Warming Southeast Australian Climate: The effects of Sea Surface Temperatures (SSTs)

Kwanshae Flenory[1], Michael Richman[2], Lance M. Leslie[2]

[1]National Weather Center Research Experience for Undergraduates, Norman, Oklahoma

Langston University, Langston Oklahoma

[2]School of Meteorology, University of Oklahoma, Norman, Oklahoma

Abstract

In the past few decades, the climate in Australian has been warming at an alarming rate when compared to historical variations. Associated with that warming, extended heat events, lasting for weeks to months have plagued the country. Climate model projections suggest that such events will occur more frequently and intensify in the future. The extreme temperatures have damages ecosystems through droughts and fire and resulted in the loss of human life.

This study examines how the combination of sea surface temperatures (SSTs) and climate drivers predict summer mean maximum temperature at selected locations in SE Australia. Ninety-one ocean grid boxes of SST surrounding Australia were used for simultaneous and lag_1 relations as well as 42 climate drivers, creating a suite of 224 potential predictors. Variable reduction using 5-fold cross validated linear regression and bagging, resulted in ~ 90% reduction in the number of variables passed to the final prediction equations. Linear multiple and nonlinear kernel regression methods were applied to predict the January anomalies of maximum temperature using this reduced set of predictors. For the nonlinear regressions, two kernels were evaluated: polynomial and radial basis function. The polynomial degree and radial basis function kernel width were optimized for sea surface temperatures and climate drivers by maximizing their 10-fold cross validated correlations with the air temperatures at the various locations in SE Australia. The key findings were (1) climate drivers had as much significant influence on the prediction accuracy as SSTs and (2) the combination of the reduced sets of SSTs and climate drivers often accounted for 40-60% of the January mean maximum temperature variance. Such a large percentage of predictable variance is expected to lead to more effective monthly temperature predictions.

1. Introduction

Australia has warmed by 0.9C since 1910 (Fig. 1) and is expected to continue warming at a rate that is proportional to the greenhouse gas concentration (Hopkin, 2014). The year of 2013 was the hottest year since 2005. The summer (November through March,) temperatures are breaking and setting new records yearly and these abnormally hot states last longer than in the past (IPCC, 2007). Extreme hot to cold daytime ratio has risen to 3:1, warming 0.8°C since 1910 (Hopkin, 2014) (Fig. 2). The hot to cold nighttime ratio is now 5:1, warming 1.1°C. These extreme temperatures are thought to take more lives than any other natural hazard over the past 150 years, and are referred to as the "silent killer." Extreme hot weather increases the number of deaths, health problems, affects infrastructure as well as socioeconomic impacts (Richman and Leslie, 2013).

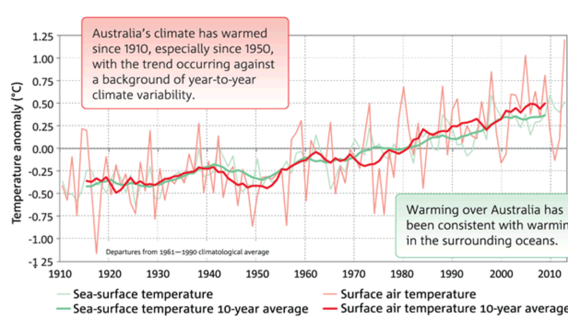


Figure 1. The frequency of warm to cool mean SST anomalies from 1961 to 1990 is show in this stacked scatterplot.

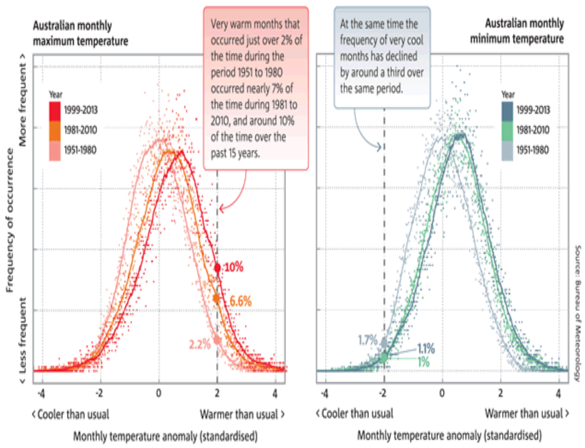


Figure 2. Mean monthly maximum and minimum temperatures.

The present research seeks to establish relationships between the mean maximum temperature in a mid-summer month, January and sea surface temperatures surrounding Australia as well as a suite of climate drivers around the globe. The goal is to create a technique that predicts the mean maximum temperature anomalies accurately, allowing notification of citizens and government entities that make temperature dependent decisions. Section 2 presents the datasets used to identify predictors and also the station temperature data response variable. The results of the analyses are communicated in Section 3 and conclusions drawn in section 4.

This study applies methods of variable reduction using linear regression and bagging trees. Bagging trees is another word for bootstrap aggregation that pulls together decision trees as base classifiers. This improves the results of the classification algorithms.

The models are given monthly anomalies of sea surface temperatures and air temperatures at a location to calculate the correlation between the two and reduce the variability by extracting the variables that had no significant attribution to the set. Irregular warming of the earth's exterior affects atmospheric movement, that produces evaporation from oceans and unequal warming of the land and oceanic surface, that causes many different climate drivers. These climate drivers are also interrelated with sea surface temperature anomalies according to each location because they affect the weather over a different area throughout different periods.

1. Data and Method

1.1 *Data*

Monthly sea surface temperature anomalies (SSTAs) from Kaplan Extended SST V2 from ESRL (Earth Systems Research Lab) were extracted from 1958 to 2013 for the present research. The data were extracted for latitudes between 2.5S and 42.5S and longitudes 107.5E to 162.5E to insure all areas around Australia were considered.

The data were in netCDF format and converted into ASCII (American Standard Code for Information Interchange), using the ncdump command.

The output contained a continuous set of monthly data from 1958 to 2013. January SSTA data were extracted in individual years and concatenated into a continuous set of 55 January months for each of the 91 SST grids.

Along with the aforementioned SST anomalies, an air temperature record at a location in SE Australia was required as a response variable. Melbourne was chosen due to it being the biggest city in Victoria and one of the most populated cities in Australia.

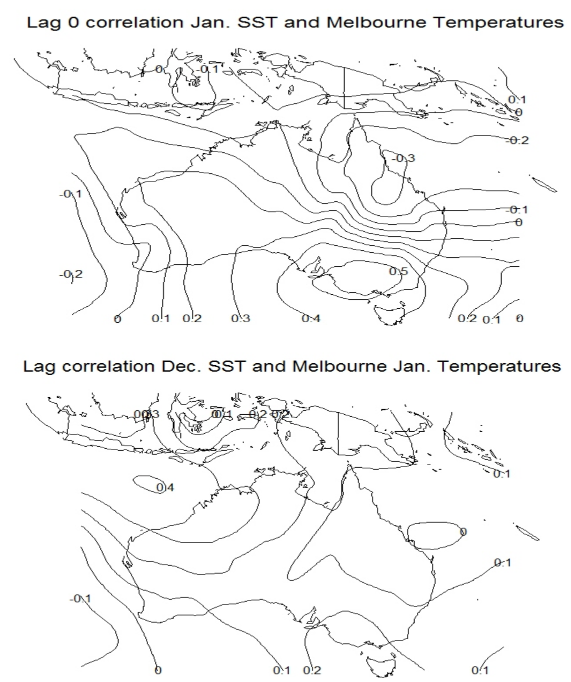


Figure 3. Correlations of sea surface temperatures according to January temperatures across Australia for (a) lag_0 and (b) lag_1.

SSTA data for the January (hereafter referred to as "lag_0" and the December months for the previous year (hereafter referred to as "lag_1") were drawn and added to the pool of potential predictors. Each SSTA time series for lag_0 and lag_1 was correlated with the Melbourne temperature time series, mapped to the domain and contoured (Figs. 3a and 3b).

The analysis procedures were part of the WEKA (Waikato Environment for Knowledge Analysis) package. WEKA programs were used for classification, variable selection, regression, and dimension reduction.

A side from the SSTA data, climate driver anomalies were also used as predictors.. The climate driver anomaly variables incorporated:

1. Southern Annular Mode(SAM)
2. Southern Oscillation Index (SOI)
3. Dipole Mode Index (DMI)
4. Pacific Decadal Oscillation (PDO)
5. Artic Oscillation (AO)
6. North Atlantic Oscillation (NAO)
7. Pacific-North American pattern (PNA)
8. Blocking 140E (B140E)
9. Blocking 160E (B160E)
10. Modoki
11. Niño 12
12. Niño 3
13. Niño 4
14. Niño 3.4
15. High Frequency Southern Annular Mode (HFSAM)
16. High Frequency Southern Oscillation Index (HFSOI)
17. High Frequency Dipole Mode Index (HFDMI)
18. High Frequency Pacific Decadal Oscillation (HFPDO)
19. High Frequency Arctic Oscillation (HFAO)
20. High Frequency North Atlantic Oscillation (HFNAO)
21. High Frequany Pacific-North American pattern (HFPNA)
22. High Frequency at Blocking 140E (HFB140E)
23. High frequency at Blocking 160E (HFB160E)
24. High Frequency Modoki (HFModoki)
25. High Frequency Niño 12A (HFNiño12)
26. High Frequency Niño 3A (HFNiño3)
27. High Frequency Niño 4A (HFNiño4)
28. High Frequency Niño 3.4A (HFNiño3.4)
29. Low Frequency Southern Annular Mode (LFSAM)
30. LowFrequency Southern Oscillation Index (LFSOI)
31. Low Frequency Dipole Mode Index (LFDMI)
32. Low FrequencyPacific Decadal Oscillation (LFPDO)
33. Low FrequencyArtic Oscillation (LFAO)
34. Low Frequency North Atlantic Oscillation (LFNAO)
35. Low Frequany Pacific-North American pattern (LFPNA)
36. Low Frequency at Blocking 140E (LFB140E)
37. Low frequency at Blovking 160E (LFB160E)
38. Low Frequency Modoki (LFModoki)
39. Low Frequency Niño 12A (LFNiño12)
40. Low Frequency Niño 3A (LFNiño3)
41. Low Frequency Niño 4 A(LFNiño4)
42. Low Frequency Niño 3.4A (LFNiño3.4)

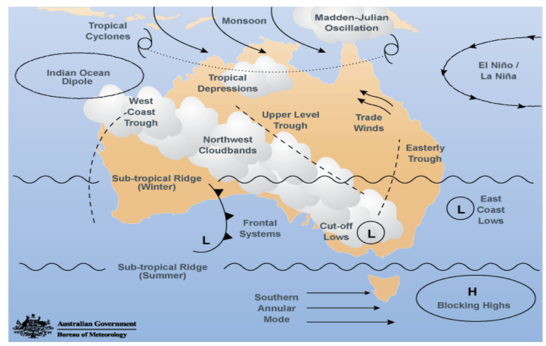


Figure 4. An illustration of some climate drivers on Australia.

In addition to Melbourne (ME), the procedure was applied to three additional air temperature sites at Tibooburra (TI), Woomera (WO), and Sydney (SY). A map including these locations are displayed below (Fig. 5)

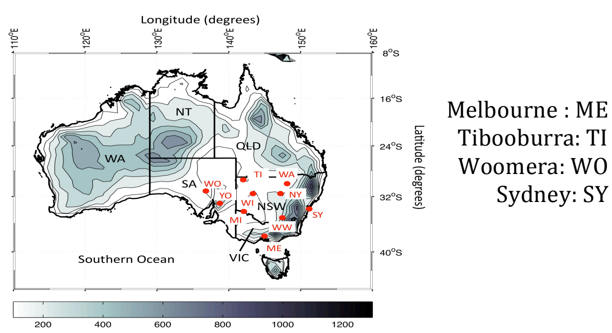


Figure 5. A map of Australia is shown to clearly show the exact locations of all four sites.

A final compilation file of SST anomalies for lag_0 and lag_1, 42 climate divers, and 4 site air temperature anomalies for the 55 years was used for the subsequent analyses in WEKA.

Since there were only 55 years of data, 224 predictors were intractable for stable prediction equations. Through attribute selection, according to correlation-based feature selection subset evaluation (CfsCubsetEval) with a greedy step-wise method, the best attributes were identified using five-fold cross-validation. The process was applied to the lag_0 SST, the lag_1 SST and the climate drivers (to predict air temperature) in three separate analysis steps. The cross-validation was important to insure that he results were not spurious (relationships that fit irreproducible noise). For the greedy stepwise multiple linear regression, the number of times any predictor appeared in the cross-validated analysis, to predict the Melbourne temperature anomalies, was noted. In instances where that value was greater than zero, the predictor was selected for further analysis. Additionally, bagging tree selection was used to identify additional variables. This process yielded a compact set of robust attributes that represented significant predictors that generalized well.. Those variables identified by the greedy stepwise and bagged tree analyses from the lag_0, lag_1 and climate drivers sets were all combined into a new compact set of predictors and subjected to a final set of predictions using linear regression and SMOregression with a exponent of 1. The advantage of this approach was that the predictor weights could be noted to determine if further variable reduction was possible for a final set of predictors. Additionally, one final bagged tree analysis was used to achieve the compact set of predictors. In the case of Melbourne, linear regression with 5-fold cross validation identified four predictors and bagging trees selected seven. For Melbourne, three of the predictors were the same for the regression and bagging methods, leaving a set of eight unique predictors.

Support Vector Regression (SVR), with sequential minimum optimization (SMOregression) , uses kernels, to replace the inner product in the regression formulation. The selection of a kernel is a critical step to obtain the solution that generalizes best to independent data. In the present analyses, a set of experiments was performed to determine the radial basis function (RBF) kernel width and the exponent weights on the polynomial (Poly) kernel. With the polynomial kernel, the exponent was changed starting with 0.1 to 3.0, in steps of every tenth of a decimal for a total of 30 kernel evaluations. For each evaluation a 10-fold cross-validation was used and the correlation change between the predictions and the observed air temperatures were noted. If the correlation continued to increase, further evaluations were made. The exponent associated with the maximum correlation was selected as the optimized model that generalized best. This process of kernel evaluation was also employed for the RBFkernel, for RBF widths ranging from .01 to .30 increasing every hundredth of decimal. The RBF weight associated with the largest correlation was noted and that was the best generalizing RBF model.

Of the two kernels, the one model with the largest correlation was kept, and the mean absolute error (MAE), and root mean absolute error (RMSE) were recorded as well. This was done for each lag_0 and lag_1 for each Melbourne and the process was repeated for each of the three remaining locations in SE Australia. The only difference for the other cities was that selection of attributes with a greedy stepwise method with a five-folds was made and the variables noted. Similarly, bagging trees were employed temperature. The one set of predictors (either greedy stepwise or bagging trees) that yielded the model that generalized best was identified. In general, it was noted that small sets of predictors yielded models with larger cross-validated correlations and smaller MAE or RMSE, suggesting that retaining too many predictors hurt the generalization properties of the predictions.

2. Results

Using the aforementioned variable selection programs in WEKA resulted in a dramatic amount of data reduction. The variables used for Melbourne alone, in lag_0 was reduced to 8.8% of the original number (8 retained out of a possible 91) and in lag_1, the 91 was reduced to 5 (5.5%) used. There were 224 potential predictor variables, for the final predictions of Melbourne air temperature, and these variable reduction techniques winnowed this to 20, for a decreased of 91.1% (8.9% selected).

After determining which variables are best correlated for each lag of SSTs and climate drivers, all of the selected attributes were used to predict the air temperature. From lag_0, eight grid points of SSTA were kept, five from lag_1, and seven from the climate drivers. All of these variables were preprocessed together with Melbourne, but this time at 10 fold cross-validation. Then the polykernel and RBFkernel ranges were changed again, keeping the best technique with the highest 10-fold cross-validated correlation. Of the two, the best method was kept and the predictions and forecasts were mapped on a scatterplot. (Fig. 6) This step was repeated for each location.

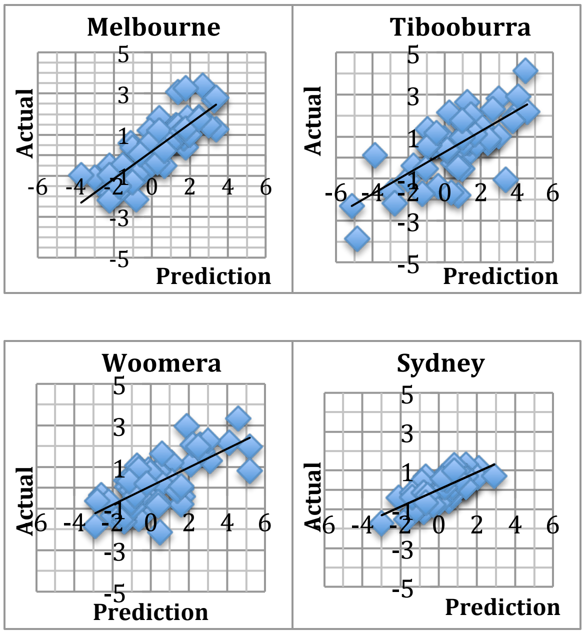
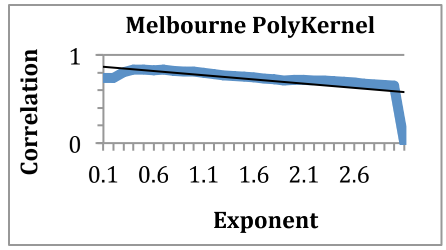
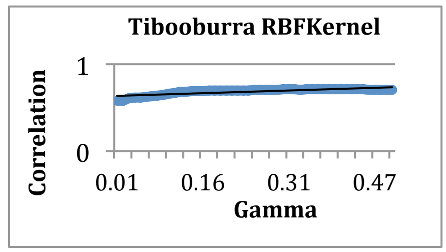


Figure 6. These scatterplots represent a 10-fold cross-validation of SSTs, Climate Drivers, and (a) Melbourne, (b) Tibboburra, (c) Woomera, (d) Sydney.

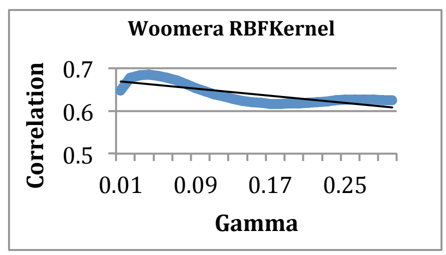
The air temperature at all of the four sites were predicted with stepwise and kernel regressions.us The evaluation of the forecasts are expressed in terms of the correlation between the predicted and observed temperatures for independent testing data in the 10-fold cross validations as well as the mean absolute error (MAE), and root mean squared error (RMSE) associated with the anomaly mean. Each location was analyzed separately and displayed as a scatterplot. Additionally, line graphs were created to show the sensitivity of the correlation accuracy to the kernel parameters (either exponent or radial basis width). The scatterplot was used to show the relation between the actual and predicted forecasts. On the forecast evaluation scatterplots (Fig. 6) a linear regression trend line or line of best fit was added to the plot and positioned on the graph to represent the data. The results from Melbourne (Fig. 6) show a 40° angle that was an over prediction of the model in comparison to the actual air temperature. The plot for Tibooburra has a regression line with a 30° angle, Woomera had a 35° angle and Sydney about 30°, indicative of overforecasting the air temperature at each location. Figure 7 shows the relation of changing the width of the polykernel to the correlation.



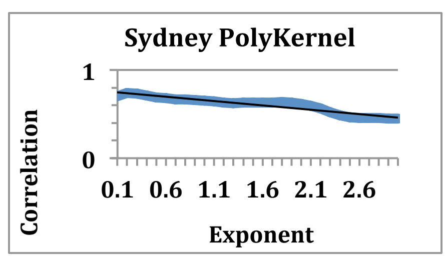
Corr.:
0.8377
MAE:
0.765 °C
RMSE:
0.976 °C



Corr.:
0.7116
MAE:
1.1841 °C
RMSE:
1.4855 °C



Corr.:
0.6839
MAE:
1.1047 °C
RMSE:
1.3971 °C



Corr.:
0.744
MAE:
0.585 °C
RMSE:
0.7649 °C

Figure 7. Line graphs show the sensitivity of correlation accuracy. According to the highest kernel method. A trend line was placed to best fit the relationship.

a) Correlation Coefficient				
	SST Lag_0	SST Lag_1	Climate Drivers	All 3 sets
Melbourne	0.650	0.600	0.669	0.838
Tibooburra	0.618	0.445	0.302	0.712
Woomera	0.626	0.568	0.593	0.684
Sydney	0.652	0.511	0.309	0.670

b) Mean Absolute Error (°C)				
	SST Lag_0	SST Lag_1	Climate Drivers	All 3 sets
Melbourne	1.116	1.166	1.108	0.765
Tibooburra	1.537	1.492	1.679	1.184
Woomera	1.177	1.672	1.254	1.105
Sydney	0.717	0.765	0.885	0.649

Table 1. Data chart that shows (a) the correlation, and (b) the MAE, of the 10-fold cross-validation for the temperature predictions at each location.

Table 1a shows that the sea surface temperatures were as significant as the climate drivers in obtaining accurate predictions. The final variance explained for each location was 70.2% at Melbourne, 50.7% at Tibooburra, 46.8% at Woomera and 42.5% at Sydney. The variability was reduced in absolute error (Fig. 1b).

In this study, it was determined that it is not optimal to put all variables into the prediction scheme; careful variable selection was vital to obtaining accurate cross-validated predictions. In terms of variable selection, bagging tree regression achieved the most accurate predictions with fewer than 10% of the available predictors. Keeping too many or all predictors always reduced the accuracy of the prediction dramatically when cross-validated on the independent data.

3. Conclusion

For the past two decades, monthly mean maximum temperatures in Australia have been increasing dramatically, putting stress on people, ecosystems and infrastructure. It is clear that the climate change has an overwhelming impact on the environment. Extreme heat over an extended period of time increases fire danger, drought,, and reduces physical structures stability as well as stressing or killing the humans, plants, and animals that inhabit Australia.

This study uses a variable reduction approach applied to SST and climate driver data to predict the January monthly maximum temperature anomalies at four diverse locations in SE Australia. Through state of the science variable selection kernel prediction techniques, the prediction variance explained for each location was 70.2% at Melbourne, 50.7% at Tibooburra, 46.8% at Woomera and 42.5% at Sydney. Such large amounts of the variance explained should make climate prediction of air temperature in SE Australia feasible.

Unfortunately, the increase in hot months has commenced and is projected to continue to become more common as greenhouse gasses increase. This makes accurate prediction of the most extreme heat critical, early detection of prolonged heat events can and will allow society to properly prepare as much possible. The methods presented herein help to define the predictability of those events.

4. Acknowledgements

The authors would like to thanks Daphne LaDue and Javier Lujan for allowing the 2014 summer Research Experience for Undergraduates (REU).

The authors would also like to thank, IT Specialist, Shawn Riley of the National Weather Center, for helping to determine the best ways to extract data as well as running commands.

This material is based upon work supported by the National Science Foundation under Grant No. 1062932.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

5. References:

Hopkin, M., 2014: Australia has warmed by 0.9C since 1910, with more in store." *Phys.org*. The Conversation, 04 Mar. 2014. Web. 20 July 2014. <http://phys.org/news/2014-03-australia-09c.html>.

IPCC, 2007: Summary for Policymakers. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Richman, M.B. and L.M. Leslie, 2013: Classification of changes in extreme heat over Southeastern Australia. *Procedia Computer Science*, **20**, 148-155.