

Valparaiso University
ValpoScholar

Geography and Meteorology Faculty Publications

Department of Geography and Meteorology

11-14-2017

Statistical Deterministic and Ensemble Seasonal Prediction of Tropical Cyclones in the Northwest Australian Region

Kevin H. Goebbert
Valparaiso University

Follow this and additional works at: https://scholar.valpo.edu/geomet_fac_pub

Recommended Citation

Goebbert, K.H., 2017: Statistical Deterministic and Ensemble Seasonal Prediction of Tropical Cyclones in the Northwest Australian Region. *Wea. Forecasting*, 32, 2117–2132, <https://doi.org/10.1175/WAF-D-17-0042.1>

This Article is brought to you for free and open access by the Department of Geography and Meteorology at ValpoScholar. It has been accepted for inclusion in Geography and Meteorology Faculty Publications by an authorized administrator of ValpoScholar. For more information, please contact a ValpoScholar staff member at scholar@valpo.edu.

Statistical Deterministic and Ensemble Seasonal Prediction of Tropical Cyclones in the Northwest Australian Region

KEVIN H. GOEBBERT

Valparaiso University, Valparaiso, Indiana

(Manuscript received 24 March 2017, in final form 26 September 2017)

ABSTRACT

Statistical seasonal prediction of tropical cyclones (TCs) has been ongoing for quite some time in many different ocean basins across the world. While a few basins (e.g., North Atlantic and western North Pacific) have been extensively studied and forecasted for many years, Southern Hemispheric TCs have been less frequently studied and generally grouped as a whole or into two primary basins: southern Indian Ocean and Australian. This paper investigates the predictability of TCs in the northwest Australian (NWAUS) basin of the southeast Indian Ocean (105°–135°E) and describes two statistical approaches to the seasonal prediction of TC frequency, TC days, and accumulated cyclone energy (ACE). The first approach is a traditional deterministic seasonal prediction using predictors identified from NCEP–NCAR reanalysis fields using multiple linear regression. The second is a 100-member statistical ensemble approach with the same predictors as the deterministic model but with a resampling of the dataset with replacement and smearing input values to generate slightly different coefficients in the multiple linear regression prediction equations. Both the deterministic and ensemble schemes provide valuable forecasts that are better than climatological forecasts. The ensemble approach outperforms the deterministic model as well as adding quantitative uncertainty that reflects the predictability of a given TC season.

1. Introduction

Tropical cyclones (TCs) can cause substantial damage and disrupt the regions they impact, making prediction of these phenomena crucial for regions that can be adversely affected (Harper et al. 2008). There are multiple time scales over which forecasts can be made of TCs, including climatic, seasonal, and individual events. This paper focuses on the seasonal prediction of three common TC metrics: TC frequency, TC days, and accumulated cyclone energy (ACE). Predictions of these metrics can be used as a general marker for whether it is expected that a region will have below-average, average, or above-average activity within the TC basin.

The relationship of TC activity to the global atmospheric circulation has been the subject of examination for a long time (Ballenzweig 1959; Namias 1969; Frank 1977) and has been used to make seasonal predictions. Specifically, seasonal TC prediction has utilized known climate indices [e.g., Southern Oscillation index (SOI), quasi-biennial oscillation (QBO), El Niño–Southern

Oscillation (ENSO)] to describe the current state of the large-scale atmospheric circulation (e.g., Gray 1984a,b; Nicholls 1979, 1984, 1985), but more recently predictors have included primary state variables from the NCEP–NCAR reanalysis dataset (Kalnay et al. 1996) that exhibit high spatial correlation with TC activity (e.g., Klotzbach and Gray 2003, 2004). Many of the predictors identified from the NCEP–NCAR reanalysis are strongly related to the classic climate indices and ultimately describe the large-scale circulation in a similar fashion (e.g., Klotzbach and Gray 2004; Goebbert and Leslie 2010).

Most of the research on TCs, especially seasonal prediction, occurs for regions in the Northern Hemisphere. Southern Hemispheric storms, which occur in the Indian and southwest Pacific Oceans, are not as well covered in the literature. There are two primary TC basins in the Southern Hemisphere: the southern Indian Ocean basin and the Australian basin. The Australian basin can be further separated into the northwest Australian (NWAUS) basin (e.g., Goebbert and Leslie 2010), the southwest Pacific Ocean basin (e.g., Diamond et al. 2013), and the Fiji region (e.g., Chand et al. 2010). This paper will focus on the prediction of TCs within the NWAUS basin between 105° and 135°E (Fig. 1).

Corresponding author: Dr. Kevin Goebbert, kevin.goebbert@valpo.edu

DOI: 10.1175/WAF-D-17-0042.1

© 2017 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](https://www.ametsoc.org/PUBSReuseLicenses) (www.ametsoc.org/PUBSReuseLicenses).

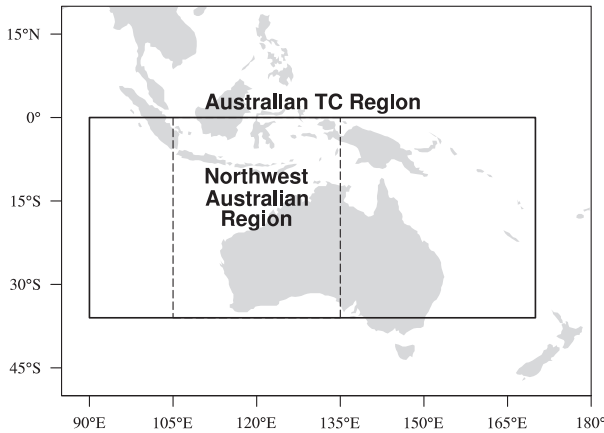


FIG. 1. Map of the Australian region with NWAUS identified between the dashed lines [based on Fig. 1a in Goebbert and Leslie (2010)].

Seasonal prediction of TC activity for particular basins began in earnest with Gray (1984a,b), where he used the QBO, a measure of ENSO, and Caribbean sea level pressure anomalies to forecast the number of hurricanes, the number of hurricanes and tropical storms, and the number of hurricane days for the upcoming season. Using a multiple linear regression (MLR) scheme, Gray (1984b) produced successful hindcasts of the TC metrics. Since that time, seasonal prediction of TC activity has increased substantially. The methods of prediction, the number of TC basins with seasonal prediction of TC activity, and the number of TC metrics being predicted have all grown (see Camargo et al. 2007a; Camargo and Barnston 2009).

There has also been increased development of ensemble approaches to seasonal TC activity (e.g., Thorncroft and Pytharoulis 2001; Chen and Lin 2013; Villarini and Vecchi 2013) with most of these predictions using an ensemble of dynamic model output in which they track TC activity for a variety of lead times and forecast lengths. A purely statistical ensemble approach to seasonal prediction of TC activity was attempted for the western North Pacific Ocean basin by Kwon et al. (2007). The authors successfully implemented an ensemble prediction scheme that used three different sets of predictors to produce the ensemble variability.

For the Australian region, seasonal prediction of TC activity was first proposed by Nicholls (1979), and that prediction scheme used winter (June–August) sea level pressure in Darwin, Australia (essentially SOI), to predict the following season’s number of TCs, which begins in November. Nicholls (1985) made changes to the initial seasonal forecasts by accounting for changes in the reported TC database and adding forecasts for the

number of TC days. The use of SOI was operationalized for the Australian region by Nicholls (1992), who also accounted for changes in the historical Australian TC database by using a first difference method to predict the change in TC activity from the previous year. Nicholls (1992, 1999) verified that seasonal prediction of Australian TC activity using the SOI method was an improvement over climatology and persistence forecasts.

More recently, Wijnands et al. (2015) summarized ongoing efforts to improve the seasonal prediction of TCs in the Australian region through the use of machine-learning algorithms, specifically support vector regression. They found that the use of this method, along with bias-corrected and accelerated bootstrapped confidence intervals, provided the most skillful forecasts. Additionally, since at least 2009, the Australian Bureau of Meteorology seasonal TC outlooks have included probabilistic chances of above- or below-average numbers of TCs for the entire Australian region; the separate western, eastern, and northern regions; as well as the northwestern subregion (BoM 2017).

This paper builds on the work of Goebbert and Leslie (2010) and investigates the success of a single deterministic MLR scheme and a new statistical MLR ensemble approach for producing seasonal forecasts of TC frequency, TC days, and ACE. To assess the skill of a seasonal prediction scheme it is considered best practice to employ more than one measure (Camargo et al. 2007a). Therefore, the skill of the deterministic prediction scheme is assessed using mean absolute error (MAE), Pearson r correlation, and 3×3 contingency tables. The skill of the ensemble prediction scheme is evaluated using the rank probability score, rank probability skill score, and multicategory reliability diagrams (Hamill 1997). Forecasts using the two methods show clear skill over the standard climatological forecast and perform as well or better than other seasonal TC prediction schemes for the NWAUS basin. In addition, the new ensemble statistical approach offers a simple and computationally inexpensive method for producing high quality, robust seasonal forecasts.

2. TC and predictor data

Australian region observations of TCs date back to the nineteenth century; however, the reliability of the Australian region TC data only dates back to 1970 (Holland 1981) with the onset of the satellite era in the region. In the Australian region a TC is defined when 10-min sustained winds reach 17 m s^{-1} (33 kt; where $1 \text{ kt} = 0.51 \text{ m s}^{-1}$), and additionally, the severity can be broken down into five different categories (BoM 2010). There are a number of TC databases that cover the

TABLE 1. Predictors used for TC frequency, TC days, and ACE. The sign indicates the predictor's impact on the next season's forecasted TC metric.

Predictors	Description	Location
TC frequency		
NA700 (−)	April–June 700-hPa geopotential height over North America	35°–45°N, 82.5°–102.5°W
HI850 (−)	May–July 850-hPa air temperature over Hawaii	10°–20°N, 132.5°–155°W
SIND850 (+)	May–July 850-hPa geopotential height over the south Indian Ocean	42.5°–54.5°S, 50°–70°E
SATL925 (−)	June–August 925-hPa geopotential height over the South Atlantic Ocean	20°–30°S, 0°–12.5°W
EPAC925 (−)	June–August 925-hPa geopotential height over the east Pacific Ocean	15°–27.5°N, 115°–130°W
TC days		
SPAC100 (+)	January–March 100-hPa <i>v</i> component of the wind over the South Pacific Ocean	40°–65°S, 140°–170°W
NA850 (−)	April–June 850-hPa geopotential heights over North America	30°–50°N, 80°–105°W
SATL1000 (−)	July–September 1000-hPa geopotential heights over the South Atlantic Ocean	5°–45°S, 35°W–10°E
ACE		
AK100 (+)	February–April 100-hPa air temperature over Alaska	42.5°–67.5°N, 105°–155°W
CN150 (−)	April–June 150-hPa <i>u</i> component of the wind over Canada	45°–65°N, 60°–125°W
SEATL1000 (−)	June–August 1000-hPa geopotential heights over the southeast Atlantic Ocean	7.5°–45°S, 20°W–15°E

NWAUS region including those of the Joint Typhoon Warning Center, the Bureau of Meteorology, and the International Best Track Archive for Climate Stewardship (IBTrACS). This paper uses the dataset presented by Harper et al. (2008) for its consistency of observation and overall quality, specifically for the NWAUS region, with the years since that dataset was published added manually by the author. The dataset was developed through a reanalysis effort lead by Woodside Energy Ltd. in an effort “to address concerns regarding the accuracy and consistency of the official BoM database and to also assemble additional critical storm-scale data, such as radius of gales and eye diameter, that were not available from the official database” (Harper et al. 2008). As of the 2015/16 season, the average number of TCs for the NWAUS basin is 5.3, with an average of 40.75 TC days and an average ACE of $27 \times 10^4 \text{ kt}^2$. A summary of many more TC metrics for the NWAUS basin, based on the Harper et al. (2008) dataset, can be found in Goebbert and Leslie (2010).

Potential predictors are chosen from the NCEP–NCAR reanalysis dataset using only the class A variables as they are strongly influenced by observational data and are considered to be the most reliable (Kalnay et al. 1996). The predictors are chosen as geographical regions with areal extents greater than 5° latitude \times 5° longitude that have high correlation (greater than 0.5 in magnitude) to the TC metric over at least a 3-month period, similar to that of Klotzbach and Gray (2004). The NCEP–NCAR reanalysis variable (e.g., air temperature at 1000 hPa) is averaged over that geographic region for a 3-month period to compose the yearly time series that will be used as a predictor. Goebbert and Leslie (2010) offer a detailed discussion on the use of NCEP–NCAR data for seasonal prediction in the

NWAUS region rather than typical climate indices such as SOI, which are commonly used for seasonal prediction in the Australian region.

3. Predictor selection and model development

a. Deterministic prediction

Potential predictors are used to develop a prediction equation using the MLR statistical method. Multiple linear regression schemes have been previously developed for other TC basins (e.g., Klotzbach and Gray 2003, 2004) and are a relatively simple approach to seasonal TC prediction. A major issue for MLR schemes is that the choice of predictors can overfit the predictand. To reduce the impact of this issue, potential predictors were removed if they had a cross correlation of greater than magnitude 0.3 with another predictor. If such cross correlation occurred, the predictor with the higher correlation to the TC metric was retained. Then, the predictors were further narrowed through a stepwise regression to minimize the number of predictors and maximize the variance explained by those predictors using the Akaike information criteria (AIC). This is done to produce the best compromise between a well-fit model while using the fewest number of predictors to describe the variance of the predictand.

The selection of predictors for each TC metric is completed using the same method. A summary of the selected predictors for the various TC metrics is found in Table 1. Goebbert and Leslie (2010) give a thorough discussion of the variability of NWAUS TC activity in relation to potential predictors including known climate indices and NCEP–NCAR reanalysis data. In general, the predictors chosen describe variations in the global

circulation pattern that relate to the variability of the TC metric being predicted.

Many of the chosen predictors correlate to some extent with one or more of the established climate indices. For example, the NA700 TC frequency predictor (Table 1) has a weak correlation with the Arctic Oscillation ($r = 0.33$), a measure of ENSO (Niño-4; $r = 0.34$), and the QBO ($r = -0.33$). Other variables do not have any strong correlation among common climate indices. For example, the SPAC100 TC days predictor (Table 1) does not correlate highly with any common climate index with a coefficient greater than magnitude 0.3. At least one predictor for each TC metric has a correlation of greater than magnitude 0.3 with an ENSO climate index. Further work on understanding the physical relationships between the predictors and known global atmospheric patterns is needed.

To accurately assess the skill of a potential prediction equation, it is imperative that the predicted year not be a part of the development dataset. This is easily accomplished by removing that year from the development dataset, then making the hindcast prediction for that year. However, TC frequency is autocorrelated from year to year with a 1-yr lag correlation greater than 0.3 for the NWAUS region. This autocorrelation could artificially inflate the accuracy of the prediction equation. Therefore, instead of leaving out just the year that is being predicted, an additional year is removed on either side of the predicted year (e.g., if the prediction for 1981 is desired, the development data would exclude the years 1980, 1981, and 1982), which will be referred to as the leave-three-out hindcast method in this paper (Elsner and Schmertmann 1994). For a limited dataset this provides some degree of data independence when constructing the prediction equation, in order to account for the autocorrelation present in TC occurrence. Even with this attempt to remove the influence of autocorrelation to more accurately assess the skill of the prediction equation, there likely remains artificial skill due to using all of the years in the development dataset in choosing the initial list of predictors from reanalysis correlations.

Another important consideration to make when developing an MLR prediction equation involved the number of years of development data needed to adequately describe the variance of the predictand. There are 46 years' worth of available data to draw from for the development of the prediction equations. The root-mean-square error (RMSE) is employed to measure the skill of the prediction scheme using different sizes of the development dataset. Using the leave-three-out hindcast method and the RMSE, the number of years used in the development dataset was varied between 5

and 35 years to determine the minimum number of years needed to provide a stable prediction (Fig. 2). Having too few years in the development dataset provides insufficient variance to the MLR scheme to accurately hindcast the variability of the predictand. For the prediction of the number of TCs, there was a sharp decrease in RMSE as the development length increased, but this pattern eventually leveled off when the development dataset contained 28 or more years (Fig. 2). Therefore, in this study the development dataset for TC frequency contains the 28 years from 1970 to 1997. A similar pattern exists for TC days and ACE, which yielded development datasets of 32 and 25 years, respectively.

b. Ensemble prediction

Ensemble prediction has been around for a long time and used in many different fields; in meteorology it has primarily been used in conjunction with numerical weather prediction models. There has been limited use of ensemble approaches for the seasonal prediction of TCs, and those that have been created primarily used dynamically based ensembles (e.g., Belanger et al. 2010, 2012), which count the number of TCs occurring during a season through the multiple deterministic model runs. A statistical ensemble approach was introduced by Kwon et al. (2007) using a large set (~ 50) of geographically averaged reanalysis variables (e.g., 500-hPa geopotential heights, 850-hPa temperature) as predictors. These predictors were then placed into three groups, and within each group 10 predictions were made to constitute the ensemble. The prediction equations were then simple regression equations between the predictand and the predictor. This ensemble approach improved the forecast ability of the deterministic predictions of Lee et al. (2007), from which the ensemble was developed.

This study follows the approach of Frank and Pfahringer (2006), who generated an ensemble by both subsampling the development years from the broader set of data with replacement (bagging) and adding noise to the input values (input smearing). The primary goal of bagging and smearing is to provide the ensemble prediction scheme with sufficient dispersion to obtain a well-calibrated prediction method. First, the 30-yr predictor set (1970–99) is resampled with replacement to generate a bagged set of predictors. Second, the training values are smeared (a_{smeared}) by adding Gaussian noise to each variable a_{original} :

$$a_{\text{smeared}} = a_{\text{original}} + p \times N(0, \sigma_a), \quad (1)$$

where σ_a is the standard deviation of the original variable from the predictor set, N is the Gaussian normal distribution, and p is a specified smearing parameter.

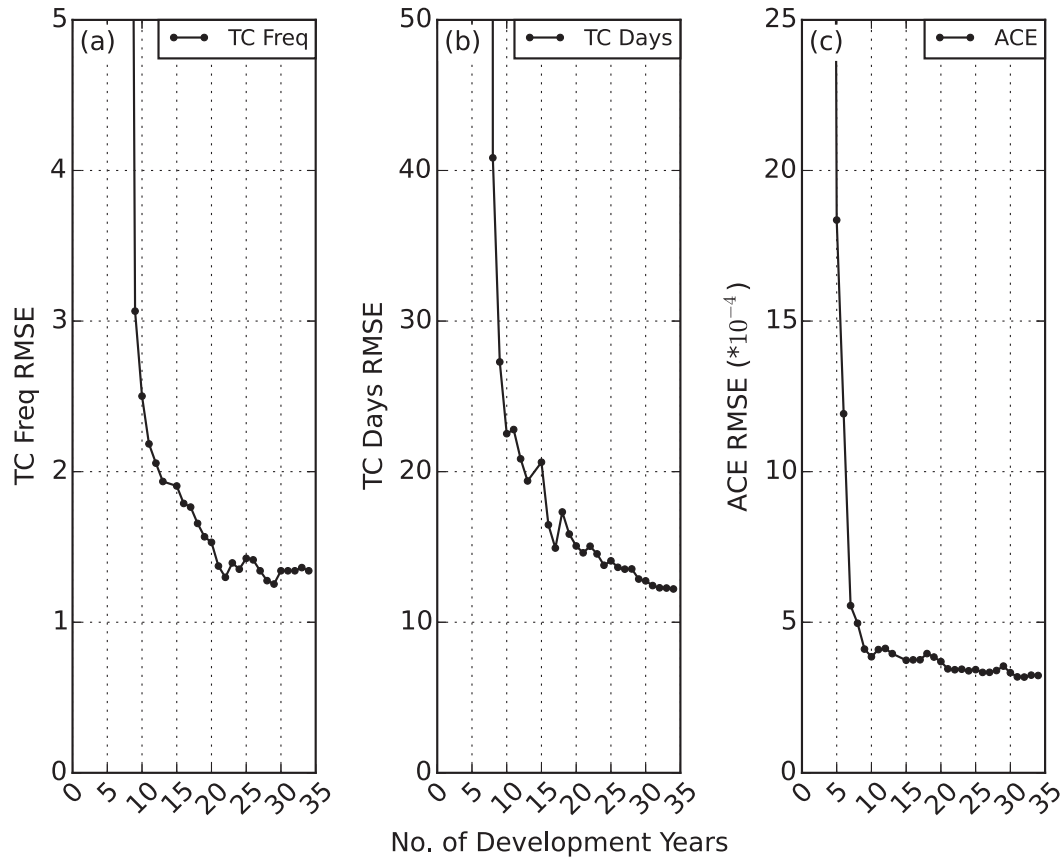


FIG. 2. A plot of the number of years used in the development prediction scheme and its reported RMSEs for (a) TC frequency, (b) TC days, and (c) ACE.

This is done for each development predictor variable and for all prediction equations. Additionally, the predictor variables for a given year are also smeared according to Eq. (1) before being used to make a prediction. The selection of the smearing parameter can be completed by a systematic grid search over a range of values (Frank and Pfahringer 2006) to ensure that adequate dispersion is achieved when selecting the value.

The determination of the smearing parameter p in Eq. (1) was found empirically by systematically searching smearing values [similar to the process of Frank and Pfahringer (2006)] from 0.5 to 1.5 to obtain a value that produced a well-calibrated prediction scheme by providing adequate dispersion to the ensemble system. A well-calibrated ensemble method with appropriate levels of dispersion can be demonstrated by the forecasted quantile falling along the unity line of a multi-category reliability diagram (Hamill 1997). Additionally, the calibration can also be measured by the sum of the absolute distance from each forecast quantile to the quantile of the forecast distribution (i.e., the distance the calibration is from the line $y = x$). As a result, there

was usually a range of smearing parameter values that had similar category errors and were well calibrated; as a result, the lowest value was retained as the smearing parameter value.

The ensemble prediction scheme for TC frequency is used as an example to illustrate the work flow of this process (Fig. 3). For ensemble prediction of TC frequency, the development dataset is set to 36 years (1970–2005), from which 28 years are chosen (with replacement; bagged) and then those bagged predictors are smeared to determine the MLR prediction equation coefficients. For hindcasting the years 1970–2005, the leave-three-out prediction method is used just as in the single MLR scheme. This procedure is repeated 100 times to create an ensemble of prediction equations, all with varying coefficients for the MLR equations. Then, the predictors for the year being predicted are first smeared, then input into each of the equations in the ensemble, resulting in many (e.g., 100) predictions of TC frequency. This ensemble approach yields a forecast that provides users some measure of uncertainty for the forecast of TC frequency. This same procedure can be followed to

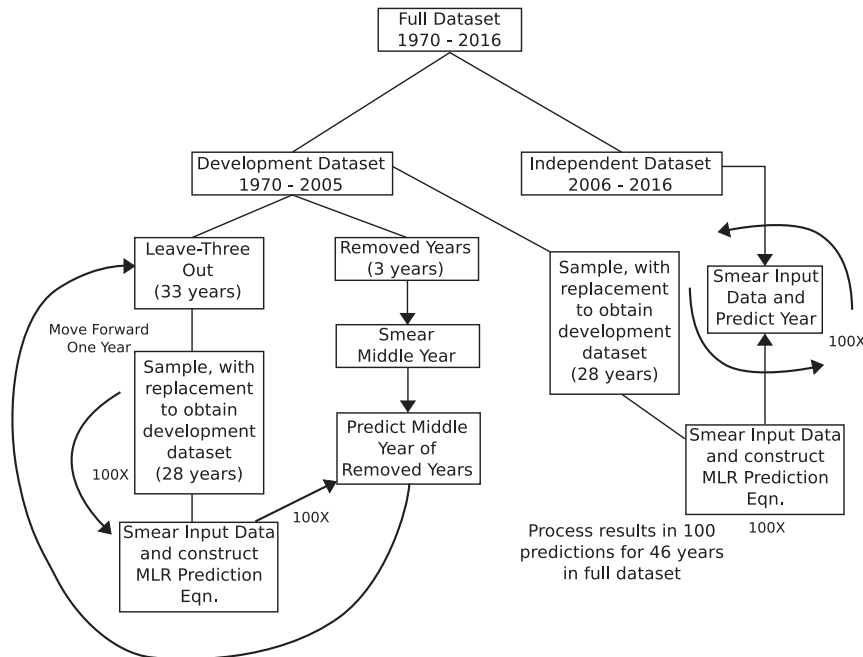


FIG. 3. A flowchart representing the workflow of developing the statistical ensemble predictions of TC metrics.

predict any TC metric. The remaining years (2006–16) are predicted using the same method and are independent of the development dataset.

4. Results and discussion

a. Deterministic prediction

Through the leave-three-out method, deterministic hindcast predictions were made for TC frequency from 1970 to 1998 and predictions from the stable developmental dataset from 1999 to 2016 (Fig. 4a). The deterministic MLR forecast equation for TC frequency that results from the 28 years of development data is

$$\begin{aligned} TC_{\text{freq}} = & 5.5714 - 0.8842 \times NA700 - 0.5762 \\ & \times HI850 + 0.5867 \times SIND850 - 0.8045 \\ & \times SATL925 - 0.4607 \times EPAC925, \end{aligned} \quad (2)$$

where the five predictors are described in Table 1. The forecasted TC frequency appears to align well with the observed values over the entire dataset, with a Pearson correlation coefficient of $r = 0.83$ and an MAE of 1.04. This is an improvement over using climatology or persistence as the forecast for the region, which have MAE values of 1.80 and 2.22, respectively. The skill score of the deterministic MLR forecast can be calculated as an improvement over a reference forecast (e.g., climatology) through the following equation:

$$SS = 1 - \frac{MAE}{MAE_{\text{Ref}}}, \quad (3)$$

where MAE_{Ref} is the reference forecast MAE (Wilks 2005). For the deterministic TC frequency prediction there is a 42% improvement over a climatological forecast (Table 2).

The MAE values and subsequent skill score are based on 46 years of hindcasts and may not represent the true statistic. Utilizing a bootstrap technique based on Efron and Tibshirani (1994), a 95% confidence interval can be calculated to verify that the MAE from the deterministic TC prediction scheme is significantly different from the MAE of the climatological prediction. The confidence interval obtained using the bootstrap technique, with 5000 replications, yields an MAE range from 0.84 and 1.24 for the deterministic model and from 1.42 to 2.22 for a climatological forecast. With no overlap in the confidence intervals, the two predictions come from different populations, and the deterministic forecasts of TC frequency are significantly better than using climatology as the forecast. This is further confirmed by a standard two-sided t -test statistic computed between the absolute error of the deterministic and climatological predictions. Doing so yields a test statistic of -3.23 and a p value of 0.002, indicating that the two population means are significantly different at the 1% level.

For the deterministic TC days prediction, hindcasts were made from 1970 to 2004, with independent predictions from 2005 to 2016 (Fig. 4b). The deterministic MLR

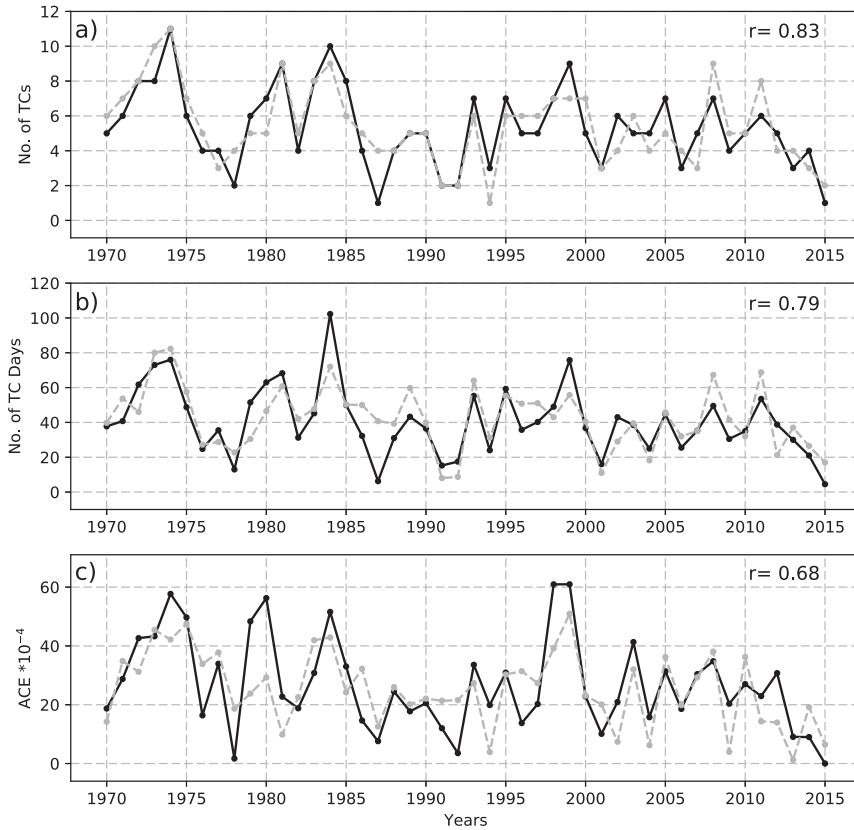


FIG. 4. Observed (black) and forecasted (dashed gray) TC metrics for the NWAUS basin from 1970 to 2015 with Pearson’s r correlation coefficients between the deterministic forecast and the actual observations for (a) TC frequency, (b) TC days, and (c) ACE.

forecast equation for TC days that results from the 32 years of development data is

$$TC_{\text{days}} = 43.648 + 7.431 \times SPAC100 - 7.930 \times NA850 - 10.847 \times SATL1000, \quad (4)$$

where the three predictors are described in Table 1. The Pearson correlation coefficient for the deterministic MLR prediction of TC days was $r = 0.79$, and the MAE was 9.92 days with the climatological prediction MAE being 14.78 days, which yielded a skill score for the deterministic forecasts of 33% (Table 2). The 90%

confidence intervals of MAE values for the deterministic MLR range from 8.33 to 11.83 and for climatological forecasts from 12.14 to 18.28. With no overlap in the confidence intervals, the two predictions come from different populations, and the deterministic forecasts of TC days are significantly better than using climatology as the forecast, although at a lower level of confidence than for the predictions of TC frequency. A two-sided t -test statistic between the deterministic and climatological forecasts was -2.25 and had a p value of 0.03, indicating that the two population means are significantly different at the 5% level.

TABLE 2. Summary of MAEs for the deterministic and ensemble prediction equations for TC frequency, TC days, and ACE ($\times 10^{-4}$). Additionally, the MAEs for forecasts of climatology, persistence, and the skill score (SS) of the deterministic forecasts against the reference forecast of climatology are also given for each TC metric studied.

	MAE by forecast method			SS _{Climo}
	Deterministic	Climatology	Persistence	
TC frequency	1.04	1.80	2.22	0.42
TC days	9.92	14.78	19.46	0.33
ACE ($\times 10^{-4}$)	9.57	12.69	15.06	0.25

TABLE 3. The 3×3 contingency table of deterministic and ensemble forecasts of TC frequency, where below average is a season with fewer than five TCs, an average season is five or six TCs, and an above-average season has seven or more TCs. Ensemble forecasts are presented as the percentage of ensemble members occurring in each category.

Forecast	Deterministic forecast			Ensemble forecast		
	Observed			Observed		
	Below avg	Avg	Above avg	Below avg	Avg	Above avg
Below avg	5	3	0	11	9	1
Avg	4	16	5	8	31	10
Above avg	0	4	9	0	10	20

Deterministic ACE prediction used the leave-three-out method to hindcast the predictions from 1970 to 1994 and independent predictions from 1995 to 2016 (Fig. 4c). The deterministic MLR forecast equation for ACE that results from the 25 years of development data is

$$\text{ACE} = 28.3330 + 4.6353 \times \text{AK100} - 6.4760 \times \text{CN100} - 5.2515 \times \text{SEATL1000}, \quad (5)$$

where the three predictors are described in Table 1, and the intercept and predictor coefficients are scaled by 10^{-4} . The Pearson correlation coefficient for the deterministic MLR prediction of ACE was $r = 0.68$, and for MAE it was $9.6 \times 10^4 \text{ kt}^2$, with a climatological prediction MAE of $12.7 \times 10^4 \text{ kt}^2$. The ACE skill score for the deterministic forecasts was 25% (Table 2), the lowest of the three predicted TC metrics. The 90% confidence intervals of MAE values for the deterministic MLR range from 8.0×10^4 to $11.3 \times 10^4 \text{ kt}^2$, and the climatological forecasts range from 10.6×10^4 to $15.1 \times 10^4 \text{ kt}^2$. In this case the confidence intervals overlap, and a direct determination cannot be made of whether the hindcast predictions made using the deterministic forecast equation is significantly better than a prediction using climatology. The two-sided t test yields a test statistic of -1.85 and a p value of 0.07, indicating that the two population means are significantly different at the 10% level. Thus, the deterministic ACE prediction equation developed in this study performs better than when using climatology as the prediction.

Forecast skill can also be assessed through forecast contingency tables, and this study utilized 3×3 contingency tables to examine below-average, average, and above-average seasons for given TC metrics. A below-average season is defined as the bottom third of the observed TC metric, an average season is the middle third, and an above-average season is the top third. For TC frequency a season is below average if less than five TCs occur, an average season is the occurrence of five or six TCs, and an above-average season has seven or more TCs (Table 3). For TC days a season is below average if less than 32 TC days occur, an average season is the occurrence of 32–45 TC days, and an above-average season has 45 or more TC days (Table 4). Below-average seasonal ACE is defined as values less than $19 \times 10^4 \text{ kt}^2$, average between 19×10^4 and $31 \times 10^4 \text{ kt}^2$, and above average when values exceed $31 \times 10^4 \text{ kt}^2$ (Table 5).

Categorical forecast skill is assessed via the Pierce skill score (PSS; Wilks 2005), which uses an unbiased random forecast to determine the skill of the forecast. For the deterministic TC frequency predictions the PSS was 0.421, and the reference forecasts of climatology and persistence yielded scores of 0.00 and 0.028, respectively. The deterministic TC days prediction yielded a PSS of 0.443, with a climatological prediction PSS of 0.00 and persistence PSS of 0.033. The deterministic ACE prediction had a lower PSS than either the deterministic TC frequency or TC days predictions at 0.252, and the climatology and persistence scores are 0.00 and -0.033 , respectively.

TABLE 4. As in Table 3, but for TC days where a below-average season is one with fewer than 32 TC days, an average season is 32–45 TC days, and an above-average season has more than 45 TC days.

Forecast	Deterministic forecast			Ensemble forecast		
	Observed			Observed		
	Below avg	Avg	Above avg	Below avg	Avg	Above avg
Below avg	9	3	1	15	6	2
Avg	6	6	1	12	13	8
Above avg	0	6	14	6	13	25

TABLE 5. As in Table 3, but for ACE where a below-average season is a season with less than $19 \times 10^4 \text{ kt}^2$, an average season is from 19×10^4 to $31 \times 10^4 \text{ kt}^2$, and an above-average season is greater than $31 \times 10^4 \text{ kt}^2$.

Forecast	Deterministic forecast			Ensemble forecast		
	Observed			Observed		
	Below avg	Avg	Above avg	Below avg	Avg	Above avg
Below avg	6	6	0	15	11	2
Avg	7	6	4	15	12	12
Above avg	3	4	11	5	9	19

Overall, the deterministic forecasts for TC frequency, TC days, and ACE all performed better than a climatological or persistence forecast. Thus, these types of forecasts would provide valuable guidance to stakeholders in the NWAUS basin on the TC activity in that region before the beginning of the season. Based on the chosen predictors, forecasts can be made for TC frequency and ACE in September (once the August data are available) and TC days in October. The NWAUS basin season runs from November to April, so the predictions give a 1–2-month lead time before the beginning of the season and an even longer time before the average start date, which is Julian day 352 (i.e., 18 December).

The deterministic MLR equations developed and used in this study achieve the same or greater forecast skill compared to other forecasts for the entire Australian basin and other basins across the world. To place the MAEs of the deterministic predictions in perspective, the results are compared to those of Nicholls (1985), Wijnands et al. (2015), McDonnell and Holbrook (2004a,b), and Klotzbach and Gray (2009).

For the entire Australian region (105° – 165°E), Nicholls (1985) constructed a linear regression equation using July–September Darwin sea level pressure and obtained an MAE of 9.4 days, compared to an MAE of 9.74 days for the deterministic MLR scheme used in this study. He obtained similar MAE values for both climatological and persistence forecasts (12.7 and 19.5 days, respectively), to the values obtained for this study (14.37 and 19.8 days, respectively). The skill of these forecasts is likely a little better as a result of only using a leave-one-out cross validation on the hindcast predictions. Additionally, ENSO indices (e.g., Niño-3.4 SSTs, Niño-4 SSTs, Darwin pressure) are more strongly correlated with TC activity over the entire Australian region (Ramsay et al. 2008, see their Fig. 3) than for the western Australian basin (Goebbert and Leslie 2010, see their Fig. 7).

Recently, Wijnands et al. (2015) constructed seasonal forecasts of TCs in various regions of the Australian basin including a region very similar to the NWAUS region (their AR-W region). The scheme that had the

lowest MAE was a seasonal prediction of TC frequency using support vector regression, and they achieved an MAE of 1.27 for their hindcast predictions from 2003 to 2012. The deterministic MLR scheme developed in this paper outperforms that over the entire hindcast range (Table 2), but has a very similar MAE of 1.3 when comparing the same forecast years. One advantage of the deterministic scheme developed here is that a forecast of similar quality (in terms of MAE) can be made at the beginning of September, as opposed to the beginning of November for their scheme, with predictors that contain data only from October.

McDonnell and Holbrook (2004a,b) used a Poisson regression method to predict the occurrence of TCs over the entire Australian region by dividing the region into 2° latitude \times 5° longitude cells and forecasting the occurrence in those cells. While many of their predictions covering the Australian basin were able to improve on a forecast of climatology (see their Table 3), for their western region (105° – 125°E) they were not able to produce successful seasonal predictions of TC activity and saw only slight improvements for their northern region (125° – 145°E), which also covers part of the NWAUS region defined in this study.

The deterministic prediction equations developed in this study obtained skill on par with or better than similarly developed prediction equations for other basins. For example, the verification statistics (1984–2008) for the forecasts for the North Atlantic Ocean (Klotzbach and Gray 2009) yielded a skill score over climatology of 29% for named tropical storms and 15% for named tropical storm days for their June forecasts of the oncoming TC season. Klotzbach and Gray (2009) have also issued forecasts in early August from 1984 to 2008, updating their previous forecasts. These forecasts performed slightly better than the June forecasts as the calculated skill scores for named tropical storms was 31% and for named storm days it was 22%.

b. Ensemble prediction

Using the ensemble approach for the seasonal prediction of TCs allows for probabilistic forecasts of TC metrics to provide a more thorough evaluation of the

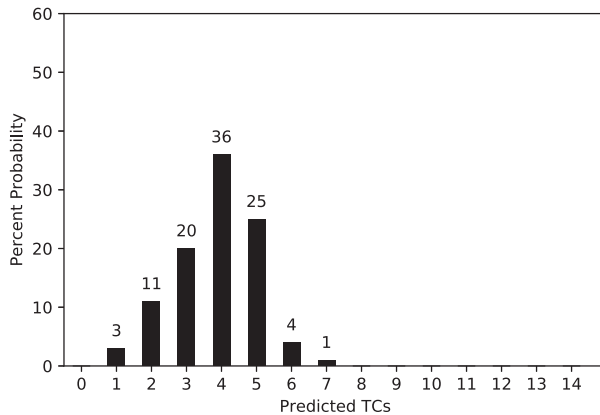


FIG. 5. Bar chart of the 100 ensemble predictions of TC frequency for 1987 with the number of times the ensemble scheme predicted the individual values plotted above the bar.

forecasted TC metric. Instead of a single forecast for TC frequency (e.g., a prediction of four TCs for the upcoming season), an ensemble with 100 members can yield probabilistic information about the number of TCs that will occur in a season. For example, an ensemble might have 48 members out of 100 that predict four or more TCs for the upcoming season. This potentially aids the forecast end user by offering more information about a range of forecasted values, similar to other ensemble forecasts.

An example of a seasonal TC frequency forecast from the ensemble method used in this paper is illustrated in Fig. 5 for the TC season from November 1987 to April 1988. This season had a large range in the number of predicted TCs, from as few as one to as many as seven. The forecast can then be expressed as a percent chance of a given number of TCs for the 1987 season (e.g., a 36% chance that there will be four TCs). An end-user then has a more clear understanding of the inherent uncertainty of the forecasted value.

Similarly, the forecast can also give a probabilistic value to whether it is expected that the seasonal TC metric will be below average, average, or above average. For the 1987 TC season (Fig. 5) there is a 70% chance that the season will have a below-average number of TCs (four or fewer TCs), a 29% chance that there will be an average number of TCs (five or six TCs), and a 1% chance that there will be an above-average number of TCs (seven or more TCs). There was only one TC that occurred during the 1987 NWAUS TC season, which verifies the 70% chance that the season will have a below-average number of TCs.

Despite the fact that only three of the ensemble members predicted only one TC occurring in 1987 (Fig. 5), the scheme correctly identified the fact that

there was a very high likelihood that there would be a below-average number of TCs. It should not be too surprising that the scheme did not have a large number of predictions for a single TC occurrence as that has only happened once in the development dataset (1987), and the very nature of the MLR scheme makes it difficult to predict the extreme values that are rarely (or not otherwise) observed in the dataset, as that is where assumptions under which MLR schemes operate (e.g., normality, linearity) begin to break down.

The ensemble predictions of TC frequency result in some years having larger predictive ranges and others having much smaller ranges (Fig. 6) over the 46 years of hindcasts with the ensemble prediction system. For TC frequency the systematic search for an appropriate smearing parameter yielded a value of 0.65. The calibration of the scheme appears good (Fig. 7a), with an average category error of 1.33, except for the lowest and highest quintiles. This is likely inherent from using MLR, which makes predicting extreme values difficult. This is better than the climatological forecast, which while well calibrated had an average category error of 2.44 (Fig. 7b).

To assess the skill of ensemble predictions, an appropriate scheme for scoring probabilistic predictions is needed. The rank probability score (RPS; Epstein 1969; Murphy 1969, 1971) is one such scheme that can be applied to ordered, multicategory predictions and is used here as defined in Weigel et al. [(2007), see their Eq. (1)] for both the ensemble probabilistic predictions and reference predictions using climatology. Then, from the RPS, a rank probability skill score (RPSS; Weigel et al. 2007) can be computed for the ensemble predictions of TC frequency, TC days, and ACE in the same fashion as the skill score using MAE with Eq. (3).

The ensemble TC frequency prediction scheme obtained an RPS of 2.33, with the climatological probabilistic prediction garnering an RPS of 3.87 (Table 6). This yields an RPSS for the ensemble prediction over a climatological prediction of 0.40, indicating an improvement in forecast skill of 40% for the ensemble prediction scheme developed in this paper. Similarly, the skill of the ensemble prediction scheme can be made based on a three-category prediction of below-average, average, and above-average TC frequency. The calculated RPS of the ensemble prediction system is 3.88 whereas the climatological prediction yields an RPS of 10.0, and the ensemble prediction system improves on climatology by 61%.

To compare the ensemble and deterministic predictions, the deterministic prediction can be assumed to be a probabilistic forecast with all 100% predicted in a single category (e.g., a prediction of five TCs) and then

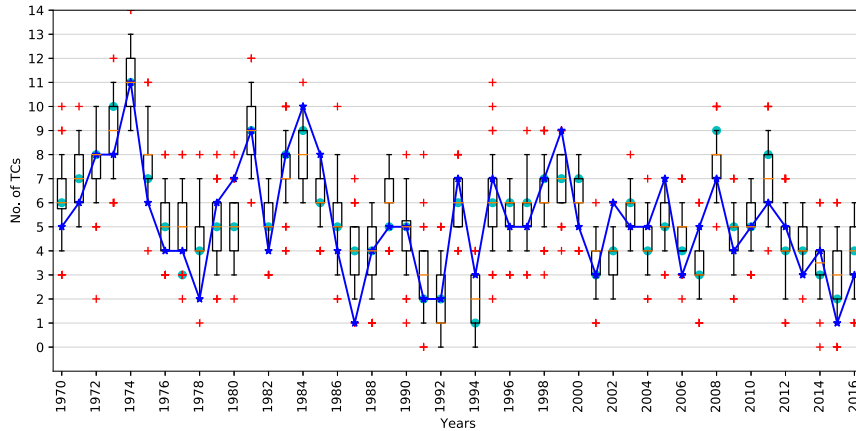


FIG. 6. Ensemble predictions of TC frequency from 1970 to 2016 with the ensemble predictions represented as box-and-whisker plots of each year, where the box represents the inner quartile range, and the whiskers are at 5% and 95% of the forecasted members. Those forecasts falling outside of the middle 90% are considered outlier forecasts and are labeled with red plus (+) signs. The deterministic forecasts are cyan-colored circles, and actual observations are labeled as blue stars connected by a line.

measured using the RPS. Doing so yields an RPS of 3.2 for the deterministic prediction of TC frequency and an improvement over the climatological prediction of 17% (Table 6). Comparing the ensemble and deterministic predictions of TC frequency with this metric indicates superior performance of the ensemble system.

The ensemble predictions for TC days performed well over the 46 years of hindcast predictions (Fig. 8), using a smearing parameter of 0.75, which produces a well-calibrated forecast (Fig. 9a) with an average category

error of 2.47 compared with the climatological prediction average category error of 3.58 (Fig. 9b). TC days were assessed in bin lengths of 5 days (e.g., 0–5, 5–10, and 15–20), which yields approximately 21 forecast categories. As with the TC frequency ensemble predictions, there is a noticeable difference in the spread of the ensemble in different years. There were some years (e.g., 2008) where the actual number of TC days for that year fell outside of the ensemble spread (Fig. 8); the trend in the forecast is in generally good agreement with the observations. The ensemble TC days

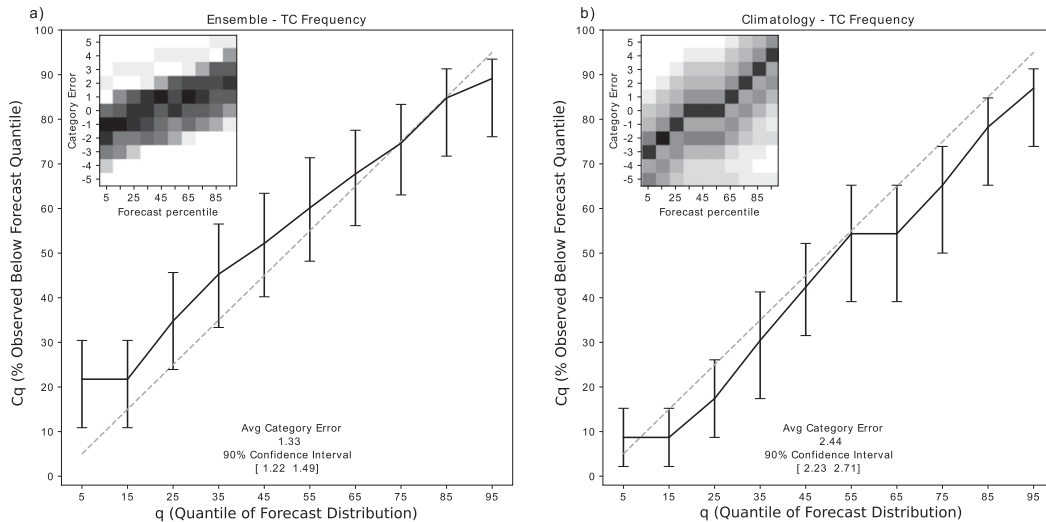


FIG. 7. Multicategory reliability diagram for (a) ensemble seasonal prediction of TC frequency and (b) climatological probabilistic prediction of TC frequency. The dashed gray line represents the perfect forecast line. Error bars represent the 90% confidence interval of the forecast as determined by use of a conventional bootstrap approach. Insets are the frequencies of category error for each forecast/observed quantile pair with darker shading representing larger percentages of frequencies in that cell.

TABLE 6. Summary of RPS results for the ensemble and deterministic prediction equations for TC frequency, TC days, and ACE. Additionally, the RPS results for forecasts using climatology are given as well as the RPSSs of the ensemble forecasts against the reference forecast for each TC metric studied.

	RPS by forecast method			RPSS _{Climo}
	Ensemble	Deterministic	Climatology	
TC frequency	2.33	3.20	3.87	0.40
TC days	2.85	3.75	4.03	0.30
ACE	2.97	4.09	3.75	0.21

prediction had an RPS of 2.85 categories, and the climatological prediction had an RPS of 4.09 categories (Table 6). This yields an improvement over climatological prediction of 30% by using the ensemble prediction scheme for TC days, whereas the deterministic prediction RPS was 3.75 categories and only a 7% improvement over the climatological prediction.

Similarly, the number of categories can be reduced to three (below average, average, above average), and the RPS assessed for this reduced category prediction. The calculated RPS of the three-category ensemble prediction scheme was 2.94 categories whereas the climatological prediction yields an RPS of 8.33, and the ensemble prediction system improved on climatology by 54%. The RPS of the deterministic prediction scheme was 5.67 for the three-category prediction and improved on the climatological prediction by 43%.

The ensemble predictions for ACE also performed well for all 46 years (Fig. 10), using a smearing parameter of 0.85, which produces a well-calibrated forecast (Fig. 11a) with an average category error of 1.18 compared with the climatological prediction average category error of 1.94 (Fig. 11b). However, the calibration is not as good overall for ACE as it was for TC frequency and TC days. The predictions resulted in RPS values of the ensemble that were similar to the deterministic predictions (Table 6) but generally had larger ranges in

the predicted values for any given year compared to the predictions of TC frequency or TC days. Similar to the TC days prediction, the prediction scheme is assessed categorically using a bin lengths of 10 (e.g., 0–10, 10–20), when the predictions are scaled by 10^{-4} . The RPS for the ensemble ACE forecasts was 2.97 categories (Table 6), with 3.75 categories for the climatology, with the ensemble producing a 21% improvement over the reference forecast of climatology. The deterministic prediction RPS was 4.09 categories and is not an improvement over the climatological forecast.

Using only the three category bins of below average, average, and above average yields substantially better results for the ensemble ACE prediction scheme. The three-bin ensemble prediction yielded an RPS of 2.07 categories; the climatology prediction had an RPS of 12.0, which is an improvement over climatology of 83%.

The ensemble TC frequency scheme can be compared to the probabilistic forecasts produced by the Australian BoM for the seasons 2009–16 (Table 7). The ensemble scheme made accurate predictions in 7 out of the 8 yr, only missing the forecast in 2012 when the actual number of TCs was five. This was not a particularly bad prediction as the scheme gave nearly equal chances to above- or below-average TC frequencies, 48% and 52%, respectively. The forecasts from the Australian BoM were correct in 4 out of the 8 yr, again with a relatively

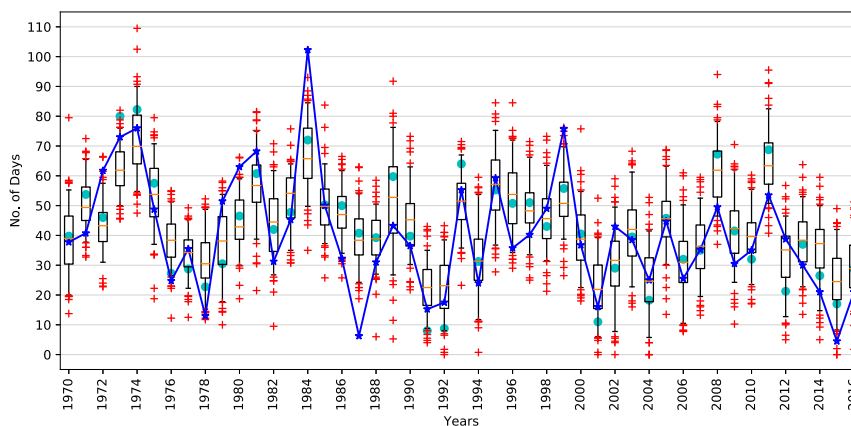


FIG. 8. As in Fig. 6, but for TC days.

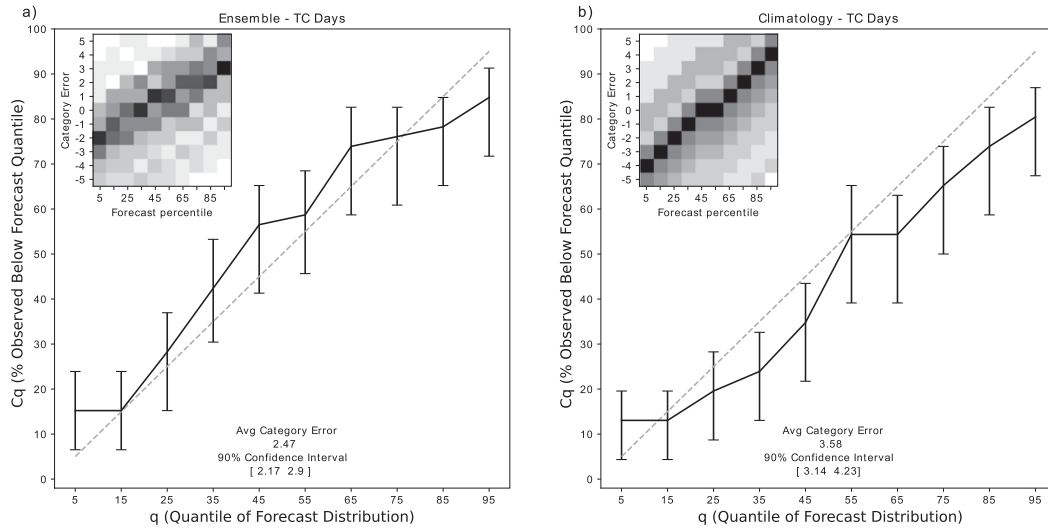


FIG. 9. As in Fig. 7, but for TC days ensemble prediction.

good prediction for 2012 and for other missed forecasts except for 2016. In general, the ensemble scheme appears to have a more dynamic range in predicted percentages when compared to the BoM forecasts, and the forecast percentages appear to better reflect the observed TC frequency.

The support vector regression scheme produced by Wijnands et al. (2015) also created 90% confidence intervals through bootstrapping to give a range of forecast values, similar to an ensemble prediction. For their western region the observed TC frequency fell outside of their forecast intervals in two of the eight predicted years. The ensemble scheme developed in this paper had a slightly better record over the same period with only one forecast falling outside of its 90% confidence interval. Additionally, these two schemes appear to be of similar quality with MAEs of 1.33 (Fig. 7) for the

ensemble scheme and 1.27 for the support vector regression scheme. An advantage of the ensemble scheme is the forecast can be produced 2 months earlier than that of Wijnands et al. (2015), allowing for more time to prepare before the onset of the TC season.

There is some level of uncertainty associated with any prediction, including statistical seasonal prediction of TCs. While traditional ensemble methods have existed for a long time, they have not been routinely used in statistical prediction methods, such as the seasonal prediction of TCs for a particular TC basin. The ensemble method developed in this paper yields forecasts that have an appropriate spread of forecasts, while producing forecasts that are a substantial improvement over the classic reference forecast of climatology. Using the RPS metric, it is also clear that the ensemble method is an improvement over the deterministic approach and

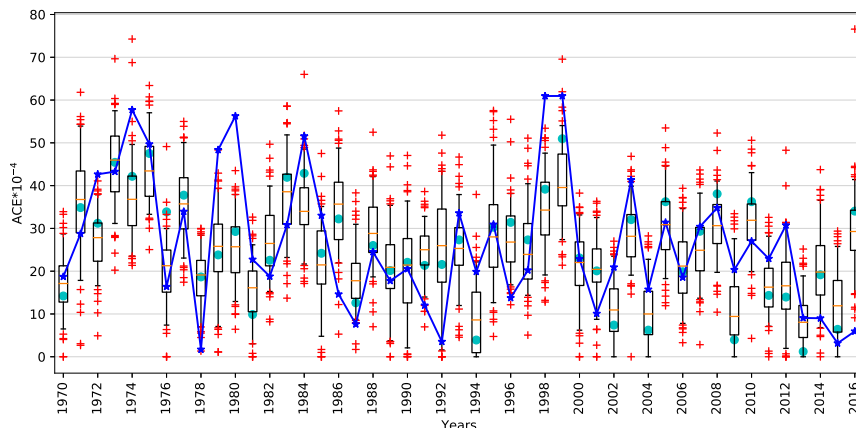


FIG. 10. As in Fig. 6, but for ACE.

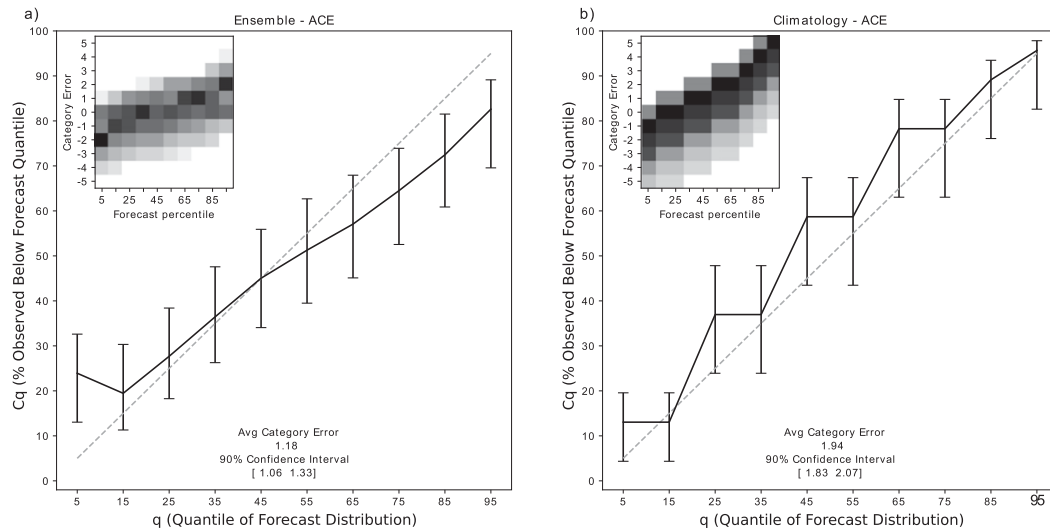


FIG. 11. As in Fig. 7, but for ACE ensemble prediction.

conveys more information, which can aid end users in their decision-making processes.

5. Conclusions

Seasonal prediction of TC activity has been ongoing in the Australian region since the late 1970s, but the NWAUS basin has received relatively little attention as it is usually aggregated into the entire Australian region or a larger subbasin prediction. This paper described a deterministic MLR seasonal prediction scheme for the NWAUS basin that has skill in predicting TC frequency and TC days, with lower skill for ACE. Bootstrapped MAE results indicate that the forecasts for TC frequency and TC days were statistically better than a climatological forecast for the basin at the 95% and 90% levels, respectively. While the confidence interval of MAE for ACE slightly overlaps with the climatological prediction confidence interval, further statistical tests indicated that the prediction was from two different populations, and therefore, the deterministic MLR method was an improvement over a climatological forecast.

Additionally, a new statistical ensemble approach to forecasting TC activity in the NWAUS region has been presented. In this approach, 100 deterministic MLR equations were developed by bagging and smearing the development dataset; these equations were then used to forecast the TC metrics. The forecast skill of the ensemble approach was higher than that of the deterministic prediction (as measured by the RPS), and all TC metric prediction schemes performed better than climatology. The added benefit of the ensemble prediction was that the usually implied uncertainty in a seasonal forecast now has some direct uncertainty

values associated with a particular forecast. The result of the ensemble prediction gives a range of possible values for the forecasted TC metrics, which inherently gives valuable information about the confidence of the prediction.

The challenge in assessing the skill of these seasonal predictions was the inherent limitation due to using a small dataset (46 years). However, that concern was taken into account through using the leave-three-out method to hindcast the years in the development dataset. Despite using the leave-three-out method, there likely remains a certain amount of artificial skill due to the method of selecting the predictor variables and by choosing a smearing parameter that ensured adequate model dispersion. Despite these caveats, the techniques described in this paper appear to be a substantial improvement over using climatology as the basis for forecasting TC activity in the basin.

TABLE 7. Comparison of probabilistic TC forecasts from Australian BoM and the ensemble seasonal prediction scheme for the NWAUS region. The forecast is the percent chance of at least an average number of TCs occurring in a season. Correct forecasts are set in boldface.

Year	Avg No. of TCs	BoM (%)	Ensemble (%)	No. of observed TCs
2009	6	44	24	4
2010	6	67	40	5
2011	5	60	99	6
2012	5	42	48	5
2013	5	55	39	3
2014	5	38	18	4
2015	5	15	9	1
2016	5	63	34	3

In general, the forecasts produced by the deterministic and ensemble schemes performed as well, if not better, than other prediction schemes for the same basin. The advantage of the schemes developed here is in the simplicity of the statistical technique coupled with the ability to forecast the TC metric one or two months prior to the beginning of the TC season. Limitations of the prediction schemes include the uncertainty in our physical understanding of the individual predictors as well as their longevity. Initial analysis of the physical understanding of the predictors, through simple correlation analysis, reveals the fact that predictors are related to some common climate indices, but are likely reflective of combinations of atmospheric patterns not well captured by a single index value. The longevity of the predictors will only be known as the length of the record increases. Overall, the deterministic and ensemble methods provide simple high quality forecasts, produced in advance of the beginning of the NWAUS TC season by one or more months, allowing end users an ample amount of time to utilize the forecasts to their fullest extent.

Acknowledgments. The manuscript was greatly improved thanks to suggestions made by Drs. Bradford Barrett, Lance Leslie, Craig Clark, and Teresa Bals-Elsholz. I would also like to thank three anonymous reviewers whose suggestions have greatly improved this manuscript. Funding for the publication of this manuscript came from the Office of the Dean, College of Arts and Sciences, Valparaiso University, as well as the Department of Geography and Meteorology.

REFERENCES

- Ballenzweig, E. M., 1959: Relation of long-period circulation anomalies to tropical storm formation and motion. *J. Meteor.*, **16**, 121–139, [https://doi.org/10.1175/1520-0469\(1959\)016<0121:ROLPCA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1959)016<0121:ROLPCA>2.0.CO;2).
- Belanger, J. I., J. A. Curry, and P. J. Webster, 2010: Predictability of North Atlantic tropical cyclone activity on intraseasonal time scales. *Mon. Wea. Rev.*, **138**, 4362–4374, <https://doi.org/10.1175/2010MWR3460.1>.
- , P. J. Webster, J. A. Curry, and M. T. Jelinek, 2012: Extended prediction of north Indian Ocean tropical cyclones. *Wea. Forecasting*, **27**, 757–769, <https://doi.org/10.1175/WAF-D-11-00083.1>.
- BoM, 2010: Bureau of Meteorology tropical cyclone intensity scale. Accessed 28 February 2017, <http://www.bom.gov.au/cyclone/faq/>.
- , 2017: Australian tropical cyclone seasonal outlook archive. Accessed 26 June 2017, <http://www.bom.gov.au/climate/cyclones/australia/archive.shtml>.
- Camargo, S. J., and A. G. Barnston, 2009: Experimental dynamical seasonal forecasts of tropical cyclone activity at IRI. *Wea. Forecasting*, **24**, 472–491, <https://doi.org/10.1175/2008WAF2007099.1>.
- , —, P. J. Klotzbach, and C. W. Landsea, 2007a: Seasonal tropical cyclone forecasts. *WMO Bull.*, **56**, 297–309.
- Chand, S. S., K. J. E. Walsh, and J. C. L. Chan, 2010: A Bayesian regression approach to seasonal prediction of tropical cyclones affecting the Fiji region. *J. Climate*, **23**, 3425–3445, <https://doi.org/10.1175/2010JCLI3521.1>.
- Chen, J.-H., and S.-J. Lin, 2013: Seasonal predictions of tropical cyclones using a 25-km-resolution general circulation model. *J. Climate*, **26**, 380–398, <https://doi.org/10.1175/JCLI-D-12-00061.1>.
- Diamond, H. J., A. M. Lorrey, and J. A. Renwick, 2013: A southwest Pacific tropical cyclone climatology and linkages to the El Niño–Southern Oscillation. *J. Climate*, **26**, 3–25, <https://doi.org/10.1175/JCLI-D-12-00077.1>.
- Efron, B., and R. J. Tibshirani, 1994: *An Introduction to the Bootstrap*. Chapman and Hall/CRC, 436 pp.
- Elsner, J. B., and C. P. Schmertmann, 1994: Assessing forecast skill through cross validation. *Wea. Forecasting*, **9**, 619–624, [https://doi.org/10.1175/1520-0434\(1994\)009<0619:AFSTCV>2.0.CO;2](https://doi.org/10.1175/1520-0434(1994)009<0619:AFSTCV>2.0.CO;2).
- Epstein, E. S., 1969: A scoring system for probability forecasts of ranked categories. *J. Appl. Meteor.*, **8**, 985–987, [https://doi.org/10.1175/1520-0450\(1969\)008<0985:ASSEPF>2.0.CO;2](https://doi.org/10.1175/1520-0450(1969)008<0985:ASSEPF>2.0.CO;2).
- Frank, E., and B. Pfahringer, 2006: Improving on bagging with input smearing. *Advances in Knowledge Discovery and Data Mining*, W.-K. Ng et al., Eds., Springer-Verlag, 97–106, https://doi.org/10.1007/11731139_14.
- Frank, N. L., 1977: Tropical systems—A ten year summary. Preprints, *11th Tech. Conf. on Hurricanes and Tropical Meteorology*, Miami, FL, Amer. Meteor. Soc., 455–458.
- Goebbert, K. H., and L. M. Leslie, 2010: Interannual variability of northwest Australian tropical cyclones. *J. Climate*, **23**, 4538–4555, <https://doi.org/10.1175/2010JCLI3362.1>.
- Gray, W. M., 1984a: Atlantic season hurricane frequency. Part I: El Niño and 30 mb quasi-biennial oscillation influences. *Mon. Wea. Rev.*, **112**, 1649–1668, [https://doi.org/10.1175/1520-0493\(1984\)112<1649:ASHFPI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1984)112<1649:ASHFPI>2.0.CO;2).
- , 1984b: Atlantic season hurricane frequency. Part II: Forecasting its variability. *Mon. Wea. Rev.*, **112**, 1669–1683, [https://doi.org/10.1175/1520-0493\(1984\)112<1669:ASHFPI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1984)112<1669:ASHFPI>2.0.CO;2).
- Hamill, T. M., 1997: Reliability diagrams for multicategory probabilistic forecasts. *Wea. Forecasting*, **12**, 736–741, [https://doi.org/10.1175/1520-0434\(1997\)012<0736:RDFMPF>2.0.CO;2](https://doi.org/10.1175/1520-0434(1997)012<0736:RDFMPF>2.0.CO;2).
- Harper, B. A., S. A. Stroud, M. McCormack, and S. West, 2008: A review of historical tropical cyclone intensity in North-Western Australia and implications for climate change trend analysis. *Aust. Meteor. Mag.*, **57**, 121–141.
- Holland, G. J., 1981: On the quality if the Australian tropical cyclone data base. *Aust. Meteor. Mag.*, **29**, 169–181.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2).
- Klotzbach, P. J., and W. M. Gray, 2003: Forecasting September Atlantic basin tropical cyclone activity. *Wea. Forecasting*, **18**, 1109–1128, [https://doi.org/10.1175/1520-0434\(2003\)018<1109:FSABTC>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<1109:FSABTC>2.0.CO;2).
- , and —, 2004: Updated 6–11-month prediction of Atlantic basin seasonal hurricane activity. *Wea. Forecasting*, **19**, 917–934, [https://doi.org/10.1175/1520-0434\(2004\)019<0917:UMPOAB>2.0.CO;2](https://doi.org/10.1175/1520-0434(2004)019<0917:UMPOAB>2.0.CO;2).
- , and —, 2009: Twenty-five years of Atlantic basin seasonal hurricane forecasts (1984–2008). *Geophys. Res. Lett.*, **36**, L09711, doi:10.1029/2009GL037580.

- Kwon, H. J., W.-J. Lee, S.-H. Won, and E.-J. Cha, 2007: Statistical ensemble prediction of the tropical cyclone activity over the western North Pacific. *Geophys. Res. Lett.*, **34**, L24805, doi:[10.1029/2007GL032308](https://doi.org/10.1029/2007GL032308).
- Lee, W.-J., J.-S. Park, and H. J. Kwon, 2007: A statistical model for prediction of the tropical cyclone activity over the western North Pacific. *J. Korean Meteor. Soc.*, **43**, 175–183.
- McDonnell, K. A., and N. J. Holbrook, 2004a: A Poisson regression model approach to predicting tropical cyclogenesis in the Australian/southwest Pacific Ocean region using the SOI and saturated equivalent potential temperature gradient as predictors. *Geophys. Res. Lett.*, **31**, L20110, doi:[10.1029/2004GL020843](https://doi.org/10.1029/2004GL020843).
- , and —, 2004b: A Poisson regression model of tropical cyclogenesis for the Australian–southwest Pacific Ocean region. *Wea. Forecasting*, **19**, 440–455, [https://doi.org/10.1175/1520-0434\(2004\)019<0440:APRMOT>2.0.CO;2](https://doi.org/10.1175/1520-0434(2004)019<0440:APRMOT>2.0.CO;2).
- Murphy, A. H., 1969: On the ranked probability skill score. *J. Appl. Meteor.*, **8**, 988–989, [https://doi.org/10.1175/1520-0450\(1969\)008<0988:OTPS>2.0.CO;2](https://doi.org/10.1175/1520-0450(1969)008<0988:OTPS>2.0.CO;2).
- , 1971: A note on the ranked probability skill score. *J. Appl. Meteor.*, **10**, 155–156, [https://doi.org/10.1175/1520-0450\(1971\)010<0155:ANOTRP>2.0.CO;2](https://doi.org/10.1175/1520-0450(1971)010<0155:ANOTRP>2.0.CO;2).
- Namias, J., 1969: On the causes of the small number of Atlantic hurricanes in 1968. *Mon. Wea. Rev.*, **97**, 346–348, [https://doi.org/10.1175/1520-0493\(1969\)097<0346:OTCOTS>2.3.CO;2](https://doi.org/10.1175/1520-0493(1969)097<0346:OTCOTS>2.3.CO;2).
- Nicholls, N., 1979: A possible method for predicting seasonal tropical cyclone activity in the Australian region. *Mon. Wea. Rev.*, **107**, 1221–1224, [https://doi.org/10.1175/1520-0493\(1979\)107<1221:APMFPS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1979)107<1221:APMFPS>2.0.CO;2).
- , 1984: The Southern Oscillation, sea-surface-temperature, and interannual fluctuations in Australian tropical cyclone activity. *J. Climatol.*, **4**, 661–670, <https://doi.org/10.1002/joc.3370040609>.
- , 1985: Predictability of interannual variations of Australian seasonal tropical cyclone activity. *Mon. Wea. Rev.*, **113**, 1144–1149, [https://doi.org/10.1175/1520-0493\(1985\)113<1144:POIVOA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1985)113<1144:POIVOA>2.0.CO;2).
- , 1992: Recent performance of a method for forecasting Australian season tropical cyclone activity. *Aust. Meteor. Mag.*, **40**, 105–110.
- , 1999: SOI-based forecast of Australian region tropical cyclone activity. *Experimental Long-Lead Forecast Bulletin*, Vol. 8, No. 4, Climate Prediction Center, Washington, DC, 71–72, <http://www.cpc.ncep.noaa.gov/products/predictions/experimental/bulletin/Dec96/art33.html>.
- Ramsay, H. A., L. M. Leslie, P. J. Lamb, M. B. Richman, and M. Leplastrier, 2008: Interannual variability of tropical cyclones in the Australian region: Role of large-scale environment. *J. Climate*, **21**, 1083–1103, <https://doi.org/10.1175/2007JCLI1970.1>.
- Thorncroft, C., and I. Pytharoulis, 2001: A dynamical approach to seasonal prediction of Atlantic tropical cyclone activity. *Wea. Forecasting*, **16**, 725–734, [https://doi.org/10.1175/1520-0434\(2002\)016<0725:ADATSP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2002)016<0725:ADATSP>2.0.CO;2).
- Villarini, G., and G. A. Vecchi, 2013: Multiseason lead forecast of the North Atlantic power dissipation index (PDI) and accumulated cyclone energy (ACE). *J. Climate*, **26**, 3631–3643, <https://doi.org/10.1175/JCLI-D-12-00448.1>.
- Weigel, A. P., M. A. Liniger, and C. Appenzeller, 2007: The discrete Brier and ranked probability skill scores. *Mon. Wea. Rev.*, **135**, 118–124, <https://doi.org/10.1175/MWR3280.1>.
- Wijnands, J. S., G. Qian, K. L. Shelton, R. J. B. Fawcett, J. C. L. Chan, and Y. Kuleshov, 2015: Seasonal forecasting of tropical cyclone activity in the Australian and the South Pacific Ocean regions. *Math. Climate Wea. Forecasting*, **1**, 21–42, <https://doi.org/10.1515/mcwf-2015-0002>.
- Wilks, D. S., 2005: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Academic Press, 648 pp.