

Article

Application of Machine Learning to Attribution and Prediction of Seasonal Precipitation and Temperature Trends in Canberra, Australia

Joshua Hartigan ^{*}, Shev MacNamara and Lance M. Leslie 

School of Mathematical and Physical Sciences, University of Technology Sydney, Ultimo, New South Wales 2007, Australia; Shev.MacNamara@uts.edu.au (S.M.); Lance.Leslie@uts.edu.au (L.M.L.)

* Correspondence: Joshua.Hartigan@student.uts.edu.au

Received: 21 May 2020; Accepted: 9 June 2020; Published: 11 June 2020



Abstract: Southeast Australia is frequently impacted by drought, requiring monitoring of how the various factors influencing drought change over time. Precipitation and temperature trends were analysed for Canberra, Australia, revealing decreasing autumn precipitation. However, annual precipitation remains stable as summer precipitation increased and the other seasons show no trend. Further, mean temperature increases in all seasons. These results suggest that Canberra is increasingly vulnerable to drought. Wavelet analysis suggests that the El-Niño Southern Oscillation (ENSO) influences precipitation and temperature in Canberra, although its impact on precipitation has decreased since the 2000s. Linear regression (LR) and support vector regression (SVR) were applied to attribute climate drivers of annual precipitation and mean maximum temperature (TMax). Important attributes of precipitation include ENSO, the southern annular mode (SAM), Indian Ocean Dipole (DMI) and Tasman Sea SST anomalies. Drivers of TMax included DMI and global warming attributes. The SVR models achieved high correlations of 0.737 and 0.531 on prediction of precipitation and TMax, respectively, outperforming the LR models which obtained correlations of 0.516 and 0.415 for prediction of precipitation and TMax on the testing data. This highlights the importance of continued research utilising machine learning methods for prediction of atmospheric variables and weather patterns on multiple time scales.

Keywords: machine learning; precipitation; temperature; southeast Australia; attribution; prediction

1. Introduction

Numerous long-term droughts have occurred in Australia throughout recorded history. They include the Federation drought (1895–1902, [1]), the World War II drought (1937–1945, [2]) and the Millennium Drought (1997–2009, [3]). The current drought affecting southeast Australia began soon after the Millennium drought, with massive impacts on agriculture and water availability, including Australia's capital city, Canberra. It resulted in extreme fire weather, culminating in devastating wildfires affecting a large portion of the region in the 2019–2020 fire season. Globally, locations at regions of similar latitudes to southern Australia, such as Cape Town and California, also have recently been affected by severe droughts (e.g., [4,5]). Drought is becoming increasingly frequent and more severe in these regions [6,7] as increasing mean temperatures increase evaporation and enhance bushfire threat. Furthermore, land surface changes and an increasing population can place greater strain on water resources, exacerbating drought impacts [8,9].

In addition to enhanced fire risk, drought can have significant socio-economic effects [10] through reduced irrigation for agriculture, serious impacts on human health including decreased sanitation,

increased energy use, and greater risk of flooding as rain falls on dry soil that inhibits runoff. Canberra is located in southeast Australia and relies on nearby, catchment area rainfall for its water supply. It is situated between Sydney and Melbourne, Australia's two most populous cities, and is representative of many aspects of southeast Australian climate warming. Following the Millennium Drought, numerous programs aimed at decreasing per capita water consumption were introduced, resulting in significant reductions [11]. However, it is only a temporary solution to the water availability problem as Canberra's population continues to increase rapidly [12]. Furthermore, annual precipitation over southeast Australia has decreased, particularly during autumn, due to the increasing strength and poleward progression of the subtropical ridge [13–15]. Projects aimed at increasing water availability include the enlargement of the Cotter dam, and development of a pipeline from the Murrumbidgee River to Googong dam [11]. The pipeline supply still ultimately is controlled by rainfall, and the current drought led only to a short period of operation of the pipeline, due to low water levels [16]. As the Murrumbidgee River is a major contributor to the Murray-Darling basin (MDB), which produces 39% of the national food supply [17], it has limited utility during severe drought conditions.

Management of future water supply in Canberra and the wider southeast Australia region requires both a detailed assessment of any trends in important variables, and the development of accurate models to predict these variables [18,19]. This study focusses on trends in precipitation and temperature, as these are key variables in meteorological drought [4]. These variables can be predicted using global circulation models, statistical models, or a combination of each model [20]. In this study, statistical models are utilised as they require less computational resources, can be used to attribute the main climate drivers associated with these variables, and have performed well in previous studies on prediction of climatic variables (e.g., [5,18,19,21]).

2. Materials and Methods

2.1. Data

Monthly mean maximum temperature (TMax), mean minimum temperature (TMin) and total monthly precipitation time series were obtained for Canberra (35.3° S, 149.1° E) from the publically available Climate Data Online provided by the Australian Bureau of Meteorology (BoM) website (<http://www.bom.gov.au/climate/data/index.shtml>). Stations with relatively long records in the area were chosen to represent Canberra. Any missing data points, which accounted for approximately 1% of the data, were filled using a moving average centred on the year for which the data was missing. The precipitation and temperature time series start at 1938 as this is the earliest period in which both data sets were available for Canberra.

2.2. Statistical Analysis

To examine trends in Canberra precipitation and temperature, the data are analysed both annually, and over each of the four seasons, to determine any variations in precipitation or temperature over shorter time-scales. An overview of trends present in the time series was first gained by plotting the time series and their associated percentiles. The data then were grouped into 20-year periods and bootstrap resampling was applied with 5000 resamples to gain a deeper understanding of any trends present in the data. Permutation testing then was applied with replacement to test for statistical significance between the means of the various two 20-year periods. Statistical tests of series breaks could have been performed in this study to determine the periods used (e.g., [22–24]), however, the aim of this study was not to identify precise breakpoints in the time series. As the climate continues to change, and natural variability can enhance or reduce changes in the climate system over different periods, the detection of breakpoints in the time series can be challenging especially when there is no abrupt change. The decision to group data into 20-year periods was made in order to have equal length periods of sufficient size to compare and identify any changes in the mean between these periods.

2.3. Attribute Selection

First, a wavelet analysis was performed on the entire detrended time series anomalies, following the approach of Torrence and Compo [25]. This provides an understanding of the time evolution of the periodicity of signals within a time series, which possibly is useful for detecting potential climate drivers, such as the ENSO phases, and analysing how the influence of these climate drivers might change over time. Wavelets are used as they efficiently resolve both high- and low-frequency signals [26]. In this study, the Morlet wavelet is used.

Potential climate drivers under consideration in this study are numerous and include the Atlantic Meridional Oscillation (AMO), the Dipole Mode Index (DMI), the global sea surface temperature anomalies (GlobalSSTA), the global temperature anomalies (GlobalT), Niño3.4, the Pacific Decadal Oscillation (PDO), the Southern Annular Mode (SAM), and the Southern Oscillation Index (SOI), obtained from the Earth System Research Laboratory (http://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/). Tasman Sea sea surface temperature anomalies (TSSST) were also used, which were obtained from the Australian Bureau of Meteorology (<http://www.bom.gov.au/climate/change/#tabs=Tracker&tracker=timeseries>). All data are from 1957–2017. It is worth noting that there are many other potential climate drivers which could be considered instead of those used here, such as the interdecadal pacific oscillation (IPO, [27]) being used over the PDO. However, the climate drivers included in this study cover all ocean basins, and many are correlated to those that were excluded. For example the correlation between the IPO and the PDO over the period 1957–2017 is 0.699.

Figure 1 shows scatter plots of annual precipitation in Canberra against each of these potential climate attributes, with a linear fit and the correlation coefficient between the climate attributes and annual precipitation. Figure 2 shows the scatter plots for annual TMax. Many of the potential climate drivers have low correlations with annual precipitation or TMax, and exhibit non-linear relationships. This study set out to attribute the climate drivers which have greatest influence on annual precipitation and TMax. As there are non-linear relationships between the variables, non-linear statistical models are used for attribution and prediction. This study performed preliminary work on selection of attributes for annual precipitation and TMax using support vector regression (SVR), similar to the approach of [28], where both polynomial (poly) and radial basis function (RBF) kernels were considered. Multiple linear regression (LR) was also performed to compare the performance of non-linear techniques with conventional linear techniques.

Two-way attribute interactions, obtained simply by multiplying one variable with another (e.g., Niño3.4*DMI), can be important in statistical models as one predictor might reinforce another. For example, a positive IOD can enhance the drying effect of an El-Niño event in Australia [29,30]. Past studies using regression and machine learning techniques have found the optimum model selected using different variable selection techniques can involve interaction terms (e.g., [5,21,31]). As such, this study has included all possible two-way interactions between the aforementioned climate drivers as potential attributes. Other possible relationships, such as an inverse relationship, could exist between two or more variables. However, in this study it was decided to only use the two-way multiplicative relationships as potential predictors due to the amount of testing required to choose between this or other relationships that might exist.

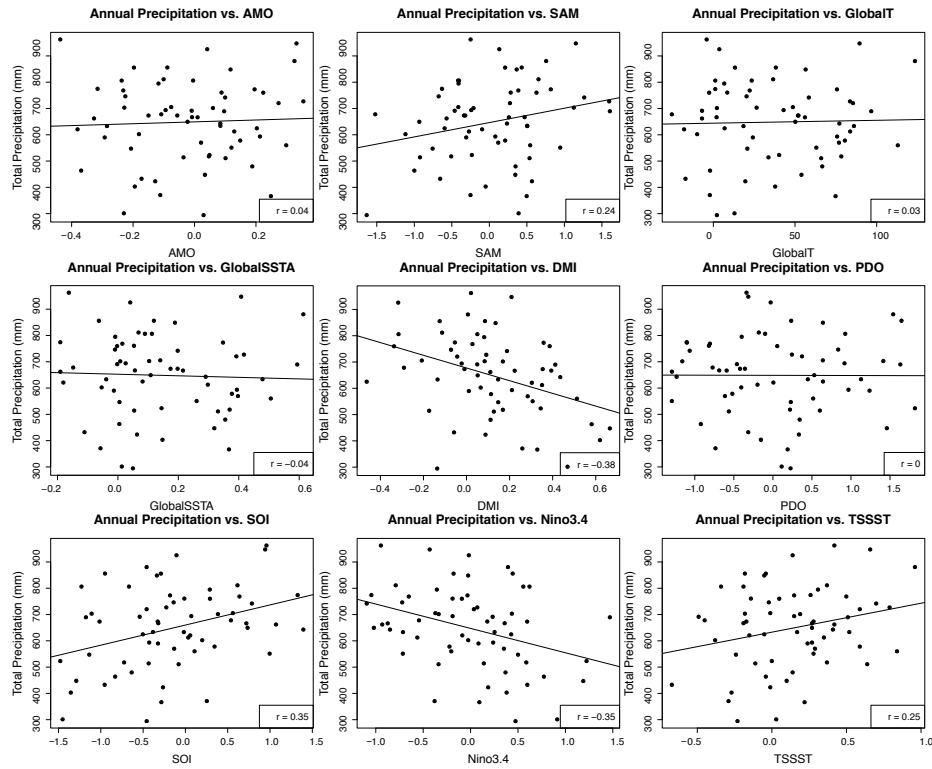


Figure 1. Scatter plot of annual precipitation in Canberra against each climate driver that serve as possible attributes, showing non-linear relationships between precipitation and the climate drivers. The thin black line is a linear fit against the data using least squares regression. Correlations of each predictor against precipitation are provided in the box to the lower right of each sub-panel, low correlations ($|r| < 0.3$) indicate little linear relationship between a predictor and precipitation.

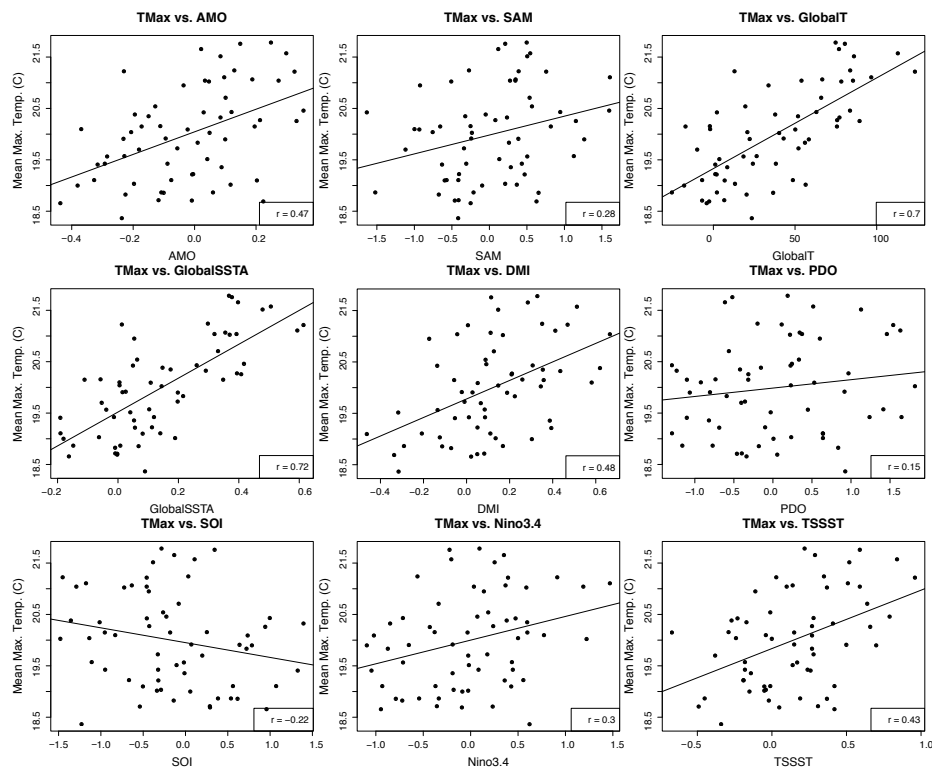


Figure 2. As in Figure 1 except for TMax. Predictors with $|r| > 0.4$ indicate moderate or strong relationships with TMax likely because they are attributes of global warming, or have been influenced by global warming.

2.4. Training and Prediction

Including interaction terms, there is a total of 45 potential attributes which could serve to predict annual precipitation and TMax. However, using all attributes would result in a model which is severely overfit. This can cause large errors when the model is applied to the test data set, and also increases the complexity of the model, making it very difficult to obtain a physical understanding underlying the predictions. As such, it is desirable to select only the attributes that generalize well [21,32]. In this study, 80% of the data (1957–2005) was selected to train each model. Ten-fold cross-validation [33] was applied to the training set using forward selection through the space of potential attributes, with a similar approach to that of [34].

For the RBF SVR, there are two parameters which need to be selected. The gamma value (G) is a constant in the RBF kernel function which relates to how much influence support vectors have on the decision boundary. The cost parameter (C) is the cost of violating constraints in the optimization problem. The exponent (E) of the poly kernel, and the cost parameter are free parameters which need to be selected in the poly SVR. The free parameters for both RBF and poly kernels are often selected by performing a grid search [35], and they influence how strongly the model fits to the training data, thereby affecting the bias and variance of the model on the testing data set. During the grid search for the final model fitting step in this study, values of C tested ranged from $2^{0.5}$ to 2^3 , with powers varying by 0.5; values of G ranged from 0 to 1, varying by 0.5; and values of E tested were 1, 2 and 3.

Table 1 lists the percentage of folds each attribute appeared in for the models trained on annual precipitation, while Table 2 lists the percentage of folds for annual TMax. The SVR models used in the development of these tables did not have the C, G and E parameters tuned, instead taking default values of 1, 1/(no. of predictors) and 3 respectively, in order to decrease computation time. Those attributes which appeared in at least 5 of the 10 folds were retained for further selection using the LR and SVR methods. The most parsimonious model, with the highest correlation and lowest root mean square error (RMSE) against the observed training data was selected and used to predict both annual precipitation and TMax on the testing data set (2006–2017).

Table 1. Percentage of folds selecting attributes for precipitation using LR, and SVR with the radial basis (RBF) and polynomial (poly) kernel functions.

Attribute	LR	SVR (RBF)	SVR (Poly)	Attribute	LR	SVR (RBF)	SVR (Poly)
AMO	10	80	40	DMI*TSSST	10	70	50
DMI	100	60	50	GlobalSSTA*GlobalT	0	80	40
GlobalSSTA	0	60	50	GlobalSSTA*Nino3.4	0	60	60
GlobalT	0	70	70	GlobalSSTA*PDO	20	50	50
Nino3.4	30	50	60	GlobalSSTA*SAM	10	70	50
PDO	10	60	60	GlobalSSTA*SOI	10	40	60
SAM	100	60	70	GlobalSSTA*TSSST	0	70	60
SOI	60	50	50	GlobalT*Nino3.4	30	50	50
TSSST	0	70	60	GlobalT*PDO	40	50	70
AMO*DMI	40	60	30	GlobalT*SAM	20	70	70
AMO*GlobalSSTA	0	90	70	GlobalT*SOI	10	50	50
AMO*GlobalT	0	90	60	GlobalT*TSSST	0	70	50
AMO*Nino3.4	20	60	40	Nino3.4*PDO	0	60	70
AMO*PDO	20	70	20	Nino3.4*SAM	20	80	60
AMO*SAM	70	70	30	Nino3.4*SOI	0	50	60
AMO*SOI	0	70	50	Nino3.4*TSSST	30	50	50
AMO*TSSST	0	70	50	PDO*SAM	10	80	40
DMI*GlobalSSTA	0	30	70	PDO*SOI	10	60	50
DMI*GlobalT	0	30	40	PDO*TSSST	10	70	60
DMI*Nino3.4	0	50	40	SAM*SOI	0	80	40
DMI*PDO	20	50	50	SAM*TSSST	60	80	60
DMI*SAM	40	70	30	SOI*TSSST	90	60	50
DMI*SOI	0	70	30				

Table 2. Same as Table 1 except for TMax.

Attribute	LR	SVR (RBF)	SVR (Poly)	Attribute	LR	SVR (RBF)	SVR (Poly)
AMO	10	50	80	DMI*TSSST	0	60	70
DMI	100	70	70	GlobalSSTA*GlobalT	10	60	60
GlobalSSTA	100	40	70	GlobalSSTA*Nino3.4	0	70	60
GlobalT	70	40	70	GlobalSSTA*PDO	0	80	60
Nino3.4	60	50	60	GlobalSSTA*SAM	0	60	80
PDO	10	60	70	GlobalSSTA*SOI	0	60	60
SAM	100	70	70	GlobalSSTA*TSSST	90	60	70
SOI	10	60	60	GlobalT*Nino3.4	0	40	60
TSSST	0	30	80	GlobalT*PDO	90	80	70
AMO*DMI	0	60	70	GlobalT*SAM	0	60	90
AMO*GlobalSSTA	30	70	70	GlobalT*SOI	0	60	60
AMO*GlobalT	10	80	60	GlobalT*TSSST	0	60	70
AMO*Nino3.4	0	60	60	Nino3.4*PDO	0	60	70
AMO*PDO	20	80	60	Nino3.4*SAM	0	80	90
AMO*SAM	90	50	90	Nino3.4*SOI	0	60	60
AMO*SOI	0	70	70	Nino3.4*TSSST	0	70	60
AMO*TSSST	0	50	50	PDO*SAM	0	100	50
DMI*GlobalSSTA	0	50	50	PDO*SOI	10	50	70
DMI*GlobalT	0	50	70	PDO*TSSST	0	70	70
DMI*Nino3.4	0	80	40	SAM*SOI	0	90	90
DMI*PDO	10	50	60	SAM*TSSST	0	60	90
DMI*SAM	10	70	50	SOI*TSSST	0	60	70
DMI*SOI	70	70	60				

3. Results

3.1. Evolution of Precipitation

Time series of precipitation and box plots of bootstrapped mean precipitation over 20-year periods, from 1939–1958 through to 1999–2018, for Canberra are provided in Figure 3. The data are annual and cover the four seasons, as any seasonal changes to precipitation can potentially be important in influencing drought. The *p*-values from permutation testing, comparing the means for the periods 1939–1958 and 1979–1998 against the mean for 1999–2018, are shown in Table 3. In Table 3 we report multiple *p*-values, so the issue of multiple testing may be considered, as discussed in chapter 15 of [33], for example. One way to control the false discovery rate is the well-known Bonferroni correction. In our example, with $N = 30$ tests, we could consider instead replacing the 0.05 threshold by the threshold $0.05/N$. If we did so, then some of the results in Table 3 would no longer be significant. However, most of those results remain significant even with a more stringent test, and we note that a Bonferroni correction can sometimes be pessimistic.

The Canberra mean annual precipitation has remained relatively stable throughout the observational record, with little change between 20-year periods, except for the slightly wetter period 1959–1978 (Figure 3), which does not have a statistically significant difference in the mean compared to other 20-year periods (not shown). However, there appears to have been a modest decline in variability since the 2000s, with a decrease in recorded precipitation above the 75th percentile and below the 25th percentile since the 2000s. Further, there has been a reduction in the interquartile range (IQR) and whisker lengths in the 1999–2018 box plot compared to previous 20-year periods. This decline in variability of rainfall between 1939–1958 and 1999–2018 is statistically significant at the 90th percentile (*p*-value = 0.093). In contrast, mean precipitation during the seasons is not always as stable. During the Canberra autumn (March–May), a decline in precipitation is apparent as the frequency of years recorded below (above) the 15th (75th) percentile has increased (decreased) since the 1990s. This decline in autumn precipitation also is clear in the box plots, with a statistically significant decrease in mean precipitation between 1939–1958 and 1999–2018 (*p*-value = 0.026). Meanwhile, mean winter (June–August) precipitation has undergone little change, with no statistically significant difference between 20-year periods. Overall, this suggests precipitation over the cooler months of the year, which are vital for catchment inflows, is decreasing.

Table 3. *p*-values for the permutation test on the difference between the mean wet season precipitation, mean maximum temperature or mean minimum temperature between 1939–1958 and 1999–2018; and 1979–1998 and 1999–2018. Tests conducted used 5000 resamples. Text in bold face highlight statistical significance at the 95% confidence level, text in italics highlights statistical significance at the 90% confidence level.

Years	Precipitation	TMax	TMin
	Annual		
1939–1958 vs. 1999–2018	0.7610	0.0000	0.0022
1979–1998 vs. 1999–2018	0.8620	0.0000	0.5210
	Autumn		
1939–1958 vs. 1999–2018	0.0262	0.0000	0.5730
1979–1998 vs. 1999–2018	<i>0.0840</i>	0.0094	0.2350
	Winter		
1939–1958 vs. 1999–2018	0.8790	0.0000	0.5760
1979–1998 vs. 1999–2018	0.3680	0.0000	0.8050
	Spring		
1939–1958 vs. 1999–2018	0.2160	0.0000	0.0002
1979–1998 vs. 1999–2018	0.6120	0.0002	0.3030
	Summer		
1939–1958 vs. 1999–2018	0.0466	0.0004	0.0000
1979–1998 vs. 1999–2018	<i>0.0978</i>	0.0150	0.0372

Both spring (September–November) and summer (December–February) precipitation appear to be more variable than the other half of the year, likely due to the convective nature of precipitation during this period. The frequency of years recording below the 25th percentile during spring has decreased since the late-1990s, with a corresponding increase in frequency of years recording above the 85th percentile. As a result, the box plots for spring show a very slight increase between 1939–1958 and 1999–2018. However this is not statistically significant (*p*-value = 0.216). The frequency of years recording below the 25th percentile in summer has also decreased since the 1990s, while the frequency of years above the 80th percentile has increased. This also is apparent in the box plots, with a potential increasing trend in summer precipitation. The difference in mean summer precipitation between 1939–1958 and 1999–2018 is statistically significant (*p*-value = 0.047). Further, there is relatively significant variability between 20-year periods during summer, with low mean precipitation periods followed by higher mean precipitation periods. For example, the difference in mean precipitation between 1979–1998 and 1999–2018 is statistically significant at the 10% confidence level (*p*-value = 0.098). The difference in mean precipitation between 1939–1958 and 1959–1978 is just outside of the 15% confidence level, with a low *p*-value of 0.152. Comparing the low and high mean precipitation periods against each other (e.g., 1959–1978 vs. 1999–2018), neither period has a statistically significant difference in mean precipitation (*p*-value = 0.485). This is despite the suggestion in the box plots that mean precipitation might have increased from 1959–1978 to 1999–2018 (Figure 3b).

While precipitation appears to have a potentially increasing trend for some seasons, the only season with a statistically significant trend is autumn (*p*-value = 0.026). As other seasons have stable or gradual increases in precipitation, annual precipitation appears to remain stable over time. Despite this, the decreasing trend in autumn precipitation is of concern because runoff over the cooler months of the year is greatest in southeast Australia [36] and autumn precipitation is necessary for saturation of the soil, allowing runoff into the catchments from the subsequent winter and spring rainfall [37,38]. In addition, the inter-decadal variability of summer precipitation could result in a relatively prolonged period of low annual precipitation, which is currently being realised.

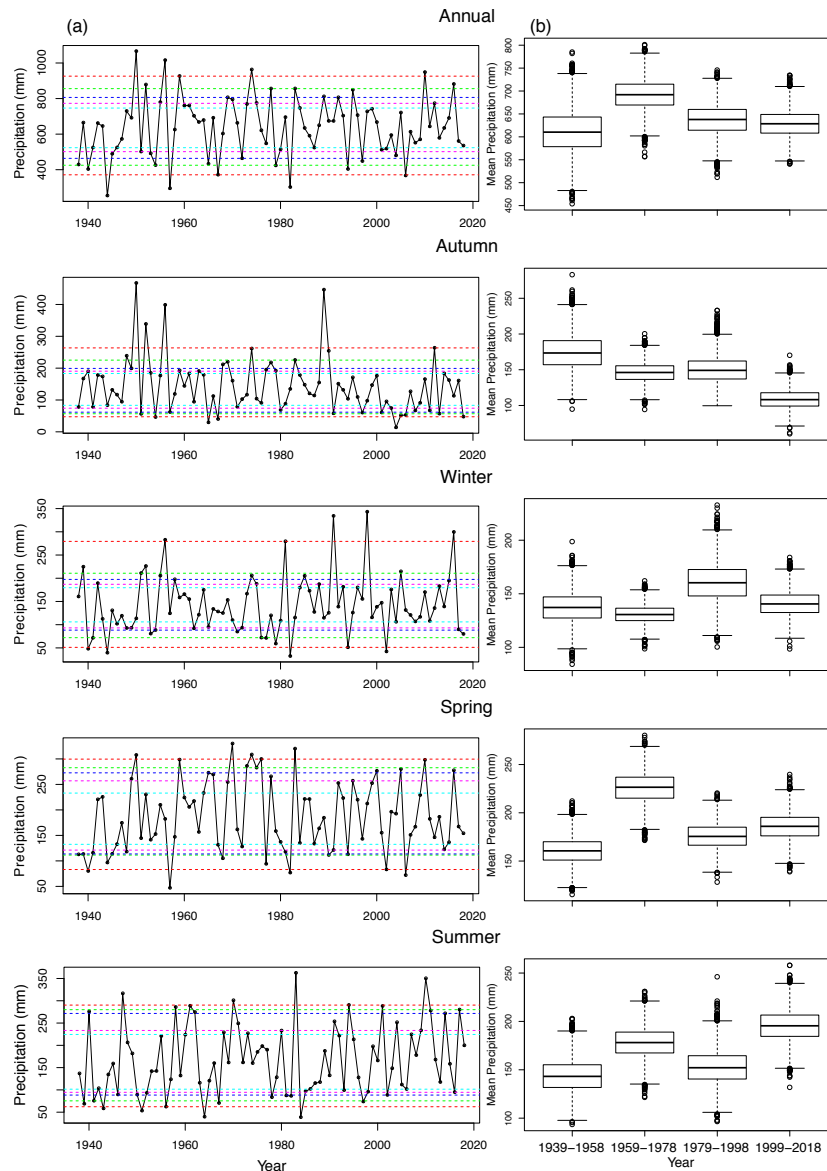


Figure 3. (a) Annual precipitation time series; and (b) box plots of the bootstrapped mean precipitation over 20-year periods for Canberra, both annually and for all four seasons. Dashed lines indicate the 5th and 95th (bottom and top red), 10th and 90th (bottom and top green), 15th and 85th (bottom and top dark blue), 20th and 80th (bottom and top pink), and 25th and 75th percentiles (bottom and top light blue).

3.2. Evolution of Temperature

Temperature plays an important role in drought as it modulates the potential evapotranspiration, creating a positive feedback, as drier soil increases air temperature [4]. While higher temperatures can allow the atmosphere to hold more moisture, sufficient increases in temperature, particularly in regions experiencing drought, will reduce the number of clouds in the atmosphere due to drying of the soil, resulting in less precipitation [5]. Thus, it is necessary to monitor how temperature might also be evolving. An approximation to mean temperature changes can be made by observing changes to the mean minimum and maximum temperatures. Figure 4 provides the time series and box plot figures for TMax in Canberra, both annually and over the four seasons, whereas Figure 5 provides the overview for TMin.

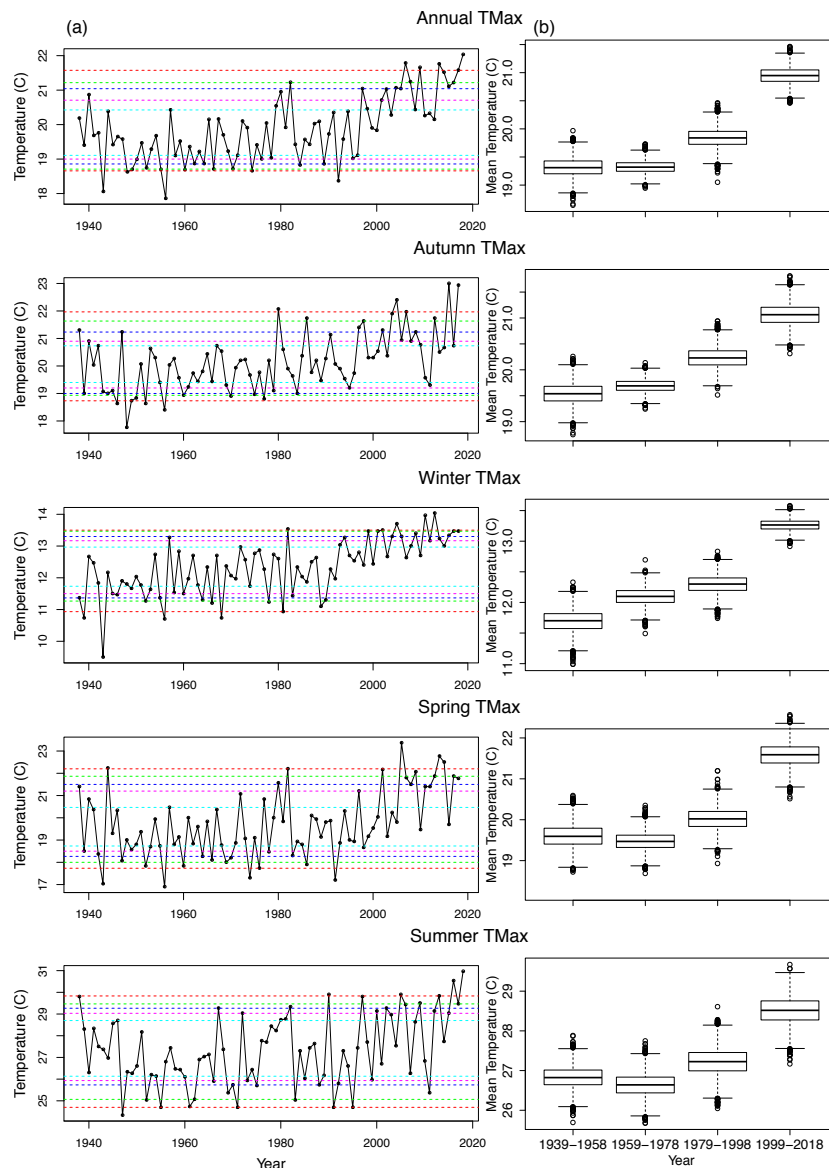


Figure 4. (a) Time series and (b) box plots as in Figure 3 except for TMax.

Annual TMax has been increasing in Canberra, with the difference in mean TMax over the past two twenty-year periods being statistically significant (p -value = 0.000; Table 3). Meanwhile, TMin has had a statistically significant increase compared to the earlier part of the record (1939–1958 vs. 1999–2018 p -value = 0.002). However, TMin over the last two twenty year periods has remained relatively stable with no statistically significant difference in the mean (p -value = 0.521). While TMin has increased since the beginning of the record, the increase is far less than TMax where mean TMax has increased by approximately 1.5 °C over the length of the record (Figure 4b). TMax over the last twenty years stands out against the rest of the time series, with no years recorded below the median, and the majority of years above the 75th percentile occurring during this period. For TMin, the past twenty years have more years above the median, but fewer years above the 95th percentile as the twenty-year period before it, which contributes to these periods having the same means. Variability for both TMax and TMin has remained approximately the same over the observational record with little change to both IQR and the range of the box plots.

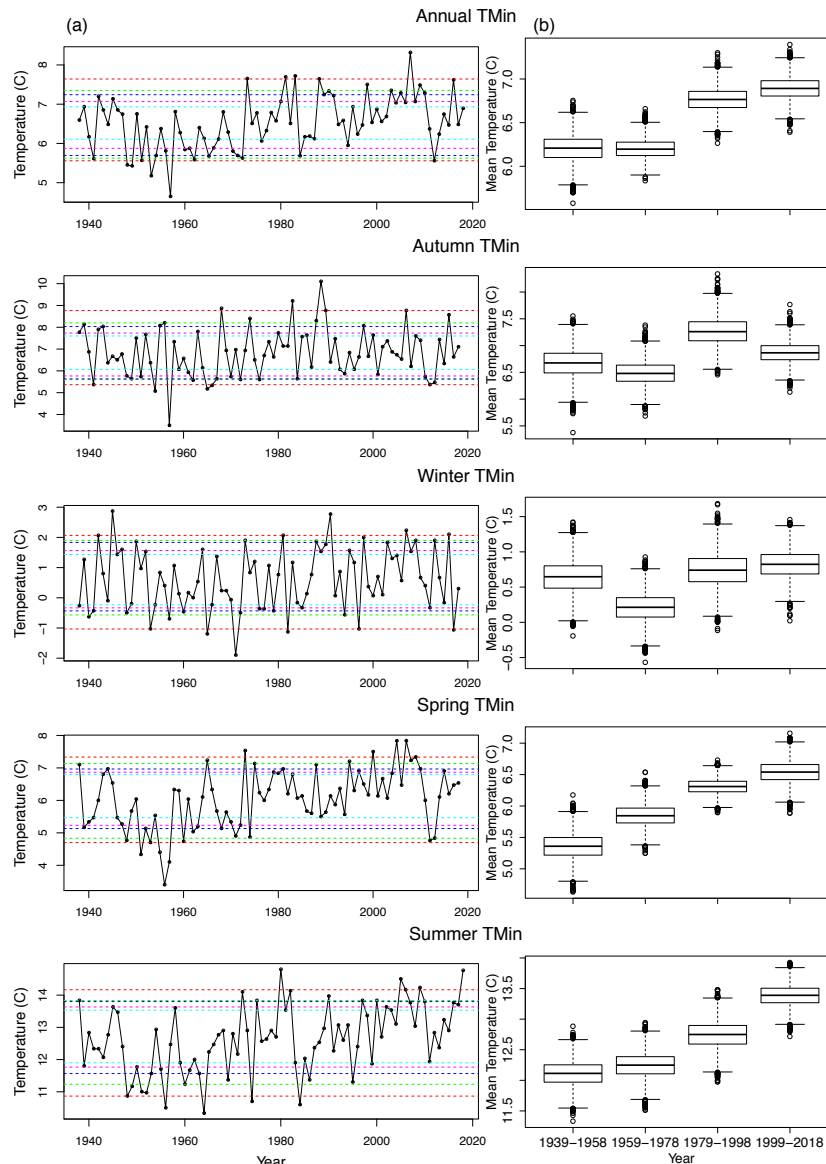


Figure 5. (a) Time series and (b) box plots as in Figure 3 except for TMin.

For the cooler autumn and winter months, there has been a clear increasing trend in TMax with the difference in mean TMax between 1939–1958 and 1999–2018, and between 1979–1998 and 1999–2018 both being statistically significant (p -values < 0.01 ; Table 3). The increasing trend for autumn TMax has been relatively gradual, with the frequency of years recorded below (above) the 25th (75th) percentile decreasing (increasing) since the 1980s (Figure 4a). In contrast, there has been a strong, sudden increase in winter TMax over the past twenty years. Since the late-1990s, there are no years below the median and almost all years above the 75th percentile have occurred during this period. This abrupt change in winter TMax also is evident in the box plots with a statistically significant difference in mean TMax for 1999–2018 compared to 1979–1998 (p -value = 0.000). Further, there has been a strong reduction in variability of winter TMax which is marked by the reduced IQR size in the box plots (Figure 4b). Meanwhile, for TMin in the cooler months there has been little change in both the mean and the variance. This is most clear in the box plots, where the only period showing a statistically significant difference in mean TMin is 1959–1978 (p -value compared to 1999–2018 = 0.042; Figure 5b). Overall, as TMax is increasing and TMin is remaining stable, this suggests the mean temperature during the cooler months is increasing. This increase is of concern as most inflow to water storage occurs during

these months, and higher temperatures can reduce these inflows and further evaporate water that is already stored [3].

For the warmer months (September–February), TMax again has an increasing trend with statistically significant differences in the mean for both spring and summer in all periods tested (Table 3). This increase is relatively gradual with an increase (decrease) in frequency of years above (below) the 75th (25th) percentile since the 1980s in both spring and summer. Notably, there are no years with TMax below the 25th percentile since 2000 in spring, and a large number of years above the 90th percentile in spring and summer (Figure 4a). TMin observations during spring show an increasing trend, although this has tapered off in the most recent 20-year period, with the mean TMin not significantly different from that of 1979–1998 (p -value = 0.303). Since the 1980s, there have only been two years with spring TMin below the 25th percentile, both during the 2000s (Figure 5a). There also was an increased frequency of TMin above the 90th percentile during the late-1990s and 2000s. TMin during summer has continued to increase, with the difference between the two most recent twenty-year periods being statistically significant (p -value = 0.037). Since the 2000s, only one year was recorded just above the 25th percentile, whereas there was a greater frequency in years above the 75th percentile compared to other twenty-year periods. Overall, the mean temperature in the warmer months also appears to be increasing. Because the annual mean temperature is increasing, along with mean temperature during each season, this suggests the region will experience higher mean temperatures during all seasons, resulting in greatly reduced runoff and increased evapotranspiration [3].

3.3. Wavelet Analysis of Precipitation

Wavelet power spectra and global power spectra for Canberra precipitation, both annually and over each season, are presented in Figure 6. There is a strong signal in the 2–7 year period for annual precipitation, which is evident in the global power spectra as a peak over the 3-year and 4-year periods within the 95% confidence band. This is suggestive of an influence from the El-Niño Southern Oscillation (ENSO). Although the influence appears to have weakened slightly over the last 20 years, it remains statistically significant. There are also suggestions at an interdecadal mode with a power spike over the 8–16 year period, although this is not statistically significant due to the length of the time series. ENSO appears to influence precipitation during all seasons, with high power over the 2–7 year period.

Consistent with previous studies, this influence is strongest during spring, with greater total power and a strong peak over the 4-year period in the global power spectrum [39]. The ENSO influence over autumn precipitation is quite strong with high global power spectra, although it has reduced over the past 20 years which could be a consequence of the decreasing trend in autumn precipitation. There is high power over the 8–16 year period in both spring and summer, suggesting these seasons might be influenced by an interdecadal mode of variability such as the PDO or IPO, which is evident in the summer box plots (Figure 3b). However, this signal is mostly not statistically significant at either the 90th or 95th confidence percentiles. Although the PDO/IPO displays an influence on precipitation and drought risk in eastern Australia (e.g., [40,41]), the decadal influence on autumn and winter precipitation in Canberra is not apparent. Given the lack of a statistically significant influence in spring and summer, this is likely why power over the 8–16 year period in annual precipitation is weak and not statistically significant.

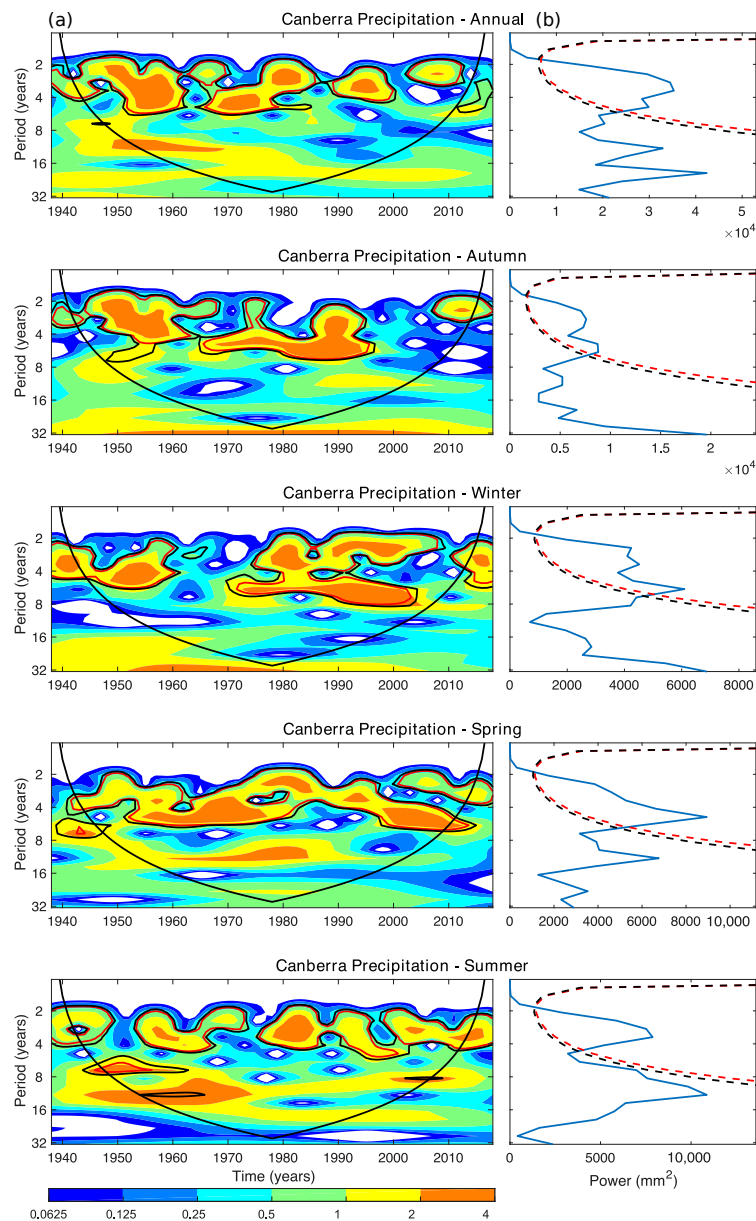


Figure 6. Wavelet analysis for total wet season precipitation for Canberra. Low values (blue) in the wavelet power spectrum (a) indicate low periodicity while high values (red) indicate high periodicity. Peaks in the global power spectrum (b) on the right panel indicate high periodicity. The dashed/solid red line indicates the 95% confidence level, while the dashed/solid black line indicates the 90% confidence level in each plot. The cone of influence is depicted in the wavelet power spectrum by the solid black cone.

3.4. Wavelet Analysis of Temperature

Wavelet power spectra and global power spectra also were computed for TMax and TMin (Figures 7 and 8, respectively). There is weak, but statistically significant, power over the 2–4 year period for annual TMax that is suggestive of an ENSO influence. However, climate drivers generally appear to have less influence on TMax than on precipitation. Occasionally, there is moderate power over the 2–7 year period during autumn, though it does not appear consistently over time. For winter, there is moderate power over the 2–7 year period until the 1990s, when this power largely disappears likely due to the sudden shift in the winter TMax time series (Figure 4). Spring and summer appear to be the most consistent seasons, with moderate power over the 2–7 year period remaining throughout

the time series. Summer TMax has weaker power during the earlier part of the observational record, although the strength of this signal has increased since the 1990s, which is seen in the time series with a clearer periodic signal in summer TMax (Figure 4a).

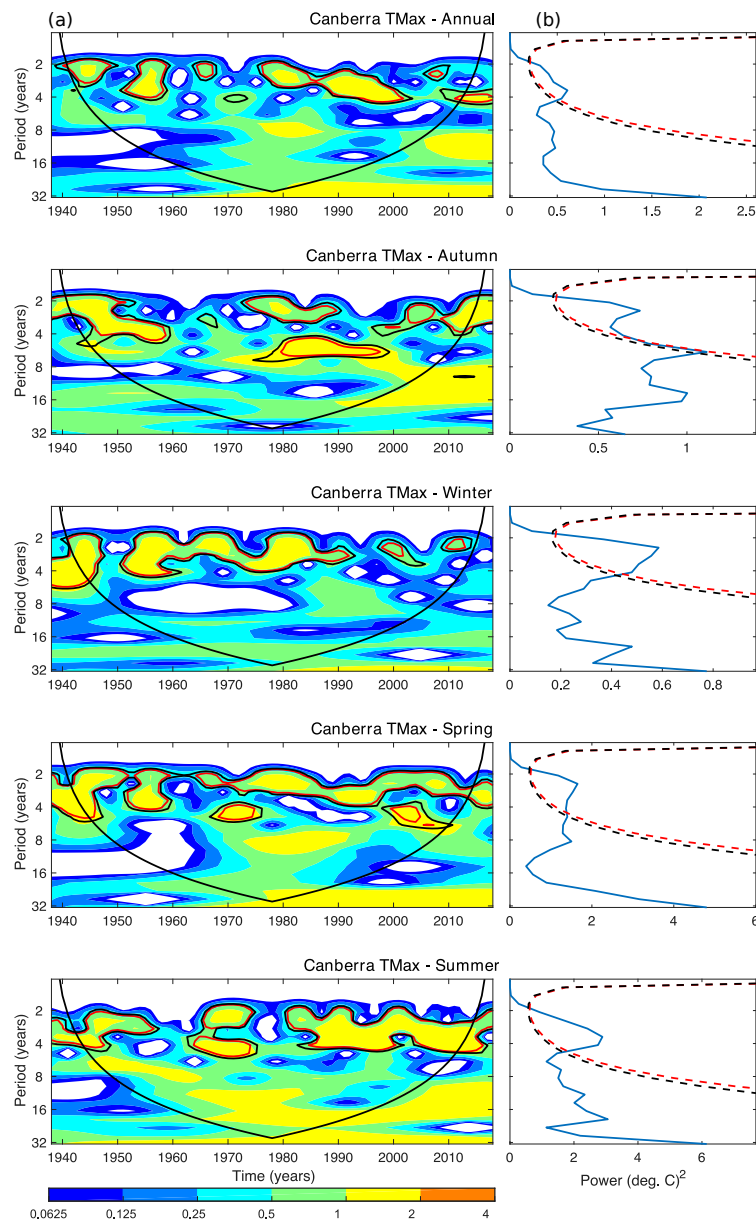


Figure 7. (a) Wavelet power spectra and (b) global power spectra for TMax, with low and high values in the power spectra indicating low and high periodicity, respectively, the cone of influence and statistical significance bands as already explained in Figure 6.

TMin appears to have greater periodicity within the time series. Annual TMin has high power over the 2–4 year period, suggestive of an ENSO influence, but weakens over the most recent 20 years. Autumn TMin has high power over the 2–7 year period, especially prior to the 1990s. Meanwhile, winter TMin has consistent moderate power over the 2–7 year period throughout the time series. TMin during spring and summer also exhibits an ENSO influence. However, in contrast to TMax, this influence begins to weaken from the 1990s for spring, and from the 2000s for summer.

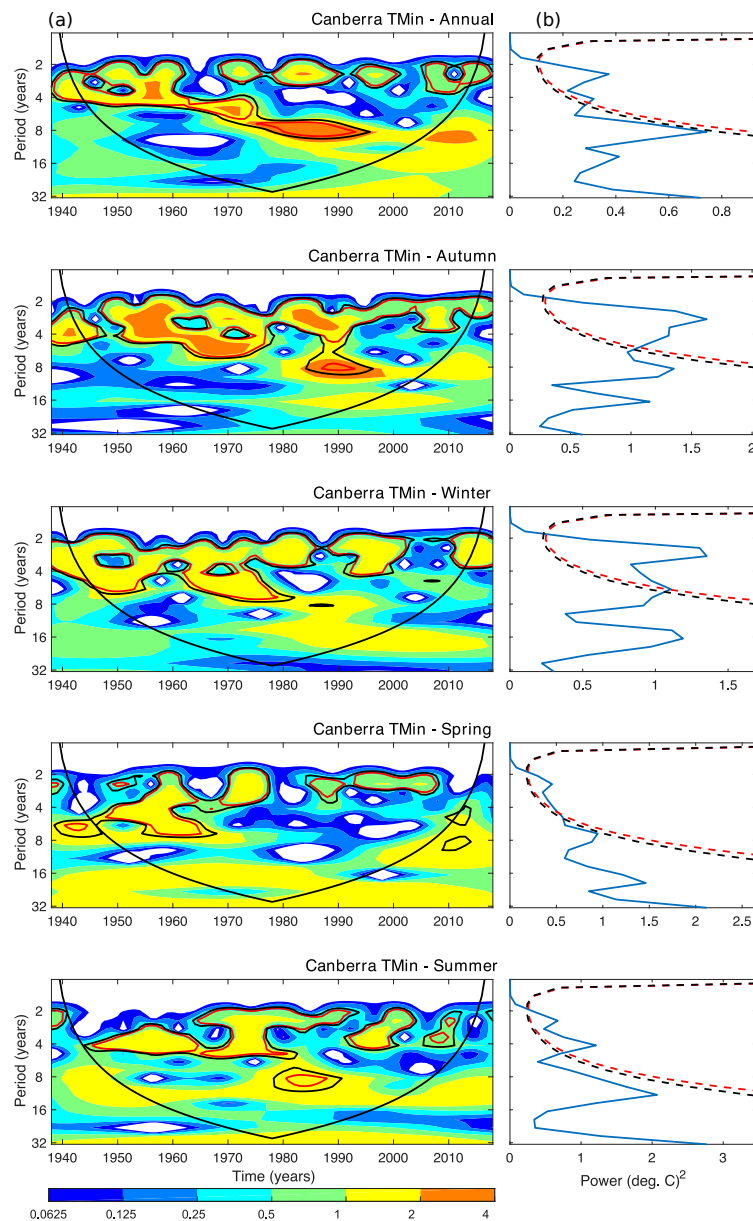


Figure 8. (a) Wavelet power spectra and (b) global power spectra for TMin, with low and high values in the power spectra indicating low and high periodicity, respectively, the cone of influence and statistical significance bands as already explained in Figure 6.

3.5. Training and Prediction of Precipitation

Table 4 shows the correlation, error statistics and skill scores for the final LR and SVR models on the testing data set. Following [42], the definition for skill is:

$$\text{Skill} = 1 - \frac{MSE_{\text{model}}}{MSE_{\text{climatology}}}, \quad (1)$$

where MSE_{model} is the mean square error (MSE) of the model against the observations, and $MSE_{\text{climatology}}$ is the MSE between the mean of the climate from 1957 up to the year prior to the observation, and the year of the observation.

For precipitation prediction, the LR model which performed best is:

$$\text{precipitation} = 689.14 + 75.07 \times \text{SAM} + 70.23 \times \text{SOI} - 238.73 \times \text{DMI} + \varepsilon, \quad (2)$$

where ϵ represents the model error. This model achieved modest correlation, and low RMSE (Table 4). In contrast, the RBF SVR model utilises AMO*SOI, DMI*TSSST and SAM*SOI as predictors. The RBF SVR model for precipitation in this study has $G = 0.5$ and $C = 2^{0.5}$. The poly SVR uses only DMI and SAM as standalone predictors, with weights of -0.65 and 0.37 respectively, and parameters $E = 1$ and $C = 2$.

Table 4. Performance of each model on the testing data set (2006–2017). The best performing model for predicting annual precipitation and TMax is highlighted, based on low RMSE, and both high skill and correlation.

Model	Precip. LR	Precip. SVR (RBF)	Precip. SVR (Poly)	TMax LR	TMax SVR (RBF)	TMax SVR (Poly)
RMSE (mm ² /°C ²)	131.5	126.0	127.502	0.871	1.113	0.783
Skill	0.246	0.308	0.291	0.594	0.337	0.673
Correlation	0.516	0.738	0.568	0.415	0.708	0.531
R ²	0.266	0.544	0.322	0.173	0.502	0.282
Error Mean (mm/°C)	−21.1	−22.6	0.2	0.435	−1.026	−0.595
Error SD (mm/°C)	135.5	129.4	133.2	0.788	0.452	0.531
Error Skewness	−0.181	−0.115	0.133	−0.395	0.387	0.032
Error Kurtosis	−1.191	−0.813	−0.639	−1.453	−1.068	−1.577

In Table 1, there are many potential attributes for each method which could have been selected by the model. The models occasionally selected attributes which appeared in less folds than others. For example, the RBF SVR for annual precipitation selected AMO*SOI, which appeared in 70% of folds. Meanwhile, other attributes like AMO*GlobalIT and AMO*GlobalSSTA, which appeared in 90% of folds, were not selected. It might seem counterintuitive to exclude these attributes, which potentially hold some physical relevance to annual precipitation in Canberra. Despite appearing in a higher percentage of folds, it is likely that these attributes are added later in the forward selection process. As the goal is to obtain a model with the least number of attributes, low RMSE and high correlation, attributes are no longer added to the model when the resulting RMSE on the training set begins to increase.

The RBF SVR model achieves slight reductions in RMSE compared with LR, but has improvements in correlation from 0.516 in the LR model to 0.738 in the SVR RBF (Table 4). Meanwhile, the poly SVR improved on the LR model but did not perform as well as the RBF SVR with slightly higher RMSE and a lower correlation of 0.568 (Table 4). Skill scores of all models were positive, suggesting they perform better than using the mean climatology in predicting precipitation for that year. The RBF SVR performed best, with an increase of approximately 25% in skill, compared to LR (Table 4). However, these skill scores still are relatively low, with the RBF SVR model only achieving a skill of 0.308, suggesting that the current models for precipitation could be improved upon in future work.

Both the LR and RBF SVR tend to underpredict precipitation. Looking at model predictions against observations (not shown), all three models considerably underpredicted annual precipitation in 2010, 2012 and 2016. Annual precipitation in both 2010 and 2016 were above the 10th percentile (Figure 3a). For the RBF SVR, extremes in both high and low precipitation years tend to be underpredicted in the training data (not shown), although the shape of the precipitation time series is mostly achieved. Therefore, some of this prediction error is likely because the model didn't train as close to precipitation extremes. Furthermore, both LR and RBF SVR models contain SOI as a predictor, which is highly correlated to ENSO and its impact on Australia [39,43]. The wavelet analysis in this study suggested ENSO influences annual precipitation (Figure 6). However, this influence has decreased over the most recent 20 years. It is likely that the reduced ENSO relationship could contribute to model errors both in the testing data set and into the future.

3.6. Training and Prediction of TMax

For TMax, the best performing LR model was:

$$TMax = 19.44 + 1.51 \times GlobalSSTA - 1.28 \times DMI \times SOI + 2.92 \times GlobalSSTA \times TSSST + \epsilon. \quad (3)$$

This model obtained the lowest correlation (0.415; Table 4) of all models. The RBF SVR model used DMI*PDO, DMI*TSSST and GlobalT*TSSST as predictors. The free parameters used in this model were $G = 0.5$ and $C = 2^{0.5}$. Meanwhile, the poly SVR model used DMI*SAM and GlobalSSTA with weights of -0.63 and 2.18 respectively, and free parameters $E = 1$ and $C = 2^{0.5}$. The RBF SVR model achieved an RMSE of 1.113, which was worse than the RMSE of 0.871 for LR (Table 4). Meanwhile, the poly SVR recorded the lowest RMSE (0.783; Table 4) on the testing data set but it obtained a lower correlation of 0.531 compared to 0.708 for the RBF SVR (Table 4). Overall, the poly SVR appears to perform best out of these three models with an increase in skill of approximately 13% over the LR model and 100% over the RBF SVR model. Again, all models show improvement compared to the mean climatology, with greater improvements in skill compared to the precipitation models. For example, the poly SVR achieves a skill of 0.673, whereas the RBF SVR for precipitation has a skill of 0.308 (Table 4).

Unlike the LR model, both SVR models tended to underpredict, with the RBF SVR largely underpredicting (Table 4). While both models include attributes related to global warming, it is likely that the contribution has been underestimated. Considering the time series and box plots of annual TMax (Figure 4), the increase in TMax is greatest in the most recent 20-year period, which covers the entire testing data set.

4. Discussion

4.1. Drought Vulnerability

Annual precipitation is not found to be decreasing in Canberra, however the seasonal evolution of precipitation exhibits a decline in autumn precipitation consistent with previous studies [13–15]. Summer precipitation shows a slight increasing trend, which was not statistically significant, and both winter and spring precipitation were stable. While annual precipitation is stable, this change in seasonal precipitation indicates the importance of splitting the time series into smaller subsets to observe any trends masked in the annual time series. Canberra water storages are reliant on cool season rainfall, as this is when runoff tends to be greater and there is less evapotranspiration [36,44]. Further, summer precipitation is highly variable as it is often related to convective storms, with short bursts of heavy rain occurring rather than long periods of soaking rain. This finding suggests that catchment inflows might become less reliable in the future as the climate warms.

Annual mean maximum (TMax) and mean minimum (TMin) temperature show an increasing trend, suggesting the mean temperature in Canberra is also increasing. TMax is increasing in every season, while TMin remains relatively stable over the cooler months and is increasing during the warmer spring and summer months. Hence, the mean temperature appears to be increasing across every season. As a result, potential evapotranspiration will increase in the region, leading to reduced water storages and less rainfall runoff [3,36–38]. Consequently, the trends in both precipitation and temperature suggest that Canberra is likely becoming more susceptible to drought conditions.

4.2. Attribution and Prediction

Attribute selection on annual precipitation and TMax was carried out using LR and SVR with 10-fold cross-validation. The SVR model was developed using both polynomial (poly) and radial basis function (RBF) kernels. Both SVR models improved on precipitation prediction compared to LR, with the greatest improvements from the RBF SVR. This outperformed the LR model, with a correlation of 0.738 compared to 0.516, and skill against the mean climatology of 0.308 compared to 0.246 (Table 4). Predictors influencing annual precipitation in Canberra included DMI, SAM, SOI and TSSST. The wavelet analysis suggested ENSO as a potential driver of precipitation (Figure 6), which can be defined using the SOI [43]. However, other selected attributes were not suggested by the wavelet analysis as their periods are shorter than a year.

Meanwhile, the SVR models for TMax showed mixed results compared to the LR model. The skill from the LR model was 0.594, but the RBF SVR had a reduced skill of 0.337 (Table 4). However, the poly

SVR appreciably improved in skill (0.673) and correlation (0.531) compared to LR (correlation = 0.415; Table 4). The wavelet analysis suggested some ENSO influence on TMax, however this was relatively weak (Figure 7). The common attributes selected included DMI and measures of global warming, such as global sea surface temperature anomalies or global temperature anomalies.

The reliance of the precipitation models on SAM is likely because this climate driver relates to the position of the belt of westerly winds in the southern hemisphere and the associated position of rain-bearing low pressure systems along southern Australia [29,45]. The DMI is recognised to have a strong influence on precipitation especially for southeast Australia as it affects how much moisture is transported into precipitation-bearing systems [29,46]. Furthermore, the DMI has been shown to influence decadal variability of precipitation in southeast Australia [47]. This influence is reflected in both the LR and poly SVR models as the DMI term has the greatest weight out of all predictors. The Tasman Sea is adjacent to eastern Australia, so the two-way interaction between DMI and TSSST used in the RBF SVR suggests that there are concurrent effects of the Tasman Sea and Indian Ocean on moisture advection into precipitation systems which appears to improve prediction. The SOI affects where precipitation is most favourable in the Pacific region. The LR and poly SVR models have either weak or no contribution from SOI, however the RBF SVR includes the two-way interaction between SOI and AMO. This suggests a teleconnection between the Pacific and Atlantic basins that may influence eastern Australia rainfall more than ENSO alone, which has been shown by [48,49]. Interestingly, despite the known effect of co-occurring ENSO and DMI events of the same sign on precipitation extremes in southeast Australia [47], the two-way interaction term between DMI and either SOI or Niño3.4 was not selected by any model. This could be a result of the models not training close to precipitation extremes. As all models utilise SAM and DMI in some form, these climate drivers appear to have the greatest influence on precipitation in Canberra.

As DMI influences moisture in the region [29,46], it might influence TMax through modulating how much cloud cover is present, thereby affecting incoming and outgoing solar radiation. Similarly, due to changing the location of low-pressure systems [29,45], SAM can also have an effect on cloud cover, and therefore TMax in the region which is likely why it was selected by the better performing poly SVR. Further, as there is a strong warming signal in the time series (Figures 4 and 5), inclusion of GlobalSSTA and GlobalT, which are attributes of global warming, would likely improve prediction of TMax. In fact, the global warming signals contribute close to 80% of the poly SVR weights and similarly make 80% of the predictor contribution in LR.

The results highlight that relatively accurate models of precipitation and TMax can be developed utilising machine learning techniques. However, only forward selection was applied to select the models in this study. Other model selection techniques, such as a backward search [34] or correlation-based feature selection [50], could be applied in the future to improve confidence in the generalizability of the results. This especially is the case for predicting annual precipitation, as the skill in the models developed in this study, compared with climatology, were relatively low (skill for LR is 0.246, for SVR RBF is 0.308; Table 4). This reduced skill could be partly due to the weakening relationship of ENSO on annual precipitation (Figure 6a). Further insight might also be gained by considering other climate drivers as potential predictors such as the Atlantic Niño or trans-basin variability index, which have also displayed influences on precipitation variability in Australia [48,49]. Despite including two-way interaction terms and predictors which are well-correlated with other climate drivers not included in this study, replacement with these different climate drivers could result in improved predictive skill over the models presented in this study. Lagged time series of climate predictors could also be used as predictor variables in future work, as they have been shown to be of benefit in past drought modelling studies (e.g., [32,51]). Furthermore, because precipitation patterns vary across the seasons, improved predictive skill might be obtained by developing models for each season [20].

5. Conclusions

In recent years heavy rainfall events, notably in 2010, 2012 and 2016, have provided relief to drought affected southeast Australia. However, it is well-known that Australia is frequently affected by drought, and for southeast Australia it is expected to become increasingly common as global warming continues [15]. Numerous initiatives aimed to improve water availability in Canberra were introduced following the Millennium drought [11] but, as the most recent drought has shown, they do not render the city immune to future drought. This study found that precipitation and temperature are evolving in a way that can decrease catchment inflows and increase the likelihood of drought in Canberra. Greater demand on water resources is expected into the future as the population continues to increase, making it vital to continue research in how factors influencing drought vary over time. This also necessitates the study of improving predictive capabilities of drought.

Multiple linear regression (LR) and support vector regression (SVR) was used to model annual precipitation and mean maximum temperature (TMax), as this can highlight the most important climate drivers and prove useful to management of future water resources [18,19]. Annual precipitation appeared most influenced by DMI, SAM, SOI and Tasman Sea SST anomalies. Meanwhile, DMI and measures of global warming influenced annual TMax. The SVR models notably improved on prediction of annual precipitation and TMax compared to LR. Overall, the results of this study highlight how beneficial continued research in machine learning model development can be for improving the prediction of, and for increasing our understanding of, the underlying dynamics of atmospheric variables and weather patterns.

Author Contributions: Conceptualization, S.M.; methodology, J.H., S.M. and L.M.L.; software, J.H.; formal analysis, J.H.; investigation, J.H.; data curation, J.H.; writing—original draft preparation, J.H.; writing—review and editing, J.H., S.M. and L.M.L.; visualization, J.H.; supervision, S.M. and L.M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by an Australian Government Research Training Program Scholarship. Funding support was provided from the Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers, and the School of Mathematical and Physical Sciences, University of Technology Sydney.

Acknowledgments: We thank three anonymous reviewers for their suggestions which have improved the quality of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Data Availability: Data sets related to this article can be found at <https://doi.org/10.5281/zenodo.3797615> hosted at Zenodo [52].

References

1. Bureau of Meteorology. The “Federation Drought”, 1895–1902. 2009. Available online: <https://webarchive.nla.gov.au/awa/20090330051442/http://pandora.nla.gov.au/pan/96122/20090317-1643/www.bom.gov.au/lam/climate/levelthree/c20thc/drought1.html> (accessed on 4 June 2020).
2. The World War II Droughts 1937–1945. Available online: <https://webarchive.nla.gov.au/awa/20090330051442/http://pandora.nla.gov.au/pan/96122/20090317-1643/www.bom.gov.au/lam/climate/levelthree/c20thc/drought3.html> (accessed on 4 June 2020).
3. Chiew, F.H.S.; Potter, N.J.; Vaze, J.; Petheram, C.; Zhang, L.; Teng, J.; Post, D.A. Observed hydrologic non-stationarity in far south-eastern Australia: Implications for modelling and prediction. *Stoch. Environ. Res. Risk Assess.* **2014**, *28*, 3–15. [CrossRef]
4. Richman, M.B.; Leslie, L.M. Uniqueness and Causes of the California Drought. *Procedia Comput. Sci.* **2015**, *61*, 428–435. [CrossRef]
5. Richman, M.B.; Leslie, L.M. The 2015–2017 Cape Town Drought: Attribution and Prediction Using Machine Learning. *Procedia Comput. Sci.* **2018**, *140*, 248–257. [CrossRef]
6. Niang, I.; Ruppel, O.C.; Abdrabo, M.A.; Essel, A.; Lennard, C.; Padgham, J.; Urquhart, P. 2014: Africa. In *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working*

- Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Technical Report; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2014; pp. 1199–1265.
7. Diffenbaugh, N.S.; Swain, D.L.; Touma, D. Anthropogenic warming has increased drought risk in California. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 3931–3936. [[CrossRef](#)] [[PubMed](#)]
 8. Narisma, G.T.; Pitman, A.J. The Impact of 200 Years of Land Cover Change on the Australian Near-Surface Climate. *J. Hydrometeorol.* **2003**, *4*, 424–436. [[CrossRef](#)]
 9. De Fraiture, C.; Wichelns, D. Satisfying future water demands for agriculture. *Agr. Water Manag.* **2010**, *97*, 502–511. [[CrossRef](#)]
 10. Wilhite, D.A.; Sivakumar, M.V.K.; Pulwarty, R. Managing drought risk in a changing climate: The role of national drought policy. *Weather Clim. Extrem.* **2014**, *3*, 4–13. [[CrossRef](#)]
 11. Icon Water. Water for Generations. 2019. Available online: <https://www.iconwater.com.au/Water-education/Water-and-sewerage-system/Water-for-Generations.aspx> (accessed on 26 February 2019).
 12. Australian Bureau of Statistics. *Australian Demographic Statistics*; Technical Report; Cat. No. 3101.0; Australian Bureau of Statistics: Canberra, Australia, 2018, p. 60.
 13. Timbal, B.; Fawcett, R. A Historical Perspective on Southeastern Australian Rainfall since 1865 Using the Instrumental Record. *J. Clim.* **2013**, *26*, 1112–1129. [[CrossRef](#)]
 14. Timbal, B.; Drosowsky, W. The relationship between the decline of Southeastern Australian rainfall and the strengthening of the subtropical ridge. *Int. J. Climatol.* **2013**, *33*, 1021–1034. [[CrossRef](#)]
 15. Post, D.A.; Timbal, B.; Chiew, F.H.S.; Hendon, H.H.; Nguyen, H.; Moran, R. Decrease in southeastern Australian water availability linked to ongoing Hadley cell expansion. *Earth's Future* **2014**, *2*, 231–238. [[CrossRef](#)]
 16. Burgess, K. Canberra's emergency water pipeline makes short-lived debut. *Canberra Times* **2018**; Article 6000370. Available online: <https://www.canberratimes.com.au/story/6000370/canberras-emergency-water-pipeline-makes-short-lived-debut/#gsc.tab=0> (accessed on 27 February 2019).
 17. Holland, J.E.; Luck, G.W.; Finlayson, C.M. Threats to food production and water quality in the Murray-Darling Basin of Australia. *Ecosyst. Serv.* **2015**, *12*, 55–70. [[CrossRef](#)]
 18. Deo, R.C.; Şahin, M. Application of the Artificial Neural Network model for prediction of monthly Standardized Precipitation and Evapotranspiration Index using hydrometeorological parameters and climate indices in eastern Australia. *Atmos. Res.* **2015**, *161*, 65–81. [[CrossRef](#)]
 19. Bagirov, A.M.; Mahmood, A.; Barton, A. Prediction of monthly rainfall in Victoria, Australia: Clusterwise linear regression approach. *Atmos. Res.* **2017**, *188*, 20–29. [[CrossRef](#)]
 20. Choudhury, D.; Mehrotra, R.; Sharma, A.; Gupta, A.; Sivakumar, B. Effectiveness of CMIP5 Decadal Experiments for Interannual Rainfall Prediction Over Australia. *Water Resour. Res.* **2019**, *55*, 7400–7418. [[CrossRef](#)]
 21. Richman, M.B.; Leslie, L.M. Machine Learning for Attribution of Heat and Drought in Southwestern Australia. *Procedia Comput. Sci.* **2019**, *168*, 3–10. [[CrossRef](#)]
 22. Buishand, T. Some methods for testing the homogeneity of rainfall records. *J. Hydrol.* **1982**, *58*, 11–27. [[CrossRef](#)]
 23. Buishand, T. Tests for detecting a shift in the mean of hydrological time series. *J. Hydrol.* **1984**, *73*, 51–69. [[CrossRef](#)]
 24. Hubert, P.; Carbonnel, J.; Chaouche, A. Segmentation des séries hydrométéorologiques—application à des séries de précipitations et de débits de l'Afrique de l'ouest. *J. Hydrol.* **1989**, *110*, 349–367. [[CrossRef](#)]
 25. Torrence, C.; Compo, G.P. A Practical Guide to Wavelet Analysis. *Bull. Am. Meteorol. Soc.* **1998**, *79*, 61–78. [[CrossRef](#)]
 26. Lau, K.M.; Weng, H. Climate Signal Detection Using Wavelet Transform: How to Make a Time Series Sing. *Bull. Am. Meteorol. Soc.* **1995**, *76*, 2391–2402. [[CrossRef](#)]
 27. Henley, B.; Gergis, J.; Karoly, D.; Power, S.; Kennedy, J.; Folland, C. A tripole index for the interdecadal Pacific oscillation. *Clim. Dyn.* **2015**, *45*, 3077–3090. [[CrossRef](#)]
 28. Richman, M.B.; Leslie, L.M.; Trafalis, T.B.; Mansouri, H. Data Selection Using Support Vector Regression. *Adv. Atmos. Sci.* **2015**, *32*, 277–286. [[CrossRef](#)]
 29. Risbey, J.S.; Pook, M.J.; McIntosh, P.C.; Wheeler, M.C.; Hendon, H.H. On the Remote Drivers of Rainfall Variability in Australia. *Mon. Weather Rev.* **2009**, *137*, 3233–3253. [[CrossRef](#)]

30. Cai, W.; van Rensch, P.; Cowan, T.; Hendon, H.H. Teleconnection Pathways of ENSO and the IOD and the Mechanisms for Impacts on Australian Rainfall. *J. Clim.* **2011**, *24*, 3910–3923. [[CrossRef](#)]
31. Ramsay, H.A.; Richman, M.B.; Leslie, L.M. Seasonal Tropical Cyclone Predictions Using Optimized Combinations of ENSO Regions: Application to the Coral Sea Basin. *J. Clim.* **2014**, *27*, 8527–8542. [[CrossRef](#)]
32. Prasad, R.; Deo, R.C.; Li, Y.; Maraseni, T. Input selection and performance optimization of ANN-based streamflow forecasts in the drought-prone Murray Darling Basin region using IIS and MODWT algorithm. *Atmos. Res.* **2017**, *197*, 42–63. [[CrossRef](#)]
33. Efron, B.; Hastie, T. *Computer Age Statistical Inference*; Cambridge University Press: Cambridge, UK, 2016.
34. Maldonado, S.; Weber, R. A wrapper method for feature selection using Support Vector Machines. *Inf. Sci.* **2009**, *179*, 2208–2217. [[CrossRef](#)]
35. Hsu, C.-W.; Chang, C.-C.; Lin, C.-J. *A Practical Guide to Support Vector Classification*; Technical Report; National Taiwan University: Taipei, Taiwan, 2003; p. 16.
36. Kirono, D.; Chiew, F.; Kent, D. Identification of best predictors for forecasting seasonal rainfall and runoff in Australia. *Hydrol. Process.* **2010**, *24*, 1237–1247. [[CrossRef](#)]
37. Cai, W.; Cowan, T. Dynamics of late autumn rainfall reduction over southeastern Australia. *Geophys. Res. Lett.* **2008**, *35*, L09708. [[CrossRef](#)]
38. Cai, W.; Cowan, T.; Briggs, P.; Raupach, M. Rising temperature depletes soil moisture and exacerbates severe drought conditions across southeast Australia. *Geophys. Res. Lett.* **2009**, *36*. [[CrossRef](#)]
39. Chiew, F.; Piechota, T.; Dracup, J.; McMahon, T. El Nino/Southern Oscillation and Australian rainfall, streamflow and drought: Links and potential for forecasting. *J. Hydrol.* **1998**, *204*, 138–149. [[CrossRef](#)]
40. Power, S.; Casey, T.; Folland, C.; Colman, A.; Mehta, V. Inter-decadal modulation of the impact of ENSO on Australia. *Clim. Dyn.* **1999**, *15*, 319–324. [[CrossRef](#)]
41. Kiem, A.; Franks, S. Multi-decadal variability of drought risk, eastern Australia. *Hydrol. Process.* **2004**, *18*, 2039–2050. [[CrossRef](#)]
42. Wilks, D.S. *Statistical Methods in the Atmospheric Sciences*; Academic Press: Oxford, UK, 2011; Volume 100.
43. McBride, J.L.; Nicholls, N. Seasonal Relationships between Australian Rainfall and the Southern Oscillation. *Mon. Weather Rev.* **1983**, *111*, 1998–2004. [[CrossRef](#)]
44. Kiem, A.S.; Johnson, F.; Westra, S.; van Dijk, A.; Evans, J.P.; O'Donnell, A.; Rouillard, A.; Barr, C.; Tyler, J.; Thyer, M.; et al. Natural hazards in Australia: Droughts. *Clim. Chang.* **2016**, *139*, 37–54. [[CrossRef](#)]
45. Hendon, H.H.; Thompson, D.W.J.; Wheeler, M.C. Australian Rainfall and Surface Temperature Variations Associated with the Southern Hemisphere Annular Mode. *J. Clim.* **2007**, *20*, 2452–2467. [[CrossRef](#)]
46. Ummenhofer, C.C.; England, M.H.; McIntosh, P.C.; Meyers, G.A.; Pook, M.J.; Risbey, J.S.; Gupta, A.S.; Taschetto, A.S. What causes southeast Australia's worst droughts? *Geophys. Res. Lett.* **2009**, *36*. [[CrossRef](#)]
47. Ummenhofer, C.; Gupta, A.; Briggs, P.; England, M.; McIntosh, P.; Meyers, G.; Pook, M.; Raupach, M.; Risbey, J. Indian and Pacific Ocean Influences on Southeast Australian Drought and Soil Moisture. *J. Clim.* **2011**, *24*, 1313–1336. [[CrossRef](#)]
48. Choudhury, D.; Gupta, A.; Sharma, A.; Taschetto, A.; Mehrotra, R.; Sivakumar, B. Impacts of the tropical trans-basin variability on Australian rainfall. *Clim. Dyn.* **2017**, *49*, 1617–1629. [[CrossRef](#)]
49. Johnson, Z.; Chikamoto, Y.; Luo, J.J.; Mochizuki, T. Ocean Impacts on Australian Interannual to Decadal Precipitation Variability. *Climate* **2018**, *6*, 61. [[CrossRef](#)]
50. Hall, M.A. Correlation-based Feature Selection for Machine Learning. Ph.D. Thesis, University of Waikato, Hamilton, New Zealand, 1999.
51. Abbot, J.; Marohasy, J. Skilful rainfall forecasts from artificial neural networks with long duration series and single-month optimization. *Atmos. Res.* **2017**, *197*, 289–299. [[CrossRef](#)]
52. Hartigan, J. Seasonal Precipitation and Temperature Data in Canberra, Australia. Zenodo. 2020. Available online: <https://zenodo.org/record/3797615#.XuBM9XERXIU> (accessed on 6 May 2020).

