Research Article **Open Access**

J.S. Wijnands, G. Qian, K.L. Shelton, R.J.B. Fawcett, J.C.L. Chan, and Y. Kuleshov*

Seasonal forecasting of tropical cyclone activity in the Australian and the South Pacific Ocean regions

DOI 10.1515/mcwf-2015-0002
Received June 12, 2015; accepted August 17, 2015

Abstract: The Australian Bureau of Meteorology (Bureau) issues operational tropical cyclone (TC) seasonal forecasts for the Australian region (AR) and the South Pacific Ocean (SPO) and subregions therein. The forecasts are issued in October, ahead of the Southern Hemisphere TC season (November to April). Improvement of operational TC seasonal forecasts can lead to more accurate warnings for coastal communities to prepare for TC hazards. This study investigates the use of support vector regression (SVR) models, exploring new explanatory variables and non-linear relationships between them, the use of model averaging, and lastly the integration of forecast intervals based on a bias-corrected and accelerated non-parametric bootstrap. Hindcasting analyses show that the SVR model outperforms several benchmark methods. Analysis of the generated models shows that the Dipole Mode Index, 5VAR index and the Southern Oscillation Index are the most frequently selected as explanatory variables for TC seasonal forecasting in all regions. The usage of ENSO-related covariates implies that definitions of regions and subregions may have to be updated to achieve optimal forecasting performance. Overall, the new SVR methodology is an improvement over the current linear discriminant analysis models and has the potential to increase accuracy of TC seasonal forecasts in the AR and SPO.

Keywords: tropical cyclones, seasonal forecasting, support vector regression, forecast interval, Australian region, South Pacific Ocean

MSC: 62G08, 62G15, 62F40, 86A10

1 Introduction

Tropical cyclones (TCs) are the most destructive weather systems that form over warm waters of the Pacific, Indian and North Atlantic Oceans. On average, about 87 TCs form each year [17] with about twice as many TCs occurring in the Northern Hemisphere (NH) as compared to the Southern Hemisphere (SH). TC activity has also been observed to distribute unevenly over various regions. In the NH, TC activity is highest in the western

***Corresponding Author: Y. Kuleshov:** School of Mathematics and Statistics, The University of Melbourne, Parkville VIC, Australia
and Australian Bureau of Meteorology, Docklands VIC, Australia
and School of Mathematical and Geospatial Sciences, Royal Melbourne Institute of Technology (RMIT) University, Melbourne VIC, Australia
and Faculty of Sciences, Engineering and Technology, Swinburne University of Technology, Melbourne VIC, Australia, E-mail: y.kuleshov@bom.gov.au
J.S. Wijnands, G. Qian: School of Mathematics and Statistics, The University of Melbourne, Parkville VIC, Australia
K.L. Shelton: JBA Consulting, Skipton, United Kingdom
R.J.B. Fawcett: Australian Bureau of Meteorology, Docklands VIC, Australia
J.C.L. Chan: School of Energy and Environment, City University of Hong Kong, PRC

North Pacific (32 TCs/year on average) [17]. In the SH, the Australian region (AR) has the highest TC activity, with about 11 TCs/year forming or entering the area annually [11]. Historically TCs have had major impacts on shipping and coastal infrastructure, agriculture and the safety and economic wellbeing of communities if they approach a coast and make landfall. The main hazards associated with TCs are strong winds, heavy rain, storm surges, flooding, landslides and tornadoes [31].

International research efforts over recent decades have contributed substantially to our collective understanding of TCs, highlighting strong links to the El Niño-Southern Oscillation (ENSO) and some trends in intensity and distribution [e.g., 4, 12, 15, 16, 28, 45, 47]. [7] conclude that the neutral ENSO phase also contains valuable information with respect to TC activity in the AR, which is uncovered by a split in negative-neutral and positive-neutral regimes. One issue emerging from previous research is the difficulty in forecasting seasonal TC activity for seasons where the predictors employed move outside their historical ranges [25]. For example, sea-surface temperatures (SSTs) now frequently move outside their historical ranges, rendering them problematic as predictors for forecasting TC activity in recent years [26].

TC seasonal forecasting is a key element of an early warning system to increase preparedness of coastal communities ahead of the TC season, and such a service is provided by several organisations around the world (an overview of these organisations is provided in [5]). For the AR, [34, 35] documented the relationship between number of TCs and ENSO indices and developed a seasonal forecast methodology based on this relationship [36]. In addition to count forecasting, [42] investigated the landfall potentiality at the moment of cyclogenesis. Currently, the Australian Bureau of Meteorology (henceforth, the Bureau) provides TC count seasonal outlooks for the AR and the South Pacific Ocean (SPO). The operational statistical model of the Bureau consists of two linear discriminant analysis (LDA) models, one based on the Southern Oscillation Index (SOI) and the other based on the Niño3.4-region SSTs. Although these models are used operationally to generate probabilistic forecasts, deterministic forecasts will be used in this study for comparison purposes.

The 2010/2011 TC season saw one of the strongest La Niña events on record develop over the Pacific. Typically, during La Niña years TC activity is high in the AR [23]. Based on historical relationships between the number of TCs in the AR and ENSO indices, the Bureau in its 2010/2011 Australian TC seasonal outlook predicted that

> *For the full Australian region, there is a high degree of confidence (98%) that the total number of tropical cyclones will be above average. The forecast values from the two models (20–22) are significantly higher than the long-term average value of 12.*[i]

Similarly, the Guy Carpenter Asia-Pacific Climate Impact Centre (GCACIC) at the City University of Hong Kong forecasted that 19 TCs would either develop within or move into the AR in the 2010/2011 TC season[ii]. In fact, the actual TC activity during that season was below average, with 10 TCs recorded instead of the 19–22 forecasted by the Bureau and the GCACIC. Although tropical depressions are not included in TC numbers and are not in scope of this study, it should be noted that the actual number of tropical depressions during this season was well above average. This example shows that there is potential for improvement of current statistical methods for TC seasonal forecasting.

Other examples exist for different basins. The Laboratory for Atmospheric Research at the City University of Hong Kong has been issuing real-time predictions of the annual number of TCs affecting the western North Pacific since 2000. Beginning in 2008, these forecasts have been issued by the GCACIC [18]. However, it was found that the TC activity over the western North Pacific has a significant decreasing trend in recent years with respect to frequency (while TC intensity is increasing) [21]. The GCACIC's prediction scheme, which was first developed in 1997, does not incorporate this trend and therefore overestimated the TC activity during the last few years. As a result, the GCACIC's prediction scheme is currently under revision and the operational forecasts for the TC activity over the western North Pacific have not been issued since 2011[iii]. For the AR, [11] reports a statistically significant decreasing trend in TC numbers in the 1981/1982 to 2012/2013 seasons after removing ENSO-related variability. This study also indicates a potential temporal change in the ENSO impact on TC numbers, with lower impact of ENSO on TC activity in recent years.

Recently we conducted pilot studies [25, 48] investigating the prospects to improve the accuracy of statistical seasonal TC forecasting in the AR and the SPO. Similarly to the western North Pacific, TC activity over

the AR has shown a decreasing trend in recent years (as noted above). Incorporating the trend, our study first demonstrated an improvement in the predictability of the linear regression model [25]. We then showed that a new methodology based on support vector regression (SVR) can improve upon the current LDA methodology for both the AR and the SPO [48]. Our results are consistent with the findings of earlier studies by [41] and [40] that investigated the application of machine-learning algorithms for seasonal prediction of TCs, concluding that SVR leads to better results than linear regression models. In the current study, we continue this lead by investigating further applications of this machine learning algorithm to improve the accuracy of TC seasonal forecasting. This includes non-linear relationships between covariates and the construction of forecast intervals.

The paper is structured as follows. Section 2 describes the data that are used and the methodologies we applied in this study. In Section 3 the performance of the new SVR methodology is evaluated in cross-validation and a hindcast analysis. Section 4 contains a discussion and a summary is presented in Section 5.

2 Data and methodology

2.1 TC data and regions under investigation

The Bureau, together with its international partners, has created a homogenised TC archive for the SH [24] which can be accessed through the Bureau's Pacific Tropical Cyclone Data Portal[iv]. This archive is used in this study. It contains best track data and intensity information for all TCs occurring in the SH for the 1969/1970 to 2012/2013 seasons. As part of the analysis related to this study, we performed additional data verification on the SH TC archive, in which data from different sources have been compared in detail and updated to improve the data quality for the investigated regions. In the data verification phase, we first identified for every TC whether at least one point of its track occurred in a certain region of interest. If this was the case and the cyclonic system attained a central pressure less than or equal to 995 hPa [25] while in this region, it was counted towards the number of TCs in that specific region. This approach differs somewhat from some of our earlier work [24] where the lifetime minimum central pressure was used to locate the TC in space.

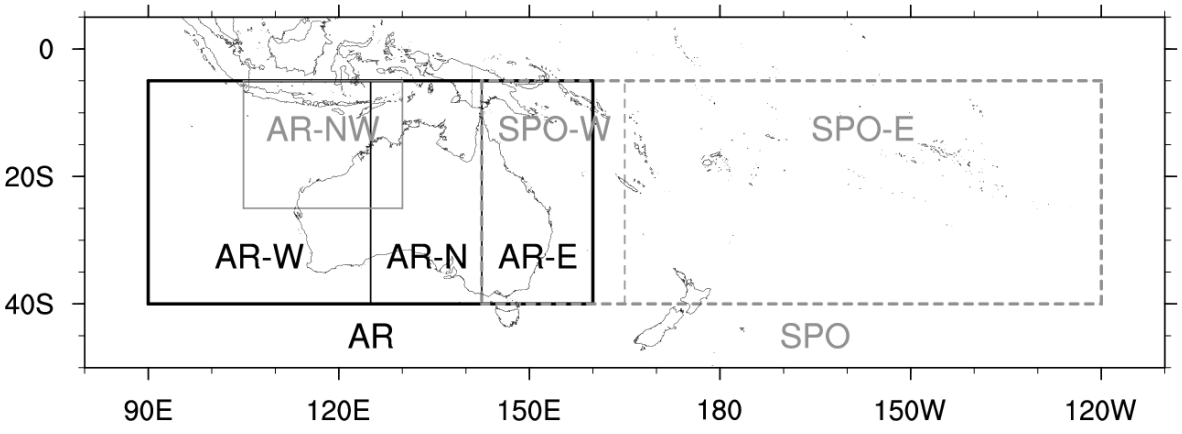

Figure 1: TC regions and subregions as defined by the Bureau for operational TC seasonal forecasting.

The regions and subregions defined by the Bureau for preparing its operational TC seasonal outlooks are shown in Figure 1. All regions have latitudes 5°S and 40°S as their northern and southern boundaries respectively except for the AR-NW subregion which is defined as the area 5°S–25°S and 105°E–130°E. The AR is defined as 90°E–160°E and the SPO as 142.5°E–120°W. Australian subregions are defined as 90°E–125°E (AR-W), 125°E–142.5°E (AR-N) and 142.5°E–160°E (AR-E). Subregions in the SPO are defined as 142.5°E–165°E

(SPO-W) and 165°E–120°W (SPO-E), and so we note that the AR-E region is contained entirely within the SPO-W region. These are the main regions that are investigated in this study.

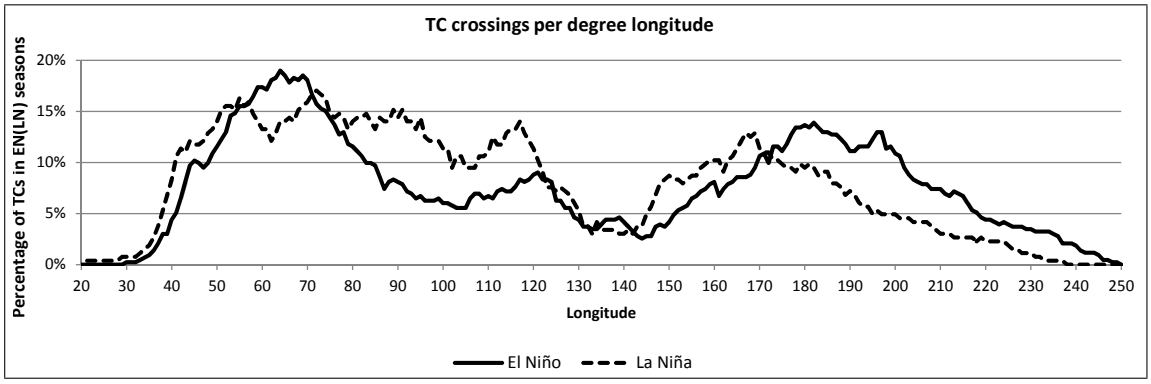

Figure 2: Crossings per degree of longitude expressed as a percentage of total TC numbers in EN (solid line) and LN (dashed line) seasons in the SH.

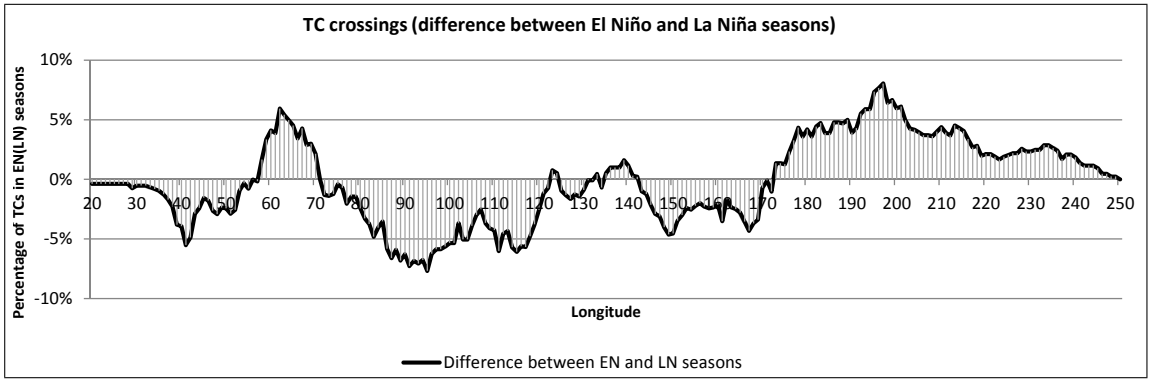

Figure 3: Difference in TC crossings per degree of longitude between EN and LN seasons. Positive (*negative*) values mean more TC crossings in EN (*LN*) seasons and fewer crossings in LN (*EN*) seasons.

2.2 ENSO connection

The plausibility of using ENSO-related covariates for TC seasonal forecasting for a specific region depends on the impact of ENSO on the spatial distribution of TCs. To explore this, ENSO phase designations have been obtained after consideration of several (combined) atmospheric and oceanic indices [23]. In this study, SH TC seasons characterised predominantly by an El Niño (EN) event are 1972/1973, 1976/1977, 1977/1978, 1979/1980, 1982/1983, 1986/1987, 1987/1988, 1991/1992, 1992/1993, 1993/1994, 1994/1995, 1997/1998, 2002/2003, 2004/2005, 2006/2007 and 2009/2010, while seasons 1970/1971, 1973/1974, 1974/1975, 1975/1976, 1988/1989, 1998/1999, 1999/2000, 2007/2008, 2010/2011 and 2011/2012 are defined as characterised by predominantly a La Niña (LN) event. The remaining seasons between 1970/1971 and 2012/2013 are classified as neutral. The ENSO-related variability in the geographical distribution of SH TCs in terms of longitude-crossing is analysed in Figure 2 and Figure 3, indicating which parts of the SH show potential predictability using ENSO-related indices. One can perform a simple analysis to identify regions in the SIO and SPO with high (*low*) expected

forecasting skill, for example 80°E–120°E and 140°E–170°E (*125°E–140°E*). Note that the AR-N subregion (5°S–40°S, 125°E–142.5°E) shows no significant difference in TC occurrence between EN and LN seasons.

2.3 Covariates used for TC seasonal forecasting

Most of the covariates used in this study to predict the number of seasonal TCs are ENSO-related. The Niño1+2 (hereafter, N12), Niño3 (N3), Niño3.4 (N34) and Niño4 (N4) SST anomalies, the El Niño Modoki Index (EMI) and the Dipole Mode Index (DMI) measure SST anomalies in various parts of the equatorial Pacific and Indian Oceans. Additionally, ENSO-event-related atmospheric pressure differences, expressed in terms of the SOI, have been included. Lastly, two indices summarizing multiple types of observations have been included. These are the Multivariate ENSO Index (MEI) and the Bureau-internal 5VAR index. The MEI is based on sea-level pressures, zonal and meridional components of the surface wind, SSTs, surface air temperatures and total cloudiness fraction of the sky [50]. The 5VAR index is based on the first principal component of monthly Darwin mean sea level pressure (MSLP), Tahiti MSLP and the N3, N34 and N4 indices [23].

N12, N3, N34 and N4 data have been obtained from the Climate Prediction Center (CPC) of the US National Oceanic and Atmospheric Administration (NOAA)[v]. MEI data have also been obtained from NOAA[vi]. EMI[vii] and DMI[viii] data have been obtained from the Japan Agency for Marine-Earth Science and Technology (JAMSTEC). SOI data have been obtained from the Bureau[ix].

Although data for many covariates are available since the early 20th century or before, we are limited by the availability of the TC data. TC data are limited for several reasons. First, the quality of TC observations has improved over time and in particular, data in the AR are considered significantly more reliable after meteorological satellites came into operation in the late 1960s [20]. An extensive discussion on the quality of TC data in the SH is provided in [24]. Second, the observed number of TCs in the 1969/1970 season has been identified as an outlier in several models by [25]. Therefore, only TC data from the 1970/1971 season onwards have been used, and in consequence there is no need to use the earlier covariate data. At the time of our analysis, the best track database for both the AR and SPO included the 2012/2013 season. The 2012/2013 season was excluded from our initial analyses as it was used for a separate recalibration analysis (not presented). Therefore, data from the seasons between 1970/1971 and 2011/2012 were used for modelling purposes.

2.4 Statistical models investigated

When constructing a mathematical model for a particular problem, many different approaches are possible. For example, [9] use a Poisson regression approach to TC count forecasting. Our approach started with the application of various machine learning algorithms to the SH TC data set in a pilot study [48], to identify which machine learning algorithms are suitable for the problem of TC forecasting in the AR and SPO. This identified SVR as a promising algorithm to improve operational forecasting performance. Since the ε-insensitive loss function (see Section 2.5) used in SVR incorporates the impact of response's outliers on the prediction process, whether the response variable in SVR is continuous numerical or discrete numerical, or whether or not the response has a normal distribution, does not have significant impact on the prediction results. Thus we choose to use SVR with the response being the raw TC counts without any transformation and specific distribution assumption in this paper. Hence our approach is non-parametric and distribution-free. We choose to explore and exploit the non-linear and non-parametric regression structure rather than the probability distribution of the response to improve the prediction performance. SVR has been used successfully by other researchers for weather forecasting, for example for seasonal prediction of winter extreme precipitation over Canada [51] and for storm surge predictions [39]. As the number of available data points for TC seasonal forecasting is low, it is important to use a model having efficient predictive power (i.e. generalisation performance, [2]). It has been shown that SVR can be efficiently specified by only a small number of support vector data points. SVR is also robust against response outliers and is flexible in terms of the regression function as said before. All these properties of SVR make it a suitable methodology for solving the seasonal TC forecasting problem.

Our present study follows the SVR approach by providing an in-depth investigation on how to optimise SVR modelling for TC seasonal forecasting. For example, we investigate the use of different kernels, e.g., radial basis function and polynomials of second and third degree, and demonstrate how they can improve upon the performance of the baseline linear kernel function.

2.5 SVR model setup

The basic idea behind the SVR using ε-insensitive loss function is that the resulting fitting errors at the training data points should be ideally at most ε while the regression surface should be as flat as possible (i.e., total variation in regression coefficients is small) so that the prediction errors at the testing sample can be under control. This idea can be formulated, subject to some tolerance of greater than ε fitting errors, by a convex optimisation problem with the associated constraints balancing the flatness of the regression surface and the deviation from the prescribed tolerance level of fitting errors. This optimisation problem can be solved more easily by its dual formulation involving Lagrange multipliers and quadratic programming. A detailed analysis of the solving process reveals that not all data points are active (i.e., used) in the process and only a subset of the data points, called the support vectors, are used in the computation process to determine the regression surface. The regression surface is not necessarily specified under the linear space of the covariates, but can be specified under the reproducible kernel Hilbert space (RKHS) which makes the regression space very flexible and computationally feasible for high-dimensional covariates.

The following describes the setup of our SVR model. We define the response variable y as the number of TCs in a season, TC. Table 1 summarises the variables that are used in this study. As operational forecasts for the regions in the SH should be issued by the end of October (prior to the beginning of a TC season which starts in November), covariate values up to this month could be used in the analysis. October values for the covariates generally have a higher correlation with the number of TCs than earlier months (not shown). Therefore, October observations preceding the TC season are used for modelling purposes. Operationally, it is further possible to issue preliminary forecasts by using forecast October values for the covariates. A definitive forecast can then be issued when October values become available in early November. Because of the high monthly variability of SOI [44], the average SOI in the months August, September and October is used instead for this covariate in place of the single-monthly October value. All $2^9 - 1 = 511$ possible variable combinations, except the empty model, have been investigated as potential models. Data were standardised before SVR models were fitted. Final predictions that are negative are set to zero.

Table 1: Description of variables.

Variable	Description
y	Number of TCs in a season
x_1	DMI: Dipole Mode Index (October)
x_2	N4: Niño 4 SST anomalies (October)
x_3	N34: Niño 3.4 SST anomalies (October)
x_4	N3: Niño 3 SST anomalies (October)
x_5	N12: Niño 1+2 SST anomalies (October)
x_6	EMI: El Niño Modoki Index (October)
x_7	5VAR index (October)
x_8	MEI: Multivariate ENSO Index (October)
x_9	SOI: Southern Oscillation Index (average Aug., Sep. and Oct.)

The SVR model being used is an extension to ε-support vector regression [46]. A key element of SVR is the choice of the loss function, which determines how prediction errors (as compared to the actual number of

TCs) are treated in the model fitting process. The loss function used in this study is the ε-insensitive loss (see Equation (1)). Namely, the loss equals zero when a residual's absolute value $|r|$ is smaller than ε; otherwise it equals $|r| - \varepsilon$. Data points where the loss function is unequal to zero are called the support vectors and model fitting only depends on these points. Since larger values for ε lead to fewer data points influencing the solution appropriate parameter tuning based on prediction performance is important, especially since the number of observations available for TC seasonal forecasting is low. Furthermore, due to the linear instead of squared relationship in the loss function, SVR as used here for TC seasonal forecasting is less sensitive to response outliers at the support vectors. That is, the ε-insensitive loss generally puts more weight on the support vectors where the resultant fitting errors are greater than ε, but not too much weight on those support vectors whose response values are outliers.

$$V_\varepsilon(r) = \begin{cases} 0 & \text{if } |r| < \varepsilon, \\ |r| - \varepsilon, & \text{otherwise.} \end{cases} \tag{1}$$

The prediction function used for a specific region is given by

$$f(\boldsymbol{x}) = \sum_{m=1}^{M} \beta_m h_m(\boldsymbol{x}) + \beta_0,$$
$$\text{with } \{h_m(\boldsymbol{x})\},\ m = 1,\ 2,\ \ldots,\ M \text{ being a set of basis functions.} \tag{2}$$

This prediction function is linear with respect to coefficients β_m, but can become non-linear in the covariates and non-parametric because of the transformations $h_m(\boldsymbol{x})$. The unknown parameters $\boldsymbol{\beta}$ and β_0 in Equation (2) are estimated by minimising

$$H(\boldsymbol{\beta},\ \beta_0) = \sum_{i=1}^{n} V_\varepsilon\left(y_i - f(\boldsymbol{x_i})\right) + \frac{1}{2C}\sum_{m=1}^{M} \beta_m^2 \text{ with } n \text{ the number of observations.} \tag{3}$$

Minimising $H(\boldsymbol{\beta},\ \beta_0)$ consists of minimising the prediction errors (using the loss function described above) and a penalty term $\frac{1}{2C}\sum_{m=1}^{M}\beta_m^2$ which penalises models that are too complex. The tuning parameter C is used to control the extent of the complexity of the SVR model.

The minimisation problem of Equation (3) can be equivalently formulated using Lagrange multipliers with addition of the Karush-Kuhn-Tucker (KKT) conditions to ensure optimality of the solution (not shown). The reformulated problem can be solved using techniques such as sequential minimal optimization [37]. The solution $\hat{f}(\boldsymbol{x}) = \sum_{m=1}^{M}\hat{\beta_m}h_m(\boldsymbol{x}) + \hat{\beta_0}$ then has the form

$$\hat{f}(\boldsymbol{x}) = \sum_{i=1}^{n} \hat{\alpha}_i K(\boldsymbol{x},\ \boldsymbol{x_i}) \text{ with } K(\boldsymbol{x},\ \boldsymbol{x}') = \sum_{m=1}^{M} h_m(\boldsymbol{x})h_m(\boldsymbol{x}') \text{ and } \hat{\alpha}_i \text{ the fitted Lagrange multipliers.} \tag{4}$$

Hence, the solution of the SVR model only depends on the covariates or the transformation of the covariates through inner products. Several kernel functions are available that provide this transformation of covariates into a higher dimensional feature space. For example, with two covariates a polynomial kernel function $K(\boldsymbol{x},\ \boldsymbol{x}') = \left(\langle \boldsymbol{x},\ \boldsymbol{x}'\rangle + 1\right)^2$ corresponds to transforming the covariates using $h_1(\boldsymbol{x}) = 1,\ h_2(\boldsymbol{x}) = \sqrt{2}x_1,\ h_3(\boldsymbol{x}) = \sqrt{2}x_2,\ h_4(\boldsymbol{x}) = x_1^2,\ h_5(\boldsymbol{x}) = x_2^2$ and $h_6(\boldsymbol{x}) = \sqrt{2}x_1x_2$. For a more detailed mathematical description, see [19].

2.6 Kernel functions

The kernel functions that we investigated in this study are given in Equations (5), (6) and (7).

$$\text{Linear kernel: } K(\boldsymbol{x},\ \boldsymbol{x}') = \langle \boldsymbol{x},\ \boldsymbol{x}'\rangle \tag{5}$$

$$\text{Radial Basis Function kernel: } K(\boldsymbol{x},\ \boldsymbol{x}') = e^{-\sigma\|\boldsymbol{x}-\boldsymbol{x}'\|^2} \tag{6}$$

$$\text{Polynomial kernel: } K(\boldsymbol{x},\ \boldsymbol{x}') = \left(\gamma\langle \boldsymbol{x},\ \boldsymbol{x}'\rangle + \beta\right)^d \tag{7}$$

While the linear kernel only considers linear relationships, non-linear relationships can be modelled using the radial basis function and polynomial kernels. The radial basis function kernel is often used in other SVR studies to investigate non-linear relationships, containing only one extra parameter compared to the linear kernel function. To cover a larger range of non-linear relationships, polynomial kernels of degree 2 and 3 are also considered in this study, which complicates the calibration process.

With a choice for one of these kernel functions, the SVR model is fully defined. Besides the parameters $\boldsymbol{\beta}$ and β_0 (or $\boldsymbol{\alpha}$), the SVR model formulation contains several other parameters, namely ε, C and any parameters used in the chosen kernel function. These other parameters are referred to as hyperparameters. Due to the number of hyperparameters involved, tuning them can be challenging. This is especially the case when a large number of models have to be calibrated. Potential calibration approaches include Grid Search (GS), which is an exhaustive search algorithm and is implemented in several software packages such as LibSVM [8]. However, this algorithm is computationally expensive as it calculates model performance for all possible combinations of hyperparameter values, using a fixed set of values for each hyperparameter. In general, GS is a feasible approach to calibrate an SVR model using a linear kernel function. However, when calibrating many SVR models using a polynomial kernel, a large amount of computing power is required. Note that five hyperparameters have to be calibrated when using the polynomial kernel (C, ε, β, γ and d). To tune the hyperparameters we therefore use a Pattern Search (PS) algorithm, as in [33], which is briefly described below.

First, a transformation of the hyperparameters is required to search the parameter space efficiently. We define $W = \ln(C)$ and $X = \ln(\frac{1}{\varepsilon})$ for all kernels. For the radial basis function kernel we denote $Y = \ln(\sigma)$ and for the polynomial kernel we use $Y = \ln(\gamma)$ and $Z = \ln(\beta)$. For the polynomial kernel we fix the degree d: a separate calibration is performed for polynomial kernels of degree 2 and degree 3 respectively.

With the above notations the PS algorithm for the polynomial kernel randomly picks a column vector $\boldsymbol{c} = (w, x, y, z)^T$ as the initial tuning point, at which the model performance to be detailed below is also evaluated. Next a pattern matrix P determining the local search neighbourhood is defined as

$$P_{\text{polynomial kernel}} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \end{bmatrix}. \tag{8}$$

An initial value δ_0 is chosen for the step size δ and a value δ_{end} is chosen for the stopping criterion. We then iterate over the following steps:

1. Explore the surrounding points defined by $A = \boldsymbol{c}\mathbf{1}^T + \delta P$, with $\mathbf{1}^T$ a row of nine 1's. Determine the model performance for the set of hyperparameters specified by each column of A.
2. If the model performance for any of the surrounding points is better than the pattern centre, use this point as the new centre $\boldsymbol{c}$ and go to (1). If the model performance does not improve, keep $\boldsymbol{c}$ and set $\delta = \frac{\delta}{2}$.
3. If $\delta \geq \delta_{\text{end}}$, return to (1). Otherwise, the PS algorithm ends and $\boldsymbol{c}$ is the local optimum that has been found.

As the PS algorithm finds a local optimum instead of the global optimum, several random starts are required to minimise the probability of ending in a local optimum which is not close to the global optimum. The set of hyperparameters associated with the PS run yielding the best performance is selected.

Values for δ_0, δ_{end} and the number of random starts have been determined using a convergence analysis. The number of random starts is set to 4 and the stopping criterion δ_{end} is set to 2^{-4}. δ_0 is set to 2 for the linear and radial basis function kernels and 4 for the polynomial kernel. Convergence analysis has shown that a large value for δ_0 of 4 is required in the polynomial kernel case to ensure convergence towards the global optimum. This could be due to the larger hyperparameter space for the polynomial kernel, as four different hyperparameters have to be optimised (for a fixed degree d). Note that using a larger δ_0 enables the search algorithm to navigate through the parameter space with larger steps at the start thus decreases the probability of ending up in a bad local optimum. Our research therefore indicates that a higher δ_0 used in the PS algorithm is advantageous when the number of hyperparameters is large.

2.7 Variable selection

To evaluate a particular model we adopt a leave-one-out cross-validation (LOOCV) approach, following [13]. This approach involves treating each season in the data as a fold, and predicting the number of TCs for every fold by using the numbers of TCs and covariates observed in all other folds. The use of LOOCV is not justified when there are significant season-to-season autocorrelations in observed TC numbers. Testing the autocorrelation and partial autocorrelation function for the AR and the SPO (not shown), indicates that there is no significant autocorrelation in our time series. These tests also indicate that a statistical approach based on time series modelling is unlikely to produce better results.

In statistics, the Kullback-Leibler (KL) divergence measures the disparity between the true distribution of the response variable and an estimated one based on the employed model [27]. For the non-parametric SVR model the KL divergence does not have a closed form and is not computable, thus we use the LOOCV mean squared error (MSE) as a *de facto* estimate of the KL distance. Variable selection is based on LOOCV MSE where superfluous covariates are eliminated with high probability.

As we use PS in combination with LOOCV, numerous SVR models are fitted in our analysis. Being conscious of computing power we set the tolerance of the stopping criterion of the KKT conditions to 10^{-2} within the PS algorithm. Although the default setting of this criterion is 10^{-3}, [29] indicates that a loose stopping criterion up to 10^{-1} still yields a good approximation of the LOOCV error, while significantly reducing training time. The 10^{-2} tolerance is only applied to find the optimal hyperparameters. Once optimal hyperparameters for a specific combination of covariates and kernel function have been determined, the LOOCV MSE for these models is recalculated using the default stopping criterion value of 10^{-3}.

After the most promising models have been determined, a model averaging approach is investigated to increase model performance. Model averaging has applications in weather and climate forecasting, where ensemble prediction is widely used to improve prediction skill and assess the confidence level of predictions [e.g., 6, 32]. In the context of this study it is applied as a method for machine learning algorithms [e.g., 10]. In a model-averaging approach, predictions of the top m models are averaged for some suitably chosen m. Analysis presented in Section 3 explores different values for m and indicates that averaging the top four models can increase performance compared to using a single model.

The methodology described above has been implemented using parallel computing in R [38], using a Linux cluster computing system of the School of Mathematics and Statistics at the University of Melbourne. We investigated several libraries including *kernlab* [22] and *e1071*, the latter containing the SVR implementation of the LibSVM package [8]. For our data set, the SVR implementation of the *kernlab* library delivered the best runtime performance. This library has been used for all analyses presented in this paper.

2.8 Forecast intervals

Prediction intervals for forecasts (hereafter, forecast intervals) enhance the applicability of the new SVR methodology to operational forecasting. To construct a forecast interval, the distribution of forecast residuals after model averaging is required. Such a distribution can be approximated by a bootstrap method using samples of forecast residuals obtained by a three-fold cross-validation on training data for each of the top four models (extra residuals could be generated by repeating this process). Three bootstrap methods were compared: a parametric bootstrap, a percentile-based non-parametric bootstrap and a bias-corrected and accelerated (BCa) non-parametric bootstrap [14]. More detail is provided in Appendix A. To incorporate model averaging, bootstrap statistics obtained from the respective four bootstrap samples are further averaged to generate an overall bootstrap statistic. From the replicates of the overall bootstrap statistic for each modelling method and region/subregion, a forecast interval can be constructed with a pre-specified confidence level. Performance of the different bootstrap methods was evaluated by comparing nominal and achieved confidence levels with a general preference for small forecast intervals. As the BCa non-parametric bootstrap performed best, this method has been chosen to construct forecast intervals. Forecast intervals can be used operationally as the forecast format, as they are more meaningful than a point prediction. Lastly, forecast

intervals have been rounded to the nearest integer. The impact of rounding confidence intervals is that the confidence level will differ from the pre-specified nominal confidence level. However, on average the forecast intervals will still be close to the pre-specified confidence level given the numerical rounding procedure that is used.

2.9 Model performance evaluation

Some of the results in Section 3 are based on the LOOCV methodology described above. In operational practice, a forecast for the next season can be obtained by using data only up to the current season. Therefore, the forecasting performance of a model should also be assessed based on forecasting for future seasons, in addition to the LOOCV forecasting assessment. Since LOOCV is not purely used for model evaluation in this study but is used for model selection as well, hindcasting will provide a more accurate assessment of actual forecasting performance. Hence we perform hindcasting to resemble future forecasting, i.e., using the data up to a certain season in the past to forecast the next season. With this idea we use data from seasons 1970/1971 to 2002/2003 (33 seasons in total) to calibrate the model and forecast TC activity in the 2003/2004 season. Then, we use data from seasons 1970/1971 to 2003/2004 for model calibration and forecast TC activity in the 2004/2005 season, and so forth until the 2012/2013 forecast. Except for using a shorter period for model fitting (and the use of *all* seasons within this period rather than the *all-but-one* seasons of the LOOCV process), the settings in our hindcast analysis are exactly the same as described before. This means the selection of the top four SVR models is performed using data of the training period only: these selected models are then used to hindcast TC activity in the next season. Note that LDA hindcasts have been determined based on the same (updated) data set as the SVR method and could therefore deviate from issued forecasts. These hindcasts are then compared against observed TC numbers.

Several statistics are used to measure model performance throughout this paper, including normalised root mean squared error (nRMSE) and mean absolute error (MAE). In addition, the forecast skill S of a forecasting scheme over climatology as defined by [49, p. 276-282] has been used. Definitions for these statistics are given below.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}(\boldsymbol{x_i})\right)^2 \tag{9}$$

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} \tag{10}$$

$$\mathrm{nRMSE} = \frac{\sqrt{\mathrm{MSE}}}{\sigma_y} \tag{11}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{f}(\boldsymbol{x_i})\right| \tag{12}$$

$$S = \left(1 - \frac{\mathrm{RMSE}_{\mathrm{scheme}}}{\mathrm{RMSE}_{\mathrm{climatology}}}\right) * 100\% \tag{13}$$

3 Results

3.1 Model performance for 1970/1971 to 2011/2012 data

This Section describes the performance of the SVR and LDA models for the 1970/1971 to 2011/2012 training data. First, nRMSE results for averaging the SVR models are described (see Table 2). Here we demonstrate that using model averaging improves the accuracy of the forecasts. For all regions, model averaging yields better results than using the best model by itself. Analysing the results in Table 2, it appears that averaging the top two to four models works well for most regions. It was decided to use the average prediction of the best four SVR models for all regions.

From Table 2 it also appears that the AR is the region with the best model performance (as measured by nRMSE), followed by the western subregion AR-W. Performance for the AR-NW and AR-N subregions is considerably poorer. Consistent with previous studies [e.g., 25], the AR shows better predictability using ENSO-related covariates than the SPO. Performance in the different subregions of the AR varies, while performance for the SPO and its subregions is quite similar.

Table 2: nRMSE for various options of model averaging. The smallest (i.e., best) nRMSE for each region is underlined. Grey shading indicates a deviation of less than 0.01 from the best result.

	AR	AR-W	AR-NW	AR-N	AR-E	SPO	SPO-W	SPO-E
Best model	0.540	0.568	0.804	0.791	0.736	0.744	0.627	0.660
Top two	0.493	0.556	0.798	0.764	0.722	0.660	0.623	0.622
Top three	0.501	0.571	0.789	0.757	0.700	0.659	0.613	0.628
Top four	0.474	0.563	0.797	0.756	0.703	0.666	0.615	0.635
Top five	0.478	0.565	0.798	0.773	0.706	0.648	0.623	0.643
Top six	0.482	0.571	0.798	0.773	0.706	0.629	0.632	0.643
Top seven	0.484	0.574	0.796	0.783	0.703	0.629	0.639	0.650
Top eight	0.487	0.568	0.799	0.788	0.702	0.628	0.641	0.648
Top nine	0.488	0.573	0.800	0.794	0.700	0.631	0.642	0.628
Top ten	0.488	0.566	0.800	0.799	0.696	0.634	0.641	0.635

Performance of the selected SVR models was also compared to the performance of the LDA models. Overall, the SVR models perform better. In order to explain the increase in performance of the SVR models compared to the LDA models, results are also benchmarked against a linear regression (LR) model for each region. We only present results for the main regions (the AR and the SPO). The LR model uses the same covariates as the LDA approach, namely N34 and the SOI. Hence, the LR approach was not optimised using a variable selection process. It is only provided in this study as a benchmark. Since the SVR methods are based on an optimisation routine applied to multiple covariates and the four best models are selected from all of the models produced, these models should perform better on average than a single LR or LDA method.

LOOCV forecasted numbers of TCs using the different modelling approaches versus actual TC numbers for the AR and SPO are shown in Figure 4 and Figure 5, respectively. For every fold the number of TCs in one specific season is forecasted, using data from all other seasons for model development. The solid black line represents the actual number of TCs in each season. Predictions of the top four SVR models in the LOOCV are averaged to obtain a single forecast for each season.

In general, for the AR (Figure 4), the forecasts of the different models follow a similar pattern as the actual number of TCs. The LR and LDA approaches have similar forecast performance. For seasons with high TC activity the SVR model usually performs better. For example, the 1983/1984 season showed a peak in TC activity with 19 TCs, the highest on record during the analysed period. The SVR model forecasts about 14 TCs, while the LR and LDA models are less successful for this season with a forecast close to 12 TCs. The SVR model does not always outperform the LDA model. For example, in the 1987/1988 season the LDA model is closest to the actual number of 5 TCs. Overall, the SVR model outperforms the LDA model in 79 percent of seasons. It is also important that the SVR method is able to capture the lower TC activity in recent years. From the 2006/2007 season onwards there is a clear distinction in performance between LR and LDA models and the SVR approach. For all these seasons, the SVR approach outperforms both LR and LDA. For the 2010/2011 season particularly the difference is large. In this season LDA shows a large overprediction of 18 TCs in LOOCV, while the actual number of TCs was 10 (as mentioned in Section 1, overprediction was also experienced in the operational forecasts). SVR model performance for the AR in recent seasons shows this methodology has the potential to overcome the poor forecasts of the LDA model in recent years.

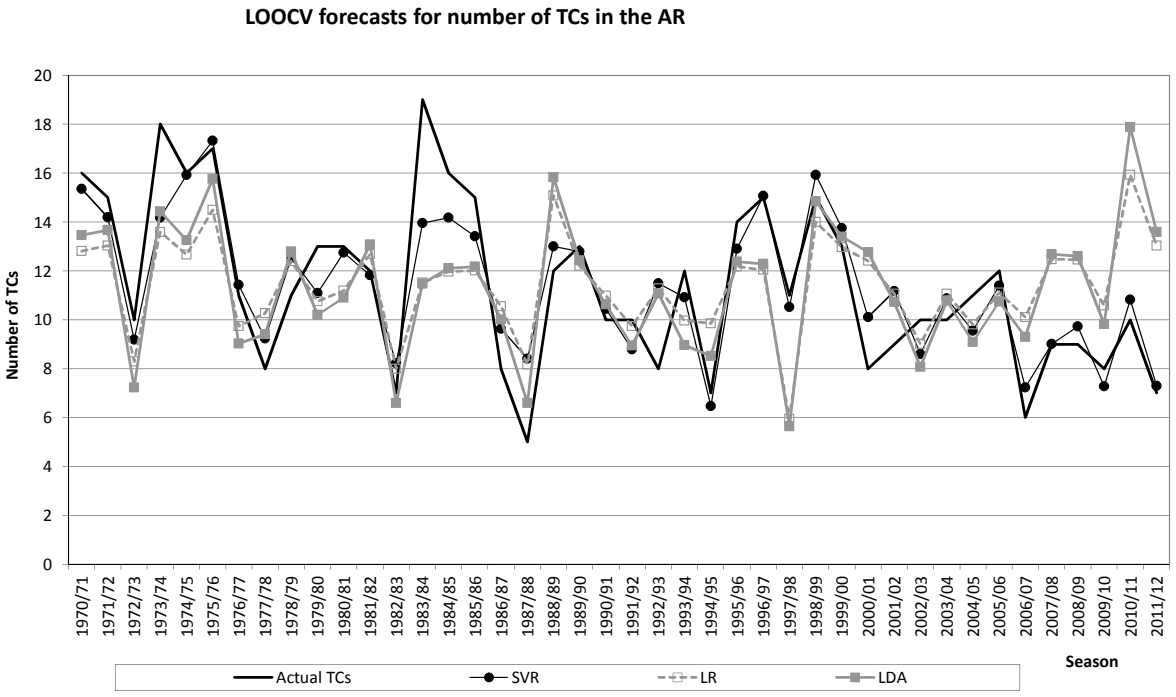

Figure 4: Cross-validated forecasts of *TC* for the AR.

For the SPO (Figure 5), it appears that all forecasts are less accurate in comparison to the AR. This was expected from the nRMSE results presented in Table 2. More specifically, it shows that the models have more difficulty capturing extremes. For example, the TC seasons 1971/1972, 1990/1991 and 2003/2004, having either very low or high TC activity, are not captured very well. The peak in the time series at the 1997/1998 season (17 TCs) is forecasted accurately by the SVR model, while the LR and LDA models forecast a season with average TC activity. As for the AR, the LR and LDA forecasts are quite close together. While the actual number of TCs reaches a minimum of 3 and a maximum of 17 TCs, the forecasts using LDA are mostly (and unhelpfully) between 8 and 10 TCs. For the 2011/2012 season the SVR approach is able to forecast the low activity with reasonable accuracy. This shows that the SVR approach also has the potential to overcome less skilful forecasts from the LDA model in the SPO in recent years. Overall, the SVR model outperforms the LDA model in 74 percent of seasons.

Table 3 and Table 4 show the nRMSE, MAE and the forecast skill over climatology (S) in LOOCV for the investigated models in the main regions (the AR and SPO). We have included the SVR, LR and LDA models and a forecast based on climatology. The forecast based on climatology is a prediction for each season based on the average observed number of TCs in the training set. To capture the effect of LOOCV, this average is calculated by excluding the actual number of TCs in the year that is forecasted. For the AR, the climatology forecast varies around 11.4 TCs per season, while for the SPO the forecasts are close to 9.4 TCs.

Table 3: Goodness-of-fit statistics for the AR under LOOCV. For each metric, the best result is underlined.

	SVR	LR	LDA	Climatology
nRMSE	0.474	0.899	0.916	1.024
MAE	1.201	2.580	2.542	2.893
S	53.8%	12.2%	10.6%	0.0%

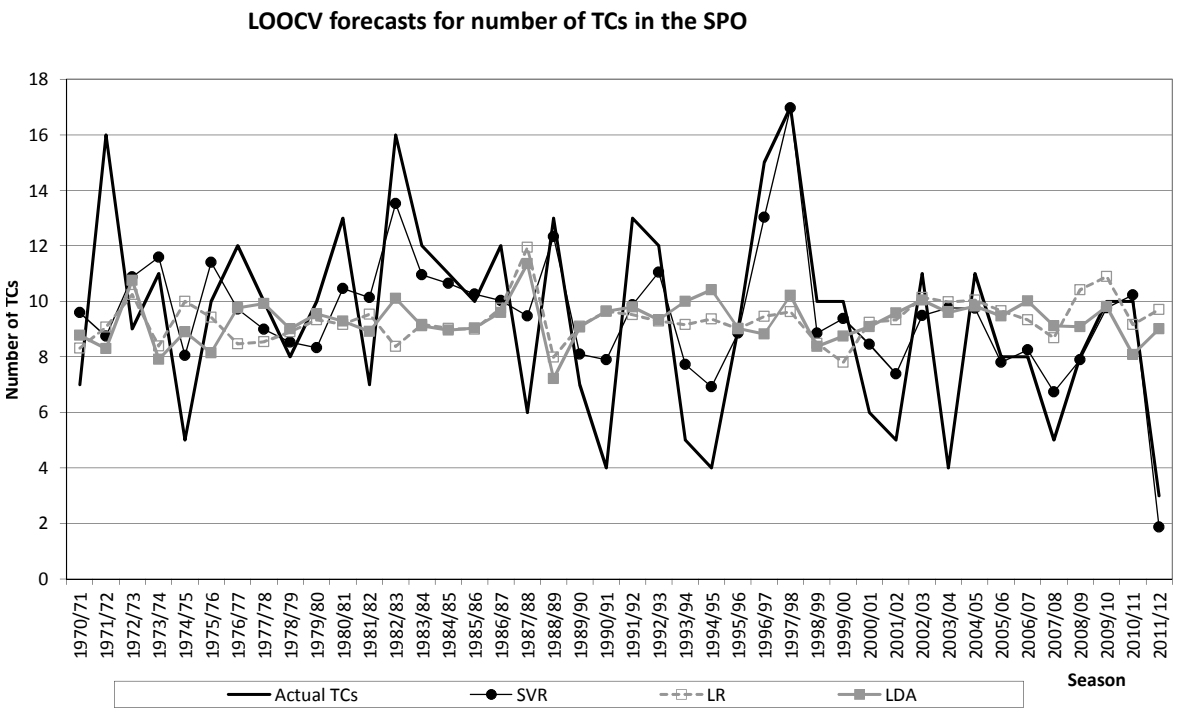

Figure 5: As per Figure 4, but for the SPO.

Table 4: As per Table 3, but for the SPO.

	SVR	LR	LDA	Climatology
nRMSE	0.666	1.074	1.057	1.024
MAE	1.765	3.101	3.044	2.916
S	35.0%	−4.9%	−3.2%	0.0%

These tables show an improvement of the SVR approach compared to the LR and LDA models for all measures in both regions. The SVR model also shows an improvement over climatology, thus pointing to the basic feasibility of seasonal TC number forecasting. The LR and LDA models perform similarly (although the LR model would likely perform better with more appropriate choices for the covariates [40]). For the SPO, LOOCV point predictions of these models are worse than climatology, shown by the negative skill *S*.

We conclude this Section with the following remarks. First, looking at the two main regions under investigation, the AR and the SPO, it appears that skilful seasonal forecasting of TC numbers for the SPO is more difficult than for the AR, as all models have better performance for the AR. Second, based on nRMSE, the subregions AR-NW and AR-N of the AR are more difficult to forecast than the other regions/subregions that were investigated. This could be due to a smaller range of variation in observed TCs which makes approximation using a continuous outcome variable more difficult, or the ENSO-related covariates used may not have strong forecasting power for these regions. Third, the graphs as well as the goodness-of-fit statistics show that the SVR model performs better in forecasting TC numbers than the LR and LDA models. Since LOOCV is also used for the model selection process, additional performance analysis is required.

3.2 Forecast bias

In this Section we perform analyses on the forecast bias of all investigated models for the main regions (the AR and SPO). For each region the actual number of TCs has been plotted versus the forecasted number of TCs. Figure 6 shows the results of this analysis.

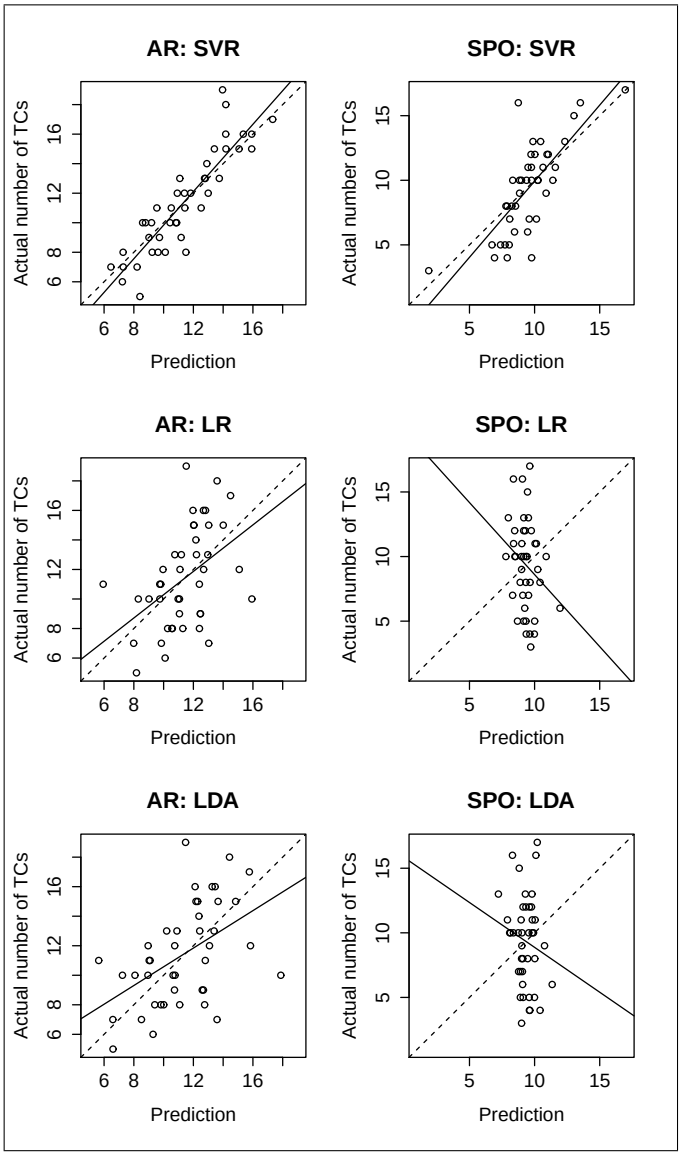

Figure 6: Comparison of observed TC (vertical axis) versus predicted TC (horizontal axis). Solid lines represent the linear regression of observed TC against predicted TC. The dashed line $y = x$ represents the line of perfect calibration.

To investigate the forecast bias, we analyse the range of actual numbers of TCs conditional on the obtained predictions. The solid lines in Figure 6, representing the linear regression of observed TC against predicted TC, show whether the actual numbers of TCs tend to be lower, equal or higher than the forecast. In practice, slight deviations will occur, but for an adequate model the regression line is expected to be close to the line $y = x$. If the solid lines are not close to the dashed line, this indicates there is a prediction bias. As TC impact is likely to increase when TC activity in the regions is very high, forecast bias in the top right corner of

these graphs is especially important and should be particularly taken into account if the models are used operationally.

For the SVR model the regression lines are close to the line $y = x$ for both regions, indicating there is no substantial bias. For the AR, when moving from the SVR model to the LR model and finally to the LDA model, a slight increase in forecast bias can be observed for each step. It should be noted that for both the LR and LDA models the highest predictions for the AR are associated with actual average TC activity. The SVR model shows a forecast bias that indicates that when a prediction is high, the corresponding actual number of TCs is even slightly higher. With respect to issuing pre-seasonal warnings, this bias is preferred over a bias in the opposite direction, as the opposite direction would yield a higher risk of incorrect TC warnings being issued. For the SPO, the forecast bias for the SVR model is limited, but the forecast bias for the LR and LDA models is large. When forecasts of the LDA model are above average (e.g., larger than 10 TCs), the actual numbers of TCs tend to be lower. This effect is even more pronounced for the LR models. Analysis of the forecasts conditional on the observations show that the expected under/overprediction of the LR and LDA models for the SPO is worse than a forecast based on climatology (not shown).

In summary, in the case when an SVR forecast is high the actual number of TCs is likely to be high as well. This means that high TC forecasts of the SVR model could be used to provide an early warning ahead of an expected active TC season to communities.

3.3 Hindcasting

This Section shows the results of the hindcasting analysis for the 2003/2004 to 2012/2013 seasons. Due to limited available computing power this analysis was performed for a selection of regions: the AR, subregions AR-W and AR-N and the SPO. Forecasts of the SVR, LR and LDA models and observed TC activity are shown in Figure 7. Also included are the forecast intervals for the SVR approach using the BCa bootstrap method. Each panel shows a 90% forecast interval for every point prediction. Due to bias corrections forecast intervals are not always symmetrical around the point prediction.

For the AR, all three methods have similar forecasts for the 2003/2004 to 2005/2006 seasons. Afterwards, forecasts start to move apart, with superior performance of SVR in more recent years. For the AR-W a similar pattern is observed. For the AR-N subregion, none of the methods provides very good hindcasting accuracy. The LDA model forecasts 3 TCs for every season, except for 2006/2007 where the forecast is rounded down to 2 TCs. For the SPO forecasts of the benchmark models are quite stable as well, with LR fluctuating between 9 and 11 TCs and LDA between 8 and 10 TCs. The SVR model forecasts the low TC activity in the 2011/2012 season, which showed a historically low number of 3 TCs which was not present in training data. Overall, SVR results show less improvement over LR and LDA models than in the LOOCV analysis, with most improvement achieved in forecasts for recent years. This could be related to the smaller training data set used for earlier years. In recent years the SVR model outperforms the benchmark models for the AR, AR-W and SPO. Tables 5 and 6 show MAE and nRMSE statistics for the full hindcast period of 2003/2004 to 2012/2013. Except for the AR-N subregion, these statistics show better hindcasting performance of SVR compared to the benchmark models.

Table 5: MAE statistics for the hindcast analyses for 2003/2004 to 2012/2013. The best result for each region is underlined.

	AR	AR-W	AR-N	SPO
SVR	<u>2.40</u>	<u>1.27</u>	0.96	<u>2.48</u>
LR	3.19	1.57	0.90	2.78
LDA	3.35	1.79	<u>0.88</u>	2.59

Table 6: As per Table 5, but showing nRMSE.

	AR	AR-W	AR-N	SPO
SVR	0.93	0.73	0.90	0.83
LR	1.14	0.77	0.82	1.02
LDA	1.25	0.88	0.79	0.94

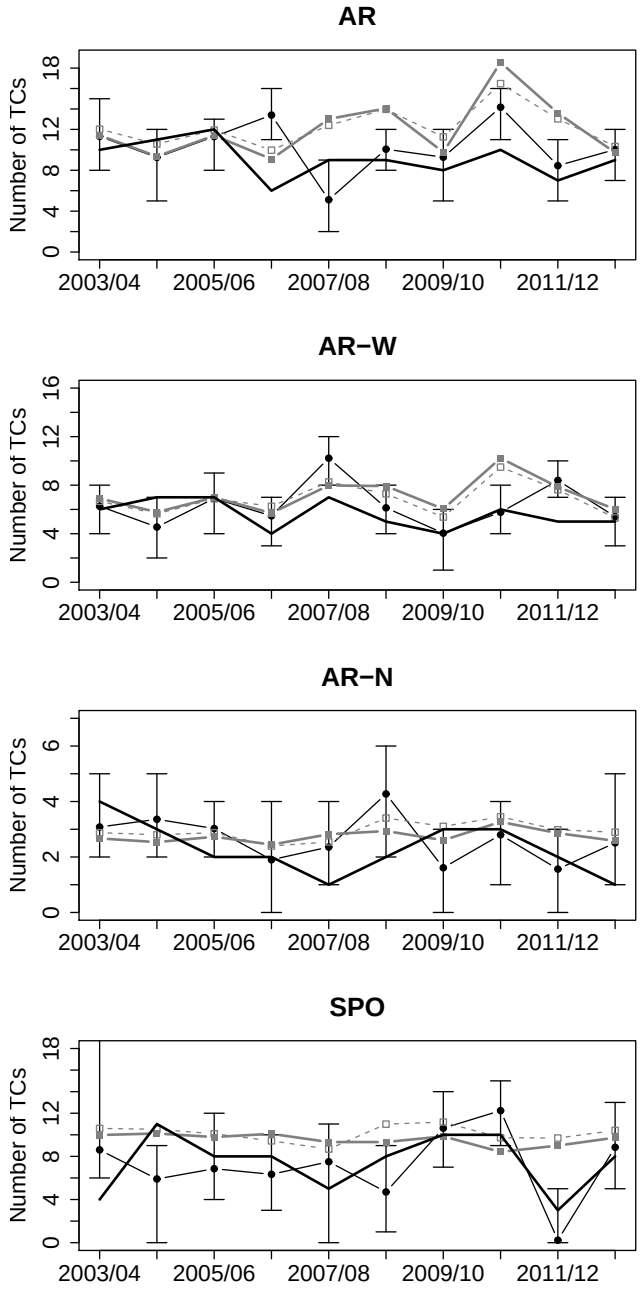

Figure 7: Hindcast analyses for the AR, AR-W, AR-N and SPO for the 2003/2004 to 2012/2013 seasons. SVR forecasts are indicated using solid round markers with integer-valued 90% BCa forecast intervals. Open grey square markers with dashed lines indicate LR forecasts and solid grey square markers indicate LDA forecasts. Thick, black solid lines indicate the actual number of TCs.

Based on Figure 7, accuracy of the generated forecast intervals has been assessed as well. Of all 40 forecasts for the 2003/2004 to 2012/2013 seasons, the actual number of TCs is inside the forecast interval 34 times (i.e. 85%) at a nominal 90% confidence level. Using the methodology to construct forecast intervals as described in Section 2.8, forecast intervals can also be generated for the LR and LDA forecasts. Note that this is for comparison purposes only, as the Bureau does not generate forecast intervals using this methodology. The forecast intervals for the SVR model are narrower than the respective forecast intervals of the LR and LDA models (not shown). Since for the SPO forecasts of the benchmark models are generally close to the average number of TCs, it is expected that the forecast intervals for this region will be wide. For example, for the 2008/2009 season, the generated 90% forecast interval for the LDA forecast for the SPO is [3, 15] TCs.

3.4 Analysis of covariate usage

On average, the best SVR models each use around four covariates. This indicates that the variable selection approach successfully discards the models which are too complex (e.g., using all nine covariates).

Table 7: Occurrence of covariates in the selected top four models.

	AR and subregions	SPO and subregions	All regions
Occurrence of DMI	70%	69%	70%
Occurrence of 5VAR	68%	44%	59%
Occurrence of SOI	61%	48%	56%
Occurrence of MEI	48%	44%	46%
Occurrence of N4	43%	46%	44%
Occurrence of N3	40%	50%	44%
Occurrence of N34	40%	44%	41%
Occurrence of N12	39%	46%	41%
Occurrence of EMI	35%	35%	35%

From Table 7, one can see that the DMI and the 5VAR are the two most frequently selected indices used in the top models. The relationship between ENSO and TC activity in the regions of the SH has been investigated in a number of studies, finding that variation in TC geographical distribution depends on the ENSO phase. [1] and [12] have demonstrated that TCs tend to form more towards the north-east in El Niño years compared with La Niña years in the SPO. The results of [3] demonstrate that warm SSTs, increased relative humidity and reduced vertical wind shear are critical components in determining areas favourable for TC genesis in the SH. Warmer than normal SSTs, positive relative humidity anomalies and reduced vertical wind shear are found in the eastern (*western*) part of the basin during El Niño (*La Niña*) events, helping to explain the observed cyclogenesis. As the 5VAR index captures relative changes in both the atmospheric and oceanic environment favourable for TC genesis over a large domain covering the equatorial central and eastern Pacific regions, it explains the more frequent selection of this index amongst the top models compared with indices describing changes in just atmospheric (e.g., the SOI) or just oceanic (e.g., N34 or N4) conditions over smaller individual regions. Analysing the skill of statistical models for TC seasonal prediction in the Coral Sea basin using multiple predictors (EMI and Trans-Niño Index) which account for the relative warming or cooling of the equatorial central Pacific in contrast with the eastern and western Pacific, [40] arrived at a similar conclusion that models which use multiple predictors demonstrate superior skill compared to the skill of statistical models which use individual SST anomaly indices such as N34 or N4.

Similar to the ENSO, the IOD is a coupled ocean-atmosphere phenomenon in the Indian Ocean. A positive IOD event is characterised by warmer than average SSTs in the western equatorial Indian Ocean and cooler

than average SSTs in the south-eastern equatorial Indian Ocean. Typically, SSTs in the north Australian-Indonesian region and in the Coral Sea are also cooler than average during a positive IOD event. Associated with a westward shift of warmer ocean waters, the area of increased convection also shifts from the eastern to the western equatorial Indian Ocean. These coupled ocean-atmosphere changes during a positive IOD mode bring less favourable conditions for TC genesis to the AR and the western SPO region. On the other hand, during a negative IOD event, warmer than average SSTs over the south-eastern equatorial Indian Ocean, the Australian-Indonesian region and the Coral Sea lead to increased atmospheric convection over these regions and an associated increase in mid-tropospheric relative humidity creating a favourable environment for enhanced tropical cyclogenesis over this area. These relative changes in the oceanic and atmospheric environment associated with negative and positive phases of the IOD which impact on TC activity help explain the frequent selection of the DMI as one of the predictors for the top models.

4 Discussion

A potential drawback of the new SVR model is the risk of overfitting by LOOCV due to the large number of models and hyperparameter combinations investigated. Nonetheless, the top models use an average of only four out of the nine available covariates, which indicates that superfluous covariates are eliminated. Hindcast analyses also show that the SVR approach has skill in independent forecasts for recent seasons, which would not be the case if there was a high degree of overfitting. A limitation of our research is that subregions are modelled independently of the main region. Hence, it can occur that the total of the forecasts for the subregions is not in line with the forecast for the entire region.

Alternative methodologies that could be considered include a generalised SVM where the TC number is treated as following a Poisson or negative binomial distribution. This would require analysis that is beyond the scope of this study.

We have explored various approaches expecting to further increase the forecasting performance. First, we used the transformation $\ln(TC + 0.5)$ in our SVR model as the probability distribution of a log-Poisson response variable is more similar to the Gaussian than the Poisson itself and we thought SVR may perform better for a Gaussian response. Actually the forecasting performance from using this transformation improved slightly for the SPO only and did not change significantly for the other regions. Thus the $\ln(TC+0.5)$ transformation was not adopted in the end. What we found here probably confirmed the robustness property of SVR with a non-Gaussian response. Therefore, we did not investigate other transformations for the TC response in our study. Second, a year-on-year changes approach was investigated. In this approach, as suggested by [36], the change in the number of TCs compared to the previous season is modelled instead of the number of TCs itself. This could reduce the confounding effect of potential secular changes in (non-linear) relationships between TC numbers and the covariates and could be more robust with respect to any downward trend in the number of observed TCs. However, we found that the quality of fit when forecasting year-on-year changes is inferior to the fit achieved when modelling the number of TCs directly. Third, we investigated whether extra information is available mid-season to increase accuracy of issued forecasts by means of a mid-season update. Updated covariate values are not likely to result in better forecasting performance (not shown). Hence, only one covariate is added, describing the number of TCs that have already occurred in the first part of the season. Although results on training data are promising, a hindcast analysis shows no improvement over the issued forecast. The main reason is a difference in temporal distribution of TCs over the season in the training set and recent seasons used for testing. However, the number of data points is too small to conclude that there is a shift of TC activity towards the first half of the season. Fourth, using recalibration analyses it was found that recalibrating the model regularly, preferably annually, can lead to better forecasting performance than using the same model to forecast several seasons. Recalibration may be required because of a potential change in (the relationship between) covariates and/or response variable due to climate variability and change not captured in the training data. Finally, a different definition of subregions in the AR was investigated. This research is motivated by our analyses in Figure 2 and Figure 3, showing that ENSO-related

covariates may not be the appropriate ones for forecasting TCs for the AR-N region. Therefore, we investigated the consequence of dividing the AR into only two subregions; 90°E–135°E, 5°S–40°S and 135°E–160°E, 5°S–40°S (also used by the GCACIC). This split did not result in a significant performance improvement, except for eliminating a subregion with limited performance. Furthermore, Figure 3 indicates that extending the eastern boundary of the AR-E region to 170°E could give better forecasting results. Note that [40] used 170°E as an eastern boundary investigating skill of TC seasonal forecasting over the Coral Sea basin. Future research based on climatological arguments may provide more insights into effective criteria for splitting regions.

5 Summary

Performance of the SVR models has improved considerably compared to the pilot study [48] by exploring polynomial and radial basis function kernels tuned using a Pattern Search algorithm and applying model averaging to the best models. To test forecasting accuracy, LR and LDA models have been used as benchmark. The benchmark models often show similar performance throughout this study. Hindcast analyses have been performed for all investigated models to assess operational performance. These analyses indicate the SVR approach performs particularly well in recent seasons. The SVR model outperforms the benchmark methods in the AR and SPO; however, it does not perform so well for smaller subregions such as the AR-NW and AR-N. Most likely this is due to (i) smaller number of TCs in these subregions, and (ii) a weak relationship between TC variability and large-scale climate processes such as ENSO and IOD. For example, as Figure 3 indicates, there appears to be no strong relationship of the total number of TCs in the AR-N subregion with ENSO. Various covariates have been explored in this study. An analysis of the best models shows that the DMI performs well as a predictor in both the AR and SPO. For the AR, 5VAR and the SOI show similar performance.

The construction of forecast intervals using a non-parametric BCa bootstrap method enables operational forecasting using this methodology. Generated forecast intervals in the hindcast analysis show coverage close to the claimed confidence level. Furthermore, forecast intervals of the SVR approach are generally more narrow than those of the benchmark models, especially for regions where LR and LDA forecasts show little variation. The outcomes of this research are consistent with previous studies using SVR for seasonal TC forecasting [41], indicating superior performance over linear regression methods.

Finally, the SVR approach for the AR and SPO inhibits minimal forecast bias. This means that high TC forecasts from the SVR model could be used to provide an early warning to communities ahead of an expected active TC season.

Appendix A: Bootstrap methods

In this paper we investigated the application of three bootstrap methods:

1. Parametric bootstrap. The parametric bootstrap approach depends on fitting a parametric distribution on a sample of residuals. In the context of using SVR models, [30] suggest the fitting of a Laplace or Gaussian distribution.
2. Percentile-based non-parametric bootstrap. Instead of using a parametric distribution, the confidence interval is determined from the bootstrap distribution itself by sampling from residuals and calculating the required percentiles.
3. Bias-corrected and accelerated (BCa) non-parametric bootstrap. Same as the percentile-based non-parametric bootstrap, but with a correction for a potential bias in the bootstrap statistic, which may not be symmetrically distributed around zero.

Of these methods, the BCa bootstrap uses a pivotal quantity for the bootstrap statistic. In statistics, given X a sample taken from a probability distribution, where θ is a parameter of that distribution, a function $\Re$ of X and

θ is called a pivotal quantity if and only if the distribution of $\Re(X, \theta)$ does not depend on the value of θ [43]. The bootstrap statistic in the parametric bootstrap is not pivotal, as an estimated scale parameter is used in the Laplace/Gaussian distribution for generating bootstrap samples. In other words, the distribution's scale parameter is unknown in this context and has to be estimated from the residuals. Similarly, the bootstrap statistic in the percentile-based non-parametric bootstrap is not pivotal as its distribution depends on the bias that has to be estimated from the generated residuals. In the BCa bootstrap the dependence on θ is removed before estimating a forecast interval and finally the forecast interval is scaled back to the scale of interest, in theory leading to more accurate forecast intervals [14].

Acknowledgement: The authors thank Dr. Andrew J. Dowdy, Dr. Paul A. Gregory and Dr. Andrew B. Watkins for reviewing the manuscript. In addition, the authors acknowledge two anonymous reviewers for their valuable comments. All these reviews helped us to improve the quality of the original manuscript.

Notes

[i] See http://www.bom.gov.au/climate/ahead/archive/tropical-cyclone/2010-2011-tc.shtml.
[ii] See http://weather.cityu.edu.hk/tc_forecast/2010_forecast_NOV.pdf.
[iii] See http://www.cityu.edu.hk/gcacic/pacific.htm.
[iv] See http://www.bom.gov.au/cyclone/history/tracks.
[v] See http://www.cpc.ncep.noaa.gov/data/indices/ersst3b.nino.mth.81-10.ascii.
[vi] See http://www.cdc.noaa.gov/people/klaus.wolter/MEI/table.html.
[vii] See http://www.jamstec.go.jp/frsgc/research/d1/iod/DATA/emi.monthly.txt.
[viii] See http://www.jamstec.go.jp/frcgc/research/d1/iod/DATA/dmi_HadISST_jan1958-dec2012.txt.
[ix] See ftp://ftp.bom.gov.au/anon/home/ncc/www/sco/soi/soiplaintext.html.

References

[1] R. E. Basher and X. Zheng. "Tropical cyclones in the southwest Pacific: Spatial patterns and relationships to Southern Oscillation and sea surface temperature." In: *Journal of Climate* 8.5 (1995), pp. 1249–1260. DOI: 10.1175/1520-0442(1995)008<1249:TCITSP>2.0.CO;2.

[2] A.I. Belousov, S.A. Verzakov, and J. von Frese. "A flexible classification approach with optimal generalisation performance: support vector machines." In: *Chemometrics and Intelligent Laboratory Systems* 64.1 (2002), pp. 15–25. DOI: 10.1016/S0169-7439(02)00046-1.

[3] S. J. Camargo, K. A. Emanuel, and A. H. Sobel. "Use of a genesis potential index to diagnose ENSO effects on tropical cyclone genesis." In: *Journal of Climate* 20.19 (2007), pp. 4819–4834. DOI: 10.1175/JCLI4282.1.

[4] S. J. Camargo and A. H. Sobel. "Western North Pacific tropical cyclone intensity and ENSO." In: *Journal of Climate* 18.15 (2005), pp. 2996–3006. DOI: 10.1175/JCLI3457.1.

[5] S. J. Camargo, A. G. Barnston, P. J. Klotzbach, and C. W. Landsea. "Seasonal tropical cyclone forecasts." In: *WMO Bulletin* 56.4 (2007), pp. 297–309.

[6] J. Camp, M. Roberts, C. MacLachlan, E. Wallace, L. Hermanson, A. Brookshaw, A. Arribas, and A. A. Scaife. "Seasonal forecasting of tropical storms using the Met Office GloSea5 seasonal forecast system." In: *Quarterly Journal of the Royal Meteorological Society* 141.691 (2015), pp. 2206–2219. DOI: 10.1002/qj.2516.

[7] S. S. Chand, K. J. Tory, J. L. McBride, M. C. Wheeler, R. A. Dare, and K. J. E. Walsh. "The different impact of positive-neutral and negative-neutral ENSO regimes on Australian tropical cyclones." In: *Journal of Climate* 26.20 (2013), pp. 8008–8016. DOI: 10.1175/JCLI-D-12-00769.1.

[8] C.-C. Chang and C.-J. Lin. "LIBSVM: A library for support vector machines." In: *ACM Transactions on Intelligent Systems and Technology* 2.3 (2011), 27:1–27:27. DOI: 10.1145/1961189.1961199.

[9] K. Davis, X. Zeng, and E. A. Ritchie. "A new statistical model to predict seasonal North Atlantic hurricane activity." In: *Weather and Forecasting* 30.3 (2015), pp. 730–741. DOI: 10.1175/WAF-D-14-00156.1.

[10] T. G. Dietterich. "Ensemble methods in machine learning." In: *Multiple classifier systems*. Vol. 1857. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2000, pp. 1–15. ISBN: 978-3-540-67704-8. DOI: 10.1007/3-540-45014-9_1.

[11] A. J. Dowdy. "Long-term changes in Australian tropical cyclone numbers." In: *Atmospheric Science Letters* 15.4 (2014), pp. 292–298. DOI: 10.1002/asl2.502.

[12] A. J. Dowdy, L. Qi, D. Jones, H. Ramsay, R. Fawcett, and Y. Kuleshov. "Tropical cyclone climatology of the South Pacific Ocean and its relationship to El Niño-Southern Oscillation." In: *Journal of Climate* 25.18 (2012), pp. 6108–6122. DOI: 10.1175/JCLI-D-11-00647.1.

[13] W. Drosdowsky and L. E. Chambers. *Near global sea surface temperature anomalies as predictors of Australian seasonal rainfall*. Tech. rep. 65. Melbourne: Bureau of Meteorology Research Centre, 1998.

[14] B. Efron. "Better bootstrap confidence intervals." In: *Journal of the American Statistical Association* 82.397 (1987), pp. 171–185. DOI: 10.1080/01621459.1987.10478410.

[15] K. Emanuel. "Increasing destructiveness of tropical cyclones over the past 30 years." In: *Nature* 436.7051 (2005), pp. 686–688. DOI: 10.1038/nature03906.

[16] J. L. Evans and R. J. Allan. "El Niño/Southern Oscillation modification to the structure of the monsoon and tropical cyclone activity in the Australasian region." In: *International Journal of Climatology* 12.6 (1992), pp. 611–623. DOI: 10.1002/joc.3370120607.

[17] W. M. Frank and G. S. Young. "The interannual variability of tropical cyclones." In: *Monthly Weather Review* 135.10 (2007), pp. 3587–3598. DOI: 10.1175/MWR3435.1.

[18] A. Z.-C. Goh and J. C. L. Chan. "An improved statistical scheme for the prediction of tropical cyclones making landfall in South China." In: *Weather and Forecasting* 25.2 (2010), pp. 587–593. DOI: 10.1175/2009WAF2222305.1.

[19] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning*. 2nd ed. Springer, 2009. ISBN: 978-0-387-84857-0.

[20] G. J. Holland. "On the quality of the Australian tropical cyclone data base." In: *Australian Meteorological Magazine* 29 (1981), pp. 169–181.

[21] N.-Y. Kang and J. B. Elsner. "Consensus on climate trends in western North Pacific tropical cyclones." In: *Journal of Climate* 25.21 (2012), pp. 7564–7573. DOI: 10.1175/JCLI-D-11-00735.1.

[22] A. Karatzoglou, A. Smola, K. Hornik, and A. Zeileis. "kernlab - An S4 package for kernel methods in R." In: *Journal of Statistical Software* 11.9 (2004), pp. 1–20. URL: http://www.jstatsoft.org/v11/i09.

[23] Y. Kuleshov, L. Qi, R. Fawcett, and D. Jones. "Improving preparedness to natural hazards: Tropical cyclone seasonal prediction for the Southern Hemisphere." In: *Advances in Geosciences, Volume 12: Ocean Science (OS)*. Ed. by J. Gan. Vol. 12. Singapore: World Scientific Publishing, 2009, pp. 127–143. DOI: 10.1142/9789812836168_0010.

[24] Y. Kuleshov, R. Fawcett, L. Qi, B. Trewin, D. Jones, J. McBride, and H. Ramsay. "Trends in tropical cyclones in the South Indian Ocean and the South Pacific Ocean." In: *Journal of Geophysical Research* 115.D01101 (2010). DOI: 10.1029/2009JD012372.

[25] 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." In: *Atmospheric and Climate Sciences* 2.3 (2012), pp. 298–306. DOI: 10.4236/acs.2012.23027.

[26] Y. Kuleshov, C. Spillman, Y. Wang, A. Charles, R. de Wit, K. Shelton, D. Jones, H. Hendon, C. Ganter, A. Watkins, J. Apajee, and A. Griesser. "Seasonal prediction of climate extremes for the Pacific: Tropical cyclones and extreme ocean temperatures." In: *Journal of Marine Science and Technology* 20.6 (2012), pp. 675–683. DOI: 10.6119/JMST-012-0628-1.

[27] S. Kullback and R. A. Leibler. "On information and sufficiency." In: *The Annals of Mathematical Statistics* 22.1 (1951), pp. 79–86. DOI: 10.1214/aoms/1177729694.

[28] M. A. Lander. "An exploratory analysis of the relationship between tropical storm formation in the western North Pacific and ENSO." In: *Monthly Weather Review* 122.4 (1994), pp. 636–651. DOI: 10.1175/1520-0493(1994)122<0636:AEAOTR>2.0.CO;2.

[29] J.-H. Lee and C.-J. Lin. *Automatic model selection for support vector machines*. Tech. rep. Department of Computer Science and Information Engineering, National Taiwan University, 2000.

[30] C.-J. Lin and R. C. Weng. *Simple probabilistic predictions for support vector regression*. Tech. rep. National Taiwan University, 2004.

[31] J. Malilay. "Tropical cyclones." In: *The public health consequences of disasters*. Ed. by E. K. Noji. Oxford University Press, 1996. Chap. 10, pp. 207–227. ISBN: 9780199747689.

[32] F. Molteni, R. Buizza, T. N. Palmer, and T. Petroliagis. "The ECMWF ensemble prediction system: Methodology and validation." In: *Quarterly Journal of the Royal Meteorological Society* 122.529 (1996), pp. 73–119. DOI: 10.1002/qj.49712252905.

[33] M. Momma and K. P. Bennett. "A pattern search method for model selection of support vector regression." In: *2002 SIAM International Conference on Data Mining*. SIAM, 2002, pp. 261–274. DOI: 10.1137/1.9781611972726.16.

[34] N. Nicholls. "A possible method for predicting seasonal tropical cyclone activity in the Australian region." In: *Monthly Weather Review* 107.9 (1979), pp. 1221–1224. DOI: 10.1175/1520-0493(1979)107<1221:APMFPS>2.0.CO;2.

[35] N. Nicholls. "The Southern Oscillation, sea-surface-temperature, and interannual fluctuations in Australian tropical cyclone activity." In: *Journal of Climatology* 4.6 (1984), pp. 661–670. DOI: 10.1002/joc.3370040609.

[36] N. Nicholls. "Recent performance of a method for forecasting Australian seasonal tropical cyclone activity." In: *Australian Meteorological Magazine* 40.2 (1992), pp. 105–110.

[37] J. C. Platt. "Fast training of support vector machines using sequential minimal optimization." In: *Advances in Kernel Methods. Support Vector Learning*. Ed. by B. Schölkopf, C. J. C. Burges, and A. J. Smola. MIT Press, 1999, pp. 185–208. ISBN: 9780262194167.

[38] R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria, 2014. URL: http://www.R-project.org/.

[39] S. Rajasekaran, S. Gayathri, and T.-L. Lee. "Support vector regression methodology for storm surge predictions." In: *Ocean Engineering* 35.16 (2008), pp. 1578–1587. DOI: 10.1016/j.oceaneng.2008.08.004.

[40] H. A. Ramsay, M. B. Richman, and L. M. Leslie. "Seasonal tropical cyclone predictions using optimized combinations of ENSO regions: Application to the Coral Sea basin." In: *Journal of Climate* 27.22 (2014), pp. 8527–8542. DOI: 10.1175/JCLI-D-14-00017.1.

[41] M. B. Richman and L. M. Leslie. "Adaptive machine learning approaches to seasonal prediction of tropical cyclones." In: *Procedia Computer Science* 12 (2012), pp. 276–281. DOI: 10.1016/j.procs.2012.09.069.

[42] K. K. Saha and S. A. Wasimi. "An index to assess the propensity of landfall in Australia of a tropical cyclone." In: *Natural Hazards* 72.2 (2014), pp. 1111–1121. DOI: 10.1007/s11069-014-1058-y.

[43] J. Shao. *Mathematical statistics*. 2nd ed. Springer, 2003. ISBN: 978-0-387-95382-3.

[44] K. E. Trenberth. "Signal versus noise in the Southern Oscillation." In: *Monthly Weather Review* 112.2 (1984), pp. 326–332. DOI: 10.1175/1520-0493(1984)112<0326:SVNITS>2.0.CO;2.

[45] K. E. Trenberth. "The definition of El Niño." In: *Bulletin of the American Meteorological Society* 78.12 (1997), pp. 2771–2777. DOI: 10.1175/1520-0477(1997)078<2771:TDOENO>2.0.CO;2.

[46] V. Vapnik. *The nature of statistical learning theory*. Springer-Verlag New York, 2000. ISBN: 978-0-387-98780-4. DOI: 10.1007/978-1-4757-3264-1.

[47] P. J. Webster, G. J. Holland, J. A. Curry, and H.-R. Chang. "Changes in tropical cyclone number, duration, and intensity in a warming environment." In: *Science* 309.5742 (2005), pp. 1844–1846. DOI: 10.1126/science.1116448.

[48] J. S. Wijnands, K. Shelton, and Y. Kuleshov. "Improving the operational methodology of tropical cyclone seasonal prediction in the Australian and the South Pacific Ocean regions." In: *Advances in Meteorology* (2014), pp. 1–8. DOI: 10.1155/2014/838746.

[49] D. S. Wilks. *Statistical methods in the atmospheric sciences*. 2nd ed. International Geophysics Series. Elsevier, 2006. ISBN: 978-0-12-751966-1.

[50] K. Wolter and M. S. Timlin. "Monitoring ENSO in COADS with a seasonally adjusted principal component index." In: *Proceedings of the 17th Annual Climate Diagnostics Workshop*. Norman, OK: NOAA/NMC/CAC, NSSL, Oklahoma Clim. Survey, CIMMS and the School of Meteor., Univ. of Oklahoma, 1993, pp. 52–57.

[51] Z. Zeng, W. W. Hsieh, A. Shabbar, and W. R. Burrows. "Seasonal prediction of winter extreme precipitation over Canada by support vector regression." In: *Hydrology and Earth System Sciences* 15.1 (2011), pp. 65–74. DOI: 10.5194/hess-15-65-2011.