Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 61 (2015) 428 – 435

Procedia Computer Science

Complex Adaptive Systems, Publication 5
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2015-San Jose, CA

Uniqueness and Causes of the California Drought

Michael B. Richman[a]* and Lance M. Leslie[a]

[a]*School of Meteorology, University of Oklahoma, 120 David L. Boren Blvd, Suite 5900, Norman, OK 73072, USA*

Abstract

The current California drought, which is part of the abnormal to extreme drought conditions affecting much of southwest USA, has lasted for 4 years (2011/12 – 2014/15). It has intensified steadily to what at present is likely the worst Californian drought since reliable instrumental records began in 1895. The uniqueness of this drought is demonstrated by assessing the Oct. – Mar. wet seasons for instances <25th percentile of precipitation and >75th percentile in average temperature. Of the 8 seasons since 1895 that met these percentile conditions, only the present drought satisfied these criteria for more than one season. Predictions of California precipitation and temperature anomalies were made using linear regression (LR), and support vector regression (SVR) with several linear and non-linear kernels, applied to a range of climate drivers and local sea surface temperatures (SSTs). Cross-validated correlations were low (LR) to moderate (SVR) for precipitation, but were high (>0.7) for both temperature LR and SVR, with SVR marginally exceeding LR. The leading predictors were global warming and local SSTs near the California coast. Finally, the cool seasons were classified as dry/not-dry and hot/not-hot using logistic regressions and k-means classification clustering. Again, it was found that predictability was low for dry/not-dry classes but was high (>70% correct) for hot/not-hot classes. This research suggests that the climate system has warmed sufficiently so that drought can no longer be assessed solely by the lack of precipitation, but must consider the combination of low precipitation and abnormal warmth.

Peer-review under responsibility of scientific committee of Missouri University of Science and Technology

Keywords: Drought; Climate Change; Prediction; Classification; Machine Learning

1. Introduction

Drought in California is not a recent phenomenon. Devastating droughts have occurred many times before, extending back thousands of years in the historical record, well before the reliable instrumental record period that runs from 1895 to the present. During the period of instrumental records, there have been droughts that have lasted

* Corresponding author. Tel.: +1-405-325-1853; fax: +1-405-325-7689.
E-mail address: mrichman@ou.edu

Peer-review under responsibility of scientific committee of Missouri University of Science and Technology
doi:10.1016/j.procs.2015.09.181

for multiple years and produced massive social and economic losses. Notable cases include the long-lasting "Dust Bowl" drought of the late 1920s – early 1930s; the short but costly drought of 1976–77; and the drought from the late 1980s to the early 1990s. Each drought resulted in attempts either to increase the water supply or limit water usage. At present, California again is in a drought period that has lasted four years and produced record cumulative rainfall deficits over much of the state. The size of the current rainfall deficit suggests that it is the worst drought during the instrumental record. Although brief compared with some historical droughts, it likely is more extreme than longer droughts [1]. As of the 28 July 2015, nearly 50% of California was encompassed by "exceptional drought" (the most severe category), over 70% of the state under extreme and exceptional drought and over 97% under one of the four drought categories (Fig. 1). The cool/wet season (Oct. – Mar.) has ended and most of California now faces ~ 6 months of low rainfall, in keeping with its Mediterranean climate. The upcoming 2015-16 cool/wet season is critical to ameliorating the drought, and it is possible that the developing El Niño conditions in the equatorial Pacific Ocean will bring relief from the drought as, in the past, strong El Niño conditions commonly have prevented, limited or ended drought conditions.

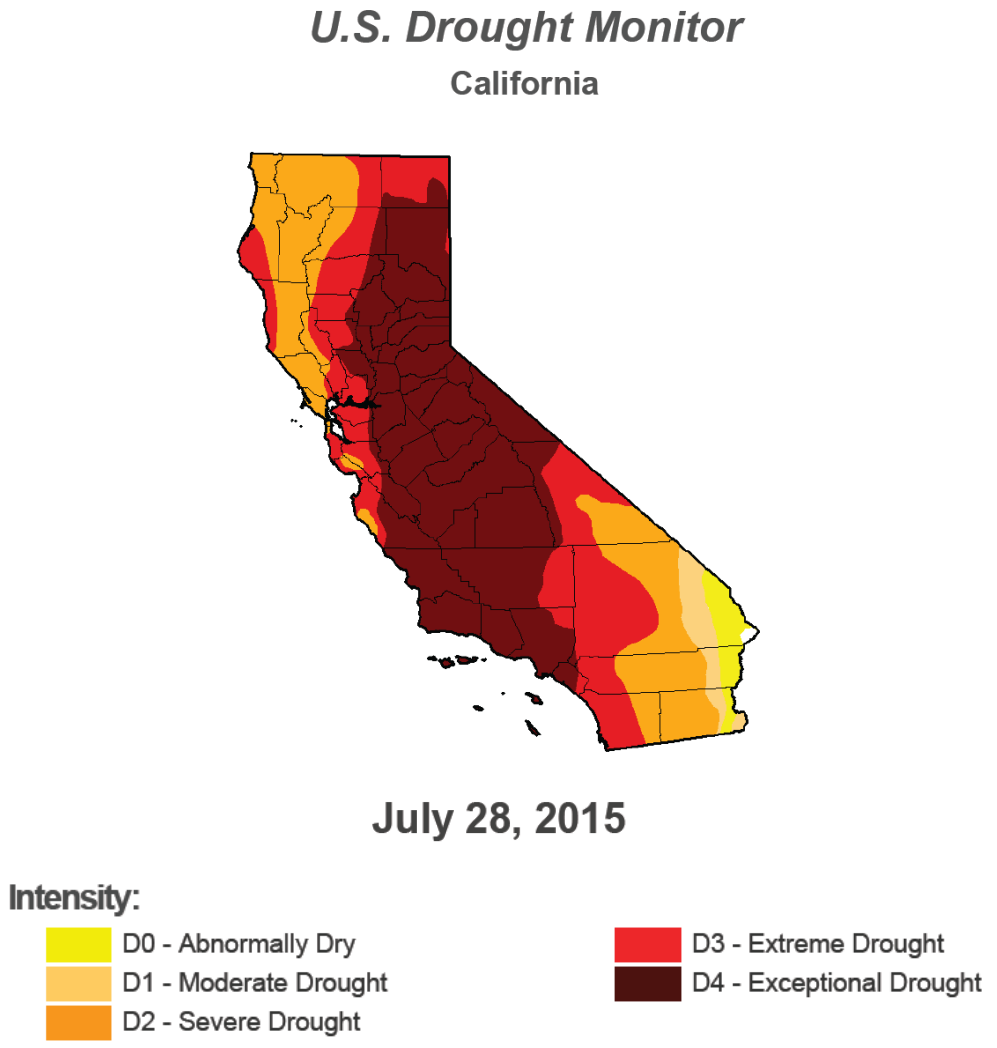

Fig. 1. USDA drought assessment for California 28 July 2015 (released 30 July 2015).

The consequences of drought in California are numerous and include mandatory water usage limitations, reduced of hydroelectricity generation, increased energy use for household cooling, transport, treatment of water and wastewater, decreased irrigation (drastically reducing food production), enhanced risk in wildfires and heat-waves, deforestation, landslides/mudslides when rain falls on dry soil loosened by parched vegetation, diminished snowpack that provides ~30% of California's water supply, loss of animal and plant species, worsening air and water quality (increased pollution), and human health impacts such as heat stress.

Despite the importance of California droughts, relatively few research studies have focused on their possible links with large-scale climate drivers, and their a priori detection (e.g., [2, 3]). Moreover, most view the current drought as a precipitation deficit event. Furthermore, the present drought, and future drought risk in general, are studied mainly in the context of human-induced climate change (e.g., [4, 5]). The role of temperature, in tandem with precipitation, has remained largely unexplored, despite higher temperatures affecting other parts of the water cycle, such as albedo, evapotranspiration, and the dramatic reduction of the above-mentioned California snowpack, which provides a late cool season water supply through melting.

2. Data and Methods

2.1. Data

Whereas there is no single definition of drought, there is agreement that drought is caused by a prolonged period of below normal precipitation, resulting is a loss of surface and groundwater. Water balance is a function of input (precipitation) and output (evaporation, water use). As evaporation is a strong function of temperature, its role in drought cannot be ignored, particularly as the climate warms [6]. Accordingly, precipitation and temperature data from the National Oceanographic and Atmospheric Administration (NOAA) have been obtained for the period 1895 to the present for California. Those data are available statewide and for subareas of the state, known as climate divisions. The statewide values are used herein. As climate patterns spanning large distances on the globe, known as teleconnections, are often used to explain persistent climate anomalies, a series of indices defining the patterns have been obtained from NOAA. California has a Mediterranean climate with the vast majority of precipitation falling in the cool season (Oct. - Mar.). Thus, data in this study are limited to the cool season: the response variables are California precipitation and temperature (1895-2015) [from http://climate.gov]. The following predictors (known as climate drivers) are applied: global air temperature (GT), Niño3.4 (sea-surface temperatures (SSTs) from 5N to 5S and from 120W to 170W), local SSTs in 5 degree latitude/longitude grid boxes centered at 25N, 180 and 40N, 130W and a set of climate indices/oscillations {Trans-Niño (TNI), Atlantic Multidecadal (AMO), Pacific Decadal (PDO), North Pacific (NPI), Tropical North Atlantic (TNA), and the Western Hemisphere warm pool (WHWP) for 1948-2015 [from http://esrl.noaa.gov]}. All predictors are preprocessed (normalized). The California precipitation data were summed for each cool season and that cool season total was expressed as an anomaly from the 1981-2010 average or climatology. The California temperature data were averaged over the cool season and expressed as an anomaly from the 1981-2010 climatology. These data served as the basis for the prediction portion of the study. Additionally, the precipitation (temperature) data were ranked from the driest (coolest) to wettest (hottest) cool seasons. Various percentiles of precipitation (i.e., 10, 20, 25, 33) and temperature (i.e., 67, 75, 90, 95) were used to convert the numerical data into a nominal scale as [0 = no drought, 1 = drought] for the classification section of the research.

2.2. Methods

Multiple linear regression (LR) and support vector regression (SVR) [7], with polynomial (poly) and radial basis function kernels (RBF), were applied to the aforementioned climate data to obtain predictions of the rainfall and precipitation anomalies. These analyses were subject to 10-fold cross validation to obtain the correlation, mean absolute and root mean squared errors. The classification for drought/no drought used logistic regression with and without a threshold adjustment [8] and classification with k-means clustering [9]. Forecast evaluation indices were applied to assess the performance of each classification model.

3. Results

3.1. Uniqueness of the present drought

Precipitation total (average temperature) data for the wet/cool seasons of 1895/96 – 2014/15 were ranked from driest to wettest (coolest to warmest). Various percentiles were examined to assess the uniqueness of the combination of dry and hot seasons. Of the 30 seasons identified at or below the 25th percentile of precipitation (Fig. 2a), there were 16 cases that lasted for one year, 5 that lasted two years, 0 that lasted three years and 1 that lasted four years (2011/12 – 2014/15). Similarly, of the 30 seasons identified at or above the 75th percentile of temperature (Fig. 2b), there were 18 cases that lasted for one year, 3 that lasted two years, and 2 that lasted three years (2001/02 – 2003/04 and 2012/13 – 2014/15). There were eight years that had the combination of the lowest quartile of precipitation and highest quartile of temperature. Of those eight, only one (2012/13 – 2014/15) persisted for more than a single cool season. This establishes the uniqueness of the current drought and provides strong evidence it is a combination of low precipitation and high temperatures (i.e., a heat drought). Additionally, there has been an upward trend in cool season temperature since the 1950's, with few years from 1990 to the present below the 75th percentile (Fig. 2b).

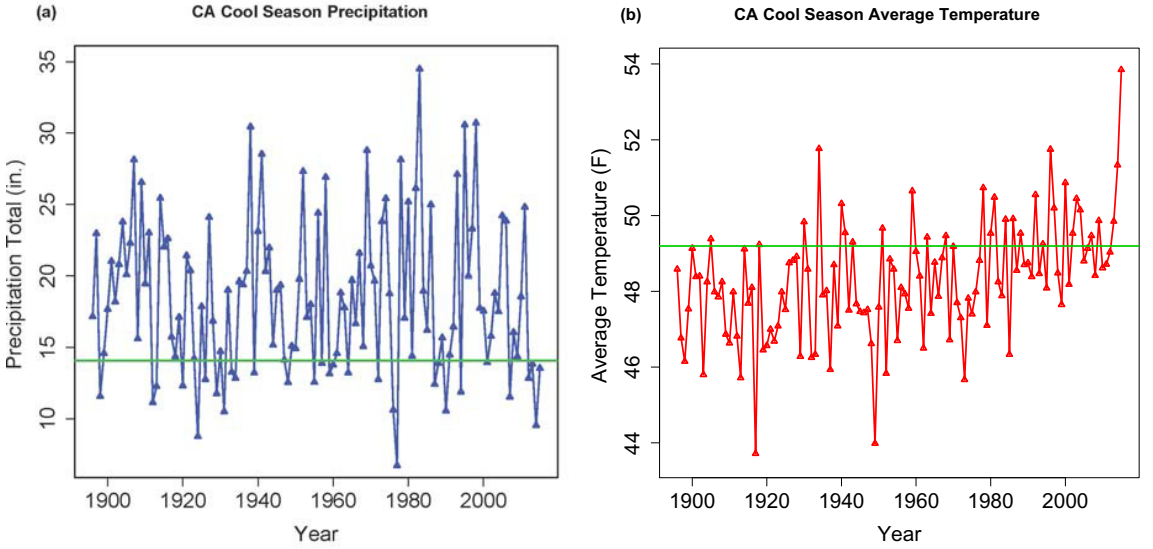

Fig. 2. (a) Precipitation totals for the wet/cool seasons (Oct. – Mar.) 1895/96 – 2014/15; (b) average temperature for the wet/cool seasons 1895/96 – 2014/15. Green line represents the 25th percentile of precipitation (a) and the 75th percentile of temperature (b).

3.2. Prediction of precipitation and temperature anomalies

3.2.1. Precipitation anomaly prediction

As climate index data exist from 1948 to the present, the wet/cool seasons of 1948/49 – 2014/15 were used to train and test all the models. Model performance was made using 10-fold cross-validated LR and SVR. The predictors were selected for LR in a stepwise fashion using the Akaike Information Criterion (AIC) with forward selection and backward elimination. The use of AIC favors small residual error in the model selected and penalizes for adding additional predictors to avoid overfitting. Its use combined with cross-validation helps to insure the results obtained are reliable (generalize). Table 1 lists the predictors selected and accuracy statistics. The SVR was optimized for two kernel classes (poly, RBF). The polynomial kernel with the smallest generalization error had an exponent value (E) of 1 and complexity or regularization parameter (C) of 0.5, whereas the RBF kernel had a gamma value (G) of 0.005 and C=50.

Table 1. Model results of prediction of California precipitation anomalies.

Model	Predictors selected	Cross-validated Correlation	Cross-validated MAE (inches)	Cross-validated RMSE (inches)
LR	SST (25N, 180), TNI, TNA	0.188	4.573	5.929
SVR (Poly, E=1, C=0.5)	SST (25N, 180), TNI, TNA	0.416	4.099	5.512
SVR (RBF, G=. 005, C= 50)	SST (25N, 180), TNI, TNA	0.381	4.105	5.527

The predictor weights for LR are divided nearly equally between the SST (representing upstream forcing of the atmosphere associated with abnormally cold water at the International Dateline in the subtropics) and the effect of the above average geopotential heights in the subtropics associated with the TNA. The TNA is associated with dry conditions in California [10]. The TNI, a pattern representing the evolution of El Niño, has a weight about ⅓rd that of TNA, suggesting a weak impact of that deep tropical forcing on west coast precipitation [11]. In contrast, for SVR with polynomial kernel, the weights were distributed more equally among the three predictors. Given that the long-term mean precipitation was 18.42 inches, MAE values of ~ 4.1 – 4.6 inches suggests none of the results of the prediction experiments were particularly accurate. However, the SVR predictions are approximately 10% (7%) more accurate in MAE (RMSE) than LR, with double of the correlation statistic.

3.2.2 Temperature anomaly prediction

Temperature anomaly prediction used the same LR and SVR techniques as noted in 3.1.1. The mean cool season temperature in California was 48.29°F. The complexity parameter on the polynomial kernel SVR is somewhat smaller for temperature (Table 2) than for precipitation. However, as was the case for the precipitation

Table 2. Model results of prediction of California temperature anomalies.

Model	Predictors selected	Cross-validated Correlation	Cross-validated MAE (°F)	Cross-validated RMSE (°F)
LR	SST (40N, 130W), GT, PDO, Niño3.4, NPI	0.706	0.920	1.090
SVR (Poly, E=1, C=0.35)	SST (40N, 130W), GT, Niño3.4	0.723	0.872	1.072
SVR (RBF, G=. 0125, C= 15)	SST (40N, 130W), GT, Niño3.4	0.724	0.873	1.069

analysis, the exponent of 1 minimized the generalization error. Similarly, for the RBF kernel, the complexity function was smaller than for the precipitation analysis. The prediction performance of the temperature anomaly analyses was considerably different than for precipitation, with much larger correlations and lower MAE, RMSE. It is noteworthy that SVR models required only three predictors to achieve slightly more accurate results than LR, which selected five variables using the AIC. To gain physical insight into the prediction of temperature anomalies, it was useful to examine the relative weights of the predictors. For the LR, the largest weights were both positive in sign, for the GT and local SST off the coast of California (40N, 170W). For the local SST, it has been established that warm water is associated with atmospheric ridging and warm air temperatures [12]. The remaining weights had magnitudes much smaller (<⅓rd that of the aforementioned two) and were all negative, suggesting Niño3.4, PDO and NPI are weakly associated with cooling. It is noteworthy that the GT and SST signals were the leading drivers of heat in California. Exploration of the relationship between the GT driver and the surface air temperature was particularly revealing (Fig. 3) showing the strongest warming off the west coast of the United States extending east into the California region.

3.3 Classification of drought

3.3.1 Classification of dry Oct. – Mar. (cool) seasons

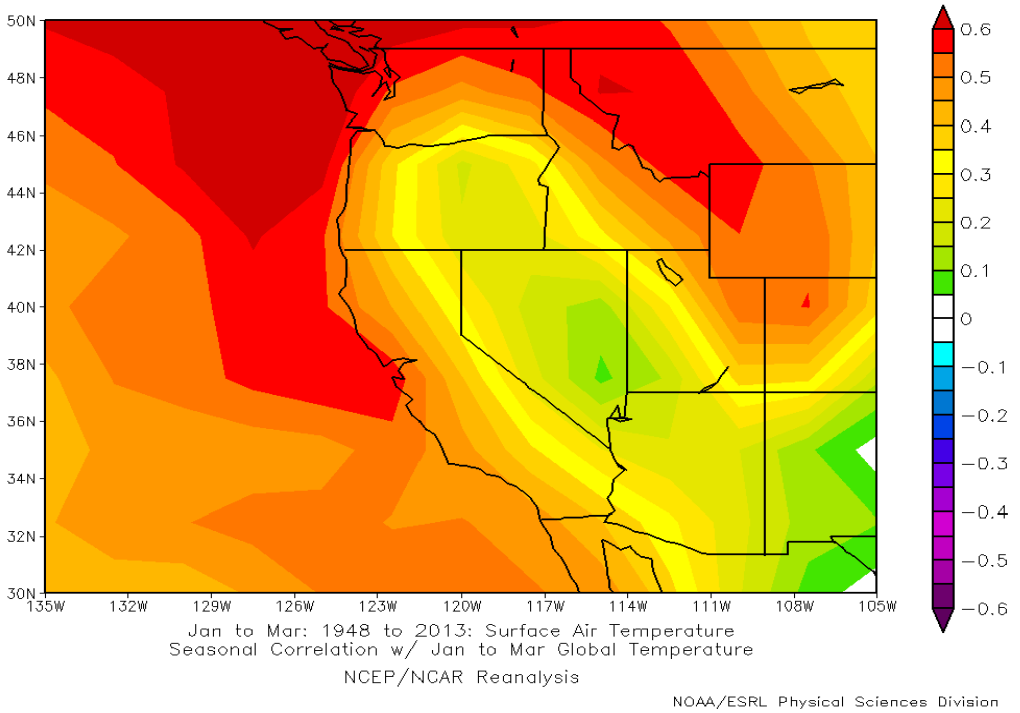

Fig. 3. Correlation between surface air temperature and global warming climate driver.

Precipitation totals for each cool season that were ≤25th percentiles of the historical values (Fig. 1a, green line) were classified as drought (1) and totals >25th percentile placed into a class of not drought (0). The aim of the classification analyses was to correctly predict the drought/no drought cases. Logistic regression using the aforementioned climate predictors was applied to the 67 wet/cool seasons of 1948/49 – 2014/15. During that period, 49 cases (~ 73.1%) had no drought and 18 (~ 26.9%) had drought. This is fairly close to the expected relative frequencies of 75 and 25%, respectively. Logistic regression classified nearly 63% of the cases correctly (Table 3). However, closer inspection of the results revealed that the no drought cases comprised the majority of the correct classifications (nearly 60% of the 49 cases) whereas only 2 (of 18) drought cases were correctly assigned to their observed class. This manifested as a large number of missed predictions, where 16 (of 18) droughts were incorrectly classified. Since the design was not balanced, a threshold adjustment was attempted to identify more droughts correctly. The results of that adjustment (Table 3) show an improvement in the correct identification of drought (nearly 78% of the 18 cases). However, there was a large false alarm problem as the majority of the no drought years were incorrectly classified as drought. Classification using k-means clustering fell somewhere between the two logistic regression methods, reducing the false alarms by 50% over the logistic regression with threshold adjustment but having three times the number of misses. The performance of the classifications is examined more formally through a set of indices, including the probability of detection (POD), the fraction of positive events that were correctly forecast. The shortcoming of logistic regression in POD is improved greatly by application of the threshold adjustment. All methods had poor performance as measured by the false alarm ratio (FAR), which is the fraction of positive forecasts in which the event did not occur. The probability of false detection (POFD) measures the ratio of false alarms to the total number of "no" (class 0) observations. Here the tradeoff in the threshold adjustment is seen as it results in a large number of false alarms. The bias measures the ratio of the frequency of forecast class 1 events to the frequency of observed class 1 events. An ideal bias value is 1.

Values substantially <1 represent a system that underforecasts dry events (e.g., logistic regression), whereas values substantially >1 are associated with systems that overforecast such events (e.g., logistic regression with the threshold adjustment and, to a lesser degree, classification clustering). Heidke skill score (HSS) measures the accuracy relative to a random forecast. None of the techniques had positive skill in forecasting dry seasons. Overall, these forecast evaluation results reinforce the difficulty in classifying seasonal precipitation using large-scale climate patterns as predictors.

Table 3. Results of classification of California for dry Oct.-Mar. seasons. Predicted (P) versus Observed (O) dry (1) and not dry (0) and various forecast evaluation statistics.

Model	Predictors selected	Correct Positive (P=1, O=1)	False Alarm (P=1, O=0)	Miss (P=0, O=1)	Correct Negative (P=0, O=0)	Correctly classified (%)	POD	FAR	POFD	BIAS	HSS
Logistic Regression	All	2	9	16	40	62.69	.111	.818	.184	.611	-.083
Logistic Regression threshold adjustment	All	14	42	4	7	31.34	.778	.750	.857	3.111	-.048
Classification Clustering	All	6	21	12	28	50.75	.333	.778	.429	1.500	-.095

3.3.2 Classification of warm Oct. – Mar. (cool) seasons

The average cool season temperature anomalies were placed into the class warm (1) if a seasonal value was ≥ the long-term 75th percentile (Fig. 1b, green line). Otherwise, the season was classified as not warm (0). For the 67 cases tested, 23 (~ 34.3%) had warm conditions and 44 (~ 65.7%) did not. Unlike the precipitation analyses, these percentages are not close to the expected relative frequencies, since the temperature record is clearly nonstationary, with a much larger number of warm cases that would be expected from a stationary process over the past 30 years. Logistic regression correctly classified over 70% of the cases (Table 4). Inspection of the 23 seasons that were warm showed that ~ 47.8% were classified correctly. Similarly, for those seasons that were not warm, ~ 81.8%

Table 4. Results of classification of California for warm Oct. – Mar. seasons. Predicted (P) versus Observed (O) warm (1) and not warm (0) and various forecast evaluation statistics.

Model	Predictors selected	Correct Positive (P=1, O=1)	False Alarm (P=1, O=0)	Miss (P=0, O=1)	Correct Negative (P=0, O=0)	Correctly classified (%)	POD	FAR	POFD	BIAS	HSS
Logistic Regression	All	11	8	12	36	70.15	.478	.421	.182	.826	.309
Logistic Regression threshold adjustment	All	14	11	9	318	70.15	.609	.440	.250	1.087	.351
Classification Clustering	All	6	18	5	26	65.67	.783	.500	.409	1.565	.329

were classified correctly. There were slightly more misses than false alarms with this method. After application of a threshold adjustment, the logistic regression decision boundary shifts to classify correctly ~ 60.9% of the warm cases and the number of misses decreases. The tradeoff is slightly lower percentage of not warm seasons was

classified correctly (75%) and the false alarms increase. Classification using k-means clustering correctly identified ~ 78.3% of the warm cases with a much smaller number of misses (Table 4), but had less accuracy in correct identification of those cases that were not warm seasons (~ 59.1%) and more false alarms. The forecast evaluation measures summarize the performance of the various techniques applied. The POD is closest to the ideal value (1) for classification clustering. However, classification clustering had more false alarms, resulting in large values of FAR and POFD. Of the three methods, the logistic regression with threshold adjustment had close to ideal bias and the largest HSS, though all methods were superior to a random forecast. Given the base rate (34.3%) is less than a random guess, the skill scores > .3 are notable.

4. Conclusions

California is in the midst of a prolonged drought. Since 1895, there have been other instances of precipitation deficit but, because of below average-to-average temperatures and much lower population, multi-year droughts were not as severe nor did they have as much impact as the present drought. The current drought is a combination of an unprecedented period of four consecutive dry years combined with three concurrent warm years, one of which was the warmest ever recorded since 1895. This combination has led to a heat drought, with little accumulated snowpack and spring temperatures rising rapidly owing to the paucity of snow. The higher temperatures lead to higher evaporation rates, and drier soil in a positive feedback with air temperature. This research suggests that the climate system has warmed sufficiently that drought can no longer be assessed solely by the lack of precipitation, but must consider the combination of low precipitation and abnormal warmth caused by the global warming signal. The support vector regression analyses predicting the magnitude of the precipitation and temperature anomalies show that temperature is far easier to predict than precipitation, given a set of climate drivers as well as local SST predictors. The leading predictors (those with the largest weights) for anomalously warm temperatures are the global warming signal and warm ocean waters off the California coast. There is a weaker contribution from the tropical Niño3.4 box SSTs. The classification of cool seasons as either dry or not dry and hot or not hot was carried out with logistic regression and k-means classification clustering. The models tested did a poor job of selecting the correct class for the precipitation (dry/not dry). However, for temperature (hot/not hot), about 70% of the seasons were correctly classified. The addition of the threshold adjustment in logistic regression shifted the decision boundary, to identify more drought (i.e., dry and hot) conditions. However, this decreased accuracy of predicting drought, when it does not occur, was the trade off. Thus, careful consideration needs to be given to the costs involved with policy decisions associated with misses and false alarms in drought classification.

References

1. Griffin D, Anchukaitis KJ. How unusual is the 2012-2014 California drought? *Geophys. Res. Lett*. 2014; **41**, 9017-9023.
2. Fierro, AO. Relationships between California rainfall and large-scale climate drivers. *Int. J. Climatol*. 2014; 34, 3626-3640.
3.Mao X, Nijssen, B, Lettenmaier DP. Is climate change implicated in the 2013-2014 California drought? A hydrologic perspetive. *Geophys. Res. Lett*. 2015; **42**, 2805-2813.
4. Mann M, Gleick PH. Climate change and California drought in the 21st century. *Proc. National Academy of Sciences* 2015; **112**, 3858-3859.
5. Diffenbaugh NS, Swain DL, Touma, D. Anthropogenic warming has increased drought risk in California. *Proc. National Academy of Sciences* 2015; **112**, 3931-3936.
6. Jeong D I, Sushami L, Khaliq MN. The role of temperature in drought projections over North America *Climatic Change* 2014; **127** 289-303.
7. Drucker H, Burges CJC, Kaufman L, Smola A and Vapnik VN, Support vector regression machines. *Adv. in Neural Information Processing Systems* 1997; **9** 155-161.
8. Richman MB, Leslie LM. Attribution and prediction of maximum temperature extremes in SE Australia. *Procedia Computer Science* 2014; **36**, 612-617.
9. Li D, Wong KD, Hu YH, Sayeed KM. Detection, classification and tracking of targets *Signal Processing Magzine (IEEE)* 2002; **19**, 17-29.
10. Barnston AG, Livezey RE. Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Mon. Wea. Rev.* 1987; **115**, 1083-1126.
11. Kennedy AM, Garen DC, Koch RW. The association between climate teleconnection indices and Upper Klamath seasonal streamflow: Trans-Niño Index *Hydrological Processes* 2009; **23** 973-984.
12. Namias J, Yuan X, Cayan DR. Persistence of North Pacific sea surface temperature and atmospheric flow patterns. *J. Climate*, 1988; **1**, 682-703.