

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

OPEN ACCESS

RECEIVED
20 August 2020REVISED
9 November 2020ACCEPTED FOR PUBLICATION
25 November 2020PUBLISHED
23 December 2020

Original Content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.

Interpretation of the statistical/dynamical prediction for seasonal
tropical storm frequency in the western North PacificNamyoung Kang¹ and James B Elsner²¹ Department of Geography, Kyungpook National University, Daegu 41566, Republic of Korea² Department of Geography, Florida State University, Tallahassee FL 32306, United States of AmericaE-mail: nkang.fsu@gmail.com**Keywords:** tropical storm, western North Pacific, seasonal prediction, hybrid model, predictive probability distribution

Abstract

Despite the improving techniques for seasonal prediction of tropical storm frequency, attention seems focused on accuracy rather than on forecast interpretation. This study aims to show how seasonal predictions from a hybrid model, i.e. statistical/dynamical model, can be interpreted with probability distributions. The tropical storm frequency in the western North Pacific is modeled with environmental predictors through multiple linear regression. For a demonstration of the probabilistic structure of the prediction result, the forty-two member ensemble predictions from the GloSea5 model for June–July–August in 2020 are used as the dynamical input. Rather than dealing with the expected frequency, this study introduces the predictive probability for a single value of the frequency. From as many probability distributions, a marginal probability distribution is obtained as the final predictive probability distribution. The probability distribution is then compared to the climatological reference by terciles. Additionally, predictive probability distributions made with the individual predictors provide helpful information on how each contributes to the final prediction. This probabilistic interpretation procedure is expected to be effectively used for improving any hybrid approach.

1. Introduction

As one of the most destructive natural disasters in the world, tropical storms have made people live in fear of their occurrences in the tropics (Mendelsohn *et al* 2012). Here the term tropical storm refers to a tropical cyclone whose maximum sustained wind speed exceeds 17 m s^{-1} (World Meteorological Organization 2017). Tropical storms are most frequent over the western North Pacific accounting for one-third of the global count and they are often the most intense (Supplementary information of Kang and Elsner (2015)). Intensity enhancement by global warming is also increasing public concerns over the potential damages (Kang and Elsner 2018). As residents potentially affected by these storms take an increasingly keen interest in next season's storm frequency, forecast agencies and research centers in the region have been developing various prediction models.

Decades ago, pioneering seasonal predictions of tropical storm frequency was done with statistical models in the Australian region (Nicholls 1979, Nicholls 1984), and the Atlantic basin (Gray 1984a,

Gray 1984b, Elsner and Schmertmann 1993). Owing to the high-power computer resources, dynamical models with a finer resolution are now available to detect tropical storms more realistically for seasonal predictions (Vitart *et al* 2007, Alessandri *et al* 2011, Vitart *et al* 2014). A hybrid method, combining the above two approaches, is a relatively new and increasingly popular approach to seasonal tropical cyclone prediction. A hybrid model is a so-called statistical/dynamical model, where the statistical model is used to relate the dynamically produced environmental factors to the tropical storm frequency. The technique has been introduced to the western North Pacific (Kim *et al* 2013, Li *et al* 2013), and most operational agencies make use of the hybrid modeling framework. The hybrid modeling technique has been further improved to utilize multi-model ensembles of the dynamical models (Kim *et al* 2017). Klotzbach *et al* (2019) describes up-to-date modeling approaches for the seasonal prediction of the storm frequency in the 12 agencies around the world. Four among the five forecast agencies in the western North Pacific produce seasonal predictions by running hybrid models.

Despite continuous improvement in prediction techniques, attention seems focused mostly on accuracy rather than on the interpretation of the predictions. This study aims to show how the prediction results of the hybrid model can be interpreted with probability distributions. Differing from the conventional analysis dealing with the expectation of the frequency, here we take special note of the predictive probability for a single value of the frequency. Then the quantitative interpretation of the future frequency is available by assigning the probability densities to 'below-normal', 'normal', and 'above-normal' ranges referenced by climatological terciles. For a demonstration of the probabilistic structure of the result, this study utilizes the environmental prediction of the GloSea5 for June–July–August (JJA) in 2020, operationally run by the Korea Meteorological Administration (KMA). GloSea5 is the Global Seasonal forecast system designed by the UK Met Office, where the ocean-atmosphere-land systems are fully coupled (MacLachlan *et al* 2015, Williams *et al* 2015). Here, the ensemble outputs from the GloSea5 model are used as the dynamical input to a statistical module. The statistical module employs the empirical relationship between the environment and the tropical storm frequency. Among the environmental factors, El Niño Southern Oscillation (ENSO) and the global ocean warmth are suggested to effectively explain the variability of tropical storm activity in the western North Pacific (Kang and Elsner 2015, Kang and Elsner 2016). Near orthogonality between the two large-scale variables over a long period provides confidence in quantitatively comparing the major contributions of the internal variability and external forcing to the seasonal predictions. Section 2 describes the data, and section 3 reviews the environmental connection to tropical storm frequency and demonstrates the predictive probability of the future frequency. Section 4 introduces the interpretation of the probability distributions, and section 5 provides a summary of the results and a discussion. Statistics and figures are made using the R programming language (www.r-project.org) and available online (<https://rpubs.com/Namyoung/P2020b>).

2. Data

The annual frequency of western North Pacific tropical storms comes from the best-track data provided by the U.S. Joint typhoon warning center. In the statistical module, the tropical storm frequency is explained by the two primary environmental factors, ENSO and global ocean warmth. ENSO is indicated by the southern oscillation index (SOI) from the National Oceanic and Atmospheric Administration (NOAA)/Climate Prediction Center, and the global ocean warmth is diagnosed by area-weighted global mean sea surface temperature (SST) of the NOAA/National Centers for Environmental Prediction

(NCEP) reanalysis (<https://psl.noaa.gov/cgi-bin/data/timeseries/timeseries1.pl>). El Niño is represented by the negative sign of SOI. Only for a naming purpose in this paper, the negative sign of SOI and global mean SST are denoted as NSOI and GMSST, respectively. The two variables are averaged over JJA for the 30-year period of 1986–2015. The year range is set consistent with supporting earlier work (Yang *et al* 2018, Kang *et al* 2019). For the predictors in the statistical module, the SOI and GMSST are obtained by the GloSea5 predictions on the environmental variables, run by KMA in operation. Predicted SOI is calculated by standardizing the pressure difference between spatially interpolated values at the two locations of Darwin (130.8°E, 12.5°S) and Tahiti (201.8°E, 17.6°S), and predicted GMSST is globally averaged with the weights on the latitude. For this, a trigonometric function, $\cos\phi$, is applied for weighting the area by its latitude (ϕ). Model predictions are scaled to have realistic magnitude by comparing the hindcast simulations of GloSea5 with the observed climatology over the 20-year period (1991–2010). For a demonstration of the probabilistic structure of the seasonal prediction, this study utilizes a case of the environmental prediction made by KMA in late May for JJA of 2020. Then, 42 ensemble pairs of NSOI and GMSST are available for the dynamical input to the statistical module as the predictors. Data can be available through the data management office at the agency.

The best-track data for the tropical storm frequency in the western North Pacific are available from the Regional Specialized Meteorological Center (RSMC) Tokyo (www.jma.go.jp/jma/jma-eng/jma-center/rsmc-hp-pub-eg/trackarchives.html). SOI from NOAA/CPC (<http://cpc.ncep.noaa.gov/data/indices/soi>), and SST from NOAA/NCEP (<https://psl.noaa.gov/data/gridded>).

3. Model structure

3.1. Predictors of the tropical storm frequency

For a statistical module in a hybrid model, this study utilizes two primary environmental predictors of ENSO and global ocean warmth based on the previous works on their connection to the inter-annual tropical storm activity. Here, El Niño status and global ocean warmth are indicated by NSOI and GMSST, respectively. Since SOI is less likely to reflect the localized global warming pattern like Niño indices do, it better represents an internal variability showing no forced trend (Kang *et al* 2019). GMSST implies more than a temperature value. It indicates the status of the forced environment made up of all collinear physical factors for global ocean warming. Here, 'collinear' means 'concurrent' from a linear perspective. The forced trend of GMSST becomes more apparent

on longer timescales, which makes the two predictors only weakly correlated (Yang *et al* 2018). NSOI and GMSST during the 30-year period (1986–2015) bears little relationship ($r = -0.19$; $[-0.51, 0.18]$, 95% confidence interval (CI)), which provides more stable statistical framework of the model.

A statistical relationship between the environment and tropical storm frequency is shown in figure 1 for the same 30-year period. PC1 and PC2 are the two principal components derived from a principal component analysis of NSOI and GMSST. PC1 (PC2) implies the in-phase (out-of-phase) mode between the variations of NSOI and GMSST. Colored inner ring shows the correlation of each directional variability with storm frequency. Only positive values are represented, and the significant correlation coefficients are colored in red. NSOI indicates a warm phase of ENSO, which is El Niño. A significant correlation ($r = +0.43$; $[0.08, 0.68]$, 95% CI) at NSOI confirms that El Niño (La Niña) increases (decreases) the storm frequency (Wang and Chan 2002, Chan 2005). On the other hand, the frequency is negatively correlated with GMSST ($r = -0.55$; $[-0.76, -0.24]$, 95% CI), meaning that frequency is significantly decreased by a globally warmer environment. Then, the highest correlation ($r = +0.64$; $[0.37, 0.82]$, 95% CI) appears along the PC2 direction. The PC2 represents the out-of-phase relationship between NSOI and GMSST. This means that PC2 represents variations where a higher tropical cyclone frequency is likely to occur in a relatively colder El Niño year, while a lower frequency is likely to occur in a relatively warmer La Niña year (Kang and Elsner 2016). It is also noted that PC1 is only weakly correlated with the frequency, implying that the contribution of NSOI is offset by that of GMSST, and vice versa. Though the essential characteristics of the variables may not be continuous normal, this study assumes the normality of the variables for linear interpretation of the results. In this paper, the frequency of tropical storms is denoted as FRQ. Values of FRQ, NSOI, and GMSST are adequately described by normal distributions as verified by a Shapiro–Wilk test (Royston 1982). A multiple regression is used to quantify the linear perspective as follows:

$$\hat{\text{FRQ}} = \beta_0 + \beta_1 \cdot \text{NSOI} + \beta_2 \cdot \text{GMSST}, \quad (1)$$

where β_0 , β_1 and β_2 are the intercept and the two regression coefficients, respectively. $\hat{\text{FRQ}}$ is the mean response corresponding to a particular pair of values of the predictor variables, i.e. NSOI and GMSST.

A generalized linear model of FRQ by assuming another probability density form such as a Poisson distribution could be an alternative for modeling the non-negative discrete frequency values. But a normal distribution is an adequate approximation to the Poisson distribution when the annual rate exceeds 20 even without a continuity correction. Moreover, this

paper places emphasis on describing the probabilistic structure represented in figure 1. Between any two standardized variables, a correlation coefficient is the regression coefficient of the OLS model, and at the same time the projection length in a geometric variability space (Supplementary information of Kang and Elsner (2015)). The correlation ring in figure 1 is again understood as the projection length of each directional variability onto FRQ. Then, equation (1) is the mathematical expression of the biggest projection length, 0.64, which implies the correlation coefficient of FRQ with $\hat{\text{FRQ}}$ made by OLS regression.

3.2. Probability distribution of the prediction on the climatological predictors

Figure 2(a) shows a bivariate normal distribution of NSOI and GMSST over the 30-year period (1986–2015). The values are plotted in standardized forms. Thirty pairs of NSOI and GMSST are dotted in gray, and the theoretical normal probability densities are contoured circles. Having no information on the future environment, the prediction would be the climatological mean (green dot). By construction, the predicted value, $\hat{\text{FRQ}}$ is an expectation over possible FRQs. The probability density of $\hat{\text{FRQ}}$ can be obtained from a t-distribution. The t-statistic can be expressed as

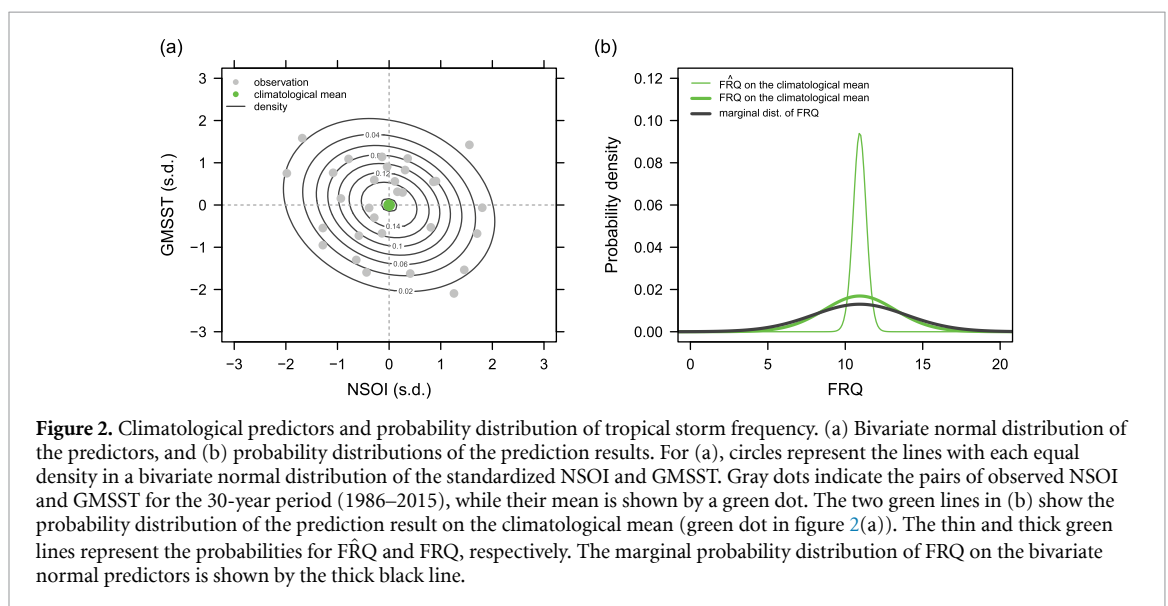
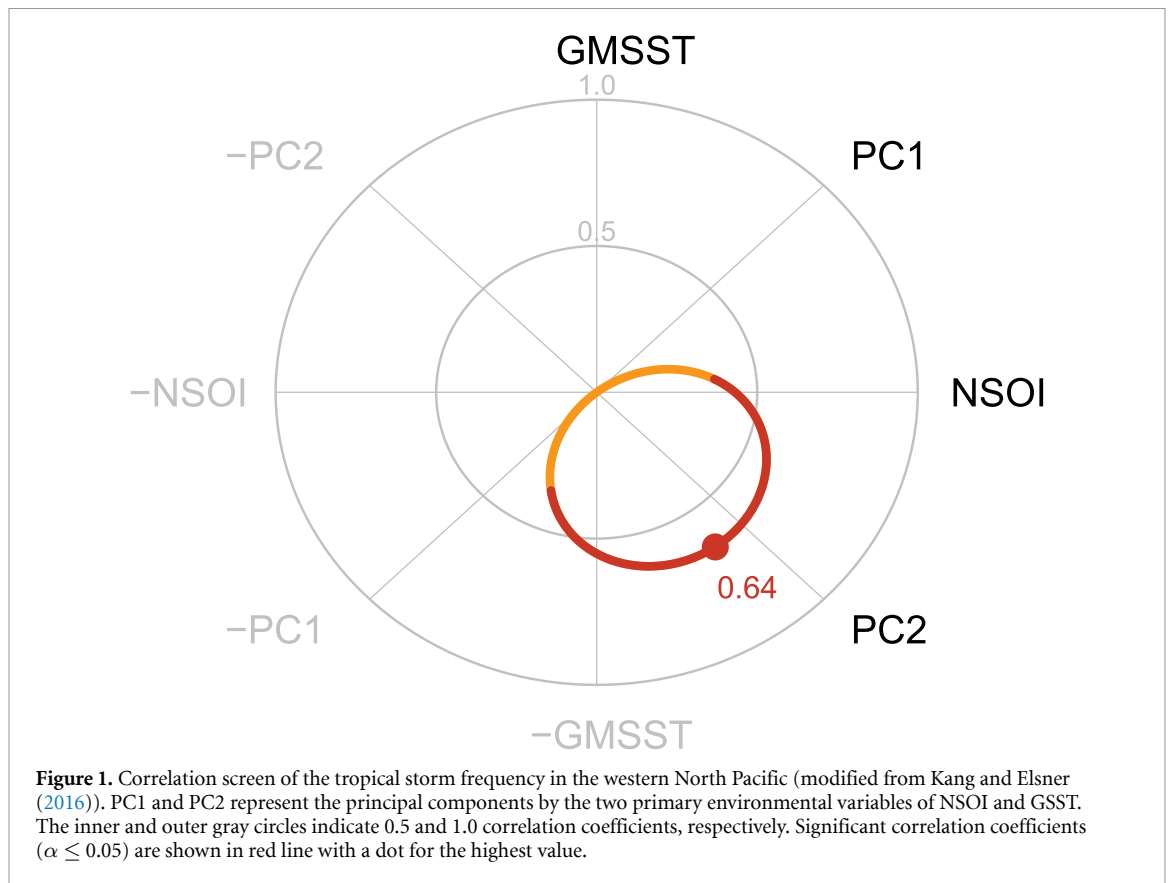
$$t_{\alpha/2} = \frac{\hat{\text{FRQ}} - \mu_{\text{FRQ}}}{\text{s.e.}(\hat{\text{FRQ}})}, \quad (2)$$

where α is the significance level, and μ_{FRQ} is the mean of $\hat{\text{FRQ}}$. The spread of the distribution depends on the standard error (s.e.) of $\hat{\text{FRQ}}$. ‘Confidence interval’ refers to a certain probability density range, i.e. 95%, of the expectation. The degrees of freedom in a multiple regression equals $N-k-1$, where N and k are the numbers of observations and variables, respectively. Thus, the t-distribution is determined by 27 degrees of freedom where $N = 30$ and $k = 2$. The probability density distribution is shown as a thin green line in figure 2(b). It is found that $\hat{\text{FRQ}}$ presents a sharper peak than FRQ. This implies that there should be a significant loss of uncertainty information, which might have made a CI-based probabilistic approach to seasonal prediction less successful so far. Since the probability is about the expectation ($\hat{\text{FRQ}}$), and not about the individual FRQ, its spread is narrower than what the probability of FRQ might show. The probability of FRQ rather than $\hat{\text{FRQ}}$ would provide more practical information on the future FRQ. Now, we want to find the probability of the t-statistic for FRQ. The t-statistic for a single value of FRQ can be expressed as

$$t_{\alpha/2} = \frac{D - \mu_D}{\text{s.e.}(D)}, \quad (3)$$

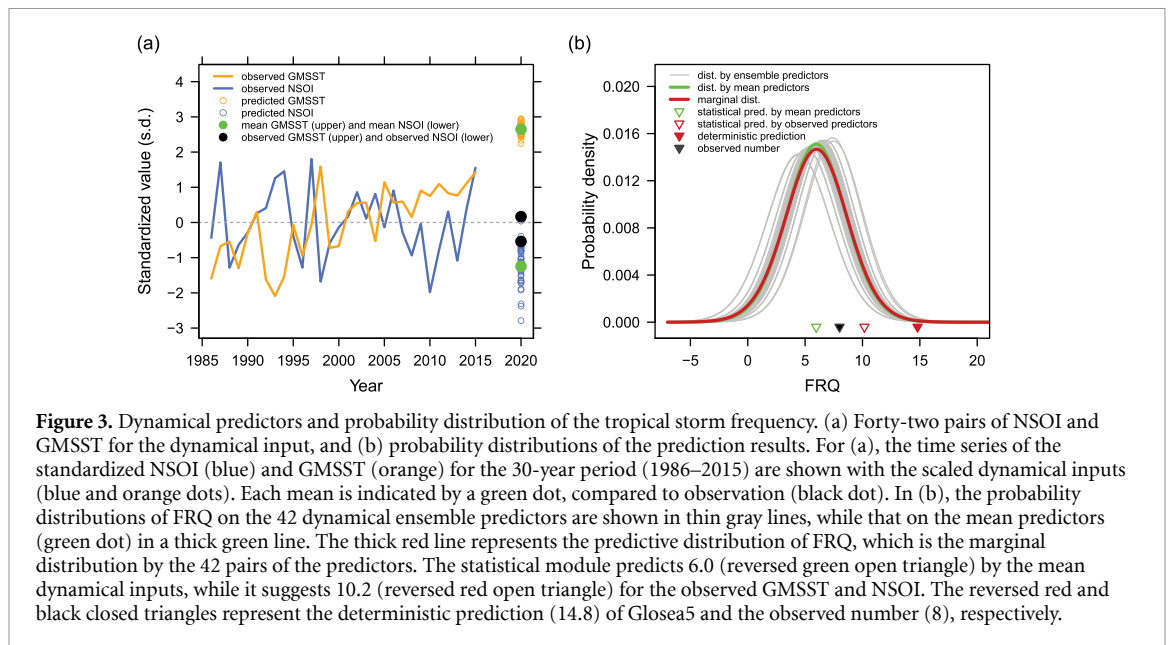
where

$$D = \text{FRQ} - \hat{\text{FRQ}}. \quad (4)$$



D denotes the difference of FRQ from its predicted value. Since in the mean FRQ equals $\hat{FRQ}$, μ_D is by definition zero. The probability distribution of FRQ is shown in a thick green line, whose spread is confirmed wider than that of $\hat{FRQ}$. ‘Prediction interval’ for a certain probability range refers to this probability distribution. Then, the probability distribution is summarized as the predictive probability of FRQ on the two climatological mean values of NSOI and GMSST.

However, the pair of climatological means is only the case with the highest density among all possible future environments. Every grid point in the density plane of the bivariate normal distribution (see figure 2(a)) is a candidate pair of the predictor values. The grid points are at 0.1 standard deviation (s.d.) intervals from -3.0 to 3.0 for each predictor, and then 61×61 pairs of the predictor values are applied to the model. The predictive probability, shown as a thick black line in figure 2(b), is the marginal distribution



of the tropical storm frequency, where each distribution on a pair of the predictor values is weighted by its density value. Since the uncertainties of the predictors are considered, the density distribution is found to have a bit larger spread than that of the prediction on the climatological means.

3.3. Probability distribution of the prediction on the dynamical predictors

Finally we have the information on the future environment from a dynamical model, GloSea5. Figure 3(a) shows the time series of NSOI (blue) and GMSST (orange) over JJA during the 30-year period (1986–2015). NSOI shows little correlation with time as expected from an internal variability ($r = -0.07$; $[-0.42, 0.29]$, 95% CI). On the other hand, GMSST is seen to be significantly increasing over time with some fluctuations ($r = +0.75$; $[0.53, 0.87]$, 95% CI). Forty-two pairs of the predictors from GloSea5 are plotted in circles for each variable. They are the values scaled to observation using the GloSea5 hindcast simulations for the past 20 years (1991–2010). The means of the predictor values are shown in green dots. The dynamical model predicts the highest level of global ocean warmth, but near-neutral ENSO status. A qualitative interpretation might be that the future frequency will be smaller than average considering the influence of the warmest environment. This interpretation can be quantitatively examined by the probability distributions. The green curve in figure 3(b) represents the probability distribution for only the pair of the means (green dots in figure 3(a)). However, this does not fully involve the spread of the ensemble members. The probability distributions for ensemble pairs of the predictors are shown in thin gray lines. The predictive probability distribution can

be obtained by the marginal distribution of the 42 ensemble members (red). This predictive probability quantifies how the dynamical model foresees the tropical storm frequency in JJA of 2020, which can be compared to the above theoretical reference (black line in figure 2(b)).

Here, the dynamical error structure can be quantitatively estimated by observed environmental predictors. As long as the environmental connection to the tropical storm frequency remains the same for the season, the prediction error can be understood by examining the difference between the predictors and the observed values. Each black dot in figure 3(a) represents observed GMSST (upper) and NSOI (lower), showing each large difference from the GloSea5 prediction. In the statistical module, the positive GMSST and the negative NSOI lead to the smaller FRQ at the same time. It is concluded that both the warm-biased GMSST and La Niña-biased ENSO status in this case excessively shifted the probability distribution to the smaller numbers. The combined contribution of the dynamical errors is captured as the gap between the reversed green open triangle (6.0) and the reversed open red triangle (10.2) in figure 3(b). The two triangles indicate the predicted values for the mean dynamical inputs and the observed predictors, respectively. In the same way, this interpretative framework can be used for understanding the potential prediction error and its origin before the observation is made. This interpretative framework also helps people infer a modification of the probability distribution. If someone believes there would be no warm bias of GMSST, the person may consider more frequency than presented by the model. This functional analysis could also apply to the validation of hybrid predictions after the season. Observation reveals that the eventual reduction of the

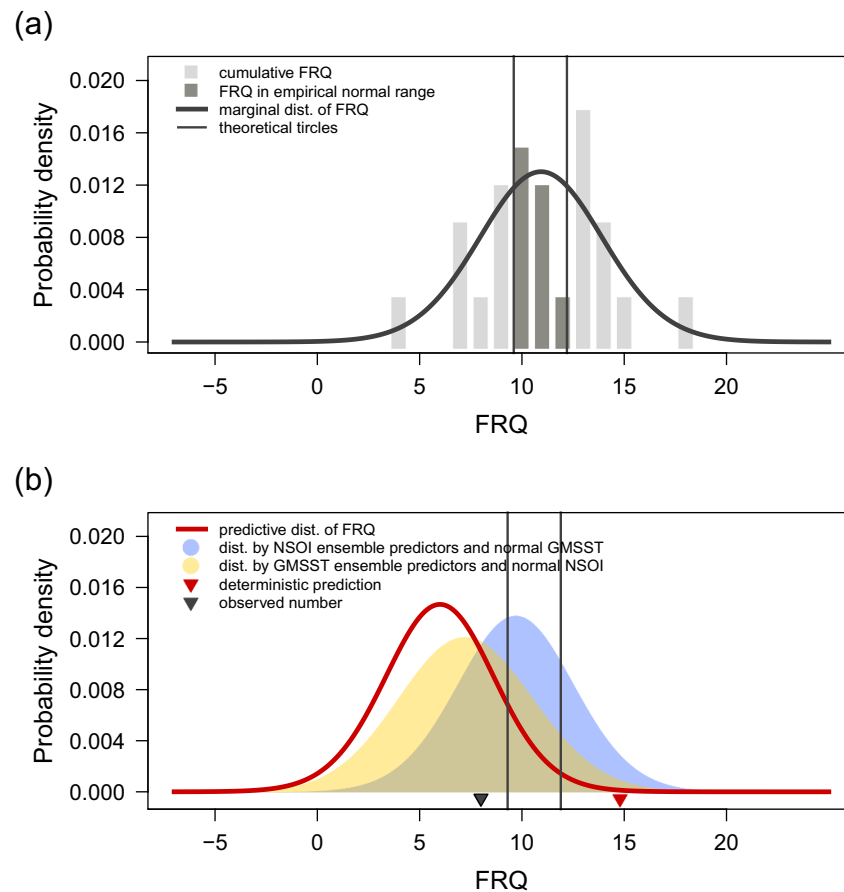


Figure 4. Portioning the predictive densities of the tropical storm frequency by climatological terciles. (a) Comparison of the terciles in the theoretical climatology and the histogram, and (b) division of the predictive densities divided by the theoretical terciles. Terciles are 0.33 and 0.66 quantiles. For (a), histogram bars in dark gray represents the empirical 'normal'. In (b), the predictive densities of FRQ are compared with the ranges defined by the theoretical terciles. In this example case, 90.3%, 8.4%, and 1.3% are assigned to 'below-normal', 'normal', and 'above-normal', respectively. Areas shaded in blue and yellow shows the predictive probability distributions where respectively NSOI and GMSST ensemble predictors are exclusively applied, providing helpful information on their individual role in the prediction. The reversed red and black closed triangles represent the deterministic prediction (14.8) of GloSea5 and the observed number (8), respectively.

tropical storm number might have been influenced by the NSOI anomaly rather than the normal GMSST. The statistical module itself is also a source of prediction errors. The best inference, however, relies on the most confident statistical module at the time the prediction is made.

4. Interpretation

Quantitative interpretation of the prediction can be made by climatological terciles, classifying the frequency portions into 'below-normal', 'normal' and 'above-normal' categories. Figure 4(a) shows the histogram of the empirical distribution of the tropical storm frequency. A conventional way of defining the terciles is to find quantiles at .33 and .66 probability levels in the observed 30 frequency samples. Since the frequency is a non-negative discrete number in nature, someone may roughly regard the 'normal' range as 10 to 12 (dark gray bars). This can be compared with the terciles (vertical lines) induced by the theoretical climatology (black line, the same as in figure 2(b)). Since all possible cases of the predictors

are theoretically taken into account, the marginal probability is considered as a theoretical climatology of the tropical storm frequency. Here, 'theoretical' means the normal form of the bivariate distribution of the predictors. The theoretical approach gives a more complete and realistic form of the uncertainty distribution than the conventional approach of simply accumulating the observations, i.e. histogram. Having the same form of the predictive probability (see figure 2(b)), this probability distribution can be used as a practical climatological reference to any predictive probability that a model produces. With a simple but less rigorous form of the distribution, the histogram-based empirical terciles are considered less effective to be used as a climatological reference for interpreting the model results. As the theoretical distribution depicts a predictive density distribution in a normal form of continuous values, its terciles are symmetrically distributed around the mean (10.9), while the empirical terciles are not. The predictive probability of the hybrid model can be divided by the terciles (figure 4(b)). The 'normal' range is from 9.6 to 12.2, inclusive.

In comparison, model prediction in figure 4(b) (red line) is seen more distributed on 'below-normal' of the frequency values, while less on 'above-normal'. This can be quantitatively described as 90.3%, 8.4%, and 1.3%, respectively for 'below-normal', 'normal', and 'above-normal' ranges. Below-normal number of tropical storms (8) were eventually observed in JJA 2020. The motivation of the current research is amplified by the unrealistic value of the deterministic prediction (14.8), which is the average of the last five ensemble simulations among the 42 members. The number for each member is determined by a detection algorithm considering the wind speed and the warm core temperature anomaly relative to the surrounding environment. It is also noted that only the detected number of tropical storms simulated by GloSea5 does not provide an explicit rationale behind the predicted values.

The two shaded distributions show the different influences NSOI and GMSST have on the prediction result. An area shaded in blue shows the predictive probability where only the NSOI ensemble predictors are applied. For each ensemble member of NSOI, the normal GMSSTs at 0.1 s.d. from -3.0 to 3.0 , i.e. 61 samples are used for the pairs of predictor values. Then, the marginal distribution comes from the 42×61 predictive probability distributions. The marginal predictive probability for ENSO status of the 42 NSOI ensemble members, describes 47.8%, 33.1%, and 19.1% probability densities for 'below-normal', 'normal', and 'above-normal', respectively. The same procedure is applied to the ensemble members of GMSST (area shaded in yellow). Here, it is assumed that only GMSST is informative but not NSOI. For each ensemble member of GMSST, the normal NSOIs at 0.1 s.d. intervals from -3.0 to 3.0 , are used for the pairs of predictor values. The marginal predictive probability by the warmest level of GMSST ensembles shows 75.2%, 17.9%, and 7.0% probability densities for 'below-normal', 'normal', and 'above-normal', respectively. Though the final predictive probability (red line) cannot be divided into the two additive distributions for NSOI and GMSST, each shaded distribution provides helpful information on how the final is brought about. In this case, the prediction results are interpreted as both the environmental conditions accompanying the ENSO and the warmest global ocean are likely to allow tropical cyclone frequency to occur in the 'below-normal' range overall.

5. Summary and discussion

Forecast agencies and research centers have been developing various models for seasonal prediction of tropical cyclones. A hybrid model has been available by combining statistical and dynamical approaches to the seasonal prediction of the western North

Pacific tropical storm frequency. However, attention seems focused mostly on accuracy rather than on the interpretation of the prediction results. This study shows how the prediction results of the hybrid model can be interpreted by probability distributions. For a demonstration of the probabilistic structure of the result, the environmental prediction of the GloSea5 for JJA in 2020 is utilized. This interpretative approach may apply to any hybrid prediction. It needs to be noted that the prediction skill of a specific hybrid model is not within the scope of the current study.

The modeling procedure and the relevant interpretations are as follows. First, the statistical module in the hybrid model utilizes two primary environmental predictors which are ENSO and global ocean warmth based on the previous works on their connection to the interannual tropical storm activity (Kang and Elsner 2016, Yang *et al* 2018, Kang *et al* 2019). Near-zero correlation between NSOI and GMSST during the 30-year period (1986–2015) provides a robust statistical framework for the model. The tropical storm frequency is modeled by the predictors through multiple linear regression. Second, the distribution of the climatological frequency is produced on a theoretical basis. Rather than dealing with the expectation of the frequency, this study introduces the predictive probability for a single value of the frequency. By applying pairs of environmental predictors in a bivariate normal distribution, a better climatological reference than a conventional histogram is obtained. Third, the predictors from a dynamical module, GloSea5 are input to the statistical module. The dynamical input consists of 42 ensemble pairs of NSOI and GMSST, and then a marginal probability distribution is obtained by as many probability distributions. Lastly, the final predictive probability is compared to the prepared climatological reference. The quantitative interpretation of the model prediction is available by assigning the probability densities to 'below-normal', 'normal', and 'above-normal' ranges referenced by climatological terciles. Auxiliary two probability distributions by separate NSOI and GMSST, provide helpful information on how the final prediction is affected by the predictors.

The merit of this interpretative seasonal prediction lies in the fact that the uncertainties in the statistical module as well as the dynamical module are explicitly expressed. Here, the predictive probability distribution itself comes from the statistical uncertainty, and the dynamical ensemble predictors give rise to the spread of the predictive probability distributions. Then, the final predictive probability distribution of the marginal densities includes all available uncertainties as it should be. Additional strength is that the cause and effect of the model prediction can be reviewed after the season. As long as the environmental connection to the tropical storm

frequency remains the same for the season, the prediction error can be understood by examining the difference of the predictors from observations.

A caveat of this model is that some negative frequency range could be included in the probability distribution. Despite the usefulness of the linear assumption, the deficiency comes from the normality of the frequency. Its influence is not considered significant for the three-month prediction (i.e. JJA), but the caveat implies that the prediction for a longer period such as six months (i.e. June to November) as having a larger frequency value might be better for the application of the current interpretation in operation. Another thing to consider is that the regression in the statistical module relies on a linear relationship between the tropical storm frequency and the environment predictors. The western North Pacific is a region with the strongest observed trends in various metrics (Knutson *et al* 2019). The interpretation in the current research should be effective as much as the linearity captures the changing values of the metrics.

The interpretative framework in this study is designed for the interannual variation of the tropical storm frequency and deals with the overall frequency in the western North Pacific. The influence of ENSO on TC frequency varies in magnitude and sign across the western North Pacific (Wang and Chan 2002, Kang *et al* 2019). The sub-regional interpretation, including the number of landfalling storms (e.g. around the South China Sea or Japan) is beyond the scope of the current study but will be a future challenge when observations are sufficient to find some reliable statistical relationship.

The validation of the hybrid model could be available by a sufficient number of forecast events. This study focuses on the interpretation of the results from a hybrid model, rather than the techniques for its accuracy. The accuracy of the model could be further improved in various ways. An increasing number of qualified observations may make a more reliable statistical module. Better and more explanatory environmental predictors can be found as well. The dynamical ensemble model can even be replaced by a new one. Each of the above options would be a research topic for the more accurate seasonal prediction, while the interpretative framework shown in this study will be valid for any result.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgment

Partial support for this work came from the Shaw Foundation in support of graduate education in geography at Florida State University, USA.

ORCID iD

Namyoung Kang  <https://orcid.org/0000-0003-3018-9038>

References

- Alessandri A, Borrelli A, Gualdi S, Scoccimarro E and Masina S 2011 Tropical cyclone count forecasting using a dynamical seasonal prediction system: sensitivity to improved ocean initialization *J. Clim.* **24** 2963–82
- Chan J C L 2005 Interannual and interdecadal variations of tropical cyclone activity over the western north pacific *Meteorol. Atmos. Phys.* **89** 143–52
- Elsner J B and Schmertmann C P 1993 Improving extended-range seasonal predictions of intense Atlantic hurricane activity *Weather and Forecasting* **8** 345–51
- Gray W M 1984a Atlantic seasonal hurricane frequency. Part I: El Niño and 30 mb quasi-biennial oscillation influences *Mon. Weather Rev.* **112** 1649–68
- Gray W M 1984b Atlantic seasonal hurricane frequency. Part II: forecasting its variability *Mon. Weather Rev.* **112** 1669–83
- Kang N and Elsner J B 2015 Trade-off between intensity and frequency of global tropical cyclones *Nat. Clim. Change* **5** 661–4
- Kang N and Elsner J B 2016 Climate mechanism for stronger typhoons in a warmer world *J. Clim.* **29** 1051–7
- Kang N and Elsner J B 2018 The changing validity of tropical cyclone warnings under global warming *NPJ Clim. Atmos. Sci.* **1** 36
- Kang N, Kim D and Elsner J 2019 The contribution of super typhoons to tropical cyclone activity in response to ENSO *Sci. Rep.* **9** 5046
- Kim H-M, Lee M-I, Webster P J, Kim D and Yoo J H 2013 A physical basis for the probabilistic prediction of the accumulated tropical cyclone kinetic energy in the western north pacific *J. Clim.* **26** 7981–91
- Kim O-Y, Kim H-M, Lee M-I and Min Y-M 2017 Dynamical–statistical seasonal prediction for western North Pacific typhoons based on APCC multi-models *Clim. Dyn.* **48** 71–88
- Klotzbach P *et al* 2019 Seasonal tropical cyclone forecasting *Tropical Cyclone Res. Rev.* **8** 134–49
- Knutson T *et al* 2019 Tropical cyclones and climate change assessment: Part I: detection and attribution *Bull. Am. Meteor. Soc.* **100** 1987–2007
- Li X, Yang S, Wang H, Jia X and Kumar A 2013 A dynamical-statistical forecast model for the annual frequency of western pacific tropical cyclones based on the NCEP climate forecast system version 2 *J. Geophys. Res. Atmos.* **118** 12061–74
- MacLachlan C *et al* 2015 Global seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system *Q. J. R. Meteor. Soc.* **141** 1072–84
- Mendelsohn R, Emanuel K, Chonabayashi S and Bakkensen L 2012 The impact of climate change on global tropical cyclone damage *Nat. Clim. Change* **2** 205–9
- Nicholls N 1979 A possible method for predicting seasonal tropical cyclone activity in the Australian region *Mon. Weather Rev.* **107** 1221–4
- Nicholls N 1984 The southern oscillation, sea surface temperature and interannual fluctuations in Australian tropical cyclone activity *J. Clim.* **4** 661–70
- Royston J P 1982 An extension of Shapiro and Wilk's w test for normality to large samples *Appl. Stat.* **31** 115–24
- Vitart F *et al* 2014 Seasonal forecasts: topic 5.2 *Int. Workshop on Tropical Cyclones VIII report*
- Vitart F, Huddleston M R, Déqué M, Peake D, Palmer T N, Stockdale T N, Davey M K, Ineson S and Weisheimer A 2007 Dynamically-based seasonal forecasts of atlantic tropical storm activity issued in june by eurosip *Geophys. Res. Lett.* **34** L16815

- Wang B and Chan J C L 2002 How strong ENSO events affect tropical storm activity over the western North Pacific *J. Clim.* **13** 1517–36
- Williams K *et al* 2015 The met office global coupled model 2.0 (gc2) configuration *Geosci. Model Dev.* **8** 1509–24
- World Meteorological Organization 2017 Typhoon Committee Operational Manual: Meteorological Component 2017th edn *Technical Report WMO TD-No. 196*
- Yang S-H, Kang N, Elsner J B and Chun Y 2018 Influence of global warming on western North Pacific tropical cyclone intensities during 2015 *J. Clim.* **31** 919–25