



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

**Season-Ahead Forecasting of Water Storage and Irrigation
Requirements**

An Application to the Southwest Monsoon in India

Arun Ravindranath^{1*}, Naresh Devineni¹, Upmanu Lall², Paulina Concha Larrauri³

¹Department of Civil Engineering , Center for Water Resources and Environmental Research (City Water Center),
NOAA Center for Earth System Sciences and Remote Sensing Technologies , City University of New York (City
College), New York, NY 10031

²Department of Earth and Environmental Engineering , Columbia Water Center, The Earth Institute, Columbia
University, New York, NY 10027, USA

³Columbia Water Center, The Earth Institute, Columbia University, New York, NY 10027, USA

Short title: A novel crop water stress index

* - Contact Author's email: aravind000@citymail.cuny.edu



29 **Abstract**

30 Water risk management is perhaps the most ubiquitous challenge a stakeholder in the
31 water or agricultural sector faces. We present a methodological framework for forecasting water
32 storage requirements and present an application of this methodology to risk assessment in India.
33 The application focused on forecasting crop water stress for potatoes grown during the monsoon
34 season in the Satara district of Maharashtra. Pre-season large-scale climate predictors used to
35 forecast water stress were selected based on an exhaustive search method that evaluates for
36 highest Rank Probability Skill Score and lowest Mean Squared Error in a leave-one-out cross
37 validation mode. Adaptive forecasts were made over the years 2001 through 2013 using the
38 identified predictors and a semi-parametric k-nearest neighbors approach. The accuracy of the
39 adaptive forecasts (2001-2013) was judged based on directional concordance and contingency
40 metrics such as hit/miss rate and false alarms. Based on these criteria, our forecasts were correct
41 nine out of thirteen times, with two misses and two false alarms. The results of these drought
42 forecasts were compared with precipitation forecasts from the Indian Meteorological Department
43 (IMD). We assert that it is necessary to couple informative water stress/risk indices with an
44 effective forecasting methodology to maximize the utility of such indices, thereby optimizing
45 water management decisions.

46
47 **Keywords:** Crop stress, water risk, seasonal forecasts, climate-information, deficit, monsoon
48 prediction, contract farming, agricultural drought risk

49

50

51

52

53

54

55

56

57

58

59

60

61

62



63 1. Introduction

64 Monitoring and forecasting systems can aid in pinpointing mitigation tactics for water
65 security and water resources management. There is a continued interest in forecasting and
66 monitoring systems that can inform planners and decision-makers in various water-dependent
67 sectors at sufficient lead times and with increasingly higher levels of accuracy and reliability.
68 The agricultural sector is perhaps the greatest example of this, being a heavily water-dependent
69 sector that serves as the economic backbone of a country. The agricultural sector consumes
70 more freshwater than any other economic sector, with an estimated 1,300 m³/cap/yr needed to
71 maintain an adequate diet (Rockstrom et al., 2009). Significant increases of water will be
72 required to produce food by 2050, ranging from 8,500 to 11,000 km³/yr, depending on to what
73 extent rainfed and irrigated agricultural systems improve (Rockstrom et al., 2009). Additionally,
74 to maintain high yields, irrigation will continue to be an important buffer to climate shocks. This
75 is especially true when one considers that almost all of the world's major agricultural lands are
76 located in the most drought-prone areas of the world (Mishra and Desai, 2006). Hence,
77 developing forecasting techniques to improve how we address irrigation requirements, water
78 storage requirements and crop water stress is a major step in dealing with the larger issue of
79 water resources management at local, regional and global scales. The present study focuses on
80 forecasting water storage and irrigation requirements in the agricultural sector as one important
81 dimension to the larger issue of drought forecasting and water resources management, with an
82 application of such forecasting to the monsoonal climate of India.

83
84 Existing forecasts either deal directly with basic hydrologic or meteorological variables,
85 such as precipitation, temperature and soil moisture, or they work with proxies of drought, often
86 in the form of indices such as the Standardized Precipitation Index, or SPI (McKee et al, 1993),
87 the Palmer Drought Severity Index, or PDSI (Palmer, 1965), the Standardized Precipitation
88 Evapotranspiration Index, or SPEI (Serrano et al, 2010), and the Normalized Difference
89 Vegetation Index, or NDVI, among others. A comprehensive list of indices used in drought
90 forecasting can be found in Heim (2002), Mishra and Singh (2010) and Liu and Pan (2016). The
91 forecast of basic variables requires subsequently integrating these forecasts into a product that
92 can estimate water storage or irrigation requirements, as these variables do not immediately
93 divulge such information. This represents a challenge by itself. In light of this limitation, in this
94 paper, we present a crop water stress index that is defined and constructed based on work by
95 Devineni et al (2013). The advantage of this particular index, hereby known as the cumulative
96 deficit index (CDI), is that it accounts for the variability in water supply and demand while
97 incorporating information specific to a particular crop of interest. CDI is derived by
98 accumulating differences in supply (rainfall) and demand (crop water requirement), and with
99 very few crop input parameters. The CDI is a determinant of water stress faced by the crop and
100 hence of the dependence of the crop yield on water availability. It can be interpreted as the water
101 that is required from external storage beyond rainfall to meet demand (Devineni et al, 2013;
102 Devineni et al, 2015). Therefore, the index directly informs water storage and irrigation
103 requirements.

104
105 The primary focus of this paper will be on exploring the possibility of providing forecasts
106 for CDI by investigating the sources of predictability and developing statistically verifiable
107 models for the season-ahead probabilistic forecasts. Significant crop water deficits can adversely
108 impact the crop production or water reserves and lead to high-energy costs for pumping



109 groundwater for irrigation to maintain yield. The seasonal forecasting of CDI provides a way for
110 institutional planning and action in this context to reduce the climate-related water risks in
111 agriculture, which is one of the largest consumers of water. An application of CDI forecasting is
112 presented for the state of Maharashtra in India to verify whether advance reliable forecasts for
113 potato-based CDI can be developed. A semi-parametric k-nearest neighbor (kNN) bootstrapping
114 algorithm as described in Lall and Sharma (1996) is employed for forecasting CDI using pre-
115 season large-scale climate indices. This is a simple probabilistic forecasting procedure that
116 captures uncertainty. We examine these forecasts and suggest ways of interpreting them in a
117 manner that can aid stakeholders in the agricultural water resources sector in addressing the
118 fundamental questions about irrigation and water storage requirements. These forecasts will then
119 be compared to precipitation forecasts for the same season in the same area of India as given by
120 the Indian Meteorological Department (IMD).

121

122 In section 2, we present a survey of the existing forecasting systems in monsoonal
123 climates and their skill and limitations. In section 3, we discuss the background and scientific
124 basis of CDI, including its explicit formulation and governing equations. In section 4, we get
125 into a thorough description of the case study and all steps involved, including background
126 information relating to the case study and location, data collection and processing, a complete
127 description of the forecasting model, methods and predictor selection scheme. Section 5 presents
128 the results of the forecast, a discussion of these results and their implications, and a comparison
129 of our results with those of IMD. Finally, section 6 summarizes and concludes the paper.

130

131 **2. A Brief Review of the Current Forecasting Systems for Water Management in** 132 **Monsoonal Climates**

133 A number of forecasting methodologies have been proposed or developed for water
134 management and agricultural planning. Shah and Mishra (2016) investigated the goodness of the
135 Global Ensemble Forecast System (GEFS) for generating medium-range (~7 day) drought
136 forecasts in India, and found that the GEFS has higher forecasting skill during the non-monsoon
137 season than monsoon season for both temperature and precipitation, largely due to intraseasonal
138 variability during the monsoon season. This forecasting system tends to forecast temperature
139 variables with higher skill than precipitation and has variable skill according to region. Hence,
140 there is sensitivity to intraseasonal variation, which monsoon climates are notorious for, and
141 regional variation as well. Mishra and Desai (2005) used well-chosen linear stochastic models
142 (ARIMA) to forecast SPI- 3, 6, 9, 12, and 24 as a drought proxy in the Kansabati River Basin, an
143 important source of water for irrigation and an area in which crops are grown, in the Purulia
144 district of West Bengal, India at lead times of 1, 2, 3, 4, 5 and 6 months. Highest skill, as
145 measured by the correlation coefficient between observed and model-predicted SPI series,
146 occurred at shorter lead times, with correlation values between 0.799 and 0.925 depending on
147 which SPI series was forecasted. Asoka and Mishra (2015) forecasted vegetation anomalies (as
148 NDVI) at the regional scale as a proxy of vegetation health, and thus moisture availability. The
149 model used NDVI, root-zone soil moisture, and sea surface temperature (SST) at one to three
150 months lead time to develop the vegetation anomaly forecast, and skill was highest at one month
151 lead time and much lower for two and three months lead time as measured in a validation phase
152 by examining the R^2 statistic and by plotting the observed NDVI against the model-interpolated
153 series for one-, two-, and three- month lead times. Skill also varied based on location in space
154 and, in particular, was lower during the monsoon season (JJAS) likely due to the effect of



155 intraseasonal variability of the monsoon system on agricultural practices. Belayneh and
156 Adamowski (2012), in the interest of drought forecasting, forecasted SPI 3 and SPI 12 over lead
157 times of one and six months in the Awash River Basin in Ethiopia using Artificial Neural
158 Network, Wavelet Neural Network and Support Vector Regression models and similarly found
159 that forecast skill was higher at the shorter lead time. Kar et al (2012) considered Multi-Model
160 Ensemble (MME) methods in both a deterministic and probabilistic context. It was found that
161 the individual member models showed poor skill in simulating monsoon interannual variability
162 and that on average spatially, a MME scheme that uses the member models as predictors in a
163 point-by-point multiple regression as a means of averaging the member model forecasts
164 outperforms the other schemes mentioned in the paper in forecasting precipitation. However, it
165 was found that even here, none of the three MME schemes had any usable skill in a certain
166 region of India, and it was concluded that a probabilistic system would work better. When
167 probabilistic forecasts were generated (probabilistic MME) and evaluated for skill, RPSS was
168 positive for the best scheme, in only the northern most parts of India and a few scattered points
169 in north and central India. Finally, Shah et al (2017) examined how different forecast products
170 can be used operationally to provide hydrologic forecasts (e.g. for precipitation, temperature) for
171 India at a 7 – 45 day accumulation period, which is critical for agricultural and water resource
172 planning. Forecast skill was evaluated on the basis of correlation with observations, median
173 absolute error (MAE) and the critical success index (CSI). Four forecast products from Indian
174 Institute of Tropical Meteorology (IITM) were compared with Climate Forecast System version
175 2 (CFSv2) and Global Ensemble Forecast System version 2 (GEFSv2) forecast products, and it
176 was found that the meteorological variables predicted from the IITM products showed superior
177 skill for all accumulation periods. The key point here is that the IITM ensemble is postulated to
178 capture intraseasonal variability of rainfall during the monsoon season.

179 As an alternative to these agricultural planning measures, we introduce a new seasonal
180 crop water stress index that is more informative than the total rainfall measure. It gives a
181 surrogate for irrigation water required and incorporates intraseasonal rainfall and temperature
182 variability along with information inherent to the specific crop and planting region.

183

184 **3. The Cumulative Deficit Index: Background and Scientific Basis**

185 Our interest in this study is to provide one-season-ahead forecasts of irrigation and water
186 storage requirements for water resources management in the agricultural sector, and
187 subsequently compare the outcomes of these forecasts with the forecasts issued by IMD. We
188 begin by developing an index for crop water stress as a means of gauging irrigation
189 requirements. The index developed and used in this study computes the maximum cumulative
190 deficit over a growing season between daily water requirement for optimal crop growth and daily
191 effective rainfall. Variants of this method have been presented in our previous studies for
192 quantifying the water stress globally (Devineni et al, 2013; Devineni et al, 2015; Chen et al,
193 2014), and drought indexing for the United States (Etienne et al, 2016; Ho et al, 2016). Given an
194 n -year record of daily data, our water stress index calculates the day-by-day accumulation of
195 deficit in rainfall in each of the n growing seasons. The maximum of these seasonal daily deficit
196 values is taken to be the value of the index for the season. Hence, we give this index the name
197 *cumulative deficit index*, abbreviated CDI. On a practical level, such an index gives a worst-
198 case scenario in terms of the seasonal water stress on the crop, and can therefore be interpreted as
199 the amount of water that should be drawn from external storage to meet water demand. This



200 may include irrigation, ground water pumping, interbasin transfers, and/or withdrawing water
201 from a storage or water-harvesting facility.

202 Deficit is estimated as the difference between the seasonal crop water requirement and
203 effective rainfall for each crop in a given location in the season. Effective rainfall is given as
204

$$205 \quad S_{t,d} = \alpha * P_{t,d} \dots (1)$$

206 In Eq. (1), $P_{t,d}$ is the rainfall for a given day d in the year t . α is the parameter that determines
207 the fraction of rainfall that can be utilized by the crops for a location. It accounts for losses to
208 direct runoff, evaporation and groundwater infiltration. In our study, we set $\alpha = 0.7$.

209 The water use for a given crop is estimated based on the expected growth stage and daily
210 evapotranspiration as

$$211 \quad D_{t,d} = k_{c,d} * ET_{0,t,d} \dots (2)$$

212
213 In Eq. (2), $k_{c,d}$ is the crop coefficient, which is the ratio of actual evapotranspiration (ET_a) of a
214 given crop under non-stressed conditions to reference crop evaporation (ET_0). It represents crop-
215 specific water use at various growth stages of the crop and is typically derived empirically based
216 on local climatic conditions (Doorenbos and Pruitt, 1977). The accumulated deficit over a
217 season is then given as
218

$$219 \quad deficit_{j,d} = \max(deficit_{j,d-1} + D_{j,d} - S_{j,d}, 0) \text{ where } deficit_{j,d=0} = 0 \dots (3)$$

$$220 \quad CDI_{j,t} = \max(deficit_{j,d(y)}: d = 1:n_s; t = 1:n); \text{ where } deficit_{j,d(0)} = 0, y=1, \dots, n \dots (4)$$

221
222
223 In equation (3), $deficit_{j,d}$ refers to the accumulated daily deficit for any given year with a crop
224 growth period of n_s days in the year, $D_{j,d}$ to total daily water demand, $S_{j,d}$ to the total daily
225 effective rainfall, for geographical location j , and day d ; t refers to a calendar or cropping year;
226 and n is the total number of years in the analysis. For an n -year record, seasonal water stress is
227 evaluated as the maximum cumulative deficit each season and defined here as $CDI_{j,t}$. CDI
228 focuses on the rainfall distribution within the season relative to the crop water demand. It
229 therefore accounts for the timing of planting, different stages of crop growth, and the timing and
230 distribution of rainfall in the season. The index may also be treated as a hydrologic index and
231 forecasted exactly as one would forecast precipitation or temperature variables, or any other
232 water stress or drought index. Depending on the lead time of such forecasts, this can give
233 farmers and other agricultural stakeholders a sufficient amount of planning and preparation time,
234 thus providing them a critical edge in hedging agricultural water risk. This is critical in irrigation
235 and water storage planning.
236
237

238 **4. Case Study: Forecasting Irrigation Requirements for Potatoes in Maharashtra, India**

239 We endeavored to forecast CDI for potatoes grown in the Satara district in Maharashtra,
240 India as an application. The Satara district in Maharashtra is one of the primary regions for
241 sourcing potatoes during the monsoon season (June - September). Satara supplies the majority of
242 the potatoes processed by the Frito-Lay manufacturing plant in Pune, Maharashtra (Economic
243 Times, 2013). Potato is a major cash crop in Maharashtra and accounts for at least 75% of total



244 production (Nikam, *et al.*, 2008). The average annual rainfall in this arid to semi-arid region is
245 around 350 mm with high inter-annual variability. The region has experienced four droughts
246 (seasonal rainfall below long-term average) since 2001. The ability to predict such droughts with
247 a reasonable accuracy at lead times of three to six months could suggest ways to adapt existing
248 agricultural operations to the anticipated conditions and minimize the impacts of droughts on the
249 agricultural supply chain. Hence, we develop, present and evaluate the results from retrospective
250 forecasts of CDI for the monsoon season over the period 2001-2013. The June-July-August-
251 September (JJAS) season is the growing season for potatoes in the Satara district. It is also the
252 core monsoon season for the Indian sub-continent. The forecasts use climate data from three to
253 six months prior to the beginning of the monsoon season as predictors, and forecasts are to be
254 issued in May, one month prior to monsoon onset.

255

256 **4.1: Data Collection and Processing**

257 *4.1.1: Precipitation and Temperature Data and the CDI*

258 Gridded daily rainfall data from 1901 – 2004 available at $1^0 \times 1^0$ spatial resolution from
259 the India Meteorological Department (Rajeevan *et al.*, 2006), and gridded daily temperature data
260 from 1948 – 2000, available at the same spatial resolution from National Center for Atmospheric
261 Research (Ngo, *et al.*, 2005) are used in this study. Since the daily temperature data is available
262 only for 53 years, we used the daily climatology, i.e. the mean daily temperature, for the
263 remaining 60 years (Devineni *et al.*, 2013). The daily climate time series grids were spatially
264 averaged over the Satara district. This process resulted in a time series of daily precipitation and
265 temperature estimates for 104 years. The daily Reference Crop Evapotranspiration (ET_0) was
266 developed based on the daily time series of minimum, mean and maximum temperature data, and
267 extraterrestrial solar radiation (Hargreaves and Samani, 1982). The Hargreaves method is used
268 globally to predict ET_0 in regions where data availability is limited to air temperature data (Allen,
269 *et al.*, 1998). Seasonal daily rainfall data from 2005 to 2013 for the Satara district were collected
270 separately from a website maintained by the Agricultural Department of Maharashtra State and
271 used to augment the 104 years of rainfall and temperature data. The CDI was computed for each
272 of these 113 seasons using the daily rainfall data and reference crop evapotranspiration. This
273 will serve as the predictand for our forecast model. The computation of CDI is illustrated in Fig.
274 1. These figures provide insights on the time-evolving vulnerability to stress arising from
275 deficient rainfall and changes in crop demand.

276

277 CDI as a water stress measure is a proxy of not only crop water stress but also irrigation
278 and water storage requirements. Consider Fig. 1. When daily seasonal rainfall is low or when
279 rainfall enters an inactive phase for a considerable period of time, as displayed by the vertical
280 cyan bars, the amount of daily accumulated water deficit increases to reflect the disparity
281 between water supplied as rainfall and the water required by the crop to sustain itself, as
282 displayed by the red curve in Fig. 1. The highest point, or peak, on the black deficit time series
283 in Fig.1 is the value of CDI, and it prepares us for the worst-case scenario of deficient water
284 supply for the crop. This can be calculated for multiple crops, each CDI value depending on the
285 specific crop's water demand and the location and time of planting. This gives the stakeholder a
286 conservative estimate of how much additional water is needed beyond what Nature is willing to
287 supply in order to maintain critical yields while apportioning water resources intelligently. Since
288 agriculture tends to be one of the largest consumers of water --- about seventy-percent of all the



289 world's freshwater withdrawals go towards irrigation use (USGS, 2017), and this is in addition to
290 what is rainfed --- this is an integral part of water resources management.

291 The annual time series of the CDI computed for the JJAS season (referred to as Kharif
292 season in India sub-continent) in Satara is presented in Fig. 2. We have standardized the CDI
293 values as the percentage difference each year from the 113-year average of CDI. The long-term
294 average CDI for growing potatoes in Satara is 241 mm. This is equivalent to approximately
295 257,644 gallons of water used for irrigating a one-acre farm of potatoes on average throughout
296 the season. The percent differences in Fig. 2 refer to percentages of this number, i.e. a 10%
297 increase in CDI indicates an additional requirement of 25,764 gallons. From Fig. 2, it is clear that
298 (a) Satara experiences recurrent droughts with intermediate wet periods and (b) there is year-to-
299 year persistence in the incidence of these droughts. Such variations and epochal changes are
300 typically modulated through large-scale global climate patterns. Investigating the relationship
301 between monsoon deficit and the large-scale climate teleconnections could enable the
302 development of models that can be used to understand and predict the variability in the CDI in
303 the region.

304

305 4.1.2: Climate Precursors and Climate Data

306 Our goal was to develop a simple statistical model for predicting CDI for potatoes grown in
307 Satara. The generalized climate forecast models available at low spatial resolution are not
308 specific enough for this task. Consequently, the first objective was to identify appropriate climate
309 predictors before the monsoon starts in June. There is an extensive history of developing long-
310 range predictions of monsoon rainfall that are based on various regional to large-scale climate
311 predictors (Walker, 1924; Thapliyal, 1987). A variety of seasonal forecasts of the all India
312 Summer Monsoon Rainfall (ISMR) are documented and available for reference (Gadgil et al.,
313 2007; Kumar et al., 1995).

314 It is well established that inter-annual climate modes such as ENSO associated with
315 anomalous Sea Surface Temperature (SST) conditions in the tropical Pacific Ocean influence the
316 inter-annual variability of ISMR (Parthasarathy and Pant, 1985; Shukla and Paolino, 1983).
317 Anomalously warm tropical eastern Pacific SSTs (El Niño) are associated with a drier-than-
318 normal ISMR, whereas anomalously cool tropical eastern Pacific SSTs (La Niña) are associated
319 with a wetter-than-normal ISMR (Sikka, 1980; Parthasarathy and Panth, 1985; Rasmusson and
320 Carpenter, 1983). Ihara, *et al.* (2007) have suggested that the ENSO warm (cool) phases shift the
321 location of the tropical Walker circulation and cause deficient (excessive) rainfall by suppressing
322 (enhancing) the convection over India. Hence, ENSO indices were chosen to be among the
323 candidate predictors for the forecast model.

324 Raw monthly SST data for the Niño 3, Niño 4, Niño 12 and Niño 34 indices were taken from
325 the KNMI climate explorer database (KNMI, 2016). For each given raw ENSO index (3, 4, 12
326 and 34), we considered three different types of derived ENSO indices: a December-January-
327 February (DJF) seasonal average, a March-April-May (MAM) seasonal average, and a MAM
328 minus DJF (MAM-DJF) differenced time series. Among the Niño indices calculated, the change
329 in the tropical Pacific SSTs from December to May (MAM-DJF trend) was found to be of
330 significance by previous investigators. Shukla and Paolino (1983) found the correlation
331 coefficient between the MAM-DJF trend pressure anomalies and the ISMR to be a significant -
332 0.42. Parthasarathy et al. (1988) found the correlation coefficient between this winter-to-spring
333 trend and ISMR over the period 1951-1980 to be between 0.40 and 0.52 in magnitude, depending
334 on the specific region within the tropical Pacific. Hence, MAM-DJF trends from Niño 3, Niño 4,



335 Niño 12 and Niño 34 were considered to be potential model predictors. Parthasarathy et al.
336 (1988) found that the MAM-averaged tropical Pacific SSTs over the box 14 N to 20 N, 176 E to
337 160 W had a correlation of -0.40 with ISMR, convincing us to consider this average as well. In
338 addition to the MAM and MAM-DJF averages, we computed the winter season (DJF) average,
339 although DJF-averaged tropical Pacific SSTs were not found to be significant in the literature.
340 However, it is worth noting that Parthasarathy et al. (1988) found that the correlation coefficient
341 between the Darwin SLP during the DJF season and ISMR was +0.39.

342 As the concurrent season (JJAS) state of ENSO has an important, well-documented impact
343 on ISMR, we also elected to include the Niño 34 JJAS average. As mentioned earlier, an El
344 Niño event during the JJAS season is strongly associated with an anomalously dry JJAS rainfall
345 season in India, while a La Niña event during the JJAS season is strongly associated with an
346 anomalously wet JJAS rainfall season in India, prompting our choice. We coupled the JJAS
347 seasonal average for the Niño 34 index with forecasts of the JJA and JAS seasonal averages for
348 the Niño 34 index. These forecasts were obtained from the International Research Institute for
349 Climate and Society (IRI) ENSO forecast page and covered the period 2002-2013. These
350 forecasts can be used to forecast JJAS monsoon CDI in place of the observed Niño 34 JJAS
351 values on a real-time basis. These forecasted values were averages of the projections from at
352 least six distinct statistical/dynamical models, with one average for the JJA season and one
353 average for the JAS season. Together, we start with a total of thirteen ENSO-based indices.

354 Other candidate predictor variables include concurrent season (JJAS) eastern Indian Ocean
355 SSTs known as the Indonesian Throughflow, or ITF. Warm, low-salinity water from the Pacific
356 is introduced into the Indian Ocean via the ITF and is considered to be an integral component in
357 the heat and hydrological budget of the Indian Ocean (Gordon et al., 1997). The ITF waters are
358 also believed to influence SSTs and associated ocean-atmosphere coupling within the Indian
359 Ocean, making it an important aspect of monsoon climate research (Gordon et al., 1997). Thus,
360 the ITF was also selected to be a candidate predictor in the model. During the JJAS monsoon
361 season, the ITF is strengthened considerably, allowing an abundant amount of relatively warm
362 water to be injected into the Indian Ocean. Eastern Indian Ocean SSTs during the JJAS season
363 correspond to enhanced (suppressed) atmospheric convection during the anomalous warming
364 (cooling) of the Indian Ocean waters, which in turn supplies (robs) the developing monsoon of
365 much-needed moisture. We found that the Spearman rank correlation coefficient between CDI in
366 Satara and the average SST anomalies over 20° N and 5° S and 100° E and 130° E (the region
367 representing ITF) during the JJAS season is around -0.35 (statistically significant at the 95%
368 level), suggesting that warm conditions in the ITF region result in below-normal CDI, or low
369 crop water stress. Figure 3 presents the field correlation map of SST anomalies with CDI. For
370 these reasons, we chose concurrent season ITF data to be a candidate predictor. The ITF data was
371 collected from the IRI data library and consists of two components: an observation component
372 and a forecasted component. The observations consist of measured eastern Indian Ocean SST
373 anomalies during the JJAS season from 1901 through 2013. The forecasts consist of JJAS-
374 season ITF values retrospective from the ECHAM4.5 global climate model and cover the period
375 2001-2013. Skillful forecasts for the tropical SSTs based on coupled ocean-atmospheric general
376 circulation models have been in operation from various climate centers since 1998. Hence, in
377 the forecasting scheme, we used the ITF derived from forecasted SST state issued in May from
378 ECHAM4.5 operational forecasting center (available from IRI data library:
379 http://iridl.ldeo.columbia.edu/SOURCES/.IRI/.FD/.ECHAM4p5/.Forecast/.ca_sst/.ensemble24/;
380 Li and Goddard, 2005; van den Dool, 2007; Roeckner et al., 1996). The observed JJAS ITF data



381 are used to train the model, while the retrospective JJAS ITF forecasts are used to make forecasts
382 for the years 2001 – 2013.

383

384 **4.2: The Forecasting Procedure**

385 *4.2.1: Predictor Selection*

386 Given a pool of candidate predictors, the next step is to select the best subset of those
387 predictors. The predictors used in the forecasting model were chosen based on an exhaustive
388 search method. In the exhaustive search method, all possible combinations of the candidate
389 predictor variables are used to develop models that are cross-validated on historical data. Skill
390 metrics are then used to compare the predictive accuracy of each combination. In the present
391 study, we began with 113 years of CDI data and fourteen candidates: Niño 3 DJF, Niño 3 MAM,
392 Niño 3 MAM-DJF, Niño 4 DJF, Niño 4 MAM, Niño 4 MAM-DJF, Niño 12 DJF, Niño 12
393 MAM, Niño 12 MAM-DJF, Niño 34 DJF, Niño 34 MAM, Niño 34 MAM-DJF, Niño 34 JJAS
394 and ITF. The exhaustive search method utilized the kNN cross-validation algorithm and forty
395 years of training data (1901-1940) to build forecast distributions for each of the years 1941-2013.
396 At each step, the training data was updated to include data from all of the years up until the year
397 being cross-validated. Thus, we always use only the historical data and update the model each
398 year with the information of the previous year, much as a regular user of the forecast system
399 would have to do. These forecasting distributions, built over a 73-year record (1941 to 2013)
400 were created successively for every unique combination of two variables, every unique
401 combination of three variables, so on and so forth until we reached the entire pool of predictors.

402 For each and every possible unique combination of the predictor variables, we obtain a
403 matrix of seventy-three columns. For each of these seventy-three (73) years, the squared error
404 and rank probability score (Epstein, 1969; Murphy, 1969, 1971; Candille and Talagrand, 2005)
405 were computed, and from this the root mean squared error (RMSE) and rank probability skill
406 score (RPSS) were computed. In this manner, a single RPSS value and MSE value were
407 calculated for every possible combination of the predictor variables. We chose the following
408 combination of predictors based on the relative optimality of both their RPSS and RMSE scores:
409 Niño 12 MAM-DJF, Niño 34 MAM-DJF, and ITF, and this set of variables had an RMSE of
410 49.25 mm of required (JJAS) seasonal water storage and RPSS of 0.26. We devised a simple but
411 effective decision rule for determining the optimal choice of predictors based on ranking the
412 metric values. This is especially useful when the number of combinations of variables is
413 unwieldy. Optimality was determined by assigning a rank number to the RMSE and RPSS
414 values in such a way that the number 1 was assigned to the lowest RMSE value, 2 to the second
415 lowest RMSE value, and so on, and the number 1 was assigned to the largest RPSS value, 2 to
416 the second largest RPSS value, and so on. For a fixed number of cross-validated predictor
417 candidates, and for each RMSE/RPSS pair, one pair for each combination of predictors, we
418 determined an RMSE and RPSS rank and took the sum of these ranks. The smallest of all of
419 these sums corresponds to the best or optimal set of predictors among all possible sets of cross-
420 validated predictors. We then compared the rank sum along with the number of predictors to
421 choose the best set of predictors. The chosen trio of predictors mentioned above had the
422 unequivocally highest value of RPSS and second lowest RMSE value out of all possible
423 combinations of the original set of seventeen candidates, the lowest RMSE being only slightly
424 smaller at 48.92 mm. Conceptually, this procedure is similar to the “best subsets regression” or
425 “step-wise regression” (Helsel and Hirsch, 2002), but in the spirit of using kNN algorithm for
426 forecasting, we designed this selection scheme to use the kNN algorithm instead.



427 CDI forecasts were subsequently made using the selected set of predictors. The forecast
 428 procedure is tested using the leave-one-out cross-validation method. Each historical observation
 429 is omitted in turn, and the model is developed using the remaining years of data. A prediction of
 430 the observation that was not kept in the model-building set is then made and compared with the
 431 actual outcome for that year. Results from a variant of this approach are presented in the next
 432 section. The CDI for the 2001 Kharif season is predicted using the model developed based on
 433 data from 1901 – 2000. Similarly, the CDI for 2002 is predicted based on the model that is
 434 developed using the data from 1901 – 2001. Thus, as we move from year to year, we update the
 435 model observations and predict the future state.

436

437 4.2.2: The *k*-Nearest Neighbors Real-Time Forecasting Model

438 The forecasts were developed using a semi-parametric *k*-nearest neighbors (k-NN)
 439 model. This is a data-driven approach that develops a conditional probability distribution of the
 440 CDI given the predictors by first identifying the *k*-historical climate conditions that are most
 441 similar to the current values of the climate predictors and then randomly drawing the vector of
 442 CDI values in the historical data that correspond to these *k* neighbors. The neighbors are
 443 weighted so that the closer or more similar neighbors are chosen more often than those further
 444 away. The key steps are as follows.

445 Let \mathbf{X} be the design matrix of size $n \times p$, where p = number of predictors selected from
 446 the original pool of candidates. Let \mathbf{x}_i denote the i^{th} row of \mathbf{X} . Hence, \mathbf{x}_i is a vector containing
 447 the values of each of the p predictor variables during year i . Denoting the current values of the
 448 predictors by \mathbf{x}_c , the idea is to find k such predictor vectors from the historical record (i.e. find k
 449 values of \mathbf{x}_i with $i < c$) that are most "similar" to the value of \mathbf{x}_c and use this information to
 450 construct a sampling distribution of CDI from which we can issue probabilistic forecasts. The
 451 number of neighbors in the model, or k , represents the number of degrees of freedom in the
 452 model, and should be chosen with care, as the choice of k affects the skewness and level of
 453 uncertainty in the sampling distributions. After trying several different values for k , we found an
 454 optimal value to be $k = 25$. Rajagopalan and Lall (1999) recommend that k be roughly equal to
 455 \sqrt{n} , where n = the total number of observations. In our situation, it was evident that we required
 456 more neighbors than this rule would allow, due to the skewness and variance apparent in the
 457 sampling distributions when using only eleven or fewer neighbors.

458 Let \mathbf{y} be the n -dimensional vector of seasonal CDI values, each component of which
 459 represents the aggregate water deficit level over the JJAS growing season of every year in the
 460 historical record. Assume that \mathbf{y} has been centered and normalized by its historical average to
 461 produce mean-normalized anomalies. The first step was to consider the individual distance
 462 values (under some specified metric) between \mathbf{x}_c and \mathbf{x}_i for $i = 1, \dots, c-1$. The chosen distance
 463 metric for our k-NN model was the Mahalanobis distance (Mahalanobis, 1936)

464

$$465 D_M(\mathbf{x}_c, \mathbf{x}_i) = \sqrt{(\mathbf{x}_c - \mathbf{x}_i)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_c - \mathbf{x}_i)} \dots (5)$$

466

467 where $\boldsymbol{\Sigma}$ is the covariance matrix of the training values in \mathbf{X} . The Mahalanobis distance measure
 468 judges point separations in a metric space based on statistical dissimilarity, as opposed to solely
 469 physical distance. Hence, the level of similarity between predictor values across different years
 470 is determined by the orientation and location of each point relative to the scatterplot of the
 471 predictor data. Large distances from \mathbf{x}_c represent predictor values that are statistically anomalous
 472 in the context of the predictor data.



473 After the Mahalanobis distances had been calculated, the k (with $k = 25$) smallest distance
474 values were selected and the corresponding years in which these distances occurred were noted.
475 These years, hereby referred to as the *analog years*, are the years during which the predictor
476 signals were most similar to those of the current year. The vector-valued predictors during these
477 analog years are referred to as the *neighbors* of \mathbf{x}_c .

478 The final step was to resample CDI values from the analog years. The resampling
479 technique employed is a nonparametric method known as the *bootstrap* (Efron, 1979; Efron and
480 Tibshirani, 1993). The idea behind the bootstrap component is to sample with replacement from
481 a pool of data using the underlying distribution that generated the data to guide the sampling
482 process. We chose not to assign a parametric family of distributions to the CDI data, and instead
483 estimated its underlying distribution semi-parametrically using a kernel density estimator. This
484 semi-parametric method of k -NN bootstrapping was first introduced in Lall and Sharma (1996).
485 Applications of the methods using different variants have since been presented (see for example,
486 Rajagopalan and Lall, 1999, Souza and Lall, 2003 and references therein). We employed the
487 same discrete resampling kernel proposed in Lall and Sharma (1996), which has the general form
488 $K(j) = 1/(j*S)$ with $S = \sum_{j=1}^k 1/j$, where j is the rank of each neighbor of \mathbf{x}_c , a rank of $j=1$
489 assigned to the closest neighbor and a rank of $j=k$ assigned to the most distant neighbor. Our
490 strategy was to build this kernel density estimator based on the ranks of the selected neighbors
491 and resample the predictand values from these analog years. We resampled from the twenty-five
492 analog CDI values 1,000 times, and each of the twenty-five values was resampled proportionally
493 to the probability of its occurrence as determined by the density estimator.

494

495 4.2.3: Analyzing the k -NN Results

496 The way in which model results are interpreted and presented is important for potential
497 stakeholders. In this case study, our interest was in forecasting the CDI for a given potato
498 growing season in Satara. The information from these forecasts can be of great use to potato
499 farmers in Satara as well as corporations with investments in these farming areas. This
500 necessitates a clear and concise communication of the forecast results.

501 The output of the k -NN model was a time series for each forecasted year consisting of
502 1,000 realizations. This is the sampling distribution for the CDI and consists of mean-
503 normalized anomaly values from the analog years converted to percentage values. As stated in
504 the previous section, the deficit value from each analog year in the sampling distribution is
505 represented proportionally to its probability of occurrence as assigned by a kernel density
506 estimator. The sampling distribution is used to issue one-season-ahead probabilistic forecasts
507 (i.e. the likelihood of a deficit for the forthcoming growing season). There are a whole slew of
508 possibilities when it comes to using these sampling distributions for probability-based forecasts.
509 Our approach includes the following for a given forecasted growing season:

- 510 1. A boxplot depicting the sampling distribution with the observed percent anomaly value
511 superimposed on the boxplot for every growing season forecasted. In using predictand
512 anomalies, the historical mean becomes the zero line in the coordinate plane of the
513 boxplot.
- 514 2. A three-category forecasting system with the categories “above normal”, “normal” and
515 “below normal”, provided that the historical mean/climatology is the threshold that is
516 desired.



- 517 3. Calculate the probabilities for the categories specified in step 2 from the sampling
518 distribution generated in step 1, and use this to evaluate the accuracy and strength of the
519 forecast based on contingency metrics such as hit rates and false alarms.
520 4. To get a sense of the spread/variability in the boxplot distribution, calculate the
521 Interquartile Range (IQR).
522 5. Compare the value of the observed percent anomaly of the predictand with the category
523 in which the majority of the probability mass of the sampling distribution lies. This is of
524 central importance in getting a basic sense of the accuracy of the forecast.

525 In general, the construction of such a sampling distribution allows the investigator the freedom to
526 calculate probabilities on many different thresholds. The thresholds should be defined by the
527 particular application and the needs of any stakeholders involved.

528 **5. Case Study: Forecast Results and Discussion**

529 **5.1: CDI Forecast Results and Comparison with IMD Monsoon Forecasts**

530 We hereby present the results of the CDI forecasts for the 2001 – 2013 JJAS seasons in
531 the Satara district, Maharashtra, India. Forecasts are specifically made in the interest of
532 irrigation requirements for potatoes grown in the Satara district, and we discuss the results in this
533 context. The output of the k-NN model is the forecasting distributions for CDI of the thirteen
534 years and a series of boxplots representing these forecast distributions as shown in Fig. 4. The
535 probabilities calculated from these distributions are shown in Table 1, columns 2 and 3.

536 Figure 4 shows a series of boxplot diagrams depicting the k-NN forecast distributions for
537 CDI over the years 2001 – 2013. All calculations in this Figure, including the construction of the
538 distributions themselves, were done using anomalies of the predictand rather than the raw
539 predictand values. The anomalies were calculated by subtracting the 1901 – 2013 mean from the
540 data and dividing by this mean value and converting the quotient to a percentage. The idea is to
541 gauge the level of seasonal crop water deficit in a forecasted year with respect to the level of
542 crop water deficit that has occurred on average over the entire historical record. This should
543 address the question: how “normal” or “abnormal” is a given level of deficit over a season with
544 respect to everything we have seen or experienced thus far. Given that the forecast is developed
545 one season ahead, the sign of a strong shift in the probability will alert the decision-makers to an
546 anticipated deficit or surplus event.

547 We have created two general possibilities: the observed percent anomaly values (triangles
548 in Fig. 4) can be positive or negative. As the forecasts have been carried out using anomalies
549 instead of raw values, the 1901 – 2013 historical average is re-positioned as the zero line in Fig.
550 4. We calculate the probability under the kNN forecast distribution of observing positive
551 (negative) deficit anomalies for each year in 2001 – 2013. These are retrospective forecasts in
552 the sense that these anomalies have already been observed and recorded but not used in building
553 the model. These probabilities, corresponding observed percent anomalies and IQR values are
554 presented in Table 1. The utility of these forecasts are discussed in section 5.2.

555 Given the above information, we judge the accuracy of the forecasts during any given
556 year on a few simple criteria: the directional agreement between the observed percent predictand
557 anomaly and the median of the forecast distribution (Fig. 4), joint consideration of the forecast
558 probabilities and the observed percent anomaly (Table 1, columns 2, 3 and 4) and the level of
559 uncertainty in the forecast distribution (Fig. 4 and Table 1, column 5). Uncertainty is measured
560 by the IQR of the boxplot distribution. In the present context, we say that a forecast for a given
561 year has *identical directionality* (with respect to the observation) if both the median of this



562 forecast and the observation (as a percent anomaly) are either positive (above the historical
563 average) or negative (at or below the historical average). The absence of identical directionality
564 will be called *dissimilar directionality*.

565 The box-and-whiskers plots shown in Fig. 4 for each year illustrates the range of possible
566 values of the CDI for that year. We have identical directionalities for the years 2001, 2004,
567 2005, 2006, 2007, 2010, 2011, 2012 and 2013. For the years 2001, 2011 and 2012, the model
568 correctly forecasted that the water stress conditions for the Maharashtra potatoes would be above
569 the CDI climatology. We can see from Fig. 4 that both the observed percent anomalies
570 (triangles) and the medians for all of these forecasted years are positive. Additionally, Table 1,
571 column 2 shows that the majority of the probability mass of the kNN distribution is placed in the
572 “Above Mean” category for 2001, 2011 and 2012, while column 4 shows that for these years, the
573 observed CDI anomalies are positive. Similarly, for the years 2004, 2005, 2006, 2007, 2010 and
574 2013, the model correctly forecasted that water stress conditions for the potatoes would be below
575 the historical average, and this can be seen from Fig. 4, where the observed anomalies and the
576 medians for all of these forecasted years are negative. Similarly, Table 1, column 3 shows that
577 the majority of the probability mass from the kNN forecasting model was placed on the “Below
578 Mean” category for these years, and the corresponding observed CDI anomalies are also
579 negative. For the years 2002, 2003, 2008 and 2009, we have dissimilar directionalities. The
580 forecasts suggest higher probability values for below average CDI during 2002 and 2003,
581 whereas positive anomalies were observed for these years. Similarly, the forecasts for 2008 and
582 2009 placed the majority of the probability mass on higher than average CDI, suggesting that
583 these years were likely to see higher than normal potato water stress. However, the observed
584 CDI anomalies were negative, implying the opposite scenario.

585 We say that a *hit* has occurred if identical directionality is observed. A *miss* occurs if the
586 forecast implies below average water stress, but the observation shows above average water
587 stress. Finally, a *false alarm* occurs if the forecast implies above average water stress while the
588 observation shows below average water stress. Table 2 shows that the hit rate of the kNN
589 forecasts is 9/13, the miss rate is 2/13 and the false alarm rate is 2/13. Table 3 shows a
590 comparison of our CDI forecasts with seasonal total precipitation forecasts of the India
591 Meteorological Department, abbreviated IMD. The IMD forecast presented here for 2001 is
592 long-range for precipitation in the JJAS season over three climatically homogeneous regions in
593 India: Northwest India, Peninsular India, and Northeast India. Maharashtra is in Peninsular
594 India, and so we refer to this forecast. For 2001, the forecast result was categorized as either
595 normal, above normal or below normal. “Normal” is defined as being within $\pm 10\%$ of the long-
596 period average, or LPA. Beginning in 2003, IMD began offering two-stage forecasts, the first
597 released in mid-April using data up to March and an update in June using data up through May.
598 For both 2011 and 2013, we used the initial country-wide forecast, as the updated forecasts for
599 JJAS could not be found. In 2003, IMD began to divide their forecast results into five
600 categories: drought/deficient, below normal, near normal/normal, above normal and excess.
601 “Deficient” (drought) is defined as JJAS total seasonal rainfall that is less than 90% of the long
602 period average (LPA). “Below normal” is defined as JJAS rainfall that is 90% – 96% of the
603 LPA, “normal” (sometimes called “near normal”) is defined as JJAS rainfall that is 96% – 104%
604 of the LPA, “above normal” is defined as JJAS rainfall that is 104% – 110% of the LPA and
605 “excess” is defined as JJAS rainfall that is more than 110% of the LPA. The IMD forecasts are
606 reported as percentages of the LPA, as shown in column 3 of Table 3. Going by the categories
607 defined by IMD, and comparing these forecasts with actual JJAS seasonal total precipitation



608 anomalies from our gridded rainfall data set, where these anomalies have been calculated with
609 respect to the long period average defined as 1901 – 2013, we classify each forecast as a hit, miss
610 or false alarm as was done with the CDI forecasts. The hit rate for IMD is 1/9, the miss rate is
611 3/9 and the false alarm rate is 5/9. We must bear in mind that the total precipitation forecasts
612 given here are for an entire region that includes the state of Maharashtra, whereas our CDI
613 forecasts are generated based on CDI calculations from the target location of Satara,
614 Maharashtra, India. Hence, our CDI anomalies reflect the conditions of Satara on a much higher
615 resolution than the coarse IMD precipitation anomalies. Furthermore, we are comparing IMD
616 forecasts with actual precipitation totals from Satara, and computed with respect to the 1901 –
617 2013 LPA instead of the 1951 – 2000 LPA of IMD, under the reasonable assumption that the
618 LPA does not change much between those two definitions. While the IMD monsoon forecasts
619 can provide a broad regional understanding of the monsoon conditions, supplementing them with
620 targeted crop-specific forecasts such as ours will help improve agricultural planning and regional
621 water management.

622 We define a *strong forecast* as a forecast in which the probability assigned to one of the
623 two categories is at least 60%. In our situation, ten out of the thirteen years witnessed strong
624 forecasts. A weak forecast runs the risk of being less informative to decision-makers, whereas a
625 strong forecast is much more assertive and definitive, and hence decisions can be made more
626 easily with a strong forecast. The forecasts were also correct for seven of these ten years, as seen
627 in Table 2. The forecasts were correct, but barely weak, for two years (2001 and 2011). If one
628 considers acting only if the probability associated with a CDI forecast is at least 60%, then the
629 forecast is correct seven out of ten times. Raising this to 66% leads to four out of six years
630 classified correctly.

631 It is important to point out that one should also consider the uncertainty (column five in
632 Table 1) when evaluating the power of the forecasts. Knowing the uncertainty is useful since
633 years in which the uncertainty in the forecast is low and there is a strong indication for CDI may
634 lead to different risk management actions than years in which the forecast has strong directional
635 change but is also marked by high uncertainty.

636

637 **5.2: Discussion of Results: The Utility of Targeted Forecasts**

638 It is natural to ask how one might go about using CDI forecasts. Here is a short example
639 of how these forecasts can facilitate decision-making. In 2001, irrigating, or ensuring water
640 storage equal to 294,745 gallons per acre for the potatoes would have been the ideal situation, as
641 this is equivalent to being 14.4% above the average CDI value of 241 mm of water storage
642 equivalent. However, this exact amount cannot be known in the absence of the observed CDI
643 anomaly, which is found in column four of Table 1. Using the median as a plausible estimate for
644 the true anomaly value, roughly 268,980 gallons per acre would have been irrigated or stored
645 instead. A more risk-averse decision-maker may choose to use the upper quartile or even
646 maximum of the kNN-generated sampling distribution as a proxy for the true anomaly value.
647 Such decisions are often made on the basis of prior experience.

648 Although total seasonal rainfall is sometimes used for agricultural water planning, CDI
649 boasts a significant advantage over total seasonal rainfall in this capacity. CDI reliably accounts
650 for water stress incurred by haphazard and erratic patterns of rainfall during the season. A total
651 seasonal rainfall forecast that indicates a growing season with sufficient rainfall will not be
652 reliable when rain throughout the season is erratically distributed in clusters of rainy days,
653 whereby all of the rainfall in a given season occurs within a portion of the season, and the



654 remainder of the season is virtually dry. This is a common occurrence in monsoonal climates,
655 and may have deleterious effects on crops that are vulnerable to prolonged dry periods and/or
656 chunks of time during which rainfall is excessive. Long dry spells throughout the season that
657 can be detrimental to drought-sensitive crops are not accounted for in a measure of total seasonal
658 rainfall, making it possible for the seasonal rainfall to appear sufficient due to sporadic
659 occurrences of large precipitation events. Consequently, it can also serve as a better indicator
660 than regional rainfall to devise index insurance products for agriculture, where crop specific
661 indices can be developed (Skees, 2016). These characteristics of crop water stress must be
662 accounted for in the proper planning and management of agricultural water resources.

663 To illustrate the above point further, we appeal to Figure 5. In this figure, the varying
664 rainfall distribution is indicated by the vertical bars, the crop demand is given by the horizontal
665 line (primary y-axis), and the time series shows the cumulative deficit. The second panel shows
666 two distinct years during which the total seasonal rainfall was 590 mm (vertical line). During
667 one of these two years, the CDI value was 111 mm of water deficit for the potato crop, while the
668 CDI value for the other year was 228 mm. This indicates that the water stress for a particular
669 crop relies on both the magnitude and frequency of seasonal rainfall. When daily seasonal
670 rainfall is more uniform, the daily deficit values do not have the chance to accumulate as much
671 as when rainfall is less uniform and, as a result, when there are persistent dry spells or long
672 precipitation-inactive periods. Panel three shows the resulting cumulative deficit when daily
673 rainfall occurs with greater frequency during the JJAS season and hence the total seasonal
674 rainfall is distributed among the days of the growing season fairly uniformly. The fourth panel,
675 immediately to the right of the third panel, shows the resulting cumulative deficit when rainfall is
676 dominant during the first and last months of the JJAS season. While rainfall events do occur in
677 between, the magnitude of the rainfall is quite low, allowing the seasonal daily CDI time series
678 to spike to a considerably higher maximum value (228 mm) than the CDI time series in panel
679 three (111 mm maximum). The CDI time series recedes and recovers at the end of the season
680 when the rainfall increases in magnitude. Hence, CDI can discriminate between two monsoon
681 seasons which have the same total rainfall, but differ in that one may have rainfall distributed
682 uniformly over the season through modest rainfall events, while the other may have a few intense
683 rain events separated by long dry periods. As we can see, the latter gives rise to a much higher
684 CDI.

685

686 **6. Summary and Conclusion**

687 A novel crop water stress index, the CDI, was developed here as a way of estimating
688 water storage and irrigation requirements in the interest of agricultural water resources. As
689 management of water resources requires advance knowledge of water risk, the main task
690 accomplished here was the forecasting of CDI as an effective method for understanding and
691 hedging risk. This concept of forecasting CDI for evaluating irrigation requirements was applied
692 to a case study in the Satara district of Maharashtra, India in which the CDI pertaining to
693 potatoes grown in Satara during the Southwest monsoon season was forecasted using large-scale
694 climate indices as predictors in a semi-parametric k-nearest neighbors stochastic model that
695 issues probabilistic forecasts. The climate indices used were defined either concurrent to the
696 monsoon season or three to six months prior. Based on the hit and false alarm rates, the results
697 achieved using our methodology were more favorable than precipitation forecasts conducted by
698 the India Meteorological Department. We also observed in our method a greater tendency
699 towards strong and informative forecasts.



700 This study developed a framework for quantifying and analyzing climate-induced
701 agricultural risks. It is based on (a) developing CDI for assessing crop-specific water risk,
702 irrigation requirements and water storage needs for the agricultural sector; (b) investigating the
703 sources of predictability for this indicator, and (c) developing statistically verifiable models for
704 issuing season-ahead probabilistic forecasts for evaluating water risk and irrigation needs. We
705 can conclude that this is a useful approach to investigating irrigation requirements and that
706 bootstrap-based uncertainty estimation is useful for developing probability-based management
707 models for optimizing agricultural decisions.

708

709 **Acknowledgements**

710 This research was supported by:

711 (a) NSF grant 1360446 (Water Sustainability and Climate, Category 3)

712 (b) PSC-CUNY award 69729-00 47

713 Partial support for the third and fourth authors is provided from PepsiCo Inc. through the
714 WATER RISKS AND SUSTAINABILITY grant. The statements contained within the
715 manuscript/research article are not the opinions of the funding agency or the U.S. government
716 but reflect the authors' opinions.

717

718 **Data Availability**

719 The CDI data used in this paper is available upon request of the contact author.

720

721 **References**

- 722 1. Allen, R.G., Pereira, L.S., Raes, D., Smith, M. Crop Evapotranspiration --- Guidelines
723 for Computing Crop Water Requirements. *FAO Irrigation and Drainage Paper 56*, FAO
724 of the UN, Rome, 15 pp., 1998.
- 725 2. Asoka, A. and Mishra, V. Prediction of Vegetation Anomalies to Improve Food Security
726 and Water Management in India, *GEOPHYS RES LETT*, V. 42, pp. 5290 – 5298, 2015.
- 727 3. Belayneh, A. and Adamowski, J. Standard Precipitation Index Drought Forecasting
728 Using Neural Networks, Wavelet Neural Networks and Support Vector Regression,
729 *Applied Computational Intelligence and Soft Computing*, 13 pp., 2012.
- 730 4. Candille, G. and Talagrand, O. Evaluation of Probabilistic Prediction Systems for a
731 Scalar Variable, *Q J ROY METEOR SOC*, V. 131, pp. 2131 – 2150, 2005.
- 732 5. Chen,
- 733 6. Devineni, N., Perveen, S., and Lall, U. Assessing Chronic and Climate Induced Water
734 Risk Through Spatially Distributed Cumulative Deficit Measures: A New Picture of
735 Water Sustainability in India, *WATER RESOUR RES*, V. 49, pp. 2135-2145, 2013.
- 736 7. Devineni, N., Lall, U., Etienne, E., Shi, D., and Xi, C. America's water risk: Current
737 demand and climate variability, *GEOPHYS RES LETT*, V. 42, 2015.
- 738 8. Doorenbos, J., Pruitt, W.O. Guidelines for Predicting Crop Water Requirements:
739 *Irrigation and Drainage Paper 24*, FAO of the UN, Rome, 154 pp., 1977.
- 740 9. The Economic Times, India Times, [http://articles.economictimes.indiatimes.com/2013-
741 09-25/news/42394669_1_drip-irrigation-farming-market](http://articles.economictimes.indiatimes.com/2013-09-25/news/42394669_1_drip-irrigation-farming-market) (Accessed: 3/1/2018), 2013.
- 742 10. Efron B. Bootstrap Methods: Another Look at the Jackknife. *Annals of Statistics*, V. 7, pp.
743 1-26, 1979.
- 744 11. Efron, B. and Tibishirani, R. An Introduction to the Bootstrap. Chapman and Hall, New
745 York, 456 pages, 1993.



- 746 12. Epstein, E.S. A Scoring System for Probability Forecasts of Ranked Categories, J APPL
747 METEOROL, V. 8, pp. 985 – 987, [https://journals.ametsoc.org/doi/pdf/10.1175/1520-0450\(1969\)008%3C0985:ASSFPF%3E2.0.CO%3B2](https://journals.ametsoc.org/doi/pdf/10.1175/1520-0450(1969)008%3C0985:ASSFPF%3E2.0.CO%3B2), 1969.
- 748
749 13. Etienne, E., Devineni, N., Khanbilvardi, R., and Lall, U. Development of a Demand
750 Sensitive Drought Index and its Application for Agriculture over the Conterminous
751 United States, J HYDROL, V. 534, 219–229,
752 <http://dx.doi.org/10.1016/j.jhydrol.2015.12.060>, 2016.
- 753 14. Gadgil, S., Rajeevan, M., and Francis, P.A. Monsoon Variability: Links to Major
754 Oscillations Over the Equatorial Pacific and Indian Oceans, CURR SCI INDIA, V. 93,
755 pp. 182 – 194, 2007.
- 756 15. Gordon, A.L., Ma, S., Olson, D.B., Hacker, P., Ffield, A., Talley, L.D., Wilson, D., and
757 Baringer, M. Advection and diffusion of Indonesian throughflow water within the Indian
758 Ocean South Equatorial Current. GEOPHYS RES LETT, V. 24, pp. 2573-2576,
759 <http://dx.doi.org/10.1029/97GL01061>, 1997.
- 760 16. Hargreaves, G.H. & Samani, Z.A. Estimating Potential Evapotranspiration. Journal of the
761 Irrigation and Drainage Division, V. 108, pp. 225-230, 1982.
- 762 17. Heim Jr., R.R. A Review of Twentieth-Century Drought Indices Used in the United
763 States. Bulletin of the American Meteorological Society, V. , pp. 1149 – 1165, 2002.
- 764 18. Helsel, D.R. & Hirsch, R.M. Statistical Methods in Water Resources, US Geological
765 Survey, 467 pages, 2002.
- 766 19. Ho, M., Parthasarathy, V., Etienne, E., Russo, T., Devineni, N., & Lall, U. America's
767 water: Agricultural water demands and the response of groundwater. GEOPHYS RES
768 LETT, V. 43, pp. 7546–7555. <http://dx.doi.org/10.1002/2016GL069797>, 2016.
- 769 20. Ihara, C., Kushnir, Y., Cane, M.A., & de la Peña, V.H. Indian summer monsoon rainfall
770 and its link with ENSO and Indian Ocean climate indices. INT J CLIMATOL, V. 27, pp.
771 179-187, <http://dx.doi.org/10.1002/joc.1394>, 2007.
- 772 21. Kar, S., Acharya, N., Mohanty, U.C. & Kulkarni, M.A. Skill of Monthly Rainfall
773 Forecasts Over India Using Multi-Model Ensemble Schemes. INT J CLIMATOL, V. 32,
774 pp. 1271 – 1286, <http://dx.doi.org/10.1002/joc.2334>, 2012.
- 775 22. KNMI Climate Explorer, <https://climexp.knmi.nl>, 1/1/2014
- 776 23. Kumar, K.K., Sonam, M.K. & Kumar, R.K. Seasonal Forecasting of Indian Summer
777 Monsoon Rainfall: A Review. WEATHER, V. 50, pp. 449 – 467,
778 <http://dx.doi.org/10.1002/j.1477-8696.1995.tb06071.x>, 1995.
- 779 24. Lall, U. & Sharma, A. A Nearest Neighbor Bootstrap for Resampling Hydrologic Time
780 Series. WATER RESOUR RES, V. 32, pp. 679 – 693, <http://dx.doi.org/10.1029/95WR02966>, 1996.
- 781
782 25. Li, S. & Goddard, L. Retrospective Forecasts with the ECHAM4.5 AGCM IRI Tech.
783 Report 05 – 02 December 2005, 2005.
- 784 26. Liu, X. & Pan, Y. Agricultural Drought Monitoring: Progress, Challenges, and
785 Prospects. J GEOGR SCI, V. 26, pp. 750 – 767, <http://dx.doi.org/10.1007/s11442-016-1297-9>, 2016.
- 786
787 27. Mahalanobis, P.C. On the Generalized Distance in Statistics. Proceedings of the National
788 Institute of Sciences of India, V. 2, pp. 49 – 55, 1936.
- 789 28. McKee, T.B., Doesken, N.J. & Kleist, J. The Relationship of Drought Frequency and
790 Duration to Time Scales. Eighth Conference on Applied Climatology, Anaheim,
791 California, 17 – 22 January 1993, 1993.



- 792 29. Mishra, A.K. & Desai, V.R. Drought Forecasting Using Stochastic Models. *STOCH*
793 *ENV RES RISK A*, V. 19, pp. 326 – 339, <http://dx.doi.org/10.1007/s00477-005-0238-4>,
794 2005.
- 795 30. Mishra, A.K. & Desai, V.R. Drought Forecasting Using Feed-Forward Recursive Neural
796 Network. *ECOL MODEL*, V. 198, pp. 127 – 138,
797 <http://dx.doi.org/10.1016/j.ecolmodel.2006.04.017>, 2006.
- 798 31. Mishra, A.K. & Singh, V.P. A Review of Drought Concepts. *J HYDROL*, V. 391, pp.
799 202 – 216, <http://dx.doi.org/10.1016/j.jhydrol.2010.07.012>, 2010.
- 800 32. Murphy, A.H. On the “ranked probability score”. *J APPL METEOROL*, V. 8, pp. 988 –
801 989, [https://doi.org/10.1175/1520-0450\(1969\)008<0988:OTPS>2.0.CO%3B2](https://doi.org/10.1175/1520-0450(1969)008<0988:OTPS>2.0.CO%3B2), 1969.
- 802 33. Murphy, A.H. A Note on the Ranked Probability Score. *J APPL METEOROL*, V. 10,
803 pp. 155 – 156, [https://doi.org/10.1175/1520-
804 0450\(1971\)010<0155:ANOTRP>2.0.CO%3B2](https://doi.org/10.1175/1520-0450(1971)010<0155:ANOTRP>2.0.CO%3B2), 1971.
- 805 34. Ngo-Duc, T., Polcher, J. & Laval, K. A 53-year Forcing Data Set for Land Surface
806 Models. *J GEOPHYS RES*, V. 110, 13 pp., <http://dx.doi.org/10.1029/2004JD005434>,
807 2005.
- 808 35. Nikam, A.V., Shendage, P.N., Jadhav, K.L. & Deokate, T.B. Economics of Production
809 of *Kharif* Potato in Satara, India. *International Journal of Agricultural Science*, V. 4, pp.
810 274 – 279, 2008.
- 811 36. Palmer, W.C. Meteorological Drought. Research Paper No. 45, U.S. Department of
812 Commerce, Washington, D.C., 65 pp., 1965.
- 813 37. Parthasarathy, B. & Pant, G.B. Seasonal Relationships Between Indian Summer
814 Monsoon Rainfall and the Southern Oscillation. *J CLIMATOL*, V. 5, pp. 369 – 378,
815 [http://dx.doi.org/551.513.7:551.553.11:551.577.32\(540\)](http://dx.doi.org/551.513.7:551.553.11:551.577.32(540)), 1985.
- 816 38. Parthasarathy, B., Diaz, H.F. & Escheid, J.K. Prediction of All-India Summer Monsoon
817 Rainfall with Regional and Large-Scale Parameters. *J GEOPHYS RES*, V. 93, pp. 5341 –
818 5350, <http://dx.doi.org/10.1029/JD093iD05p05341>, 1988.
- 819 39. R Core Team (2018). R: A language and environment for statistical computing. R
820 Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- 821 40. Rajagopalan, B. & Lall, U. A k-nearest neighbor simulator for daily precipitation and
822 other weather variables. *WATER RESOUR RES*, V. 35, pp. 3089 – 3101,
823 [http://dx.doi.org/1999WR9000280043-1397/99/1999WR900028\\$09.00](http://dx.doi.org/1999WR9000280043-1397/99/1999WR900028$09.00), 1999.
- 824 41. Rajeevan, M., Bhate, J., Kale, J.D. & Lal, B. High Resolution Daily Gridded Rainfall
825 Data for the Indian Region: Analysis of Break and Active Monsoon Spells. *CURR SCI*
826 *INDIA*, V. 91, pp. 296 – 306, 2006.
- 827 42. Rasmusson, E.M. & Carpenter, T.H. The Relationship Between Eastern Equatorial
828 Pacific Sea Surface Temperature and Rainfall Over India and Sri Lanka. *MON*
829 *WEATHER REV*, V. 111, pp. 517 – 528, [http://dx.doi.org/10.1175/1520-
830 0493\(1983\)111%3C0517:TRBEEP%3E2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1983)111%3C0517:TRBEEP%3E2.0.CO;2), 1983.
- 831 43. Rockstrom, J., Karlberg, L., Wani, S.P., Barron, J., Hatibu, N., Oweis, T., Bruggeman,
832 A., Farahani, J. & Qiang, Z. Managing Water in Rainfed Agriculture – The Need for a
833 Paradigm Shift, *AGR WATER MANAGE*, V. 97, pp. 543 – 550,
834 <http://dx.doi.org/10.1016/j.agwat.2009.09.009>, 2009.
- 835 44. Roeckner, E. and Coauthors. The atmospheric general circulation model ECHAM5:
836 Model description and simulation of present-day climate. Max-Planck-Institut für
837 Meteorologie Rep. 218, Hamburg, Germany, 90, 1996.



- 838 45. Serrano-Vicente, S.M., Beguería, S. & López-Moreno, J.I. A Multiscalar Drought Index
839 Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. J
840 CLIMATE, V. 23, pp. 1696 – 1718, <http://dx.doi.org/10.1175/2009JCLI2909.1>, 2010.
841 46. Shah, R.D. & Mishra, V. Utility of Global Ensemble Forecast System (GEFS)
842 Reforecast for Medium-Range Drought Prediction in India. J HYDROMETEOROL, V.
843 17, pp. 1781 – 1800, <http://dx.doi.org/10.1175/JHM-D-15-0050.1>, 2016.
844 47. Shah, R.D., Sahai, A.K. & Mishra, V. Short to Sub-Seasonal Hydrologic Forecast to
845 Manage Water and Agricultural Resources in India. Hydrol. Earth Syst. Sci., V. 21, pp.
846 707 – 720, <http://dx.doi.org/10.5194/hess-21-707-2017>, 2017.
847 48. Shukla, J. & Paolino, D.A. The Southern Oscillation and Long-Range Forecasting of the
848 Summer Monsoon Rainfall over India. MON WEATHER REV, V. 111, pp. 1830 – 1837,
849 [http://dx.doi.org/10.1175/1520-0493\(1983\)111%3C1830:TSOALR%3E1.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1983)111%3C1830:TSOALR%3E1.0.CO;2), 1983.
850 49. Skees, J.R. Innovations in Index Insurance for the Poor in Lower Income Countries.
851 Agriculture and Resource Economics Review, V. 37, pp. 1 – 15, [http://doi.org/](http://doi.org/10.1017/S1068280500002094)
852 [10.1017/S1068280500002094](http://doi.org/10.1017/S1068280500002094), 2016.
853 50. Souza, F.A. & Lall, U. Seasonal to Interannual Ensemble Streamflow Forecasts for
854 Ceara, Brazil: Applications of Multivariate, Semiparametric Algorithm. WATER
855 RESOUR RES, V. 39, 13 pp., <http://dx.doi.org/10.1029/2002WR001373>, 2003.
856 51. Thapliyal, V. Prediction of Indian Monsoon Variability Evaluation and Prospects
857 Including Development of a New Model. China Ocean Press, pp. 397 – 416, 1987.
858 52. Irrigation Water Use, <https://water.usgs.gov/edu/wuir.html>, accessed 3/14/2018, 2017.
859 53. van den Dool, H.M. Empirical Methods in Short-Term Climate Prediction, Oxford
860 University Press, 215 pp., 2007.
861 54. Walker, G.T. Correlations in seasonal variations of weather, IX: A further study of world
862 weather (World Weather II), Memoirs of India Meteorological Department, V. 24, pp.
863 275 – 332, 1924.

864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886



887

Tables

888

889

Table 1

Year	Probability of Above Mean	Probability of Below Mean	Observed CDI Anomaly (%)	Boxplot IQR (vertical axis units of %-anomalies)
2001	0.59	0.41	+14.4	10.9
2002	0.42	0.58	+15.5	21.0
2003	0.20	0.80	+37.8	23.1
2004	0.35	0.65	-20.1	7.70
2005	0.25	0.75	-51.3	12.1
2006	0.37	0.63	-47.9	10.0
2007	0.37	0.63	-20.5	2.60
2008	0.75	0.25	-6.33	19.1
2009	0.64	0.36	-30.0	5.10
2010	0.18	0.82	-56.4	31.1
2011	0.58	0.42	+2.72	0.19
2012	0.68	0.32	+25.4	9.90
2013	0.18	0.82	-9.36	24.6

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913



914 **Table 2**

Year	Forecast	Actual Observation	Result
2001	AM (59%)	AM	Hit
2002	BM (58%)	AM	Miss
2003	BM (80%)	AM	Miss
2004	BM (65%)	BM	Hit
2005	BM (75%)	BM	Hit
2006	BM (63%)	BM	Hit
2007	BM (63%)	BM	Hit
2008	AM (75%)	BM	False Alarm
2009	AM (64%)	BM	False Alarm
2010	BM (82%)	BM	Hit
2011	AM (58%)	AM	Hit
2012	AM (68%)	AM	Hit
2013	BM (82%)	BM	Hit

915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943



944 **Table 3**

Year	CDI Forecast Results	IMD Precipitation Forecast	Actual Precipitation	IMD Forecast Results
2001	Hit	96% of LPA	93% of LPA	Hit
2002	Miss	Not Available	68% of LPA	NA
2003	Miss	99% of LPA	40% of LPA	Miss
2004	Hit	103% of LPA	160% of LPA	False Alarm
2005	Hit	Not Available	160% of LPA	NA
2006	Hit	90% of LPA	141% of LPA	False Alarm
2007	Hit	96% of LPA	163% of LPA	False Alarm
2008	False Alarm	Not Available	95% of LPA	NA
2009	False Alarm	Not Available	212% of LPA	NA
2010	Hit	99% of LPA	199% of LPA	False Alarm
2011	Hit	98% of LPA	85% of LPA	Miss
2012	Hit	96% of LPA	46% of LPA	Miss
2013	Hit	98% of LPA	150% of LPA	False Alarm

945

946

947

948

949

950

951

952

953

954

955

956

957

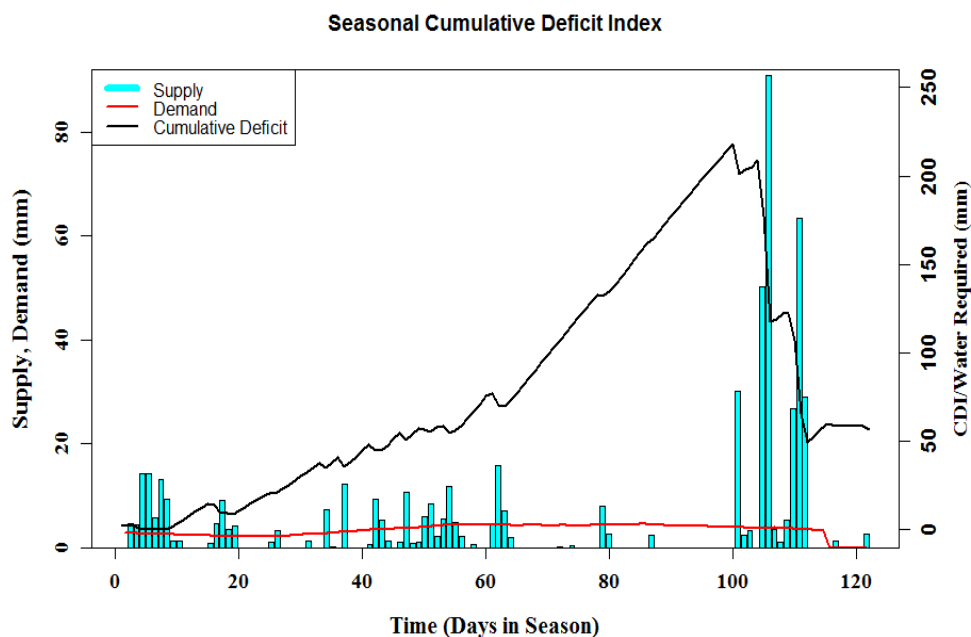
958

959



960

Figures



961

Figure 1

962

963

964

965

966

967

968

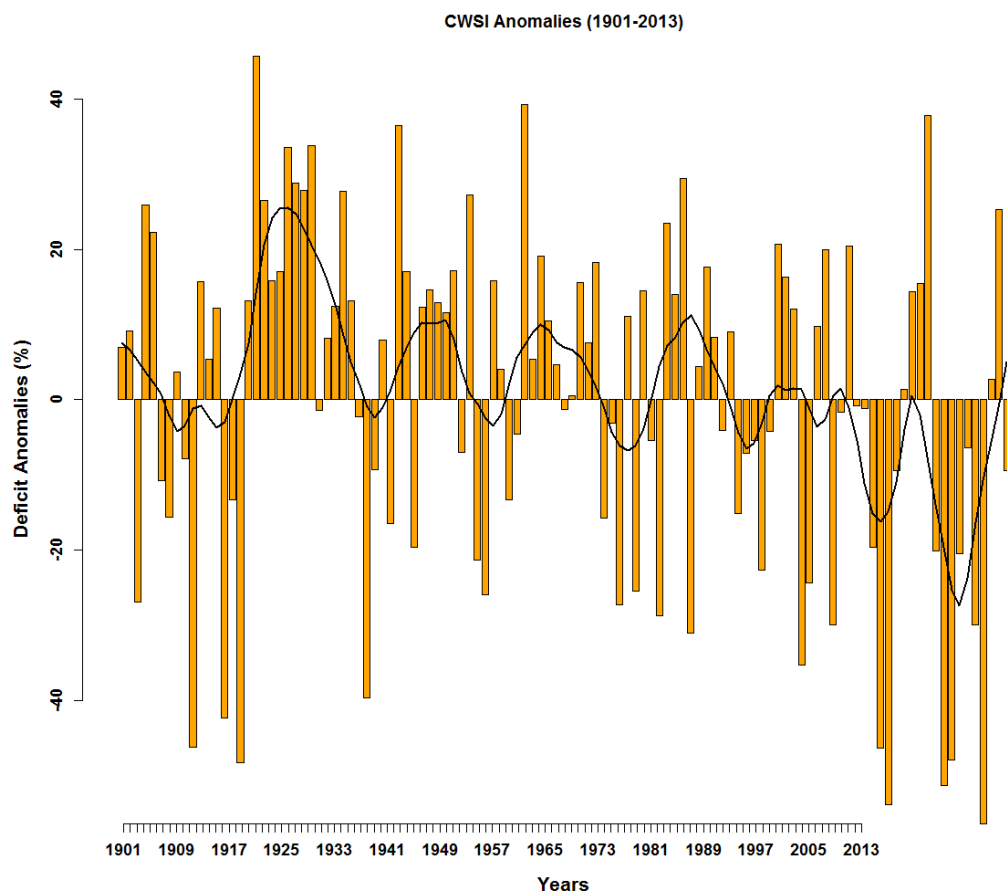
969

970

971

972

973



974

975 **Figure 2**

976

977

978

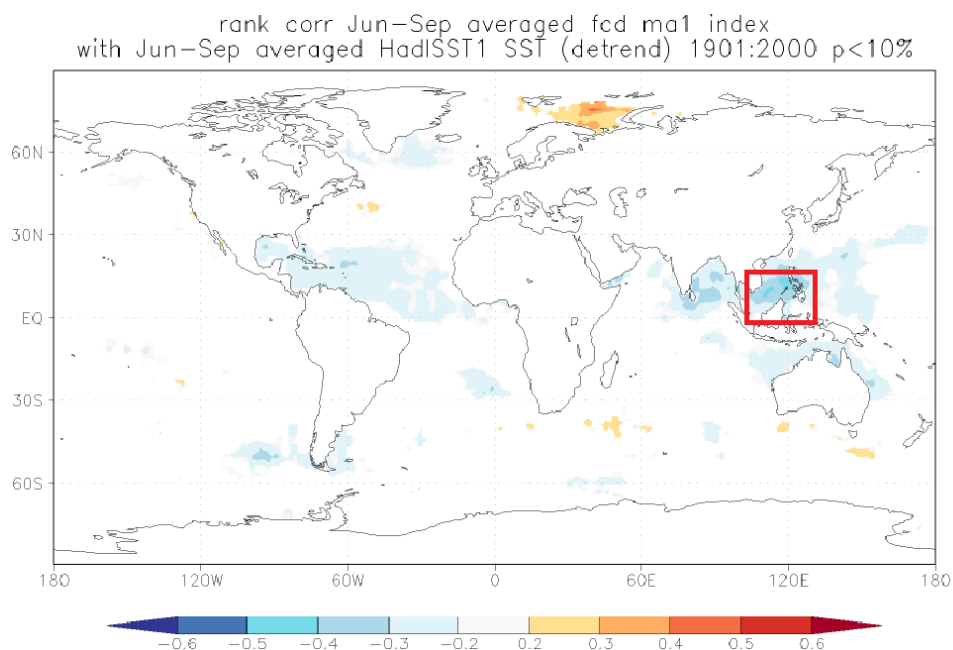
979

980

981

982

983



984

985 **Figure 3**

986

987

988

989

990

991

992

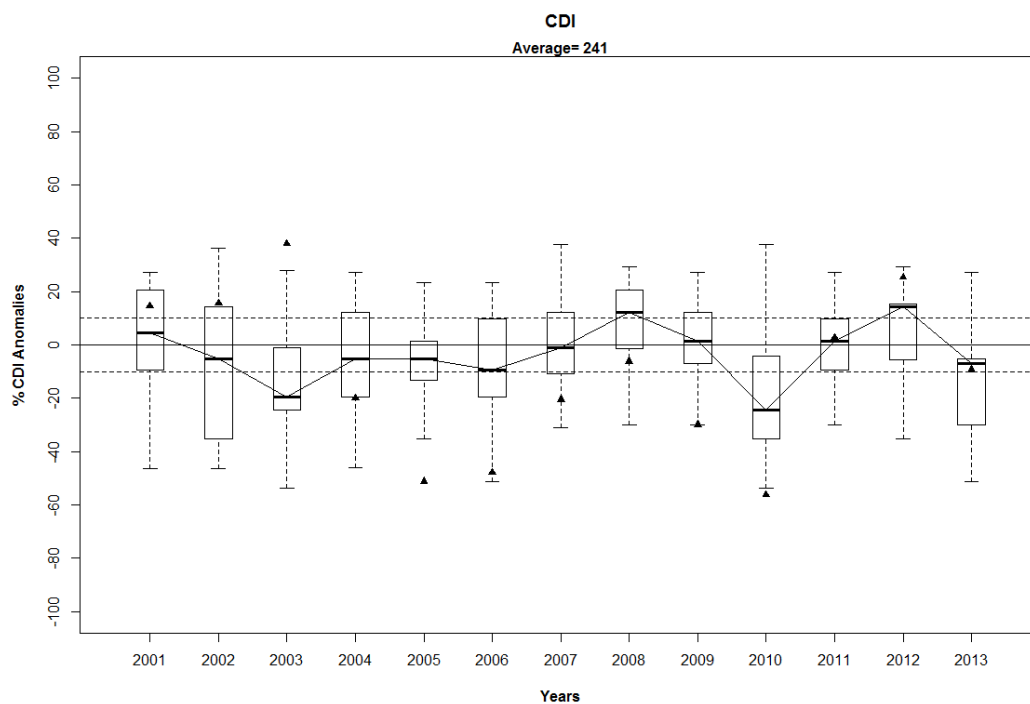
993

994

995

996

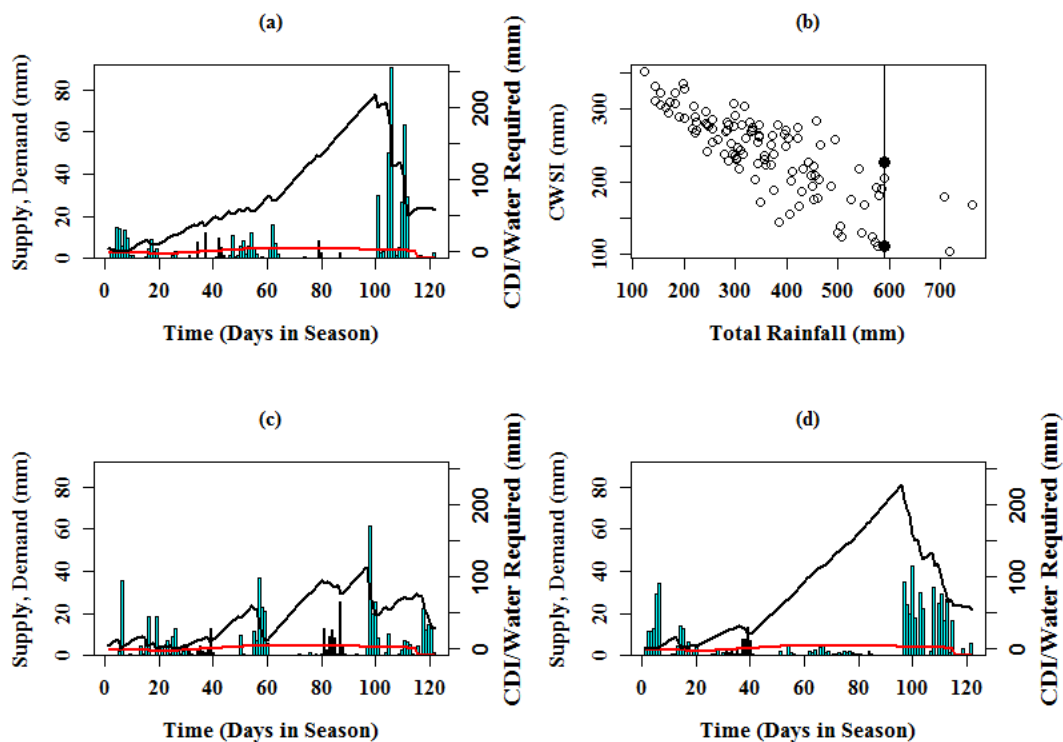
997



998
999

Figure 4

1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010



1011
1012 **Figure 5**

1013

1014

1015

1016

1017

1018

1019

1020

1021

1022

1023



Figure and Table Captions

1024

1025

1026

1027

1028

1029

1030

1031

1032

1033

1034

1035

1036

1037

1038

1039

1040

1041

1042

1043

1044

1045

1046

1047

1048

1049

1050

1051

1052

1053

1054

1055

1056

1057

1058

1059

1060

1061

1062

1063

1064

1065

1066

1067

1068

1069

Table 1: The table below shows important statistics calculated from kNN forecasts of CDI. In particular, column 2 displays the probabilities of the CDI for a particular season being above the CDI climatology. These probabilities are calculated from the kNN sampling distribution, which in turn is simulated from historical values of the CDI based on the nearest neighbors determined in the predictor variable space. Column 3 shows the complementary probabilities of being below this historical average. The forecasts for years 2001-2013 are retrospective and may serve as a cross-validation for the kNN model. Column 4 shows the values of the actual (observed) CDI anomalies with respect to the 1901-2013 climatology as percentages. A negative value implies that the actual CDI value was below the historical average by the given percentage. The rounded IQR values are shown in the final column of the table.

Table 2: The results of the kNN-generated CDI forecasts, including the most likely category (AM = Above Mean, BM = Below Mean) along with the corresponding kNN-assigned probability value expressed as a percentage in parentheses next to it (column 2), the category in which the observed anomaly value resides (column 3), and the hit/miss/false alarm designations corresponding to these results (column 4).

Table 3: A comparison of the CDI forecasts and the JJAS total seasonal precipitation forecasts generated by the India Meteorological Department (IMD). Column 2 is a repeat of column 4 in Table 2; a record of the accuracy of CDI forecasts expressed in terms of hits and misses. Column 3 contains the forecasts issued by IMD, and column 4 are the actual observations of JJAS seasonal total rainfall using rainfall data from the Satara district itself. The fifth and final column of Table 3 shows the accuracy of the IMD forecasts in terms of hits and misses using their own 5-category system.

Figure 1: A plot of the cumulative deficit index (CDI) for the JJAS season in a randomly selected year in our data set. The plot depicts the change in CDI as rainfall distribution and crop water requirement varies over the given monsoon season. The vertical cyan bars are the daily rainfall magnitudes, the slowly-changing red line is the crop water requirement (demand) and the black time series is the CDI itself. Notice how CDI increases as rainfall is either low in magnitude or sparsely distributed in certain blocks of time in the season.

Figure 2: Bar plot showing the CDI percent deficit anomalies for each of the years/growing seasons under consideration (1901 – 2013). The black, smooth time series is produced by an 11-year LOWESS smoothing of the CDI percent deficit anomalies and is meant to show the critical trends in the CDI over the entire 1901 – 2013 period.

Figure 3: Spearman rank correlation between CDI in Satara and SST field during the same JJAS season. SST region in the Indian Ocean (red box) that influences the CDI has a statistically significant correlation at the 95% significance level.

Figure 4: Boxplot diagrams depicting the kNN forecast distributions for CDI over the years 2001 – 2013 for potatoes grown in the Satara district, Maharashtra, India. Longer, more stretched out boxes indicate a greater degree of variability, or uncertainty, in the forecast



1070 distribution. Boxes in which the median is grossly off-center indicates that the forecast
1071 distribution is heavily skewed. Anomalies with respect to the climatology of the predictand were
1072 used in the boxplot calculations. As the results are presented in terms of the percent anomalies,
1073 the historical average is located at zero. The triangles represent the observations as percent
1074 anomalies about the mean.

1075

1076 **Figure 5:** The four panels pictured here depict the CDI in various ways. In panels (a), (c) and
1077 (d), the blue bars represent daily seasonal rainfall levels (in mm), the red curve represents crop
1078 evaporative water demand (ET_0) and the black time series is the CDI calculated based on this
1079 data. Panel (a) illustrates the basic nature of CDI using the daily seasonal CDI time series from
1080 the JJAS growing season of 2013. Note that this time series is specifically calculated for
1081 potatoes grown in the Satara district of Maharashtra, India during the 2013 JJAS growing season.
1082 Panel (b) shows a scatterplot of total rainfall across all growing seasons (1901 – 2013) and CDI
1083 across all growing seasons. A significant negative correlation between them is apparent from
1084 this scatterplot (Pearson correlation is -0.8, Spearman rank correlation is -0.812, Kendall rank
1085 correlation is -0.623). This panel demonstrates two different growing seasons, with two different
1086 CDI values, during which the total seasonal rainfall was the same. Panel (c) is a seasonal CDI
1087 time series plot corresponding to the growing season with the lower CDI value on the vertical
1088 line in panel (b). Panel (d) is a seasonal CDI time series plot corresponding to the growing
1089 season with the higher CDI value on the vertical line in panel (b).