Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 114 (2017) 314–323

Procedia Computer Science

www.elsevier.com/locate/procedia

Complex Adaptive Systems Conference with Theme: Engineering Cyber Physical Systems, CAS October 30 – November 1, 2017, Chicago, Illinois, USA

Reducing Tropical Cyclone Prediction Errors Using Machine Learning Approaches

Michael B. Richman[a]*, Lance M. Leslie[a], Hamish A. Ramsay[b], Philip J. Klotzbach[c]

[a]*School of Meteorology, University of Oklahoma, Norman, OK 73072 United States*
[b]*School of Earth, Atmosphere and Environment, Monash University, Melbourne, 3800 Australia*
[c]*Department of Atmospheric Science, Colorado State University, Fort Collins, CO 80523 United States*

Abstract

Tropical cyclones (TCs) in the North Atlantic region are predictions for the June 1-November 30 season using predictors from Atlantic, Indian and Pacific Oceans sea surface temperature anomalies. Here, the aim is to reduce TC seasonal prediction errors and is realized by applying support vector regression (SVR) to an initial predictor pool and using the model prediction errors to iteratively identify additional attributes that reduce those errors. Prediction errors from this approach are compared with those from an existing, statistical seasonal prediction model, developed at Colorado State University (CSU). The SVR approach was optimized using attribute selection with wrapper selection techniques and by testing various kernels over a range of complexity parameter. Results of the comparison between seasonal SVR and the CSU TC model indicate that proper attribute selection lowers prediction errors significantly. Compared with the CSU model, the SVR model increases correlations between prediction and observed annual TC count from 0.62 to 0.83; the mean absolute error (MAE) is reduced from 2.9 to 1.8 and the root mean squared error (RMSE) drops from 3.8 to 2.7. Furthermore, the approach of using prediction errors to improve machine learning models is flexible and can be adapted readily to other TC basins.

Peer-review under responsibility of the scientific committee of the Complex Adaptive Systems Conference with Theme: Engineering Cyber Physical Systems.

Keywords: Machine Learning; Support Vector Regression; Kernels; Tropical Cyclones; Prediction Errors; Prediction; Attribute Selection

* Corresponding author. Tel.: +1-405-325-1853; fax: +1-405-325-7689.
E-mail address: mrichman@ou.edu

1877-0509
Peer-review under responsibility of the scientific committee of the Complex Adaptive Systems Conference with Theme: Engineering Cyber Physical Systems.
10.1016/j.procs.2017.09.048

1. Introduction

Tropical cyclones (TCs) rank high among the most devastating natural hazards. They are rotating low pressure tropical weather phenomena characterized by destructive winds and intense rainfall. When approaching coastal regions TCs are at their most threatening as, in addition to the strong winds and heavy rainfall, they can generate flooding from a combination of both massive rainfall amounts over land and storm surges generated over the oceans. TCs are an annual threat to many countries, with the potential to cause massive human casualties, damage to infrastructure, loss of crops and livestock, and irrevocable harm to ecological systems.

Given the immense socio-economic impact of TCs, especially those near land or making landfall, much research effort has been made to provide timely and skillful seasonal predictions of the number (and frequently the intensity and landfall locations) of TCs occurring in the TC seasons of the many regions affected by these storms. TC seasonal predictions for the North Atlantic region, based on statistical modeling, are issued by Colorado State University (CSU). These predictions commenced in 1984 [1], and have been posted every season from 1984 to the present. The statistical models used in prediction mode by CSU will be referred to hereafter as the CSU model. The oceanic and atmospheric predictors have changed since 1984 and the changes are summarized by [2]. An important aspect of the CSU model and the machine learning model developed here is that both utilize significant statistical relationships established between the total North Atlantic seasonal TC numbers and sea surface temperature (SST) anomalies (SSTA) as discussed, for example, by [3]. The SSTA are measured relative to long-term monthly mean SST values (e.g., the 1981-2010 mean SST). More recently, CSU's statistical models predict Accumulated Cyclone Energy (ACE; an integrated metric that accounts for frequency, intensity and duration of TCs), and the prediction for the number of TCs is derived from a linear regression between historical TCs and ACE.

Additional CSU predictions are generated during the TC season, but not addressed here, as our focus is on predictions issued before or at the beginning of the North Atlantic TC season (June 1). The CSU model predictions are used as the baseline relative to which the machine learning model developed in this study is compared. The CSU predictions were chosen primarily because of their lengthy prediction record of 32 years. It is noted also that the present study focusses only on the seasonal count of all TCs and does not provide separate count predictions of those TCs that reach hurricane strength or ACE.

2. Data and Methodology

2.1. Data

As mentioned in the Introduction, the TC season in the North Atlantic Ocean begins on June 1 and ends on November 30, and one of the CSU TC seasonal predictions is issued around June 1 for the upcoming TC season. Here, the CSU TC prediction issued around June 1 is compared to the observed number of TCs for the period 1984 to 2015, with the differences between the predictions and observations defining the CSU prediction errors. Monthly SST data are obtained from the Hadley Centre Sea Ice and Sea Surface Temperature dataset (HadISST; [4]) for the 32 year period 1984-2015, and anomalies are computed relative to the 1981-2010 monthly climatology.

Rank correlations were formed between the vector of annual total TCs and the set of monthly average SSTAs for January through May 1984-2015. The monthly rank correlations at every location in the global oceans between 60°N and 60°S were isoplethed to define monthly spatial fields. Rank correlation was used, rather than the more traditional Pearson correlation, as it measures nonlinear relationships in the values that are linear in the ranks (i.e., monotone relationships). Rank correlation also helps to reduce the impact of outliers, for example, the 2005 Atlantic hurricane season where 28 TCs were observed. Rank correlation coefficients with the most statistical significance and spatial coherence were represented as rectangles (Fig. 1). Each rectangle is treated as an attribute, lettered by basin and numbered (e.g., A1 for Atlantic region 1). For each rectangle, all gridpoints were averaged to obtain the mean monthly SSTA. These attributes served as the basis for a predictor pool, along with the trends in SSTAs for each rectangle for April to May. These two-way interactions between rectangles is defined as the multiplication of the SSTA between all sets of two boxes (e.g., A1*P1). The process and rationale of defining interaction terms in this manner is outlined in [5]. This combination of 5 months of candidate attributes times 6 attributes per month, 6 trend attributes and 75 interaction attributes results in 111 candidate attributes.

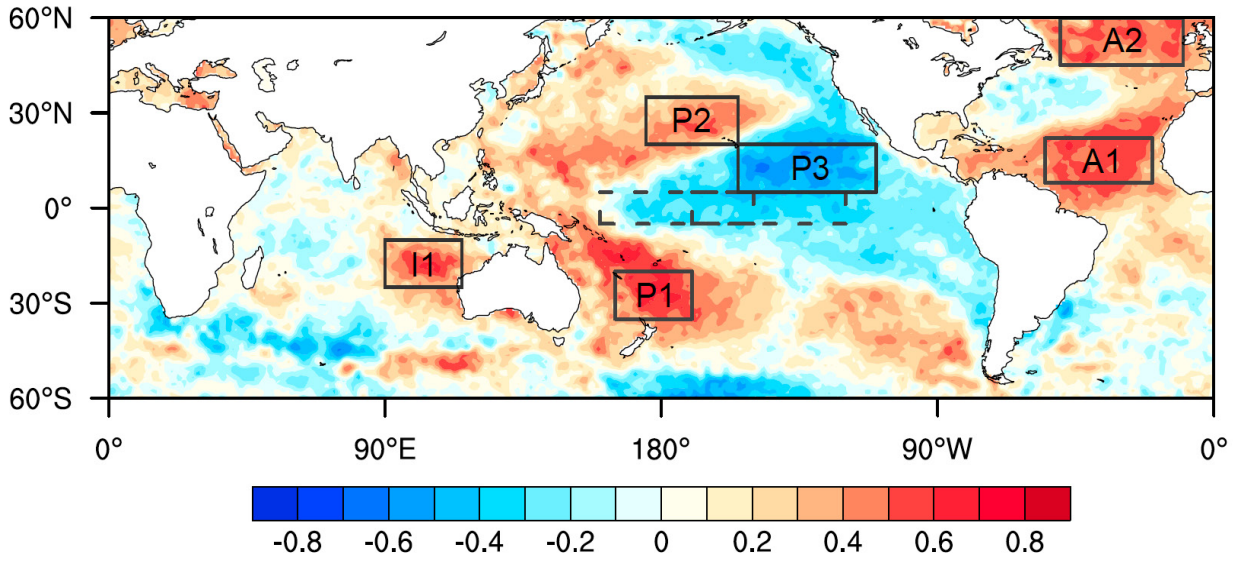

Fig. 1. Rank correlation map of global sea-surface temperature anomalies for May with TCs (tropical storms and hurricanes). Atlantic Ocean attributes are denoted by the letter A, Pacific Ocean attributes by the letter P and Indian Ocean attributes by the letter I followed by the region number. The dashed rectangles show the well-established Niño 4 and Niño 3.4 regions used to define ENSO events.

2.2. Support Vector Regression (SVR)

In SVR, the input, x, is mapped onto a m-dimensional feature space using some fixed (nonlinear) mapping, then a linear model is constructed in this feature space. The linear model, in the feature space, $f(x,\omega)$, is

$$f(x,\omega) = \sum_{j=1}^{m} \omega_j \, g_j(x) + b, \tag{1}$$

where $g_j(x),\ j = 1,\dots,m$ denotes a set of nonlinear transformations, and b is the "bias" term [6]. Often the data are assumed to have a zero mean (this can be achieved by preprocessing), so the bias term is dropped. The quality of estimation is measured by the loss function $L(y, f(x,\omega))$. SVR uses an ε-insensitive loss function by [7]:

$$L_\varepsilon\big(y, f(x,\omega)\big) = \begin{cases} 0 & if\ |y\text{-}f(x,\omega)| \le \varepsilon \\ |y\text{-}f(x,\omega)|\text{-}\varepsilon & otherwise \end{cases} \tag{2}$$

The empirical risk is:

$$R_{emp}(\omega) = \frac{1}{n}\sum_{i=1}^{n} L_\varepsilon\left(y_i, f(x_i,\omega)\right),\ i = 1,\dots,n \tag{3}$$

SVR performs linear regression in high-dimension feature space using ε-insensitive loss, and also tries to reduce model complexity by minimizing $\|\omega\|^2$. This is described by introducing (non-negative) slack variables ξ_i, ξ_i^*, to measure the deviation of training samples outside ε-insensitive zone. Thus, SVR is formulated as minimization of:

$$\min\ \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(\xi_i, \xi_i^*)$$

$$\text{s.t.}\ \begin{cases} y_i\text{-}f(x_i,\omega) \le \varepsilon + \xi_i^* \\ f(x_i,\omega)\text{-}y_i \le \varepsilon + \xi_i \\ \xi_i, \xi_i^* \ge 0,\ i = 1,\dots,n \end{cases} \tag{4}$$

where C is the complexity parameter. This optimization problem becomes the dual problem and its solution is:

$$f(x) = \sum_{i=1}^{n_{sv}} (\alpha_i - a_i^*)\, K(x_i, x) \tag{5}$$

s.t. $0 \le \alpha_i^* \le C,\ 0 \le \alpha_i \le C$

where n_{sv} is the number of support vectors and the kernel function is

$$K(x, x_i) = \sum_{j=1}^{m} g_j(x)\, g_j(x_i) \tag{6}$$

2.3. SVR Attribute Selection

A wrapper selection process [8] using greedy SVR with 10-fold cross validation to select only attributes that appeared in ≥ 5 of the 10 folds. The initial set of attributes surviving the process were from May, the month just prior to TC season, onset and included A1, P2, P3, and I1. Additionally, the interaction terms between (i) A1 and P2 and (ii) P1 and P3 were retained (Fig. 1). Predictions of the 1984-2015 number of TCs in the Atlantic basin were made using a 10-fold cross validated SVR. The resulting SVR predictions were compared to the observed number of TCs and an error vector was created. Rank correlations between this error vector and oceanic SSTA was formed, isoplethed and a second set of rectangles was constructed (Fig. 2) that were statistically significant and spatially extensive. The process is outlined in Fig. 3. An additional Atlantic Ocean predictor, four additional Pacific Ocean predictors and two additional Indian Ocean predictors. The same wrapper attribute selection process with cross-validated SVR identified a final set of predictors. This final set of attributes identified all were from May: A1, P3, P5, P6, I1, I3 and two trend terms identifying SSTAs from April to May for A1 (A1tr) and P2 (P2tr).

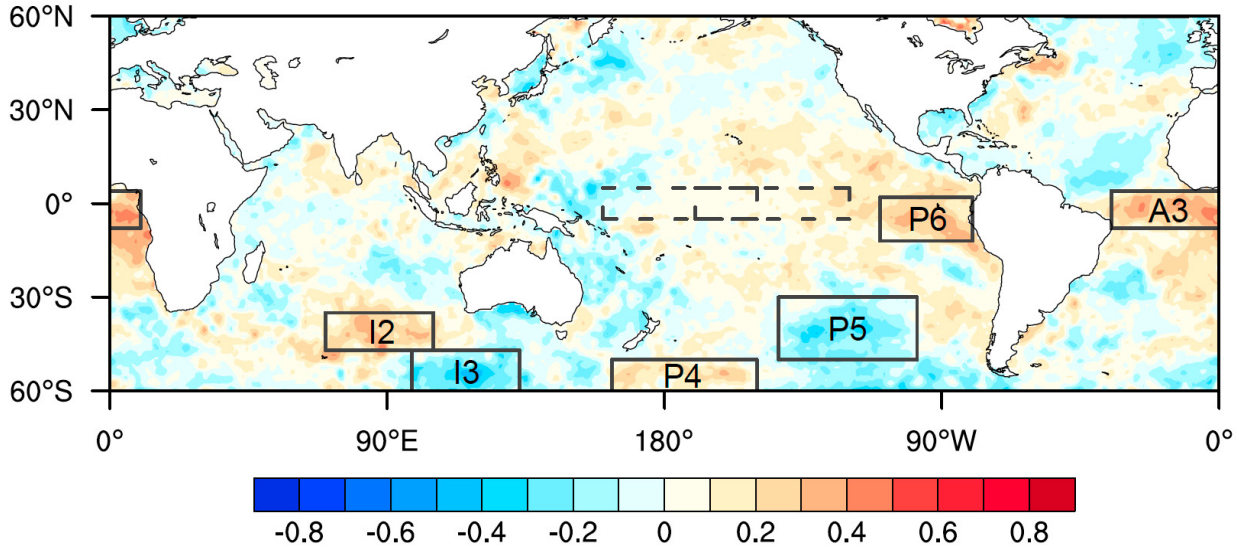

Fig. 2. Rank correlation map of global SSTAs for May with initial SVR error vector. Atlantic Ocean attributes are denoted by the letter A, Pacific Ocean by P and Indian Ocean by I followed by region number. Dashed rectangles are as in Fig. 1.

3. Results and Discussion

3.1. Optimization of the SVR Complexity Parameter, C

Optimization using polynomial and radial basis function kernels was undertaken. Altering the complexity parameter, *C* in (4), on the accuracy of the solution was investigated by starting the cross-validation with different random seeds to assemble information on the variation at discrete values of *C*. *C* determines the trade-off between the model complexity (flatness) and the degree to which deviations larger than ε are tolerated in optimization

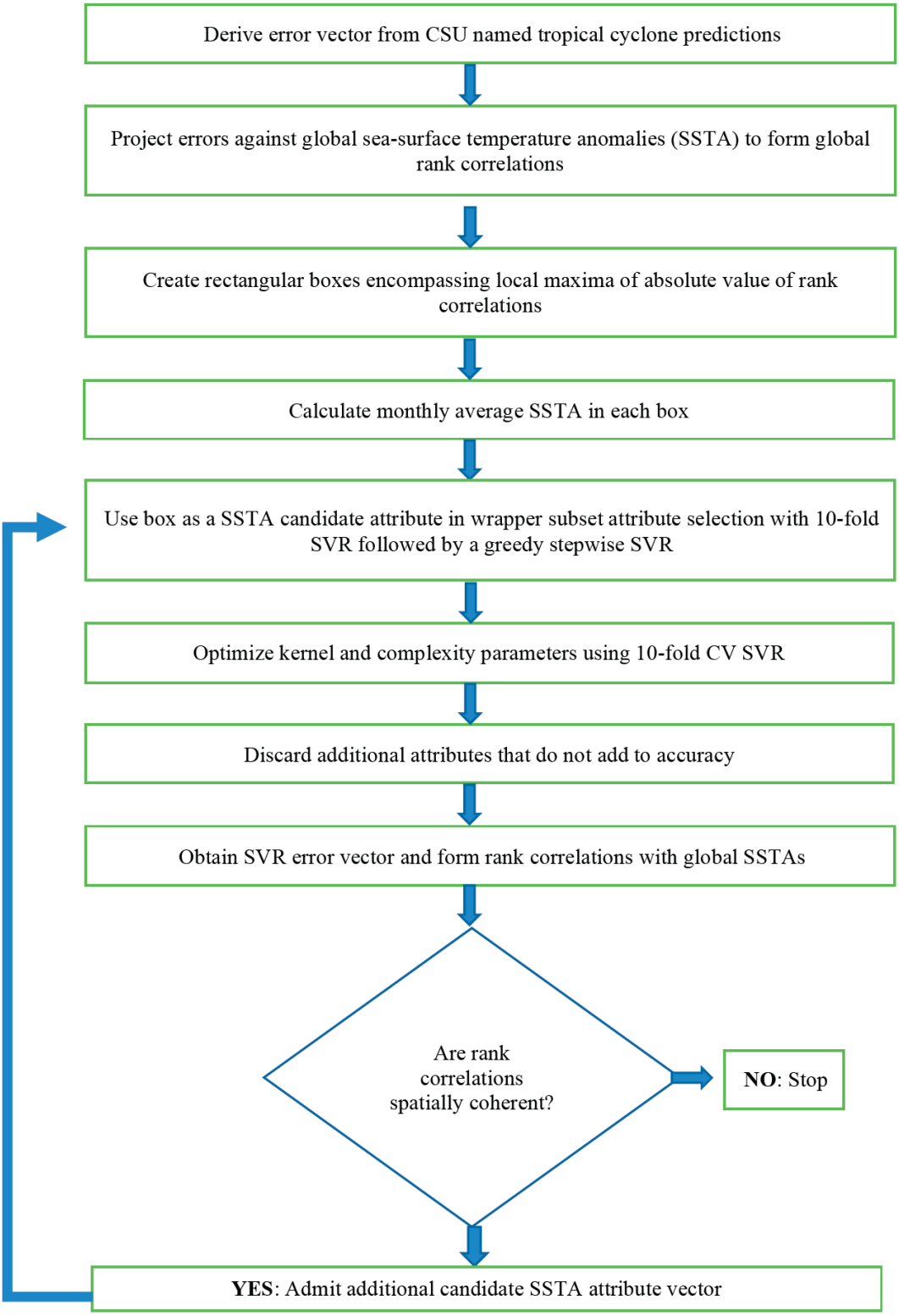

Fig. 3. Flowchart of model error attribute selection process.

formulation. If C is too large, then the objective is to minimize the empirical risk too heavily, without regard to the model complexity part in the optimization formulation. For large values of C, the optimization chooses a smaller-margin hyperplane, if that hyperplane captures all the training points that fit well. Conversely, a very small value of C will cause the optimization to seek a larger-margin separating hyperplane, even if that hyperplane does not fit well using more points. For near-zero values of C, more points are poorly fit, even if the training data are linearly separable.

3.2. Using an Optimal Range of C to Investigate the Variability of SVR Weights

Experimentation over a large range of C (0.1 to 3.0) revealed that (i) there was a broad minimum for most random seeds in the range of $0.5 < C < 2.0$ and, (ii), the mean absolute error (MAE) values and minimum vary as a function of the random seed in the 10-fold cross-validation. Therefore, to establish a reliable mean and confidence interval on the MAE, analyses from ten random seeds were used to form ten sets of MAE results and those results for each value of C was bootstrapped 5000 times. Fig. 4 shows the boxplots of that bootstrap interval.

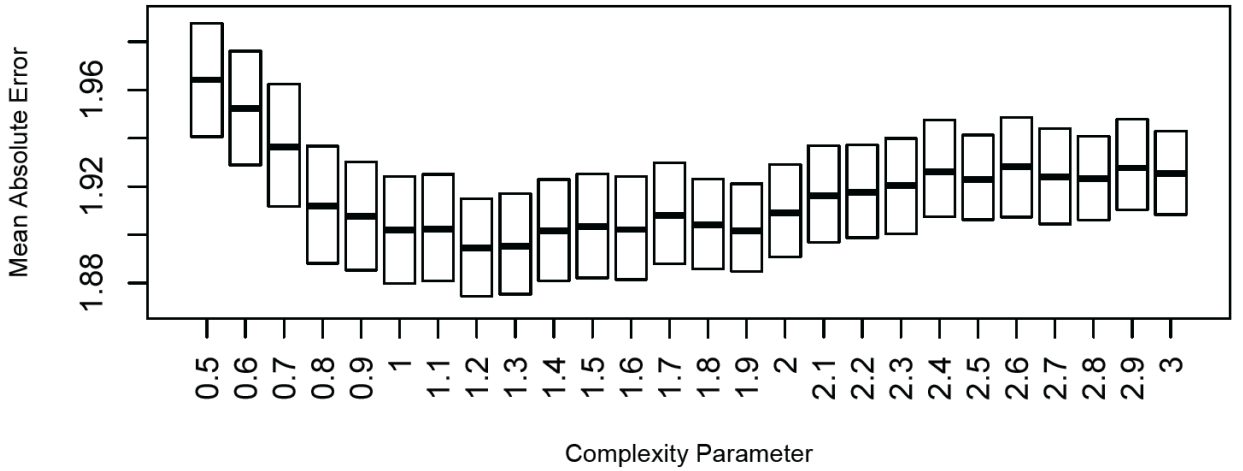

Fig. 4. Boxplots of MAE accuracy of the SVR predictions of TCs as a function of complexity parameter, C.

The bootstrapped mean MAE from Fig. 4 was used to investigate the effect of complexity parameter on the attribute weights in the SVR. Each value of C from 0.5 to 3.0 was used to form a separate solution, and the resulting attribute weights and their variability were summarized in the boxplots for each weight (Fig. 5). A noteworthy finding was that each attribute weight remained stable (showed little variation) over the tested range of C.

3.3. Physical Interpretation of Attributes

A1 (Fig. 5) is the most important predictor, as expected. There is already an established statistical (linear) relationship between local SSTAs in North Atlantic (i.e., A1) and seasonal TC counts [9]. The second most important attribute is the P3 region, which is immediately north of the main El Niño 3.4 region (dashed rectangle in Fig. 1). A1tr, the trend in SSTs from April through May, is the next most important variable. As in A1, warmer waters are more conducive for TC development. Anomalous warming from April to May is typically associated with an anomalously weak subtropical high and reduced trade wind strength, characteristic of a positive Atlantic Meridional Mode [10]. P2tr, is the change in SSTs from April to May in the western North Pacific Ocean. This region often warms as La Niña matures. The P6 region, known as the "cold tongue" in the Pacific Ocean off the equatorial South American coast, is sensitive to El Niño onset conditions. When El Niño occurs, the upwelling of cold waters is reduced and SSTA becomes positive. As El Niño matures, wind shear in the Atlantic TC genesis region and especially in the Caribbean increases, thereby reducing the development of TCs. P6 is a South Pacific Ocean predictor. I3 is a high latitude Indian Ocean region where SSTAs are controlled by the strength of the westerly winds that control ocean gyres and currents. I1 is also an Indian Ocean attribute, in a region known to

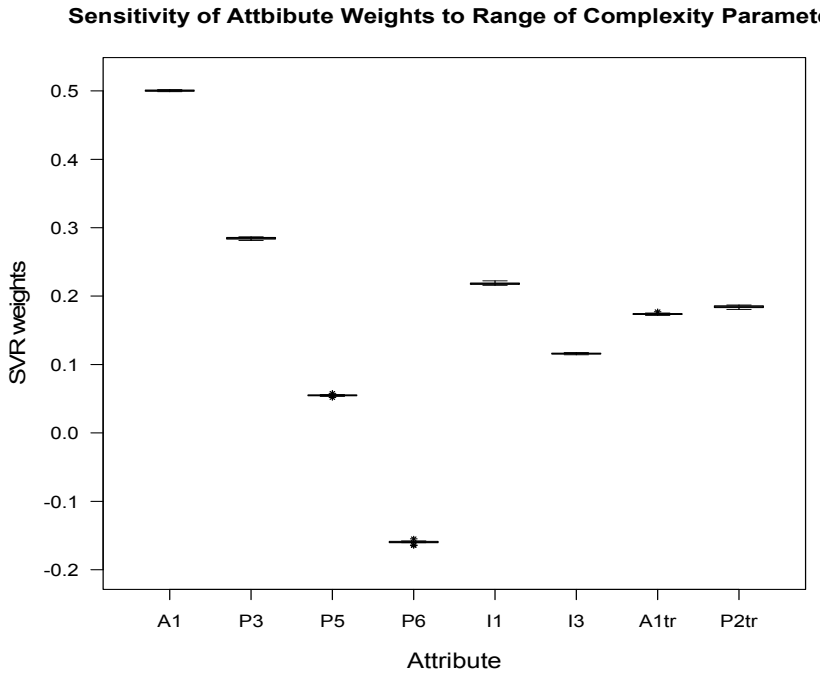

Fig. 5. Boxplots of SVR weights and for each attribute as a function of varying the complexity parameter.

modulate Australia TCs; however, physical link to Atlantic TCs is not known. P5 is a mid-latitude South Pacific rectangle. Like I3, it is in a region that likely is modulated by the strength of the westerly winds.

3.4. Accuracy of Results Compared to CSU Model

This experiment investigates if a machine learning approach could perform at or above the level of accuracy of existing predictions for Atlantic TCs. After fitting RBF kernels and polynomial kernels of various exponents through optimization of the complexity parameter and epsilon, two models were selected. The CSU model is a baseline to compare SVR models. A comparison of the RBF SVR model to the CSU model predictions (Fig. 6) reveals yearly CSU model predictions had four large errors (>5 storms in 1995, 2005, 2006 and 2012), whereas the SVR model had large errors in two years (>5 storms in 1995 and 2005). In all years with larger errors, the SVR prediction was closer to observed TC numbers than the CSU prediction (Fig. 7). The variance explained increases from 38% (CSU) to 68% and 68% for the SVR Polynomial and RBF kernel models, both 80% improvements over the CSU model. Other evaluation statistics (MAE, RMSE) showed a similar improvement. The CSU model has an MAE of 2.9 TCs per year, which decreased to 1.8 (a 65% reduction). for the polynomial kernel and to 1.8 (a 64% reduction) for the RBF kernel. The scatterplots (Fig. 7) visually confirm these improvements obtained from the SVR approach. The error distributions, when binned by the nearest whole number, are shown in Fig. 8. The sharpness of the predictions values, with a much larger number of zero errors and small errors (defined as number of errors $\leq |2|$ TCs) is evident for the SVR models.

Prediction error statistics are shown in Table 1. The mean error (bias) for the CSU model -.0.97, indicating underprediction, is reduced by 57% and 60% for the two SVR models. The median discards the actual numbers, in favor of ranks, eliminating effects of the outliers (e.g., 2005 errors), hence is a resistant bias statistic. The CSU model median error is -1; SVR models are essentially zero. Similarly, the CSU model mode (-3) suggests an underprediction bias, but SVR models' modes both are zero. The amount of error, as a deviation from the mean error, is shown in the variance statistic. CSU model variance is 14.1, suggesting a wide range in prediction errors. Both SVR models damp these error swings with values of 5.9 and 5.7, decreases of 58% and 60% over the CSU model. Discounting the largest errors, applying the interquartile range (IQR), the errors of the middle 50 percent of

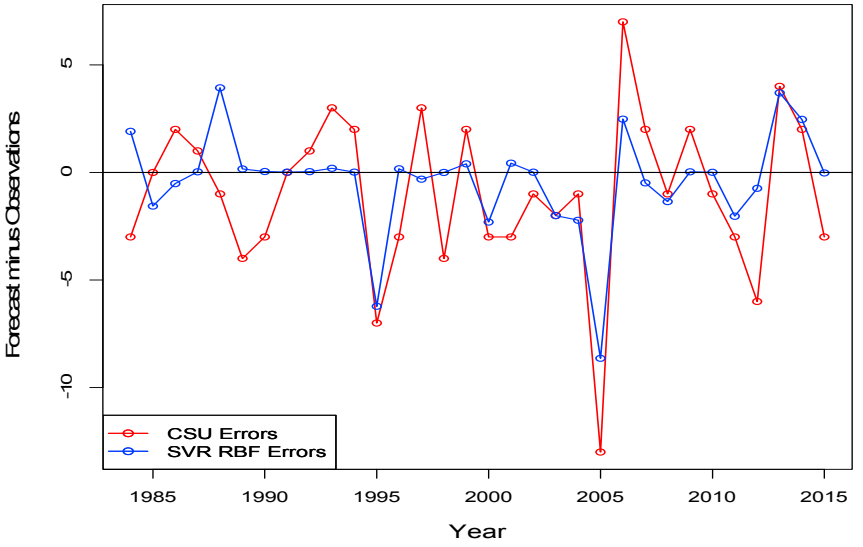

Fig. 6. Time series of CSU TC predictions versus SVR RBF predictions for the same years.

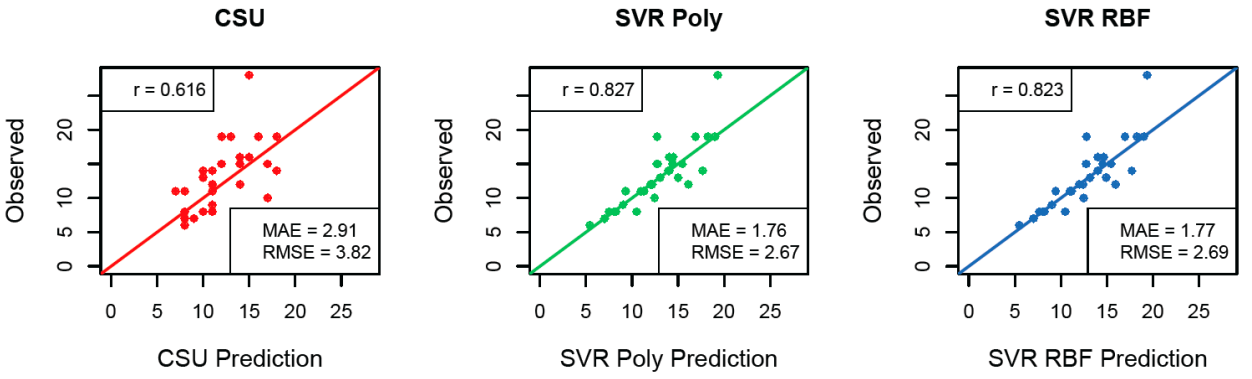

Fig. 7. Scatterplots of the number of observed TCs compared to the number predicted for the CSU (left), SVR Poly (middle) and SVR RBF(right) models. The correlation between predicted and observed TCs as well as the MAE and RMSE are shown.

the distribution can be examined. The CSU IQR is 5 whereas the SVR models' IQRs are 1.2 and 1.1, improvements of 77% and 79%. The direction of the larger errors can be measured by the skewness statistic. All three models are left skewed, because larger errors are underpredictions. The CSU model has the smallest negative skewness and that can be interpreted as an advantage over the SVR models; however, skewness has standard deviation in the denominator and SVR standard deviations (see variance statistic) are much smaller than the CSU standard deviation. Because of the smaller variance in the SVR models, a smaller error causes the skewness to increase. Therefore, limiting the deviations to the middle 50% of the distribution can be investigated with the Yule-Kendall statistic. The CSU error directionality in the middle 50% becomes slightly positive, whereas the middle 50% of the error directionality in the SVR is moderately negative. As Yule-Kendall statistic has the IQR in the denominator and the SVR models have IQRs 20-25% as large as the CSU model, the results of the Yule-Kendall statistic do not capture the importance of the errors. The fourth central moment, kurtosis, is often interpreted as the tailedness of the distribution. Since the distribution is comprised of errors, if the mean lies near zero, then a large kurtosis reflects an accurate solution but, if the mean does not lie near zero, then a large value of kurtosis is indicative of large errors. Therefore, kurtosis must be interpreted after examining the mean. All three models are highly leptokurtic but the CSU model mean is centered near -1, whereas for the SVR models the peaks are ~ -0.4. The kurtosis values of the

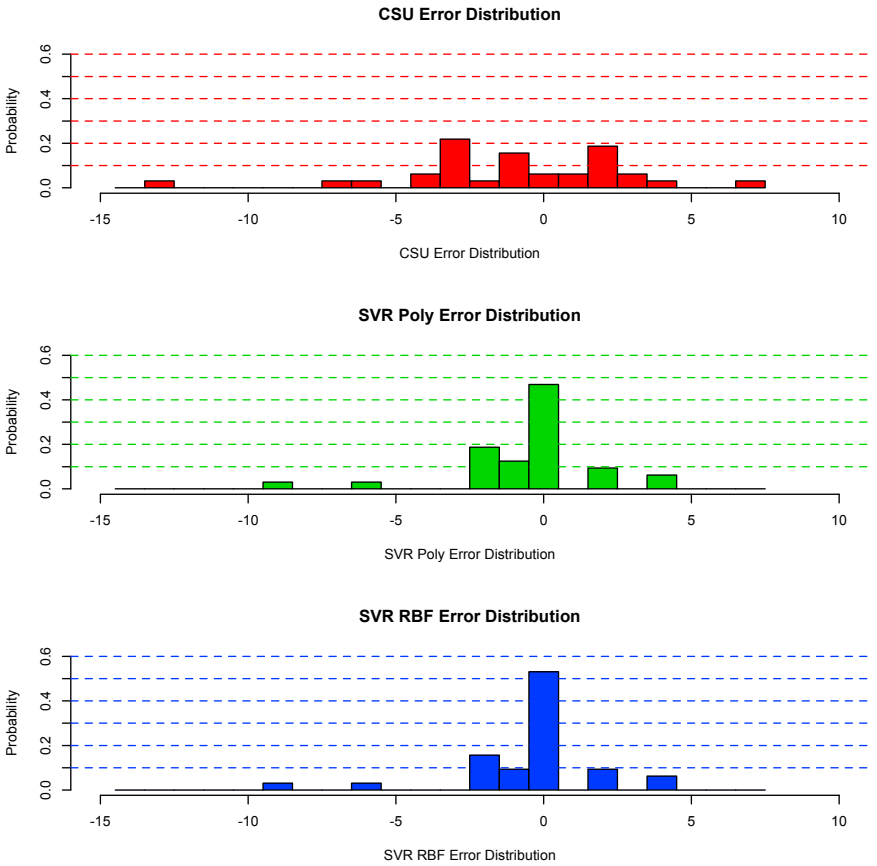

Fig. 8. Histograms of the TC prediction errors for the CSU, SVR Poly and SVR RBF models.

Table 1. Comparison of CSU to SVR prediction errors using distributional moments, resistant statistics and prediction evaluation indices (CSI, POD, FAR, Bias). A correct positive prediction defined as predictions within ±2.49 (rounded to nearest whole number) TCs of that observed. False alarms (false positives) are defined as prediction errors of ≥+2.5 TCs (rounded to nearest whole number) compared to number observed and misses (false negatives) are defined as prediction errors of ≤-2.5 TCs (rounded to nearest whole number).

Statistic	CSU Error	SVR Poly Error	SVR RBF Error
Mean	-0.969	-0.413	-0.389
Median	-1	0.012	0.012
Mode	-3	0	0
Variance	14.1	5.9	5.7
IQR	5	1.17	1.07
Skewness	-0.78	-1.29	-1.35
Yule-Kendall Statistic	0.2	0.67	0.70
Kurtosis	4.68	-0.67	0.70
CSI = a/(a+b+c)	0.5	0.781	0.812
POD = a/(a+c)	0.571	0.893	0.897
FAR = b/(a+b)	0.2	0.138	0.103
Bias = (a+b)/(a+c)	0.714	1.036	1

SVR are 36% and 39% larger than for the CSU model. Given the larger kurtosis and smaller mean error, these statistics imply a superior error distribution for the SVR models.

If the prediction errors are categorized appropriately, their practical significance can be assessed. Comparing the CSU to SVR predictions using evaluation indices with correct positives ("hits"), i.e., within ±2.5 (rounded to nearest whole number) TCs (Table 1). Hence, an error of 2.3 TCs is rounded to 2 TCs whereas 2.8 rounds to 3 to obtain integer values for the number of TCs. False alarms (false positives) are defined as errors of ≥+2.5 TCs (rounded to nearest whole number, 3) compared to number observed. Misses (false negatives) are defined as errors of ≤-2.5 TCs (rounded to nearest whole number, -3). Since there are no "correct nulls" using this approach, only cells a, b, and c of the confusion matrix are considered [11]. The available prediction evaluation statistics of these cells include CSI, POD, FAR and Bias. The CSI measures the number of hits divided by all the number of observations. The range is 0 to 1, with values close to 1 desirable. SVR models show improvements of 56% and 62% over the CSU predictions. The POD measures the number of hits divided by the number of observed storms in each category with a range of 0 to 1 with 1 being the most desirable. The two SVR models have improvements of 56% and 57% over the CSU model. Since POD can be artificially high when models overpredict, the FAR also must be considered to assess the goodness of the prediction classifications. FAR is false alarms divided by predicted storms, ranging from 0 to 1 with 0 the most desirable. SVR models have FARs lower than the CSU model, improving by 31% and 49% over the CSU model. Bias measures the number of overpredictions divided by the number of underpredictions. Its range of 0 to +∞, with an ideal value of one. As bias does not measure the number of hits directly, it is part of the evaluation process. The SVR models have biases closer to 1 than does the CSU model, with improvements of 35% and 40%. The suite of prediction evaluation indices shown in Table 1 provides a clear indication that the SVR models have a categorical performance advantage over the CSU model.

4. Conclusions

This study evaluates the predictive skill of a machine learning model for the seasonal prediction of tropical cyclone counts for the North Atlantic region. The skill is measured relative to an established seasonal TC statistical prediction model developed at and maintained by Colorado State University (CSU). The machine leaning method, which is based on a careful selection of predictors is a support vector regression (SVR) model, utilizing two different kernels. The SVR model predictions outperform the CSU model predictions, issued on June 1, using a wide range of metrics, including an increased correlation between observed and predicted TC count from 0.62 to 0.83; and reductions of the MAE (RMSE) from 2.9 (3.8) to 1.8 (2.7). For reasons outlined in the introduction, further comparisons using the most recent CSU models and SVR models are planned. To address the challenge of determining the accuracy of these models in true forecast mode, predictions will be archived.

References

[1] Gray, W.M. (1984). Atlantic seasonal hurricane frequency. Part II: Forecasting its variability. *Mon. Wea. Rev.*, **112**, 1669–1683.

[2] Klotzbach, P.J. and Gray, W.M. (2009). Twenty-five years of Atlantic basin seasonal hurricane forecasts (1984–2008). *Geophys. Res. Lett.*, **36**, L09711, doi:10.1029/2009GL037580.

[3] Ramsay H.A., Leslie L.M., Lamb P.J., Richman M.B. and Leplastrier M. (2008). Interannual variability of tropical cyclones in the Australian region: role of large-scale environment. *J. Climate*, **21,** 1083–1103.

[4]. Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, C. K., Rowell, D. P., Kent, E. C. and Kaplan, A. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407.

[5] Ramsay, H.A., Richman, M.B. and Leslie, L.M. (2014). Seasonal tropical cyclone predictions using optimized combinations of ENSO regions: application to the Coral Sea basin. *J. Climate*, **27**, 8527–8542

[6]. Smola, A. and Schölkopf, B. (1998). A Tutorial on Support Vector Regression. NeuroCOLT Technical Report NC-TR-98-030, Royal Holloway College, University of London, UK.

[7] Vapnik, V.N. (1982). Estimation of Dependences Based on Empirical Data. Springer-Verlag. Berlin, Germany.

[8] Kohavi, R and John G.H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, **97**, 273-324.

[9] Zhao, M., Held, I.M. and Vecchi, G.A. (2010). Forecasts of the hurricane season using a global atmospheric model assuming persistence of SST anomalies. *Mon. Wea. Rev.*, **138**, 3858–3868.

[10] Vimont, D. J., and Kossin, D.J. (2007). The Atlantic meridional mode and hurricane activity. *Geophys. Res. Lett.*, **34**, L07709, doi:10.1029/2007GL029683.

[11] Wilks, D.S. (2011). Statistical Methods in the Atmospheric Sciences. Third Edition. Academic Press, Cambridge, MA.