremes. *Environmental Research Letters*,  
360 **10**(9), 94015. <https://doi.org/10.1088/1748-9326/10/9/094015>
- 361 41. Pfahl, S., O’Gorman, P. A., & Fischer, E. M. (2017). Understanding the regional pattern of projected  
362 future changes in extreme precipitation. *Nature Climate Change*, **7**(6), 423– 427.  
363 <https://doi.org/10.1038/nclimate3287>
- 364 42. Shepherd, T. G. (2014), Atmospheric circulation as a source of uncertainty in climate change  
365 projections, *Nature Geoscience*, **7**, 703– 708, doi:10.1038/ngeo2253
- 366 43. Hamed, K. H., and A. R. Rao (1998), A modified Mann-Kendall trend test for autocorrelated data,  
367 *Journal of Hydrology*, **204**, 182– 196.
- 368 44. Wills, R. C. J., Battisti, D. S., Armour, K. C., Schneider, T., & Deser, C. (2020). Pattern Recognition

- 369 Methods to Separate Forced Responses from Internal Variability in Climate Model Ensembles and  
370 Observations, *Journal of Climate*, 33(20), 8693-8719. <https://doi.org/10.1175/JCLI-D-19-0855.1>
- 371 45. Bador, M., Alexander, L. V., Contractor, S., & Roca, R. (2020). Diverse estimates of annual maxima  
372 daily precipitation in 22 state-of-the-art quasi-global land observation datasets. *Environmental*  
373 *Research Letters*, 15(3), 35005. <http://doi.org/10.1088/1748-9326/ab6a22>
- 374 46. Alexander, L. V., Bador, M., Roca, R., Contractor, S., Donat, M., & Nguyen, P. (2020).  
375 Intercomparison of annual precipitation indices and extremes over global land areas from in situ,  
376 space-based and reanalysis products. *Environmental Research Letters*, 15(5), 055002.  
377 <https://doi.org/10.1088/1748-9326/ab79e2>
- 378 47. Theil, H. (1950), A rank-invariant method of linear and polynomial regression analysis, *Indagationes*  
379 *Mathematicae*, 12, 85–91.
- 380 48. Sen, P. K. (1968), Estimates of the regression coefficient based on Kendall's tau. *Journal of the*  
381 *American statistical association*, 63, 1379–1389.
- 382 49. Borodina, A., Fischer, E. M., & Knutti, R. (2017). Models are likely to underestimate increase in heavy  
383 rainfall in the extratropical regions with high rainfall intensity. *Geophysical Research Letters*, 44,  
384 7401– 7409. <https://doi.org/10.1002/2017GL074530>
- 385 50. Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., & Müller, K. R. (Eds.). (2019). *Explainable AI:*  
386 *Interpreting, explaining and visualizing deep learning*. Springer Nature.

387

## 388 **Data and Methods**

### 389 **Data**

390 We use daily precipitation rate output from a collection of climate models participating in  
391 CMIP5 and CMIP6 (Supplementary Table 1). Data from each ensemble's historical forcing scenario is  
392 combined with future projections following a high-emissions scenario to create a time-series from 1920

393 to 2099 for each model. Future projections from CMIP5 follow the Representative Concentration  
394 Pathway 8.5 (RCP 8.5)<sup>51</sup>, while CMIP6 projections follow the Shared Socioeconomic Pathway 5–8.5 (SSP  
395 5-8.5)<sup>52</sup>. To increase our sample size, we combine both CMIP5 and CMIP6 model subsets into one  
396 ensemble, which is justifiable considering the very similar time evolution of the total anthropogenic  
397 forcing in RCP 8.5 and SSP 5-8.5 scenarios (ref. 52, their Figure 3c). We regrid all daily precipitation data  
398 to a 2° x 2° spatial grid and compute the Rx1day value for each year at each land grid point.

399 We use four datasets of observational estimates of daily precipitation rate with global  
400 coverage: Multi-Source Weighted-Ensemble Precipitation, version 2 (MSWEP)<sup>53</sup>, Global Precipitation  
401 Climatology Centre (GPCC) version 2018<sup>54</sup>, and Rainfall Estimates on a Gridded Network (REGEN)<sup>55</sup>,  
402 including both REGEN\_ALL and REGEN\_LONG . MSWEP is a hybrid reconstruction using *in situ*, satellite  
403 and reanalysis data, whereas GPCC and the REGEN datasets are developed from ground-based  
404 measurements. REGEN\_ALL is developed by interpolating all considered station data whereas  
405 REGEN\_LONG is developed using only the stations with a data record of 40 years or longer. We further  
406 use seven widely used reanalysis products for comparison : ECMWF ERA5<sup>56</sup>, Japanese 55-year Reanalysis  
407 (JRA55)<sup>57</sup>, Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA2)<sup>58</sup>,  
408 NCEP Climate Forecast System Reanalysis (CFSR)<sup>59</sup> , Global Soil Wetness Project Phase 3 (GSWP3)<sup>60</sup>,  
409 NCEP-DOE Reanalysis 2 (NCEP2)<sup>61</sup> and NOAA-CIRES-DOE Twentieth Century Reanalysis version 3  
410 (20CRv3)<sup>62</sup>. These observational and reanalysis datasets are selected considering the availability of full  
411 global land coverage and data for at least three decades (Table S2). All observation and reanalysis data  
412 were regridded to the same 2° x 2° spatial grid as the models, and then Rx1day was calculated at each  
413 grid point for each year.

414

415 **Neural network based detection method**

416 Here we apply the method in ref. 29 (see their Figure 1a) to predict the year with which given  
417 annual Rx1day maps from GCMs are associated, a regression task. This requires the ANN to learn the  
418 signature of the forced response in simulated Rx1day. By feeding the ANN multicentury data from  
419 forced simulations, it learns to distinguish the forced signal from internal climate variability. The use of  
420 multiple GCMs helps the ANN learn the common elements of the forced response most relevant to the  
421 prediction task, a process that fully considers model uncertainty as well as internal climate variability.  
422 Input to the ANN from each model is a vectorized spatial map of Rx1day ( $2^\circ \times 2^\circ$  spatial grid = 16200 grid  
423 values) for each year from 1920 to 2099. Our primary goal is to detect the anthropogenic signal in  
424 extreme precipitation over land. Thus, we mask out data over the ocean at this stage, resulting in 6082  
425 land grid values. The ANN architecture consists of two hidden layers with ten nodes each. The *Rectified*  
426 *Linear Unit* activation function is used for all hidden units.

427 Approximately 80% of the models (35) are used for training the ANN, while the rest (9) are used  
428 for testing. K-fold validation is applied to split the initial training dataset into training and validation  
429 datasets. Here we set K=2. The *mean squared error* between the actual and predicted year of Rx1day is  
430 used as the loss function to be minimized during the training. For the optimizer which updates the ANN  
431 based on the gradient of the loss, we select *rmsprop*. Climate variables inherently contain spatial  
432 autocorrelation. To account for this dependence among adjacent input data points, we use L2  
433 regularization between inputs and the first hidden layer, which adds the sum of squared weights as a  
434 penalty term to the loss function. By iterating over L2 values of leading order of magnitudes and  
435 inspecting the tradeoff between low prediction error and generalizability (Figure S1), we found L2=0.001  
436 to be a suitable value for our analysis.

437 We could have made the ANN more complex to achieve higher accuracy. But we elected not to  
438 do so, partly because increasing the number of hidden units or changing the other hyperparameters  
439 (except for L2 regularization) did not result in a substantial increase in accuracy. More importantly, we

440 aimed to keep the ANN simple, with a reasonable degree of accuracy. This is because the main goal is  
 441 not to obtain a perfect prediction, but rather to reveal the forced patterns the ANN learns (e.g. ref. 29,  
 442 30). As we show in Section 3, imperfections in the prediction also can be physically interpreted within  
 443 the D&A research framework (Figure 2).

444

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

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

455

$$456 \quad \sum_{i=1}^d P_i = \dots = \sum_j P_j = \sum_k P_k = \dots = f(\mathbf{x}) \quad (1)$$

457

$$458 \quad \mathbf{a}_k = \text{ReLU}(\sum_j \mathbf{a}_j w_{jk} + \mathbf{b}_k) \quad (2)$$

459

$$460 \quad P_j = \sum_k \left( \alpha \frac{a_j w_{jk}^+}{\sum_j a_j w_{jk}^+} - \beta \frac{a_j w_{jk}^-}{\sum_j a_j w_{jk}^-} \right) \quad (3)$$

461

462 The relevance propagation rule from layers  $j$  to  $k$  is given in eq. 3. This general form is also  
463 known as the  $\alpha\beta$ -rule<sup>35,36</sup>. The components  $()^+$  and  $()^-$  indicate only positive and negative weights are  
464 being considered, respectively. The  $\alpha$  and  $\beta$  coefficients represent the relative amount of positive and  
465 negative relevance to be propagated, respectively. As shown in eq. 3, positive relevance (i.e. excitatory  
466 influence) and negative relevance (i.e. inhibitory influence) are associated with positive and negative  
467 weights, respectively. The  $\alpha$  and  $\beta$  coefficients are to be chosen with the constraints  $\alpha - \beta = 1$  and  $\beta \geq$   
468 0. The combination  $\alpha = 2$  and  $\beta = 1$  (LRP <sub>$\alpha 2\beta 1$</sub> ) have been experimentally inferred as suitable values,  
469 and have been adopted in previous research<sup>35,36,63-65</sup>. Here we adopt the LRP <sub>$\alpha 2\beta 1$</sub>  rule.

470 The  $\alpha\beta$ -rule rule with  $\alpha = 1$  and  $\beta = 0$  (LRP <sub>$\alpha 1\beta 0$</sub> , also known as the *Deep Taylor Decomposition*)  
471 is a special case where for each input, the sum of LRP relevance heatmaps is equivalent to  $f(x)$ . This is a  
472 valuable property for the interpretation of results. As  $\beta = 0$ , LRP <sub>$\alpha 1\beta 0$</sub>  only considers the information  
473 which positively contributes to the final decision. For regression tasks such as the problem at hand here,  
474 inputs which contribute to a decrease in  $f(x)$  (i.e. an earlier predicted year; negative relevance) are  
475 equally as important as inputs which contribute to an increase (i.e. a later predicted year; positive  
476 relevance) to understand what the ANN has learned. Moreover, when  $\alpha > 1$ , the  $\alpha\beta$ -rule does not  
477 conserve the relevance from the output value back to the input layer. For these reasons, ref. 33 pointed  
478 out that caution should be exercised when applying the  $\alpha\beta$ -rule with 1)  $\alpha = 1$  for regression and 2)  $\alpha >$   
479 1 in general. We find that for our simple ANN, applying LRP <sub>$\alpha 2\beta 1$</sub>  results in a 1:1 relationship between the  
480 resultant relevance heatmaps and  $f(x)$  for each input (Figure S1d). This allows the visualization of input  
481 that contributes to a decrease in  $f(x)$  while maintaining a direct relationship between the ANN  
482 predicted value and LRP heatmaps. Therefore, we proceed with rescaled relevance heatmaps derived  
483 from LRP <sub>$\alpha 2\beta 1$</sub>  for interpreting our ANN. More details on LRP can be found in previous work (ref.  
484 33,35,36,66).

485

486 **References**

- 487 51. Meinshausen, M., *et al.* (2011), The RCP greenhouse gas concentrations and their extensions from  
488 1765 to 2300, *Climatic Change*, **109**(1–2), 213– 241, doi:10.1007/s10584-011-0156-z
- 489 52. O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler,  
490 E., Lamarque, J. F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., & Sanderson, B. M. (2016). The  
491 Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model*  
492 *Development*, **9**(9), 3461– 3482. <https://doi.org/10.5194/gmd-9-3461-2016>
- 493 53. Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., van Dijk, A. I., & Adler, R. F. (2018).  
494 MSWEP V2 global 3-hourly 0.1° precipitation: Methodology and quantitative assessment. *Bulletin of*  
495 *the American Meteorological Society*, **100**, 473– 500. <https://doi.org/10.1175/BAMS-D-17-0138.1>
- 496 54. Ziese, M, Rauthe-Schöch, A, Becker, A, Finger, P, Meyer-Christoffer, A, & Schneider, U. (2018). GPCC  
497 full data daily version. 2018 at 1.0°: Daily land-surface precipitation from rain-gauges built on GTS-  
498 based and historic data: Global Precipitation Climatology Centre (GPCC, <http://gpcc.dwd.de/>) at  
499 Deutscher Wetterdienst.
- 500 55. Contractor, S., Donat, M. G., Alexander, L. V., Ziese, M., Meyer-Christoffer, A., Schneider, U.,  
501 Rustemeier, E., Becker, A., Durre, I., & Vose, R. S. (2020). Rainfall Estimates on a Gridded Network  
502 (REGEN)—A global land-based gridded dataset of daily precipitation from 1950–2013. *Hydrology*  
503 *and Earth System Sciences*, **24**(2), 919– 943. <https://doi.org/10.5194/hess-24-919-2020>
- 504 56. Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horanyi, A., Muñoz-Sabater, J., Nicolas, J., Peubey,  
505 C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P.,  
506 Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R.,  
507 Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Holm, E.,  
508 Janiskova, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I.,  
509 Vamborg, F., Villaume, S., & Thepaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of*

510 *the Royal Meteorological Society*, **2020**, 1– 51. <https://doi.org/10.1002/qj.3803>

511 57. Kobayashi, S., *et al.* (2015), The JRA-55 reanalysis: General specifications and basic characteristics,  
512 *Journal of the Meteorological Society of Japan*, **93**(1), 5– 48, doi:10.2151/jmsj.2015-001

513 58. Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., ... Zhao, B. (2017). The  
514 Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of*  
515 *Climate*, **30**(14), 5419– 5454. <https://doi.org/10.1175/JCLI-D-16-0758.1>

516 59. Saha, S., S. Moorthi, H.-L. Pan, and 49 others (2010), The NCEP climate forecast system reanalysis,  
517 *Bulletin of the American Meteorological Society*, **91**(8), 1015– 1057, doi:10.1175/2010BAMS3001.1

518 60. Kim, H. (2017), Global Soil Wetness Project Phase 3 Atmospheric Boundary Conditions (Experiment  
519 1), *Data Integration and Analysis System (DIAS)*. <https://doi.org/10.20783/DIAS.501>

520 61. Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. Hnilo, M. Fiorino, and G. L. Potter (2002),  
521 NCEP-DOE AMIP-II reanalysis (R-2), *Bulletin of the American Meteorological Society*, **83**, 1631– 1643.

522 62. Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., Allan, R., Yin,  
523 X., Vose, R., & Titchner, H. (2019). Towards a more reliable historical reanalysis: Improvements for  
524 version 3 of the Twentieth Century Reanalysis system. *Quarterly Journal of the Royal Meteorological*  
525 *Society*, **145**(724), 2876– 2908.

526 63. Böhle, M., Eitel, F., Weygandt, M., & Ritter, K. (2019). Layer-wise relevance propagation for  
527 explaining deep neural network decisions in MRI-based Alzheimer's disease classification. *Frontiers*  
528 *in aging neuroscience*, **11**, 194. <https://doi.org/10.3389/fnagi.2019.00194>

529 64. Dobrescu, A., Valerio Giuffrida, M., & Tsafaris, S. A. (2019). Understanding deep neural networks for  
530 regression in leaf counting, *Proceedings of the IEEE Conference on Computer Vision and Pattern*  
531 *Recognition Workshops* (pp. 2600– 2608). Long Beach, CA, USA: IEEE.

532 65. Grigorescu, I., Cordero-Grande, L., Edwards, A. D., Hajnal, J. V., Modat, M., & Deprez, M. (2019).  
533 Investigating Image Registration Impact on Preterm Birth Classification: An Interpretable Deep

534 Learning Approach. In *Smart Ultrasound Imaging and Perinatal, Preterm and Paediatric Image*  
535 *Analysis*, pp. 104-112, Springer, Cham.

536 66. Montavon, G., Binder, A., Lapuschkin, S., Samek, W., & Müller, K. R. (2019). Layer-wise relevance  
537 propagation: an overview. In *Explainable AI: interpreting, explaining and visualizing deep learning*  
538 (pp. 193-209). Springer, Cham. doi: 10.1007/978-3-030-28954-6\_10.

539

540

#### 541 **Acknowledgements**

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

548 **Competing financial and non-financial interests:** The authors declare no competing interests. **Data and**  
549 **materials availability:** Data and materials that support the findings of this study are available upon  
550 reasonable request from the authors.