1

2 **Towards a robust, impact-based, predictive drought metric**

3

4 **Sanaa Hobeichi[1,2], Gab Abramowitz[1,2], Jason P. Evans[1,2], and Anna Ukkola[1,2]**

5 [1]Climate Change Research Centre, UNSW Sydney, NSW 2052, Australia.

6 [2]ARC Centre of Excellence for Climate Extremes, UNSW Sydney, NSW 2052, Australia.

7 Corresponding author: Sanaa Hobeichi (s.hobeichi@unsw.edu.au)

8 **Key Points:**

9 • A new approach to defining drought, providing information for all impacted sectors.

10 • The metric is based on an empirical relationship between droughts documented in impact
11 reports, and a range of observed climate features.

12 • The metric quantifies the conditional probability of drought considering climate features,
13 and can be used for forecasting.

14

15

Abstract

This work presents a new approach to defining drought, establishing an empirical relationship between historical droughts (and wet spells) documented in impact reports, and a broad range of observed drought-related climate features. A Random Forest (RF) algorithm was trained to identify the particular combinations of predictors – such as precipitation, soil moisture and potential evapotranspiration – that led to categorical, documented drought or non-drought events. Unlike traditional drought definitions, the new RF drought indicator combines meteorological, hydrological, agricultural, and socioeconomic drought, providing drought information for all impacted sectors. The metric also quantifies the conditional probability of drought (rather than being threshold-based), considering multiple climate features and their interactive effect, and can be used for forecasting.

The approach was validated out-of-sample across several random selections of training and testing datasets, and demonstrated better predictive capabilities than commonly used drought indicators in a range of performance metrics. Furthermore, it showed a comparable performance to the (expert elicitation-based) US Drought Monitor (USDM) which is the current state-of-the-art record of historical drought in the USA. As well as providing an alternative historical drought indicator to USDM, the RF approach offers additional advantages by being automated, by providing drought information at the grid-scale, and by having predictive capacity.

As a proof-of-concept case, the RF drought indicator was trained on Texan climate data and droughts, and validated in all Texas ecoregions. However, the introduced approach can be easily implemented to develop a RF drought indicator for new regions if adequate information on historical droughts is available.

1 Introduction

By the mid-1980s, drought had been defined in the scientific literature in more than 150 ways (Wilhite & Glantz, 1985). Existing definitions reflect perception differences across various disciplines (e.g. meteorology, hydrology, agriculture, society and economy) of the most important impacts of droughts (Wilhite & Glantz, 1985). Research in the late 1990s grouped existing conceptual definitions into four forms of drought (AMS, 1997). Meteorological drought (also termed climatological drought) refers to a period of below normal precipitation. Agricultural or soil moisture drought is concerned with the deficiency in water available for agriculture or natural ecosystem as a result of subsequent soil moisture depletion. Hydrological drought is concerned with the direct or indirect impacts of shortfall in surface and subsurface water supply. Socioeconomic drought refers to the effect of any of the meteorological, agricultural or/and hydrological droughts on people and water-dependent economies. More recently, the IPCC report defined drought as 'a period of abnormally dry weather long enough to cause a serious hydrological imbalance' (IPCC, 2014; Seneviratne et al., 2012).

Drought indicators typically assess anomalies in a particular climate feature and make drought conclusions based on pre-defined thresholds (Heim, 2002; J. Keyantash & Dracup, 2002; Yihdego et al., 2019). Among the most common indicators used in drought analysis are the Standardised Precipitation Index (SPI; McKee et al. 1993) and the Palmer drought severity index (PDSI; Palmer 1965). SPI is based solely on precipitation (P) anomaly, while PDSI simulates soil moisture anomaly from the difference of potential evapotranspiration (PET) and P. More recently, (Hobbins et al. 2016) developed the Evaporative Demand Drought Index (EDDI), a drought indicator that is based solely on PET anomaly.

Drought indicators typically define a drought event as statistically anomalous in a distribution of a specific climate feature (e.g. McKee et al. 1993; Stagge et al. 2015). There are however circumstances where near-normal conditions of several climate variables occurring simultaneously lead to impactful droughts even though they wouldn't necessarily be labelled as droughts using common drought indicators. For example, in the agricultural context, moderate pre-existing soil moisture shortages combined with a moderate precipitation shortage will likely result in a drought. None of these hydroclimatic variables, when considered in isolation, needs to be an extreme anomaly for a drought to occur (IPCC, 2014). Similarly, a pre-existing soil moisture surplus combined with abnormally low precipitation might not lead to a drought. Therefore, looking for droughts only in the extremes of a distribution can be misleading.

Furthermore, drought indicators usually focus on a narrow selection of climate or agro-hydrological variables (and sometimes a single one) and so ultimately cannot identify all forms of droughts (Van Loon & Van Lanen, 2012). For effective drought planning and response, it is important to develop monitoring tools capable of providing drought information for all sectors impacted by droughts (Wilhite, 2009). This requires simultaneous assessment of several drought-related variables (Brown et al., 2008). Several approaches integrate various aspects of the land-atmosphere-ocean system (e.g. Azmi et al. 2016; Brown et al. 2008; Fernando et al. 2019; Keyantash and Dracup, 2004; Li et al. 2015; Zhang and Jia, 2013), improving drought identification. However, they were not designed to detect all forms of drought, although some exceptions exist (Azmi et al., 2016). The development of a comprehensive drought index was described by the United States Western Governors' Association (WGA) as a top priority for improving monitoring capabilities and assisting sectors at risk in planning mitigation activities (AWG, 2004).

Recent research applied machine learning techniques to predict existing drought indices using a number of climate variables as predictor variables (Deo & Şahin, 2015; Khan et al., 2020; Park et al., 2016; Soh et al., 2018). These efforts enabled the reconstruction of drought indices over time and space where the original drought indices could not be developed mainly due to lack of data needed to derive them. Machine learning-based indictors developed this way at best mimic the predictive capabilities of the drought indicator they are trying to emulate. However, as the drought indicators are themselves not perfect, fail to accurately depict drought events. Ultimately, the enhancement brought by most of these machine learning-based indicators is limited to extrapolation in time (i.e. future predictions and past reconstruction) and/or space (areas with no data) – the quality of prediction offered by the drought index did not improve.

Very little effort has been made to incorporate real drought impacts data in the development of drought indices. This is curious since, in reality, the main purpose of using drought indicators is to enable governments and water-dependent sectors to better address impacts associated with droughts (AWG, 2004). Arguably, for better decision-making in water resources and agricultural management, it is important that drought definitions only include droughts that have impacts, and avoid the very real possibility of giving false warnings about events simply because they were found in the extreme of a distribution.

The aim of this paper is to introduce a new approach to defining drought using machine learning that can address some of the limitations of existing approaches. Texas is used as the test region, taking advantage of the wealth of drought information available from drought impact

106 reports and other resources. The following section describes the data used to train the Random
107 Forest (RF) algorithms, and the applied methods to test and validate the developed RF drought.
108 Results are provided in section 3, then discussed and summarized afterwards.

109 　　　The main text should start with an introduction. Except for short manuscripts (such as
110 comments and replies), the text should be divided into sections, each with its own heading.
111 Sections are numbered (1, 2, 3, etc.). A maximum of four levels of heads may be used, with
112 subsections numbered 1.1., 1.2.; 1.1.1., 1.2.1; 1.1.1.1., and so on. Headings should be sentence
113 fragments. Examples of headings are:

114 **2 Materials and Methods**

115 A RF binary classification algorithm was trained to discern 'drought' and 'no drought'
116 conditions from monthly climate data. The labelled data that was used to build and test the RF
117 model comprised monthly climate data as predictor variables (or features), and a binary class of
118 'drought' and 'no drought' labels as a response variable. We developed a database of binary
119 labels by compiling several hundred reports that provide information on drought impacts and
120 monthly weather conditions at 30 Texan counties. Corresponding climate data was extracted
121 from several global datasets of drought-related variables.

122

123 　　　Training the RF algorithm was conducted on 75% of the labelled data, while the
124 remaining 25% of the data was used for out-of-sample testing of the trained model. The
125 performance of the RF algorithm was assessed across 100 different random selection of training
126 and testing subsets and compared with commonly used drought indicators along with the US
127 Drought Monitor (USDM), which is the state-of-the art drought monitor in Texas. A detailed
128 description of the methodology is provided below.

129 　　　2.1 Predictor variables

130 　　　Predictor variables comprise a range of drought-related climate variables and phenomena
131 that describe the land-atmosphere-ocean system. These include monthly estimates of
132 precipitation (P), soil moisture (SM), potential evapotranspiration (PET), actual
133 evapotranspiration (ET), change in water storage (CWS), Normalised Difference Vegetation
134 Index (NDVI), and El Niño–Southern Oscillation (ENSO). Soil moisture of the previous month
135 (SM_{prev}) and the calendar month were also incorporated as predictor variables. The source and
136 reference of each dataset are provided in Table 1. The spatial resolution of all the employed
137 gridded datasets is 0.25° except PET and CWS which have a coarser resolution of 0.5°. All the
138 gridded datasets were resampled to a common 0.5° grid using nearest neighborhood
139 interpolation. Predictor variables were then extracted at 30 grid points (Figure 1) in all time steps
140 during 1982-2016 where matching drought event labels are available. The 30 grid points are
141 located in 30 counties, most of them are about the size of a grid cell, i.e. 0.5°. These are
142 distributed over all 12 Texan eco-regions identified by the United States Environmental
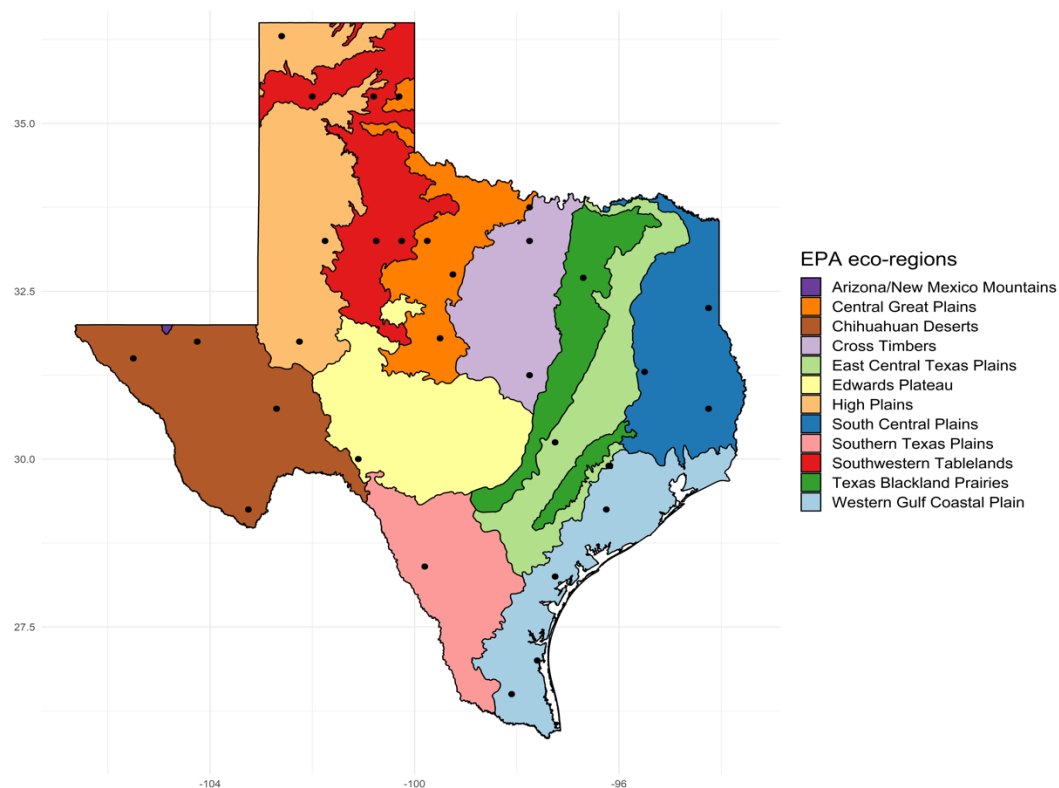143 Protection Agency (EPA, https://www.epa.gov/; Figure 1).

144
Figure 1: Location of 30 grid cells used in this study over a layer of Texas ecoregion map (level 3) developed by the EPA (https://www.epa.gov/).

Table 1: Climate variables used as predictor variables

Climate variable and unit × month⁻¹	Name and Reference	Temporal and spatial coverage and resolution	Data description and access link
Change in total water storage (mm)	GRACE-REC (Humphrey & Gudmundsson, 2019)	1979-2016 monthly 0.5° global land	JPL_MSWEP – 1st member: Statistical model trained with GRACE JPL mascons and forced with MSWEP precipitation. The change in total water storage in a given month was computed by subtracting the total water storage anomalies of the previous month from the current month. https://figshare.com/
Evapotranspiration (mm)	DOLCE V2.1 (Hobeichi, 2020) (Hobeichi et al., 2020)	1980-2018 monthly 0.25° global land	Observationally constrained hybrid evapotranspiration product derived by merging 11 available ET products. http://dx.doi.org/10.25914/5eab8f533aeae
Precipitation (mm)	GPCC V2018 (Schneider et al., 2018)	1891-2016 monthly 0.25° global land excluding Antarctica	Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historical Data https://psl.noaa.gov/data/gridded/data.gpcc.html
Potential Evapotranspiration (mm)	Priestley-Taylor PET	1901-2017 monthly	Calculated from CRU TS4.02 monthly cloud cover and mean temperature using the R package

The unit for the *Climate variable and unit × month⁻¹* column is written as "unit $\times$ month^{-1}".

		0.5° global land excluding Antarctica	*rstash* (https://github.com/rhyswhitley/r_stash; Davis et al. 2017)
Soil moisture of the current and previous months $(m^3\ m^{-3})$	CCI-SM (Gruber et al., 2019) (Gruber et al., 2017) (Dorigo et al., 2017) (Dorigo et al., 2017)	1979-2019 daily 0.25° daily global land excluding land covered with snow	COMBINED CCI Soil Moisture product datasets v04.7 https://esa-soilmoisture-cci.org/
month		1980-2016	Calendar month
ENSO Index	(Smith & Sardeshmukh, 2000)	1870-2020 1-month running mean	A Bivariate EnSo Timeseries or the "BEST" ENSO Index it combines (i) SOI: Southern Oscillation Index (based on the observed sea level pressure differences between Tahiti and Darwin) and (ii) Niño 3.4 SST (NINO3.4 is the average sea surface temperature anomaly in the region bounded by 5°N to 5°S, from 170°W to 120°W) based on the mean climatology for the period 1871-2001. https://psl.noaa.gov/
NDVI	NASA-GIMMS v1.1 (Pinzon & Tucker, 2014)	July 1981 to Dec 2017 0.0833° bimonthly	NDVI from Advanced Very High Resolution Radiometer, averaged to monthly by taking the maximum of bimonthly values https://gimms.gsfc.nasa.gov/

150
151
152   Most of these predictor variables appear in existing drought monitoring approaches
153 (Beguería et al. 2014; Brown et al. 2008; Karnieli et al. 2010; McKee et al. 1993; Nanzad et al.
154 2019). Incorporating ENSO as a predictor variable was guided by studies showing droughts in
155 Texas are related to La Niña events, which affect Pacific moisture patterns (Pu et al., 2016;
156 Schubert et al., 2004; Seager et al., 2014). SM_{prev} was used to provide information on the
157 resilience of the system to withstand drought.

158   2.2 Binary database of 'drought' and 'no drought' events

159 'Drought' and 'no drought' events attributed to a grid cell during a period of time are based on
160 information extracted from two main sources. The source that contributed to most of the
161 'drought' events is the Drought Impacts Reporter (DIR), a national interactive drought impact
162 database developed and maintained by the U.S. National Drought Mitigation Center (NDMC)
163 (Wilhite et al., 2007). Sources contributing to the DIR database include news articles, scientific
164 publications, National Weather Service Drought Information Statements, agency reports, and
165 reports submitted by government officials and the public. The DIR comprises information on
166 drought impacts reported by a wide range of drought-impacted sectors. Submitted reports from
167 any source are then reviewed for drought impact information and verified by NDMC before
168 becoming publicly available at https://droughtreporter.unl.edu/. Reported impacts include the
169 agricultural sector, livestock, water, energy, and fire sectors, social impacts, forestry, recreation
170 and tourism, and more.
171
172   A major source of 'no-drought' events are the Texas Climate Monthly Reports (TCMR),
173 monthly bulletins produced by the Office of the State Climatologist at Texas A&M University.

174 They provide a summary of weather conditions throughout Texas, describe big weather events
175 such as floods, storms, and hurricanes, and report the number of days with rain and monthly
176 precipitation totals picked up in several locations. Monthly bulletins are produced from 1990
177 onwards and can be accessed at https://climatexas.tamu.edu/products/texas-climate-
178 bulletins/index.html.
179
180   Building the database of Texas drought events involved a careful assessment of the DIR,
181 TCMR, and relevant literature. Periods where regions were trending toward drought or
182 recovering from it are not marked as events. Furthermore, we excluded reports of small scale
183 impacts and only included county scale impacts; this ensured scale consistency between
184 observed drought impacts and the measured drivers described in Section 2.1.
185   The final (spatiotemporally incomplete) database for this test case comprises a total of
186 1005 records in 500/505 split for 'no drought' and 'drought' respectively. Each record consists
187 of a location (a county), a time (year and month) and a label ('drought' or 'no drought'). Table
188 S1 in the supplementary material shows these records along with the relevant source.

189   2.3 Random Forest Algorithm

190   2.3.1 Building a Random Forest classification and probability Model

191   Random forest algorithm (Breiman 2001) grows a collection of classification trees (or
192 alternatively probability trees) each fitted on bootstrap samples (samples are drawn with
193 replacement) of labelled data (predictor variables and associated labels) available for training. As
194 a result of the bootstrapping procedure, trees in the forest are trained on different - but not
195 mutually exclusive - subsets of labelled observations. In each tree, data undergo recursive binary
196 splits based on the predictor variables. The sample data at a parent node is split on a predictor's
197 cutoff value (e.g $P=100$ mm) and results into exactly two child nodes. A subset of predictors of
198 predefined size is available for the split at each node. The RF algorithm carries out an
199 optimization procedure that controls the selection of an appropriate predictor at each node, the
200 cutoff values at which the data will split, and whether there will be further splitting. These
201 decisions are based on a metric known as the Gini index (Breiman et al., 1984) which measures
202 the relevance and consequence of each feature available for split at each node, and that ensures
203 that as the trees grow, the impurity decreases, i.e. the variance within subsequent child nodes
204 decreases. Each tree keeps growing until the impurity does not decrease further, or until the
205 number of samples in the terminal node – also called leaf node - falls below a threshold.
206
207   Each terminal node in the forest is assigned a class 'drought' or 'no drought' and a
208 probability of drought. The class represents the majority label in the terminal node. The
209 probability of drought is equal to the proportion of 'drought' labels at the terminal node, and it
210 represents the conditional probability of drought emergence given the features described from
211 the top of the tree down to this terminal node. The reliability of conditional probabilities
212 computed by the RF approach is examined and demonstrated by Malley et al. (2012).
213
214   This work applies a new implementation of random forest developed in the "RANdom
215 forest GEneRator" (ranger; Wright and Ziegler, 2017), an open source software package in R.
216 Ranger provides a higher computational speed and better memory storage efficiency compared to
217 other available implementations (e.g. Random Jungle (Kruppa et al., 2014), and randomForest

218 (Liaw & Wiener, 2002)) while maintaining a similar performance (Wright & Ziegler, 2017). We
219 used the default parameters described in the ranger package to build both the RF classification
220 model and a RF probability model. These involve 500 trees, 3 predictor variables available for
221 split at each node (i.e. $\sqrt{number\ of\ features}$), and the same size as the training dataset is used
222 for number of bootstrap samples.
223
224   It is important to note that the sub-sampling of predictors at each node along with the
225 bootstrapping procedure and the fact that trees are built in parallel force variation between trees
226 and ensure that they have a small pairwise correlation.
227
228   The outcome from training the RF algorithm on drought event data can be either a RF
229 binary classification model or a RF probability model. This is determined during the training
230 process and is based on whether the purpose is to classify new samples as 'drought' or 'no
231 drought', or to compute the conditional probability of drought. Here we developed and used both
232 models.
233

234   2.3.2 Prediction

235   To predict the binary class and the drought probability of a given new sample, its driver
236 values are propagated through all the trees in the forest and the terminal node values at each tree
237 – for both class and the probability – are collated. The final class assigned to the new sample is
238 based on the majority class from all trees, and the estimated conditional probability of drought is
239 the average probability estimate over all trees.

240   2.3.2 Variable importance

241   We use conditional permutation to assess the importance of each predictor variable as
242 described in (Strobl et al., 2008). To measure the importance of a particular predictor variable,
243 for example ET, ET is randomly permuted, then predictions are made using the remaining
244 variables and the permuted variable (substitute of ET). The difference in prediction accuracy
245 before and after permuting ET averaged over all permutations in the forest is used as a metric of
246 its importance. The most important variable is the one that achieves the largest reduction in
247 prediction accuracy when randomly permuted. Conditional permutation variable importance
248 reflects the true impact of each predictor variable more reliably than the default variable
249 importance scheme in the Ranger package, namely Gini importance (Sandri & Zuccolotto,
250 2008). For each predictor variable, Gini importance measures the reduction in impurity on the
251 response variable achieved by each predictor at every split across all nodes in all trees. The
252 conditional permutation importance was proven more reliable than the Gini importance in
253 situations where some predictor variables are highly pairwise correlated (Strobl et al., 2008),
254 and/or have different scales of measurement and categories (Strobl et al., 2007). Conditional
255 permutation variable importance was derived using the R party package (http://party.R-forge.R-
256 project.org).

257   2.4 Comparison of drought indicators

258   We compared the prediction skill of the RF drought indicator (tested out-of-sample) with
259 commonly used drought indicators. We provide a quick summary of these, and refer readers to
260 the associated publications for further details.

261

262 SPI: Assesses drought solely from precipitation. At a given location, long term monthly
263 precipitation is transformed into a normal distribution, and the computed SPI value represents the
264 unit standard normal deviate. Previous studies have associated droughts with SPI values of less
265 than −1 e.g. (Bachmair et al., 2015), −0.8 (in USDM) or 0 (McKee, 1995). We calculated
266 monthly SPI using the SPEI R package for each grid point presented in Figure 1 from the same
267 precipitation dataset used to develop the RF model. We derived SPI for several accumulation
268 periods including 1, 3, 6, 9 and 12 months. In this study we carry out the analysis using each of
269 the three drought cutoffs, i.e. -1, -0.8 and 0.

270

271 Evaporative Demand Drought Index (EDDI) (Hobbins et al., 2016): monitors drought solely
272 from PET anomalies, where PET is derived using the American Society of Civil Engineers
273 standardized reference ET equation (Walter et al., 2000), which estimates PET by simplifying
274 the Penman–Monteith equation mainly from satellite-based estimates of temperature, humidity,
275 windspeed, and solar radiation. Unlike SPI, the probability distribution of PET is computed
276 empirically using an inverse normal approximation. Positive (negative) EDDI values are
277 commonly used to discern drought (no drought) conditions. We downloaded EDDI maps for the
278 period 1980-2016 from https://psl.noaa.gov/eddi/ using the R package 'eddi'.

279

280 PDSI: assesses droughts using anomalies of soil moisture, where soil moisture is calculated from
281 *P* and PET using a simple soil moisture balance model. Negative (positive) PDSI values are used
282 to discern drought (wet) conditions. In this work we calculated PDSI from the same *P* and PET
283 datasets used to develop the RF drought indicator. We used the R package scPDSI to calculate a
284 self-calibrated version of PDSI.

285

286 The U. S. Drought Monitor (USDM) (Svoboda et al., 2002): is currently the state-of-the-practice
287 for drought monitoring in the U.S. It consists of weekly maps that show regions where land has
288 been Abnormally Dry (D0), or in drought with intensity ranging from moderate (D1) to
289 exceptional (D4). Drought categories are produced from blending i) several drought indices
290 including SPI and PDSI, (ii) the analysis of various observed and modelled climate variables
291 such as *P*, temperature, snow water equivalent, water in the soil, streams, lakes and others, (iii)
292 reported drought impacts, and (iv) experts assessment of i), ii) and iii) and judgments. In this
293 sense USDM is a retrospective, assimilated observationally-based product, that could not, for
294 example, be applied to climate projections. The spatial resolution of the USDM Maps is the
295 approximate scale of a climate division, that is 10 regions in Texas. USDM maps are available
296 from 2000. We downloaded USDM maps from https://www.drought.gov/drought/ and
297 aggregated weekly maps into monthly binary 'drought' / 'no-drought' maps whenever possible.
298 Regions consistently in drought (non-drought) during a month were labelled 'drought' (no-
299 drought), whereas regions that were in drought during part of the month were not used in the
300 comparison.

301

302   2.5 Out-of-sample testing and performance metrics

303   We assessed the performance of the RF algorithm by testing its ability to correctly classify
304 unseen events (not used in training). To achieve this, 75% of events were used to train the RF
305 model, and the remaining 25% of events used to test it. The 75/25 sampling was randomized 100
306 times to create 100 different RF models. The performance of the RF approach was then assessed
307 by comparing the performance of each RF model at its 25% of out-of-sample events, and
308 aggregating across the 100 cases. Six statistical metrics commonly used in binary classification
309 were then used to compare the out-of-sample success of the RF model compared to existing
310 drought metrics:

311 • Accuracy: correct predictions expressed as a fraction of total predictions.
312 • False alarm rate: incorrect 'drought' predictions expressed as a fraction of all 'drought'
313   predictions.
314 • Success ratio or precision: correct 'drought' predictions expressed as a fraction of all
315   'drought' predictions.
316 • Threat Score or Critical Success Index: measures how well 'drought' predictions correspond
317   to 'drought' observations. It is calculated as correct 'drought' predictions expressed as a
318   fraction of both 'drought' predictions and 'drought' observations combined.
319 • True positive rate or sensitivity (also known as recall and hit rate): correct 'drought'
320   predictions expressed as a fraction of 'drought' observations.
321 • True negative rate of specificity: correct 'no-drought' predictions expressed as a fraction of
322   'no-drought' observations.

323 A perfect score is 0 for the "False alarm rate", and 1 for all the other performance metrics.
324
325   We computed these performance metrics for the RF-drought indicator, EDDI, PDSI, SPI,
326 and USDM at all 100 testing datasets. We also assessed the predictive ability of 8 other well
327 known machine learning classifiers (Balakrishnama & Ganapathiraju, 1998; Breiman, 2001;
328 Friedman, 1991; Kuhn, 2008; Mitchell, 1997; Nelder & Wedderburn, 1972; Scholkopf et al.,
329 1997; Swain & Hauska, 1977; Wilhite et al., 2007; Zou & Hastie, 2005) trained with the same
330 training datasets as the RF classifier, by computing these performance metrics across the same
331 100 out-of-sample testing iterations. The other machine learning algorithms are listed in Table
332 S2 in the supplementary material, we refer the reader to the associated publications for
333 description of each algorithm.

334 **4 Results**

335   3.1 Performance of RF and other ML classifiers out-of-sample

336   Figure 2 shows the performance results of the random forest and other ML classification
337 algorithms, each trained on 75% of events and tested out-of-sample at 25%, across 100 random
338 selections of training and testing samples. Random forest achieves above 90% score in accuracy,
339 true positive, true negative and success ratio across the majority of iterations. The median threat
340 score exceeds 80%, and the median false alarm rate is about 10%. In comparison with the other
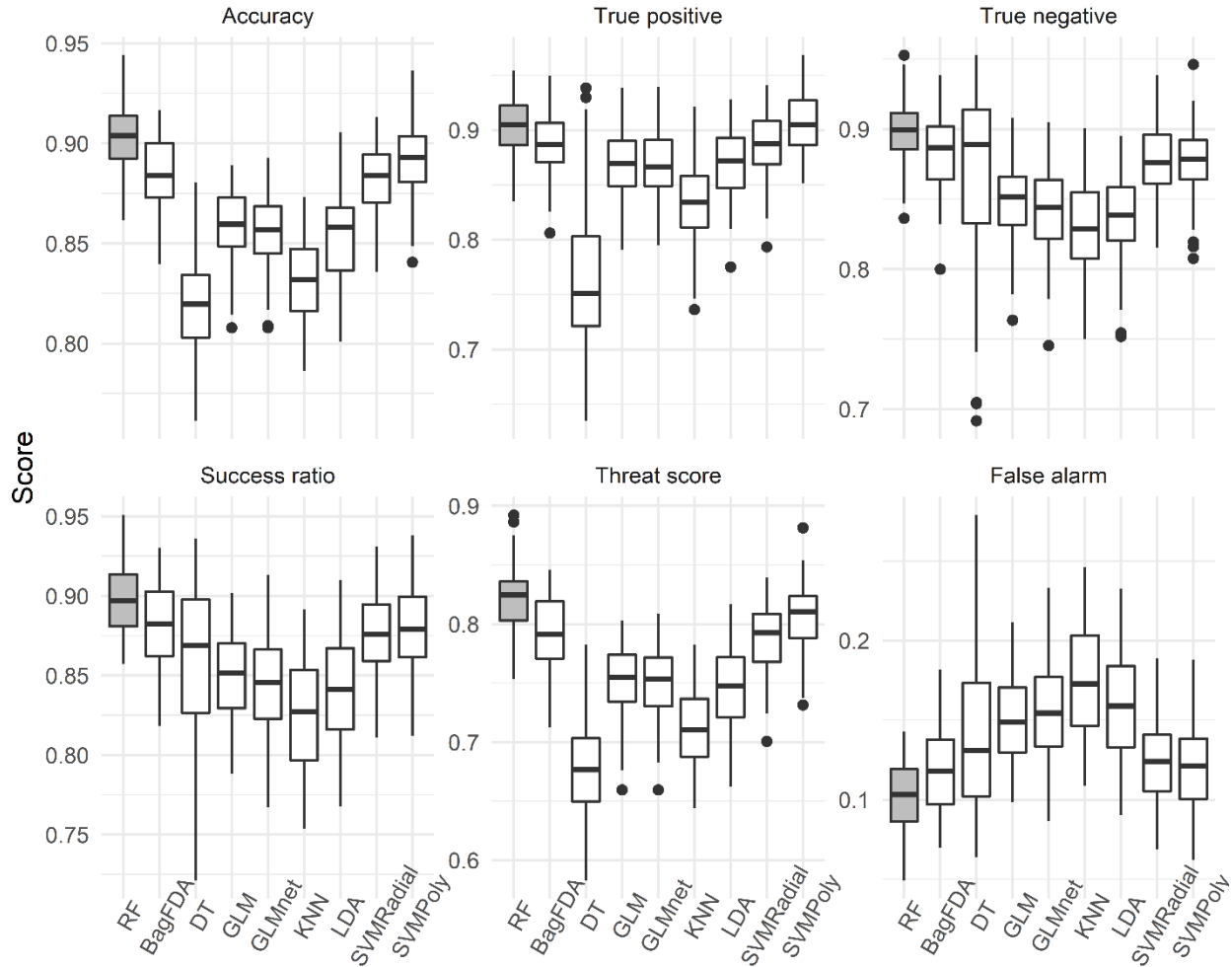341 ML approaches, overall, the random forest algorithm performs the best across all metrics.

Figure 2: Performance results of RF classification algorithm and 9 other ML classifiers at testing samples across 100 different sub-sampling of training and validating samples. Performance scores are explained in section 2.5.

The competitiveness of RF with the best available ML algorithms has been demonstrated across a range of applications (e.g. (Cutler et al., 2007; Fernández-Delgado et al., 2014; McGovern et al., 2017; Park et al., 2016; Rodriguez-Galiano et al., 2012)). Figure 2 shows that RF stands out as much more capable than the other employed ML algorithms in identifying teleconnections between climate features and droughts. There are two additional benefits in using random forests. First, RF is capable of quantifying the conditional probability of drought, a very important feature that is not found in most other classifiers. Also, as highlighted in section 2.3.2, RF allows the assessment of the importance of its predictor variables, which gives insight into the factors influencing droughts, as well as the least important climate features in explaining and quantifying droughts in different circumstances.

3.2 Performance of RF out-of-sample, compared to SPI, PDSI, EDDI and USDM

Figure 3 illustrates the performance results of the RF drought indicator relative to EDDI, PDSI, USDM and SPI computed for 6 months accumulation period (at two drought cutoffs, −0.8

360   and 0, denoted by SPI$_{-0.8}$ and SPI$_0$ respectively). The drought indicators are computed across the
361   100 different testing datasets. Overall, RF and USDM achieve the highest scores across all
362   metrics followed by SPI$_{-0.8}$ and SPI$_0$. We exclude SPI at -1 drought cutoffs from the plot as it
363   consistently shows inferior performance than each of SPI$_{-0.8}$ and SPI$_0$.
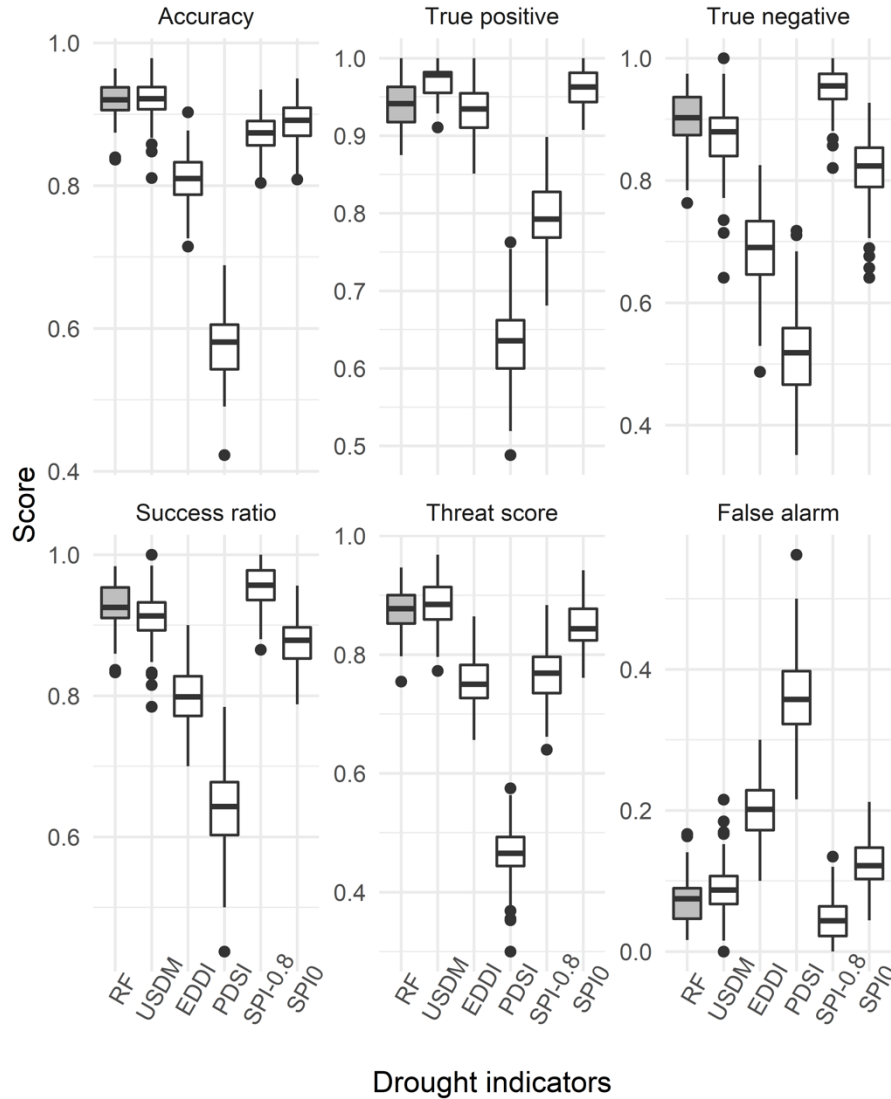

364
365   **Figure 3:** Performance scores of RF classifier and commonly used drought indicators i.e.
366   USDM, EDDI with drought threshold value of 0I, PDSI with drought threshold value of 0, and
367   SPI with drought threshold values of 0 (SPI$_0$) and -0.8 (SPI$_{-0.8}$) and computed for a six-month
368   accumulation period. Scores are computed at testing samples across 100 different sub-sampling
369   of training and validating samples. Performance scores are explained in section 2.5.
370
371   Figure 3 shows that the RF approach is more accurate than EDDI, SPI (at both
372   thresholds), and PDSI, and has comparable accuracy to USDM. While the accuracy metric
373   provides a summary of performance, the true positive and true negative scores compare the
374   ability to correctly predict drought and no drought, respectively. USDM, EDDI, SPI$_0$ and PDSI
375   appear to do significantly better in identifying 'drought' compared to 'no drought'. This indicates
376   that most of the inaccuracy in these three indicators come from their tendency to mistakenly

377 predict 'drought' when there is actually 'no drought'. The RF approach scores higher than
378 USDM in True negative and lower in True positive. The difference in score between True
379 positive ratio and True negative ratio is the smallest in the RF approach and the highest in EDDI.
380 Overall, the score of the RF approach is the least variable across the six performance metrics
381 among all the indicators. The RF approach gives fewer false alarms of droughts than the other
382 indicators and has the best success ratio. In comparison, USDM stands out in the 'threat score',
383 scoring slightly higher than the RF drought indicator.
384
385 PDSI shows poor performance overall. It was previously reported that monthly PDSI do
386 not capture droughts on short time scales, i.e. less than a year (Dai, 2017). SPI computed for 6
387 months accumulation period performed the best compared to the other examined accumulation
388 periods (i.e. 1, 3, 9 and 12), and its performance varies according to the drought cutoff. At a -0.8
389 cutoff, where droughts correspond to SPI ≤ -0.8, $SPI_{-0.8}$ scored low in True positive and threat
390 score, which indicates that $SPI_{-0.8}$ tends to miss droughts. This explains why $SPI_{-0.8}$ achieved a
391 near optimal score in the True negative metric. In contrast, at a 0 cutoff, SPI_0 scored low in True
392 negative and a near optimal score in True positive, which indicates that SPI_0 tends to predicts
393 drought when there is actually no drought.
394

395   3.3 RF drought probability maps

396 We built the final RF drought indicator for Texas on all event data without excluding a
397 proportion for validation. In Figure 2 and Figure 3, the purpose of training the RF algorithm on a
398 subset (75%) of the labelled of data was to validate the RF algorithm on unseen data and get a
399 robust estimate of the derived RF model. The RF drought indicator is then used to derive drought
400 probability maps for Texas.
401 In the following, we reference a Texas Climate Monthly Reports (TCMR) of a given month, for
402 example January 2010 as TCMR/1-2010, where the actual reference is
403 https://climatexas.tamu.edu/products/texas-climate-bulletins/january-2010.html.
404 We reference an impact report from the DIR database as DIR followed by its impact ID, e.g.
405 DIR4115.

406   3.3.1 The 2011 drought

407 We examined a drought episode over Texas during 2010-2012 (known as the 2011
408 drought) using drought probability maps derived by the new RF drought indicator for the period
409 spanning from January 2010 to April 2012. The 2011 drought was considered one of the most
410 catastrophic short-term droughts in the US and caused tremendous agricultural, hydrologic,
411 economic and socio-economic losses (Combs, 2014; Grigg, 2014). It was thought to be linked to
412 strong La Niña conditions in the Pacific which were established in the fall of 2010 and were
413 responsible for the below normal rain received during 2010-2012 (Folger et al., 2013; Texas
414 Water Development Board, 2012). The drought probability maps in Figure 4 illustrate how the
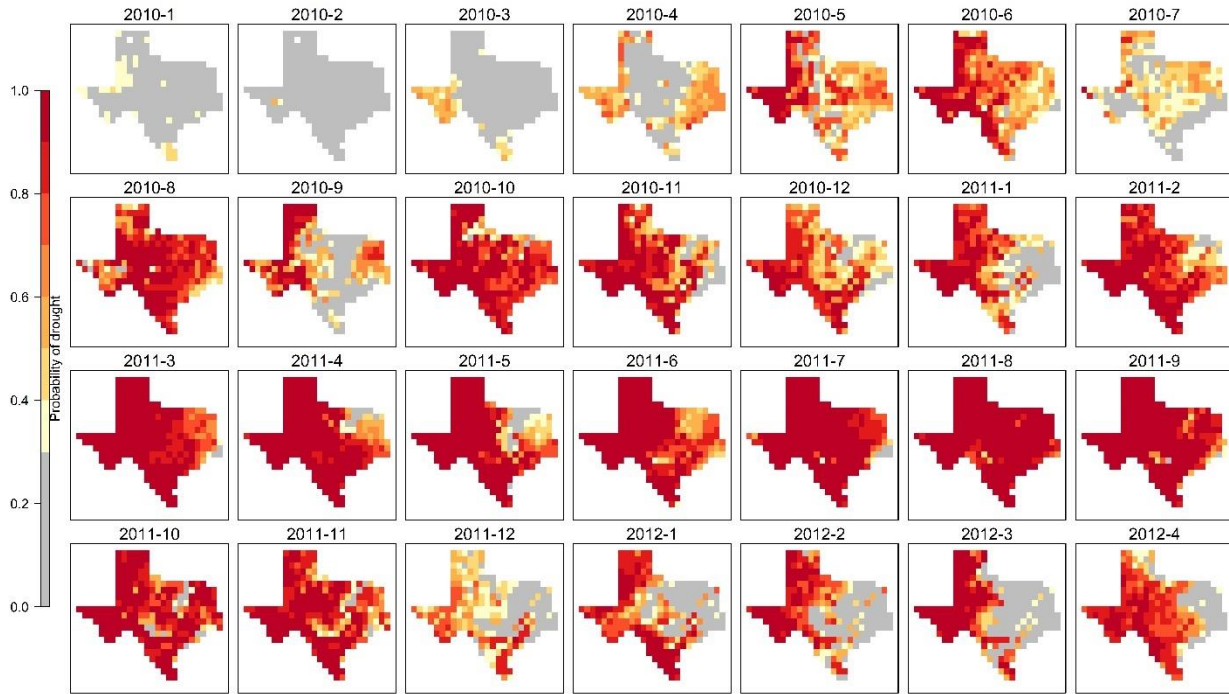415 2011 drought progressed in time and space throughout the examined period.

Figure 4: Drought probability maps predicted by RF during a drought episode.

Weather stations across Texas reported abundant precipitation during winter 2010 (TCMR/1-2010, TCMR/2-2010, TCMR/3-2010, TCMR/12-2010). As soon as the spring began, dry conditions were felt statewide. According to impact reports, dry conditions were reported in the South central plains, Western Gulf Coastal Plain (DIR4115) and Panhandle from March 2010. In the next months, dry conditions worsened and caused severe impacts on the growing season (DIR25697). The drought probability maps in Figure 4 show an increase in drought probability from April through June, starting in Panhandle, west and south Texas and expanding gradually to the entire state. The first half of July brought substantial rain (TCMR/7-2010) due to Hurricane Alex, which according the probability map has temporarily obliterated drought in most of Texas. The very dry and very hot August (TCMR/7-2010) appeared to have quickly wiped out the moisture brought by the wet spell in July; this is reflected in the increase in drought probabilities. In September 2010, a tropical storm brought significant rain along the Western Gulf Coastal Plain, Southern Texas Plains and East Central Texas plain (TCMR/9-2010), which as indicated in the September 2010 map temporarily broke the drought in these regions. Rain was also picked up by areas in the west and in the Panhandle, however, due to the very high temperatures, these areas were not relieved from the drought as observed in the drought probability map of September 2010. Very dry and very warm conditions returned in October (TCMR/10-2010) and quickly elevated drought probabilities. The drought areas, and many parts of Texas did not receive a single trace of rain. By the end of fall, drought exacerbated in Bastrop (DIR14853), Austin (DIR25214), Panhandle (DIR 3667), and many areas across the state were reported as natural disaster areas (DIR4115). The eastern part of the state experienced cold weather and rainy respite in January 2011 (TCMR/1-2011), which lowered the percentage of the land in drought. In February 2011, Texas experienced sub-zero temperatures with scarce precipitation (TCMR/2-2011), which put most of the state under drought. Probability maps show that drought conditions continued throughout Texas in March 2011. In April 2011, abnormally

444   dry and warm weather continued across the entire state. According to drought reports, since the
445   beginning of 2011, bushfires devastated thousands of acres almost everywhere (DIR4160, 4158,
446   4167, 3937, 4199, 4166, 4120, 4167). By April, the water level in lakes, wetlands and rivers had
447   reached very low levels (DIR3667, 4212, 25155), and voluntary and compulsory reduction in
448   water use was imposed in many areas across the state (DIR 24648, 3879). In April and May, the
449   Dallas region in northern Texas picked up drought breaking rains (TCMR/4-2011, TCMR/5-
450   2011) which helped reduce the probability of drought before the abnormally warm summer had
451   started (TCMR/6-2011). Drought continued during the summer causing more wildfires
452   (DIR4465) and tremendous losses in agriculture statewide (DIR29694, 26744, 4019, 4022,
453   14864, 3965). The drought persisted the entire 2011, however there were a few cold fronts that
454   brought important rain over many areas in the eastern part of the state (TCMR/11-2011) in
455   November, and the relieved areas experienced temporary decrease in drought probability during
456   that month. December 2011 was in general wetter than usual in most of the state except in the far
457   west (TCMR/12-2011). This is reflected in the significant decrease in drought probability during
458   this month. January 2012 was another wetter than usual month. Substantial rain was observed in
459   all weather stations except in the Panhandle, Rio Grande Valley and most of the Far West
460   (TCMR/1-2012). The drought probability maps for the months of January to April 2012 show a
461   drought free area stretching from the Central Great Plains to the South Central Plains.
462

463   3.3.2 Comparing RF drought indicator with EDDI and SPI indices in representing the
464   2011 drought

465   We assessed the agreement between the RF drought indicator and EDDI and SPI in
466   representing the 2011 drought during January 2010 and April 2012 using two metrics:
467   correlation and difference in drought onset. The correlation between the RF drought probabilities
468   and SPI is very strong everywhere (Figure 5a). In fact, unsurprisingly, precipitation was found to
469   be the most explanatory variable in discerning 'drought' and 'non drought' as described in more
470   detail below. Negative correlations were obtained because drought is denoted by negative values
471   in SPI and higher (positive) probabilities in RF. In comparison, the correlation between RF and
472   EDDI in Figure 5d is high (0.5-0.8) in the western half of the state but weakens in the eastern
473   half of the state, with the lowest correlation observed in the Cross Timbers regions.
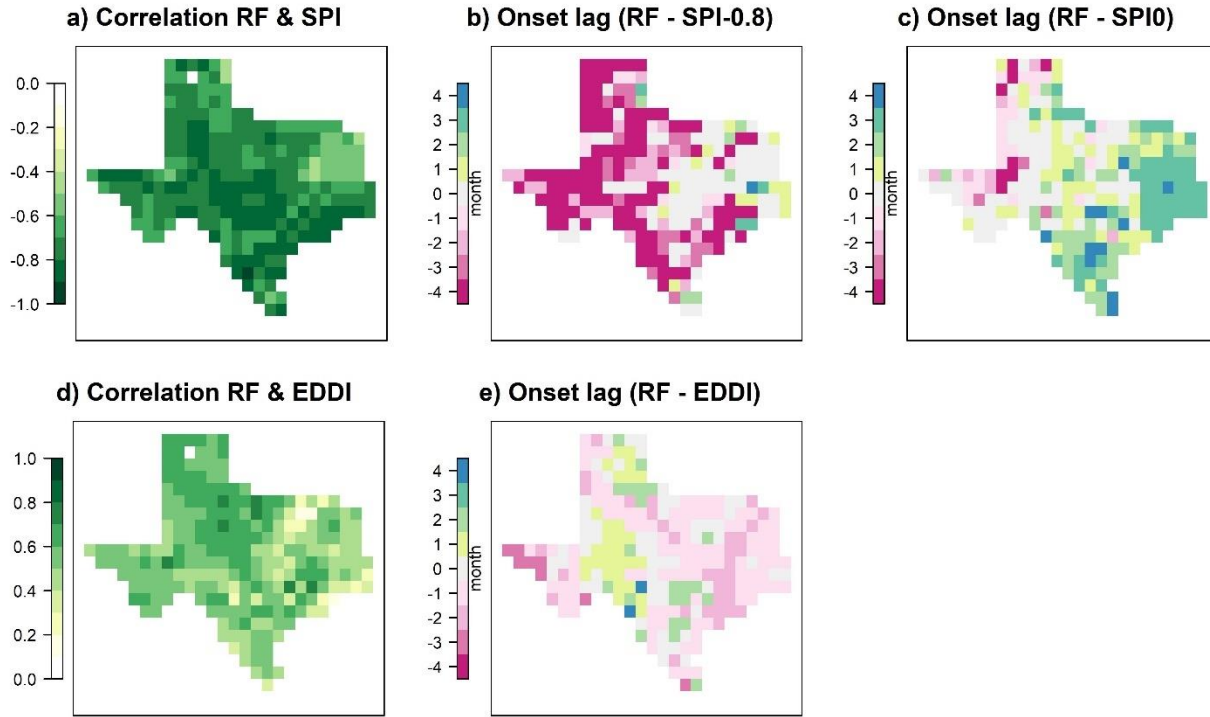
474
Figure 5: Correlation between RF drought probabilities and a) SPI, and d) EDDI. Difference in RF drought onset and each of b) SPI-6 with a drought threshold of -0.8 (i.e. $Onset_{RF} - onset_{SPI-0.8}$), c) SPI with a drought threshold of 0 (i.e. $onset_{RF} - onset_{SPI0}$) and e) EDDI (i.e. $onset_{RF} - onset_{EDDI}$). Correlations and onsets are computed for the period spanning January 2010 – April 2012.

We examined the difference in drought onset with SPI at the two drought thresholds and over several accumulation periods. Figure 5b and Figure 5c display the results for $SPI_{-0.8}$ and SPI_0 respectively, both computed for 1-month accumulation period. Drought appears in RF drought index well in advance of $SPI_{-0.8}$ across the dry western half of the state and the majority of the state. One finding from Figure 3 is that $SPI_{-0.8}$ tends to miss droughts, which according to Figure 5b results from a delayed start of droughts. In contrast, drought appears in RF after SPI_0 over the majority of the state, with the largest difference observed in the wettest part of the state. The reason is likely that SPI does not know how resilient the system is. For example, after several rainy months, water is abundant, and a month of abnormally low rain would not necessarily lead to a drought. While SPI accumulated over longer time periods than 1 month is likely to better capture the resilience of the system since it has longer *P* memory, at 1 month accumulation period the SPI has higher correlation with RF and a smaller drought onset difference (Figure S1 in the supplementary material). It has been reported that SPI computed for a short accumulation period is more suitable for use as a drought indicator for immediate impacts (European Comission, 2020). Figure 5b and c suggest that neither of the two drought thresholds is optimal, and a better threshold value is likely to be between 0 and −0.8.

Figure 5e shows that the drought appears in RF with a small lag of ±1 month compared to EDDI. RF shows drought emergence before EDDI in the majority of the state except areas in the west central and the southwest. Considering the low correlation in the wet parts of the state and

501 the low 'True negative' score achieved by EDDI in Figure 3, EDDI appears to not capture
502 drought dynamics under drought-breaking flash events such as tropical storms and hurricanes
503 that hit the eastern part of the state.
504
505   The RF drought indicator quantifies the probability of drought rather than its categorical
506 severity as in EDDI, SPI, PDSI and USDM. Drought probability represents the conditional
507 probability given the current climate (see section 2.3.2 for details). Monitoring drought
508 probabilities and how they are evolving in time allows for recognizing a drought before it occurs
509 (probability increases to near 0.5) or intensifies. We argue that drought probabilities provide a
510 more reliable quantification of drought than severity categories, as they are not based on
511 distribution assumptions nor are they computed in reference to a climatology. This is unlike the
512 other drought indicators which assume a fixed number of droughts (percentile) falling in each
513 drought category during a climatological period. Furthermore, the derived drought probabilities
514 take into account the interaction of a range of climate variables in the land-ocean-atmosphere
515 system that can influence droughts.
516

517   3.4 Importance of climate features in explaining droughts

518 We generated 100 RF models and computed the importance of each predictor variable as the
519 average of its conditional permutation importance across all forests. As described in section
520 2.3.2, the importance of a given predictor variable, for example ET, is the difference in
521 prediction accuracy before and after permuting ET averaged over all permutations. Table 2
522 shows the mean and the range of importance of each predictor variable across the 100 RF
523 models, and its ranking. All the variables appear to offer useful information to discern 'drought'
524 and 'no drought', since they all have non-zero importance. Also, as expected, precipitation is the
525 climate feature that provides the maximum information about drought, followed by ENSO and
526 SM. SM_{prev} comes next, its high importance is likely to come from its provision of moisture
527 memory and a signal of system resilience. ET and CWS empower drought predictions equally,
528 followed by PET and NDVI. The month feature was the least important variable.
529
530 **Table 2:** Importance of climate features in discerning 'drought' and 'no drought' measured using
531 conditional permutation scheme (Strobl et al. 2008). 'Mean' (Range) is the mean (range of)
532 importance computed across 100 generated RFs.
533

Importance Rank	1	2	3	4	5	6	7	8	9
Climate feature	P	ENSO	SM	SM_{prev}	ET	CWS	PET	NDVI	Month
Mean	0.089	0.069	0.058	0.0165	0.0073	0.0073	0.0058	0.0038	0.0028
Range	[0.084 – 0.096]	[0.066 – 0.073]	[0.053 – 0.064]	[0.0138 - 0.019]	[0.006 – 0.0088]	[0.0057 – 0.008]	[0.0045 – 0.0072]	[0.0027– 0.0052]	[0.002 – 0.0037]

534
535   Despite P being more important to drought than PET, PET anomalies can depict the
536 beginning of drought better than P anomalies, at least as embodied in EDDI and SPI
537 respectively, as inferred from Figure 5. One example of a situation where relying on P anomalies
538 can be misleading is when abnormally low precipitation occurs after several wet months. In this
539 case a drought will appear in the SPI signal, whereas in reality water is abundant and the lack of

540 rain will not necessarily lead to drought emergence. Another example is that abnormally high
541 PET can lead to drought even when precipitation is near normal (Lukas et al., 2017) in which
542 case, drought will not be indicated by SPI.
543

544   3.5 RF forecast models

545   In a further analysis, we use RF to build three forecast models – RF F1, RF F2 and RF F3
546 – that quantify drought 1, 2 and 3 months ahead, respectively. In the training process, each event
547 record consists of a label ('drought', 'no drought') observed at a month, and climate features
548 observed 1 (RF F1), 2 (RF F2) and 3 (RF F3) months before. We assess the predictive skill of
549 these forecast models following the same out-of-sample testing approach described in Section
550 2.5. Figure 6 illustrates the results of the out-of-sample performance of RF drought indicator and
551 each forecast model across 100 different testing datasets. The three forecast models score above
552 83% in 'Accuracy', 'True positive', 'True negative', and 'Success ratio' across the majority of
553 the out-of-sample testing, but as expected, could not beat the scores of the RF drought indicator
554 with concurrent predictor variables. These values are comparable or better than EDDI, PDSI or
555 SPI with concurrent predictor variables (Figure 3) and so offer hope for successful short-term
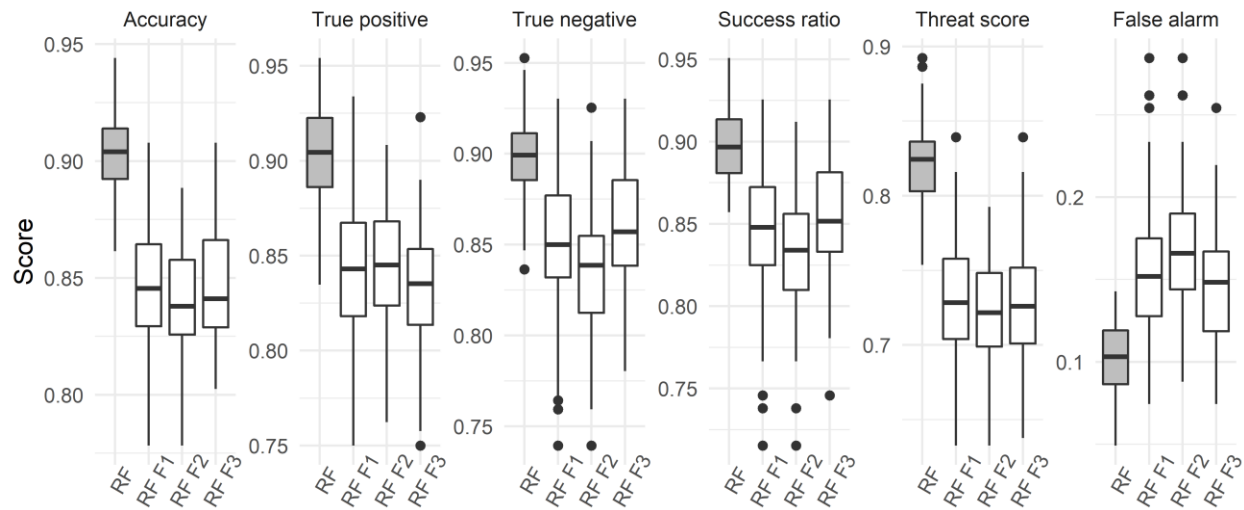556 predictive capacity.
557

558

559 **Figure 6:** Performance results of RF classifier and RF drought indicators, RF F1, RF F2 and RF
560 F3 at testing samples across 100 different sub-sampling of training and validating samples.
561 Performance scores are explained in section 2.5.
562

563   We also assessed how well the forecast models replicate the probability derived by the
564 RF drought indicators. For this analysis, we calculate 4 new performance metrics at each of the
565 testing events, and 100 testing datasets to measure the discrepancy of the forecast models with
566 the RF drought indicator. The employed metrics are root mean squared error (RMSE), standard
567 deviation (SD) difference, correlation and mean absolute bias. The results in Figure 7 show that
568 the discrepancy between forecasted drought probabilities and the actual drought probability
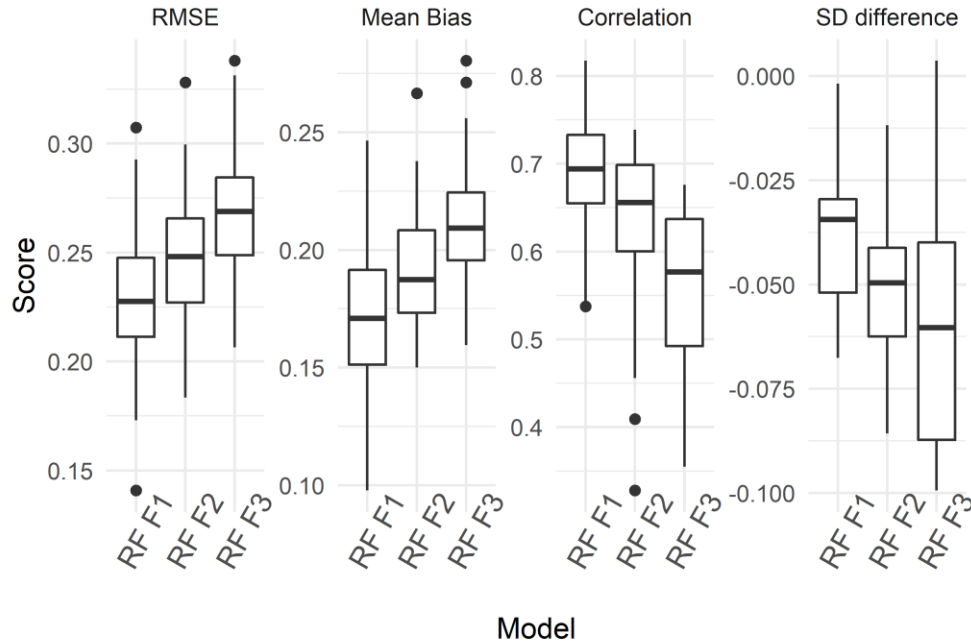569 slightly increases as the forecast period increases as expected.

Figure 7: Performance of the three forecast models RF F1, RF F2 and RF F3 relative to RF drought indicator.

Finally, Figure 8 shows maps of the correlation of the three forecast models with the RF drought indicator during the drought episode January 2010 – April 2012. Similar to our previous findings from Figure 7, correlation decreases as the forecast period increases, particularly in the wet east of the state. The lag in the drought onset is presented in Figure 8 d,e and f for RF F1, RF F2 and RF F3 respectively. The onset difference maps show that the onset lag is in the range ±1 in the west for all the three forecast models, whereas in the east the forecast models tend to delay drought as the forecast period increases.
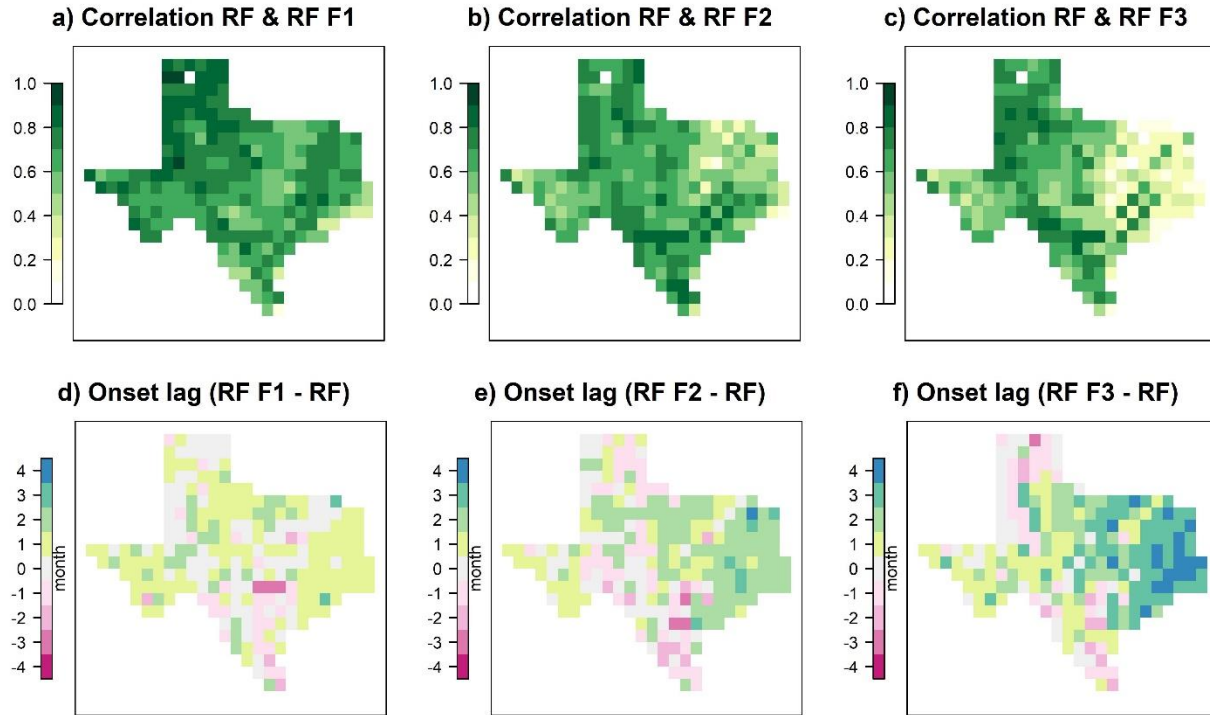
Figure 8: Correlation between RF drought probabilities and a) RF F1, b) RF F2, and b) RF F3. Difference in RF drought onset and each of d) RF F1, e) RF F2, and f) RF F3

5 Discussion

4.1 The new RF drought indicator versus USDM

The advanced capabilities of the RF approach and USDM in discerning 'droughts' and 'non droughts' compared to EDDI, SPI and PDSI highlight the importance of analysing the collective changes in climate features to better support drought quantification.

USDM is the current state of the art index of the weekly drought conditions in the U.S.; the new RF drought indicator provides a valuable counterpart to USDM for drought monitoring at the monthly scale. There are however several advantages in using the RF approach: a) the RF algorithm is developed once, then building drought probability maps from current climate data is an automated process. In comparison, deriving USDM maps is not automated as it incorporates subjective opinion and experts' interpretation; b) The spatial resolution of the RF drought indicator is 0.5° (or higher where finer resolution inputs are available), whereas USDM provides a big picture of the drought conditions over 10 Texan climate regions. The sparse resolution of USDM did not allow it to resolve droughts at the grid scale and resulted in prediction errors in the out-of-sample tests (Figure 3); c) USDM provides discrete drought categories, with limited ways for analysing them, and no clear method on how to aggregate them from weekly to other temporal scales (e.g. monthly). In comparison, the RF algorithm can be trained on data aggregated over several months and then applied to quantify droughts with longer time frames; d) The RF approach shows good forecast capabilities, while USDM does not have any forecast

605 capabilities. This is true both in terms of the lag models demonstrated here, and the applicability
606 of the RF approach to climate model projection data.

607 4.2 Transferability of the derived RF drought indicator to new regions

608 The new RF drought indicator was developed by training a RF algorithm on patterns
609 within the Texan region. Therefore, the particular RF drought indicator derived here is specific to
610 Texas and should not be used to monitor and quantify droughts in new locations outside Texas.
611 Clearly the physical processes linked with the initiation and persistence of drought are different
612 over different regions around the world. One obvious example is that droughts in Texas are
613 related to the cold phase of ENSO, whereas in many regions on land, droughts are related to the
614 warm phase of ENSO (i.e. El Niño, e.g. Australia). However, the approach is entirely portable,
615 assuming new RF models are developed for new locations and historical drought data of
616 sufficient quantity and reliability exist in those locations.
617

618 4.3 Future research directions

619 There are a number of key processes linked with the initiation and persistence of drought
620 that could be incorporated to improve the predictive skills of the RF drought indicator but were
621 not included here, for example zonal moisture advection (Erfanian & Fu, 2019). Nevertheless, as
622 new relevant climate variables become available, it is easy to test their ability to improve
623 predictions, and if justified, incorporate them as additional predictors.
624

625 We used a random forest to generate spatial predictions of drought. However, the spatial
626 location of points was ignored in the modeling process, so that spatial autocorrelation was not
627 accounted for. Hengl et al. (2018) developed a new framework called Random Forest for spatial
628 data (RFsp) that extends RF to account for spatial dependence. The RFsp framework
629 incorporates distances from observation points as predictor variables and therefore, adds
630 geographical proximity effects into the prediction process. More recently, (Georganos et al.,
631 2019) developed a novel geographical implementation of RF, named Geographical Random
632 Forest (GRF) that addresses spatial heterogeneity by disaggregating RF into geographical space
633 in the form of local sub-models. GRF is implemented in the R package SpatialML
634 (http://lctools.science/). We anticipate that applying any of the RFsp or the GRF approach in the
635 future will further improve the performance of the RF drought indicators and the predictive skills
636 of the RF forecasting models. It is important to note that both approaches require a larger number
637 of grid cells than what was used here.
638

639 Another topic for future research is using deep learning as an alternative, and more
640 powerful approach than RF to capture the spatio-temporal characteristics of droughts (Reichstein
641 et al., 2019). A few studies implemented deep learning for drought quantification (e.g. Deo and
642 Şahin, 2015; Shen et al. 2019). These studies used drought indicators as spatially and temporally
643 continuous labels. However, this approach is not optimal as drought indicators suffer from biases
644 and should not be used as 'ground-truth' labels. Given the absence of spatially and temporally
645 continuous drought data, using deep learning to quantify droughts remains challenging.

6 Conclusions

In contrast to most scientific drought metrics, in this work we used recorded drought impacts as our observational definition of drought, and used a random forest model to establish an empirical relationship between drought impact and a broad range of drought-related climate predictors. This approach was able to predict unseen drought impact events with far greater success than existing climate-variable based drought metrics, such as SPI, PDSI or EDDI, and performed as well out-of-sample as the assimilated drought product USDM. However, unlike USDM, the approach offers considerable predictive ability, both in the short-term drought predictions and use with climate projections. While Texas was used as a test case here, the approach is applicable to any region with sufficient spatiotemporal drought records.

Acknowledgments

References

AMS. (1997). Policy statement. *Society, Bulletin of the American Meteorological*, *78*(5), 847–852. https://doi.org/https://doi.org/10.1175/1520-0477-78.5.847

AWG. (2004). Creating a drought early warning system for the 21st century: The national integrated drought information system. Western Governors Report. Denver, CO: Western Governors Association.

Azmi, M., Ruediger, C., & Walker, J. P. . (2016). A data fusion-based drought index. *Water Resources Research*, *52*, 2222–2239. https://doi.org/10.1002/2015WR017834

Bachmair, S., Kohn, I., & Stahl, K. (2015). Exploring the link between drought indicators and impacts. *Natural Hazards and Earth System Sciences*, *15*(6), 1381–1397. https://doi.org/10.5194/nhess-15-1381-2015

Balakrishnama, S., & Ganapathiraju, A. (1998). Linear discriminant analysis-a brief tutorial. *Institute for Signal and Information Processing*, *18*(1998), 1–8.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/doi.org/10.1023/A:1010933404324

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. CRC press.

Brown, J. F., Wardlow, B. D., Tadesse, T., Hayes, M. J., & Reed, B. C. (2008). The Vegetation Drought Response Index (VegDRI): A New Integrated Approach for Monitoring Drought Stress in Vegetation. *GIScience & Remote Sensing*, *45*(1), 16–46. https://doi.org/10.2747/1548-1603.45.1.16

Combs, S. (2014). Going deeper for the solution. *Texas Water Report*.

Cutler, D. R., Edwards Jr, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, *88*(11), 2783–2792.

Dai, A. N. C. for A. R. S. (Eds). (2017). The Climate Data Guide: Palmer Drought Severity Index (PDSI). Retrieved January 21, 2021, from https://climatedataguide.ucar.edu/climate-data/palmer-drought-severity-index-pdsi

Davis, T. W., Prentice, I. C., Stocker, B. D., Thomas, R. T., Whitley, R. J., Wang, H., et al. (2017). Simple process-led algorithms for simulating habitats (SPLASH v. 1.0): robust indices of radiation, evapotranspiration and plant-available moisture. *Geoscientific Model Development*, *10*(2), 689–708.

Deo, R. C., & Şahin, M. (2015). Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia. *Atmospheric Research*, *153*, 512–525. https://doi.org/10.1016/j.atmosres.2014.10.016

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., et al. (2017). ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sensing of Environment*, *203*, 185–215.

Erfanian, A., & Fu, R. (2019). The role of spring dry zonal advection in summer drought onset over the US Great Plains. *Atmospheric Chemistry & Physics*, *19*(24).

European Comission. (2020). *EDO indicator Factsheet: Standardized Precipitation Index (SPI)*. Retrieved from https://edo.jrc.ec.europa.eu/

Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, *15*, 3133–3181. https://doi.org/10.1117/1.JRS.11.015020

Fernando, D. N., Chakraborty, S., Fu, R., & Mace, R. E. (2019). A process-based statistical seasonal prediction of May–July rainfall anomalies over Texas and the Southern Great Plains of the United States. *Climate Services*, *16*(September 2018), 100133. https://doi.org/10.1016/j.cliser.2019.100133

Folger, P., Cody, B. A., & Carter, N. T. (2013). Drought in the United States: Causes and issues for Congress. *Selected Issues in Water Resources and Management*, 32–78.

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, 1–67.

Georganos, S., Grippa, T., Niang Gadiaga, A., Linard, C., Lennert, M., Vanhuysse, S., et al. (2019). Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto International*, *36*(2), 121–136. https://doi.org/10.1080/10106049.2019.1595177

Grigg, N. S. (2014). The 2011–2012 drought in the United States: new lessons from a record event. *International Journal of Water Resources Development*, *30*(2), 183–199.

Gruber, A., Dorigo, W. A., Crow, W., & Wagner, W. (2017). Triple collocation-based merging of satellite soil moisture retrievals. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(12), 6780–6792.

Gruber, A., Scanlon, T., van der Schalie, R., Wagner, W., & Dorigo, W. (2019). Evolution of the ESA CCI Soil Moisture climate data records and their underlying merging methodology. *Earth System Science Data*, 1–37.

Heim, R. R. (2002). A Review of Twentieth-Century Drought Indices Used in the United States. *Bulletin of the American Meteorological Society*, (August), 1149–1166.

Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M., & Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, *2018*(8). https://doi.org/10.7717/peerj.5518

Hobbins, M. T., Wood, A., McEvoy, D. J., Huntington, J. L., Morton, C., Anderson, M., & Hain, C. (2016). The evaporative demand drought index. Part I: Linking drought evolution to variations in evaporative demand. *Journal of Hydrometeorology*, *17*(6), 1745–1761. https://doi.org/10.1175/JHM-D-15-0121.1

Hobeichi, S. (2020). Derived Optimal Linear Combination Evapotranspiration - DOLCE v2.1.

735   CLEX. https://doi.org/10.25914/5f1664837ef06

736 Hobeichi, S., Abramowitz, G., & Evans, J. (2020). Robust historical evapotranspiration trends
737   across climate regimes. *Hydrol. Earth Syst. Sci. Discuss*, (December), 1–32.
738   https://doi.org/doi.org/10.5194/hess-2020-595

739 Humphrey, V., & Gudmundsson, L. (2019). GRACE-REC: a reconstruction of climate-driven
740   water storage changes over the last century. *Earth System Science Data Discussions*, 1–41.
741   https://doi.org/10.5194/essd-2019-25

742 IPCC. (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II*
743   *and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*
744   *[Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]*. Geneva, Switzerland,.

745 Keyantash, J., & Dracup, J. A. . (2002). The quantification of drought : An evaluation of drought
746   indices. *American Meteorological Society*, *BAMS*, 1167–1180.

747 Keyantash, J. A., & Dracup, J. A. (2004). An aggregate drought index: Assessing drought
748   severity based on fluctuations in the hydrologic cycle and surface water storage. *Water*
749   *Resources Research*, *40*(9).

750 Khan, N., Sachindra, D. A., Shahid, S., Ahmed, K., Shiru, M. S., & Nawaz, N. (2020). Prediction
751   of droughts over Pakistan using machine learning algorithms. *Advances in Water*
752   *Resources*, *139*(January), 103562. https://doi.org/10.1016/j.advwatres.2020.103562

753 Kruppa, J., Liu, Y., Diener, H., Holste, T., Weimar, C., König, I. R., & Ziegler, A. (2014).
754   Probability estimation with machine learning methods for dichotomous and multicategory
755   outcome: Applications. *Biometrical Journal*, *56*(4), 564–583.

756 Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical*
757   *Software*, *28*(5), 1–26.

758 Li, Q., Li, P., Li, H., & Yu, M. (2015). Drought assessment using a multivariate drought index in
759   the Luanhe River basin of Northern China. *Stochastic Environmental Research and Risk*
760   *Assessment*, *29*(6), 1509–1520.

761 Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3),
762   18–22.

763 Van Loon, A. F., & Van Lanen, H. A. J. (2012). A process-based typology of hydrological
764   drought. *Hydrology and Earth System Sciences*, *16*(7), 1915.

765 Lukas, J., Hobbins, M., & Rangwala, I. (2017). *The EDDI User Guide v1.0*. Retrieved from
766   https://www.esrl.noaa.gov/psd/eddi/pdf/EDDI_UserGuide_v1.0.pdf

767 Malley, J. D., Kruppa, J., Dasgupta, A., Malley, K. G., & Ziegler, A. (2012). Probability
768   machines: consistent probability estimation using nonparametric learning machines.
769   *Methods of Information in Medicine*, *51*(1), 74–81. https://doi.org/10.3414/ME00-01-0052

770 McGovern, A., Elmore, K. L., Gagne, D. J., Haupt, S. E., Karstens, C. D., Lagerquist, R., et al.
771   (2017). Using artificial intelligence to improve real-time decision-making for high-impact
772   weather. *Bulletin of the American Meteorological Society*, *98*(10), 2073–2090.

773 McKee, T. B. (1995). Drought monitoring with multiple time scales. In *Proceedings of 9th*
774   *Conference on Applied Climatology, Boston, 1995*.

775 McKee, T. B., Doesken, N. J., & Kleist, J. (1993). The relationship of drought frequency and
776   duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology* (Vol.
777   17, pp. 179–183). Boston.

778 Mitchell, T. M. (1997). Does machine learning really work? *AI Magazine*, *18*(3), 11.

779 Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal*
780   *Statistical Society: Series A (General)*, *135*(3), 370–384.

781   Palmer, W. C. (1965). Meteorological drought. *Office of Climatology, US Department of*
782   *Commerce, Washington DC*, *45*, 58 pp.
783   Park, S., Im, J., Jang, E., & Rhee, J. (2016). Drought assessment and monitoring through
784   blending of multi-sensor indices using machine learning approaches for different climate
785   regions. *Agricultural and Forest Meteorology*, *216*, 157–169.
786   Pinzon, J. E., & Tucker, C. J. (2014). A non-stationary 1981–2012 AVHRR NDVI3g time series.
787   *Remote Sensing*, *6*(8), 6929–6960.
788   Pu, B., Fu, R., Dickinson, R. E., & Fernando, D. N. (2016). Why do summer droughts in the
789   Southern Great Plains occur in some La Niña years but not others? *Journal of Geophysical*
790   *Reseach: Atmospheres*, *121*, 1120–1137. https://doi.org/10.1002/2015JD023508
791   Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat.
792   (2019). Deep learning and process understanding for data-driven Earth system science.
793   *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1
794   Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P.
795   (2012). An assessment of the effectiveness of a random forest classifier for land-cover
796   classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *67*, 93–104.
797   Sandri, M., & Zuccolotto, P. (2008). A bias correction algorithm for the gini variable importance
798   measure in classification trees. *Journal of Computational and Graphical Statistics*, *17*(3),
799   611–628. https://doi.org/10.1198/106186008X344522
800   Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., & Ziese, M. (2018). GPCC Full
801   Data Monthly Product Version 2018 at 0.5°: Monthly land-surface precipitation from rain-
802   gauges built on GTS-based and historical data. *Global Precipitation Climatology Centre*.
803   Retrieved from https://psl.noaa.gov/data/gridded/ data.gpcc.html
804   Scholkopf, B., Sung, K.-K., Burges, C. J. C., Girosi, F., Niyogi, P., Poggio, T., & Vapnik, V.
805   (1997). Comparing support vector machines with Gaussian kernels to radial basis function
806   classifiers. *IEEE Transactions on Signal Processing*, *45*(11), 2758–2765.
807   Schubert, S. D., Suarez, M. J., Pegion, P. J., Koster, R. D., & Bacmeister, J. T. (2004). Causes of
808   long-term drought in the US Great Plains. *Journal of Climate*, *17*(3), 485–503.
809   Seager, R., Goddard, L., Nakamura, J., Henderson, N., & Lee, D. E. (2014). Dynamical causes of
810   the 2010/11 Texas-northern Mexico drought. *Journal of Hydrometeorology*, *15*(1), 39–68.
811   https://doi.org/10.1175/JHM-D-13-024.1
812   Seneviratne, S. I ., Nicholls, N ., Easterling, D ., Goodess, C. M ., Kanae, S ., Kossin, J ., et
813   al. (2012). *Changes in climate extremes and their impacts on the natural physical*
814   *environment. In: Managing the Risks of Extreme Events and Disasters to Advance Climate*
815   *Change Adaptation [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi,*
816   *M.D. Mastr*. Cambridge, UK, and New York, NY, USA.
817   Shen, R., Huang, A., Li, B., & Guo, J. (2019). Construction of a drought monitoring model using
818   deep learning based on multi-source remote sensing data. *International Journal of Applied*
819   *Earth Observation and Geoinformation*. https://doi.org/10.1016/j.jag.2019.03.006
820   Smith, C. A., & Sardeshmukh, P. D. (2000). The effect of ENSO on the intraseasonal variance of
821   surface temperatures in winter. *International Journal of Climatology: A Journal of the*
822   *Royal Meteorological Society*, *20*(13), 1543–1557.
823   Soh, Y. W., Koo, C. H., Huang, Y. F., & Fung, K. F. (2018). Application of artificial intelligence
824   models for the prediction of standardized precipitation evapotranspiration index (SPEI) at
825   Langat River Basin, Malaysia. *Computers and Electronics in Agriculture*, *144*, 164–173.
826   https://doi.org/https://doi.org/10.1016/j.compag.2017.12.002

827 Stagge, J. H., Tallaksen, L. M., Gudmundsson, L., Van Loon, A. F., & Stahl, K. (2015).
828   Candidate distributions for climatological drought indices (SPI and SPEI). *International*
829   *Journal of Climatology*, *35*(13), 4027–4040.

830 Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable
831   importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, *8*.
832   https://doi.org/10.1186/1471-2105-8-25

833 Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable
834   importance for random forests. *BMC Bioinformatics*, *9*(1), 307.
835   https://doi.org/10.1186/1471-2105-9-307

836 Svoboda, M., LeComte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., et al. (2002). The
837   drought monitor. *Bulletin of the American Meteorological Society*, *83*(8), 1181–1190.

838 Swain, P. H., & Hauska, H. (1977). The decision tree classifier: Design and potential. *IEEE*
839   *Transactions on Geoscience Electronics*, *15*(3), 142–147.

840 Texas Water Development Board. (2012). Chapter 4 Climate of Texas. *Water For Texas 2012*
841   *State Water Plan*, 145–155. Retrieved from
842   https://www.mendeley.com/viewer/?fileId=0ec0c4ff-06ba-832f-0042-
843   1a4fa3e46231&documentId=0abfb413-0a9d-3fa4-8c83-87957a3263c7

844 Walter, I. A., Allen, R. G., Elliott, R., Jensen, M. E., Itenfisu, D., Mecham, B., et al. (2000).
845   ASCE's standardized reference evapotranspiration equation. In *Watershed management and*
846   *operations management 2000* (pp. 1–11).

847 Wilhite, D. A. (2009). Drought monitoring as a component of drought preparedness planning. In
848   *Coping with drought risk in agriculture and water supply systems* (pp. 3–19). Springer.

849 Wilhite, D. A., & Glantz, M. H. (1985). Understanding: The drought phenomenon: The role of
850   definitions. *Water International*, *10*(3), 111–120.
851   https://doi.org/10.1080/02508068508686328

852 Wilhite, D. A., Svoboda, M. D., & Hayes, M. J. (2007). Understanding the complex impacts of
853   drought: A key to enhancing drought mitigation and preparedness. *Water Resources*
854   *Management*, *21*(5), 763–774. https://doi.org/10.1007/s11269-006-9076-5

855 Wright, M. N., & Ziegler, A. (2017). ranger: A fast implementation of random forests for high
856   dimensional data in C++ and R. *Journal of Statistical Software*, *77*, 1–17.
857   https://doi.org/10.18637/jss.v077.i01

858 Yihdego, Y., Vaheddoost, B., & Al-weshah, R. A. (2019). Drought indices and indicators
859   revisited. *Arabian Journal of Geosciences*, *12:69*, 12.

860 Zhang, A., & Jia, G. (2013). Monitoring meteorological drought in semiarid regions using multi-
861   sensor microwave remote sensing data. *Remote Sensing of Environment*, *134*, 12–23.

862 Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of*
863   *the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301–320.

864