

Research Article

Improving the Operational Methodology of Tropical Cyclone Seasonal Prediction in the Australian and the South Pacific Ocean Regions

J. S. Wijnands,¹ K. Shelton,² and Y. Kuleshov^{1,2}

¹ *The University of Melbourne, Parkville, VIC 3010, Australia*

² *Bureau of Meteorology, Docklands, VIC 3008, Australia*

Correspondence should be addressed to Y. Kuleshov; y.kuleshov@bom.gov.au

Received 23 July 2013; Revised 20 December 2013; Accepted 6 January 2014; Published 17 March 2014

Academic Editor: Jean-Pierre Barriot

Copyright © 2014 J. S. Wijnands et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Tropical cyclones (TCs) can have a major impact on the coastal communities of Australia and Pacific Island countries. Preparedness is one of the key factors to limit TC impacts and the Australian Bureau of Meteorology issues an outlook of TC seasonal activity ahead of TC season for the Australian Region (AR; 5°S to 40°S, 90°E to 160°E) and the South Pacific Ocean (SPO; 5°S to 40°S, 142.5°E to 120°W). This paper investigates the use of support vector regression models and new explanatory variables to improve the accuracy of seasonal TC predictions. Correlation analysis and subsequent cross-validation of the generated models showed that the Dipole Mode Index (DMI) performs well as an explanatory variable for TC prediction in both AR and SPO, Niño4 SST anomalies—in AR and Niño1+2 SST anomalies—in SPO. For both AR and SPO, the developed model which utilised the combination of Niño1+2 SST anomalies, Niño4 SST anomalies, and DMI had the best forecasting performance. The support vector regression models outperform the current models based on linear discriminant analysis approach for both regions, improving the standard deviation of errors in cross-validation from 2.87 to 2.27 for AR and from 4.91 to 3.92 for SPO.

1. Introduction

Tropical cyclones (TCs) are extreme weather events that form each season over tropical oceans. In the Australian Region (AR; 5°S to 40°S, 90°E to 160°E) and the South Pacific Ocean (SPO; 5°S to 40°S, 142.5°E to 120°W) TCs mainly occur during the six months from November to April; however, TCs do occur outside the TC season. TCs can have a major impact on human life when they make land-fall. For example, in 1974 the Australian city of Darwin was devastated by TC Tracy, killing 71 people [1]. The main dangers of TCs are destructive winds, heavy rainfall, flooding, and storm surges. Besides the impact on human life, TCs can have a severe impact upon Pacific island countries and their economies. For example, the small island of Tikopia located in the SPO (12° 18'S 168° 50'E) was completely devastated in 2002 by TC Zoe [2].

Preparation for TCs is an important element to reduce the destructive impacts of TC land-falls. Therefore, the Bureau of

Meteorology aims to predict the likelihood of TC activity for several regions at the start of each TC season. Currently the operational statistical model of the Bureau of Meteorology consists of two linear discriminant analysis (LDA) models, one based on the Southern Oscillation Index (SOI) and one based on Niño3.4 Sea Surface Temperatures (SST) [3]. The results have been disseminated in a form of TC seasonal outlooks, to provide local communities and government authorities with early advice about expected TC activity. For public release, the models produce probabilities of above median tropical cyclone activity; however, they can also produce the most likely number of TCs in the upcoming season.

A different statistical model-based approach is to use machine learning algorithms. A recent pilot study by Richman and Leslie [4] indicates that support vector regression performs well for TC prediction. In this paper we investigate the use of machine learning algorithms for the Bureau of Meteorology TC forecasting regions.

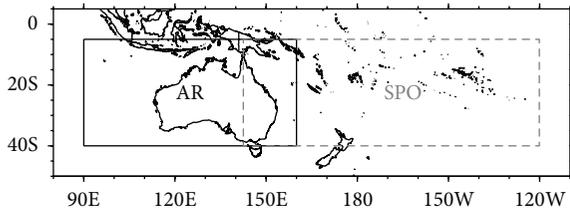


FIGURE 1: Map of two study areas: AR is defined as the area 90°E - 160°E and 5°S - 40°S and SPO is defined as the region 142.5°E - 120°W and 5°S - 40°S .

2. Data and Methodology

The Bureau of Meteorology's National Climate Centre (NCC) has developed a TC archive for the Southern Hemisphere, in close collaboration with international partners [5]. The number of TCs for each season can be obtained from the Southern Hemisphere TC archive via the Pacific Tropical Cyclone Data Portal of the Bureau of Meteorology (<http://www.bom.gov.au/cyclone/history/tracks/>). The Southern Hemisphere TC archive has been developed by consolidating best track data prepared by Regional Specialised Meteorological Centres (RSMCs) Nadi (Fiji) and Tropical Cyclone Warning Centres (TCWCs) Brisbane, Darwin and Perth (Australia) and Wellington (New Zealand). To estimate TC intensity, Dvorak methodology is used operationally by the RSMCs and TCWCs with the World Meteorological Organization's responsibilities for issuing TC warning and preparing best track data [6]. To keep consistency with previous studies, the genesis of a TC is defined when a cyclonic system first attains a central pressure equal to or less than 995 hPa [7–11]. AR is defined as the area 90°E - 160°E and 5°S - 40°S and SPO is defined as the region 142.5°E - 120°W and 5°S - 40°S (Figure 1). NCC of the Australian Bureau of Meteorology has operational responsibilities to issue TC seasonal outlooks for both the AR and the SPO region and to provide Australians and population of countries in the South Pacific Ocean with early warning advice about TC activity expected in the coming season. Consequently, there is an overlap between two areas under investigation. This overlap was taken in consideration in this study; if a TC was recorded in both AR and SPO region, we included it in our analysis for both areas.

Operational tropical cyclone seasonal prediction in AR started with pioneering work by Nicholls [12, 13] who developed statistical methodology for forecasting TC activity in the upcoming season based on state of the El Niño—Southern Oscillation (ENSO). ENSO is a large-scale climate phenomenon that occurs across the tropical Pacific Ocean and has two distinctly different phases: warm (El Niño) and cold (La Niña), with an intervening neutral phase. Relationships between ENSO and TC activity in the Western Pacific and its smaller subregions such as AR are well understood and described in the literature (e.g., [8, 14, 15]). These relationships allow forecasting TC seasonal activity in October for six months ahead using values of ENSO indices which describe the state of the atmosphere and ocean in

the central Pacific in the preceding months of July–August–September. Explanatory variables that have been investigated in this study include a number of ENSO indices such as Niño1+2, Niño3, Niño3.4 and Niño4 SST anomalies, 5VAR index, Multivariate ENSO Index (MEI), and El Niño Modoki Index. Detailed description of the ENSO indices used in this study can be found in Trenberth [16], Ashok et al. [17], Wolter and Timlin [18], and Kuleshov et al. [9]. Two other well-established indices which have been used as potential predictors of TC activity in AR and SPO are the Dipole Mode Index (DMI) and the Southern Oscillation Index (SOI).

Values for Niño1+2, Niño3, Niño3.4, and Niño4 SST anomalies have been obtained from the Climate Prediction Center (CPC) of the National Oceanic and Atmospheric Administration (NOAA) (<http://www.cpc.ncep.noaa.gov/data/indices/ersst3b.nino.mth.81-10.ascii>). 5VAR is an NCC-internal ENSO index, based on the first principal component of monthly Darwin mean sea level pressure (MSLP), Tahiti MSLP, and Niño3, Niño3.4, and Niño4 SST indices [9–11]. Data for MEI have been obtained from NOAA (<http://www.esrl.noaa.gov/psd/enso/mei/table.html>). The Japan Agency for Marine–Earth Science and Technology (JAMSTEC) provides data for the El Niño Modoki Index (<http://www.jamstec.go.jp/frsgc/research/d1/iod/DATA/emi.monthly.txt>) and the Dipole Mode Index (http://www.jamstec.go.jp/frsgc/research/d1/iod/DATA/dmi_HadISST_jan1958-dec2012.txt). Data for SOI have been obtained from the Bureau of Meteorology (<ftp://ftp.bom.gov.au/anon/home/ncc/www/sco/soi/soiplaintext.html>).

Since meteorological satellites came into operational use at the end of the 1960s, the reliability of TC data has improved significantly [19]. Earlier data are not considered sufficiently accurate for modelling purposes and therefore data prior to the 1969–70 season have not been included in our analysis. Kuleshov et al. [10, 11] have identified this point as an outlier in several models. The SPO best track database contains data up until the 2010–11 season, whereas AR contains best track information up to the 2011–12 season. Therefore, seasons from 1970–71 to 2010–11 data have been used for SPO, while 1970–71 to 2011–12 data have been used for AR.

Using Weka [20] the performance of several machine learning algorithms has been investigated. Weka (Waikato Environment for Knowledge Analysis) is a collection of machine learning algorithms for data mining tasks, developed by the University of Waikato in New Zealand. This open source software package includes algorithms such as decision trees, nearest neighbour classifiers, and regression models. We applied various machine learning algorithms to the TC data set, including isotonic regression, least median squared regression, linear regression, multilayer perceptron, pace regression, normalised Gaussian radial basis function network, support vector regression, K-nearest neighbours classifier, K^* instance-based classifier, locally weighted learning, additive regression, conjunctive rule, decision table majority classifier, M5 rules, M5P, decision stump, REP-Tree decision tree, bagging in combination with decision tree algorithms, and random subspace in combination with decision tree algorithms. To avoid artificial skill we did not use all training data to fit the models, but instead

performed a cross-validation analysis for each of the potential prediction models. Therefore, the model prediction for each TC season was an out of sample prediction. First, we utilised all explanatory variables for each algorithm to make an initial selection of algorithms with potential skill. This selection was made by ranking all models based on the root mean squared error (RMSE) statistic. Then, we focussed on improving the performance of promising algorithms for seasonal TC prediction by selecting different combinations of explanatory variables.

There was a large difference between all models, as some showed very little skill in the seasonal prediction of TCs. Most decision trees that we investigated did not perform very well, possibly due to the small number of years of data that is currently available. Overall, out of the algorithms that we investigated, the machine learning algorithm that performed best for the seasonal prediction of TCs was support vector regression. This algorithm exhibited the lowest RMSE in cross-validation and will be the focus of the remainder of this paper.

The support vector regression model is an extension to ε -support vector regression [21]. We assume we have a set of observations y_i and several explanatory variables $x_{1,i}$, $x_{2,i}$, $x_{3,i}$, ..., $x_{n,i}$ with i the number of observations and n the number of explanatory variables. In ε -support vector regression the goal is to find a function $f(\underline{x})$ that predicts the observations and has a maximum error smaller than or equal to ε , while being as flat as possible, meaning none of the coefficients is very large relative to the rest of the coefficients. For a linear function $f(\underline{x})$ this can be described by:

$$f(\underline{x}) = \langle \underline{w}, \underline{x} \rangle + b \text{ with } \underline{w}, \underline{x} \in R^n, b \in R \quad (1)$$

and n the number of explanatory variables.

In this equation \underline{x} is a vector with explanatory variables, \underline{w} is a vector containing coefficients, and b is the bias term. We search for optimal solution by changing coefficients vector \underline{w} . The notation $\langle \underline{w}, \underline{x} \rangle$ denotes the dot product between \underline{w} and \underline{x} . Flatness for $f(\underline{x})$ means searching for a set of small coefficients in vector \underline{w} . The idea behind this is to make the model less sensitive to errors in measurement or random shocks in explanatory variables, leading to better prediction performance.

To solve (1) the following is used:

$$\begin{aligned} &\text{Minimise } \frac{1}{2} \langle \underline{w}, \underline{w} \rangle \\ &\text{subject to } \begin{cases} y_i - \langle \underline{w}, \underline{x}_i \rangle - b \leq \varepsilon \\ \langle \underline{w}, \underline{x}_i \rangle + b - y_i \leq \varepsilon. \end{cases} \end{aligned} \quad (2)$$

The formulation described above can sometimes be infeasible, when no function exists where the maximum error is ε . To solve this problem the methodology is extended by adding slack variables ξ_i and ξ_i^* to the model. In case the absolute error for observation y_i , given by $|y_i - f(x_i)|$, is larger than ε , one of the slack variables ξ_i or ξ_i^* will be equal to the excess error and the other slack variable will be set to 0. Instead of minimising $(1/2)\langle \underline{w}, \underline{w} \rangle$, we minimise $(1/2)\langle \underline{w}, \underline{w} \rangle$ plus a penalty for errors exceeding the threshold ε .

TABLE 1: Correlation with number of TCs in AR.

Explanatory variable	Month with highest correlation	Correlation
Niño4 SST anomalies	September	-0.672
Dipole Mode Index	September	-0.615
Niño3.4 SST anomalies	Average Aug/Sep	-0.611
5VAR index	August	-0.597
El Niño Modoki Index	September	-0.590
SOI	August	0.583
Niño3 SST anomalies	September	-0.523
MEI	September	-0.520
Niño1 + 2 SST anomalies	September	-0.289

Consider

$$\begin{aligned} &\text{Minimise } \frac{1}{2} \langle \underline{w}, \underline{w} \rangle + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ &\text{subject to } \begin{cases} y_i - \langle \underline{w}, \underline{x}_i \rangle - b \leq \varepsilon + \xi_i \\ \langle \underline{w}, \underline{x}_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0. \end{cases} \end{aligned} \quad (3)$$

The complexity factor $C > 0$ is used to mathematically enforce more importance on the flatness of $f(\underline{x})$ or the amount up to which deviations larger than ε are tolerated. Equation (3) shows the approach that is used for seasonal TC prediction. Further information on support vector regression is provided by Smola and Schölkopf [22].

3. Results

3.1. Selection of Explanatory Variables. To decide which explanatory variables to use, an initial correlation analysis has been performed. As seasonal prediction has to be performed in October before the TC season starts, only values for September or earlier months can be used for forecasting. The values for the preceding months July, August, and September were investigated and also the two and three month averages for August and September and July, August and September. Results for several potential explanatory variables in AR are shown in Table 1 ranked by absolute correlation. Table 2 presents the results of correlation analysis for SPO; however, explanatory variable are listed in the same order as in Table 1 rather than ranked by absolute correlation, to assist readers in comparing the results for both regions. The explanatory variables with a high absolute correlation would presumably work well when used in a model to predict TC activity. This analysis has been performed separately for AR and SPO region to select the most promising predictors for each region. As for AR the best track data were available until the 2011-12 season; correlations between number of TCs and the explanatory variables have been investigated over the 1970-71–2011-12 period. For SPO this period was 1970-71–2010-11.

This analysis assisted us in selecting the indices with high correlation with regional TC activity. High correlation with some indices could be easily traced to teleconnections with

TABLE 2: Correlation with number of TCs in SPO.

Explanatory variable	Month with highest correlation	Correlation
Niño4 SST anomalies	August	-0.157
Dipole Mode Index	August	-0.216
Niño3.4 SST anomalies	September	0.056
5VAR index	September	0.055
El Niño Modoki Index	Average Aug/Sep	-0.307
SOI	August	0.064
Niño3 SST anomalies	September	0.149
MEI	July	0.127
Niño1+2 SST anomalies	Average Jul/Aug/Sep	0.252

the environment favourable for TC genesis and development in the selected regions of the Southern Hemisphere [8]; for example, negative (positive) Niño4 SST anomalies in AR (SPO) could be explained through cooler (warmer) SST in the TC genesis region. On the other hand, for some indices it is not so straight forward, for example, negative (positive) Niño1+2 SST anomalies in AR (SPO) which we found in this study. In addition, we found that combination of explanatory variables gives better results than using only one variable (details are in following sections).

Concurrently, exploring variants of ENSO Modoki for the seasonal prediction of Coral Sea TC activity, Ramsay et al. [23] also found that predictive skill is maximised when indices capturing the relative changes to equatorial SSTs in the Pacific are included; hence, the use of Niño 4 and Niño1+2 together in their study too. While further detailed research is required to explain teleconnections which have some influence there, it is clear that the combination of Niño indices describes better the basinwide equatorial Pacific SST anomaly variations than any Niño index alone. However, correlation between the indices should be taken in consideration when combining them. For example, Niño4, Niño3, and Niño3.4 indices are strongly correlated (using the data from <http://www.cpc.ncep.noaa.gov/data/indices/ersst3b.nino.mth.81-10.ascii> one can find a correlation of 0.77 for September Niño4 with Niño3 (1970–2012) indicating that these indices indeed similarly explain SST variability in the central Pacific). On the other hand, correlation for Niño1+2 with Niño4 is only 0.59 which suggests that a combination of these two indices would add value to more comprehensive description of basinwide changes related to ENSO. Note the Niño1+2 and Niño4 indices have previously been combined into the Trans-Niño index (http://www.cgd.ucar.edu/cas/catalog/climind/TNI_N34/) so there is a history of using these together. However, to the best of our knowledge, there is no literature on relationship between Niño1+2 variability and Australian climate. The physical link between Niño1+2 and the variability of TCs in Australian and SPO regions will be a subject of our future research but is beyond the scope of the current study.

The aim of this analysis was to select the months with the highest correlation for all potential model variables. In the

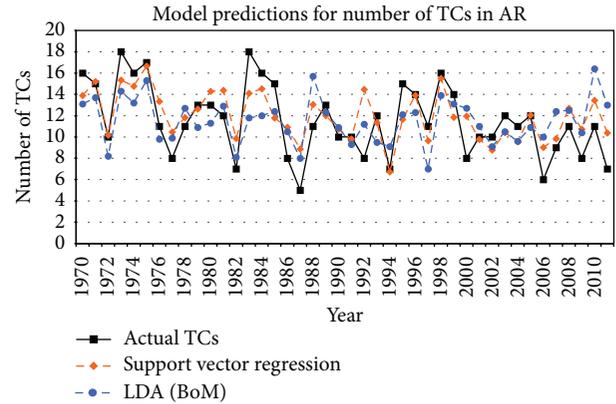


FIGURE 2: Leave-one-out cross-validation results AR.

TABLE 3: Goodness-of-fit statistics AR.

	Support vector regression	LDA (Bureau of Meteorology)
R^2	53.3%	27.4%
Adjusted R^2	49.6%	23.6%
St. dev. errors	2.266	2.874

remainder of this paper the months in Tables 1 and 2 will be used to model TC activity for AR and SPO, respectively.

3.2. Support Vector Regression Models. For both AR and SPO the combination of explanatory variables Niño1+2 SST anomalies, Niño4 SST anomalies, and DMI has the best forecasting performance. This was tested using leave-one-out cross-validation. Although the El Niño Modoki Index had the highest correlation with number of TCs in SPO, utilising this variable in the support vector regression resulted in worse forecasting performance than the model that was finally selected. As a result of the correlation analysis the values of the explanatory variables are taken from different months for AR and SPO. The support vector regression model for AR uses September values for all three explanatory variables, while for SPO the three-month average over July, August, and September is used for Niño1+2 SST anomalies and the August value for Niño4 SST anomalies and DMI. The explanatory variables have been normalised before support vector regression was applied.

Both AR and SPO support vector regression models used the polynomial kernel $K(x, y) = \langle x, y \rangle^p$. In support vector regression input variables are mapped into a feature space using this kernel function. The regression function then aims to separate data points using boundaries in this feature space. The polynomial kernel allows for linear separation (exponent $p = 1$) or nonlinear separation boundaries (exponent $p \geq 2$) [24]. For our data set we obtained the best results using linear separation ($p = 1$). The algorithm for support vector machines for regression used the adaption of the stopping criterion by Shevade et al. [25]. The complexity parameter C was optimised using cross-validation and is 1.1 for AR and 2.5 for SPO.

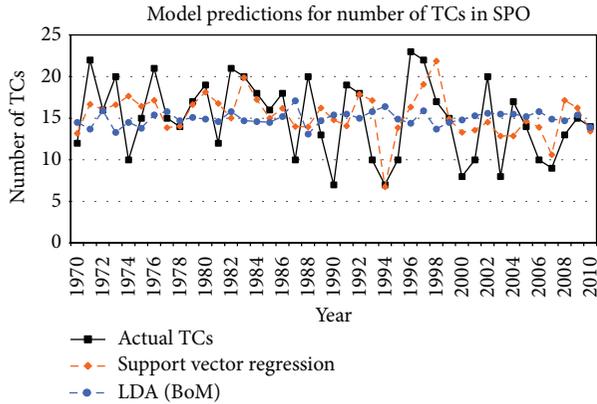


FIGURE 3: Leave-one-out cross-validation results SPO.

TABLE 4: Goodness-of-fit statistics SPO.

	Support vector regression	LDA (Bureau of Meteorology)
R^2	26.6%	-13.7%
Adjusted R^2	20.7%	-19.7%
St. dev. errors	3.925	4.910

3.3. *Cross Validation Results.* Figures 2 and 3 show the results of the leave-one-out cross-validation for AR and SPO, respectively. In other words, for every year, the target season is left out of the training period and all other seasons are used to create the model and then a forecast is made for the season left out.

Goodness-of-fit statistics are given in Tables 3 and 4. In these tables R^2 , which is the percentage of explained variance, is calculated as $R^2 = 1 - (SS_{res}/SS_{tot})$. In this formula $SS_{res} = \sum_{i=1}^n (y_i - f(x_i))^2$ is the sum of squares of the residuals, with y_i the actual number of TCs for year i , $f(x_i)$ the out of sample prediction for the number of TCs in year i , and $n = 42$ for AR and $n = 41$ for SPO. $SS_{tot} = \sum_{i=1}^n (y_i - \bar{y})^2$ is the total sum of squares, with \bar{y} the average actual number of TCs. A negative value of R^2 for SPO indicates that the variance of the errors is larger than the variance of the observations. Hence, it means that a model predicting the period average number of TCs, in this case 15, will yield a better prediction than the current LDA (Bureau of Meteorology) methodology. Adjusted R^2 is a measure that adjusts R^2 to account for the number of explanatory variables that are utilised. As R^2 increases when extra explanatory variables are added to a model, adjusted R^2 allows for a better comparison of models with a different number of explanatory variables. Adjusted $R^2 = 1 - (1 - R^2)((n - 1)/(n - p - 1))$ with $n = 42$ for AR, $n = 41$ for SPO, $p = 3$ the number of explanatory variables of the support vector regression model, and $p = 2$ for the LDA model.

First, it is clear that forecasting TC numbers for SPO is more difficult than for AR, as both models have significantly better performance for AR. The correlation analysis that was performed already showed this. Second, the graphs as well

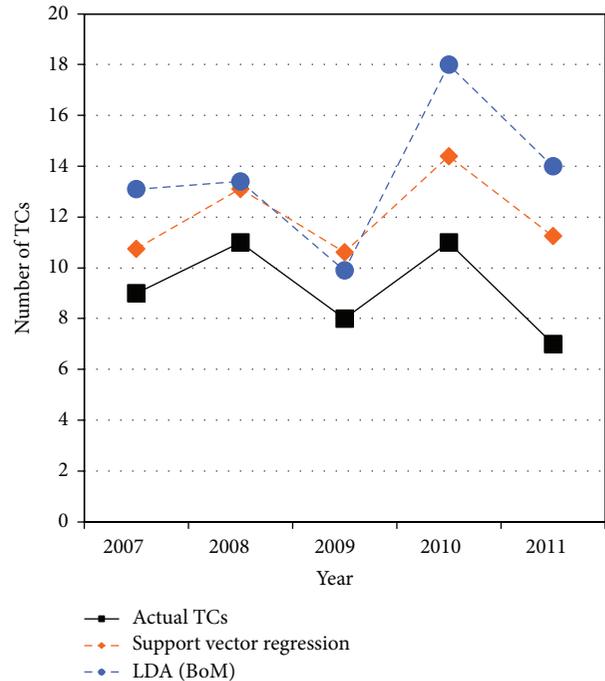


FIGURE 4: Independent forecasts for AR (2007–2011).

TABLE 5: MAE for independent forecasts (seasons from 2007-08).

Region	Support vector regression	LDA (Bureau of Meteorology)
AR	2.823	4.480
SPO	1.893	2.375

as the goodness-of-fit statistics show that the support vector regression models perform better in predicting TC numbers than the LDA method.

3.4. *Independent Forecasts 2007–2011.* With cross-validation the number of TCs for each fold is predicted using data from years that occur after the predicted year (except for the prediction for the final season), which is not possible in reality. To simulate the operational forecasting performance of the model, a forecast is made for the most recent seasons. Data from the 1970-71–2006-07 seasons are used to train the support vector regression and linear discriminant analysis models, in order to predict the number of TCs for the 2007-08–2011-12 seasons. For SPO the 2011-12 predictions were ignored as best track data were not yet available for this season. The results are shown in Figures 4 and 5. (In Figures 4 and 5, a TC season is indicated as a year when the season begins, e.g., for 2007-08 TC season which lasts from November 2007 to April 2008 inclusive; the season is indicated as 2007.)

This graph for AR shows that support vector regression is more consistent in forecasting TC activity than linear discriminant analysis, as it follows a similar pattern as the observed number of TCs. For SPO support vector regression also performs better than the LDA methodology, which

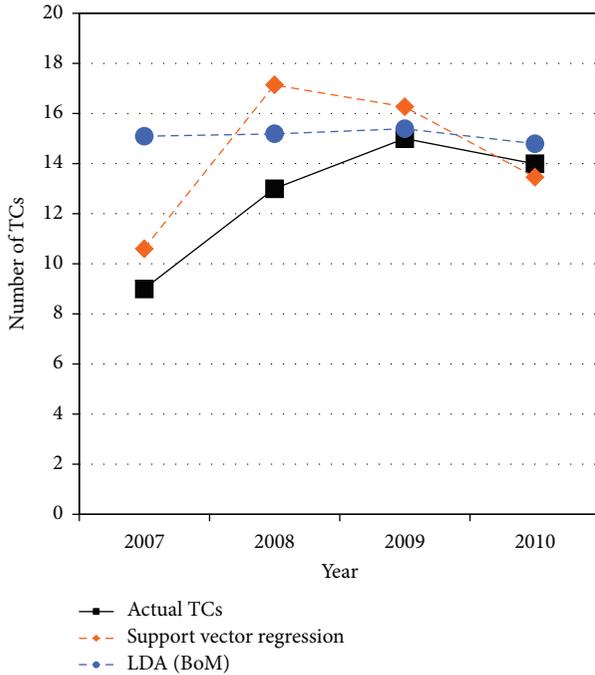


FIGURE 5: Independent forecasts for SPO (2007–2010).

TABLE 6: Standard deviation of errors for independent forecasts (seasons from 2007–08).

Region	Support vector regression	LDA (Bureau of Meteorology)
AR	1.011	2.441
SPO	1.930	2.600

predicts around 15 TCs for every year and fails to capture the interannual variability in TC activity. Support vector regression for SPO performs better, although the 2008–09 TC season prediction shows it is not a very consistent model. For both models the mean absolute error (MAE) of the model prediction versus actual TC number for AR and SPO is shown in Table 5 and the standard deviation of errors in Table 6. These tables show that the MAE for AR is larger, although the standard deviation of errors for AR is smaller than for SPO. It should be noted that the sample size for this analysis is small.

In order to quantify the impact of utilising different ensembles of seasons for model development, we calculated these statistics as well (Tables 7 and 8) utilising the predictions from the cross-validation analysis. Only the seasons from 2007–08 until 2011–12 for AR and 2007–08 until 2010–11 for SPO are included to calculate MAE and standard deviation of errors.

For SPO, the MAE and standard deviation of errors are similar in the hindcasts and forecasts for the seasons after 2007–08. However, for AR we observe that the errors in the forecast are larger than in the hindcast. This could indicate that errors increase for predictions further into the future. A larger sample would be required to test this.

TABLE 7: MAE in cross validated hindcasts (seasons from 2007–08).

Region	Support vector regression	LDA (Bureau of Meteorology)
AR	2.223	3.740
SPO	1.896	2.025

TABLE 8: Standard deviation of errors in cross validated hindcasts (seasons from 2007–08).

Region	Support vector regression	LDA (Bureau of Meteorology)
AR	0.979	1.923
SPO	1.942	2.724

4. Discussion

The aim of this study is to improve the accuracy of seasonal TC predictions, as the performance of the Bureau of Meteorology’s LDA models has been less efficient in recent years [3]. The results show that a support vector regression approach can indeed improve upon the current methodology for both AR and SPO. This research is consistent with past studies where support vector regression was used. For example, Richman and Leslie [4] have investigated the application of machine learning algorithms for seasonal prediction of TCs, concluding that support vector regression leads to better results than linear regression models. Furthermore, their study identified the Quasibiennial Oscillation (QBO) as an explanatory variable that boosts model performance. This variable is only recorded from 1979 onwards [26]. Therefore, less training data will be available for model development, as the 1970–71–1978–79 seasons cannot be used. Moreover, the physical reasons for improved model performance with QBO included as an additional variable are not clear. Analysing the influence of the QBO on TC activity, Camargo and Sobel [27] concluded that although there was a statistically significant relationship between the QBO and TCs in the Atlantic from the 1950s to the 1980s that relationship is no longer present in later years. As for other regions, only in AR is the relationship of TCs with the QBO significant for 1953–1982; however, similar to the case of the Atlantic, the significance disappears in 1983–2008. This change could possibly be attributed to changes in observational procedures [27]. Thus, inclusion of the QBO in the support vector regression model requires caution and needs further detailed investigation.

Different methods for variable selection can also be explored. The current correlation analysis ranks explanatory variables based on the correlation with observed TCs and selects the most promising month per predictor based on correlation. However, in the subsequent variable selection the variables with the highest correlation with observed TCs are sometimes not even selected in the final support vector regression model. In addition, for the region where the correlation of explanatory variables with the observed number of TCs was highest (AR), the final model utilises September values for all selected explanatory variables. For the region where the correlation was not that high (SPO),

the final model also uses values from earlier months. For example, in the support vector regression model for the SPO the three month average over July, August, and September is used for Niño1+2 SST anomalies and the August value for Niño4 SST anomalies and DMI. From a methodological perspective, it could be decided to use September values for all explanatory variables instead.

Our analysis gives some indications that errors could increase for predictions further into the future, although a larger sample size is required to obtain a conclusive answer. A prudent approach would be to update the model annually. In case annually updating the model is not possible, an approach researched by Nicholls [13] can be investigated. He suggests that predicting the change in TC numbers from last season to the upcoming season, rather than predicting the expected numbers directly from the explanatory variables, could reduce the confounding effect of possible secular changes in TC numbers, explanatory variables, or of relationships between them. A recent study by Dowdy and Kuleshov [28] confirms there has been a significant downward trend in the number of TCs in the AR over the 32-year period 1981-82 to 2011-12. Using the approach suggested by Nicholls [13] the forecasting errors for later years could possibly be reduced.

As the statistical models for AR give better results than for SPO, different explanatory variables for SPO could be explored in future research. Current explanatory variables used for SPO have a low correlation with observed TC activity. The consideration of new variables may increase model performance for SPO. An alternative approach to seasonal prediction of TCs is using coupled ocean-atmosphere dynamical climate model [10, 11, 29]. At the Australian Bureau of Meteorology, the Predictive Ocean Atmosphere Model for Australia (POAMA) is currently used operationally for preparing seasonal climate outlooks for AR. Superior skill of POAMA compared to statistic model for predicting seasonal rainfall in AR and SPO has been demonstrated [30]. A pilot study to explore POAMA-based methodology to predict TC seasonal activity in AR and SPO conducted under the Pacific Australia Climate Change Science and Adaptation Planning program (PACCSAP) demonstrated potential skill of the model to improve accuracy of TC forecasting comparing with current operational model [31]. These two avenues—improving statistical model-based methodology and developing new dynamical climate model-based methodology—will be further explored in our future research with the aim to improve skill of operational seasonal TC prediction in AR and SPO.

5. Summary

For AR a support vector regression model outperforms the current Bureau of Meteorology methodology based on linear discriminant analysis. In cross-validation results support vector regression reduces the standard deviation of the errors from 2.874 to 2.266. This methodology can be used to improve the accuracy of current TC predictions for AR. Similarly, the current linear discriminant analysis model has limited capability in accurately predicting SPO TC activity.

Although the developed support vector regression model for SPO gives a significant improvement in performance over the current linear discriminant analysis methodology, performance is still quite low and further research is necessary to improve the skill of seasonal TC predictions in SPO.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The Australian Agency for International Development (AusAID) and the Australian Government Department of Climate Change and Energy Efficiency (DCCEE) provided support for this research through the PACCSAP's project "Seasonal prediction of tropical cyclones." Dr. Andrew Watkins from the National Climate Centre, Australian Bureau of Meteorology provided useful comments on an earlier version of the paper.

References

- [1] J. Mottram, *Report on Cyclone Tracy*, prepared by J. Mottram, Australian Government Publishing Service, Canberra, Australia, 1977.
- [2] L. Anderson-Berry, C. Iroi, and A. Rangi, "The environmental and societal impacts of cyclone Zoe and the effectiveness of the tropical cyclone warning systems in Tikopia and Anuta," James Cook University Centre for Disaster Studies, 2003.
- [3] K. L. Shelton and Y. Kuleshov, "Evaluation of statistical predictions of seasonal tropical cyclone activity," *Journal of Climate*. In preparation.
- [4] M. B. Richman and L. M. Leslie, "Adaptive machine learning approaches to seasonal prediction of tropical cyclones," *Procedia Computer Science*, vol. 12, pp. 276–281, 2012.
- [5] Y. Kuleshov, R. Fawcett, L. Qi et al., "Trends in tropical cyclones in the South Indian Ocean and the South Pacific Ocean," *Journal of Geophysical Research D*, vol. 115, no. 1, Article ID D01101, 2010.
- [6] V. F. Dvorak, "Tropical cyclone intensity analysis using satellite data," NOAA Tech. Report NESDIS 11, 1984.
- [7] N. Nicholls, L. Landsea, and J. Gill, "Recent trends in Australian region tropical cyclone activity," *Meteorology and Atmospheric Physics*, vol. 65, no. 3-4, pp. 197–205, 1998.
- [8] Y. Kuleshov, F. C. Ming, L. Qi, I. Chouaibou, C. Hoareau, and F. Roux, "Tropical cyclone genesis in the Southern Hemisphere and its relationship with the ENSO," *Annales Geophysicae*, vol. 27, no. 6, pp. 2523–2538, 2009.
- [9] Y. Kuleshov, L. Qi, R. Fawcett, and D. Jones, "Improving preparedness to natural hazards: tropical cyclone prediction for the Southern Hemisphere," in *Advances in Geosciences*, J. Gan, Ed., vol. 12 of *Ocean Science*, pp. 127–143, World Scientific Publishing, Singapore, 2009.
- [10] Y. Kuleshov, Y. Wang, J. Apajee, R. Fawcett, and D. Jones, "Prospects for improving the operational seasonal prediction of tropical cyclone activity in the southern hemisphere," *Atmospheric and Climate Sciences*, vol. 2, no. 3, pp. 298–306, 2012.
- [11] Y. Kuleshov, C. Spillman, Y. Wang et al., "Seasonal prediction of climate extremes for the Pacific: tropical cyclones and extreme

- ocean temperatures,” *Journal of Marine Science and Technology*, vol. 20, no. 6, pp. 675–683, 2012.
- [12] N. Nicholls, “A possible method for predicting seasonal tropical cyclone activity in the Australian region,” *Monthly Weather Review*, vol. 107, no. 9, pp. 1221–1224, 1979.
- [13] N. Nicholls, “Recent performance of a method for forecasting Australian seasonal tropical cyclone activity,” *Australian Meteorological Magazine*, vol. 40, no. 2, pp. 105–110, 1992.
- [14] J. C. L. Chan, “Tropical cyclone activity over the western North Pacific associated with El Niño and La Niña events,” *Journal of Climate*, vol. 13, no. 16, pp. 2960–2972, 2000.
- [15] Y. Kuleshov, L. Qi, R. Fawcett, and D. Jones, “On tropical cyclone activity in the Southern Hemisphere: trends and the ENSO connection,” *Geophysical Research Letters*, vol. 35, no. 14, Article ID L14S08, 2008.
- [16] K. E. Trenberth, “The definition of el Niño,” *Bulletin of the American Meteorological Society*, vol. 78, no. 12, pp. 2771–2777, 1997.
- [17] K. Ashok, S. K. Behera, S. A. Rao, H. Weng, and T. Yamagata, “El Niño Modoki and its possible teleconnection,” *Journal of Geophysical Research C*, vol. 112, no. 11, 2007.
- [18] K. Wolter and M. S. Timlin, “El Niño/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI.ext),” *International Journal of Climatology*, vol. 31, no. 7, pp. 1074–1087, 2011.
- [19] G. J. Holland, “On the quality of the Australian tropical cyclone data base,” *Australian Meteorological Magazine*, vol. 29, pp. 169–181, 1981.
- [20] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The WEKA data mining software: an update,” *ACM SIGKDD Explorations Newsletter*, vol. 11, pp. 10–18, 2009.
- [21] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, 2000.
- [22] A. J. Smola and B. Schölkopf, “A tutorial on support vector regression,” *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.
- [23] H. Ramsay, M. Richman, and L. Leslie, “Exploring variants of ENSO Modoki for the seasonal prediction of Coral Sea tropical cyclone activity,” Submitted to *Journal of Climate*.
- [24] A. Ben-Hur, C. S. Ong, S. Sonnenburg, B. Schölkopf, and G. Rätsch, “Support vector machines and kernels for computational biology,” *PLOS Computational Biology*, vol. 4, no. 10, Article ID e1000173, 2008.
- [25] S. K. Shevade, S. S. Keerthi, C. Bhattacharyya, and K. R. K. Murthy, “Improvements to the SMO algorithm for SVM regression,” *IEEE Transactions on Neural Networks*, vol. 11, no. 5, pp. 1188–1193, 2000.
- [26] Climate Prediction Center—Noaa, “30 mb zonal wind index—CDAS,” 2013, <ftp://ftp.cpc.ncep.noaa.gov/wd52dg/data/indices/qbo.u30.index>.
- [27] S. J. Camargo and A. H. Sobel, “Revisiting the influence of the quasi-biennial oscillation on tropical cyclone activity,” *Journal of Climate*, vol. 23, no. 21, pp. 5810–5825, 2010.
- [28] A. J. Dowdy and Y. Kuleshov, “An analysis of tropical cyclone occurrence in the Southern Hemisphere derived from a new satellite-era dataset,” *International Journal of Remote Sensing*, vol. 23, no. 10, pp. 7382–7397, 2012.
- [29] A. Charles, Y. Kuleshov, and D. Jones, “Managing climate risk with seasonal forecasts,” in *Book Management—Current Issues and Challenges*, chapter 23, pp. 557–584, InTech, 2012.
- [30] A. Cottrill, H. Hendon, E. -P. Lim et al., “Seasonal forecasting in the Pacific using the coupled model POAMA-2,” *Weather and Forecasting*, vol. 28, pp. 668–680, 2013.
- [31] K. Shelton, A. Charles, H. Hendon, and Y. Kuleshov, “Dynamical seasonal tropical cyclone prediction for the Australian and South Pacific Regions,” *Journal of Climate*. In preparation.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

