Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 36 (2014) 612 – 617

Procedia Computer Science

Complex Adaptive Systems, Publication 4
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2014- Philadelphia, PA

Attribution and Prediction of Maximum Temperature Extremes in SE Australia

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

[a]*School of Meteorology, University of Oklahoma, 120 David L. Boren Blvd., Norman, OK 73072, United States*
[b]*Cooperative Institute for Mesoscale Meteorological Studies, 120 David L. Boren Blvd., Norman, OK 73072, United States*

Abstract

Over half of Australia's population occupy its southeastern quadrant. Temperature records for the 56-year period 1958-2013 reveal increasingly hot summers since the 1990s, with daily maximum temperatures reaching 10°C above normal. The change in monthly mean maximum temperatures (~1°C to 1.5°C above the long term mean) far exceeds the natural variability expected over a half-century. Numerous maximum temperature records have been set and the extreme heat poses a major socioeconomic threat. This work seeks climate drivers that are useful predictors of the warm mean monthly values of maximum daily temperatures for January, in southeastern Australia. The data for January 1958-2013 from one representative site, Tibooburra, is coded, in a binary sense (excessive heat – yes/no), and for actual temperature anomalies. One challenge in analyzing these data is the short records relative to the numerous possible climate drivers of excessive heat. The variables are a combination of ocean and atmospheric climate drivers plus their high and low frequency filtered values from wavelet analysis. Several feature selection methods are applied to produce a compact set of predictors exhibiting good generalization properties. Results of cross-validation of logistic regression, with and without threshold adjustment, show that cold air blocking, and teleconnection patterns, such as the Southern Annular Mode (SAM), have statistical skill (best classification Heidke skill score = 0.34) in forecasting extreme heat for binary forecasts, with correct forecasts exceeding 75% of cases. For predicting actual monthly anomalies, support vector regression and bagged trees explain anomaly temperatures with mean absolute error of 1.4°C and 1.3°C.

Keywords: Climate change; feature selection; prediction; cross-validation; wavelets

1. Introduction

Since the early 1990s, extreme maximum temperature records in Australia, particularly southeastern Australia, have been set and then subsequently broken in following years [1]. The increasing mean monthly maximum temperatures at observation sites over much of southeastern Australia, in both the warmer (November-April) and cooler seasons (May-October), motivate this study. The present study seeks a set of predictors, known as climate drivers, which show skill in predicting if a summer month is extremely hot in a binary sense (yes/no) and in the

* Corresponding author. Tel.: +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.2014.09.063

numerical value of the temperature anomaly. Notably, 2014 saw maximum temperature records set in the warm season with impacts such as 139 deaths over the average expected for the period from 13 to 23 January [2].

1.1. Feature selection

A challenge in classification and prediction, specifically, and machine learning in general, is to find ways to reduce the dimensionality n of the feature space to minimize the risk of "overfitting". Data overfitting arises when the number n of input features is large (e.g., dozens of climate drivers) and the number ℓ of training instances is comparatively small (e.g., 56 summer months). When overfit, a decision function that separates the training data is identified; however, that function will perform poorly on independent test data, resulting in poor generalization. Techniques, such as support vector machines (SVMs) and support vector regression (SVR) that use regularization [3] reduce overfitting of the data without requiring space dimensionality reduction. Despite this, SVM and SVR benefit from dimensionality reduction to improve the computational efficiency and generalization properties. The most common form of dimensionality reduction preprocesses the data with principal components (PCs) to reduce the number of predictors. However, that procedure has a disadvantage as the new features (the PCs) are linear combinations of the original data, and all n features must be used to define them. Therefore, PC analysis is not particularly useful for feature selection. A more efficient approach is to prune those features that do not support the best generalization in the classification or prediction. The techniques to accomplish the pruning include those computationally impractical for a large number of predictors (e.g., all possible combinations the features), greedy algorithms that divide the problem into pieces, finding the optimal solution to each piece [4] and SVM recursive feature elimination (RFE) [5].

1.2. Documentation of extreme heat

As shown in Figure 1, and discussed in [1] for southeast Australia, the extreme heat showed a distinct upward trend over the period of record and a higher incidence of extreme heat after 1992. Therefore, the 1958-1992 data were used to create empirical percentiles of maximum monthly temperature in for January, representing the climate prior to the majority of the warming. Those months with mean maximum temperature values exceeding the 90th percentile (p90) were identified (Fig. 1) as exceeding a temperature anomaly of +1.32°C.

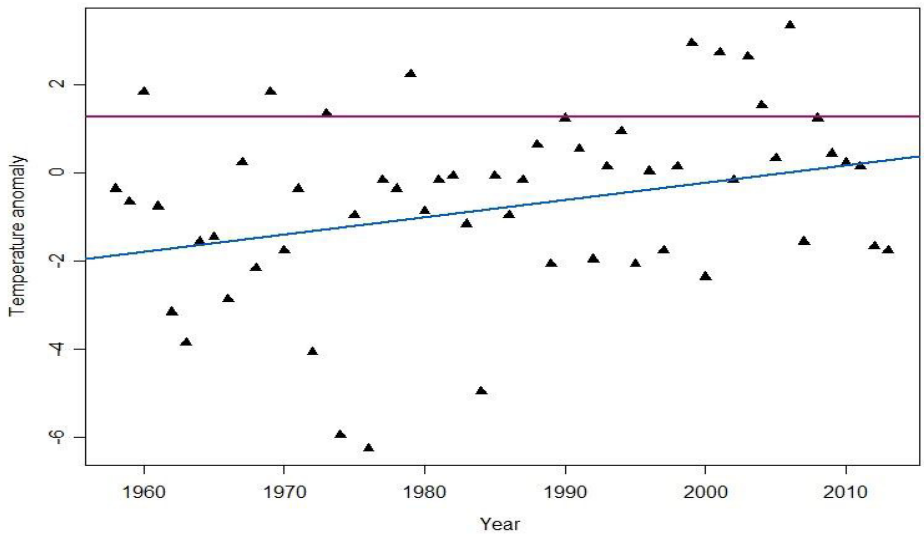

Fig. 1. Tibooburra January maximum temperature anomaly ($^{\circ}$C). The horizontal red line is +1.32°C temperature anomaly representing the 90% threshold used to classify the months with extreme heat (those months identified above the red line). The blue line represents the linear fit.

2. Data and methodology

2.1. Data

The data set consists of the monthly means of the daily maximum temperature at Tibooburra, a site in southeast Australia. The period of record comprises the 56 January months for 1958-2013. An atmospheric pattern, associated with above normal temperature in southeast Australia, is investigated through a blocking pattern index [6] centered at longitudes 140°E and 160°E, respectively, for the same months as the temperature data. The climate drivers included the following indices: Southern Annular Mode (SAM), Arctic Oscillation (AO), Pacific North American teleconnections (PNA), Dipole Mode Index (DMI) for the Indian Ocean, Pacific Decadal Oscillation (PDO), North Atlantic Oscillation (NAO), Modoki (a variant of El Niño), four El Niño indices (Niño1+2, Niño3, Niño4 and Niño3.4) and the Southern Oscillation (SO).

2.2. Classification and prediction methods

The classification was accomplished by application of logistic regression with a threshold adjustment with 5-fold cross validation where those January months with maximum temperatures exceeding +1.32°C were labeled as the class having extreme heat, and the remainder of the data were labeled as the class not having extreme heat. Prediction was accomplished by application of SVR and bagging trees. Evaluation of the models was made through a series of forecast evaluation indices and error measurements applied to testing data.

3. Results and discussion

3.1. Attribute Selection

A wavelet analysis of each climate driver identified significant peaks in global wavelet power for each of the climate drivers and blocking indices. The leading high and low frequency peaks in power were included in the predictor pool, resulting in a total of 64 potential predictors. As the number of potential predictors (attributes) was greater than the number of cases, feature selection was important. Two approaches, the SVR RFE technique and a greedy stepwise procedure, ranked the attributes. The common predictors selected by both approaches were: (1) blocking at 140°E, (2) SAM, and (3) the high frequency portion of the SAM identified by wavelet analysis. The SAM time series was deseasonalized by removing November to April mean value and then standardized. The resulting time series for both time series is shown in Fig. 2. An ordinary least squares trend line and a resistant Theil-Sen [7] line (using medians) were fit, with virtually identical results.

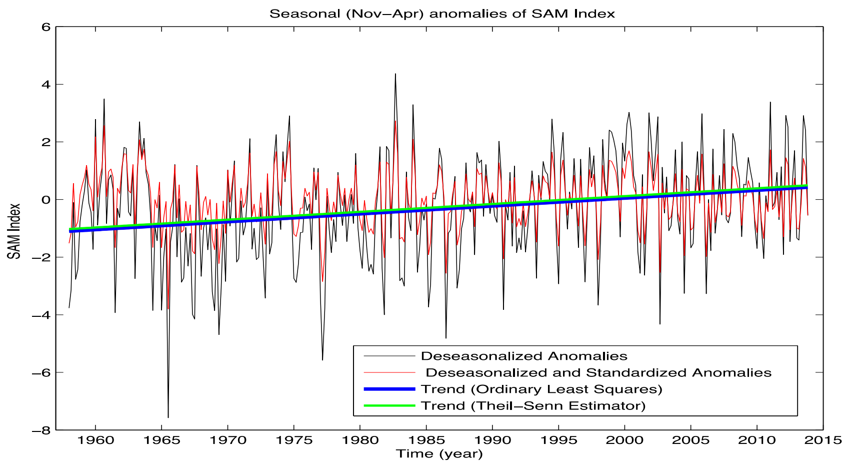

Fig. 2. Time series of November - April SAM anomalies and deseasonalized anomalies shown for the data period.

The local wavelet power spectrum as a function of time is shown in Fig. 3 (left panel). Noteworthy features are a consistently large power between 1 to 2 years (high frequency) and a second peak centered at ~14 years (low frequency). Note the dominant periods of the signal are enclosed by the black contour lines (e.g., the 1980s) and recur throughout the full period. In contrast, the low frequency dominant period peaked earlier in the record (e.g., 1958-1975). The periodocity of these signals is confirmed by the global power spectrum (Fig. 3, right panel), which shows clear peaks at the aforementioned frequencies. Note that Fig. 3b shows the standard 95% confidence level as a dashed line – at 90%, the dashed line shifts so that the high frequency peak becomes significant (not shown).

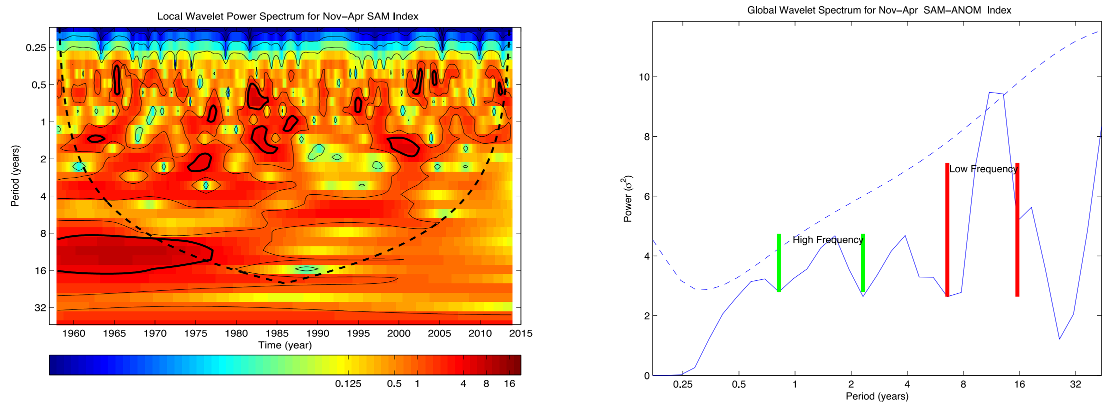

Fig. 3. Power spectra of the time series from Fig. 2.. The left panel shows the local power spectrum and the right panel shows the global power spectrum. The dashed line represents the 95% confidence levels. The green and red vertical lines show the high and low frequency bands.

3.2. Modeling the Classification of Extreme Heat

The blocking at 140°E, SAM and high frequency filtered portion of the SAM served as predictors for logistic regression, with and without a threshold adjustment [8], to model the monthly classifications. A decision threshold approach can be applied to assign class membership (e.g., extreme heat, no extreme heat) for logistic regression. Through optimization of the F-statistic, a classifier is found that selects a midpoint threshold on the probability output of the logistic regression. Performance of the classifier is measured using cross-validation. In this experiment, given the relative scarcity of months with extreme heat, the number of folds was reduced to three for the threshold adjustment. However, five folds were used to evaluate the model forecasts.

3.3. Evaluation of the Classification Forecasts of Extreme Heat

The forecasts for the unadjusted logistic regression were not particularly accurate in predicting those months when excessively hot January occurred (only 33% of those extreme events in testing data were classified correctly). A reason for this can be seen in the low climatological incidence of extreme heat (the base rate of 0.16 in Table 1) resulting in an imbalanced number of yes/no cases in the training data. However, the model was accurate in predicting months without extreme heat and, as most of the months in the data set were not associated with extreme heat, approximately 82% of the months in the testing data were classified correctly by the logistic regression without the threshold adjustment. After application of the threshold adjustment, the 67% of the testing data months with extreme heat were identified correctly.

To more fully evaluate the goodness of each model, several forecast evaluation indices were assessed using a 2x2 contingency table of forecasts of extreme heat (Y/N) versus observations (Y/N) [9]. A forecast of extreme heat in January that coincides with an observed extreme heat for that January is defined as a true positive. A forecast of extreme heat that coincides with a January not reaching the threshold that defines extreme heat is a false positive (false alarm). Similarly, a forecast of no extreme heat that coincides with a January not reaching the threshold that defines extreme heat that is a true negative. A January forecast of no extreme heat that coincides with an observed month of extreme heat is as a false negative (miss). Each type of missed forecast (a false alarm or false positive and a miss or false negative) has an associated cost. As there has been an upward trend in the incidence of extreme heat in the past few decades, there is a considerable and growing cost [2] in allocating resources to help mitigate its impacts. Accordingly, the decision cost of a miss (false negative) is substantial and the decision threshold should reflect this specific cost. The idea of the threshold adjustment is to move the decision boundary consistent with these costs. By examining the various forecast evaluation metrics for logistic regression and threshold adjusted logistic regression, the benefits and disadvantages of each model can be assessed (Table 1).

The probability of detection measures the fraction of observed extreme heat months that were forecast as extreme heat months. After application of the threshold adjustment, the probability of detection of extreme heat is improved 100% over the logistic regression without the threshold, indicating the substantial advantage of that adjustment. However, improving the probability of detection carried a penalty in creating more false alarms and there was a slight increase (8%) in the false alarm ratio (prediction of heat that does not materialize) (Table 1). The threat score ignores correct negatives or the cases where extreme heat is not forecast and it does not materialize. Such an omission of correct negative cases is considered advantageous when the climatological probability is low, as in these data. There is nearly a 40% improvement in the threat score (Table 1). Neither model had an unbiased (defined as bias =1) set of forecasts, with the unadjusted logistic regression having a bias value < 1, indicating it underforecasts extreme heat, and the adjusted logistic regression, with bias > 1, overforecasting extreme heat.

Skill statistics measure the relative improvement of the two models compared to randomly guessing if a month will exhibit extreme heat. Four skill scores were appraised. The Heidke score, which is considered a standard measure of skill by meteorologists [10], improved over 25% after the threshold adjustment was applied (Table 1). The Hanssen and Kuippers skill statistic was introduced to reduce perceived bias in the Heidke skill score [10] and it shows an increase of 80% in the adjusted threshold model. A skill statistic recommended for experiments with low climatological incidences is the equitable threat score [10] that discounts the forecasts of no extreme heat that result in no extreme heat. That skill score shows a positive change of 30% after the threshold adjustment is applied. A newer statistic, created for detection of extreme events, the extremal dependency index [11] was applied and exhibited over a 50% improvement for the threshold adjusted logistic regression. When viewed as a suite of forecast evaluations, the analyses summarized in Table 1 support the finding that threshold adjustment dramatically improves the classification accuracy for extreme heat.

Statistic	LogR	LogR threshold	Change
Base rate	0.16	0.16	0%
Probability of detection	0.33	0.67	103.0%
False Alarm Ratio	0.57	0.62	8.8%
Threat Score	0.23	0.32	39.1%
Bias	0.78	1.78	128.2%
Heidke Skill	0.27	0.34	25.9%
Hanssen-Kuiper Skill Score	0.25	0.45	80.0%
Equitable Threat Skill Score	0.16	0.21	31.3%
Extremal Dependency Index	0.38	0.58	52.6%

Table 1. Base rate and forecast evaluation statistics for logistic regression (LogR) and LogR with a threshold adjustment.

3.4. Predictions and Evaluation of Extreme Heat Anomalies

Prediction of the monthly anomalies was made using two methods: (1) SVR with over 1500 kernel evaluations and (2) bagging trees [12]. Both methods were subject to 5-fold cross validation and the testing data provided the generalization errors. The SVR found to provide the most accurate results used a radial basis function kernel. The mean absolute error was 1.4°C and the root mean squared error was 1.9°C. The correlation between the predicted and observed temperature anomalies was 0.37. The predictions arising from the bagged trees had a mean absolute error of 1.3°C and a root mean squared error of 1.8°C. The correlation between the predicted and observed testing data was 0.48, suggesting improved relative accuracy with the bagged trees.

4. Summary and conclusions

Over the past several decades, Australia has seen unprecedented high temperatures that are documented by analysis of the monthly mean maximum temperature. The extreme nature of the change in monthly mean maximum temperatures has had a negative impact on Australian society and ecosystems. The relationship between the heat and climate drivers was examined herein at a representative site, Tibooburra, for all January months during 1958-2013. Empirical percentiles of the monthly mean maximum temperatures were used to define the 90^{th} percentile of the distribution for years prior to the onset of the heat, resulting in an anomaly threshold of $+1.32^{\circ}$C. All January months with maximum temperature anomalies exceeding this value were assigned to a binary class of extreme heat. One challenge in analyzing these data is the relatively short 56 year period of record, relative to the numerous climate drivers (84) that may be related to the excessive heat. Feature selection using SVM RFE was applied to determine a compact set of predictors that demonstrated good generalization properties.

Classification algorithms are often designed to maximize the number of correct predictions. However, such an approach makes two assumptions: (1) the design is balanced with an equal number of "yes" and "no" events and (2) Incorrect forecasts of false negatives and false positives have equal costs. Both assumptions are violated in this study, so a threshold adjustment was applied to mitigate many of the negative effects of the imbalanced design. Results of cross-validation of logistic regression show that blocking of cold air, the SAM and the high frequency filtered portion of the SAM, provided correct extreme heat forecasts for over 75% of the total cases, after the threshold adjustment was applied. Comparison of the logistic regression with the threshold adjustment to that without the adjustment, through an extensive set of forecast evaluation measures, show that the adjustment improved the model considerably in nearly all aspects measured, with a slight negative tradeoff in false alarm ratio. For prediction of the actual monthly anomalies, SVR and bagged trees are explained the anomaly temperature values with a mean absolute error of 1.4°C and 1.3°C, respectively. The correlation between the predicted values of heat and those observed in testing data was larger for bagged trees (0.48) than for SVR (0.37). Given the upward trend in extreme heat, the threshold adjusted logistic regression model, SVR and bagged trees show promise in accurate prediction, allowing for resources to be allocated, helping minimize the impact of these events.

5. Acknowledgement

We thank Dr. Zewdu T. Segele for his expert assistance generating the wavelet analyses used in this study.

6. References

1. M.B. Richman and L.M. Leslie, Classification of changes in extreme heat over southeastern Australia, Complex Adaptive Systems, Publication 3, Procedia Computer Science 20 (2013) 148.
2. J. Medew, Anger over spike in deaths during record Victorian heatwave. The Age Victoria (2014).
3. N. Cristianini and J. Shawe-Taylor, An introduction to support vector machines, Cambridge University Press (1999).
4. R.K. Ahuja, J.B. Orlin and A. Tiwari, A greedy genetic algorithm for the quadratic assignment problem, Computers and Operations Research 27 (2000) 917.
5. I. Guyon, J. Watson, S. Barnhill and V. Vapnik, Gene selection for cancer classification using support vector machines, Machine Learning 46 (2002) 389.
6. M. Pook and T. Gibson, Atmospheric blocking and storm tracks during SOP-1 of the FROST project. Aust. Meteor. Mag. 1 (Special Edition) (1995) 51.
7. P.K. Sen, Estimates of the regression coefficient based on Kendall's tau, J. American Statist. Assoc. 63 (1968) 1379.
8. T.B. Trafalis, I. Adrianto, M.B. Richman and S. Lakshmivarahan, Machine-learning classifiers for imbalanced tornado data. Computational Management Science 10 (2014) 1.
9. A.H. Murphy, Forecast verification: Its Complexity and Dimensionality, Monthly Weather Review 119 (1991) 1590.
10. D.S. Wilks, Statistical Methods in The Atmospheric Sciences, Third Ed., Academic Press (2011).
11. C.A.T. Ferro and D. B. Stephenson, Extremal dependence indices: improved verification measures for deterministic forecasts of rare binary events, Weather and Forecasting 26 (2011) 699.
12. L. Breiman, Bagging predictors, Machine Learning 24 (1996) 123.