ADVANCES IN ATMOSPHERIC SCIENCES, VOL. 30, NO. 5, 2013, 1260–1274

Technical Note on a Track-pattern-based Model for Predicting Seasonal Tropical Cyclone Activity over the Western North Pacific

Chang-Hoi HO[1], Joo-Hong KIM*[2], Hyeong-Seog KIM[3], Woosuk CHOI[1], Min-Hee LEE[1], Hee-Dong YOO[4], Tae-Ryong KIM[5], and Sangwook PARK[5]

[1]*School of Earth and Environmental Sciences, Seoul National University, Seoul, Korea*

[2]*Korea Polar Research Institute, Incheon, Korea*

[3]*Atmospheric and Oceanic Science, Princeton University and NOAA/GFDL, Princeton, NJ, USA*

[4]*Forecast Policy Division, Korea Meteorological Administration, Seoul, Korea*

[5]*National Typhoon Center, Korea Meteorological Administration, Jeju-do, Korea*

(Received 20 September 2012; revised 1 December 2012; accepted 8 January 2013)

ABSTRACT

Recently, the National Typhoon Center (NTC) at the Korea Meteorological Administration launched a track-pattern-based model that predicts the horizontal distribution of tropical cyclone (TC) track density from June to October. This model is the first approach to target seasonal TC track clusters covering the entire western North Pacific (WNP) basin, and may represent a milestone for seasonal TC forecasting, using a simple statistical method that can be applied at weather operation centers. In this note, we describe the procedure of the track-pattern-based model with brief technical background to provide practical information on the use and operation of the model. The model comprises three major steps. First, long-term data of WNP TC tracks reveal seven climatological track clusters. Second, the TC counts for each cluster are predicted using a hybrid statistical–dynamical method, using the seasonal prediction of large-scale environments. Third, the final forecast map of track density is constructed by merging the spatial probabilities of the seven clusters and applying necessary bias corrections. Although the model is developed to issue the seasonal forecast in mid-May, it can be applied to alternative dates and target seasons following the procedure described in this note. Work continues on establishing an automatic system for this model at the NTC.

Key words: tropical cyclone, western North Pacific, seasonal forecast, track-pattern-based model, hybrid statistical–dynamical approach

Citation: Ho, C.-H., J.-H. Kim, H.-S. Kim, W. Choi, M.-H. Lee, H.-D. Yoo, T.-R. Kim, and S. Park, 2013: Technical note on a track-pattern-based model for predicting seasonal tropical activity over the western North Pacific. *Adv. Atmos. Sci.*, **30**(5), 1260–1274, doi: 10.1007/s00376-013-2237-6.

1. Introduction

One of the major functions of the meteorological agencies responsible for forecasting tropical cyclones (TCs) around the world is to issue accurate and informative seasonal cyclone predictions. It has become necessary to develop a seasonal TC prediction system that ensures high predictability, in order to minimize the social and economic damage to the target area caused by landfalling (and/or approaching) TCs. Various seasonal prediction models of TC activity have been developed using statistical methods, based on the empirical relationship between observed TC activity and pre-seasonal large-scale fields (e.g. Gray et al., 1992; Chan et al., 1998; Chu and Zhao, 2007; Kim et al., 2010; Zhao et al., 2010b). Parallel to the statistical methods, the dynamic approach of detecting and tracking simulated TCs in high-resolution nu-

*Corresponding author: Joo-Hong KIM, jhkim004@gmail.com

merical models has shown substantial recent progress, along with increasing computational power Vitart et al., 2007; Camargo and Barnston, 2009; Zhao et al., 2010a). Thus far, these two approaches have been widely applied to existing predictive models of seasonal TC activity, with particular emphasis on the frequency of seasonal TCs.

Recently, a track-pattern-based model that targets seasonal (June through October, JJASO hereafter) TC activity in the western North Pacific (WNP) was developed at Seoul National University of Korea (Kim et al., 2012a). Methodologically, this model is classified as a statistical–dynamical approach. Development of the statistical–dynamical model is the topic of several recent studies Wang et al., 2009; Vecchi et al., 2011). The hybrid approach has the potential to improve predictions in that it combines the respective advantages of both statistical and dynamic approaches. The key idea of this model is to separately predict seasonal TC counts for a finite number of track clusters, and build up a forecast map of track density of TCs by combining all track clusters over the entire basin. The track-pattern-based model therefore has unique merit in providing predictions for the spatial TC track density anomaly in the WNP.

The track-pattern-based model was implemented at the National Typhoon Center (NTC) of the Korean Meteorological Administration, and its first experimental forecast for the TC season of 2010 was found to be successful (Kim et al., 2012b). Subsequently, there has been continuous demand to provide technical guidelines for the track-pattern-based model. Separately from such requests, it is further necessary to reintroduce the model via this technical note, because several revisions were made for the input data from the the National Centers for Environmental Prediction (NCEP) Climate Forecast System (CFS) forecasts and the method of incorporating the predictors into the statistical model, which differ from the earlier version described in Kim et al. (2012a, b). In addition, these revisions permit the model to operate automatically in quasi-real-time. To efficiently operate this seasonal TC forecast system, it is essential for users (e.g. TC forecasters) to have knowledge of the model structure and the capability to handle the model code or process autonomously. As such, this note will assist TC forecasters who wish to run and modify the model.

This note is organized as follows. Section 2 provides an overview of the hierarchical model structure and forecast flow chart. Section 3 describes the model pre-processing, which involves the clustering technique and the preparation of dynamic forecast datasets. Section 4 introduces the core of the hybrid statistical–dynamical prediction module. Section 5 details the post-processing for combining the model predictions and observed track clusters. Section 6 introduces two examples of TC forecast in 2010 and 2011. Finally, section 7 summarizes the content of this note.

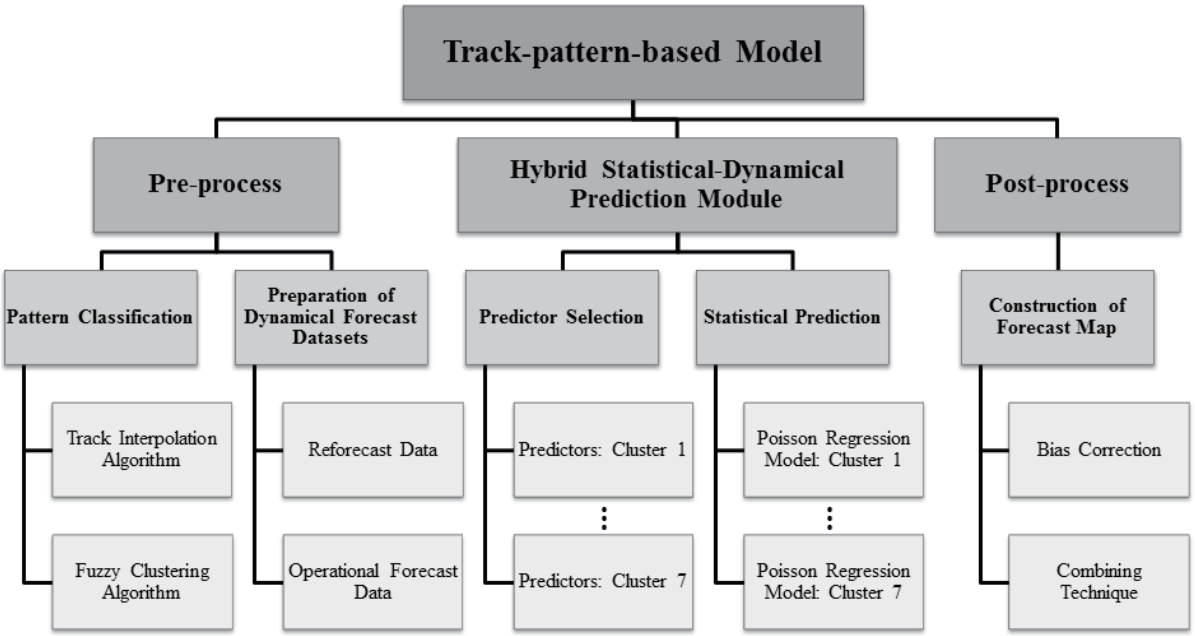

Fig. 1. Hierarchical structure of the track-pattern-based model.

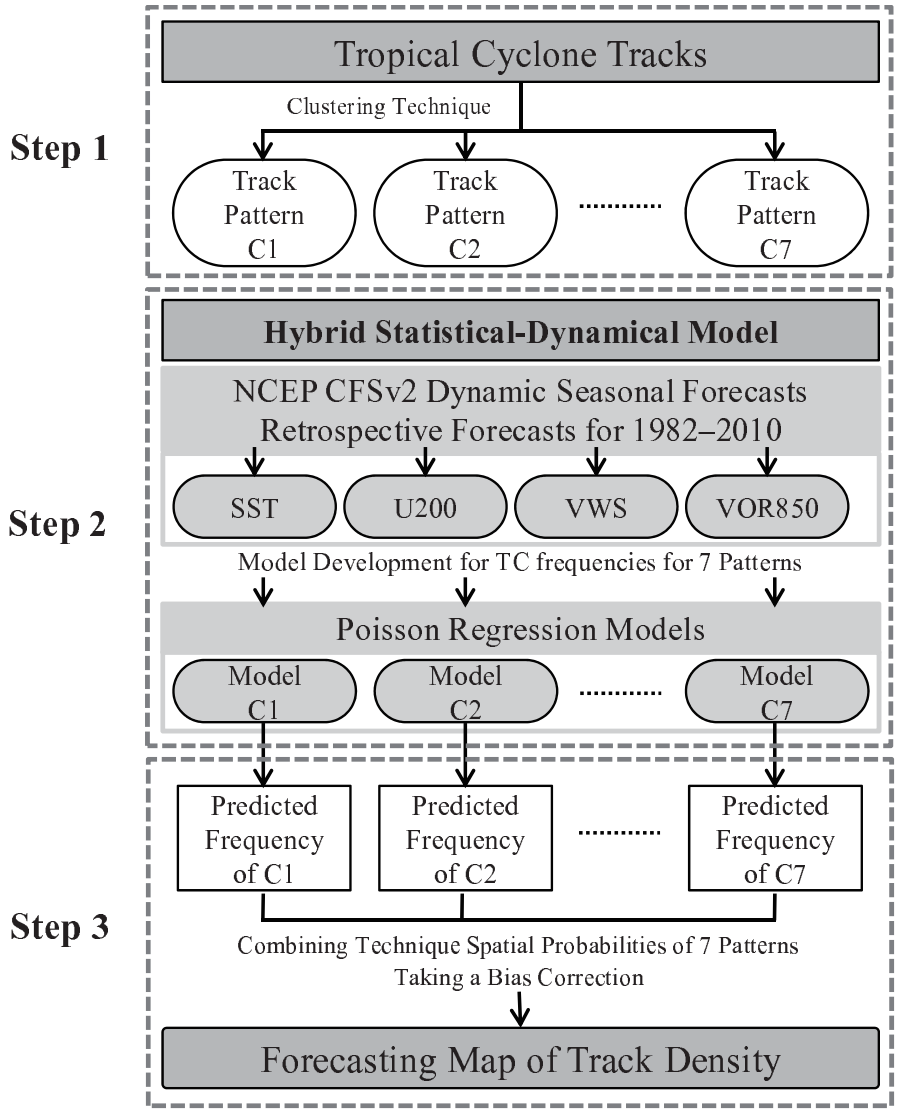

Fig. 2. Flow diagram for the three main steps of the track-pattern-based model.

2. Overview: hierarchical model structure and forecast flow chart

To begin this technical note, this section describes the hierarchical structure of the forecast system and the forecast flow chart. Figure 1 depicts the hierarchical structure of the track-pattern-based model. The model sub-programs can be classified into pre-processing, hybrid statistical–dynamical prediction module, and post-processing. The pre-processing includes programs for the pattern classification of TC tracks by a fuzzy clustering algorithm and scripts to download dynamical forecast datasets. The hybrid statistical–dynamical prediction module can be operated with the input datasets prepared in the pre-processing. With respect to each cluster, there is a program that selects the best predictors and statistical prediction model based on a Poisson regression method in the prediction module. The post-processing comprises the construction of a forecast map through a bias correction and combining technique.

The forecast flow chart of the track-pattern-based model is shown in Fig. 2. The flow is a step-by-step operation of the three modules described in Fig. 1. First, TC track clusters are obtained through a fuzzy clustering technique to objectively divide the WNP basin into multiple forecast areas according to individual track clusters (Kim et al., 2011). Second, the hybrid statistical–dynamical model was developed to predict seasonal TC counts per track cluster, which is based on the relationship between observed TC counts in each cluster and forecasted large-scale environments from the NCEP CFS. Third, we construct the final forecast map of the basin-wide TC track density by merging

the resultant track densities of all clusters. The track density of each cluster is obtained by weighting the climatological track density of the cluster by the forecast TC counts. In the following sections, these three steps are described with technical instructions to facilitate the model operation by users.

3. Pre-processing

3.1 *Fuzzy clustering of TC tracks* (*Step 1*)

3.1.1 *Description*

Step 1 involves pattern classification of TC tracks over the WNP to provide the basis of the track-pattern-based model (Fig. 2). This step implies that the model is set to focus on forecasting TC track density (i.e. the frequency of the TC pathway) based on the finite number of clusters of TC tracks. There have been several approaches to clustering TC tracks: a mixture Gaussian model (Camargo et al., 2007; Chu et al., 2010b); a k-means clustering by mass moments (Nakamura et al., 2009), and a fuzzy c-means clustering (Kim et al., 2011). Among these methods, this model adopts the fuzzy c-means clustering algorithm. The fuzzy clustering can yield more reliable classification for fuzzy objects such as TC tracks that have varying shapes and geographical pathways. Kim et al. (2011) applied the fuzzy clustering method to 855 historical WNP TC tracks during JJASO seasons from 1965 to 2006, producing seven representative TC track clusters. The present model was developed from these seven track clusters.

Figure 3 illustrates a flow chart of the fuzzy clustering procedure to establish the seven TC track clusters. At first, an interpolation algorithm was used along the track, which retained the original track shape as much as possible, and all TCs of various observational durations are converted to have 21 uniform longitude/latitude points (i.e. 20 line segments) at regular intervals. Twenty-one points are used with simple consideration of the mean lifetime of TCs ($\sim$ five days) and the observational interval of four times daily (Kim et al., 2011). All interpolated TC positions are then transposed to one-column vectors ($\boldsymbol{x}_k, k = 1, \ldots, 855$), thereby providing a suitable format for the fuzzy clustering algorithm. Once the input data are prepared, the fuzzy clustering algorithm computes the cluster centers ($\boldsymbol{c}_i$) and membership coefficients (μ_{ik}) of each TC track to all clusters. Here, the membership coefficient (range 0–1) represents the strength of membership of the kth TC (i.e. $\boldsymbol{x}_k$) to the ith cluster; the cluster center ($\boldsymbol{c}_i$) denotes the mean TC track of the ith cluster, which is the membership coefficient-weighted average of all TC tracks for the ith cluster. These parameters depend on each other, and so they can be obtained through an iterative method. The iteration is continued until the c-means functional (J) is minimized under two constraints: the membership coefficients should be greater than 0; and their sum for a certain cluster must be equal to 1 [see Kim et al., 2011 for details]. The result of the fuzzy clustering algorithm is the seven cluster centers and the membership coefficients of all TC tracks to the seven clusters. To clearly divide the TC tracks into seven mutually exclusive clusters, each is allocated to a cluster where the TC track shows the largest membership coefficient. Figure 4 shows the tracks and densities of the seven clusters (C1–C7). Each cluster has inherent characteristics for its shape and statistical properties (Kim et al., 2011).

Once the basic TC track clusters are determined, it is not necessary to repeat the clustering procedures whenever the prediction is issued, since the TC track clusters are nearly static in a climatological sense (Chu et al., 2010a; Kim et al., 2011; Kim et al., 2012b). Even if TC track data are updated every year with new TC observations, the seven basic TC track clusters can be conserved by just assigning the new TC tracks to one of the pre-existing clusters. This is easily achieved by calculating the seven membership coefficients of a new TC track relative to the seven existing cluster centers, and then assigning it to the cluster that shows the largest membership coefficient (Kim et al., 2012b).

3.1.2 *Technical instruction*

Users need to prepare historical TC best tracks in the WNP basin. There are four best track sources for the target basin: the Regional Specialized Meteorological Center, Tokyo (http://www.jma.go.jp/en/ typh); the Joint Typhoon Warning Center (http://www.usno.navy.mil/JTWC, Chu et al., 2002); the Shanghai Typhoon Institute/China Meteorological Administration (http://www.typhoon.gov.cn); and the Hong Kong Observatory (http:// www.hko.gov.hk). The interpolation program reads the six-hourly best track locations (N) per TC from the first location that achieved tropical storm (TS) intensity to the last TS location, and then interpolates them into 21 segments (M) with equal length (Kim et al., 2011). The FORTRAN-style pseudocode of the core algorithm of the track interpolation procedure is provided in the supplementary material (Fig. S1, available online from http://www.iapjournals.ac.cn/aas). To test this program, users may change the input best track data and the number of interpolated segments.

The MATLAB "Fuzzy Clustering Toolbox" is used to perform the fuzzy c-means clustering (Balasko et al., 2005). The main program incorporates the input data

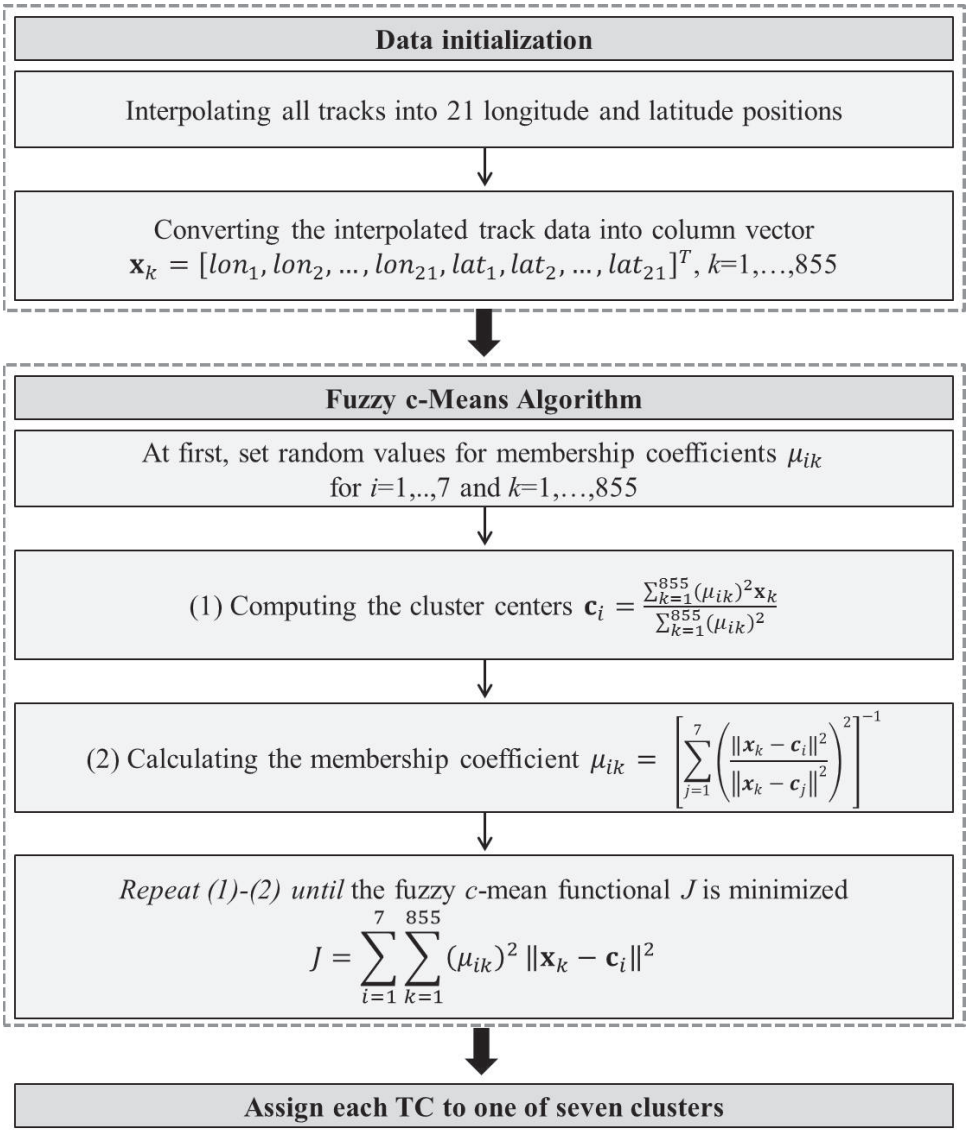

Fig. 3. Flow diagram for Step 1.

of interpolated locations of historical TCs and calls the fuzzy *c*-means clustering algorithm illustrated in Fig. 3. An additional routine for measuring cluster validity determines the optimal cluster number, which is seven in our case. As all programs are publicly available, anybody can repeat our clustering analysis without much difficulty.

3.2 *Preparation of dynamic forecast datasets (first part of Step 2)*

3.2.1 *Description*

This model uses NCEP CFS datasets as the dynamic component to develop the hybrid statistical–dynamical process. The NCEP CFS is a fully coupled global dynamic model to predict monthly-to-seasonal timescale environments. This model is operated at the NCEP Climate Prediction Center to provide seasonal forecasts. The earlier CFS version 1 model (CFSv1) was recently updated to version 2 (CFSv2) in March 2011 (Saha et al., 2012). Following this upgrade, the track-pattern-based model, which was originally developed using CFSv1, was re-built based on the CFSv2 data. The description of this step is therefore based on the upgraded CFSv2.

The NCEP CFSv2 provides two types of seasonal forecast data: one is the "reforecast" for 1982–2010, and the other is the "real-time operational forecast" in the current year. To operate the track-pattern-based model, both datasets are required. The reforecast is utilized to construct the hybrid statistical–dynamical

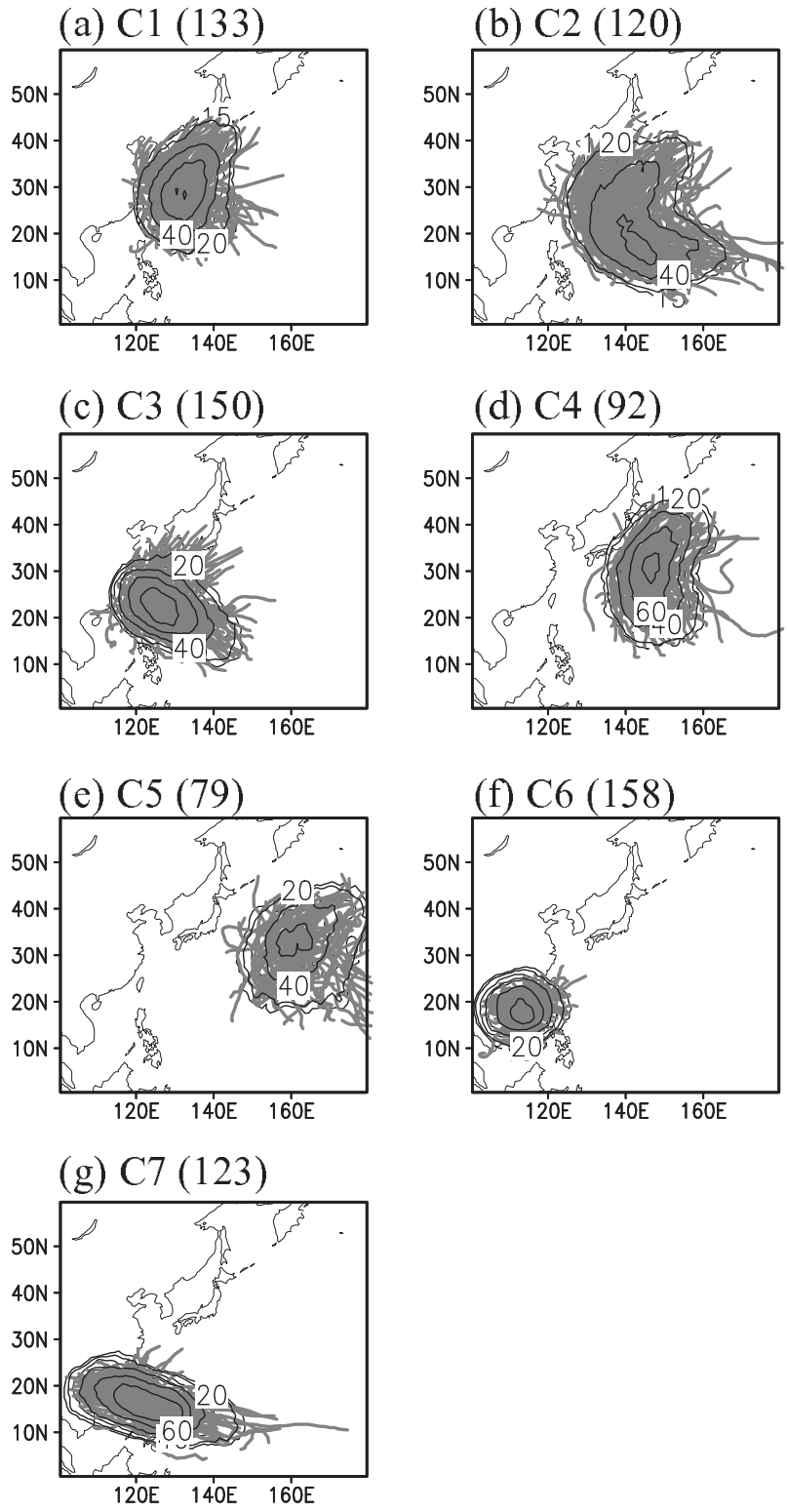

Fig. 4. Climatological TC tracks and their gridded densities for seven clusters (a-g) during 1965–2006. The total number of TCs in each cluster is shown in parentheses.

model to predict the TC counts in each cluster, and the operational forecast data are necessary to issue the seasonal forecast in the current year. CFSv2 yields nine-month forecasts that consist of four ensembles per day with different initial conditions at 0000, 0600, 1200, and 1800 UTC. The reforecast data have nine-month forecasts issued every five days, beginning from 1 January of every year for the period 1982–2010, whereas the operational 9-month forecasts in the current year are issued every day. To construct the track-pattern-based model, 12 ensemble members of CFSv2 issued on 26 April and 1 and 6 May are used.

3.2.2 *Technical instruction*

The CFSv2 datasets can be accessed via the official website (http://cfs.ncep.noaa.gov/cfsv2/downloads.html). While the monthly-mean of nine-month reforecast runs are archived at the National Climatic Data Center (http://nomads.ncdc.noaa.gov/modeldata/cmd_mm_9mon), and the most recent operational forecast data are posted on the NCEP web server (http://nomads.ncep.noaa.gov/pub/data/nccf/com/cfs/prod/cfs) for seven days after issuing the forecast.

The NCEP CFSv2 data are archived as a General Regularly-distributed Information in Binary form version 2 (GRIB-2), which is standardized by the World Meteorological Organization's Commission for Basic Systems. It is a mathematically concise format to efficiently store data, and it is therefore advantageous to release it online. After downloading these datasets, they should be converted to 4-byte binary format to be read by the prediction module. Among many ways to extract GRIB-2 format to binary form, users can choose any appropriate method via the NCEP GRIB-2 use website (http://www.nco.ncep.noaa.gov/pmb/docs/grib2).

4. The hybrid statistical–dynamical prediction module

As the core of the model, the hybrid statistical–dynamical prediction module forecasts TC counts for each track cluster. This method is developed based on the statistical relation between the observed seasonal TC counts during JJASO and the simultaneous large-scale environmental fields taken from dynamical forecasts of a global climate model (Kim et al., 2012a), which is the optimal choice to maximize predictability, as well as to develop a physically meaningful forecast model. In the revised module based on NCEP CFSv2, the model training period for the statistical prediction is extended to the year 2010 (i.e. 1982–2010). These changes in the dynamic forecast data and training period have a substantial influence on the selected predictors and their critical regions, and hindcast skills for all clusters. This section will introduce the major changes in the updated model version.

4.1 ***Predictor selection*** (***second part of Step*** 2)

4.1.1 *Description*

The selection of appropriate predictors is a crucial factor in yielding better prediction performance of the track-pattern-based model. In this model, the predictors are selected by the following four rules.

(1) The predictor candidates are sea surface temperature (SST), 200-hPa zonal wind (U200), vertical wind shear (VWS), and 850-hPa relative vorticity (VOR). These environmental parameters are known to affect the TC genesis and tracks (e.g. Gray, 1968; Wang and Chan, 2002; Kim et al., 2005), and have been widely utilized in previous models to assess seasonal TC activity (e.g. Chan et al., 1998; Kim et al., 2010; Chu et al., 2010a).

(2) Critical regions are determined for each predictor and each cluster by considering correlation patterns between the observed TC counts and the large-scale environments during JJASO. Correlation analysis is performed for the 12-member ensemble mean of NCEP CFSv2 reforecasts as well as the observed environments of the NCEP reanalysis data. The two correlation maps are compared in order to identify statistically significant and physically reasonable regions for candidate predictors of each cluster (boxes in Figs. 5 and 6). The predictor candidates without a definitive critical region are omitted from the predictor set. If there are multiple significant regions in a cluster, the cross-validation tests are performed by changing the critical region, by which the set of critical regions showing the better hindcast skill is determined.

(3) Once the critical regions are selected, final predictors are obtained with respect to each member of the NCEP CFSv2 ensemble forecasts. For each ensemble member, the spatial average of candidate predictor variables within their individual critical regions is calculated using only the grid points for which the correlation coefficient is significant at the 95% confidence level, and where the sign of the correlation is the same as that determined from the correlation patterns for the ensemble mean (as shown in Figs. 5 and 6). It is noted that the significant grid points used to construct final predictors are all different between the 12 ensemble members. For instance, grid points with a significant negative correlation inside a rectangular area are chosen to form the final SST predictor of C2, because negative correlations prevail in the critical domain (Fig. 5c).

(4) The cross-validation tests are also performed to find the optimal combination of predictors showing the better hindcast skill. In addition, a variance infla-

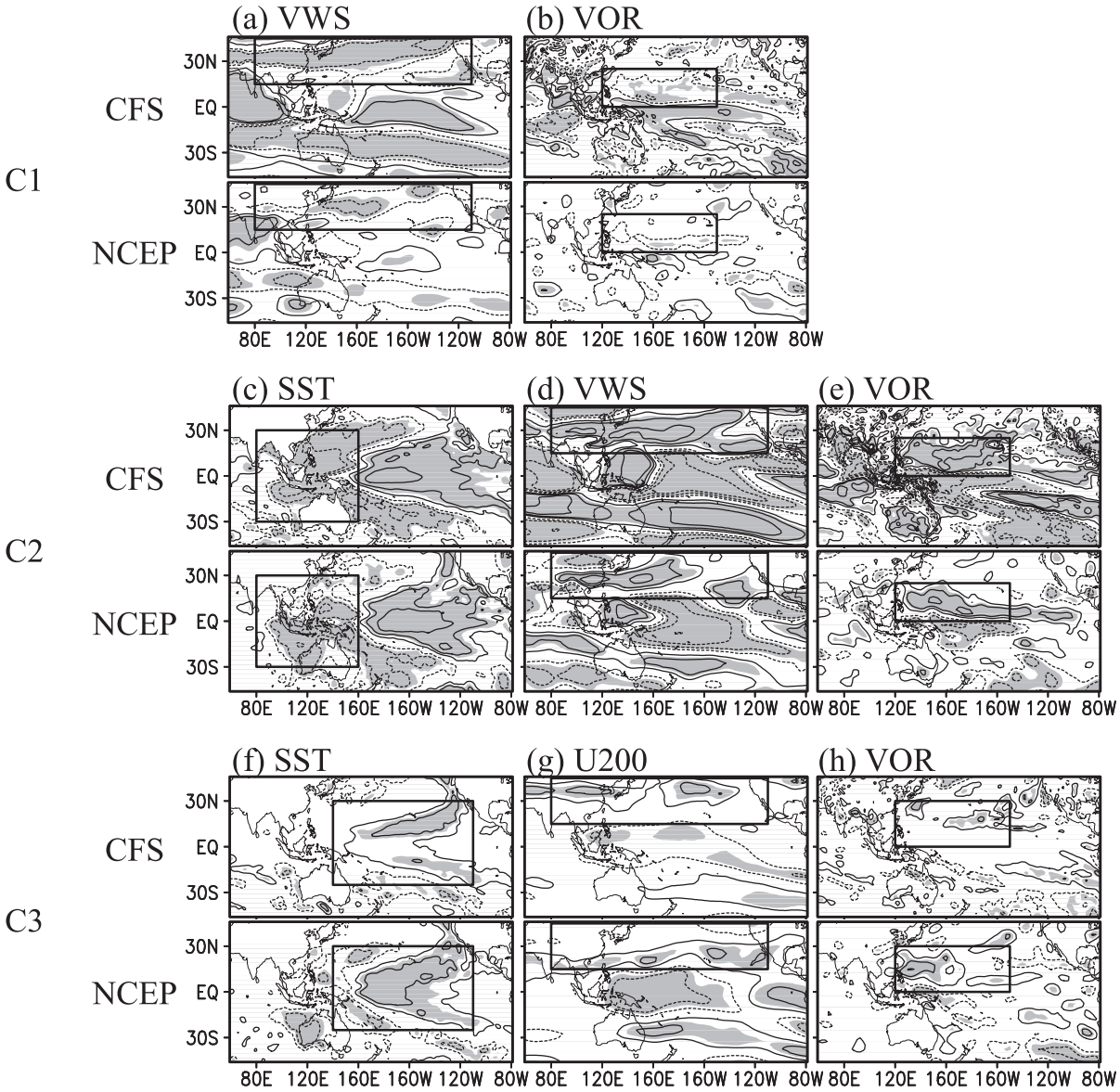

Fig. 5. Critical regions (boxes) and correlations between observed TC counts for C1–C3 and ensemble mean of the NCEP CFSv2 reforecasts (upper panel)/NCEP reanalysis data (lower panel) for each predictor. Contour interval is 0.2, and the zero line is omitted. Statistically significant regions ($p < 0.10$) are shaded.

tion factor (VIF) (Davis et al., 1986) is examined for each ensemble predictor set to avoid multicollinearity among the predictors. If a predictor has a VIF greater than 10, the predictor is dropped from the final predictor set in order to ensure the stability of the regression-based forecast model (e.g. Davis et al., 1986; O'Brien, 2007; Villarini et al., 2011). The final best predictor sets and their critical regions are shown in Figs. 5 and 6.

Figures 5 and 6 show the critical regions selected for each cluster using the described selection method. Those figures also illustrate the correlation maps between the TC counts in each cluster and the ensemble mean of the NCEP CFSv2 reforecasts for JJASO seasons of 1982–2010 (upper box in each panel), as well as those between the TC counts in each cluster and the NCEP reanalysis data (bottom box in each panel). For C1, the VWS and VOR are chosen as the final predictors, and their critical regions are determined, whereby the best hindcast skill is achieved (Figs. 5a and b). Although the SST is discarded from the predictor set due to the multicollinearity, the overall large-scale pattern associated with C1 reflects developing La Niña. In contrast, C2 is strongly related with El Niño, in which the SST, VWS, and VOR are selected as the final predictors (Figs. 5c–e). Next, C3 uses the SST, U200,

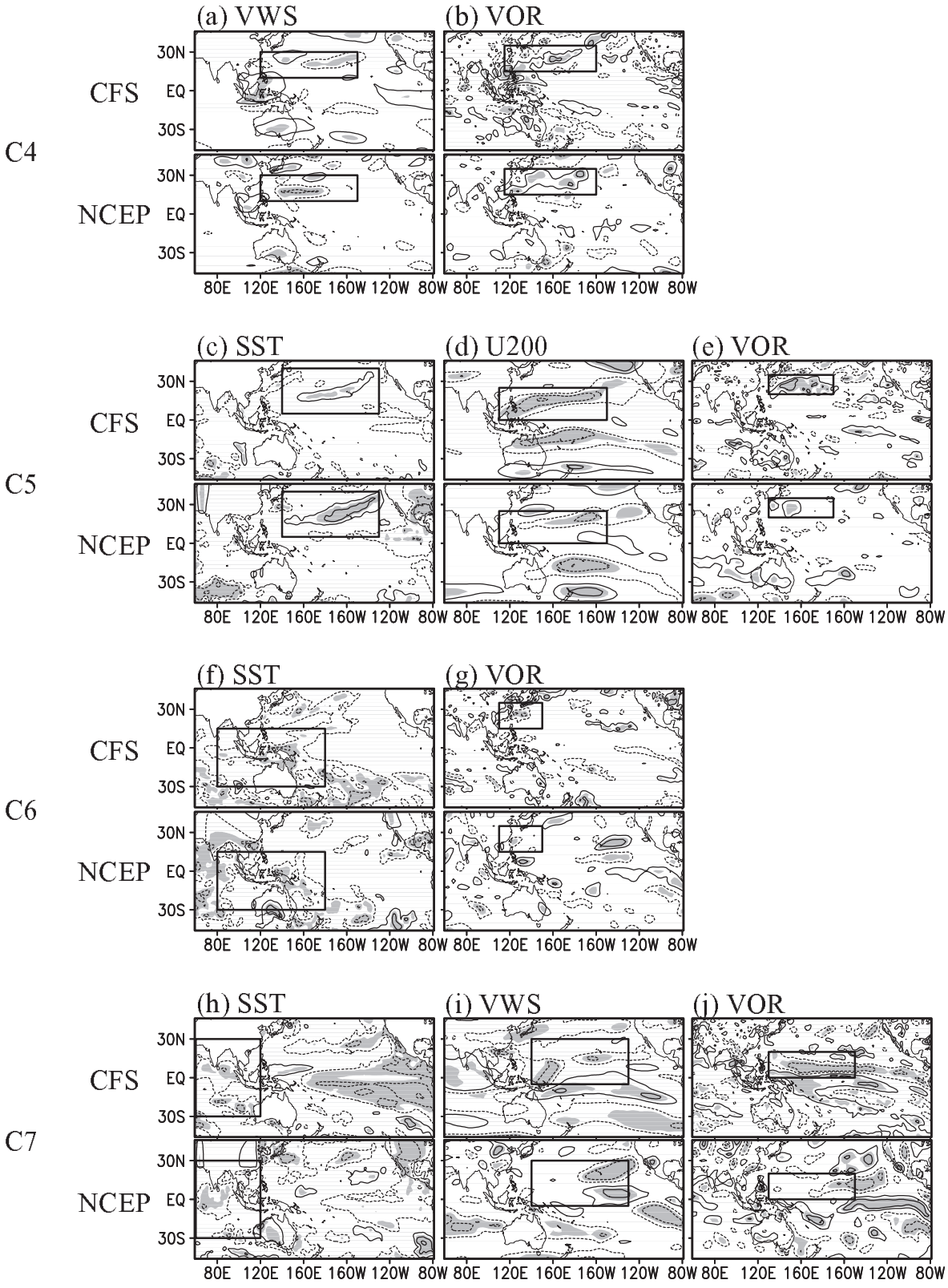

Fig. 6. The same as Fig. 5, but for C4–C7.

and VOR (Figs. 5f–h). In Kim et al. (2012a), based on NCEP CFSv1, all the predictor candidates were incorporated into a final predictor set for C1-C3, whereas some candidate predictors are omitted in this updated version since the changes in the dynamic input data and the training period significantly affect the hindcast results.

Next, the SST and U200 are excluded from the predictor sets for C4 because they have no significant correlations with the TC counts of C4. Accordingly, the VWS and VOR are selected as effective predictors for C4 (Figs. 6a, b). In the previous version, the U50 data over the tropics were added to incorporate the quasi-biennial oscillation (QBO)-relevant predictor for TCs passing offshore southeast of Japan due to its high correlation with the TC counts in C4 (Ho et al., 2009). However, the QBO relationship with TC activity in C4 has weakened recently in both the reanalysis and reforecast data (Fig. 7). In particular, the correlation changes the sign rapidly in the mid-2000s for NCEP CFSv2. The unstable relationship with the tropical U50 in NCEP CFSv2 leads us to discard the U50 data as a predictor for C4.

For C5, the VWS is removed from the predictor sets by the cross-validation test of the predictor combination. As a result, the SST, U200, and VOR are selected, and their critical regions are determined by considering the local impact of some predictors on the TC activity of the cluster (Figs. 6c–e). Meanwhile, the critical regions of C6 comprise only SST and VOR as in the previous model version (Figs. 6f, g). For the SST, we select the negatively correlated region over the Maritime Continent and that east of Taiwan for VOR, because both the observations and the dynamic forecasts yield statistically significant relationships at the selected critical region. Lastly, C7 appears to be related with the La Niña-like SST distribution (Figs. 6h–j), but it is only significant in NCEP CFSv2. Rather, the significant correlation with the SST is found in the tropical Indian Ocean, so the critical region over the Indian Ocean is selected. The critical regions for the VWS and VOR are delineated over the significant regions in the tropical North Pacific. Though the significant correlations for the VWS and VOR are confined in the eastern Pacific domain in the observation due to weaker La Niña, we decide to include them since the inclusion of them improve the hindcast skill.

It is again noted that the final predictor sets are obtained from each ensemble member of the NCEP CFSv2 forecasts, although the critical regions are selected using their ensemble means. Therefore, the grid points used for the final predictor can vary depending on the correlation maps for the individual CFSv2 ensemble members, whereas the critical regions are invariant. As NCEP CFSv2 reforecasts provide four ensembles every fifth day, the same ensemble members are obtained for the final predictor sets from the NCEP CFSv2 reforecasts as well as operational forecasts. Using these ensemble predictor sets, the hybrid statistical–dynamical prediction module can provide ensemble predictions. The ensemble prediction of the hybrid statistical–dynamical prediction module yields more accurate forecasts than the use of single predictions (Kim et al., 2012a).

4.1.2 *Technical instruction*

The predictor selection procedure is not yet automatic, but needs a manual repetition to find a best

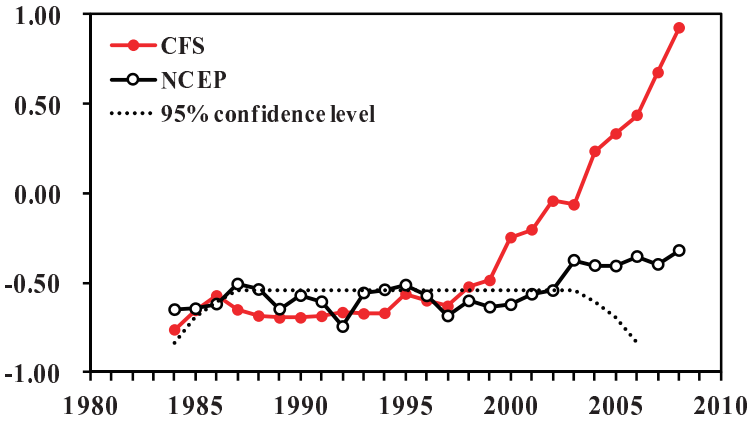

Fig. 7. Eleven-year sliding correlation coefficients between the observed TC counts in C4 and the tropical-mean (10°S–10°N) U50. The number of years used to calculate the correlation coefficients decreases for early three (1982, 1983, and 1984) and late three years (2004, 2005, and 2006) due to the data starting and ending boundaries, respectively. Dotted line denotes the 95% confidence level.

predictor set. Once the correlation maps between TC counts of each cluster and predictor candidate variables are obtained, critical regions for individual predictor candidates can be designed by Rule 2 described above. A somewhat subjective step is involved because there are mostly multiple regions with statistically significant correlations. It is recommended to select a single region as wide as possible under the condition that the region includes statistically significant grids and a same sign with the observed correlation. Shown in Figs. 5 and 6 are our best choices through a careful comparison with the observed correlation maps and further cross-validation tests. However, it should be noted that our choices can neither fully avoid the subjectivity, nor be absolute. Users may explore better critical regions by resizing (or changing) regions. After finishing the selection of critical regions for individual predictor candidate variables, the predictors are made by area-averaging over the grid points where the correlations are statistically significant at the 95% confidence level. The flow of the predictor selection program, in which all user-chosen options are listed and an example of how to set namelist input parameters in a case selecting 12 ensemble members on 26 April, 1 May, and 6 May are illustrated in the supplementary material (Fig. S2, available online from http://www.iapjournals.ac.cn/aas).

With the selection of predictors and observed TC counts, users are ready to compute coefficients of the Poisson regression model. The Poisson regression model will be described in the next subsection. Before freezing model coefficients for the forecast mode, users are advised to cross-validate various combinations of predictors through hindcasts. For example, if there are four candidate variables (e.g. SST, U200, VWS, and VOR), it is possible to get 15 combinations of the predictor set (i.e. ${}_4C_1 + {}_4C_2 + {}_4C_3 + {}_4C_4$, where C is a combination). With respect to each predictor set, all historical years are hindcast by jackknife cross-validation (Wilks, 2006), and then the predictability is assessed by calculating the hindcast correlation with the observed TC counts. In addition, a VIF Davis et al., 1986 is calculated for each predictor, which is necessary to know whether the predictors selected are largely independent. It is recommended to choose the predictor set which shows the highest hindcast correlation with the observed TC counts and has relatively low VIFs for all predictors.

4.2 *Statistical prediction of TC counts for each cluster (third part of Step 2)*

4.2.1 *Description*

Poisson regression is employed as a statistical prediction method for the TC counts of track clusters. This method was shown to be appropriate when the dependent variable is a non-negative integer and individual events being counted are rare, such as the TC counts, and has therefore been widely utilized to predict seasonal TC counts (Elsner and Schmertmann, 1993; McDonell and Holbrook, 2004; Chu and Zhao, 2007; Kim et al., 2010).

The Poisson regression equation can be simply expressed as a log-linear regression form:

$$\widetilde{y} = e^{\sum_{j=1}^{l} \beta_j x_j + \beta_0}, \tag{1}$$

where $\widetilde{y}$ is the predicted value (i.e. the expectation of TC counts); l is the number of predictors; and β_0 is the regression constant. Further, x_j are the predictors (e.g. SST, VWS, U200, VOR, and U50); and β_j the regression coefficients. The regression coefficients and constant can be obtained by maximum likelihood estimation (Wilks, 2006) using the observed TC counts in the seven clusters in addition to the predictor sets from the NCEP CFS reforecast for the training period (1982–2010). Once the regression coefficients and constant are calculated for each cluster, the TC counts for the target forecasting season is calculated using the predictor obtained from the NCEP CFS operational forecast.

4.2.2 *Technical instruction*

The Poisson regression is realized by the generalized linear model regression from the MATLAB "Statistics Toolbox" (MathWorks, 2010). To process input data for the generalized linear model regression function, simple main and sub-programs are developed. The main program reads the predictors and observed TC counts for the training period, while the sub-program incorporates them into the generalized linear model regression function to get the coefficients of the Poisson regression model. Users just need to type the right input data filenames of the predictors and observed TC counts, and then run the main program to get the forecast TC counts for the seven clusters.

5. Post-processing: construction of the forecast map of track density (Step 3)

5.1 *Description*

The post-processing (Step 3) of the track-pattern-based model constructs the final forecast map of the total TC track density and its anomaly from the climatology. This step consists of two procedures: one is the production of the forecast map by combining the climatological TC track densities (shown in Fig. 3) with predicted seasonal TC frequencies; the other is

the production of the final forecast map through bias corrections of the mean and variation. Since the procedures in the post-processing have not been modified from the early model described by Kim et al. (2012a), this note only deals with the essential techniques that help users to obtain the final product of the model, i.e. the forecast map of the TC track density in the WNP.

The first procedure is expressed by the equation:

$$\widetilde{P} = \frac{\sum_{i=1}^{7} N_{\mathrm{C}i} P_{\mathrm{C}i}}{\sum_{i=1}^{7} N_{\mathrm{C}i}} = \frac{\sum_{i=1}^{7} N_{\mathrm{C}i} P_{\mathrm{C}i}}{N_{\mathrm{Total}}} , \qquad (2)$$

where $\widetilde{P}$, $P_{\mathrm{C}i}$, and $N_{\mathrm{C}i}$ are the predicted track density, climatological track density, and predicted TC counts [i.e. $\tilde{y}$ in Eq. (1)], respectively, and i denotes the index for the ith track cluster. This equation represents a simple weighted average of predicted track densities for the seven clusters, where the weighting coefficient is the predicted TC counts of each cluster.

The second procedure is devised to correct potential biases in the mean and variation, which are inherent in the prediction using the climatology. In the track-pattern-based model, these biases originate from the use of the climatological track density. We adopt a guideline for bias correction [Eq.(3)] from Saha (2008) when applying the climatological data to real-time forecasts as

$$\widetilde{P}_{\mathrm{BC}} \widetilde{=} \left(\frac{\tilde{P} - \mu_{\mathrm{model}}}{\sigma_{\mathrm{model}}} \right) \sigma_{\mathrm{obs}} + \mu_{\mathrm{obs}} , \qquad (3)$$

where $\widetilde{P}_{\mathrm{BC}}$ is the bias-corrected track density, μ_{model} and μ_{obs} denote the model-hindcast and observed climatology, respectively, and σ_{model} and σ_{obs} are the model-hindcast and observed standard deviation, respectively.

Via the aforementioned steps, we can produce the forecast map of the total TC track density. Given this total spatial distribution, the anomaly forecast map is easily drawn by subtracting the climatological TC track density.

5.2 *Technical instruction*

The post-processing does not include any sophisticated work. The map construction procedure [i.e. Eq. (2)] is formulated with a FORTRAN code, which reads the climatological TC track density and the forecast TC counts for each cluster. Finally, the bias correction [Eq. (3)] is applied before visualizing the final forecast map.

6. Examples: TC seasonal forecasts in 2010 and 2011

In the three sections above, we described updated technical details of the three main processes of the track-pattern-based model. In this section, the procedures for producing the final forecast map and the forecast verification step are presented in a more easily understandable way, with a schematic flow diagram, using actual step-by-step forecast maps for the two most recent years, i.e. 2010 and 2011 (Fig. 8). In practice, the method of forecast and verification for those years is the same. In Fig. 8, the upper box illustrates the combining technique in Step 3, as formulated in Eq. (2). It is noted that combining the seven predicted TC counts with the climatological TC track densities over the seven clusters is essential to forecasting the spatial distribution of TC activity. Next, the step-by-step forecast processes for 2010 and 2011 are shown in Fig. 8, which better illustrates Step 3 of the model. Figure 8 also shows the verification stage of the seasonal forecast following the TC season.

For the 2010 TC season, Kim et al. (2012b) reviewed the forecast that was produced in May 2010 from the previous version of the model based on CFSv1. The 2010 forecast was evaluated as being successful, because the model predicted the above-normal TC migration near Korea and Japan, as well as the below-normal basin-wide seasonal TC counts in the WNP (Kim et al., 2012b). The successful forecast in 2010 was attributed to the superior performance of the track-pattern-based model in constructing the anomalous track density during ENSO events (Kim et al., 2012a, b) as well as the accurate dynamic forecast of La Niña development by CFSv1. In this note, the 2010 TC season is predicted again by the updated CFSv2 model, as shown in the panels on the left side of Fig. 8. If the spatial pattern of the forecast is compared with that of the observation, the 2010 forecast by the updated model appears better than that of the earlier version [compare with Fig. 6a in Kim et al. (2012b)]. The greater similarity with the observed data in 2010 implies that the updated CFSv2 model shows better performance than the earlier version with CFSv1; the hindcast accuracy of TC counts (i.e. the correlation between the observed and predicted TC counts) for each cluster is slightly improved in the updated version (Table 1); that is, the statistics (e.g. correlation coefficients, RMSEs, and mean squared skill scores) that can be used to measure the reliability of the model forecast (Wilks, 2006) are generally better in the majority of the patterns.

As illustrated on the right side of the flow chart, the forecast map for the 2011 TC season was assessed

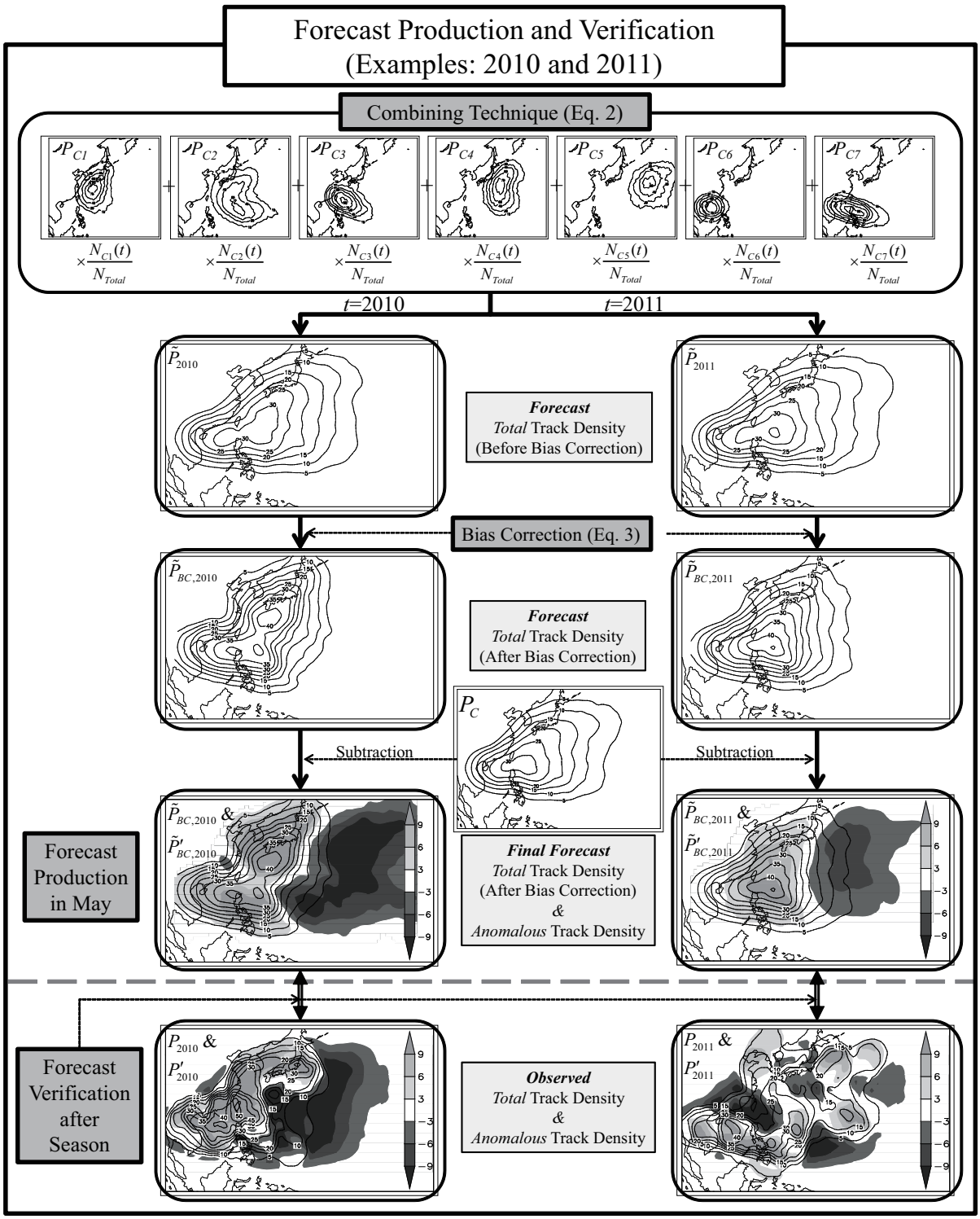

Fig. 8. Illustration of the step-by-step processes of constructing the final forecast map before the TC season, and the forecast verification step after the TC season, for the 2010 and 2011 TC seasons. Contours for the total track density start from 5 with an interval of 5.

Table 1. Correlation coefficients (COR), RMSEs, and mean square skill scores (MSSS) between the observed TC counts and the ensemble mean of hindcasts for the period 1982–2010 for CFSv1 and CFSv2.

Pattern	COR		RMSE		MSSS	
	CFSv1	CFSv2	CFSv1	CFSv2	CFSv1	CFSv2
C1	0.75	0.81	1.30	1.31	0.54	0.53
C2	0.74	0.85	1.44	1.12	0.53	0.72
C3	0.72	0.77	1.21	1.17	0.51	0.58
C4	0.81	0.83	0.85	1.02	0.63	0.64
C5	0.74	0.84	0.96	0.87	0.51	0.61
C6	0.77	0.77	1.28	1.26	0.50	0.53
C7	0.71	0.78	1.11	0.96	0.49	0.59

to be unsuccessful because the observed track density was too complex to be reconstructed by this model, based on the finite number of climatological track clusters. Results from this year demonstrate the limitations of statistical–dynamical predictions of TC counts and construction of forecast maps based on a finite number of climatological track clusters. To mitigate anxiety arising from a potential forecasting failure, future studies should seek to devise a fail-safe method (i.e. ensemble model output statistics, Bayesian approach) within a probabilistic forecasting framework.

7. Summary

This note describes the technical procedures of the track-pattern-based model that has been developed to predict seasonal (JJASO) TC track density over the WNP. This model provides the spatial distribution of seasonal TC tracks covering the entire WNP basin by separately predicting the representative TC track clusters and combining them into one TC-density map. The model shows improved predictive capacity compared with previous models that predict TC counts as a single number over the vast WNP basin, or within limited areas of the basin. We believe that the present model represents a significant advance in the development of seasonal TC forecasting.

The track-pattern-based model is built on seven track clusters, classified using a fuzzy c-means algorithm. TC counts associated with the seven clusters are predicted using a hybrid statistical–dynamical approach, based on the statistical relationship between the observed TC activity and the forecast fields of large-scale environments from NCEP CFSv2. The NCEP CFSv2 operational 9-month forecasts are provided every fifth day from 1 January, allowing the predictions for JJASO to be made from February onwards, and to be continuously updated until late May. After the forecasts are completed for the seven clusters, a prediction map of seasonal TC track density is constructed by weighted average of track densities over the seven clusters [Eq. (2)]. This TC density map then undergoes bias correction [Eq. (3)] before the final forecast is issued.

Recently, dynamic forecasting of seasonal TC activity was attempted using a high-resolution dynamic model that can retrieve the TC trajectory (e.g. Chen and Lin, 2011). However, to date, such dynamic forecasts have been limited to basin-wide TC counts due to the limited ability to forecast seasonal TC tracks. In comparison, the track-pattern-based model proposed here was developed to forecast the distribution of seasonal TC tracks over the WNP (Kim et al., 2012a, b) without requiring a high-resolution dynamic model. Moreover, this proposed model can be operated at lower cost than high-resolution dynamic models. As a result, this track-pattern-based model will provide a good alternative prediction technique until more reliable dynamic prediction models for seasonal TC tracks are developed and implemented.

Acknowledgements. This work was funded by the Korea Meteorological Administration Research and Development Program under Grant CATER 2012–2040. Mr. W. CHOI was supported by the BK21 project of the Korean government.

REFERENCES

Balasko, B., J. Abonyi, and B. Feil, 2005: Fuzzy clustering and data analysis toolbox: For use with Matlab. Department of Process Engineering, University of Veszprem, Hungary, 77pp. [Available online at http://www.mathworks.com/matlabcentral/fileexchange/7486.]

Camargo, S. J., and A. G. Barnston, 2009: Experimental dynamical seasonal forecasts of tropical cyclone activity at IRI. *Wea. Forecasting*, **24**, 472–491.

Camargo, S. J., A. W. Robertson, S. J. Gaffney, P. Smyth, and M. Ghil, 2007: Cluster analysis of typhoon tracks. Part I: General properties. *J. Climate*, **20**, 3635–3653.

Chan, J. C. L., J.-E. Shi, and C.-M. Lam, 1998: Seasonal forecasting of tropical cyclone activity over the west-

ern North Pacific and the South China Sea. *Wea. Forecasting*, **13**, 997–1004.

Chen, J. H., and S.-J. Lin, 2011: The remarkable predictability of inter-annual variability of Atlantic hurricanes during the past decade. *Geophys. Res. Lett.*, **38**, L11804, doi:10.1029/2011GL047629.

Chu, J.-H., C. R. Sampson, A. S. Levine, and E. Fukada, 2002: The Joint Typhoon Warning Center tropical cyclone best-tracks, 1945–2000. Naval Research Laboratory Tech. Rep. NRL/MR/7540-02-16, 112pp.

Chu, P.-S., and X. Zhao, 2007: A Bayesian regression approach for predicting seasonal tropical cyclone activity over the central North Pacific. *J. Climate*, **20**, 4002–4013.

Chu, P.-S., X. Zhao, C.-H. Ho, H.-S. Kim, M.-M. Lu, and J.-H. Kim, 2010a: Bayesian forecasting of seasonal typhoon activity: A track-pattern-oriented categorization approach. *J. Climate*, **23**, 6654–6668.

Chu, P.-S., X. Zhao, and J.-H. Kim, 2010b: Regional typhoon activity as revealed by track patterns and climate change. Vol. 2, *Hurricanes and Climate Change*, J. B. Elsner et al., Eds., Springer, 137–148.

Davis, C. E., J. E. Hyde, S. I. Bangdiwala, and J. J. Nelson, 1986: An example of dependencies among variables in a conditional logistic regression. *Modern Statistical Methods in Chronic Disease Epidemiology*, S. H. Moolgavkar and R. L. Prentice, Eds., Wiley, New York, 140–147.

Elsner, J. B., and C. P. Schmertmann, 1993: Improving extended-range seasonal predictions of intense Atlantic hurricane activity. *Wea. Forecasting*, **8**, 345–351.

Gray, W. M., 1968: Global view of the origin of tropical disturbances and storms. *Mon. Wea. Rev.*, **96**, 669–700.

Gray, W. M., C. W. Landsea, P. W. Mielke, Jr., and K. J. Berry, 1992: Predicting Atlantic seasonal hurricane activity 6-11 months in advance. *Wea. Forecasting*, **7**, 440–455.

Ho, C.-H., H.-S. Kim, J.-H. Jeong, and S.-W. Son, 2009: Influence of stratospheric quasi-biennial oscillation on tropical cyclone tracks in the western North Pacific. *Geophys. Res. Lett.*, **36**, L06702, doi: 10.1029/2009GL037163.

Kim, H.-S., C.-H. Ho, P.-S. Chu, and J.-H. Kim, 2010: Seasonal prediction of summertime tropical cyclone activity over the East China Sea using the least absolute deviation regression and the Poisson regression. *Int. J. Climatol.*, **30**, 210–219.

Kim, H.-S., J.-H. Kim, C.-H. Ho, and P.-S. Chu, 2011: Pattern classification of typhoon tracks using the fuzzy c-means clustering method. *J. Climate*, **24**, 488–508.

Kim, H.-S., C.-H. Ho, J.-H. Kim, and P.-S. Chu, 2012a: Track-pattern-based model for predicting seasonal tropical cyclone activity in the western North Pacific. *J. Climate*, **25**, 4660–4678.

Kim, J.-H., C.-H. Ho, C.-H. Sui, and S. K. Park, 2005: Dipole structure of interannual variations in summertime tropical cyclone activity over East Asia. *J. Climate*, **18**, 5344–5356.

Kim, J.-H., C.-H. Ho, H.-S. Kim, and W. Choi, 2012b: 2010 western North Pacific typhoon season: Seasonal overview and forecast using track-pattern-based model. *Wea. Forecasting*, **27**, 730–743.

MathWorks, 2010: MATLAB statistics toolbox user's guide—version 7.4. The MathWorks, Inc., Natick, MA, USA.

McDonell, K. A., and N. J. Holbrook, 2004: 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.

Nakamura, J., U. Lall, Y. Kushnir, and S. J. Camargo, 2009: Classifying North Atlantic tropical cyclone tracks by mass moments. *J. Climate*, **22**, 5481–5494.

O'Brien, R. M., 2007: A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, **41**, 673–690, doi: 10.1007/s11135-006-9018-6.

Saha, S., 2008: Documentation of operational NCEP CFS data files. NCEP, 7pp. [Available online at http://cfs.ncep.noaa.gov/menu/doc/.]

Saha, S., and Coauthors, 2012: The NCEP climate forecast system version 2. [Available online at http://cfs.ncep.noaa.gov/.]

Vecchi, G. A., and Coauthors, 2011: Statistical-dynamical predictions of seasonal North Atlantic hurricane activity. *Mon. Wea. Rev.*, **139**, 1070–1082.

Villarini, G., G. A. Vecchi, T. R. Knutson, and J. A. Smith, 2011: Is the recorded increase in short-duration North Atlantic tropical storms spurious? *J. Geophys. Res.,* **116**, D10114, doi: 10.1029/2010JD 015493.

Vitart, F., and Coauthors, 2007: Dynamically-based seasonal forecasts of Atlantic tropical storm activity issued in June by EUROSIP. *Geophys. Res. Lett.*, **34**, L16815, doi: 10.1029/2007GL030740.

Wang, B., and J. C. L. Chan, 2002: How strong ENSO events affect tropical storm activity over the western North Pacific. *J. Climate*, **15**, 1643–1658.

Wang, H., and Coauthors, 2009: A statistical forecast model for Atlantic seasonal hurricane activity based on the NCEP dynamical seasonal forecast. *J. Climate*, **22**, 4481–4500.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Science.* 2nd Ed., Academic Press, 627pp.

Zhao, M., I. M. Held, and G. A. Vecchi, 2010a: Retrospective forecasts of the hurricane season using a global atmospheric model assuming persistence of SST anomalies. *Mon. Wea. Rev.*, **138**, 3858–3868.

Zhao, H. K., L. G. Wu, and W. C. Zhou, 2010b: Assessing the influence of the ENSO on tropical cyclone prevailing tracks in the western North Pacific, *Adv. Atmos. Sci.*, **27**(6), 1361–1371, doi: 10.1007/s00376-010-9161-9.

Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00376-013-2237-6.