

1 **Changes in United States summer temperatures**
2 **revealed by explainable neural networks**

3 **Zachary M. Labe¹, Nathaniel C. Johnson², and Thomas L. Delworth²**

4 ¹Atmospheric and Oceanic Sciences Program, Princeton University, NJ, USA

5 ²NOAA/OAR/Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA

6 **Key Points:**

- 7 • Forced temperature signals have emerged in observations during summer in the
8 United States as detected by an artificial neural network
- 9 • Increasing spatial resolution improves neural network skill for predicting the year
10 of a given summer temperature map
- 11 • Trends in western United States land surface fields are linked to an early timing
12 of emergence in a state-of-the-art climate model

Abstract

To better understand the regional changes in summertime temperatures across the conterminous United States (CONUS), we adopt a recently developed machine learning framework that can be used to reveal the timing of emergence of forced climate signals from the noise of internal climate variability. Specifically, we train an artificial neural network (ANN) on seasonally-averaged temperatures across the CONUS and then task the ANN to output the year associated with an individual map. In order to correctly identify the year, the ANN must therefore learn time-evolving patterns of climate change amidst the noise of internal climate variability. The ANNs are first trained and tested on data from large ensembles and then evaluated using observations from a station-based dataset. To understand how the ANN is making its predictions, we leverage a collection of ad hoc feature attribution methods from explainable artificial intelligence (XAI). We find that anthropogenic signals in seasonal mean minimum temperature have emerged by the early 2000s for the CONUS, which occurred earliest in the Eastern United States. While our observational timing of emergence estimates are not as sensitive to the spatial resolution of the training data, we find a notable improvement in ANN skill using a higher resolution climate model, especially for its early 20th century predictions. Composites of XAI maps reveal that this improvement is linked to temperatures around higher topography. We find that increases in spatial resolution of the ANN training data may yield benefits for machine learning applications in climate science.

Plain Language Summary

While temperatures around the world continue to warm due to human-caused climate change, some areas have observed smaller temperature trends than others. Understanding this regional variability in the rate of warming is important when assessing future projections. One location that has observed less warming is across the United States during their summer season. To evaluate temperature variability in this region using real-world observations and climate model simulations, we use a statistical method from artificial intelligence called neural networks. The goal of the neural network setup is to learn temperature patterns across the United States and then identify whether climate change effects have exceeded the range of natural variability that has occurred in the past. This is called the timing of emergence, which is the first year that the effect has clearly appeared. We find that the average United States minimum temperature increase has al-

ready emerged in historical records. However, we find no timing of emergence for the average maximum temperature, other than in the Western United States. Another important finding of this study is that by using higher resolution climate model data (i.e., more latitude and longitude points), we find better accuracy in the neural network predictions.

1 Introduction

The detection of a forced signal rising above the background of internal climate variability - often referred to as the ‘timing of emergence’ (ToE) - can be a potentially useful metric for societal and ecological planning to account for changes in weather and climate that exceed the known variability over a particular region (IPCC et al., 2021). Quantifying the ToE for a number of variables has been widely addressed within climate science research over the last few decades (e.g., Giorgi & Bi, 2009; Hawkins & Sutton, 2012; Mahlstein et al., 2012; Hawkins et al., 2014; Fischer & Knutti, 2014; King et al., 2015; Mora et al., 2013; Schlunegger et al., 2020; Hawkins et al., 2020; Satoh et al., 2022), but its precise definition is still quite sensitive to the choice of dataset, future greenhouse emission scenario, baseline reference period, spatial scale, temporal filtering, consideration of internal climate variability, and statistical testing for ToE consistency. The ToE can even be examined through a lens of compound events and combined variables (Mahony & Cannon, 2018; Rader et al., 2022; François & Vrac, 2023). While it is often implied to be associated with anthropogenic climate change, the ToE definition can also be influenced by multidecadal variability in the climate system (Lehner et al., 2017). To partially alleviate this issue, recent ToE work (e.g., Rodgers et al., 2015; Lehner et al., 2017; Schlunegger et al., 2020; Wyser et al., 2021) has examined the use of initial condition large ensembles which better isolate the role of uncertainty due to internal variability within a single global climate model (GCM) system. This is done to avoid conflating different uncertainties, such as when calculating a multi-model mean (Hawkins et al., 2014; Lehner et al., 2020). In this context, the spread of ToE estimates can be compared across individual ensemble members (i.e., range of internal variability) relative to the overall ensemble mean (i.e., radiatively forced signal).

One example of an extratropical region that experiences large variances in simulated decadal temperature trends by these climate model large ensembles is across the contiguous United States (Deser et al., 2012; Milinski et al., 2020; Lehner & Deser, 2023). While this variability is often analyzed for boreal winter months (McKinnon & Deser,

2018; Deser et al., 2016; B. Yu et al., 2020), substantial regional variability in CONUS summertime temperature trends has also been observed (Program, 2018). Of particular interest is a portion of the Southeast to Central CONUS in which little to no warming has been observed (Kunkel et al., 2006), or even has slightly cooled depending on the choice of trend period, duration, or observational dataset (Grotjahn & Huynh, 2018; Partridge et al., 2018). This ‘warming hole’ has been most prominently found in trends of daytime maximum temperatures (Pan et al., 2004; Mueller et al., 2016). It remains unclear to what extent this persistent CONUS warming hole is driven by an unusual realization of internal climate variability (a. Meehl et al., 2012; Meehl et al., 2015), land-use surface feedbacks (Mueller et al., 2016; Alter et al., 2018), changes in anthropogenic aerosols (Leibensperger et al., 2012; S. Yu et al., 2014; Banerjee et al., 2017), an indirect response to external greenhouse gas forcing (Eischeid et al., submitted), or a combination of several of these factors (Pan et al., 2004; Mascioli et al., 2017). In contrast, the Western United States has observed larger warming trends during the last few decades, which has contributed to the formation of prolonged drought risk and favorable environments for wildfire ignition (Diffenbaugh et al., 2015; Abatzoglou & Williams, 2016; Parks & Abatzoglou, 2020; Williams et al., 2020). Though the ratio of new record high temperatures compared to record lows continues to widen (Meehl et al., 2022), the overall detectability of CONUS temperature signals continues to remain challenging, partially due to the anomalous warmth observed in the Dust Bowl era (Hansen et al., 2001; Peterson et al., 2013; Donat et al., 2016). Given the broad range of consequences associated with future projected warming over the CONUS (Wuebbles et al., 2014; Program, 2018), it remains urgent to better characterize the ToE of summertime mean extreme temperatures in order to better aid in future decision-making on regional health hazards and other impacts that could fall outside of historical climate variability (Mankin et al., 2020; Deser, 2020; Schwarzwald & Lenssen, 2022; Bevacqua et al., 2023).

The aim of this study is to evaluate whether patterns of radiatively-forced temperature change have emerged across the CONUS in summer, despite substantial internal variability and only a marginally positive overall mean temperature trend in the recent observational record. To confront these challenges, we turn to a novel explainable artificial neural network (ANN) approach (Barnes et al., 2018, 2019) that can spatially leverage temperature signals across given geographic maps for identifying the ToE of individual climate variables. The advantage is that we are not limited to traditional signal-

110 to-noise metrics that only consider point-by-point statistics for disentangling the role of
111 internal variability from anthropogenic climate change, because here the ANNs can in-
112 stead potentially learn nonlinear relationships in the data. We also apply this framework
113 to different regions of the CONUS, which supports that the Central United States warm-
114 ing hole has contributed to a lack of visible anthropogenic signal in mean observed sum-
115 mertime temperatures. Although this machine learning ToE indicator may not neces-
116 sarily reflect local conditions, such as those considered for adaptation planning for ecosys-
117 tem services (Weiskopf et al., 2020) or for rapid changes to biodiversity (Henson et al.,
118 2017; Trisos et al., 2020), the spatial aggregate metric as designed here remains relevant
119 for industries that assess climate risk over larger geographic domains (e.g., Lempert et
120 al., 2004; Mills, 2005; Fischer et al., 2013; Lawrence et al., 2020; Ignjacevic et al., 2021).

121 Notably, we also reveal that the ANN can distinguish the year of temperature maps
122 during a climate model’s historical simulation of the 20th century. This is surprising given
123 the greater influence of internal variability during this period that continues until around
124 1980 when the forced greenhouse gas signal begins to more clearly emerge. These skill-
125 ful predictions by the ANN indicate that changing temperature patterns within a cli-
126 mate model may be distinguished from the noise of internal variability well before a sta-
127 tistically significant mean temperature trend is detected. This early 20th century detec-
128 tion skill improves even more when training on maps of higher spatial resolution com-
129 pared to a lower resolution configuration of the same large ensemble climate model. Us-
130 ing methods from explainable artificial intelligence (XAI), we conclude this study by ex-
131 amining the regional patterns of temperature change linked to this ANN performance
132 and its possible dependence on the size of the training data climate maps.

133 **2 Data**

134 **2.1 Climate Model Large Ensembles**

135 For the main results of this study, we use a collection of 30-member initial condi-
136 tion large ensemble simulations from a fully-coupled global climate model (GCM) called
137 the Seamless System for Prediction and EArth System Research (SPEAR; Delworth et
138 al., 2020). SPEAR is the newest seasonal to multidecadal prediction and projection sys-
139 tem from the National Oceanic and Atmospheric Administration (NOAA) Geophysical
140 Fluid Dynamics Laboratory (GFDL). SPEAR uses the same atmospheric model code

141 (AM4) and land model code (LM4) as in the GFDL CM4 model (Zhao et al., 2018a, 2018b;
 142 Held et al., 2019), but employs a different configuration of the MOM6 ocean code (Adcroft
 143 et al., 2019) so as to optimize the model for seasonal to decadal predictions and projec-
 144 tions. SPEAR incorporates 33 vertical atmospheric levels and can be designed for dif-
 145 ferent atmosphere/land resolution configurations ranging from 0.25° to 1.0° . All model
 146 versions include a common ocean grid of approximately 1.0° spacing (though refined to
 147 0.33° around the equator). The SPEAR system has already been successfully used for
 148 several studies in evaluating the predictability of temperature variability and heat ex-
 149 tremes across North America (e.g., Jia et al., 2022; Yang et al., 2022).

Table 1. Summary of the GFDL SPEAR_MED large ensemble simulations used in this study (Delworth et al., 2020). Additional details on SPEAR can be found at https://www.gfdl.noaa.gov/spear_large_ensembles/.

Name	Scenario Forcing	Years	# Members	Horizontal Resolution (Atmosphere / Ocean)
SPEAR_MED_SSP119	Historical to 2014, SSP1-1.9	1921-2100	30	nominal 0.5° /nominal 1.0°
SPEAR_MED_SSP245	Historical to 2014, SSP2-4.5	—	—	—
SPEAR_MED_SSP585	Historical to 2014, SSP5-8.5	—	—	—
SPEAR_MED_NATURAL	Natural Forcing Only	—	—	—

150 Due to the availability of more individual ensemble members to train, validate, and
 151 test our neural network than in the configuration with the highest atmospheric resolu-
 152 tion, we focus on simulations from only the SPEAR_MED (atmosphere/land of 0.5°) and
 153 SPEAR_LO (atmosphere/land of 1.0°) configurations. Ensemble members in each are
 154 initialized from conditions in an 1850 control simulation that are branched 20 years apart.
 155 Both SPEAR_MED and SPEAR_LO are forced with CMIP6 historical forcing through
 156 2014 (Eyring et al., 2016). From 2015 to 2100, they are then forced with future projec-
 157 tions from the Shared Socioeconomic Pathway 5-8.5 scenario (SSP5-8.5; Kriegler et al.,
 158 2017; Riahi et al., 2017). Recent work has shown that SSP5-8.5 is likely an unrealisti-
 159 cal extreme future emission scenario (e.g., Burgess et al., 2020; Hausfather & Peters, 2020;
 160 Peters & Hausfather, 2020). Although our study here is mostly focused on the ToE in
 161 the recent and historical past, we also compare the sensitivity of our machine learning
 162 results to training and testing on SPEAR_MED simulations conducted with more prob-

163 able SSP scenarios (Pielke et al., 2022), including the low-end SSP1-1.9 and more mod-
164 erate SSP2-4.5 (O'Neill et al., 2016).

165 Finally, we compare our SPEAR climate change simulations with a large ensem-
166 ble experiment starting from the same initial conditions in 1921 as SPEAR_MED, but
167 holding all anthropogenic forcings (i.e., greenhouse gases, anthropogenic aerosols, land
168 use/land change) fixed at 1921 levels. This experiment, referred to as SPEAR_MED_NATURAL
169 is instead prescribed with only natural radiative forcings, such as those due to solar ir-
170 radiance and volcanoes (historical to 2014, hypothetical thereafter; see Delworth et al.
171 (2022)). By comparing SPEAR_MED and SPEAR_MED_NATURAL, we can extract the
172 role of anthropogenic forcing on changes in summertime temperatures in the climate model
173 and gain insights for understanding how well the neural network performs by training
174 on data without a long-term anthropogenic signal.

175 To summarize, we consider collections of 30-member large ensembles from either
176 SPEAR_MED or SPEAR_LO for designing our machine learning architecture. These sim-
177 ulations are conducted on two different horizontal atmosphere/land grids, which we will
178 now refer to as either MED (nominal 0.5° grid) or LOW (nominal 1.0° grid) through-
179 out the text. We point this out since a key focus of this work is on comparing the effect
180 of higher spatial resolution on the performance of the neural network. Given the limited
181 availability of other fully-coupled large ensembles with high-resolution atmospheric mod-
182 els, we can only assess the MED grid using simulations conducted by SPEAR_MED or
183 its previous generation version called the Forecast-Oriented Low Ocean Resolution (FLOR)
184 system (Vecchi et al., 2014). FLOR is a fully-coupled global climate model based upon
185 GFDL's CM2.5 (Delworth et al., 2012); its large ensemble includes 30 members with CMIP5
186 historical forcing from 1921 to 2004 and Representative Concentration Pathway 8.5 (RCP8.5;
187 Riahi et al., 2011; Vuuren et al., 2011) thereafter from 2005 to 2100. FLOR has a land-
188 atmosphere resolution of 0.5° using the AM2.5 and LM3 model components (Milly et
189 al., 2014), but includes a coarser ocean from OM2.1 (Gnanadesikan et al., 2006) at a nom-
190 inal resolution of 1.0° . Since FLOR does not offer a corresponding LOW version of the
191 large ensemble like SPEAR, we simply bilinearly interpolate its corresponding temper-
192 ature maps to the LOW grid (denoted as FLOR (LO)) for again attempting to compare
193 the advantage of inputting more (or less) spatial information into our neural network
194 framework.

195 We also briefly make use of three more GCM large ensembles, which offer a sim-
196 ilar horizontal resolution and data availability as SPEAR_LO (i.e., data from at least 1921
197 to 2100 and a grid size of approximately $1.0^\circ \times 1.0^\circ$). These include the 40-member Com-
198 munity Earth System Model Large Ensemble Project Version 1 (CESM1-LE; Hurrell et
199 al., 2013; Kay et al., 2015) (CMIP5 class; RCP8.5), 100-member CESM2-LE (Danabasoglu
200 et al., 2020; Rodgers et al., 2021) (CMIP6 class; SSP3-7.0), and the 50-member large en-
201 semble using the sixth version of the Model for Interdisciplinary Research on Climate
202 (MIROC6-LE; Tatebe et al., 2019; Shiogama et al., 2023) (CMIP6 class; SSP5-8.5). While
203 other climate model large ensembles are available from the multi-model large ensemble
204 archive that contain at least 30 members (NCAR, 2020; Deser et al., 2020), their hor-
205 izontal resolution is generally too coarse for our regional deep learning approach, par-
206 ticularly when considering the three smaller geographic areas of the United States. In
207 other words, an insufficient number of grid points nearly reduces the problem to a change-
208 point time series task. This then limits the real utility of the neural network method-
209 ology, which here is to exploit any possible (non)linear temperature patterns across the
210 maps in order to identify the emergence of forced climate signals.

211 For the climate model data, we leverage monthly mean near-surface daily maxi-
212 mum, minimum, and average temperature data (i.e., TMAX, TMIN, TAVG) and then
213 calculate the seasonal mean over June to August (JJA) using only grid points across the
214 conterminous United States. A summary of the large ensemble data can be found in Ta-
215 bles 1 and S1, which include the final horizontal resolution elected to be used as input
216 to the neural network.

217 **2.2 Observations**

218 To evaluate the ToE of summertime surface temperatures in the United States, we
219 primarily use the NOAA Monthly U.S. Climate Gridded Dataset (NClimGrid; Vose et
220 al., 2014), which is a station-based gridded product of temperature (and precipitation)
221 across land areas of the conterminous United States since 1895. NClimGrid is based on
222 the interpolation of quality-controlled station data onto 5 km latitude/longitude grids
223 using records from the Global Historical Climatology Network (GHCN; Durre et al., 2010;
224 Menne et al., 2012). The homogenized dataset from NClimGrid also accounts for bias
225 correction of artificial station breaks, such as for changes in weather station locations,
226 instruments, and other temporal inconsistencies (Menne & Williams, 2009). Area-average

227 NCLimGrid temperature errors are larger over CONUS prior to 1990, but generally still
 228 within 1°C for both TMAX and TMIN (Vose et al., 2014). Here we focus on the period
 229 of 1921 to 2022, which overlaps with the output from the SPEAR climate model sim-
 230 ulations.

Table 2. Summary of the station-based and atmospheric reanalysis datasets (i.e., observa-
 tions) used in this study. The observational maps are then regridded to make inferences using
 the neural networks onto either the same LOW (1.0°; ERA5/NCLimGrid/20CRv3) or MED (0.5°;
 ERA5/NCLimGrid) atmospheric resolution as from GFDL SPEAR. See Section 2.1.

Name	Data	Years	Horizontal Resolution	Reference
ERA5	ECMWF Reanalysis v5	1940-2022	~ 0.25°	Hersbach et al. (2020)
NCLimGrid	NOAA Monthly U.S. Climate Gridded Dataset	1895-2022	~ 1/24°	Vose et al. (2014)
20CRv3	NOAA-CIRES-DOE Twentieth Century Reanalysis v3	1836-2015	~ 1.0°	Slivinski et al. (2019, 2021)

231 For comparing the sensitivity of our observational neural network predictions to
 232 the use of NCLimGrid, we also briefly evaluate our results with two atmospheric reanal-
 233 ysis datasets: the European Center for Medium-Range Weather Forecasts (ECMWF)
 234 fifth generation of atmospheric reanalysis (ERA5) available from 1940 to 2022 and the
 235 NOAA/Cooperative Institute for Research in Environmental Sciences/Department of En-
 236 ergy Twentieth Century Reanalysis (20CR) version 3 (20CRv3) available from 1836 to
 237 2015. Using ECMWF’s Integrated Forecast System release 41r2 and four-dimensional
 238 variational analysis as a data assimilation scheme, ERA5 provides global data at a hor-
 239 izontal resolution of 31 km in near-real time (Hersbach et al., 2020). It is constrained
 240 by numerous observations, like land-based weather stations, satellites, radiosondes, and
 241 other aircraft records. We focus on near-surface temperature (2-m height) from ERA5
 242 through its entire available temporal period (1940 to 2022). To compare with a longer
 243 reanalysis record, we use near-surface temperature (2-m height) from 20CRv3 (Slivinski
 244 et al., 2019) during the overlapping period from 1921 to 2015. Unlike ERA5, this 20CR
 245 product only assimilates surface pressure observations (Compo et al., 2011), which is com-
 246 pleted through four-dimensional incremental analysis updates and an 80-member ensem-
 247 ble Kalman filter approach (Lei & Whitaker, 2016; Slivinski et al., 2019). 20CRv3 uses
 248 the coupled atmosphere-land National Centers for Environmental Prediction (NCEP)
 249 Global Forecasting System (GFS) version 14.0.1 with boundary conditions from prescribed
 250 sea surface temperatures and sea-ice concentration. Overall, 20CRv3 is an improvement

251 over its predecessor (20CRv2c) for simulating synoptic dynamics and other long-term
252 surface climate fields (Slivinski et al., 2021), though work continues to improve temper-
253 ature data earlier in the twentieth century due to greater uncertainties (Gillespie et al.,
254 2023).

255 We use monthly mean temperature output from all observational and reanalysis
256 datasets to calculate the JJA seasonal mean. The neural network used here requires the
257 input data to have the same latitude and longitude dimensions. Therefore, we bilinearly
258 regrid NCLimGrid and ERA5 onto the MED (0.5°) and LOW (1.0°) spatial maps which
259 are also used by the climate model large ensemble data. The coarser 20CRv3 dataset
260 is instead only interpolated onto the LOW resolution grid. Importantly, these three ob-
261 servationally based products encompass a wide range of different structural methodolo-
262 gies and uncertainties, which therefore provide ample opportunity to test the robustness
263 of the neural network results on out-of-sample data (Table 2). A comparison of average
264 JJA CONUS temperature anomalies is also presented in Figure S1 for NCLimGrid, ERA5,
265 and 20CRv3.

266 **3 Methods**

267 **3.1 Neural Network Framework**

268 We adopt a machine learning ToE method first proposed by Barnes et al. (2019),
269 which uses a neural network to input geographic maps of climate variables and then to
270 output the year associated with each map. While this is quite a simple prediction prob-
271 lem, it has been shown that the neural network must learn to leverage time-evolving pat-
272 terns of forced climate signals in order to correctly identify the year with a single map
273 (Barnes et al., 2020). This attribution method has since been used in a wide range of
274 climate applications (e.g., Anderson & Stock, 2022), such as for disentangling the role
275 of aerosols and greenhouse gases in single-forcing large ensembles (Labe & Barnes, 2021),
276 quantifying anthropogenic signals in extreme precipitation (Madakumbura et al., 2021),
277 and identifying the ToE of combined variables like precipitation and temperature (Rader
278 et al., 2022). For this work, we take a similar approach, but build upon these previous
279 efforts by focusing on a narrower application. Here we train on high-resolution climate
280 model data and evaluate the ToE on a smaller spatial region during a period with an ob-
281 served absence of daytime warming (e.g., boreal summer in the Central United States)

282 (Partridge et al., 2018). While previous machine learning efforts in climate science have
 283 usually interpolated data to coarser grids for reasons such as computational limitations,
 284 we are particularly interested in considering whether the neural network skill (or the actual
 285 ToE) changes by training and evaluating on higher resolution data.

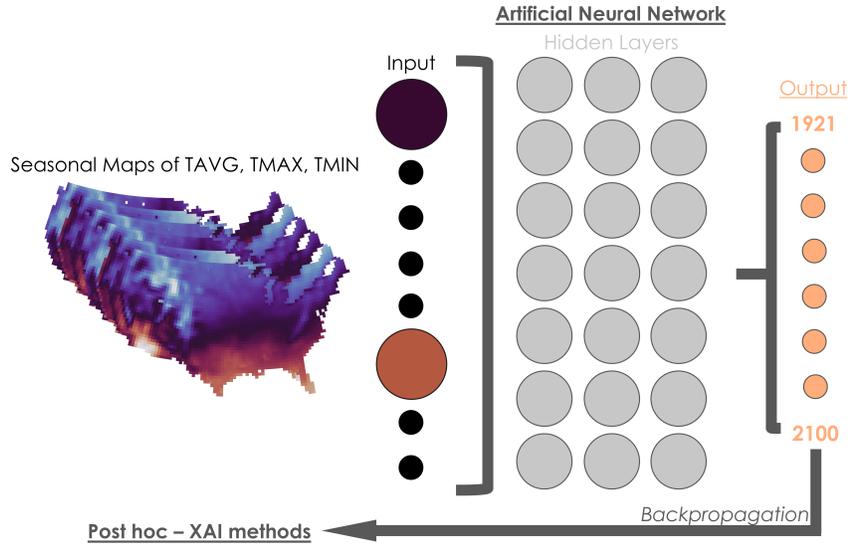


Figure 1. Schematic of the artificial neural network (ANN) used to take an input map of average June to August (JJA) temperatures over the contiguous United States (CONUS) and then output the likelihood that the map is from a particular decade. Fuzzy classification (Zadeh, 1965) is used to decode this decadal likelihood to associate each map with a single year (i.e., our final predicted output). The ANN consists of different combinations of hidden layers and nodes depending on the horizontal resolution of the training data. (see Section 3 for the architecture and hyperparameter choices). Explainable artificial intelligence (XAI) attribution methods are then used to reveal the regions that acted to increase or decrease the likelihood of the ANN’s predicted year.

286 For this study, we use an ANN, which is a statistical algorithm that can learn to
 287 approximate nonlinear functions from large quantities of data. They have become in-
 288 creasingly popular tools for Earth science prediction problems and in numerical mod-
 289 eling (Boukabara et al., 2021; Chantry et al., 2021; Irrgang et al., 2021). ANNs are fully-
 290 connected networks, which in their simplest form are comprised of an input layer, a set
 291 number of hidden layers and nodes, and an output layer (i.e., the final prediction). Ev-

292 ery node in this feed-forward architecture receives information from the previous layer
293 and can be individually computed by weighting the sum of the inputs and an added bias
294 term. The weights and biases are updated iteratively until the training process is fin-
295 ished, such as when the loss function is minimized (i.e., a measure of machine learning
296 model error). Given enough available data and limited overfitting, the ANN can then
297 be used to make skillful predictions on data it did not see during training (i.e., testing
298 data). More thorough introductions to neural networks can be found in e.g., Lecun et
299 al. (2015); Goodfellow et al. (2016); Neapolitan and Jiang (2018). Domain-specific tu-
300 torials for machine learning applications, such as in atmospheric science, are also pro-
301 vided by Chase, Harrison, Burke, et al. (2022) and Chase, Harrison, Lackmann, and Mc-
302 Govern (2022).

303 Figure 1 shows the ANN architecture used for this study. The ANN receives vec-
304 torized latitude by longitude maps of JJA temperatures (TMAX, TMIN, or TAVG) with
305 either 10,080 input values (70×144) per sample for the MED resolution maps or 2,520
306 input values (35×72) for the LOW resolution maps. In addition, we also evaluate ANNs
307 using regional map inputs for the Western, Central, and Eastern CONUS as depicted
308 in Figure 2. We focus on only land areas and therefore mask all other areas by assign-
309 ing values of zero, which the ANN then learns to ignore. This vector is fed into the ANN
310 hidden layers, and the output is the probability of a particular decade midpoint (a clas-
311 sification problem). These output values are then translated to a particular year (a re-
312 gression problem) using fuzzy classification (Zadeh, 1965), which is described in detail
313 in Barnes et al. (2020). Briefly, by denoting the central year of a particular decade (e.g.,
314 1995 for 1990-1999), a particular output (e.g., 1994) can be mapped to more than one
315 decade class. Using triangular membership functions (Zadeh, 1965) with a width of one
316 decade, the weighted sum of the decadal class probabilities can finally be mapped to a
317 specific year. For example, the year 1994 has a probability of 0.9 for falling within the
318 decade class midpoint of 1995 and a probability of 0.1 for the decade class midpoint of
319 1985. We refer to the regression problem of the predicted year throughout the rest of the
320 study for evaluating the ANN skill and ToE calculations.

321 We select our ANN architecture for different input maps by considering the effect
322 of spatial region and grid resolution. The final ANNs are selected by identifying the low-
323 est median Mean Absolute Error (MAE) on validation data after considering 20 ANNs
324 (randomized combinations of training, testing, and validation data and initialization seeds)

325 over architectures that range in complexity by the number of hidden layers and nodes
 326 (see Figures S6 and S10). In other words, this is related to the number of parameters
 327 that the model can use to learn the relevant climate patterns to more accurately pre-
 328 dict the year of a map (i.e., variations of deeper or shallower neural networks). For CONUS
 329 inputs on the MED grid, we use an ANN with 3 hidden layers of 10 nodes each. For CONUS
 330 inputs on the LOW grid, we use an ANN with 3 hidden layers of 20 nodes each. Finally,
 331 for the regional CONUS maps on both the MED and LOW grids, we use an ANN with
 332 2 hidden layers of 100 nodes each. Despite selecting different ANN architectures, we find
 333 that our results are generally robust across minor changes in hyperparameter options.
 334 The rectified linear unit (ReLU; Agarap, 2018) activation function is used for the non-
 335 linear transformation in the hidden layers, and a softmax operator is included in the out-
 336 put layer to ensure that the decadal class probabilities of the output vector sum up to
 337 one. All ANNs here use the binary cross-entropy loss function, stochastic gradient de-
 338 scent optimizer (Ruder, 2016) with Nesterov momentum set to 0.9 (Nesterov, 1983), a
 339 learning rate of 0.01, and a batch size set to 32.

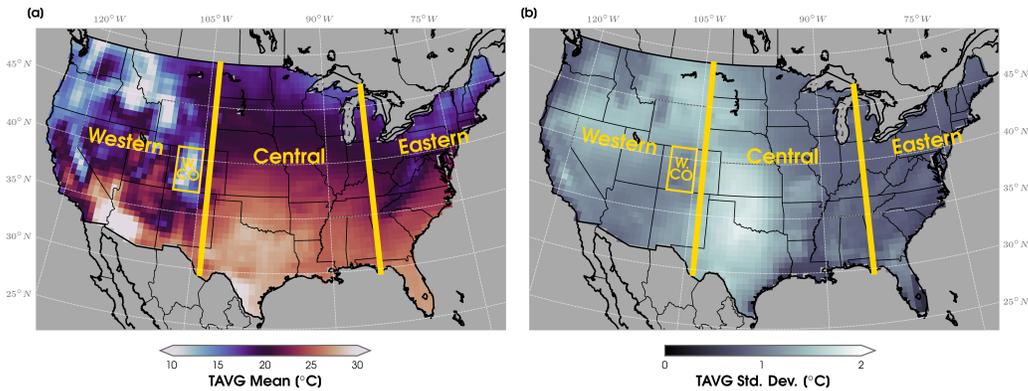


Figure 2. (a) Composite of JJA average near-surface temperature (TAVG) from an example of the training data mean used to standardize the input maps, which is calculated here using 24 ensemble members from SPEAR_MED over the period of 1981-2010. (b) As in (a), but for the training data standard deviation. See Section 3.1 for more details. The vertical yellow longitude lines are displayed at 104°W and 85°W to differentiate the three regions of the CONUS considered for this work (i.e., the Western USA, Central USA, and Eastern USA). The thin yellow box outlines the western Colorado (W. CO) region of interest used for this study (37°N - 41°N and 108°W - 105°W).

340 Unless otherwise stated, we train on 24 ensemble members, validate on 4 ensemble
341 members, and test on 2 ensemble members. To limit overfitting on the training data,
342 we apply a few different methods. First, we apply early stopping, which ends the ANN
343 training process if there is no improvement in the validation loss for 25 consecutive epochs.
344 The epoch with the best performance is ultimately selected. Next, we apply ridge reg-
345 ularization (set to 0.001) to the weights of the first hidden layer (L_2 ; Friedman, 2012),
346 which acts to limit how sensitive the weights are to outliers in the input data. This also
347 helps to smooth out any spatial autocorrelation that exists in the temperature maps and
348 improve overall interpretability (Sippel et al., 2019, 2020; Barnes et al., 2020). The sen-
349 sitivity of our ANN results to the choice of the L_2 parameter are shown in Figures S6
350 and S10.

351 Before inputting the data into the ANN, we standardize all climate model temper-
352 ature maps by subtracting the training data mean and dividing by the training data stan-
353 dard deviation over the 1981 to 2010 climatological baseline. This is computed separately
354 at every grid point. Note that similar skill is found for training and testing data using
355 other reference periods, such as 1951-1980. An example of the training mean and stan-
356 dard deviation for 24 ensemble members in SPEAR_MED is shown in Figure 2. Due to
357 mean state biases that may exist between the climate model large ensembles and obser-
358 vations, we separately standardize the observations by their own mean and standard de-
359 viation over 1981 to 2010 before making ANN inferences. Though, as we discuss later,
360 it is still possible that differences in the amount of mean warming between the climate
361 model simulations and observations could influence the machine learning skill and re-
362 lated ToE results.

363 In addition to evaluating our ANN and ToE predictions, we consider several ad hoc
364 attribution methods of XAI. Explainability methods have increasingly been shown to aid
365 in building trust and understanding for the decision-making process of neural networks,
366 including for climate science applications (e.g., Sonnewald & Lguensat, 2021; Labe & Barnes,
367 2022; Shin et al., 2022; Rampal et al., 2022; Diffenbaugh & Barnes, 2023; Mamalakis et
368 al., 2023). Output from XAI attribution methods describe the contribution of every in-
369 put sample’s latitude and longitude grid point (described here as “relevance”) to the over-
370 all prediction of the ANN. In other words, the XAI algorithms return a relevance heatmap
371 (unitless) for every input year. To evaluate the sensitivity our explainability results across
372 different methods (Mamalakis et al., 2022; Bommer et al., 2023), we consider three dif-

ferent attribution techniques: the layerwise relevance propagation z-rule (LRP_z; Bach et al., 2015), LRP epsilon-rule (LRP_ε; Bach et al., 2015) and Integrated Gradients (Sundararajan et al., 2017). Though given the similarity of the results across the XAI methods, we only show relevance figures from LRP_z and Integrated Gradients for brevity in the main results. A more detailed overview on an application of LRP to a geoscience problem can be found in Toms et al. (2020). Lastly, we caution that although these XAI techniques are very useful for outlining the important climate patterns learned by the ANN, they do not imply causation, such as for the specific physical drivers.

Our XAI heatmaps are based on composites of the testing ensemble members (or observations), where positive areas of relevance can be interpreted as regions that pushed the ANN toward its predicted year. Negative areas of relevance are subsequently interpreted as vice versa, i.e., locations that tried to push the ANN away from making its yearly prediction. We only consider relevance maps for testing data predictions that are accurate to within 5 years of the actual year.

3.2 Calculated Timing of Emergence for Observations

An annotated graphic of our ToE approach is shown in Figure 3. The ANN in this example is trained and tested on SPEAR_MED for maps of mean JJA TAVG. The testing ensemble predictions are shown using green scatter points with the actual year of a TAVG map on the x-axis, and the ANN predicted year is on the y-axis. Predictions for maps of mean JJA TAVG from NClimGrid data are then shown with red markers. The fundamentals of this ToE estimate are based on Mora et al. (2013), which is calculated as the year that a map of temperature first departs the bounds of historical climate variability and continues to do so for all future years (e.g., the red shading in Figure 3). This general approach was also demonstrated using ANNs in Barnes et al. (2019) and Rader et al. (2022). Here we define the ToE as the first predicted year that is greater than the maximum prediction during our historical baseline of 1921-1950 (gray shading in Figure 3). In addition to 1921-1950 being the earliest 30-year period available from SPEAR_MED, it more importantly overlaps with the observed anomalous warmth of the 1930s in the United States (Program, 2018) (Figure 4). Thus, we can directly compare whether an observed JJA temperature forced signal has emerged outside of this historical record that includes the extreme heat of the Dust Bowl era.

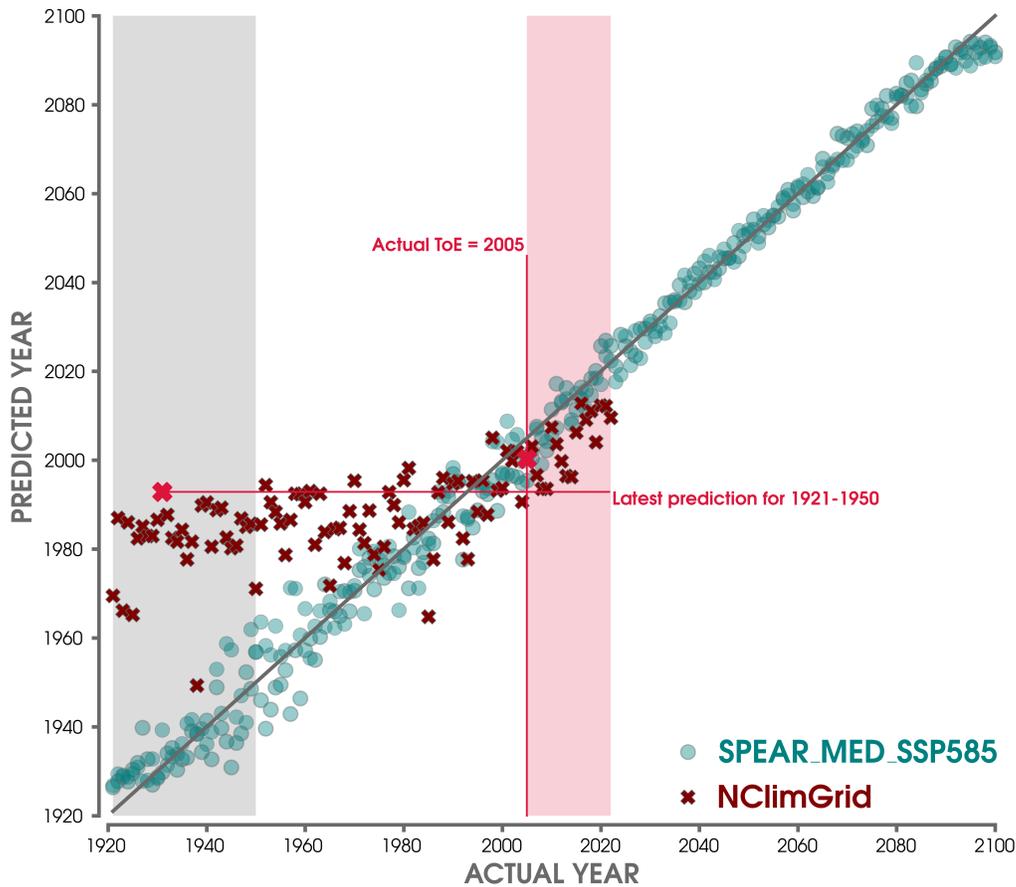


Figure 3. Schematic of the output provided by the ANN and the subsequent calculation of the timing of emergence (ToE) based on CONUS maps from observations. Blue scatter points denote ANN predictions based on SPEAR_MED (historical + SSP5-8.5 forcing) testing ensemble members for the inputs of TAVG averaged over JJA. The actual year is shown on the x-axis, and the predicted year on shown on the y-axis. Red markers are used for ANN predictions after inputting maps from NCLimGrid. A perfect prediction (1:1 slope) is annotated behind all ANN predictions with a solid gray line. To first calculate the ToE of the NCLimGrid maps, the latest predicted year in the 1921 to 1950 climatological period (vertical gray shading) is identified (left, bright red marker). The actual ToE (right, bright red marker) is then the first year when all proceeding predictions (vertical red shading) exceed the year of the 1921-1950 maximum. The observed ToE from this ANN is 2005.

404 We also compare our results with a more traditional baseline ToE estimate using
405 SPEAR_MED data. For a given ensemble member, the ToE is the first year that the 10-
406 year running JJA mean consistently stays above the 1921 to 1950 climatological period
407 by greater than two standard deviations (Lehner et al., 2017). This variability is again
408 based on the 10-year running mean temperature in 1921-1950. The actual ToE compos-
409 ites from this method are then calculated from the mean across all ensemble members
410 and at every grid point.

411 4 Results

412 4.1 Changes in United States Summertime Temperatures

413 Before estimating the ToE by the neural network framework, we start by assess-
414 ing changes in temperature in observations and as simulated by SPEAR_MED. Figure
415 4 shows the time-mean JJA temperature anomalies averaged over the CONUS for TMAX,
416 TMIN, and TAVG from 1921 to 2022. We find that observations from NCLimGrid lie out-
417 side the ensemble spread of SPEAR_MED in all three temperatures metrics during the
418 Dust Bowl of the mid-1930s. This is especially prominent for JJA TMAX, which reaches
419 values of more than 1°C greater than the warmest ensemble member from SPEAR_MED.
420 The peak TAVG during this early 20th century period was reached in 1936 (Cowan et
421 al., 2017, 2020), but it is now statistically tied with 2021 as the hottest summer on record
422 (within 0.01°C) over the CONUS (Thompson et al., 2022) (Figure 4c). The large climate
423 response following the eruption of Mount Pinatubo (Parker et al., 1996), however, is well
424 captured by SPEAR_MED. 1992 subsequently remains the coldest mean summer TMAX
425 in the NCLimGrid observational record (since at least 1921) (Figure 4a). In more recent
426 years, temperatures from NCLimGrid have remained consistently below the ensemble mean
427 and therefore exhibit less net warming than the forced response in SPEAR_MED, espe-
428 cially for TMAX. We also compare NCLimGrid with ERA5 and 20CRv3 reanalysis prod-
429 ucts for TAVG anomalies in Figure S1. While NCLimGrid and ERA5 agree well in cap-
430 turing the interannual variability and long-term trends, we find a larger discrepancy in
431 TAVG prior to 1975 where 20CRv3 shows larger warm anomalies. These discontinuities
432 have been pointed out in previous studies using older model generations of 20CR (Ferguson
433 & Villarini, 2014), which were found to be largest in the mid-20th century for the Cen-
434 tral United States (Ferguson & Villarini, 2012).

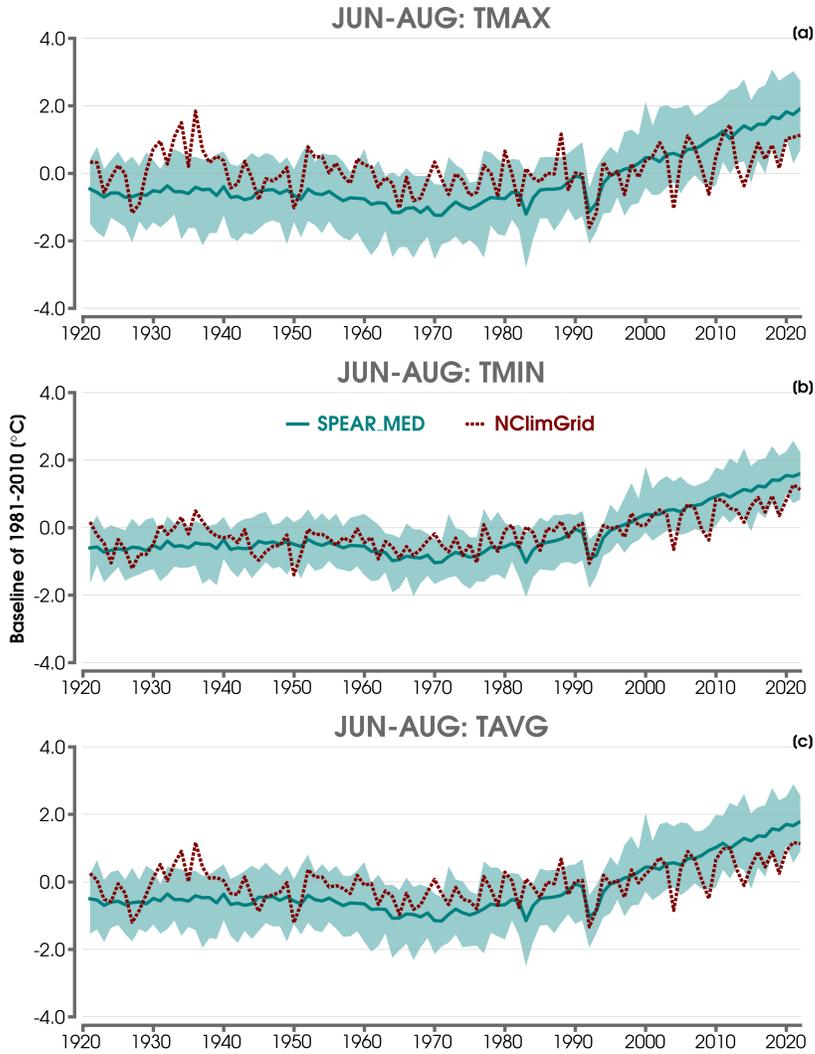


Figure 4. (a) Time series of mean JJA maximum temperature (TMAX) anomalies over the CONUS from 1921 to 2022 for the ensemble mean of SPEAR_MED (dark green line) compared to observations from NCLIMGRID (dashed red line). The spread across SPEAR_MED ensemble members is shown with the light green shading. Anomalies are computed for each dataset with respect to their own 1981-2010 climatological mean. (b) As in (a), but for the mean JJA minimum temperature (TMIN). (c) As in (a), but for the mean JJA average temperature (TAVG)

435 Figure S2 shows the time series of TAVG anomalies broken down by the three re-
 436 gions of interest, including the Western, Central, and Eastern United States. This more
 437 clearly distinguishes the anomalous heatwaves of the mid-1930s across the Central United
 438 States, which again fall outside the ensemble spread of SPEAR_MED. All three regional
 439 domains experience substantial interannual temperature variability in the NClmGrid
 440 record and reveal less long-term warming compared to the SPEAR_MED ensemble mean
 441 over the 1990 to 2022 period.

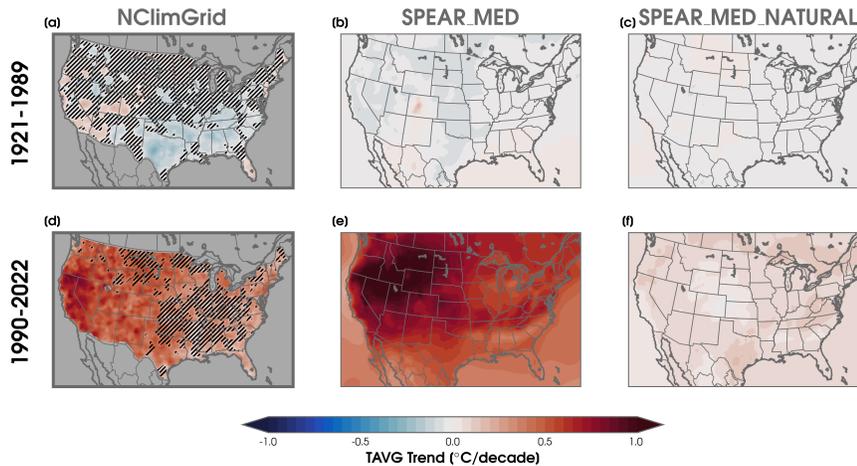


Figure 5. Linear least squares trends of average JJA TAVG from 1921 to 1989 (a, b, c) and 1990 to 2022 (d, e, f) for NClmGrid (a, d), the ensemble mean from SPEAR_MED (b, e), and the ensemble mean from SPEAR_MED_NATURAL (c, f). For maps of NClmGrid, black hatch marks indicate TAVG trends that are not statistically significant following a Mann-Kendall test (Bevan & Kendall, 1971; Mann, 1945) for the 95% confidence level.

442 Figure 5 provides spatial maps of the TAVG trends for 1921 to 1989 and 1990 to
 443 2022. Statistically significant cooling trends in NClmGrid are found over the southern
 444 United States (Figure 5a), which are close to the warming hole region for this period (Mascioli
 445 et al., 2017). The observational trends are also compared to SPEAR_MED and its par-
 446 allel natural forcing-only simulation (SPEAR_MED_NATURAL) in Figure 5b-c, which
 447 does not simulate any long-term TAVG trends in their ensemble means. However, greater
 448 warming is found in observations for the more recent past (1990-2022; Figure 5d), which
 449 is largest over the Southwestern United States. The warming hole spatial pattern is again
 450 found over the Central United States (Figure 5d). In comparison to observations, SPEAR_MED

451 simulates greater ensemble mean warming over the entire CONUS and shows the largest
452 TAVG trends over the northern Rocky Mountains (Figure 5e). Without anthropogenic
453 forcings, such as greenhouse gases or aerosols, Figure 5f reveals little to no warming sim-
454 ulated over CONUS as found in the ensemble mean of SPEAR_MED_NATURAL. Lastly,
455 comparing over the entire 1921 to 2022 historical record, we find that the ensemble mean
456 of SPEAR_MED simulates slightly greater warming trends over CONUS (Figure S3b)
457 and does not show evidence of the warming hole as in Figure S3a. Though it remains
458 unclear whether this is simply due to internal variability in the observational record, which
459 would not be captured in a composite of the ensemble mean trends of a large ensemble
460 (Eischeid et al., submitted).

461 Focusing on JJA mean maximum and minimum temperatures, Figure S4 shows larger
462 recent trends in TMAX than TMIN across the Northern and Western United States in
463 the observed record. Consistent with previous findings, the warming hole is also more
464 prominent in TMAX. This is reflected by an area of insignificant cooling across the South-
465 eastern United States (Figure S4a). Although SPEAR_MED again simulates greater posi-
466 tive trends in TMAX (Figure S4b) and TMIN (Figure S4d) for the CONUS (exceed-
467 ing 1°C per decade at its local maximum), there are similarities in the spatial pattern
468 of mean warming compared to NclimGrid. This includes a relative maximum in warm-
469 ing over the Western United States and relative minimum over the Southeastern United
470 States (Figure S4d,b).

471 Lastly, we show in Figure S5 the JJA mean CONUS TAVG for NclimGrid com-
472 pared to a collection of 30-member large ensembles of SPEAR_MED, but using differ-
473 ent radiative forcing scenarios from 2015 to 2100. This includes future projections from
474 SSP1-1.9, SSP2-4.5, and SSP5-8.5, which are compared to the natural-only forcing ex-
475 periment of SPEAR_MED_NATURAL from 1921 to 2100. The forced response in the
476 SPEAR_MED historical simulations only begins to clearly rise outside the variability in
477 the natural forcing simulation between 1990 and 2000. This occurs a decade later when
478 comparing NclimGrid to SPEAR_MED_NATURAL. While there is a large range in pro-
479 jected ensemble mean JJA TAVG change across the climate change scenarios to 2100,
480 the uncertainty due to internal variability alone is almost 2°C across the ensemble spreads.
481 Notably, we also find that the ensemble mean TAVG begins to cool by 2040 for the ag-
482 gressive climate mitigation scenario of SSP1-1.9. This continues through the end of the
483 21st century for SSP1-1.9. We also point out that there are only negligible differences

484 across future forcing scenarios in the ensemble means of JJA TAVG until between 2030
485 and 2040, but by after 2080, there is very little overlap in their ensemble spreads due to
486 the greater effects of projection scenario uncertainty.

487 **4.2 Predictions by Neural Networks**

488 We now turn to the machine learning results in Figure 6, which shows the skill of
489 the ANN for predicting the year of CONUS maps of TAVG, TMAX, or TMIN. Note that
490 Figure 6a is from the same ANN as the one displayed in the annotated schematic in Fig-
491 ure 3. Again, we focus our results on only the testing ensemble members from SPEAR_MED,
492 which are data that the ANN has never seen before. The testing data predictions (blue
493 scatter points) closely follow the 1:1 line (or perfect prediction) in all three ANNs, which
494 suggests that the ANN is able to distinguish individual JJA temperature maps despite
495 the background noise of internal climate variability. The robustness of these results to
496 different ANN architectures and ridge regularization parameters are shown in Figure S6
497 for TAVG, which are each assessed for 20 ANN iterations that used different combina-
498 tions of training, validation, and testing ensemble members and random initialization
499 seeds. The median MAE score from this distribution of ANNs is displayed in Figure S6h
500 for the architecture used to produce Figure 6 (see Section 3.1). This helps to ensure that
501 our high skill is not simply due to the chance that our ANN performed well on only one
502 subset of testing data or overfit on the training ensemble members.

503 The results for the observational predictions are also shown in Figure 6. To restate
504 from earlier, these predictions are obtained by inputting maps of JJA temperature into
505 the ANN after it has already been trained and tested on the climate model large ensem-
506 ble data. However, unlike the predictions for SPEAR_MED, we do not find that the ANN
507 can correctly predict the year during most of the 20th century for TAVG, TMAX, or TMIN.
508 As described in Rader et al. (2022), since the ANN is not confident in predicting the year
509 of a given temperature map, it tends to predict around the same year in the middle of
510 the entire time series (i.e., to reduce its potential error penalty in the loss function). How-
511 ever, especially for TAVG and TMIN, we find that the ANN observational predictions
512 begin to lie on the 1:1 line after around 1995. One measure that can be used to reveal
513 whether the ANN has identified patterns of forced change is by evaluating the order of
514 the predicted years (Labe & Barnes, 2021). Thus, this suggests that the ANN is begin-
515 ning to identify common patterns of forced climate change in NCLimGrid that were learned

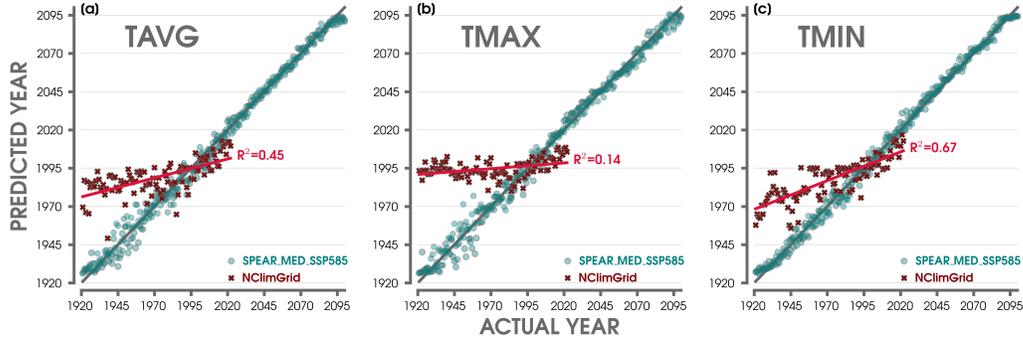


Figure 6. (a) Predictions of SPEAR_MED testing ensemble members by the ANN for input maps of TAVG averaged over JJA. The actual year is denoted on the x-axis and the predicted year on the y-axis. Red markers are shown for ANN predictions after inputting maps from NCLimGrid. A red line is displayed showing the linear least squares regression through the NCLimGrid predictions along with its corresponding R^2 value. A perfect prediction (1:1 slope) is annotated behind all ANN predictions with a solid gray line. (b) As in (a), but for an ANN trained and tested on maps of TMAX. (c) As in (a), but for an ANN trained and tested on maps of TMIN.

516 from the SPEAR_MED temperature maps in more recent years. As discussed later, we
 517 relate this point to the ToE of observed temperature change.

518 Since the SSP5-8.5 radiative forcing may be an unrealistic future climate scenario
 519 (Peters & Hausfather, 2020), we examine our results using ANNs trained on TAVG maps
 520 from the same historical forcing in SPEAR_MED, but then following either the SSP1-
 521 1.9 pathway (Figure S7a) or SSP2-4.5 pathway (Figure S7b). Overall, we find very sim-
 522 ilar skill for the testing ensemble member predictions across the SSP scenarios relative
 523 to SSP5-8.5 (Figure S7c), which is used throughout the remainder of the study. The pre-
 524 dictions for inputs of TAVG from NCLimGrid are also strikingly similar. We do point out
 525 that there is some higher testing data error toward the end of the 21st century, especially
 526 for SSP1-1.9 (Figure S7a), which suggests that the forced patterns of change may be-
 527 come less prominent after climate mitigation efforts (Figure S5). This implies more ev-
 528 idence for highlighting that the ANNs are learning to extract time-evolving climate sig-
 529 nals, including from within a single ensemble member’s realization of internal climate
 530 variability. Both of these detection outcomes are not as easily addressed by traditional
 531 signal-to-noise time-mean statistics.

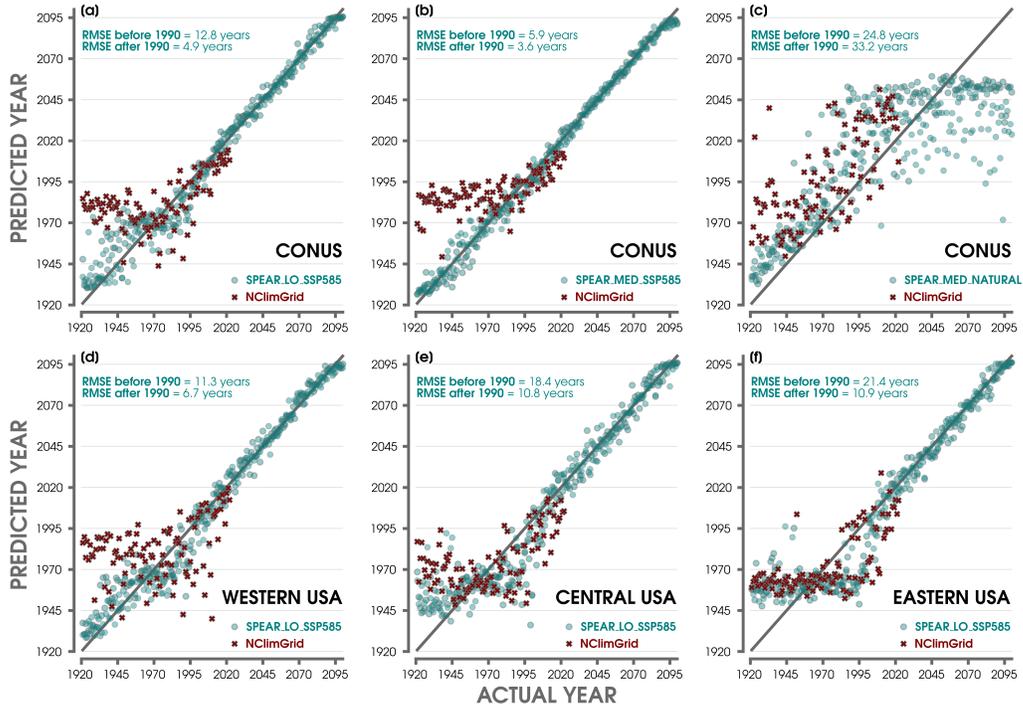


Figure 7. (a) Predictions of SPEAR_LO testing ensemble members by the ANN for input CONUS maps of TAVG averaged over JJA. The actual year is denoted on the x-axis and the predicted year on the y-axis. Red markers are shown for ANN predictions after inputting maps from NCLIMGRID. A perfect prediction (1:1 slope) is annotated behind all ANN predictions with a solid gray line. The Root Mean Squared Error (RMSE) for the SPEAR_LO testing ensemble members is included for predictions over the actual years of before and after the year 1990. (b) As in (a), but for SPEAR_MED. (c) As in (a), but for SPEAR_MED_NATURAL. (d) As in (a) but for SPEAR_LO predictions based on input maps of only the Western USA. (e) As in (a), but for SPEAR_LO predictions based on input maps of only the Central USA. (f) As in (a) but for SPEAR_LO predictions based on input maps of only the Eastern USA.

532 Another possibility is that our ANN inferences made on the observational maps
533 are sensitive to the choice of data product. We evaluate this prospect in Figure S8 by
534 inputting maps from either ERA5 reanalysis (Figure S8a) or 20CRv3 reanalysis (Fig-
535 ure S8b). The results from ERA5 are very similar to NClimGrid and once again indi-
536 cate that only by about 2000 is the ANN able to identify the order of the years of the
537 CONUS maps with a high degree of accuracy. On the other hand, we find greater di-
538 vergence in predicting the year for maps of 20CRv3. This could be due to the lower res-
539 olution of the training and testing data (i.e., LOW grid; Section 2) and/or a result of
540 the discontinuity in the TAVG anomalies prior to 1980 (Figure S1). As a result of the
541 greater uncertainties in the early 20CR data and lack of available data for the Dust Bowl
542 era in ERA5, we focus on NClimGrid for the remainder of this study.

543 To identify the spatial regions that are important for the ANN testing skill for SPEAR_MED,
544 we evaluate composites of the XAI relevance maps in Figure S9 for the LRP_z, LRP_e, and
545 Integrated Gradients methods. These composites are assessed over the entire time se-
546 ries from 1921 to 2100. Again, positive areas can be interpreted as regions that were more
547 relevant for the overall ANN yearly predictions. Although there are some small differ-
548 ences outlined between the LRP methods and Integrated Gradients, all three compos-
549 ites reveal that the northern Rocky Mountains, such as in western Montana, are an im-
550 portant indicator region. Other relevant temperature regions include areas in the East-
551 ern United States, especially in southern Florida and on the leeward side of the Appalachian
552 Mountains. There is also a notable gradient along this topographic boundary, with neg-
553 ative areas of relevance (i.e., locations that pushed the ANN to predict another year/decade)
554 stretching from the Ohio Valley to western New York State.

555 Although the ANNs are clearly able to learn a climate signal that distinguishes one
556 climate model temperature map year from another, this does not necessarily imply it is
557 related to anthropogenic forcing. We therefore explore this possibility in Figure 7b, which
558 shows the predictions for an ANN trained on SPEAR_MED compared to the simulation
559 with only natural forcing (Figure 7c). This reveals that the ANN is no longer able to make
560 an accurate prediction of the year when trained on maps from SPEAR_MED_NATURAL.
561 Similarly, there is a much larger temporal spread in predictions after inputting NClimGrid
562 data into this trained ANN (Figure 7c) compared to the SPEAR_MED network. That
563 is, the ANN is likely using the response to external forcings, such as those prescribed in
564 SPEAR_MED, to more skillfully predict the year of summertime temperature maps even

565 when temperature trends are weaker prior to 1990 (Figure 2). Having said that, there
 566 is slightly smaller spread in the earlier yearly testing predictions of SPEAR_MED_NATURAL,
 567 which could be due to the ANN detecting a minor influence of solar or volcanic forcings.
 568 We did briefly explore training on an simulation of SPEAR with anthropogenic aerosols
 569 held fixed to 1921 levels (not shown), but found similar yearly map predictions as those
 570 from using SPEAR_MED, which implies a limited role for anthropogenic aerosols on our
 571 ANN ToE results.

572 As noted by the results when evaluating 20CRv3, a last possibility is that train-
 573 ing on the high spatial resolution of SPEAR_MED is having an important role in the skill
 574 of the testing ensemble members. Put in another way, the ANN could be more likely to
 575 weight spatial information, such as temperatures around topographical gradients, for iden-
 576 tifying the relevant climate indicators. We can compare this effect by training on data
 577 with the LOW grid from the SPEAR_LO configuration, which is demonstrated in Fig-
 578 ure 7a. In addition to higher Root Mean Squared Error (RMSE) for SPEAR_LO pre-
 579 dictions before and after 1990, there is also greater spread in prediction years after in-
 580 putting NCLIMGRID TAVG maps. Similar to earlier, the robustness of the ANN skill for
 581 training and testing on SPEAR_LO across different architectures is shown in Figure S10.
 582 Thus, the overall effect of grid size is explored more in Section 4.2.2.

583 *4.2.1 Regional Variations in Timing of Emergence*

584 So far, we have demonstrated that an ANN can distinguish the year of a given map
 585 of summertime temperatures across the contiguous United States after training on a high-
 586 resolution climate model large ensemble (SPEAR_MED). Consistent with recent work
 587 (e.g., Barnes et al., 2020; Labe & Barnes, 2021; Rader et al., 2022), the ANNs here are
 588 learning time-evolving temperature patterns associated with external forcing to differ-
 589 entiate each individual year and in the correct sequential order. Moreover, the ANNs can
 590 make skillful predictions on the order of temperature map years from out-of-sample ob-
 591 servations, but only in the last decade or two.

592 To associate the period when the ANN predictions for observations begin to fall
 593 along the 1:1 prediction line, we compute the observed ToE following the methods in Sec-
 594 tion 3.2 and outlined in Figure 3. In short, we find the maximum predicted year dur-
 595 ing the 1921-1950 reference period and then identify the ToE as the point where the forward-

596 looking ANN predictions no longer fall below this historical maximum. This is first cal-
 597 culated using NCLimGrid maps of CONUS that are seasonally averaged for JJA. To en-
 598 sure the robustness of our observed ToE estimates, we conduct 100 ANNs that are trained
 599 on SPEAR_MED and use the same architecture as previously outlined. The uncertainty
 600 spread in these ToE predictions, as displayed in Figure 8a for TAVG, TMAX, and TMIN,
 601 can be attributed to differences in ensemble members used for training and validation
 602 and through the choice of 100 different random initialization seeds. Figure 8b displays
 603 the Spearman's rank correlation calculated between the actual and predict years across
 604 the NCLimGrid inputs as a measure of skill for the ANN to correctly identify the order
 605 of the years (Labe & Barnes, 2021). Albeit, by construction, earlier ToEs will also cor-
 606 respond to higher correlation coefficients.

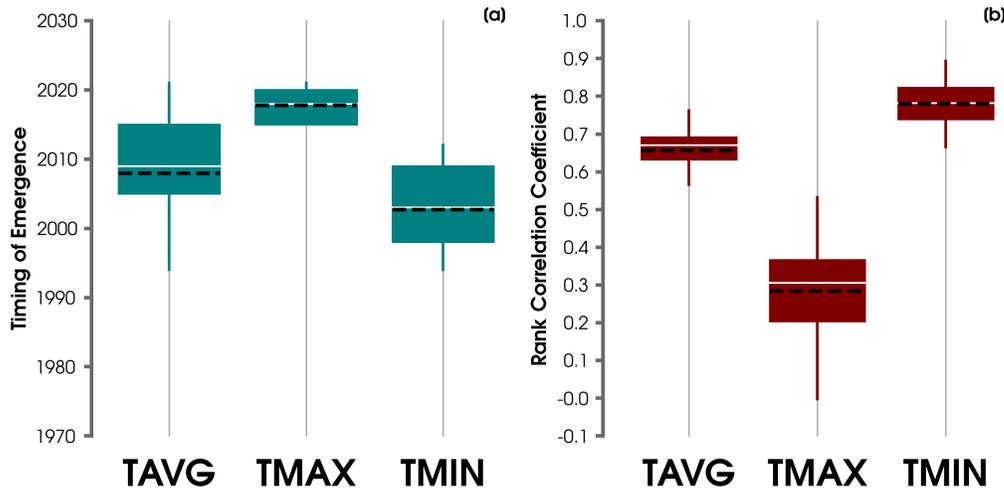


Figure 8. (a) Distribution of ToE predictions for inputs of CONUS maps using NCLimGrid after training ANNs on SPEAR_MED data of TAVG, TMAX, or TMIN. The median ToE is shown with a thin white horizontal line. The mean ToE is shown with a dashed black line. Each distribution of ToE years is constructed from 100 ANN iterations (use of random initialization seeds and different combinations of training, testing, and validation ensemble members). (b) As in (a), but for the distribution of Spearman's Rank correlation coefficients between all the actual years and predicted ANN years for NCLimGrid data.

607 The earliest median ToE is found for maps of TMIN, which is calculated to be 2003
 608 ($R=0.78$). On the other hand, the latest median ToE occurs in 2018 for TMAX ($R=0.31$).
 609 An important caveat, however, is that these ToE estimates could be biased early. This

610 is mainly an issue for late ToE predictions, like for TMAX, where by construction there
611 are few future years to compare with against the historical 1921-1950 maximum. We also
612 cannot rule out temporary reductions in temperature for a single JJA future year as a
613 result of internal climate variability (Maher et al., 2020) or from the influence of an event
614 like an explosive volcanic eruption (Sear et al., 1987).

615 We now investigate the ToE for the three selected regions across the United States
616 by separately training on mean JJA maps of TAVG, TMAX, and TMIN from SPEAR_MED,
617 but only over each smaller domain (outlined in Figure 2). Figure 9 shows the yearly pre-
618 dictions for these testing ensemble members and for regional inputs from NClimGrid.
619 If a forced signal has emerged in the observational record according to our definition, then
620 the ToE is annotated per each region. Note that if the estimated ToE is within five years
621 of present day (2022), then cautionary asterisks are included next to the ToE year, given
622 the greater uncertainty that future predictions over the next several years will remain
623 above the base period maximum. Across all three regions of CONUS, the earliest ToE
624 occurs for TMIN, especially for the Eastern United States at an estimate of 1998 ($R=0.84$)
625 (Figure 9f). We also find that the results for TMIN closely follow the 1:1 line, especially
626 after 1950 for the Central and Eastern United States (Figures 8e-f). This suggests that
627 these regional climate signals learned by the ANN after training on SPEAR_MED are
628 generalizable to NClimGrid. Most regions have not observed the emergence of a signal
629 in JJA TMAX (Figure 9a-c), and the earliest possible estimate here is for the Western
630 United States (ToE=2014, $R=0.28$) (Figure 9a). These results are consistent with re-
631 cent station-based studies finding greater warming rates in TMIN than TMAX (Meehl
632 et al., 2009; Abatzoglou & Barbero, 2014; Meehl et al., 2016; Program, 2018).

633 In addition to the observational results, we find that the testing predictions from
634 SPEAR_MED closely follow the 1:1 line, especially after 1990. The lowest testing RMSE
635 is found for the Western United States, and generally the worst ANN ensemble mem-
636 ber skill is found for TMAX and TAVG in the Eastern United States (Figures 8c and
637 8i). At the same time, the overall skill in the early to mid-20th century continues to be
638 surprising, especially given the lack of ensemble mean warming for all three regions (Fig-
639 ure S2). Therefore, in order to provide a baseline with a more traditional linear method
640 of calculating the ToE at each grid point, we compare the ANN predictions with the es-
641 timated ToE as shown in Figure S11 following Lehner et al. (2017) (see Section 3.2). For
642 most regions of CONUS, the more conventional ToE in SPEAR_MED for TAVG, TMAX,

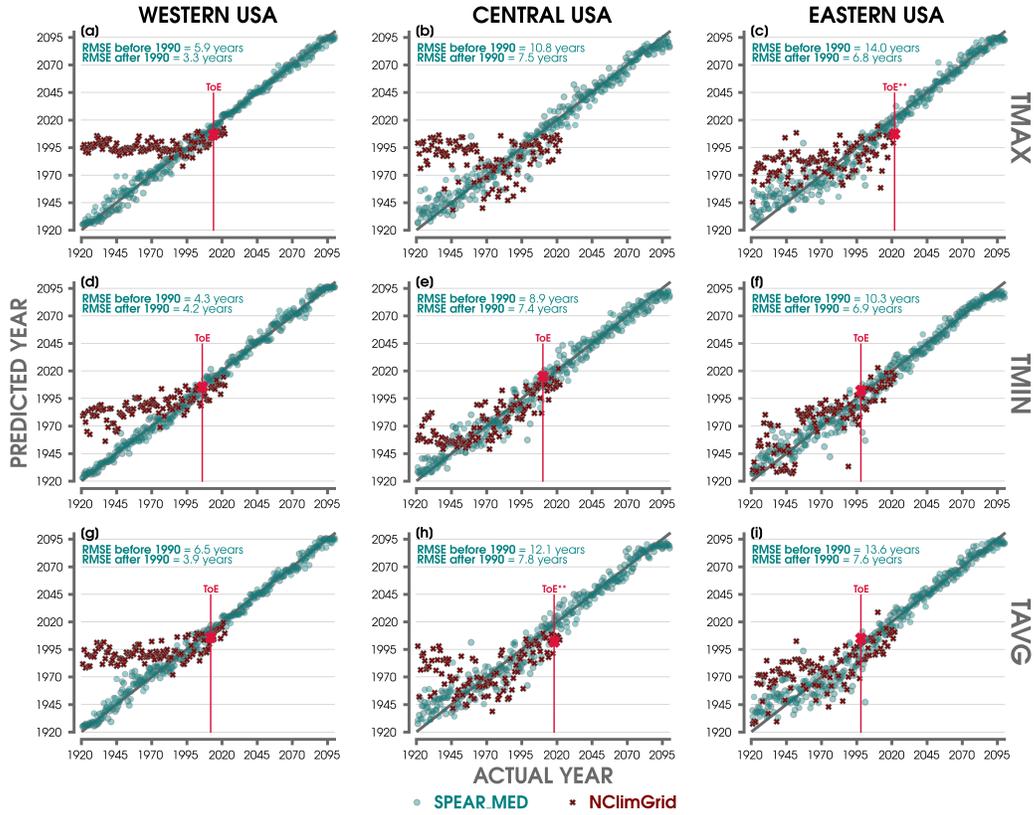


Figure 9. ANN predictions of SPEAR_MED testing ensemble members of TMAX maps for the Western USA (a), Central USA (b), and Eastern USA (c). The actual year is denoted on the x-axis and the predicted year on the y-axis. Red markers are shown for ANN predictions after inputting maps from NClmGrid. A perfect prediction (1:1 slope) is annotated behind all ANN predictions with a solid gray line. The RMSE for the SPEAR_MED testing ensemble members is included for predictions over the actual years of before and after the year 1990. If the observed ToE occurs for the NClmGrid predictions, then it is denoted for each region with a bright red marker and vertical line. If five or less years of predictions exist after this calculated ToE, then it is annotated with two added asterisks. (d-f) As in (a-c), but for input maps of TMIN. (g-i) As in (a-c), but for input maps of TAVG.

643 and TMIN occurs in the 1990s or 2000s, though there is some evidence of an earlier ToE
 644 across the higher elevations of the Western United States (Figure S11).

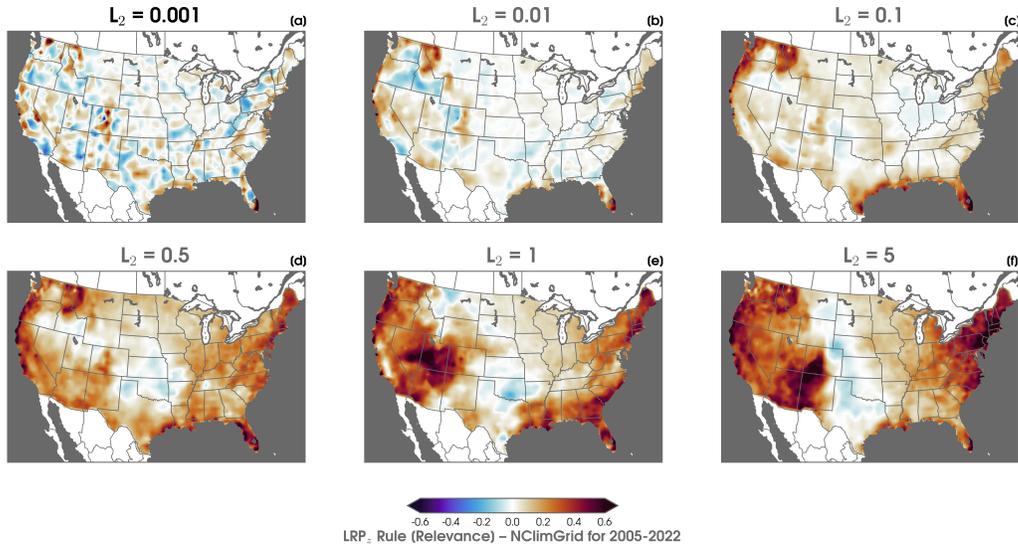


Figure 10. (a-f) Composite of LRP_z heatmaps from predictions based on NCLimGrid, which are averaged over 2005 to 2022 for ANNs trained on SPEAR_MED data using different L_2 regularization values (0.001, 0.01, 0.1, 0.5, 1, 5). The L_2 value used for the main results of the paper is labeled in bold font. Positive relevance indicates regions that pushed the ANN to make its predicted year. Negative relevance suggests areas that tried to push the ANN away from its predicted year.

645 Finally, we evaluate the XAI conclusions for NCLimGrid (after training on SPEAR_MED)
 646 by compositing those heatmaps over 2005 to 2022 in Figure 10a. This temporal range
 647 corresponds more closely to the period when a temperature signal has emerged (Figure
 648 8). As discussed in Section 3.1, the inclusion of ridge regularization can be a useful pa-
 649 rameter to limit the amount of overfitting. Likewise, it is also useful for interpreting the
 650 explainability results for how an ANN is making its predictions, as it is analogous to spa-
 651 tial smoothing for removing spurious outliers (Sippel et al., 2019; Barnes et al., 2020).
 652 These NCLimgrid relevance maps are shown in Figure 10 for ANNs trained on SPEAR_MED
 653 but using different regularization parameters. For lower ridge parameters, we find greater
 654 noise when interpreting the XAI maps, but overall higher positive relevance areas over
 655 portions of the Western United States in the vicinity of topography (Figure 10a-c). As
 656 the ridge parameter increases (i.e., penalizing larger weights to spread out the impor-

657 tance more evenly to reduce noisy collinearity), the spatial patterns of the XAI maps are
 658 smoothed, but note that the error of the ANN testing data skill begins to subsequently
 659 grow too (Figure S6). For the NClimGrid relevance maps in Figure 10d-f, we find that
 660 ANN is leveraging temperature patterns in the Eastern and Western United States (pos-
 661 itive relevance) to make its yearly predictions. Notably, this spatial pattern of relevance
 662 somewhat resembles the summertime warming hole (Mascioli et al., 2017).

663 ***4.2.2 Influence of Horizontal Resolution***

664 As this stage, it is evident that the ANNs are leveraging different spatial features
 665 to predict the year depending on the availability of fine detail on a given map grid, es-
 666 pecially at temperature gradients around high and low geographic elevations. To eval-
 667 uate this finding more closely, we now compare our regional results in Figure 7 with ANNs
 668 trained and tested on coarser regional maps from SPEAR_LO. Figure 7d-f shows these
 669 regional predictions of JJA TAVG and overall indicates poorer testing ensemble mem-
 670 ber skill for the Western, Central, and Eastern United States. In fact, the ANN is un-
 671 able to find any time-evolving signals in SPEAR_LO for the Eastern United States un-
 672 til the late 1990s. There is also greater spread in the yearly predictions after inputting
 673 LOW grid size maps from NClimGrid. As such, this result is consistent with Section 4.2
 674 that found lower skill in predicting the year of CONUS maps after training on SPEAR_LO
 675 compared to SPEAR_MED. To better compare the prediction skill between the two dif-
 676 ferent spatial resolutions of SPEAR_MED and SPEAR_LO, we show in Figure 11 the
 677 distribution of MAE scores across 20 different ANNs with the same architecture, but for
 678 different combinations of training, testing, validation data and random initialization seeds.
 679 We also compare the MAE of ANNs trained on the FLOR climate model large ensem-
 680 ble (MED grid) compared to ANNs trained on maps of CONUS from FLOR, but inter-
 681 polated onto the LOW grid spacing (FLOR (LO); see Section 2.1). For both of these ANN
 682 experiments, we again find lower MAE scores for the ANNs trained on the maps with
 683 higher spatial resolution (MED grid). This performance is also contrasted to the exper-
 684 iment without any anthropogenic forcing (not shown). The results of SPEAR_MED_NATURAL
 685 show a median MAE score over 1921 to 2100 for its best performing ANN architecture
 686 that is about 22 years, more than four times the MAE of the ensemble with anthropogenic
 687 forcing (lowest MAE on one ANN iteration/seed is still more than 18 years).

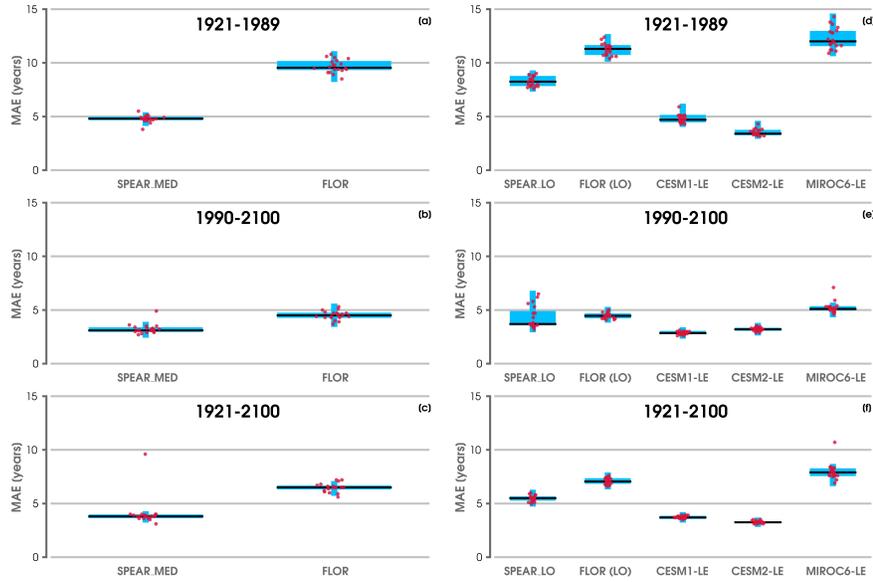


Figure 11. (a) Distribution of MAE scores for validation predictions over 1921-1989 based on inputs of CONUS maps for the overall ANN architecture with the lowest median MAE (e.g., Figure S6) after training neural networks on individual climate model large ensembles with the MED resolution (see Section 3.1) (SPEAR_MED or FLOR). Each distribution of scores (red points) is constructed from 20 ANN iterations (different combinations of training, testing, and validation ensemble members and random initialization seeds). The median score is shown with a thin black horizontal line. (b) As in (a), but for MAE scores calculated over 1990-2100. (c) As in (a), but for MAE scores calculated over 1921-2100. (d-f) As in (a-c), but for ANNs trained on individual large ensembles with the LOW resolution (SPEAR_LO, FLOR (LOW), CESM1-LE, CESM2-LE, MIROC6-LE).

688 Figure 11 also compares the skill using SPEAR and FLOR to other climate model
689 large ensembles, but this is only possible for their coarser resolution (LOW grid). Over-
690 all, the largest error in predicting the year of summertime TAVG maps is found for MIROC6-
691 LE. Notably, and despite the coarser resolution, the lowest error is found for CESM2-
692 LE across all climate model large ensembles for both before and after 1990. This is ex-
693 amined more closely in Figure S12 for a single ANN using MIROC6-LE, CESM1-LE, and
694 CESM2-LE along with the predictions after inputting observations from NclimGrid af-
695 ter training each network. Taking into consideration that there are more ensemble mem-
696 bers for these three climate models (compared to 30 total members in SPEAR and FLOR),
697 one possibility for the better skill in CESM2-LE is greater availability of training data.
698 Therefore, we conduct three more ANNs that are shown in Figure S12d-f, but use the
699 same number of training and testing data ensemble members as done with SPEAR. Close
700 results are found, and for that reason it is unlikely that the differences in skill are due
701 to more training data ensemble members. Similar to the testing data results, there is greater
702 spread in the NclimGrid predictions after training on MIROC6-LE (Figure S12a,d). Mean-
703 while, the NclimGrid predictions after training on CESM1-LE and CESM2-LE agree
704 broadly well with those from the SPEAR ANNs, but a TAVG signal again only emerges
705 by the late 1990s (i.e., predictions closer to the 1:1 line).

706 4.3 Early 20th Century Temperature Signals in SPEAR

707 A remaining question still across all of the results is how the ANN is able to dis-
708 tinguish summertime temperature maps for the climate model large ensemble data prior
709 to the late 20th century. Stated another way, what signals existed prior to 1990 when
710 the forced ensemble mean warming trend hasn't started to clearly emerge yet (e.g., Fig-
711 ure 4)? One advantage of using feature attribution XAI methods is that a relevance map
712 is obtained for each input sample. Accordingly, it's then possible to take XAI compos-
713 ites over different temporal periods to better understand the time evolution of the most
714 relevant temperature signals. Figure 12 shows the relevance maps using the LRP_z and
715 Integrated Gradients methods for composites of SPEAR_MED testing data before and
716 after 1990. For the 1921 to 1989 period, both XAI methods reveal a hotspot over west-
717 ern Colorado and generally muted relevance elsewhere across the United States. A sim-
718 ilar relevance pattern is also found for the XAI results based on inputs of TMAX (Fig-

719 ure S13). Apart from that, the relevance maps for 1990 to 2100 are more similar to those
720 described in Figure S9.

721 We compare the effect of regularization parameter and ANN architecture choices
722 on ANN skill for predicting the year of temperature maps from only 1921 to 1989. This
723 is displayed in Figures S14 and S15 for ANNs trained on SPEAR_MED and SPEAR_LO,
724 respectively. As previously discussed in Section 4.2.2, we again find better skill for the
725 higher resolution spatial grid and when using a lower ridge parameter. Relevance maps
726 for ANNs using different regularization parameters are composited for SPEAR_MED in
727 Figure S16, which reveals this effect of the ANN using smaller regional information, like
728 over the Rocky Mountains, for predicting the year of JJA temperature maps. With larger
729 ridge parameters comes worse ANN skill (Figure S6) but improved interpretability for
730 the XAI maps (Figure S16d-f). Correspondingly, it is then likely that the ANN is learn-
731 ing temperature indicators from finer spatial information, especially across western Col-
732 orado, during this early 20th century period.

733 In addition to testing data from SPEAR, the ANNs were also able to predict the
734 year of CONUS maps from other climate model large ensembles (Figure 11 and Figure
735 S12). We again turn to XAI methods in Figure S17, but alternatively for evaluating the
736 temperature signals leveraged by the ANN after inputs of CESM2-LE. The spatial pat-
737 terns of positive relevance in the early 20th century are different for CESM2-LE over 1921-
738 1989 (Figure S17a-b) compared to SPEAR_MED (Figure 12a-b). Instead, there are pos-
739 itive areas of relevance derived from CESM2-LE over the Southeastern United States and
740 again across portions of the Rocky Mountains but located north of the previous Colorado
741 hotspot.

742 Although there are no long-term temperature trends in the time-mean of SPEAR_MED
743 for the Western United States prior to around 1990 (Figure S2a), it is still possible that
744 there are spatial patterns of temperature change. Figure S18 shows the linear TAVG trends
745 for NCLIMGRID, SPEAR_MED, and SPEAR_MED_NATURAL from 1921 to 1950. While
746 observations reveal a few areas of cooling across the West, most of the CONUS does not
747 have any statistically significant temperature trends. In contrast, we find a patch of warm-
748 ing over Western Colorado in SPEAR_MED, which closely aligns with the previously iden-
749 tified XAI relevance hotspot (Figure 12a-b). Conspicuously, this warming is absent for
750 TAVG trends in SPEAR_MED_NATURAL (Figure S18c), which indicates it could be

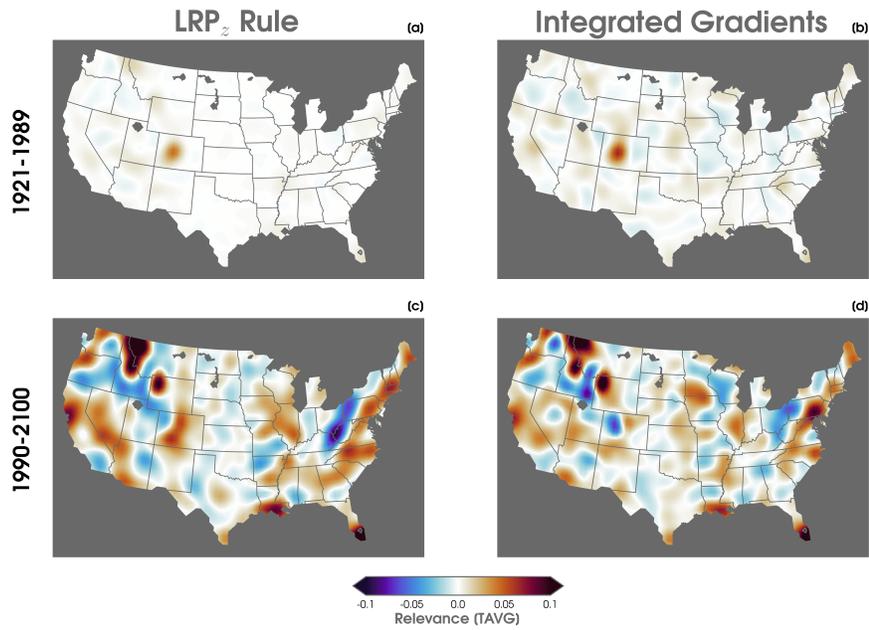


Figure 12. (a) Relevance heatmaps of TAVG using the layerwise relevance propagation z-rule (LRP_z) and (b) Integrated Gradients method for testing ensemble members from SPEAR_MED composited over 1921 to 1989. The composited heatmaps are smoothed using a Gaussian filter to improve interpretability. Positive relevance indicates regions that pushed the ANN to make its predicted year. Negative relevance suggests areas that tried to push the ANN away from its predicted year. (c-d) As in (a-b), but for composites over 1990-2100.

751 caused by anthropogenic forcing in SPEAR. We compare trends for TMAX and TMIN
752 in Figure S19 and also find mainly similar results for the small warming trend over west-
753 ern Colorado in the ensemble mean of SPEAR_MED.

754 To understand the possible physical drivers for this local warming in SPEAR_MED,
755 we calculate averages over a small box in western Colorado (outlined in Figure 2). We
756 find that this warming in SPEAR is strongly correlated with decreasing evaporation rates
757 (Figure S20a and Figure S21a), greater surface runoff, and even higher land surface tem-
758 perature warming (not shown). Despite this evidence, we do not find any correspond-
759 ing changes to precipitation or antecedent wintertime snowfall that could completely ex-
760 plain this effect (not shown). It is necessary to consider, however, that this small warm-
761 ing trend is still insignificant compared to the large spread of internal climate variabil-
762 ity as simulated by the individual ensemble members over the region (Figure S20b-d).
763 These trends are not found for the ensemble mean of SPEAR_MED_NATURAL (Fig-
764 ure S21b), as it is important to recall that land use and land change fields are also set
765 to 1921 levels in SPEAR_MED_NATURAL. We hence hypothesize that this small warm-
766 ing signal over western Colorado could be related to the land surface forcing in SPEAR_MED,
767 such as through the prescription of interactive vegetation influencing the surface energy
768 budget, but this is outside the scope of this study, and more work would be needed to
769 answer this question.

770 Finally, one last point we wish to make is that despite the ANN using this tem-
771 perature signal over western Colorado to help predict the year of a given map, it is not
772 the only reason for the high 1921-1989 skill (Figure 10). This is reflected in the results
773 from Figure 9, which shows that the ANN can still reasonably predict the year even for
774 regions of CONUS that do not include western Colorado.

775 **5 Discussion and Conclusions**

776 In this study, we used a machine learning approach to identify the ToE of summer-
777 time mean temperature extremes across the CONUS. There are several differences in this
778 methodology compared to more traditional signal-to-noise metrics used in earlier ToE
779 works. One advantage is that the ANN needs to learn time-evolving patterns of forced
780 change to make an accurate prediction, instead of only comparing time-mean metrics over
781 different epochs. In fact, we show that after training, the ANN is able to resolve these

782 forced signals even within a single ensemble member, i.e., one realization of internal cli-
783 mate variability. Changes in these temperature signals can then be visualized by apply-
784 ing XAI methods from one year to the next. Further, the ANN has the ability to learn
785 and identify (non)linear relationships across the entire spatial map to associate with the
786 ToE. These temperature patterns could differ from metrics calculated at the point-by-
787 point level or aggregate over larger domains, which all could subsequently impact ToE
788 estimates. By design, if observational predictions fall close to the 1:1 line, this suggests
789 that climate change patterns in the training data are generalizable to the real-world. This
790 is found to be the case in our ANN framework for the last two recent decades across dif-
791 ferent large ensemble climate models used for training.

792 We calculate the ANN-derived ToE by comparing to an early 20th century base-
793 line period, which encompasses the record heatwaves of the Dust Bowl-era. Nonetheless,
794 we find the emergence of a forced signal as early as the late 1990s for the observed TMIN
795 in the Eastern United States. More broadly, we also find the emergence of summertime
796 TAVG and TMIN across the entire conterminous United States. In other words, the ANN
797 can still distinguish a climate signal in JJA temperature maps during recent years, de-
798 spite the overall observed mean not exceeding the record warmth of the mid 1930s. It
799 is also possible that if the size of the spatial domain of the temperature maps were in-
800 creased, such as to consider all of North America, that the ToE may be identified even
801 earlier than found in this work (e.g., Barnes et al., 2019; Sippel et al., 2020; Labe et al.,
802 2023).

803 The ANNs are also able to make accurate predictions of the year for a given tem-
804 perature map using the climate model large ensemble data, and they are skillful before
805 the temperature response to greenhouse gas forcing overwhelms later in the 20th cen-
806 tury. This suggests that the ANN is still able to leverage patterns of temperature indi-
807 cators for distinguishing these temperature maps. We find that this is related to the higher
808 spatial resolution of the training data particularly in the vicinity of complex topogra-
809 phy. We perceive that this is not simply related to more available data samples, as we
810 do not find any skill improvement when training on additional individual ensemble mem-
811 bers. Rather it is more likely this is related to the ANN learning information from the
812 high-resolution grid and thus the ability for a climate model to represent finer temper-
813 ature structures and gradients.

814 To consider this point further, these results indicate that there is considerable po-
815 tential to use the machine learning methods described in this study to detect the emer-
816 gence of climate signals much earlier than with conventional methods, but the ability
817 to realize this potential is likely hindered by climate model error or biases. For exam-
818 ple, this could be analogous to the difference between potential predictability and fore-
819 cast skill in climate prediction studies. If you calculated the ToE in the same way de-
820 scribed here but using individual ensemble members as truth, you could define a “po-
821 tential ToE” that would be much earlier than the actual ToE (again, analogous to pre-
822 dictability). The difference between the actual and potential ToE likely would reflect er-
823 rors in the simulation of certain physical processes that the climate model deems impor-
824 tant for distinguishing forced changes from internal variability. The ability to realize that
825 potential ToE depends, however, on whether those processes are realistic and if the gap
826 between simulated and observed changes can be narrowed. In this work, the gap also seems
827 to be related to GCM resolution, as the earliest potential ToE generally occurs in the
828 higher-resolution simulations. Similarly, we are also limited by the ability of most cli-
829 mate models to accurately simulate the temperature variability realized over the 20th
830 century across the CONUS (i.e., the warming hole spatial pattern) (Eischeid et al., sub-
831 mitted).

832 Moving forward, our findings have several potential broader implications for future
833 work related to using machine learning methods on climate science applications. While
834 many XAI applications have regridded to coarser inputs because of lower computational
835 cost, this may come at the expensive of better machine learning model performance that
836 potentially could be achieved if using high-resolution data. For example, it could be in-
837 teresting to further explore this effect for applications of machine learning in subseasonal
838 to decadal prediction (Merryfield et al., 2020; Meehl et al., 2021), where neural networks
839 may be able to derive more information from features such as simulated mesoscale eddy
840 activity. But this remains an active area of research even for the climate model devel-
841 opment community (Hewitt et al., 2017; Scaife et al., 2019). It could also be interest-
842 ing to leverage these XAI tools for diagnosing biases in GCMs, such as the utility briefly
843 explored here for identifying an unexpected temperature response in western Colorado
844 which may be related to the land surface forcing fields.

845 However, having a greater number of input samples (e.g., higher resolution input
846 map) can also raise the risk of statistical overfitting. It also tends to result in lower in-

847 interpretability for understanding the machine learning models, even after applying XAI
 848 attribution methods (Samek et al., 2019; Barnes et al., 2020; Toms et al., 2020). There
 849 are thus tradeoffs to balance in all of these machine learning design choices. Despite our
 850 observational ToE estimates, which are found to be robust across a range of ANN ex-
 851 periments and training data sets, we propose that it would be helpful for more work to
 852 investigate the sensitivity of machine learning model skill to variations in input data across
 853 a variety of climate science applications.

854 Open Research

855 Atmospheric reanalysis data are openly available for ERA5 ([https://doi.org/](https://doi.org/10.24381/cds.6860a573)
 856 [10.24381/cds.6860a573](https://doi.org/10.24381/cds.6860a573)), which is supported by the Copernicus Climate Change Ser-
 857 vice (C3S; Thépaut et al., 2018) Climate Data Store (CDS). Twentieth Century Reanal-
 858 ysis Project version 3 (20CRv3) data are provided by the NOAA/OAR/ESRL PSL, Boul-
 859 der, Colorado, USA (https://psl.noaa.gov/data/gridded/data.20thC_ReanV3.html).
 860 Monthly U.S. Climate Gridded Dataset (NClimGrid; version 1) data are provided by NOAA/NCEI
 861 (<https://doi.org/10.7289/V5SX6B56>). Climate model large ensembles used in this study
 862 are available from the Climate Data Gateway at NCAR ([https://www.earthsystemgrid](https://www.earthsystemgrid.org/)
 863 [.org/](https://www.earthsystemgrid.org/)), Earth System Grid Federation (<https://esgf-node.llnl.gov/search/cmip6/>),
 864 Facility for Weather and Climate Assessments (FACTS; Murray et al., 2020) ([https://](https://www.psl.noaa.gov/repository/factsdocs)
 865 www.psl.noaa.gov/repository/factsdocs), and GFDL’s data portal ([https://nomads](https://nomads.gfdl.noaa.gov/)
 866 [.gfdl.noaa.gov/](https://nomads.gfdl.noaa.gov/)). References for all data sets are provided throughout the main text.

867 Data preprocessing was completed using NCO v5.0.1 (Zender, 2008), CDO v1.9.10
 868 (Schulzweida, 2019), and NCL v6.6.2 (NCAR, 2019). Computer code for the exploratory
 869 data analysis and machine learning architecture is available at [https://github.com/](https://github.com/zmlabe/TOE_TMIN-TMAX)
 870 [zmlabe/TOE_TMIN-TMAX](https://github.com/zmlabe/TOE_TMIN-TMAX). (*Reviewers, please note that this GitHub URL will transition*
 871 *to an archived-DOI repository at Zenodo if this paper is considered for publication*). Python
 872 v3.9.13 (Rossum & Drake, 2009) packages used for the main results of this work include
 873 Numpy v1.22.4 (Harris et al., 2020), SciPy v1.8.1 (Virtanen et al., 2020), Scikit-learn v1.1.1
 874 (Pedregosa et al., 2011), TensorFlow v2.7.0 (Abadi et al., 2016), and iNNvestigate v2.0.2
 875 (Alber et al., 2019). Matplotlib v3.5.2 (Hunter, 2007) and Basemap v1.3.6 (*Basemap*,
 876 2022) were used for plotting figures. Colormaps were provided by CMasher v1.6.3 (van der
 877 Velden, 2020), cmoclean v2.0 (Thyng et al., 2016), Palettable’s cubehelix v3.3.0 (Green,
 878 2011), and Scientific v7.0.0 (Crameri, 2018; Crameri et al., 2020).

879 Conflict of Interest

880 The Authors declare no conflicts of interest in regard to this study.

881 Acknowledgments

882 We thank Drs. Liwei Jia and Yanda Zhang for helpful reviews on a previous draft of this
 883 manuscript. We also thank Dr. Bor-Ting Jong for other helpful discussions on land sur-
 884 face feedbacks. Zachary Labe acknowledges support under award NA18OAR4320123 from
 885 the National Oceanic and Atmospheric Administration (NOAA), U.S. Department of Com-
 886 merce. This project is also supported by the NOAA Climate Program Office. Compu-
 887 tational resources for this study was through base funding of the Geophysical Fluid Dy-
 888 namics Laboratory (GFDL) provided by NOAA. The statements, findings, conclusions,
 889 and recommendations are those of the authors and do not necessarily reflect the views
 890 of NOAA, or the U.S. Department of Commerce.

891 References

- 892 Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., . . . Zheng, X.
 893 (2016). Tensorflow: A system for large-scale machine learning..
- 894 Abatzoglou, J. T., & Barbero, R. (2014). Observed and projected changes in abso-
 895 lute temperature records across the contiguous united states. *Geophysical Re-*
 896 *search Letters*, *41*. doi: 10.1002/2014GL061441
- 897 Abatzoglou, J. T., & Williams, A. P. (2016). Impact of anthropogenic cli-
 898 mate change on wildfire across western us forests. *Proceedings of the Na-*
 899 *tional Academy of Sciences of the United States of America*, *113*. doi:
 900 10.1073/pnas.1607171113
- 901 Adcroft, A., Anderson, W., Balaji, V., Blanton, C., Bushuk, M., Dufour, C. O., . . .
 902 Zhang, R. (2019, 10). The gfdl global ocean and sea ice model om4.0: Model
 903 description and simulation features. *Journal of Advances in Modeling Earth*
 904 *Systems*, *11*, 3167-3211. Retrieved from [https://agupubs.onlinelibrary](https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2019MS001726)
 905 [.wiley.com/doi/10.1029/2019MS001726](https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2019MS001726) doi: 10.1029/2019MS001726
- 906 Agarap, A. F. (2018, 3). Deep learning using rectified linear units (relu). *arXiv*. Re-
 907 trieved from <http://arxiv.org/abs/1803.08375>
- 908 Alber, M., Lapuschkin, S., Seegerer, P., Hägele, M., Schütt, K. T., Montavon, G., . . .
 909 Kindermans, P. J. (2019). Innvestigate neural networks! *Journal of Machine*

- 910 *Learning Research*, 20.
- 911 Alter, R. E., Douglas, H. C., Winter, J. M., & Eltahir, E. A. (2018). Twentieth
912 century regional climate change during the summer in the central united states
913 attributed to agricultural intensification. *Geophysical Research Letters*, 45.
914 doi: 10.1002/2017GL075604
- 915 a. Meehl, G., Arblaster, J. M., & Branstator, G. (2012, 9). Mechanisms con-
916 tributing to the warming hole and the consequent u.s. east–west differen-
917 tial of heat extremes. *Journal of Climate*, 25, 6394-6408. Retrieved from
918 <http://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-11-00655.1> doi:
919 10.1175/JCLI-D-11-00655.1
- 920 Anderson, C., & Stock, J. (2022, 12). An interpretable model of climate change
921 using correlative learning.. Retrieved from [https://arxiv.org/abs/](https://arxiv.org/abs/2212.04478v1)
922 2212.04478v1
- 923 Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K. R., & Samek, W.
924 (2015, 7). On pixel-wise explanations for non-linear classifier decisions by
925 layer-wise relevance propagation. *PLoS ONE*, 10, e0130140. Retrieved from
926 <http://www.hfsp.org/>, doi: 10.1371/journal.pone.0130140
- 927 Banerjee, A., Polvani, L. M., & Fyfe, J. C. (2017). The united states “warming
928 hole”: Quantifying the forced aerosol response given large internal variability.
929 *Geophysical Research Letters*, 44. doi: 10.1002/2016GL071567
- 930 Barnes, E. A., Anderson, C., & Ebert-Uphoff, I. (2018, 11). An ai approach to de-
931 termining the time of emergence of climate change. In (p. 19-22). NCAR. Re-
932 trieved from <https://opensky.ucar.edu/islandora/object/technotes:571>
933 doi: <http://dx.doi.org/10.5065/D6BZ64XQ>
- 934 Barnes, E. A., Hurrell, J. W., Ebert-Uphoff, I., Anderson, C., & Anderson, D. (2019,
935 11). Viewing forced climate patterns through an ai lens. *Geophysical Research*
936 *Letters*, 46, 13389-13398. Retrieved from [https://onlinelibrary.wiley](https://onlinelibrary.wiley.com/doi/abs/10.1029/2019GL084944)
937 [.com/doi/abs/10.1029/2019GL084944](https://onlinelibrary.wiley.com/doi/abs/10.1029/2019GL084944) doi: 10.1029/2019GL084944
- 938 Barnes, E. A., Toms, B., Hurrell, J. W., Ebert-Uphoff, I., Anderson, C., & An-
939 derson, D. (2020, 9). Indicator patterns of forced change learned by an
940 artificial neural network. *Journal of Advances in Modeling Earth Systems*,
941 12. Retrieved from [https://onlinelibrary.wiley.com/doi/10.1029/](https://onlinelibrary.wiley.com/doi/10.1029/2020MS002195)
942 2020MS002195 doi: 10.1029/2020MS002195

- 943 *Basemap*. (2022, 10). Retrieved from <https://github.com/matplotlib/basemap>
- 944 Bevacqua, E., Suarez-Gutierrez, L., Jézéquel, A., Lehner, F., Vac, M., Yiou, P.,
945 & Zscheischler, J. (2023, 4). Advancing research on compound weather and
946 climate events via large ensemble model simulations. *Nature Communications*
947 *2023 14:1, 14*, 1-16. Retrieved from [https://www.nature.com/articles/](https://www.nature.com/articles/s41467-023-37847-5)
948 [s41467-023-37847-5](https://www.nature.com/articles/s41467-023-37847-5) doi: 10.1038/s41467-023-37847-5
- 949 Bevan, J. M., & Kendall, M. G. (1971). Rank correlation methods. *The Statistician*.
950 doi: 10.2307/2986801
- 951 Bommer, P., Kretschmer, M., Hedström, A., Bareeva, D., & Höhne, M. M. C. (2023,
952 3). Finding the right xai method – a guide for the evaluation and ranking of
953 explainable ai methods in climate science. *arXiv*. Retrieved from [https://](https://arxiv.org/abs/2303.00652v1)
954 arxiv.org/abs/2303.00652v1
- 955 Boukabara, S.-A., Krasnopolsky, V., Penny, S. G., Stewart, J. Q., McGovern, A.,
956 Hall, D., . . . Hoffman, R. N. (2021, 5). Outlook for exploiting artificial
957 intelligence in the earth and environmental sciences. *Bulletin of the Amer-*
958 *ican Meteorological Society*, *102*, E1016-E1032. Retrieved from [https://](https://journals.ametsoc.org/view/journals/bams/102/5/BAMS-D-20-0031.1.xml)
959 journals.ametsoc.org/view/journals/bams/102/5/BAMS-D-20-0031.1.xml
960 doi: 10.1175/BAMS-D-20-0031.1
- 961 Burgess, M. G., Ritchie, J., Shapland, J., & Pielke, R. (2020, 12). Ipc baseline sce-
962 narios have over-projected co2 emissions and economic growth. *Environmental*
963 *Research Letters*, *16*, 014016. Retrieved from [https://iopscience.iop.org/](https://iopscience.iop.org/article/10.1088/1748-9326/abcdd2)
964 [article/10.1088/1748-9326/abcdd2](https://iopscience.iop.org/article/10.1088/1748-9326/abcdd2) doi: 10.1088/1748-9326/ABCDD2
- 965 Chantry, M., Christensen, H., Dueben, P., & Palmer, T. (2021, 4). Opportuni-
966 ties and challenges for machine learning in weather and climate modelling:
967 hard, medium and soft ai. *Philosophical Transactions of the Royal Society A*,
968 *379*. Retrieved from [https://royalsocietypublishing.org/doi/10.1098/](https://royalsocietypublishing.org/doi/10.1098/rsta.2020.0083)
969 [rsta.2020.0083](https://royalsocietypublishing.org/doi/10.1098/rsta.2020.0083) doi: 10.1098/RSTA.2020.0083
- 970 Chase, R. J., Harrison, D. R., Burke, A., Lackmann, G. M., & McGovern, A. (2022,
971 8). A machine learning tutorial for operational meteorology. part i: Tra-
972 ditional machine learning. *Weather and Forecasting*, *37*, 1509-1529. Re-
973 trieved from [https://journals.ametsoc.org/view/journals/wefo/37/8/](https://journals.ametsoc.org/view/journals/wefo/37/8/WAF-D-22-0070.1.xml)
974 [WAF-D-22-0070.1.xml](https://journals.ametsoc.org/view/journals/wefo/37/8/WAF-D-22-0070.1.xml) doi: 10.1175/WAF-D-22-0070.1

- 975 Chase, R. J., Harrison, D. R., Lackmann, G., & McGovern, A. (2022, 10). A
 976 machine learning tutorial for operational meteorology, part ii: Neural net-
 977 works and deep learning. *arXiv*. Retrieved from [https://arxiv.org/abs/](https://arxiv.org/abs/2211.00147v1)
 978 [2211.00147v1](https://arxiv.org/abs/2211.00147v1) doi: 10.48550/arxiv.2211.00147
- 979 Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin,
 980 X., ... Worley, S. J. (2011, 1). The twentieth century reanalysis project.
 981 *Quarterly Journal of the Royal Meteorological Society*, *137*, 1-28. doi:
 982 10.1002/qj.776
- 983 Cowan, T., Hegerl, G. C., Colfescu, I., Bollasina, M., Purich, A., & Boschat,
 984 G. (2017). Factors contributing to record-breaking heat waves over the
 985 great plains during the 1930s dust bowl. *Journal of Climate*, *30*. doi:
 986 10.1175/JCLI-D-16-0436.1
- 987 Cowan, T., Undorf, S., Hegerl, G. C., Harrington, L. J., & Otto, F. E. (2020,
 988 5). Present-day greenhouse gases could cause more frequent and longer
 989 dust bowl heatwaves. *Nature Climate Change*, *10*, 505-510. Retrieved
 990 from <https://www.nature.com/articles/s41558-020-0771-7> doi:
 991 10.1038/s41558-020-0771-7
- 992 Crameri, F. (2018, 1). Scientific colour maps. *Zenodo*. Retrieved from [https://](https://zenodo.org/record/4153113)
 993 zenodo.org/record/4153113 doi: 10.5281/ZENODO.4153113
- 994 Crameri, F., Shephard, G. E., & Heron, P. J. (2020, 12). The misuse of colour
 995 in science communication. *Nature Communications*, *11*, 1-10. Re-
 996 trieved from <https://doi.org/10.1038/s41467-020-19160-7> doi:
 997 10.1038/s41467-020-19160-7
- 998 Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K.,
 999 Edwards, J., ... Strand, W. G. (2020, 2). The community earth system
 1000 model version 2 (cesm2). *Journal of Advances in Modeling Earth Systems*, *12*,
 1001 e2019MS001916. Retrieved from [https://agupubs.onlinelibrary.wiley](https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2019MS001916)
 1002 [.com/doi/10.1029/2019MS001916](https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2019MS001916) doi: 10.1029/2019MS001916
- 1003 Delworth, T. L., Cooke, W. F., Adcroft, A., Bushuk, M., Chen, J.-H., Dunne,
 1004 K. A., ... Zhao, M. (2020, 3). Spear: The next generation gfdl modeling
 1005 system for seasonal to multidecadal prediction and projection. *Journal of*
 1006 *Advances in Modeling Earth Systems*, *12*, e2019MS001895. Retrieved from
 1007 <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2019MS001895>

- 1008 doi: 10.1029/2019MS001895
- 1009 Delworth, T. L., Cooke, W. F., Naik, V., Paynter, D., & Zhang, L. (2022, 8). A
 1010 weakened amoc may prolong greenhouse gas-induced mediterranean dry-
 1011 ing even with significant and rapid climate change mitigation. *Proceed-*
 1012 *ings of the National Academy of Sciences of the United States of Amer-*
 1013 *ica*, 119, e2116655119. Retrieved from [https://www.pnas.org/doi/abs/](https://www.pnas.org/doi/abs/10.1073/pnas.2116655119)
 1014 [10.1073/pnas.2116655119](https://www.pnas.org/doi/abs/10.1073/pnas.2116655119) doi: 10.1073/PNAS.2116655119/SUPPL_FILE/
 1015 PNAS.2116655119.SAPP.PDF
- 1016 Delworth, T. L., Rosati, A., Anderson, W., Adcroft, A. J., Balaji, V., Benson, R., ...
 1017 Zhang, R. (2012, 4). Simulated climate and climate change in the gfdl cm2.5
 1018 high-resolution coupled climate model. *Journal of Climate*, 25, 2755-2781.
 1019 Retrieved from [https://journals.ametsoc.org/view/journals/clim/25/8/](https://journals.ametsoc.org/view/journals/clim/25/8/jcli-d-11-00316.1.xml)
 1020 [jcli-d-11-00316.1.xml](https://journals.ametsoc.org/view/journals/clim/25/8/jcli-d-11-00316.1.xml) doi: 10.1175/JCLI-D-11-00316.1
- 1021 Deser, C. (2020, 11). Certain uncertainty: The role of internal climate variability
 1022 in projections of regional climate change and risk management. *Earth's Fu-*
 1023 *ture*. Retrieved from [https://onlinelibrary.wiley.com/doi/10.1029/](https://onlinelibrary.wiley.com/doi/10.1029/2020EF001854)
 1024 [2020EF001854](https://onlinelibrary.wiley.com/doi/10.1029/2020EF001854) doi: 10.1029/2020EF001854
- 1025 Deser, C., Lehner, F., Rodgers, K. B., Ault, T., Delworth, T. L., DiNezio, P. N.,
 1026 ... Ting, M. (2020, 3). Insights from earth system model initial-condition
 1027 large ensembles and future prospects. *Nature Climate Change*, 1-10. Re-
 1028 trieved from <http://www.nature.com/articles/s41558-020-0731-2> doi:
 1029 [10.1038/s41558-020-0731-2](https://doi.org/10.1038/s41558-020-0731-2)
- 1030 Deser, C., Phillips, A., Bourdette, V., & Teng, H. (2012, 2). Uncertainty in climate
 1031 change projections: the role of internal variability. *Climate Dynamics*, 38, 527-
 1032 546. Retrieved from [http://link.springer.com/10.1007/s00382-010-0977](http://link.springer.com/10.1007/s00382-010-0977-x)
 1033 [-x](http://link.springer.com/10.1007/s00382-010-0977-x) doi: 10.1007/s00382-010-0977-x
- 1034 Deser, C., Terray, L., & Phillips, A. S. (2016). Forced and internal compo-
 1035 nents of winter air temperature trends over north america during the past
 1036 50 years: Mechanisms and implications. *Journal of Climate*, 29. doi:
 1037 [10.1175/JCLI-D-15-0304.1](https://doi.org/10.1175/JCLI-D-15-0304.1)
- 1038 Diffenbaugh, N. S., & Barnes, E. A. (2023, 2). Data-driven predictions of the
 1039 time remaining until critical global warming thresholds are reached. *Pro-*
 1040 *ceedings of the National Academy of Sciences of the United States of Amer-*

- 1041 *ica*, 120, e2207183120. Retrieved from <https://www.pnas.org/doi/abs/>
 1042 10.1073/pnas.2207183120 doi: 10.1073/PNAS.2207183120/SUPPL_FILE/
 1043 PNAS.2207183120.SAPP.PDF
- 1044 Diffenbaugh, N. S., Swain, D. L., Touma, D., & Lubchenco, J. (2015). Anthro-
 1045 pogenic warming has increased drought risk in california. *Proceedings of the*
 1046 *National Academy of Sciences of the United States of America*, 112. doi:
 1047 10.1073/pnas.1422385112
- 1048 Donat, M. G., King, A. D., Overpeck, J. T., Alexander, L. V., Durre, I., & Karoly,
 1049 D. J. (2016, 1). Extraordinary heat during the 1930s us dust bowl and asso-
 1050 ciated large-scale conditions. *Climate Dynamics*, 46, 413-426. Retrieved from
 1051 <https://link.springer.com/article/10.1007/s00382-015-2590-5> doi:
 1052 10.1007/S00382-015-2590-5/FIGURES/7
- 1053 Durre, I., Menne, M. J., Gleason, B. E., Houston, T. G., & Vose, R. S. (2010,
 1054 8). Comprehensive automated quality assurance of daily surface observa-
 1055 tions. *Journal of Applied Meteorology and Climatology*, 49, 1615-1633. Re-
 1056 trieved from [https://journals.ametsoc.org/view/journals/apme/49/8/](https://journals.ametsoc.org/view/journals/apme/49/8/2010jamc2375.1.xml)
 1057 2010jamc2375.1.xml doi: 10.1175/2010JAMC2375.1
- 1058 Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., &
 1059 Taylor, K. E. (2016, 5). Overview of the coupled model intercomparison
 1060 project phase 6 (cmip6) experimental design and organization. *Geoscientific*
 1061 *Model Development*, 9, 1937-1958. doi: 10.5194/gmd-9-1937-2016
- 1062 Ferguson, C. R., & Villarini, G. (2012). Detecting inhomogeneities in the twen-
 1063 tieth century reanalysis over the central united states. *Journal of Geophysical*
 1064 *Research Atmospheres*, 117. doi: 10.1029/2011JD016988
- 1065 Ferguson, C. R., & Villarini, G. (2014, 11). An evaluation of the statistical ho-
 1066 mogeneity of the twentieth century reanalysis. *Climate Dynamics*, 42, 2841-
 1067 2866. Retrieved from [https://link.springer.com/article/10.1007/s00382](https://link.springer.com/article/10.1007/s00382-013-1996-1)
 1068 -013-1996-1 doi: 10.1007/S00382-013-1996-1/FIGURES/16
- 1069 Fischer, E. M., Beyerle, U., & Knutti, R. (2013, 11). Robust spatially aggregated
 1070 projections of climate extremes. *Nature Climate Change*, 3, 1033-1038.
 1071 Retrieved from <https://www.nature.com/articles/nclimate2051> doi:
 1072 10.1038/nclimate2051

- 1073 Fischer, E. M., & Knutti, R. (2014, 1). Detection of spatially aggregated changes
 1074 in temperature and precipitation extremes. *Geophysical Research Letters*, *41*,
 1075 547-554. Retrieved from [https://agupubs.onlinelibrary.wiley.com/doi/](https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2013GL058499)
 1076 [10.1002/2013GL058499](https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2013GL058499) doi: 10.1002/2013GL058499
- 1077 François, B., & Vrac, M. (2023, 1). Time of emergence of compound events: con-
 1078 tribution of univariate and dependence properties. *Natural Hazards and Earth*
 1079 *System Sciences*, *23*, 21-44. doi: 10.5194/NHESS-23-21-2023
- 1080 Friedman, J. H. (2012, 7). Fast sparse regression and classification. *International*
 1081 *Journal of Forecasting*, *28*, 722-738. doi: 10.1016/j.ijforecast.2012.05.001
- 1082 Gillespie, I., Haimberger, L., Compo, G. P., & Thorne, P. W. (2023, 2). Assessing
 1083 homogeneity of land surface air temperature observations using sparse-input
 1084 reanalyses. *International Journal of Climatology*, *43*, 736-760. Retrieved from
 1085 <https://rmets.onlinelibrary.wiley.com/doi/10.1002/joc.7822> doi:
 1086 [10.1002/JOC.7822](https://rmets.onlinelibrary.wiley.com/doi/10.1002/JOC.7822)
- 1087 Giorgi, F., & Bi, X. (2009, 3). Time of emergence (toe) of ghg-forced precipita-
 1088 tion change hot-spots. *Geophysical Research Letters*, *36*. Retrieved from
 1089 <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2009GL037593>
 1090 doi: 10.1029/2009GL037593
- 1091 Gnanadesikan, A., Dixon, K. W., Griffies, S. M., Balaji, V., Barreiro, M., Beesley,
 1092 J. A., ... Dunne, J. P. (2006, 3). Gfdl's cm2 global coupled climate models.
 1093 part ii: The baseline ocean simulation. *Journal of Climate*, *19*, 675-697. Re-
 1094 trieved from [https://journals.ametsoc.org/view/journals/clim/19/5/](https://journals.ametsoc.org/view/journals/clim/19/5/jcli3630.1.xml)
 1095 [jcli3630.1.xml](https://journals.ametsoc.org/view/journals/clim/19/5/jcli3630.1.xml) doi: 10.1175/JCLI3630.1
- 1096 Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*.
- 1097 Green, D. A. (2011). A colour scheme for the display of astronomical intensity im-
 1098 ages. *Bulletin of the Astronomical Society of India*, *39*.
- 1099 Grotjahn, R., & Huynh, J. (2018). Contiguous us summer maximum temperature
 1100 and heat stress trends in cru and noaa climate division data plus comparisons
 1101 to reanalyses. *Scientific Reports*, *8*. doi: 10.1038/s41598-018-29286-w
- 1102 Hansen, J., Ruedy, R., Sato, M., Imhoff, M., Lawrence, W., Easterling, D., ...
 1103 Karl, T. (2001). A closer look at united states and global surface tem-
 1104 perature change. *Journal of Geophysical Research Atmospheres*, *106*. doi:
 1105 [10.1029/2001JD000354](https://doi.org/10.1029/2001JD000354)

- 1106 Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P.,
 1107 Cournapeau, D., ... Oliphant, T. E. (2020, 9). Array programming with
 1108 numpy. *Nature*, *585*, 357. Retrieved from [https://doi.org/10.1038/
 1109 s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2) doi: 10.1038/s41586-020-2649-2
- 1110 Hausfather, Z., & Peters, G. P. (2020). *Rcp8.5 is a problematic scenario for near-*
 1111 *term emissions* (Vol. 117). doi: 10.1073/pnas.2017124117
- 1112 Hawkins, E., Anderson, B., Diffenbaugh, N., Mahlstein, I., Betts, R., Hegerl, G.,
 1113 ... Vecchi, G. (2014, 7). Uncertainties in the timing of unprecedented
 1114 climates. *Nature* *2014 511:7507*, *511*, E3-E5. Retrieved from [https://
 1115 www.nature.com/articles/nature13523](https://www.nature.com/articles/nature13523) doi: 10.1038/nature13523
- 1116 Hawkins, E., Frame, D., Harrington, L., Joshi, M., King, A., Rojas, M., & Sutton,
 1117 R. (2020, 3). Observed emergence of the climate change signal: From the
 1118 familiar to the unknown. *Geophysical Research Letters*, *47*. Retrieved from
 1119 <https://onlinelibrary.wiley.com/doi/abs/10.1029/2019GL086259> doi:
 1120 10.1029/2019GL086259
- 1121 Hawkins, E., & Sutton, R. (2012). Time of emergence of climate signals. *Geophysical*
 1122 *Research Letters*. doi: 10.1029/2011GL050087
- 1123 Held, I. M., Guo, H., Adcroft, A., Dunne, J. P., Horowitz, L. W., Krasting, J., ...
 1124 Zadeh, N. (2019, 11). Structure and performance of gfdl's cm4.0 climate
 1125 model. *Journal of Advances in Modeling Earth Systems*, *11*, 3691-3727. Re-
 1126 trieved from [https://agupubs.onlinelibrary.wiley.com/doi/10.1029/
 1127 2019MS001829](https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2019MS001829) doi: 10.1029/2019MS001829
- 1128 Henson, S. A., Beaulieu, C., Ilyina, T., John, J. G., Long, M., Séférian, R., ...
 1129 Sarmiento, J. L. (2017, 3). Rapid emergence of climate change in environ-
 1130 mental drivers of marine ecosystems. *Nature Communications* *2017 8:1*, *8*,
 1131 1-9. Retrieved from <https://www.nature.com/articles/ncomms14682> doi:
 1132 10.1038/ncomms14682
- 1133 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J.,
 1134 ... Thépaut, J.-N. (2020, 5). The era5 global reanalysis. *Quarterly Journal*
 1135 *of the Royal Meteorological Society*. Retrieved from [https://onlinelibrary
 1136 .wiley.com/doi/abs/10.1002/qj.3803](https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.3803) doi: 10.1002/qj.3803
- 1137 Hewitt, H. T., Bell, M. J., Chassignet, E. P., Czaja, A., Ferreira, D., Griffies, S. M.,
 1138 ... Roberts, M. J. (2017). *Will high-resolution global ocean models bene-*

- 1139 *fit coupled predictions on short-range to climate timescales?* (Vol. 120). doi:
1140 10.1016/j.ocemod.2017.11.002
- 1141 Hunter, J. D. (2007, 5). Matplotlib: A 2d graphics environment. *Computing in Sci-*
1142 *ence and Engineering*, 9, 99-104. doi: 10.1109/MCSE.2007.55
- 1143 Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J.,
1144 ... Marshall, S. (2013). The community earth system model: A framework
1145 for collaborative research. *Bulletin of the American Meteorological Society*, 94.
1146 doi: 10.1175/BAMS-D-12-00121.1
- 1147 Ignjacevic, P., Estrada, F., & Botzen, W. J. (2021, 7). Time of emergence of eco-
1148 nomic impacts of climate change. *Environmental Research Letters*, 16, 074039.
1149 Retrieved from [https://iopscience.iop.org/article/10.1088/1748-9326/](https://iopscience.iop.org/article/10.1088/1748-9326/ac0d7a/meta)
1150 [ac0d7a/meta](https://iopscience.iop.org/article/10.1088/1748-9326/ac0d7a/meta) doi: 10.1088/1748-9326/AC0D7A
- 1151 IPCC, Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., ... B.,
1152 Z. (2021). *Climate change 2021: The physical science basis. contribution of*
1153 *working group i to the sixth assessment report of the intergovernmental panel*
1154 *on climate change.*
- 1155 Irrgang, C., Boers, N., Sonnewald, M., Barnes, E. A., Kadow, C., Staneva, J., &
1156 Saynisch-Wagner, J. (2021, 8). Towards neural earth system modelling by
1157 integrating artificial intelligence in earth system science. *Nature Machine In-*
1158 *telligence*, 3, 667-674. Retrieved from [https://www.nature.com/articles/](https://www.nature.com/articles/s42256-021-00374-3)
1159 [s42256-021-00374-3](https://www.nature.com/articles/s42256-021-00374-3) doi: 10.1038/s42256-021-00374-3
- 1160 Jia, L., Delworth, T. L., Kapnick, S., Yang, X., JOHNSON, N. C., Cooke, W.,
1161 ... Tseng, K. C. (2022, 7). Skillful seasonal prediction of north ameri-
1162 can summertime heat extremes. *Journal of Climate*, 35, 4331-4345. Re-
1163 trieved from [https://journals.ametsoc.org/view/journals/clim/35/13/](https://journals.ametsoc.org/view/journals/clim/35/13/JCLI-D-21-0364.1.xml)
1164 [JCLI-D-21-0364.1.xml](https://journals.ametsoc.org/view/journals/clim/35/13/JCLI-D-21-0364.1.xml) doi: 10.1175/JCLI-D-21-0364.1
- 1165 Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., ... Vertenstein,
1166 M. (2015, 8). The community earth system model (cesm) large ensemble
1167 project: A community resource for studying climate change in the presence of
1168 internal climate variability. *Bulletin of the American Meteorological Society*,
1169 96, 1333-1349. Retrieved from [http://journals.ametsoc.org/doi/10.1175/](http://journals.ametsoc.org/doi/10.1175/BAMS-D-13-00255.1)
1170 [BAMS-D-13-00255.1](http://journals.ametsoc.org/doi/10.1175/BAMS-D-13-00255.1) doi: 10.1175/BAMS-D-13-00255.1

- 1171 King, A. D., Donat, M. G., Fischer, E. M., Hawkins, E., Alexander, L. V., Karoly,
 1172 D. J., ... Perkins, S. E. (2015, 9). The timing of anthropogenic emergence
 1173 in simulated climate extremes. *Environmental Research Letters*, *10*, 094015.
 1174 Retrieved from [https://iopscience.iop.org/article/10.1088/1748-9326/](https://iopscience.iop.org/article/10.1088/1748-9326/10/9/094015/meta)
 1175 [10/9/094015/meta](https://iopscience.iop.org/article/10.1088/1748-9326/10/9/094015/meta) doi: 10.1088/1748-9326/10/9/094015
- 1176 Kriegler, E., Bauer, N., Popp, A., Humpenöder, F., Leimbach, M., Strefler, J., ...
 1177 Edenhofer, O. (2017, 1). Fossil-fueled development (ssp5): An energy and
 1178 resource intensive scenario for the 21st century. *Global Environmental Change*,
 1179 *42*, 297-315. doi: 10.1016/J.GLOENVCHA.2016.05.015
- 1180 Kunkel, K. E., Liang, X. Z., Zhu, J., & Lin, Y. (2006). Can cgcms simulate the
 1181 twentieth-century "warning hole" in the central united states? *Journal of Cli-*
 1182 *mate*, *19*. doi: 10.1175/JCLI3848.1
- 1183 Labe, Z. M., & Barnes, E. A. (2021). Detecting climate signals using ex-
 1184 plainable ai with single-forcing large ensembles. *Journal of Advances in*
 1185 *Modeling Earth Systems*, *13*, e2021MS002464. Retrieved from [https://](https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021MS002464)
 1186 agupubs.onlinelibrary.wiley.com/doi/10.1029/2021MS002464 doi:
 1187 [10.1029/2021MS002464](https://doi.org/10.1029/2021MS002464)
- 1188 Labe, Z. M., & Barnes, E. A. (2022, 7). Comparison of climate model large en-
 1189 sembles with observations in the arctic using simple neural networks. *Earth*
 1190 *and Space Science*, *9*, e2022EA002348. Retrieved from [https://doi.org/](https://doi.org/10.1029/2022EA002348)
 1191 [10.1029/2022EA002348](https://doi.org/10.1029/2022EA002348) doi: 10.1029/2022EA002348
- 1192 Labe, Z. M., Barnes, E. A., & Hurrell, J. W. (2023, 3). Identifying the regional
 1193 emergence of climate patterns in the arise-sai-1.5 simulations. *Environmental*
 1194 *Research Letters*, *18*, 1-12. Retrieved from [https://iopscience.iop.org/](https://iopscience.iop.org/article/10.1088/1748-9326/acc81a)
 1195 [article/10.1088/1748-9326/acc81a](https://iopscience.iop.org/article/10.1088/1748-9326/acc81a) doi: 10.1088/1748-9326/ACC81A
- 1196 Lawrence, J., Blackett, P., & Cradock-Henry, N. A. (2020, 1). Cascading climate
 1197 change impacts and implications. *Climate Risk Management*, *29*, 100234. doi:
 1198 [10.1016/J.CRM.2020.100234](https://doi.org/10.1016/J.CRM.2020.100234)
- 1199 Lecun, Y., Bengio, Y., & Hinton, G. (2015, 5). *Deep learning* (Vol. 521). Na-
 1200 ture Publishing Group. Retrieved from [https://www.nature.com/articles/](https://www.nature.com/articles/nature14539)
 1201 [nature14539](https://www.nature.com/articles/nature14539) doi: 10.1038/nature14539
- 1202 Lehner, F., & Deser, C. (2023, 4). Origin, importance, and predictive limits
 1203 of internal climate variability. *Environmental Research: Climate*. Re-

- 1204 trieved from <https://iopscience.iop.org/article/10.1088/2752-5295/>
 1205 accf30<https://iopscience.iop.org/article/10.1088/2752-5295/accf30/>
 1206 meta doi: 10.1088/2752-5295/ACCF30
- 1207 Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E., Brunner, L., ...
 1208 Hawkins, E. (2020). Partitioning climate projection uncertainty with mul-
 1209 tiple large ensembles and cmip5/6. *Earth System Dynamics Discussions*, 1-28.
 1210 doi: 10.5194/esd-2019-93
- 1211 Lehner, F., Deser, C., & Terray, L. (2017). Toward a new estimate of "time of
 1212 emergence" of anthropogenic warming: Insights from dynamical adjustment
 1213 and a large initial-condition model ensemble. *Journal of Climate*, 30. doi:
 1214 10.1175/JCLI-D-16-0792.1
- 1215 Lei, L., & Whitaker, J. S. (2016, 7). A four-dimensional incremental analysis update
 1216 for the ensemble kalman filter. *Monthly Weather Review*, 144, 2605-2621. Re-
 1217 trieved from <https://journals.ametsoc.org/view/journals/mwre/144/7/>
 1218 mwr-d-15-0246.1.xml doi: 10.1175/MWR-D-15-0246.1
- 1219 Leibensperger, E. M., Mickley, L. J., Jacob, D. J., Chen, W. T., Seinfeld, J. H.,
 1220 Nenes, A., ... Rind, D. (2012). Climatic effects of 1950-2050 changes in us
 1221 anthropogenic aerosols-part 2: Climate response. *Atmospheric Chemistry and*
 1222 *Physics*, 12. doi: 10.5194/acp-12-3349-2012
- 1223 Lempert, R., Nakicenovic, N., Sarewitz, D., & Schlesinger, M. (2004, 7). Charac-
 1224 terizing climate-change uncertainties for decision-makers. an editorial essay.
 1225 *Climatic Change*, 65, 1-9. Retrieved from [https://link.springer.com/](https://link.springer.com/article/10.1023/B:CLIM.0000037561.75281.b3)
 1226 article/10.1023/B:CLIM.0000037561.75281.b3 doi: 10.1023/B:
 1227 CLIM.0000037561.75281.B3/METRICS
- 1228 Madakumbura, G. D., Thackeray, C. W., Norris, J., Goldenson, N., & Hall, A.
 1229 (2021, 7). Anthropogenic influence on extreme precipitation over global
 1230 land areas seen in multiple observational datasets. *Nature Communications*
 1231 2021 12:1, 12, 1-9. Retrieved from [https://www.nature.com/articles/](https://www.nature.com/articles/s41467-021-24262-x)
 1232 s41467-021-24262-x doi: 10.1038/s41467-021-24262-x
- 1233 Maher, N., Lehner, F., & Marotzke, J. (2020, 5). Quantifying the role of internal
 1234 variability in the temperature we expect to observe in the coming decades. *En-*
 1235 *vironmental Research Letters*, 15, 054014. Retrieved from [https://doi.org/](https://doi.org/10.1088/1748-9326/ab7d02)
 1236 10.1088/1748-9326/ab7d02 doi: 10.1088/1748-9326/ab7d02

- 1237 Mahlstein, I., Hegerl, G., & Solomon, S. (2012). Emerging local warming
1238 signals in observational data. *Geophysical Research Letters*, *39*. doi:
1239 10.1029/2012GL053952
- 1240 Mahony, C. R., & Cannon, A. J. (2018, 2). Wetter summers can intensify departures
1241 from natural variability in a warming climate. *Nature Communications*, *9*, 1-
1242 9. Retrieved from <https://www.nature.com/articles/s41467-018-03132-z>
1243 doi: 10.1038/s41467-018-03132-z
- 1244 Mamalakis, A., Barnes, E. A., & Ebert-Uphoff, I. (2023, 1). Carefully choose the
1245 baseline: Lessons learned from applying xai attribution methods for regression
1246 tasks in geoscience. *Artificial Intelligence for the Earth Systems*, *2*. Re-
1247 trieved from [https://journals.ametsoc.org/view/journals/aies/2/1/](https://journals.ametsoc.org/view/journals/aies/2/1/AIES-D-22-0058.1.xml)
1248 [AIES-D-22-0058.1.xml](https://journals.ametsoc.org/view/journals/aies/2/1/AIES-D-22-0058.1.xml) doi: 10.1175/AIES-D-22-0058.1
- 1249 Mamalakis, A., Ebert-Uphoff, I., & Barnes, E. A. (2022, 6). Neural network attribu-
1250 tion methods for problems in geoscience: A novel synthetic benchmark dataset.
1251 *Environmental Data Science*, *1*, e8. Retrieved from [https://www.cambridge](https://www.cambridge.org/core/journals/environmental-data-science/article/neural-network-attribution-methods-for-problems-in-geoscience-a-novel-synthetic-benchmark-dataset/DDA562FC7B9A2B30710582861920860E)
1252 [.org/core/journals/environmental-data-science/article/neural](https://www.cambridge.org/core/journals/environmental-data-science/article/neural-network-attribution-methods-for-problems-in-geoscience-a-novel-synthetic-benchmark-dataset/DDA562FC7B9A2B30710582861920860E)
1253 [-network-attribution-methods-for-problems-in-geoscience-a-novel](https://www.cambridge.org/core/journals/environmental-data-science/article/neural-network-attribution-methods-for-problems-in-geoscience-a-novel-synthetic-benchmark-dataset/DDA562FC7B9A2B30710582861920860E)
1254 [-synthetic-benchmark-dataset/DDA562FC7B9A2B30710582861920860E](https://www.cambridge.org/core/journals/environmental-data-science/article/neural-network-attribution-methods-for-problems-in-geoscience-a-novel-synthetic-benchmark-dataset/DDA562FC7B9A2B30710582861920860E) doi:
1255 10.1017/EDS.2022.7
- 1256 Mankin, J. S., Lehner, F., Coats, S., & McKinnon, K. A. (2020, 10). The value of
1257 initial condition large ensembles to robust adaptation decision-making. *Earth's*
1258 *Future*, *8*. Retrieved from [https://onlinelibrary.wiley.com/doi/10.1029/](https://onlinelibrary.wiley.com/doi/10.1029/2020EF001610)
1259 [2020EF001610](https://onlinelibrary.wiley.com/doi/10.1029/2020EF001610) doi: 10.1029/2020EF001610
- 1260 Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica*. doi: 10
1261 .2307/1907187
- 1262 Mascioli, N. R., Previdi, M., Fiore, A. M., & Ting, M. (2017). Timing and season-
1263 ality of the united states 'warming hole'. *Environmental Research Letters*, *12*.
1264 doi: 10.1088/1748-9326/aa5ef4
- 1265 McKinnon, K. A., & Deser, C. (2018). Internal variability and regional climate
1266 trends in an observational large ensemble. *Journal of Climate*, *31*. doi: 10
1267 .1175/JCLI-D-17-0901.1
- 1268 Meehl, G. A., Arblaster, J. M., & Chung, C. T. (2015). Disappearance of the
1269 southeast u.s. "warming hole" with the late 1990s transition of the in-

- 1270 terdecadal pacific oscillation. *Geophysical Research Letters*, *42*. doi:
 1271 10.1002/2015GL064586
- 1272 Meehl, G. A., Richter, J. H., Teng, H., Capotondi, A., Cobb, K., Doblas-Reyes, F.,
 1273 ... Xie, S.-P. (2021, 4). Initialized earth system prediction from subseasonal
 1274 to decadal timescales. *Nature Reviews Earth and Environment*, *2*, 340-357.
 1275 Retrieved from <https://www.nature.com/articles/s43017-021-00155-x>
 1276 doi: 10.1038/s43017-021-00155-x
- 1277 Meehl, G. A., Tebaldi, C., & Adams-Smith, D. (2016). Us daily temperature records
 1278 past, present, and future. *Proceedings of the National Academy of Sciences of*
 1279 *the United States of America*, *113*. doi: 10.1073/pnas.1606117113
- 1280 Meehl, G. A., Tebaldi, C., Walton, G., Easterling, D., & McDaniel, L. (2009). Rel-
 1281 ative increase of record high maximum temperatures compared to record low
 1282 minimum temperatures in the u.s. *Geophysical Research Letters*, *36*. doi:
 1283 10.1029/2009GL040736
- 1284 Meehl, G. A., Teng, H., Rosenbloom, N., Hu, A., Tebaldi, C., & Walton, G. (2022).
 1285 How the great plains dust bowl drought spread heat extremes around the
 1286 northern hemisphere. *Scientific Reports*, *12*. doi: 10.1038/s41598-022-22262-5
- 1287 Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., & Houston, T. G. (2012,
 1288 7). An overview of the global historical climatology network-daily database.
 1289 *Journal of Atmospheric and Oceanic Technology*, *29*, 897-910. Retrieved
 1290 from [https://journals.ametsoc.org/view/journals/atot/29/7/](https://journals.ametsoc.org/view/journals/atot/29/7/jtech-d-11-00103_1.xml)
 1291 [jtech-d-11-00103_1.xml](https://journals.ametsoc.org/view/journals/atot/29/7/jtech-d-11-00103_1.xml) doi: 10.1175/JTECH-D-11-00103.1
- 1292 Menne, M. J., & Williams, C. N. (2009, 4). Homogenization of temperature
 1293 series via pairwise comparisons. *Journal of Climate*, *22*, 1700-1717. Re-
 1294 trieved from [https://journals.ametsoc.org/view/journals/clim/22/7/](https://journals.ametsoc.org/view/journals/clim/22/7/2008jcli2263.1.xml)
 1295 [2008jcli2263.1.xml](https://journals.ametsoc.org/view/journals/clim/22/7/2008jcli2263.1.xml) doi: 10.1175/2008JCLI2263.1
- 1296 Merryfield, W. J., Baehr, J., Batté, L., Becker, E. J., Butler, A. H., Coelho,
 1297 C. A. S., ... Yeager, S. (2020, 6). Current and emerging developments in sub-
 1298 seasonal to decadal prediction. *Bulletin of the American Meteorological Soci-*
 1299 *ety*, *101*, E869-E896. Retrieved from [https://journals.ametsoc.org/view/](https://journals.ametsoc.org/view/journals/bams/101/6/bamsD190037.xml)
 1300 [journals/bams/101/6/bamsD190037.xml](https://journals.ametsoc.org/view/journals/bams/101/6/bamsD190037.xml) doi: 10.1175/BAMS-D-19-0037.1
- 1301 Milinski, S., Maher, N., & Olonscheck, D. (2020, 10). How large does a large en-
 1302 semble need to be? *Earth System Dynamics*, *11*, 885-901. Retrieved from

- 1303 <https://esd.copernicus.org/articles/11/885/2020/> doi: 10.5194/esd-11
1304 -885-2020
- 1305 Mills, E. (2005, 8). Insurance in a climate of change. *Science*, *309*, 1040-1044. Re-
1306 trieved from <https://www.science.org/doi/10.1126/science.1112121>
1307 doi: 10.1126/SCIENCE.1112121/ASSET/F852A595-6EC5-4DC5-A9C8
1308 -D41482F9C058/ASSETS/GRAPHIC/309_1040_F4.JPEG
- 1309 Milly, P. C., Malyshev, S. L., Shevliakova, E., Dunne, K. A., Findell, K. L., Gleeson,
1310 T., ... Swenson, S. (2014, 10). An enhanced model of land water and energy
1311 for global hydrologic and earth-system studies. *Journal of Hydrometeorol-*
1312 *ogy*, *15*, 1739-1761. Retrieved from [https://journals.ametsoc.org/view/](https://journals.ametsoc.org/view/journals/hydr/15/5/jhm-d-13-0162_1.xml)
1313 [journals/hydr/15/5/jhm-d-13-0162_1.xml](https://journals.ametsoc.org/view/journals/hydr/15/5/jhm-d-13-0162_1.xml) doi: 10.1175/JHM-D-13-0162.1
- 1314 Mora, C., Frazier, A. G., Longman, R. J., Dacks, R. S., Walton, M. M., Tong, E. J.,
1315 ... Giambelluca, T. W. (2013, 10). The projected timing of climate depart-
1316 ure from recent variability. *Nature*, *502*, 183-187. Retrieved from [https://](https://www.nature.com/articles/nature12540)
1317 www.nature.com/articles/nature12540 doi: 10.1038/nature12540
- 1318 Mueller, N. D., Butler, E. E., Mckinnon, K. A., Rhines, A., Tingley, M., Holbrook,
1319 N. M., & Huybers, P. (2016). Cooling of us midwest summer temperature
1320 extremes from cropland intensification. *Nature Climate Change*, *6*. doi:
1321 10.1038/nclimate2825
- 1322 Murray, D., Hoell, A., Hoerling, M., Perlwitz, J., Quan, X. W., Allured, D., ...
1323 Webb, R. S. (2020, 7). Facility for weather and climate assessments
1324 (facts): A community resource for assessing weather and climate variability.
1325 *Bulletin of the American Meteorological Society*, *101*, E1214-E1224. Re-
1326 trieved from [https://journals.ametsoc.org/view/journals/bams/101/7/](https://journals.ametsoc.org/view/journals/bams/101/7/bamsD190224.xml)
1327 [bamsD190224.xml](https://journals.ametsoc.org/view/journals/bams/101/7/bamsD190224.xml) doi: 10.1175/BAMS-D-19-0224.1
- 1328 NCAR. (2019). *The ncar command language (version 6.6.2)*. Retrieved from
1329 <http://dx.doi.org/10.5065/D6WD3XH5> doi: [http://dx.doi.org/10.5065/](http://dx.doi.org/10.5065/D6WD3XH5)
1330 [D6WD3XH5](http://dx.doi.org/10.5065/D6WD3XH5)
- 1331 NCAR. (2020). *Us clivar multi-model le archive*. Retrieved from [https://www.cesm](https://www.cesm.ucar.edu/projects/community-projects/MMLEA/)
1332 [.ucar.edu/projects/community-projects/MMLEA/](https://www.cesm.ucar.edu/projects/community-projects/MMLEA/)
- 1333 Neapolitan, R. E., & Jiang, X. (2018, 11). *Neural networks and deep learning*. Chap-
1334 man and Hall/CRC. doi: 10.1201/b22400-15

- 1335 Nesterov, Y. (1983). A method for unconstrained convex minimization problem with
1336 the rate of convergence $o(1/k^2)$. *Doklady AN USSR*, 269.
- 1337 O'Neill, B. C., Tebaldi, C., Vuuren, D. P. V., Eyring, V., Friedlingstein, P., Hurtt,
1338 G., . . . Sanderson, B. M. (2016, 9). The scenario model intercomparison
1339 project (scenariomip) for cmip6. *Geoscientific Model Development*, 9, 3461-
1340 3482. doi: 10.5194/GMD-9-3461-2016
- 1341 Pan, Z. T., Arritt, R. W., Takle, E. S., Gutowski, W. J., Anderson, C. J., & Se-
1342 gal, M. (2004, 9). Altered hydrologic feedback in a warming climate intro-
1343 duces a "warming hole". *Geophysical Research Letters*, 31, L17109. doi:
1344 10.1029/2004GL020528
- 1345 Parker, D. E., Wilson, H., Jones, P. D., Christy, J. R., & Folland, C. K. (1996).
1346 The impact of mount pinatubo on world-wide temperatures. *International*
1347 *Journal of Climatology*, 16. doi: 10.1002/(SICI)1097-0088(199605)16:5<487::
1348 AID-JOC39>3.0.CO;2-J
- 1349 Parks, S. A., & Abatzoglou, J. T. (2020, 11). Warmer and drier fire seasons con-
1350 tribute to increases in area burned at high severity in western us forests
1351 from 1985 to 2017. *Geophysical Research Letters*, 47, e2020GL089858. Re-
1352 trieved from [https://agupubs.onlinelibrary.wiley.com/doi/10.1029/](https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020GL089858)
1353 [2020GL089858](https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020GL089858) doi: 10.1029/2020GL089858
- 1354 Partridge, T. F., Winter, J. M., Osterberg, E. C., Hyndman, D. W., Kendall, A. D.,
1355 & Magilligan, F. J. (2018, 2). Spatially distinct seasonal patterns and forc-
1356 ings of the u.s. warming hole. *Geophysical Research Letters*, 45, 2055-2063.
1357 Retrieved from [https://agupubs.onlinelibrary.wiley.com/doi/10.1002/](https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2017GL076463)
1358 [2017GL076463](https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2017GL076463) doi: 10.1002/2017GL076463
- 1359 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . .
1360 Édouard Duchesnay (2011). Scikit-learn: Machine learning in python. *Journal*
1361 *of Machine Learning Research*, 12.
- 1362 Peters, G. P., & Hausfather, Z. (2020). Emissions - the 'business as usual' story is
1363 misleading. *Nature*, 577.
- 1364 Peterson, T. C., Heim, R. R., Hirsch, R., Kaiser, D. P., Brooks, H., Diffenbaugh,
1365 N. S., . . . Wuebbles, D. (2013). Monitoring and understanding changes
1366 in heat waves, cold waves, floods, and droughts in the united states: State
1367 of knowledge. *Bulletin of the American Meteorological Society*, 94. doi:

- 1368 10.1175/BAMS-D-12-00066.1
- 1369 Pielke, R., Burgess, M. G., & Ritchie, J. (2022, 2). Plausible 2005–2050 emissions
1370 scenarios project between 2c and 3c of warming by 2100. *Environmental Re-*
1371 *search Letters*, *17*, 024027. Retrieved from [https://iopscience.iop.org/](https://iopscience.iop.org/article/10.1088/1748-9326/ac4ebf/meta)
1372 [article/10.1088/1748-9326/ac4ebf/meta](https://iopscience.iop.org/article/10.1088/1748-9326/ac4ebf/meta) doi: 10.1088/1748-9326/
1373 AC4EBF
- 1374 Program, U. G. C. R. (2018). *Climate science special report: Fourth na-*
1375 *tional climate assessment, volume i* (Vol. 1). Retrieved from [https://](https://science2017.globalchange.gov/downloads/CSSR2017_FullReport.pdf)
1376 science2017.globalchange.gov/downloads/CSSR2017_FullReport.pdf%
1377 [0Ascience2017.globalchange.gov](https://science2017.globalchange.gov) doi: 10.7930/J0J964J6
- 1378 Rader, J. K., Barnes, E. A., Ebert-Uphoff, I., & Anderson, C. (2022, 7). Detection
1379 of forced change within combined climate fields using explainable neural net-
1380 works. *Journal of Advances in Modeling Earth Systems*, *14*, e2021MS002941.
1381 Retrieved from [https://onlinelibrary.wiley.com/doi/full/10.1029/](https://onlinelibrary.wiley.com/doi/full/10.1029/2021MS002941)
1382 [2021MS002941](https://onlinelibrary.wiley.com/doi/full/10.1029/2021MS002941) doi: 10.1029/2021MS002941
- 1383 Rampal, N., Gibson, P. B., Sood, A., Stuart, S., Fauchereau, N. C., Brandolino, C.,
1384 ... Meyers, T. (2022). High-resolution downscaling with interpretable deep
1385 learning: Rainfall extremes over new zealand. *Weather and Climate Extremes*,
1386 *38*. doi: 10.1016/j.wace.2022.100525
- 1387 Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., ... Rafaj, P. (2011,
1388 11). Rcp 8.5-a scenario of comparatively high greenhouse gas emissions. *Cli-*
1389 *matic Change*, *109*, 33-57. Retrieved from [https://link.springer.com/](https://link.springer.com/article/10.1007/s10584-011-0149-y)
1390 [article/10.1007/s10584-011-0149-y](https://link.springer.com/article/10.1007/s10584-011-0149-y) doi: 10.1007/S10584-011-0149-Y/
1391 FIGURES/12
- 1392 Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fuji-
1393 mori, S., ... Tavoni, M. (2017, 1). The shared socioeconomic path-
1394 ways and their energy, land use, and greenhouse gas emissions implica-
1395 tions: An overview. *Global Environmental Change*, *42*, 153-168. doi:
1396 10.1016/J.GLOENVCHA.2016.05.009
- 1397 Rodgers, K. B., Lee, S.-S., Rosenbloom, N., Timmermann, A., Danabasoglu,
1398 G., Deser, C., ... Yeager, S. G. (2021, 12). Ubiquity of human-induced
1399 changes in climate variability. *Earth System Dynamics*, *12*, 1393-1411. Re-
1400 trieved from <https://esd.copernicus.org/articles/12/1393/2021/> doi:

- 1401 10.5194/ESD-12-1393-2021
- 1402 Rodgers, K. B., Lin, J., & Frölicher, T. L. (2015, 6). Emergence of multiple ocean
1403 ecosystem drivers in a large ensemble suite with an earth system model. *Bio-*
1404 *geosciences*, *12*, 3301-3320. doi: 10.5194/BG-12-3301-2015
- 1405 Rossum, G. V., & Drake, F. L. (2009). *Python 3 reference manual*. CreateSpace.
- 1406 Ruder, S. (2016, 9). An overview of gradient descent optimization algorithms. *arXiv*.
1407 Retrieved from <http://arxiv.org/abs/1609.04747>
- 1408 Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., & Müller, K.-R. (2019).
1409 Explainable ai: Interpreting, explaining and visualizing deep learning. *Lecture*
1410 *Notes in Computer Science (LNCS)*, *11700*.
- 1411 Satoh, Y., Yoshimura, K., Pokhrel, Y., Kim, H., Shiogama, H., Yokohata, T., ...
1412 Oki, T. (2022, 6). The timing of unprecedented hydrological drought under
1413 climate change. *Nature Communications 2022 13:1*, *13*, 1-11. Retrieved
1414 from <https://www.nature.com/articles/s41467-022-30729-2> doi:
1415 10.1038/s41467-022-30729-2
- 1416 Scaife, A. A., Camp, J., Comer, R., Davis, P., Dunstone, N., Gordon, M., ... Vidale,
1417 P. L. (2019). Does increased atmospheric resolution improve seasonal climate
1418 predictions? *Atmospheric Science Letters*, *20*. doi: 10.1002/asl.922
- 1419 Schlunegger, S., Rodgers, K. B., Sarmiento, J. L., Ilyina, T., Dunne, J. P., Takano,
1420 Y., ... Lehner, F. (2020, 8). Time of emergence and large ensemble intercom-
1421 parison for ocean biogeochemical trends. *Global Biogeochemical Cycles*, *34*,
1422 e2019GB006453. Retrieved from [https://agupubs.onlinelibrary.wiley](https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2019GB006453)
1423 [.com/doi/10.1029/2019GB006453](https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2019GB006453) doi: 10.1029/2019GB006453
- 1424 Schulzweida, U. (2019, 2). Cdo user guide. *Zenodo*. Retrieved from [https://zenodo](https://zenodo.org/record/2558193)
1425 [.org/record/2558193](https://zenodo.org/record/2558193) doi: 10.5281/ZENODO.2558193
- 1426 Schwarzwald, K., & Lenssen, N. (2022). The importance of internal climate variabil-
1427 ity in climate impact projections. *Proceedings of the National Academy of Sci-*
1428 *ences of the United States of America*, *119*. doi: 10.1073/pnas.2208095119
- 1429 Sear, C. B., Kelly, P. M., Jones, P. D., & Goodess, C. M. (1987). Global surface-
1430 temperature responses to major volcanic eruptions. *Nature*, *330*. doi: 10.1038/
1431 330365a0
- 1432 Shin, N.-Y., Ham, Y.-G., Kim, J.-H., Cho, M., & Kug, J.-S. (2022). Application
1433 of deep learning to understanding enso dynamics. *Artificial Intelligence for the*

- 1434 *Earth Systems*, 1. doi: 10.1175/aies-d-21-0011.1
- 1435 Shiogama, H., Tatebe, H., Hayashi, M., Abe, M., Arai, M., Koyama, H., ... Watanabe, M. (2023). Miroc6 large ensemble (miroc6-le): experimental design and initial analyses. *Earth Syst. Dynam. Discuss.* Retrieved from
1436
1437 <https://doi.org/10.5194/esd-2023-12> doi: 10.5194/esd-2023-12
1438
- 1439 Sippel, S., Meinshausen, N., Fischer, E. M., Székely, E., & Knutti, R. (2020, 1).
1440 *Climate change now detectable from any single day of weather at global scale*
1441 (Vol. 10). Nature Research. doi: 10.1038/s41558-019-0666-7
- 1442 Sippel, S., Meinshausen, N., Merrifield, A., Lehner, F., Pendergrass, A. G., Fischer,
1443 E., & Knutti, R. (2019). Uncovering the forced climate response from a single ensemble member using statistical learning. *Journal of Climate*, 32. doi:
1444 10.1175/JCLI-D-18-0882.1
1445
- 1446 Slivinski, L. C., Compo, G. P., Sardeshmukh, P. D., Whitaker, J. S., McColl, C.,
1447 Allan, R. J., ... Wyszynski, P. (2021, 2). An evaluation of the performance of
1448 the twentieth century reanalysis version 3. *Journal of Climate*, 34, 1417-1438.
1449 Retrieved from [https://journals.ametsoc.org/view/journals/clim/34/4/](https://journals.ametsoc.org/view/journals/clim/34/4/JCLI-D-20-0505.1.xml)
1450 [JCLI-D-20-0505.1.xml](https://journals.ametsoc.org/view/journals/clim/34/4/JCLI-D-20-0505.1.xml) doi: 10.1175/JCLI-D-20-0505.1
- 1451 Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S.,
1452 McColl, C., ... Wyszynski, P. (2019, 10). Towards a more reliable historical
1453 reanalysis: Improvements for version 3 of the twentieth century reanalysis
1454 system. *Quarterly Journal of the Royal Meteorological Society*, 145, 2876-
1455 2908. Retrieved from [https://onlinelibrary.wiley.com/doi/abs/10.1002/](https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.3598)
1456 [qj.3598](https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.3598) doi: 10.1002/qj.3598
- 1457 Sonnewald, M., & Lguensat, R. (2021, 8). Revealing the impact of global heating
1458 on north atlantic circulation using transparent machine learning. *Journal of*
1459 *Advances in Modeling Earth Systems*, 13, e2021MS002496. Retrieved from
1460 <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021MS002496>
1461 doi: 10.1029/2021MS002496
- 1462 Sundararajan, M., Taly, A., & Yan, Q. (2017, 3). Axiomatic attribution for deep
1463 networks. *34th International Conference on Machine Learning, ICML 2017*, 7,
1464 5109-5118. Retrieved from <https://arxiv.org/abs/1703.01365v2> doi: 10
1465 .48550/arxiv.1703.01365

- 1466 Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., ... Ki-
 1467 moto, M. (2019, 7). Description and basic evaluation of simulated mean state,
 1468 internal variability, and climate sensitivity in miroc6. *Geoscientific Model*
 1469 *Development, 12*, 2727-2765. doi: 10.5194/GMD-12-2727-2019
- 1470 Thompson, V., Kennedy-Asser, A. T., Vosper, E., Lo, Y. T. E., Huntingford, C.,
 1471 Andrews, O., ... Mitchell, D. (2022). The 2021 western north america heat
 1472 wave among the most extreme events ever recorded globally. *Science Advances*,
 1473 8. doi: 10.1126/sciadv.abm6860
- 1474 Thyng, K., Greene, C., Hetland, R., Zimmerle, H., & DiMarco, S. (2016, 9).
 1475 True colors of oceanography: Guidelines for effective and accurate colormap
 1476 selection. *Oceanography, 29*, 9-13. Retrieved from [https://tos.org/](https://tos.org/oceanography/article/true-colors-of-oceanography-guidelines-for-effective-and-accurate-colormap)
 1477 [oceanography/article/true-colors-of-oceanography-guidelines-for](https://tos.org/oceanography/article/true-colors-of-oceanography-guidelines-for-effective-and-accurate-colormap)
 1478 [-effective-and-accurate-colormap](https://tos.org/oceanography/article/true-colors-of-oceanography-guidelines-for-effective-and-accurate-colormap) doi: 10.5670/oceanog.2016.66
- 1479 Thépaut, J. N., Pinty, B., Dee, D., & Engelen, R. (2018). The copernicus
 1480 programme and its climate change service. In (Vol. 2018-July). doi:
 1481 10.1109/IGARSS.2018.8518067
- 1482 Toms, B. A., Barnes, E. A., & Ebert-Uphoff, I. (2020, 9). Physically interpretable
 1483 neural networks for the geosciences: Applications to earth system variabil-
 1484 ity. *Journal of Advances in Modeling Earth Systems, 12*. Retrieved from
 1485 <https://onlinelibrary.wiley.com/doi/10.1029/2019MS002002> doi:
 1486 10.1029/2019MS002002
- 1487 Trisos, C. H., Merow, C., & Pigot, A. L. (2020, 4). The projected timing of abrupt
 1488 ecological disruption from climate change. *Nature 2020 580:7804*, 580, 496-
 1489 501. Retrieved from <https://www.nature.com/articles/s41586-020-2189-9>
 1490 doi: 10.1038/s41586-020-2189-9
- 1491 van der Velden, E. (2020, 2). Cmasher: Scientific colormaps for making accessible,
 1492 informative and 'cmashing' plots. *Journal of Open Source Software, 5*, 2004.
 1493 Retrieved from <https://joss.theoj.org/papers/10.21105/joss.02004>
 1494 doi: 10.21105/JOSS.02004
- 1495 Vecchi, G. A., Delworth, T., Gudgel, R., Kapnick, S., Rosati, A., Wittenberg,
 1496 A. T., ... Zhang, S. (2014, 11). On the seasonal forecasting of regional
 1497 tropical cyclone activity. *Journal of Climate, 27*, 7994-8016. Retrieved
 1498 from <https://journals.ametsoc.org/view/journals/clim/27/21/>

- 1499 [jcli-d-14-00158.1.xml](#) doi: 10.1175/JCLI-D-14-00158.1
- 1500 Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cour-
 1501 napeau, D., ... Vázquez-Baeza, Y. (2020). Scipy 1.0: fundamental al-
 1502 gorithms for scientific computing in python. *Nature Methods*, 17. doi:
 1503 10.1038/s41592-019-0686-2
- 1504 Vose, R. S., Applequist, S., Squires, M., Durre, I., Menne, C. J., Williams, C. N.,
 1505 ... Arndt, D. (2014, 5). Improved historical temperature and precipi-
 1506 tation time series for u.s. climate divisions. *Journal of Applied Meteorol-
 1507 ogy and Climatology*, 53, 1232-1251. Retrieved from [https://journals](https://journals.ametsoc.org/view/journals/apme/53/5/jamc-d-13-0248.1.xml)
 1508 [.ametsoc.org/view/journals/apme/53/5/jamc-d-13-0248.1.xml](https://journals.ametsoc.org/view/journals/apme/53/5/jamc-d-13-0248.1.xml) doi:
 1509 10.1175/JAMC-D-13-0248.1
- 1510 Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., ...
 1511 Rose, S. K. (2011, 8). The representative concentration pathways: an overview.
 1512 *Climatic Change*, 109, 5-31. Retrieved from [http://link.springer.com/](http://link.springer.com/10.1007/s10584-011-0148-z)
 1513 10.1007/s10584-011-0148-z doi: 10.1007/s10584-011-0148-z
- 1514 Weiskopf, S. R., Rubenstein, M. A., Crozier, L. G., Gaichas, S., Griffis, R., Halof-
 1515 sky, J. E., ... Whyte, K. P. (2020, 9). Climate change effects on biodiver-
 1516 sity, ecosystems, ecosystem services, and natural resource management in
 1517 the united states. *Science of The Total Environment*, 733, 137782. doi:
 1518 10.1016/J.SCITOTENV.2020.137782
- 1519 Williams, A. P., Cook, E. R., Smerdon, J. E., Cook, B. I., Abatzoglou, J. T.,
 1520 Bolles, K., ... Livneh, B. (2020). Large contribution from anthropogenic
 1521 warming to an emerging north american megadrought. *Science*, 368. doi:
 1522 10.1126/science.aaz9600
- 1523 Wuebbles, D., Meehl, G., Hayhoe, K., Karl, T. R., Kunkel, K., Santer, B., ...
 1524 Sun, L. (2014). Cmpip5 climate model analyses: Climate extremes in the
 1525 united states. *Bulletin of the American Meteorological Society*, 95. doi:
 1526 10.1175/BAMS-D-12-00172.1
- 1527 Wyser, K., Koenigk, T., Fladrich, U., Fuentes-Franco, R., Karami, M. P., &
 1528 Kruschke, T. (2021, 7). The smhi large ensemble (smhi-lens) with
 1529 ec-earth3.3.1. *Geoscientific Model Development*, 14, 4781-4796. doi:
 1530 10.5194/GMD-14-4781-2021

- 1531 Yang, X., Delworth, T. L., Jia, L., Johnson, N. C., Lu, F., & McHugh, C. (2022,
1532 9). On the seasonal prediction and predictability of winter surface temper-
1533 ature swing index over north america. *Frontiers in Climate*, *4*, 167. doi:
1534 10.3389/FCLIM.2022.972119/BIBTEX
- 1535 Yu, B., Li, G., Chen, S., & Lin, H. (2020, 8). The role of internal variability in
1536 climate change projections of north american surface air temperature and tem-
1537 perature extremes in canesm2 large ensemble simulations. *Climate Dynamics*,
1538 *55*, 869-885. Retrieved from [https://link.springer.com/article/10.1007/](https://link.springer.com/article/10.1007/s00382-020-05296-1)
1539 [s00382-020-05296-1](https://link.springer.com/article/10.1007/s00382-020-05296-1) doi: 10.1007/S00382-020-05296-1/FIGURES/15
- 1540 Yu, S., Alapaty, K., Mathur, R., Pleim, J., Zhang, Y., Nolte, C., ... Nagashima, T.
1541 (2014). Attribution of the united states "warming hole": Aerosol indirect effect
1542 and precipitable water vapor. *Scientific Reports*, *4*. doi: 10.1038/srep06929
- 1543 Zadeh, L. A. (1965, 6). Fuzzy sets. *Information and Control*, *8*, 338-353. doi: 10
1544 .1016/S0019-9958(65)90241-X
- 1545 Zender, C. S. (2008). Analysis of self-describing gridded geoscience data with netcdf
1546 operators (nco). *Environmental Modelling and Software*, *23*. doi: 10.1016/j
1547 .envsoft.2008.03.004
- 1548 Zhao, M., Golaz, J. C., Held, I. M., Guo, H., Balaji, V., Benson, R., ... Xiang, B.
1549 (2018a, 3). The gfdl global atmosphere and land model am4.0/lm4.0: 1. sim-
1550 ulation characteristics with prescribed ssts. *Journal of Advances in Modeling*
1551 *Earth Systems*, *10*, 691-734. Retrieved from [https://agupubs.onlinelibrary](https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2017MS001208)
1552 [.wiley.com/doi/10.1002/2017MS001208](https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2017MS001208) doi: 10.1002/2017MS001208
- 1553 Zhao, M., Golaz, J. C., Held, I. M., Guo, H., Balaji, V., Benson, R., ... Xiang,
1554 B. (2018b, 3). The gfdl global atmosphere and land model am4.0/lm4.0:
1555 2. model description, sensitivity studies, and tuning strategies. *Jour-*
1556 *nal of Advances in Modeling Earth Systems*, *10*, 735-769. Retrieved from
1557 <https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2017MS001209>
1558 doi: 10.1002/2017MS001209