

Article

An Alternative Multi-Model Ensemble Forecast for Tropical Cyclone Tracks in the Western North Pacific

Sanghee Jun ¹ , Nam-Young Kang ^{2,*}, Woojeong Lee ² and Youngsin Chun ²

¹ Numerical Modeling Center, Korea Meteorological Administration, Seoul 07062, Korea; appjsh@gmail.com

² National Typhoon Center, Korea Meteorological Administration, Jeju 63614, Korea; lwj@korea.kr (W.L.); hwangsa@korea.kr (Y.C.)

* Correspondence: nkang.fsu@gmail.com; Tel.: +82-70-7850-635

Received: 7 June 2017; Accepted: 12 September 2017; Published: 15 September 2017

Abstract: This study introduces an unequally weighted technique for Multi-model Ensemble (MME) forecasting for western North Pacific Tropical Cyclone (TC) tracks. Weights are calculated by partial least square regression, and members are selected by paired *t*-test. The performances for shorter forecast time ranges, such as 24, 48 and 72 h, are examined in order to improve the MME model, in which the weights for members are equally assigned. For longer forecast time ranges, such as 96 and 120 h, weights for MME members are thought to be less reliable, since the modeling is more likely to be influenced by the climate variability in the data period. A combination of both techniques for the shorter and the longer forecast time ranges is suggested as an alternative MME forecast procedure in operational meteorological agencies.

Keywords: multi-model ensemble; tropical cyclone; track forecast; partial least squares regression

1. Introduction

Climatology records the strongest and the most active Tropical Cyclones (TCs) in the western North Pacific (WNP), among other global ocean basins. TCs in this ocean basin are becoming even stronger in response to global warming [1]. A tropical cyclone is a destructive weather system that causes huge losses of both life and property. Nevertheless, the damage could be reduced by effective preparedness, as long as TCs are predictable [2]. Timely and accurate forecast for TC tracking is most necessary for the successful prevention of potential threats [3].

Multi-Model Ensemble (MME) of the forecast tracks from various prediction models have been exploited as a useful forecast technique in many meteorological centers [4]. An MME process requires a pool of participating members as prediction sources. One of the merits of the MME technique is that more reliable track forecasts are available, based on the forecast spread of the ensemble members. Previous studies have revealed that the prediction performance of the MME technique can be largely influenced by various factors in consideration. Firstly, a Selective MME SMME, from a subset of the best members outperforms the simple consensus of all possible members in the pool [5]. SMME, in consideration of predefined error mechanisms, was utilized by the U.S. Joint Typhoon Warning Center (JTWC) [6–8]. JTWC's approach, however, was not very successful, since the selection of the members could be easily influenced by forecaster's subjectivity [9]. Subsequently, Qi et al. [10] and Jun et al. [11] each independently developed and automated an objective approach to SMME, which directly makes use of model errors when selecting members. Relative independence among the members is another factor to consider [12], as is employed for the operational TC track forecast by the U.S. National Hurricane Center and JTWC [13,14]. Assigning weight to each member forecast can also be beneficial for MME performance. Unequal weights could be estimated by multiple linear regression method, whose regression coefficients represent individual contributions to the MME performance.

This attempt was made by Krishnamurti et al. [15] under the name of the ‘Superensemble’ technique. Verification of MME forecasts in operation have mostly suggested smaller errors when unequal weights are applied [16,17].

In the previous studies, SMME and weighted methods have been separately advanced. This study takes both into account for more accurate forecasting. Thus, this study tries to improve both SMME and weighted methods, and merges them into an alternative forecasting technique, Weighted SMME (WSMME). Data and methodologies in association with WSMME are addressed in Section 2. Here, statistical interval estimation using paired *t*-test is employed for selecting participating members. Then, Partial Least Squares Regression (PLSR) is used to assign weight to each selected member, avoiding the multicollinearity problem among members. Also, ‘noisy’ parts of predictors could be eliminated by PLSR when unequal weights are produced [18]. In Section 3, the forecast skills of WSMME for WNP TCs over 4 years (2012–2015) are compared with those of numerical models and other MME approaches. In addition, Section 3 provides more detailed information of WSMME based on a TC named ‘Tembin’ in 2012. Results are summarized and discussed in Section 4.

2. Data and Methods

2.1. Data

This study experiments with statistical modeling for WNP TCs whose lifetime-maximum intensity exceeds 17 ms^{-1} . Four different numerical models (Table 1) are used for developing an MME model based on 127 TC cases during 2011–2015. Results are evaluated by best-track data from Japan Meteorological Agency (JMA). When any forecast value of the numerical model is not available, an ensemble version of the model is used to provide an averaged value of the member forecasts as a substitute. Global Spectrum Model (GSM)’s forecast is not available after 84 h, so the ensemble version is used after that time. In other models, there are few irregular missing values. When a deterministic value is missing for a forecast, then its ensemble outputs are exploited. For Global Data Assimilation Prediction System (GDAPS), the 2015 version of Met Office Global and Regional Ensemble Prediction System (MOGREPS) is exploited, since GDAPS has the same origin as the Unified Model, whose ensemble version is MOGREPS.

Table 1. The list of numerical models used in this study.

Name	Produce Center	Horizontal Resolution (km)	Output Time Step	Model Cycle (Per Day)	Version
Global Data Assimilation Prediction System (GDAPS)	KMA	25	3/6 hourly (0–87/288 h)	4	Unified model vn7.8–7.9
Global Forecast System (GFS)	NCEP	28/70	3/12 hourly (0–384 h)	4	GFS 9.0.1
Integrated Forecast System (IFS)	ECMWF	16	6 hourly (0–240 h)	2	Cycle 36r4–41r1
Global Spectrum Model (GSM)	JMA	20	6 hourly (0–84 h)	4	Changed in 2014

Korea Meteorological Administration (KMA), National Centers for Environmental Prediction (NCEP), European Centre for Medium–Range Forecasts (ECMWF), Japan Meteorological Agency (JMA).

2.2. Methods

The experiment is designed for operational forecasts under real-time conditions. The most recently produced model forecasts from the forecast initial time are used [19,20]. It takes 6 to 12 h or more to arrange model track forecasts for an MME forecast. The position of minimum mean sea level pressure in the forecast field is defined as the TC center. In the tracking algorithm, first guess position is given by the previous forecasts with 6 or 12 h (when 6 h forecast is not available) lead time. Then, the TC position is extracted from a quadratic surface around the center of the first guess.

2.2.1. Selecting Stage

$$H_0 : e_r = e_a \tag{1}$$

$$H_a : e_r < e_a \tag{2}$$

Statistical interval estimation can be applied to exclude members that are very likely to degrade MME forecast skill. Here, paired *t*-test is used to provide a reliable estimation of the true performance of models than point estimation [21]. This test investigates the statistically significant difference between two models, usually correlated [22,23]. A null hypothesis such as (1) is rejected, and an alternative hypothesis such as (2) is chosen at 95% confidence level (α (significance level) ≤ 0.05). Then, a member is removed from the participant members when it is likely to increase WSMME error at 95% probability level or above. MME members are selected through the process separately at each forecast time range, i.e., 24, 48, 72, 96 and 120 h. The MME performance is attributed to the errors and the relative independence among members [12]. This implies that a selected factor would not be the error of members but the error of MME for better performances. e_a denotes the error of MME forecasts containing all members (*a*) such as GDAPS, GFS, GSM, and IFS. e_r represents the error of MME forecast made by remaining members (*r*) after removing a subject members [13]. Input values are updated continuously for evaluation. This selecting stage is practically similar to the cross-validation process in regression analysis. The error of the track forecast indicates the distance between forecast and observed positions. The position errors produced for each time step (00, 06, 12, 18 UTC) are averaged for each forecast lead time (24, 48, 96, 120 h), and *t* statistic is tested. The position error distance is calculated on a great circle. *T*-statistic provides statistical significance of the difference between MME members.

2.2.2. Weighting Stage

Unequal weights for latitude and longitude per forecast time range are calculated through the PLSR technique. For instance, if 120 h track forecasts of unequal weighting MME are produced at 6 h intervals, then the number of total unequal weights is 40. The PLSR regresses the response variable (observed TC position; *o*) on the latent variables (*T*) containing variations of both *o* and predictors (member track forecasts; *F*), not directly using *F* for the explanatory variable [24]. *F*, referred to as forecast data matrix, consists of model member and each time step for column and row, respectively. *T* is useful for preventing multicollinearity problems assigning unequal weights based on the orthogonal relationship among the latent variables. The PLSR produces the latent variable matrix (*T*), where the information on variation in both *o* and *F* is maintained (Figure 1). The number of new variables is less than or equal to that of members, since ‘noisy’ variables might be removed through the cross-validation procedure [18]. The number of variables is determined to have minimum forecast error through a ‘leave-one-out’ cross-validation method. This uses all possible datasets of training data.

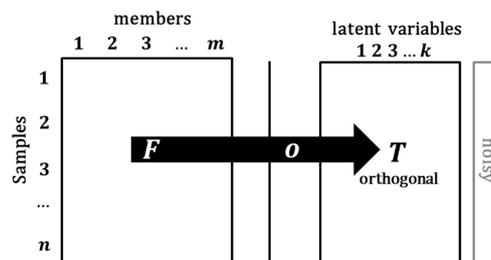


Figure 1. The schematic plot of Partial Least Squares Regression (PLSR). *F* is the matrix for forecast data of selected members (predictors) and *o* is the vector of tropical cyclone position (a response). PLSR deprives *T*, referred to as latent variables, of *F* maximizing correlation with *o*. The noisy matrix is not used for calculating unequal weights.

F and o are the training datasets consisting of increments of respective forecast longitudes and latitudes. For example, 12° N at forecast initial time and 27° N at 120 h return 15° N as its latitudinal increment [17]. The training dataset may be less effective for assigning skillful weights on WMME members, if any participant model undergoes changes. By this reason, the data range for the training datasets is limited to the last year, in order not to involve long-past characteristics of models [16]. The data set for evaluation and training may contain the forecasts made in current and the previous years. This happens to TC cases over two calendar years. This implies that the model members and weights could differ by forecast times. MME forecasts are produced for the years 2012–2015. Thus, the forecasts from 2011 are used for evaluation and training, while those from 2015 are not.

3. Results

3.1. Forecast Skill of WSMME

3.1.1. Comparison with MME

Errors are calculated for the selecting stage and weighting stage. Comparisons are made among the results from the different three approaches named as Equally Weighting Overall MME (EWOM), Equally Weighting Selective MME (EWSM) and Unequally Weighting Selective MME (UWSM) 2012–2015 (Figure 2).

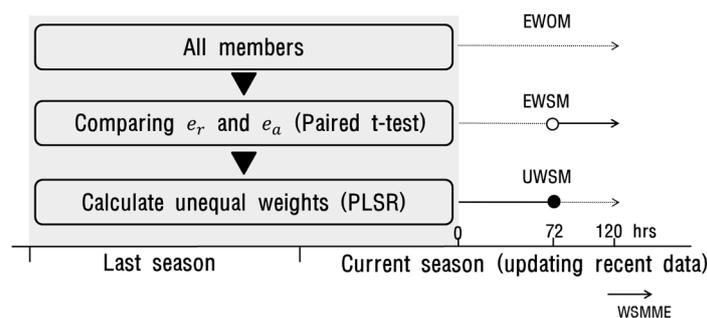


Figure 2. The schematic plot for three Multi-Model Ensemble (MME) tests and Weighted Selective MME (WSMME). WSMME forecasts are the solid lines of Equally Weighting Selective MME (EWSM) and Unequally Weighting Selective MME (UWSM).

Skill score as (3) depicts the proportion of improvement over base forecast [9,14]. Table 2 shows that the track forecast errors from EWSM are significantly smaller than EWOM at 95% confidence level, as the improvement through selecting stage ranges from 2.2 to 5.4% at 24, 48, 72, 96 and 120 h. UWSM outperforms EWSM, which is confirmed by the improved skill ranging from 1.6 to 3.1% at 24, 48 and 72 h at 95% confidence level. The WSMME forecast, which is a combination of UWSM until 72 h and EWSM for the rest, is confirmed to have the most improved forecast skill over EWOM (Figure 2).

$$skill\ score = \frac{error - the\ error\ of\ base\ forecast}{the\ error\ of\ base\ forecast} \times 100\% \quad (3)$$

Table 2. Mean track forecast errors (km) of three different experiments for MME during 2012–2015. The EWSM (UWSM)’s skill scores (%) is relative to EWOM (EWSM).

Forecast Time Range (h)	Number of Cases	EWOM		EWSM		UWSM	
		km	km	%	km	%	
24	1803	75.3	72.7	3.4 **	70.5	3.1 **	
48	1438	130.5	125.2	4.0 **	122.1	2.5 **	
72	1098	191.6	185.9	3.0 **	183.0	1.6 **	
96	808	279.0	263.9	5.4 **	263.6	0.1	
120	568	377.2	369.0	2.2 **	400.8	−8.6	

** Indicate EWSM (UWSM)’s errors are significantly smaller than EWOM (EWSM)’s at 99% confidence level.

3.1.2. Comparison with Numerical Models

Table 3 shows the comparison of the forecast skills between WSMME and each individual model. WSMME exhibits significantly smaller error distances than any other model even at 99% confidence level for 24, 48, 72, 96 and 120 h during 2012–2015. The improvement of WSMME skill is seen to be 5.7–40.9% over the numerical models. Boxplots help assess the difference of the model performances in detail [25]. A boxplot contain 5 values and outliers. The 0.5 quantile (median) is located in the center of the boxplot and the 0.25 and 0.75 quantiles (hinges) are also noted in the box. Whiskers from the box extend to the minimum and maximum values that do not exceed a length of 1.5 times the box length from each hinge. WSMME indicates a narrower range of error distribution than any other model at 24, 48, 72, 96 and 120 h (Figure 3). Also, the outliers of WSMME are mostly smaller than those of other numerical models.

Table 3. Mean track forecast errors (km) of WSMME and each model during 2012–2015. WSMME’s skill scores (%) is relative to each model.

Forecast Time Range (h)	Number of Cases	WSMME			GDAPS		GFS		IFS		GSM	
		km	km	%	km	%	km	%	km	%		
24	1722	69.9	118.1	40.9 **	82.7	15.5 **	82.2	15.0 **	106.5	34.4 **		
48	1360	122.2	196.3	37.8 **	139.3	12.3 **	135.6	9.9 **	185.8	34.2 **		
72	1017	183.1	288.0	36.4 **	209.9	12.8 **	201.8	9.3 **	278.1	34.2 **		
96	734	259.6	381.1	31.9 **	300.6	13.7 **	291.0	10.8 **	406.0	36.1 **		
120	494	367.7	490.8	25.1 **	447.3	17.8 **	389.9	5.7**	523.7	29.8 **		

** Indicates WSMME’s errors are significantly smaller than each other model’s at 99% confidence level.

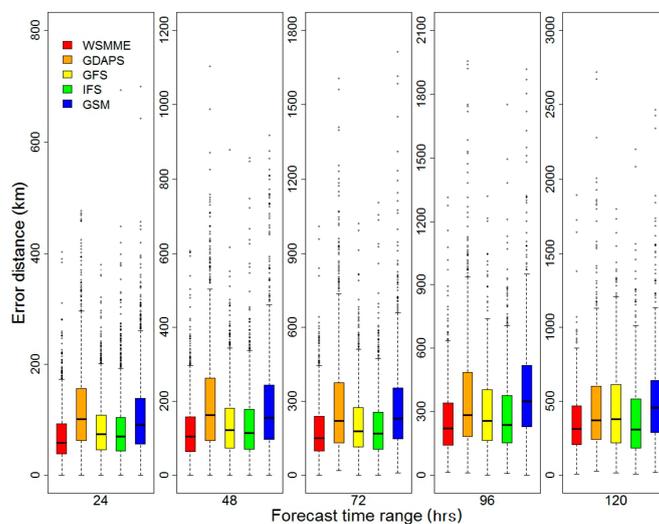


Figure 3. Boxplots for track forecast errors from WSMME and models during 2012–2015.

3.2. Case Study

For showing the details of WSMME application, the track forecast for typhoon Tembin, issued on 25 August 2012, at 00 UTC, is chosen as an example case (Figure 4). In the selecting stage, GDAPS, having the largest error distance among the members for 24 and 48 h forecasts, is removed from participating MME members, while GSM for 72 h and after is removed owing to the relatively slow movement of the forecast TCs. UWSM has a notable westward bias at 120 h than EWSM. UWSM' represents the 120 h forecast position by UWSM when the current case is excluded from the training dataset during 2011–2012. The 120 h forecast position is seen to move eastward closer to the track position of EWSM using equal weights as defined. The larger 120 h forecast errors of UWSM than EWSM imply that UWSM is easily influenced by the training dataset at this longer forecast time range.

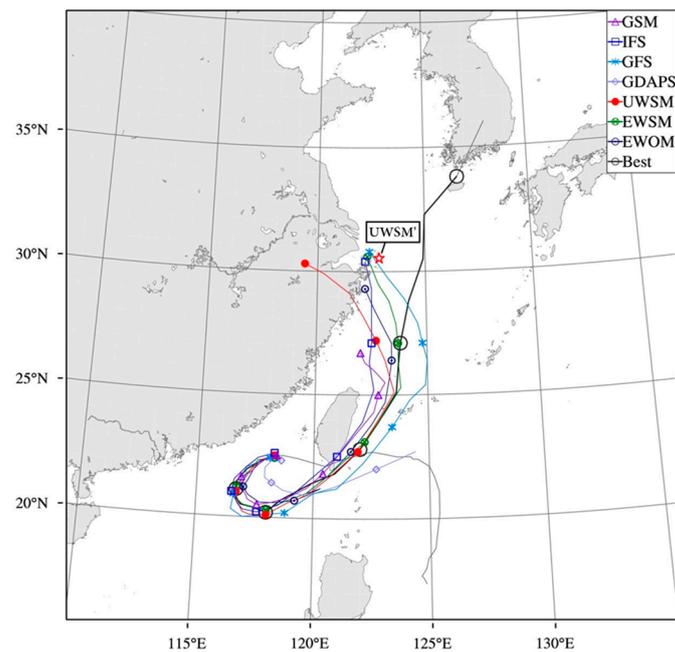


Figure 4. Best track (black line) and track forecast for Tembin (1214) at 0000 UTC 25 August 2012. Markers of each line indicate locations every 24 h. UWSM' (a star shape marker) was produced by using training datasets during 2011–2012 except for this case.

Figure 5 shows the kernel density distributions of longitudinal increments at forecast ranges between 24 and 120 h. Kernel density estimation is made by smoothing the frequency distribution of standardized longitudinal increments [23]. Gaussian kernel is used for this. Bandwidth (h), controlling smoothness of kernel density estimation, is calculated by Silverman's rule of thumb [26] as Equation (4). s is standard deviation and the Interquartile Range (IQR) is the difference between the upper and lower quartiles of the values in data. Minimum value between $0.9s$ and $\frac{2}{3}IQR$ divided by $n^{0.2}$ is h , where n is the number of values. Small (large) h results in rough (monotonous) density distribution. All kernel density estimations at 24 h show similar distributions (Figure 5a). On the other hand, the annual variation of density distributions at 120 h is seen to be larger than at 24 h (Figure 5). This confirms that the track pattern at a longer forecast time range significantly varies by year, which could be understood as the influence of interannual climate variability.

$$h = \frac{\min\{0.9s, \frac{2}{3}IQR\}}{n^{0.2}} \quad (4)$$

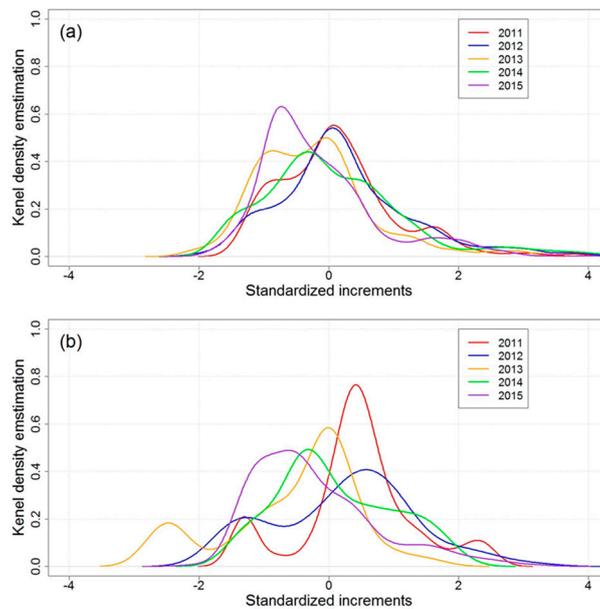


Figure 5. Kernel density estimation for standardized longitude increments at (a) 24 and (b) 120 h.

Conclusively, the UWSM outperforms EWSM at shorter forecast time ranges, but not at longer forecast time ranges, possibly due to larger influence of climate variability as well as the history of model's own improvements. Both influential factors could make the training datasets less effective to find realistic weights on participating MME members.

4. Summary and Discussion

In this study, WSMME is developed and evaluated in pursuit of more accurate TC track forecasting in the WNP. It is found that an approach using unequal weights through the PLSR method significantly improves MME skill for up to 72 h forecasts, but not thereafter. For this approach to improve its forecast skill at 96 and 120 h, consistent climate environments might be needed. The less skillful MME at 96 and 120 h by an unequal weighting approach is thought of as the consequence of influences by interannual climate variability. As pointed out by Krishnamurti et al. [27], the unequal weighting approach has a limitation to new observations by inexperienced values contained in the training dataset. In future studies, synoptic environment fluctuating by years could be considered for better MME forecast after 96 h.

WSMME forecast is a combination of UWSM until 72 h and EWSM after that. WSMME approach significantly outperforms individual models at all forecast time ranges. This suggests that WSMME could provide a good track reference for operational TC forecasters at least in the WNP. It is certain that the MME technique based on a larger spread of the forecasts provides more reliable uncertainty information. Forecast TC position is taken from the mean value of each probability distribution. Here, the spread also represents the confidence of the track forecast. A forecast error is normally small when the forecast spread is small [5,28]. MME spread can also be useful in estimating the regions of TC passage. Subsequent studies may include probabilistic information in association with WSMME forecast, which will greatly enhance the value of MME forecast.

Acknowledgments: This research was supported by the National Typhoon Center and the National Institute of Meteorological Sciences at Korea Meteorological Administration (NIMS-2016-3100).

Author Contributions: S.J., N.-Y.K. and W.L. conceived and designed the experiments; S.J. performed the experiments and analyzed the data; S.J. and W.L. contributed materials; S.J. wrote the initial manuscript; N.-Y. K. modified the manuscript; N.-Y.K., W.L. and Y.C. contributed to manuscript improvement and revisions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kang, N.-Y.; Elsner, J.B. Climate mechanism for stronger typhoons in a warmer world. *J. Clim.* **2016**, *29*, 1051–1057. [[CrossRef](#)]
2. Mitchell, T.; Guha-Sapir, D.; Hall, J.; Lovell, E.; Muir-Wood, R.; Norris, S.L.; Wallemacq, P. Setting, measuring and monitoring targets for reducing disaster risk. In *Recommendations for Post-2015 International Policy Frameworks*; Overseas Development Institute: London, UK, 2014; pp. 12–28.
3. Elsberry, R.L. Advances in research and forecasting of tropical cyclones from 1963–2013. *Asia Pac. J. Atmos. Sci.* **2014**, *50*, 3–16. [[CrossRef](#)]
4. Heming, J.; Goerss, J. Chap.10 Track and structure forecasts of tropical cyclones. In *Global Perspectives on Tropical Cyclones*; Chan, J.C.L., Kepert, J.D., Eds.; World Scientific: Singapore, 2010; Volume 4, pp. 287–323.
5. Elsberry, R.L.; Carr, L.E. Consensus of Dynamical Tropical Cyclone Track Forecasts-Errors versus Spread. *Mon. Weather Rev.* **2000**, *128*, 4131–4138. [[CrossRef](#)]
6. Carr, L.E.; Elsberry, R.L. Dynamical tropical cyclone track forecast errors Part I: Tropical region error sources. *Weather Forecast.* **2000**, *15*, 647–661. [[CrossRef](#)]
7. Carr, L.E.; Elsberry, R.L. Dynamical tropical cyclone track forecast errors part II: Midlatitude circulation influences. *Weather Forecast.* **2000**, *15*, 662–681. [[CrossRef](#)]
8. Carr, L.E.; Elsberry, R.L. Beta test of the systematic approach expert system prototype as a tropical cyclone track forecasting aid. *Weather Forecast.* **2001**, *16*, 355–368. [[CrossRef](#)]
9. Sampson, C.R.; Knaff, J.A.; Fukada, E.M. Operational evaluation of a selective consensus in the western North Pacific basin. *Weather Forecast.* **2007**, *22*, 671–675. [[CrossRef](#)]
10. Qi, L.; Yu, H.; Chen, P. Selective ensemble-mean technique for tropical cyclone track forecast by using ensemble prediction systems. *Q. J. R. Meteorol. Soc.* **2013**, *140*, 805–813. [[CrossRef](#)]
11. Jun, S.; Lee, W.; Kang, K.R.; Yun, W.T. Development of the selected multi-model consensus technique for the tropical cyclone track forecast in the western North Pacific. *Atmosphere* **2015**, *25*, 375–387. (In Korean) [[CrossRef](#)]
12. Goerss, J.S. Tropical cyclone track forecasts using an ensemble of dynamical models. *Mon. Weather Rev.* **2000**, *128*, 1187–1193. [[CrossRef](#)]
13. Sampson, C.R.; Goerss, J.S.; Weber, H.C. Operational performance of a new barotropic model (WBAR) in the western North Pacific basin. *Weather Forecast.* **2006**, *21*, 656–662. [[CrossRef](#)]
14. Sampson, C.R.; Franklin, J.L.; Knaff, J.A.; DeMaria, M. Experiments with a simple tropical cyclone intensity consensus. *Weather Forecast.* **2008**, *23*, 304–312. [[CrossRef](#)]
15. Krishnamurti, T.N.; Kishtawal, C.M.; LaRow, T.E.; Bachiochi, D.R.; Zhang, Z.; Williford, C.E.; Gadgil, S.; Surendran, S. Improved weather and seasonal climate forecasts from multimodel superensemble. *Science* **1999**, *285*, 1548–1550. [[CrossRef](#)] [[PubMed](#)]
16. Williford, C.E.; Krishnamurti, T.N.; Correa Torres, R.; Cocke, S.; Christidis, Z.; Vijaya Kumar, T.S. Real-time multimodel superensemble forecasts of Atlantic tropical systems of 1999. *Mon. Weather Rev.* **2003**, *131*, 1878–1894. [[CrossRef](#)]
17. Jun, S.; Lee, W.; Kang, K.R.; Byun, K.-Y.; Kim, J.; Yun, W.T. Applicability of the Superensemble to the tropical cyclone track forecasts in western North Pacific. *Asia Pac. J. Atmos.* **2015**, *51*, 39–48. [[CrossRef](#)]
18. Wold, S.; Sjostrom, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemometr. Intell. Lab.* **2001**, *58*, 109–130. [[CrossRef](#)]
19. Goerss, J.S.; Sampson, C.R.; Gross, J.M. A history of western North Pacific tropical cyclone track forecast skill. *Weather Forecast.* **2004**, *19*, 633–638. [[CrossRef](#)]
20. Cangialosi, J.P.; Franklin, J.L. *2014 National Hurricane Center Forecast Verification Report*; National Hurricane Center: Miami, FL, USA, 2015.
21. Gilleland, E. *Confidence Intervals for Forecast Verification: Practical Considerations*; National Center for Atmospheric Research: Boulder, CO, USA, 2010.
22. Abernethy, S.D.; DeMaria, M. Verification of a nested barotropic hurricane track forecast model (VICBAR). *Mon. Weather Rev.* **1994**, *122*, 2804–2815. [[CrossRef](#)]
23. Wilks, D.S. *Statistical Methods in the Atmospheric Sciences*, 2nd ed.; Dmowska, R., Hartmann, D., Rossby, H.T., Eds.; Academic Press: Cambridge, MA, USA, 2006; Volume 91.

24. Mevik, B.H.; Wehrens, R. The PLS package: Principal component and partial least squares regression in R. *J. Stat. Softw.* **2007**, *18*, 1–24. [[CrossRef](#)]
25. World Meteorological Organization. *Verification Methods for Tropical Cyclone Forecasts*; World Meteorological Organization: Geneva, Switzerland, 2013.
26. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; Chemical Rubber Company: Boca Raton, FL, USA, 1986; Volume 26.
27. Krishnamurti, T.N.; Kumar, V.; Simon, A.; Bhardwaj, A.; Ghosh, T.; Ross, R. A review of multimodel superensemble forecasting for weather, seasonal climate, and hurricanes. *Rev. Geophys.* **2016**, *54*, 336–377. [[CrossRef](#)]
28. Goerss, J.S. Prediction of consensus tropical cyclone track forecast error. *Mon. Weather Rev.* **2007**, *135*, 1985–1993. [[CrossRef](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).