Prediction and Predictability of Northern Hemisphere Persistent Maxima of 500-hPa Geopotential Height Eddies in GEFS

Bian He[12], Ping Liu[2], Yuejian Zhu[3], Wenting Hu[12]

[1]*State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China*
[2]*School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, New York 11794-5000, USA*
[3]*Environmental Modeling Center, NCEP, NOAA, College Park, Maryland, USA*

Corresponding author: Dr. Bian He, ORCID # 0000-0002-7290-2201

Email: heb@lasg.iap.ac.cn

Abstract

This study estimates the prediction skills associated with the persistent maxima of 500-hPa geopotential height (Z500; PMZ) zonal eddies over the Northern Hemisphere in the long forecast datasets of the Global Ensemble Forecast System (GEFS) version 10. PMZ patterns include not only closed blocking anticyclones that occur more frequently in the Euro-Atlantic-Asia sector (EAAS), but also persistent open ridges and omega-shape blockings that prevail more often over the Pacific-North America sector (PNAS); they potentially extend the predictability of severe weather events such as drought, heat wave and flooding.

PMZ occurrence frequencies in both EAAS and PNAS are predicted overall decreasing with lead time, contrasting the nearly constant frequency for classical blockings in a recent relevant diagnosis. The Brier Skill Score associated with PMZ frequencies is generally higher in the PNAS than in the EAAS, indicating better predictions in the PNAS. The reliability of the forecasts is decreased with lead time in both sectors, particularly at the tail of probability distributions, suggesting some limitations of this GEFS. PMZ events longer than one week have a mean useful skill of nearly 10 lead days by anomaly correlation coefficients (ACCs) being greater than 0.6 in the Northern Hemisphere, about 0.5-1 day more than the average skill of all cases. Among these events, 50% extend useful ACC skills up to 12 days, and another 25% go further beyond. How the better prediction skill of PMZs in the PNAS can help week 2-3 predictions over North America is discussed.

Keywords: persistent atmospheric pattern, prediction skills, 500-hPa geopotential height, medium-range forecast

1. Introduction

The Next Generation Global Prediction System (NGGPS) is being developed by the National Weather Service (NWS) of National Oceanic and Atmospheric Administration (NOAA) in collaboration with other agencies, laboratories and universities. The system aims at addressing growing service demands and improving weather forecasts up to four weeks which is beyond the first kind of prediction limit (Lorenz 1963; 1982). The NGGPS will adopt most of the packages representing physical processes in the current operational system, the Global Ensemble Forecast System (GEFS; Zhu 2005; 2008; Zhou et al., 2017); and the GEFS is newly tested to extend the lead time of operational forecasts from day 16 to 35, the same as the NGGPS' goal (Zhu et al., 2017a; 2017b). Calibrating and evaluating the ensemble forecasts of GEFS at these new ranges is thus a valuable step towards optimizing the NGGPS.

The prediction skill of 500-hPa geopotential height (Z500) is one of the most important metrics to measure the capability of a system for short and medium-range forecasts in operational numerical weather prediction centers. The skill of predicting Z500 for 7 days by the anomaly correlation coefficient (ACC) was improved from 0.4 in 1981 to 0.7 in 2017 in the European Centre for Medium-Range Weather Forecasts (ECMWF). The updated predictability of Z500 is up to 10 days when the ACC of 0.6 is considered as the lower limit of a useful skill (https://www.ecmwf.int/en/forecasts). Compatible predictability was 8.9 days in the ensemble mean of the GEFS v10 in 2014 and improved to 10.5 days in the v11 in 2016, better than the 7.9 days of the deterministic forecasts by the GFS in 2014 and the 8.5 days in 2016 (Fig. 1 in Zhu et al. 2017b).

Persistent Z500 anomaly patterns are even more prediction-meaningful, because they tend to induce meteorological hazards such as heat waves, wildfires, droughts, flooding and snow storms (Quiroz 1984; Dole et al. 2011; Sillmann et al. 2011; Chen and Zhai 2014; Whan et al. 2016). Their predictions were evaluated and calibrated in medium-range weather forecasts by many studies (Tibalti and Monlteni 1990; Anderson 1996; Molteni et al. 1996; Krishnamurti et al. 2003; Hamill and Whitaker 2007; Ardilouze et al. 2017), with a focus on the predictability of atmospheric blocking that typically consists of a closed anticyclone and a cutoff low (Rex 1950; Dole and Gordon 1983; Lejenäs and Økland 1983; Metz 1986; Tibalti and Monlteni 1990; Kaas and Branstator 1993; Pelly and Hoskins 2003; Schwierz et al. 2004; Barriopedro et al. 2010; Barnes et al. 2012).

Among the blocking indices used for those evaluation and calibration studies, the most popular one was proposed by Tibaldi and Molteni (1990; TM90 hereafter). The TM90 approach identifies a blocking pattern as the reversal of 500-hPa geopotential height (Z500 hereafter) meridional gradients, which dynamically coincides with a split of westerly flow (Rex 1950) or easterly in place of dominant westerly in mid-high latitudes. This index has been used to estimate the predictability of blocking patterns since 1990s. In TM90, the blocking frequency was severely underestimated and the blocking onset was poorly predicted even a couple of days beforehand in the early model versions of ECMWF. After initial conditions included the blocking patterns, however, the prediction became more skillful. The blocking prediction by the ECMWF model in early 2000s was improved by as much as 50%, although the predicted blocking frequency was still 30% smaller than that in the analysis (Mauritsen and Källén 2004). Similarly smaller frequency (Watson and Colucci 2002) was predicted in all lead ranges by the operational system of the National Centers for Environmental Prediction (NCEP) at that time. The blocking prediction in a recent

NCEP model, the Global Ensemble Forecast System version 10 (GEFSv10), was substantially improved over the Euro-Atlantic sector (Hamill and Kiladis 2014) with occurrence frequencies nearly constant and only slightly smaller than those in the analysis even at lead times out to 16 days. The predicted blocking frequency, however, remains very small in the Pacific sector, partly because blockings occur relatively rarely there (e.g., Pelly and Hoskins 2003).

Some persistent high systems other than closed blockings are also important in inducing severe weathers (IPCC 2013; Grotjahn et al. 2015; Liu et al. 2017). These systems include omega-shape blockings especially in their early stages and persistent open ridges, which cannot be clearly identified by classical blocking indices such as the TM90. They appear as closed anticyclones in the zonal anomaly fields of Z500, but may not meet the criteria for blocking by their time anomaly field (e.g., Dole and Gordon 1983), and their predictability has not been investigated yet. To avoid confusion, the eddy refers to the departure from the zonal mean as in previous papers (e.g., Chen and Wallace 1993; L'Heureux et al. 2008), and the anomaly specifically is the departure from the time mean.

The maxima of Z500 eddies (PMZs) in the Northern Hemisphere coincide with the location of regional open ridges and closed anticyclones (Liu et al. 2017). Some of them persist for several days to weeks, move slowly and substantially impact surface weather conditions. Because a PMZ does not require a reversal of pressure gradients in the middle troposphere, it represents more persistent weather events than closed blockings by definition. For example, a PMZ event occurred over the northeast Pacific during January 2013 (Appendix Table.1) and persisted for 17 days (Figs. 1a-h), leading to a cold surge over the western United States (Figs. 1i-p). This persistent pattern started on 9 January from an open ridge of Z500 (color shading in Fig. 1c) near 160^{o}W over the northeast Pacific where a weak Z500 eddy center (contours) simultaneously resided. The ridge developed and intensified in the next five days, moving slowly eastward (Figs. 1d-h); it persisted for eleven more days in that area and its mass-weighted average locations are listed in Appendix Table 1. The impact areas of this PMZ event, derived by expanding the PMZ centers to include all adjacent grid points with Z500 eddy values no smaller than 100 geopotential meters (slashes in Fig. 1), clearly covered the surface cold anomalies over North America (Fig. 1p), because the northerly wind component was persistently advecting cold air mass from higher latitudes. In this case, the persistent atmospheric flow pattern cannot be identified as a blocking by a typical algorithm (e.g., TM90), because the Z500 is apparently characterized as strong open ridges before it develops into a mature phase (Figs. 1g and 1h). However, this PMZ influences the surface weather similarly as atmospheric blockings do and it needs to be considered in forecasts.

PMZs are connected with blockings but they are different. The statistics of PMZs in observational data were recently derived and compared (Liu et al. 2017) with those of blocking events (Hamill and Kiladis 2014). Two substantial differences are evident in the climatology over the Pacific-North America sector during wintertime. Firstly, the PMZs occur close to the U.S. West Coast (Fig. 9a in Liu et al. 2017), while the typical blockings occur farther westward and near the date line (Fig. 1 in Hamill and Kiladis 2014), suggesting more direct impact of PMZ events on the U.S. weather. Secondly, the occurrence frequency for the PMZs has a center of 33% along the West Coast of U.S. (Fig. 9a in Liu et al. 2017); it is twice large as the center of the blocking frequency (Fig. 2a in Hamill and Kiladis 2014) in mid Pacific. Therefore, it is worth examining the predictability of PMZs, potentially different from the predictability of traditional blockings, for the medium-range forecasts in the GEFS. Such an estimate would be helpful to improve week 2-4 predictions.

In this study, we estimate the prediction skills and predictability of PMZs in occurrence frequencies and of individual PMZ events by employing GEFS v10 forecasts and several objective verification metrics. Section 2 summarizes the tracking algorithm of PMZ in Liu et al. (2017), and introduces the data sets and methods. Section 3 presents the results for the predicted PMZ frequencies, Brier skill score and its reliability component, the probability of detection, the mean square error, and the skills in anomaly correlation coefficient. Section 4 summarizes and discusses the results.

2. Datasets and methods

2.1 Datasets

The NCEP GEFS v10 forecasts are investigated in this study, partly because this version was used to generate reforecasts and estimate the predictability of typical blockings by Hamill and Kiladis (2014). This operational system consisted of one control run and twenty perturbed members. Each member run was integrated four times daily starting at 0000, 0600, 1200, and 1800 UTC. After eight days of integration, the model changed its horizontal resolution of triangular truncation at wavenumbers 254 (T254; ~55 km) to T190 (~70km), while the physical parameterizations (Zhu et al. 2007) and the vertical resolution of 42 hybrid levels remained unchanged. The Global Data Assimilation System (GDAS) prepared analysis data for initializing the control run, and this initial condition was perturbed using the ensemble transform with rescaling (ETR) technique (Wei et al. 2008) to initialize other ensemble members. The uncertainty therein was estimated using the stochastic total tendency perturbation (STTP) method (Hou et al. 2008).

The GEFS v10 forecasts between 1 January 1985 and 14 February 2012 were regenerated off-line at the Earth System Research Laboratory (ESRL; Hamill et al. 2013); and the forecasts successive until present were made in real time. The off-line forecasts (or reforecasts) included a control run and only ten perturbed members due to limited computing resources. Each run started daily at 0000 UTC and ran out to 16 days. The forecasts until 31 December 2015 are combined for the present study after being bilinearly interpolated onto $2.5^{o}\times2.5^{o}$ grids from the native resolutions mentioned above. A detailed description of the model and reforecast datasets can be found in Hamill et al. (2013).

2.2 Tracking PMZ

An objective algorithm was developed by Liu et al. (2017) to track PMZ patterns, including persistent open ridges, immature omega-shape and mature blocking highs. This algorithm identifies and connects the local maxima of zonal eddies of Z500 after zonal means are removed. The algorithm tracks PMZs in the GEFS analysis (ANL) at each 00Z, slightly different from the daily mean data in Liu et al. (2017). It identifies a PMZ event as consecutive eddy maxima lasting for two days and longer, shorter than the four-day limit in Liu et al. (2017). As a result more PMZs are tracked in the GEFS forecasts for verification. The tracking steps in a consecutive order are summarized below.

a. A core at each 00Z UTC is identified to include a local maximum of zonal eddies at Z500 and its neighboring grids whose values are greater than 100 geopotential meters (GPMs) and decrease radially to 20 GPMs smaller than the maximum value.

b. Two cores on consecutive days belong to a PMZ event if they share at least one grid point and move at a pace of at most 10^{o} longitude per day.

c. The PMZ ends at the core without a successor.

d. Each of the tracked cores is expanded to include more contiguous points whose zonal eddy values are above 100 GPMs as the impact area. A none-tracked core is finally absorbed if it is surrounded by the expanded area. The larger number of expanded points better represent the actual area impacted by the PMZ.

The PMZ events in the initial conditions (ANL) at 00Z UTC from 1985 to 2015 serve as the reference for verification, because their statistics are very similar to those based on the daily data in Liu et al. (2017; not shown). The PMZs in each forecast are tracked differently for probabilistic and deterministic verifications using a time-lagged forecasting approach, as shown schematically in Fig. 2 where each black arrow starts from the referenced initial condition and extends to 17^{th} day for one forecast. For the probabilistic forecast verification, the forecast datasets are regrouped into 16 time series starting at the initial date on 1 to 16 January 1985, respectively. Each time series contains 11322 time slices for which PMZ events are tracked and their impact areas are objects to verify. For the deterministic forecast verification, however, the PMZs are tracked in each 17-day forecast time series covering at least 1 day of observations. This prefixed day guarantees that a tracked PMZ has an onset on or after the initial conditions, i.e., as early as on -1 day (open circle in Fig. 2), and extends to at least +1 day (blue dot in Fig. 2). As a result PMZs with onset on +1 to +15 days will be used to estimate the deterministic prediction skills and predictability.

2.3 Evaluation metrics

The prediction skills of PMZs are evaluated using five objective metrics

(Brankvoic 1990; Wilks 2006; Hamill and Juras 2006) among others: Brier Skill Score (BSS), reliability diagram, Probability of Detection (POD), Mean Square Error (MSE), and Anomaly Correlation Coefficient (ACC). Each metric is summarized below.

a. Brier Skill Score

The probability of a binary ensemble forecast $p_f(j)$ for the j^{th} sample is calculated as

$$p_f(j) = \frac{1}{n}\sum_{i=1}^{n} I_i(j), \qquad (1)$$

where $I_i(j)$ is 1 if an event occurs or 0 if not, and n is the number of forecasts in the j^{th} sample. The Brier Score (BS_f) of the forecasts is defined as

$$BS_f = \frac{1}{m}\sum_{j=1}^{m}\left[p_f(j) - I_o(j)\right]^2, \qquad (2)$$

where the subscript o denotes observations, and m is the number of samples. A Brier Skill Score (BSS) is finally computed as

$$BSS = 1 - \frac{BS_f}{BS_c}, \qquad (3)$$

where BS_c is the Brier Score of the reference probability forecast. The reference is generally the averaged climatic probability of an observed event p_c and defined as

$$p_c = \frac{1}{m}\sum_{j=1}^{m} I_o(j), \qquad (4),$$

and

$$BS_c = \frac{1}{m}\sum_{j=1}^{m}\left[p_c - I_o(j)\right]^2. \qquad (5).$$

When an ensemble member is used as the reference, the BSS becomes the skill of a perfect model.

b. Reliability

The Brier Score in Equation (2) can be decomposed into three components

$$BS = \frac{1}{n}\sum_{i=1}^{K} N_i \left(p_i - o_i\right)^2 - \frac{1}{n}\sum_{i=1}^{K} N_i \left(o_i - \bar{o}\right)^2 + \bar{o}\left(1 - \bar{o}\right), \qquad (6)$$

where $m = \sum_{i=1}^{K} N_i$; K denotes the frequency bins evenly from 0.0 to 1.0 (0% to 100%) for the forecast probability $p_f\left(j\right)$ (Eq. 1); N_i is the total number of samples in each bin; o_i is the observed frequency of events corresponding to p_i for each bin; and $\bar{o}$ equals p_c in equation (4). The three terms on the right-hand side of (6) are known as reliability, resolution, and uncertainty, respectively. The reliability diagram comprehensively assesses the forecast quality by representing a joint distribution of forecasts and observations.

c. Probability of Detection

The POD, or hit rate of forecasts, evaluates the probabilistic forecast of rare events, i.e., PMZs. It is expressed as

$$POD = \frac{H}{H + M}, \qquad (7)$$

where H (hits) denotes the number of samples predicted and observed, and M (misses) is for the number of samples observed but not predicted. The POD clearly ranges from 0 to 1. Since the occurrence frequencies of PMZs are predicted reasonably well in the first several days and decreasing with later lead times (to be shown below), the false alarm rate of PMZs is not discussed here.

d. Mean Square Error

The Mean Square Error (MSE) for an ensemble forecast of N members and the F_i for the i[th] member is expressed as

$$MSE=\frac{1}{N}\sum_{i=1}^{N}\left|F_i-X\right|^2=\left|\overline{F}-X\right|^2+\frac{1}{N}\sum_{i=1}^{N}\left|F_i-\overline{F}\right|^2, \qquad (8)$$

where $\overline{F}=\frac{1}{N}\sum_{i=1}^{N}F_i$ represents the ensemble mean, and X denotes the reference (or observational analysis) irrelevant to either N or i. The MSE can be decomposed into two terms as the square errors from the ensemble mean ($\left|\overline{F}-X\right|^2$) and the variance from the ensemble mean ($\frac{1}{N}\sum_{i=1}^{N}\left|F_i-\overline{F}\right|^2$).

e. Anomaly Correlation Coefficient

The ACC is a conventional measure of skills for a single or ensemble-mean forecast. It is defined as

$$ACC=\frac{F\cdot X}{|F|\cdot|X|}=\frac{|F|^2+|X|^2-|F-X|^2}{2|F|\cdot|X|}, \qquad (9)$$

where F is the forecast, and X is the reference.

3. Results

3.1 Statistical frequency verification

The climatological statistics of tracked PMZs in the NCEP analyses (ANL) is first presented. The frequency distributions of PMZs in different seasons are shown in Fig. 3. The PMZs are mainly located near subtropical jet areas in both the Northern and Southern Hemisphere (NH and SH; Fig. 3a-d). The frequency has a clear annual cycle, larger in winter (Fig. 3a) than in summer (Fig. 3c) in both NH and SH. In December-January-February (DJF) season, two maximum frequency centers are located in the northeast Pacific and northeast Atlantic coasts. The northeast Atlantic center expands from the Atlantic to Euro-Asian continent with a size larger than that in the northeast Pacific-North America sector, similar to the blocking frequency distribution documented in previous studies (TM90; Pelly and Hoskins 2003; Barriopedro et al. 2010; Hamill and Kiladis 2014). The locations of frequency centers, however, are different: the maximum blocking frequency is over central Pacific and close to 180^{o}E (Fig. 1 in TM90), while the maximum PMZ frequency is along the west coast of North America (Fig. 3a), corresponding to persistent northerly winds and potentially colder weather in the west U.S. (Fig. 1). In addition, the maximum frequency of the PMZs reaches as much as 40% near the West Coast of U.S., more than double to that of the blocking in the central Pacific, suggesting that the extreme events over the west US would be connected with PMZ patterns more than with traditional blocking events. In June-July-August (JJA; Fig. 3c) season, the PMZ center over North America shifts inland, different from that in other three seasons. This eastward shift may lead to persistent high pressure systems over the western U.S. for potential droughts and heat waves.

We next assess the skills of GEFS forecasts in predicting the PMZ occurrence frequencies in the Northern Hemisphere. The seasonality of PMZ frequencies (40-60^{o}N mean) in the forecasts is compared with that in the ANL by lead time (different colored curves in Fig. 4). Predicted PMZ frequencies in all seasons have distributions overall similar to those in the ANL and decreasing with lead time. The decreasing rates, however, are different in each season. In winter (Fig. 4a), two maxima in the ANL and reforecasts are located near 120^{o}E and 10^{o}W, respectively. The predicted frequencies decrease from 40% on +3 day to 30% on +15 day in both Pacific and Euro-Atlantic sectors, with a gap between +9 and +12 day decreasing sharply from 35% to 30%. This decrease is notably different from the predicted blocking as documented in Hamill and Kiladis (2014), in which the blocking frequencies do not change much with lead time except for a slight decrease in the Euro-Atlantic sector. One possible reason for such a distinction is that the persistent open ridges and immature omega-shape blockings in the PMZs have predictability shorter than traditional blocking patterns. In MAM (Fig. 4b) and SON (Fig. 4d), the predicted PMZ frequencies decrease from 25% on +3 day to less than 10% on +15 day. In JJA (Fig. 4c), the predicted PMZ frequencies are the smallest with a maximum about 18% over the Pacific and decreasing to less than 4% on +15 day when the rest of frequencies become nearly zero on other longitudes (blue cure in Fig. 4c).

The BSS is then computed to assess the probabilistic forecasts of PMZ frequencies in the Norther Hemisphere. The Northern Hemisphere was divided into Pacific and Euro-Atlantic sectors in TM90 and Hamill and Kiladis (2014). Since the occurrence frequency of PMZs is overall larger than that of blockings, the ranges of the two sectors are extend somewhat. The Pacific sector extends to include North America (PNAS) covering 180^{o}E-60^{o}W and the Euro-Atlantic becomes Euro-Atlantic-Asia sector (EAAS) in 60^{o}W-120^{o}E. The corresponding BSS in DJF for both sectors is shown in Fig. 5, with the dashed lines for the BSS of the perfect model -- the first ensemble member. The BSS for PNAS (red line) is overall higher than that for EAAS (blue line) on all lead times, suggesting the PMZ is more predictable in

PNAS than in EAAS. For both sectors, the BSS decreases rapidly from 0.9 to 0.5 in the first three days. This rapid decrease contrasts that of the perfect model where the BSS decreases more slowly and remains above 0.2 after +9 day (the dashed lines). Compared with the BSS of blocking frequency (Fig. 3 in Hamill and Kiladis 2014), the BSS of PMZs for both sectors is overall similar except for a faster decrease around day +5.

The reliability diagram is constructed for more evaluations. Forecast frequencies are divided into 10 bins from 0.0 to 1 and the observed occurrence in each bin on each lead day is derived according to Equation (6). The frequency distributions in the reliability diagram are shown in Fig. 6. On day +1 (Fig. 6a), the samples are concentrated on the 0.0 bin with 14 $\times 10^5$ and 24×10^5 grids for PNAS and EAAS, respectively. The numbers of samples in other bins are overall smaller than 2×10^5 grids and nearly equally distributed. From days +3 to +6 (Figs. 6b-c), the numbers clearly decrease in the 0.0 bin. From day +9 to +15 (Figs. 6d-f), the numbers decrease notably in 0.6~1.0 bins and increase in 0.1~0.5 bins. The sample sizes are overall larger in EAAS than in PNAS because of the larger range in EAAS. These histograms of PMZs are similar to those of blocking (Fig. 5 in Hamill and Kiladis 2014).

Reliability diagrams for the PMZ probabilistic forecasts by lead time are shown in Fig. 7. The forecasts for both PNAS and EAAS are quite reliable on day +1 (Fig. 7a) with the reliability (red and blue curves) nearly perfect (black line), a slight underestimate on 0.1~0.6 bins and a slight overestimate on 0.9~1.0 bins. From day +3 to +15 (Fig. 7b-f), the forecasts become less reliable and overestimated in EAAS than PNAS on 0.6~1.0 bins. For smaller-frequency bins such as 0.0 to 0.4, the forecast probability tends to be more underestimated, especially on forecast days +9~+15 (Fig. 7d-f) in both sectors, and the difference between PNAS and EAAS can be neglected.

Although the reliability decreases with lead time on larger frequency bins, they remain between the “no skill” (green) and “perfect skill” (black) lines. Changes of the forecast probability in these bins still contribute positively to the BSS according to Equation (6), while low reliabilities in 0.1~0.3 bins contribute negatively to the BSS. The results indicate that the decreasing BSS in Fig.5 is mainly contributed by the changes in 0.1 to 0.4 bins as the forecast time increases. In addition, the reliability for PMZs is very similar to those for blockings (Fig.5 of Hamill and Kiladis 2014), indicating similarly reliable prediction skills shared by the two weather systems.

3.2 *Forecast verification of ensemble mean*

This section presents the skills of GEFS in capturing individual PMZ events at different stages and the predictability of Z500 eddies conditioned by individual PMZ events.

The statistics of the PMZ events in observations are shown in Table 1. The PMZ events are counted separately by their central locations in PNAS and EAAS, and they are classified into three types in each sector based on their durations of 4~7 days, 8~14 days, and greater than 14 days. The total number of PMZ events is 1255 in PNAS and 1657 in EAAS. The number of PMZs with the lifetime of 4~7 days is 1002 in PNAS and 1244 in EAAS, representing 80% and 75% of total events in the corresponding sector. The events persistent for 8~14 days occur less frequently, with 226 (20% of the total) over PNAS and 375 (23% of the total) over EAAS. The number of PMZs persistent for longer than 14 days is 27 in PNAS and 38 in EAAS, about 2% of the total over both sectors. The PMZs with the longest lifetime persist for 24 days over PNAS and 32 days over EAAS. Their dates, locations and intensities are listed on the Appendix Table 1 for PNAS and 2 for EAAS. These PMZ events include some well-known blocking cases, such as the event in Fig. 1 and the Euro-Russian blocking in summer 2010 (Matsueda 2011).

The prediction skills of the three types of PMZ events in GEFS are evaluated separately, with a focus on the events persistent for 8~14 days and longer than 14 days over both PNAS and EAAS, since these events are potential sources of predictability for subseasonal predictions. Verification metrics of POD, MSE, and ACC are used. The POD for each lead time is derived according to Equation (7) by counting the grid points of PMZ impact areas on each day in the GEFS forecasts and observations. The MSE and ACC for different cases are reasonably compared by choosing a common region that encloses nearly all PMZ cases for verification. Figure 8a shows the impact areas of all PMZs over PNAS on the onset day for the 15-day cases (PNAS_15). Each contour denotes one PMZ event, and all the events are enclosed by the black dashed rectangle in (25-85^{o}N, 140-300^{o}E). Similarly, Figure 8b shows the PMZs over EAAS on the onset day for the 15-day cases (EAAS_15). The region of active PMZs over EAAS is larger than that over PNAS, with the black dashed rectangle enclosing (25-85^{o}N, 90^{o}W-140^{o}E). These two regions cover most of the PMZ events persistent for 8-14 days in PNAS and EAAS (not shown). Subsequently the MSE and ACC are calculated over (25-85^{o}N, 140-300^{o}E) for the cases in PNAS (PNAS_15) and over (25-85^{o}N, 90^{o}W-140^{o}E) for EAAS (EAAS_15; 4th and 5th rows on Table 1).

The verification of POD for the GEFS ensemble mean in different PMZ groups is shown in Fig. 9. For the 15-day cases in PNAS (Fig. 9a), the mean POD on the onset day (red line) decreases clearly as the lead time increases from +1 to +15 days. The POD is 0.8 on lead day +1, decreases rapidly to 0.5 on lead day +2, and reduces to less than 0.2 after lead day +6. The POD for +1 to +14 days after onset (green lines) shows an overall higher skill than for the onset day, which indicates that the prediction skill of PMZs is higher when the PMZs are already established in the initial conditions. The POD of +15 day (blue line) is about 0.75 on lead day +1, close to that of the onset day. However, the POD of +15 day increases to 0.85 on lead day +2, much higher than that of the onset day (red line). In the meantime the POD of +15 day shows a higher score than that of the onset day until lead day +11, which indicates that the forecast skill is higher in capturing the duration of PMZ than the onset at the same skill of +10 lead days.

For the cases over EAAS (Fig. 9b), the mean POD on the onset day (red line) is 0.65 for lead day +1, smaller than those over PNAS (Fig. 9b). The PODs are close to each other from lead day +1 to +3 in EAAS, and they are clearly higher than those in PNAS for lead days +2 and +3. Meanwhile, the POD of +15 day (blue line in Fig. 9b) is 0.85 on lead day +1, higher than that in PNAS (blue line in Fig. 9a). The PODs of +15 day (blue line) are higher than those of the onset day (red line) at various lead times until day +12 in EAAS, similar to those in PNAS. These results indicate that the GEFS has a better skill in capturing the PMZ onset with a lifetime longer than 15 days in EAAS than PNAS. For the 8~14-day cases (Fig. 9c, d), the mean PODs evolve smoother for both PNAS and EAAS than for the 15-day cases, partly because more PMZ events are sampled (cf. Table. 1). The prediction skill of the PMZ onset (red line) is notably lower than that of the PMZ duration (green and blue lines) in both sectors, while the POD of +7 day (blue line) is 0.1 higher than that of onset (red line) until day +10 in PNAS and day +8 in EAAS.

Forecast errors become larger with lead time in predicting individual PMZ events as indicated by the POD analysis above. These errors can be random in nature originating from the variability within the ensemble or from the model’s systematic bias. Such errors can be measured by the MSE (Brankvoic 1990) for the ensemble mean of predicted 500hPa geopotential height anomalies (Fig. 10). Figure 10a shows the MSE for the 15-day cases in PNAS. The red curve denotes the mean MSE for the onset day (0 day) with lead time; and the black, green, and blue curves denote MSEs for one day before onset, +1 to +13 days after onset, and +14 day, respectively. Clearly the MSE increases gradually with lead time at development stages of all PMZs, and it grows to 0.9 $\times 10^4$ GPM^2 per grid on day +16. After lead day +4, the MSEs grow rapidly. They become larger from +1 to +14 days (green and blue curves) at PMZ developing stages than those for -1 and 0 days (black and red curves) and after day +5. This result indicates that the forecast error in the GEFS reforecast increases when long-lived PMZ events develop into mature stages in PNAS. For the 15-day cases in EAAS (Fig. 10c), the MSEs differ slightly at PMZ developing stages: they are close to each other for -1 to +14 days (black, red, green, and blue curves). The mean MSEs in EAAS reach 0.8 $\times 10^4$ GPM^2 per grid on day +16, roughly equivalent to those in PNAS (Fig. 10a).

The MSEs for individual PMZ cases on the onset day are shown as gray curves in Figs. 10a and 10c. Compared with the mean (red), the MSEs among the cases exhibit large differences after day +4, with a range from 0.4 to 1.5 $\times 10^4$ GPM^2 per grid on day +16. To locate the error sources, the MSEs of the onset day (red) are decomposed into two terms: one for a mean-squared error [first term on the right hand side of equation (8); MSE_ens in black in Fig. 10b], and the other for an ensemble variance error [second term on the right hand side of equation (8), MSE_spread in blue in Fig. 10b]. The MSE_spread is overall smaller than the MSE_ens from days +3 to +16 in both sectors (Fig. 10b, d).

The MSEs are further investigated for the 8-14-day cases in both sectors (Fig.

11). In PNAS (Fig. 11a), the mean MSEs are close to each other at different PMZ development stages (black, red, green, and blue curves). This indicates that the growth of forecast errors does not depend on PMZ stages, similar to that in EAAS (Fig. 11c). For the random error and model bias in the 8-14-day cases, the MES_ens is also larger than the MES_spread in both PNAS (Fig. 11b) and EAAS (Fig. 11d). This suggests that the model bias is a dominant source of the forecast error in GEFS and it is not dependent on PMZ stages either.

The ACC is a classic metric to quantify the deterministic forecast skills of 500hPa geopotential height. It generally ranges from 0 (worst) to 1 (best). An ACC greater than 0.6 in general indicates a useful forecast with troughs and ridges at Z500 properly placed (Krishnamurti et al. 2003). The ACCs of Z500 eddy fields for the 15-day PMZ cases in PNAS with forecast lead time are shown in Fig. 12a. The mean ACC of the GEFS ensemble mean for the PMZ onset day is represented by the red curve. It is close to 1.0 from days 0 to +2, and decreases notably from days +3 to +16. This decrease is inherently associated with the increase of MSE (cf. Fig. 10). The ACC predictability of PMZ onset is 8.5 and 10.5 days when 0.6 and 0.5 are used as the threshold of useful skills, respectively. The ACC for the predictions starting on -1 day (black curve) shows similar evolution as that on the onset day, and the predictability is close to 9 days with the ACC of 0.6 as a useful skill. However, the predictability at PMZ development stages (green and blue curves) is notably extended. The ACC skill is extended to 10 days with 0.6 as the lower limit for the forecasts initialized on +14 day after PMZ onset (blue curve). It is noteworthy, however, that capturing individual PMZ events is case dependent, especially after lead day +5. Similar to MSEs, the uncertainty of ACCs for the onset day (gray curves) increases with lead time, and the ACCs for individual cases vary substantially from

0.8 to less than 0.2 on day +16.

The ACC skill associated with the PMZ events is next compared with the averaged ACC skill of Z500 eddy fields in NH to investigate possible improvement due to the persistence. The ACC for NH is based on the same samples over PNAS in Fig. 12a. Results are shown in Fig. 12b: the solid curves are exactly the same as those in Fig. 12a, representing the ACCs in the predefined PNAS region for the day before the PMZ onset (-1 day, black), on the PMZ onset day (0 day, red), and +14 days after the PMZ onset (+14 day, blue). The dashed curves use the same samples but the calculated region extends to the Northern Hemisphere. In addition the averaged ACC is computed during 1 January 1985 to 31 December 2015 as the green dashed curve. The ACCs of the Northern Hemisphere are overall better than those of the PNAS by a half day. In the meantime the ACC skill for the PMZ onset is lower than the average (ACC > 0.6) with lead days of +8.5 for the PMZ onset (red) and about a half day shorter than the total (green). In contrast, the ACC skill for +14 days after PMZ onset (blue) is nearly 1.5 days better than that for the average, which indicates that the PMZ persistence effectively enhances the predictability of subseasonal signals in GEFS. For the 15-day cases in EAAS (Fig. 12c), the ACCs at different PMZ development stages are overall similar to those in PNAS with several small differences. For example, the ACCs of the onset day for individual cases (gray) have a limit of 0.6 on lead day +16 in EAAS, lower than those in PNAS. The ACC of the onset day in EAAS has a useful skill on lead day +9 (ACC > 0.6), about a half day better than in PNAS. Moreover, the ACC of +14 day after the PMZ onset (blue) is close to the onset day (red) with the ACC greater than 0.6, and it extends to day +15 with the ACC greater than 0.5 in EAAS, about 3 days better than in PNAS at the same threshold. For the 15-day cases in EAAS (Fig. 12d), the skills of -1, 0, +14 days (black, red, and blue dashed curves) are all higher than the averaged skills of the NH (green dashed), and they are over a half day better than those in each sector.

The ACC skills for the 8-14-day PMZ cases (Fig. 13) are discussed next in more detail. The mean ACCs at different PMZ development stages are close to each other in both sectors. The ACCs of -1, 0, and +7 days are overlaid on lead day +9.5 in PNAS (Fig. 13a) at the threshold of 0.6. In EAAS (Fig. 13c), the ACCs of -1 and 0 days are overlaid on lead day +9 for the threshold of 0.6 and the ACC of +7 day reaches that of lead day +9.5 at the same threshold. We also compare the ACCs for 8-14-day cases with the averaged ACC for all the days from 1985 to 2015. For the cases in PNAS (Fig. 13b), the ACCs of NH (dashed lines) are very close to the regional values (solid curves) for all PMZ development stages (-1, 0, and +7 days). The ACC skills can reach lead day +9.5 at the threshold of 0.6, slightly better than the ACC skill of the total days (green dashed). For the cases in EAAS (Fig. 13d), the ACC skills of -1 and 0 days (black and red dashed line) for the NH are overall close to the regional values (black and red solid), and the ACC skill of +7 day for the Northern Hemisphere extends to lead day +10 at the threshold of 0.6. These results indicate that the GEFS can achieve a better ACC score in PNAS than in EAAS for the PMZ events persistent longer than one week. The ACC skill in predicting the PMZ onset is still lower than predicting the PMZ development, similar to that in predicting blockings (TM90). However, the prediction skill of PMZs in GEFS is extended to 9~10 days, nearly 3 days longer than the blocking in earlier ECMWF model (Fig. 15 in TM90).

From the above verifications of POD, MSE, and ACC, the largest different skill in GEFS occurs in capturing different PMZ events instead of different stages of individual PMZ cases. We next quantify the uncertainty of GEFS in predicting PMZ onsets by sorting their ACC scores. We show the boxplots of ACC scores on the onset day for all the four PMZ groups in Fig. 14 where the black dots denote the mean values for all cases. The upper and lower boundaries denote upper and lower quartiles, and the horizontal line within each box denotes the median value. The horizontal line above and on the top of a box denotes the maximum value, and the horizontal below and on the bottom of a box denotes the minimum value. For the PMZ events longer than 15 days in PNAS (Fig. 14a), the prediction on lead day +1 is the best. All predicted cases show consistently high ACC values, and the box becomes a horizontal line. The uncertainty increases with lead time and the box expands in the vertical direction. The ACCs are lower than 0.5 for less than 25% cases on lead day +6 and above 0.6 until lead day +9 for more than half of the cases, indicating the prediction skill is about 9 days in PNAS for the PMZs persistent longer than two weeks. From lead days +12 to +16, the ACCs are above 0.6 for about 25% cases and lower than -0.3 for other 25% cases with reversed forecast patterns. For the 15-day cases in EAAS (Fig. 14b), the ACC skill is extended to 11 days for more than half of the cases with the skill above 0.6. Similarly the useful ACC skill reaches 10 days for the 8-14-day cases in both PNAS (Fig. 14c) and EAAS (Fig. 14d).

It is noteworthy that there are still about 25% useful cases (ACCs above 0.6) on lead days longer than 10 in all of the four groups. These cases can potentially improve the model's prediction skills at weeks 2 to 4. Among the 15-day cases in both sectors, we select the PMZ events with an ACC greater than 0.6 for the forecasts of the onset and +15 day on lead days +10 to +16 (Appendix Tables 3 and 4). The ACCs for these cases represent the prediction skills of the PMZ onset and duration. For the onset day (Appendix Table 3), four cases are selected with three in PNAS (26 February 2011, 10 January 2013, and 15 May 2015) and one in EAAS (24 November 2012). The first two cases are strong with intensity above 300 GPMs. For the PMZ duration

(Appendix Table 4), four cases are also selected with three in PNAS (18 December 1993, 13 December 1999, and 19 February 2005) and one in EAAS (24 February 1986). The eight cases are different in time, suggesting that the ACC skills of the PMZ onset and duration are different. The duration prediction of long-lived PMZs is overall poor even when the onset is predicted well; and relatively better predicted cases all occur in winter.

4. Summaries and Discussions

The daily forecast data sets of GEFS v10 from 1 January 1985 to 31 December 2015 are used in this study to evaluate the skills in predicting persistent maximum patterns at Z500. The persistent maxima of Z500 eddies (PMZ; Liu et al. 2017) in PNAS and EAAS sectors are tracked and their impact areas serve as the targets for verification, because the PMZs include persistent open ridges, omega-shape blockings, and closed blocking anticyclones. Skills in both probabilistic and deterministic predictions are evaluated and the main results are summarized below.

(1) The PMZ frequency is underestimated at longer lead times in both PNAS and EAAS sectors. The predicted frequency on lead day +16 is underestimated by 30% in winter and 90% in summer. The BSS skill is higher in PNAS than in EAAS and the skill of the perfect model is higher even on lead days longer than 10. The reliability skill drops considerably at larger frequency bins (0.6~0.8) in both sectors.

(2) The POD decreases with lead time in both sectors, consistent with the increasing MSE and decreasing ACC. The MSE becomes large after lead day +4 and is contributed comparably by the ensemble mean and spread. The ACC skill indicates that the predictability of the PMZ onset and duration is different, especially for those with a lifetime longer than 15 days. The onset has a useful skill up to 8.5 days for the

PNAS and 9 days for the EAAS, while the duration has a useful skill up to 10 days for the PNAS and 9.5 days for the EAAS.

(3) The uncertainty is different for predicting the PMZ onset and duration. The PMZ onset has useful ACC skills (> 0.6) about a half day in lead time shorter than the average of all forecasts; and the PMZ duration has useful ACC skills about 1 day longer in lead time. The uncertainty among different PMZ cases is even larger. Half of the PMZ cases are predictable with an ACC skill greater than 0.6 on lead day +9 to +10 in both sectors; and about 25% cases are still predictable on lead day +10 to +16.

(4) Compared with classical blocking patterns (Hamill and Kiladis 2014), the PMZ events occur more frequently in the Northeast Pacific close to North America, and their frequencies are predicted with considerable skills by the GEFS v10. The predicted frequencies of PMZs decrease notably as the lead time increases, which differs from the unchanged frequencies of predicted blockings (Fig. 2 in Hamill and Kiladis 2014). The reliability of PMZ forecast demonstrates better skills in PNAS than EAAS in high frequency bins (0.6~0.8), opposite to those of predicted blockings (Fig. 5 in Hamill and Kiladis, 2014). In addition, the skill is overall better in predicting the duration than the onset of PMZs, which is similar to predicting traditional blockings (TM90; Hamill and Kiladis 2014)

Improving the forecast skills of GEFS at weeks 3-4 is a challenging task and the PMZs are sources for potential improvements. Evaluations of the skills in predicting the PMZ events disclose some intriguing topics for future investigation. For instance, the prediction skill is more sensitive to PMZ cases with uncertainty of over 10 days than to PMZ stages with uncertainty of about 2 days. The uncertainty can be further traced to various sources, such as the limited spatial and temporal resolutions, initial conditions, and model’s physical parameterizations. Understanding these sources will be helpful in developing and evaluating the NGGPS as well.

Acknowledgments

B. He, W. Hu and P. Liu were supported by the National Weather Service under the grant NA15NWS4680015 and NA15OAR4320064. B. He and W. T. Hu were partially supported by the NSFC under the grant 41405091.

References

Anderson JL (1996) A method for producing and evaluating probabilistic forecasts from ensemble model integrations. J Clim 9:1518-1530

Ardilouze C, Batté L, Déqué M (2017) Subseasonal-to-seasonal (S2S) forecasts with CNRM-CM: a case study on the July 2015 West-European heat wave. Adv Sci Res 14:115-121

Barnes EA, Slingo J, Woollings T (2012) A methodology for the comparison of blocking climatologies across indices, models, and climate scenarios. Clim Dyn 38:2467-2481 doi:10.1007/200382-011-124306

Barriopedro D, Garcia-Herrera R, Gonzalez-Rouco JF, and Trigo RM (2010) Application of blocking diagnosis methods to general circulation models. Part I: a novel detection scheme. Clim Dyn 35:1373-1391

Brankovic C, Palmer TN, Molteni F, Tibaldi S, Cubasch U (1990) Extended-range predictions with ECMWF models: Time lagged ensemble forecasting. Quart J Roy Meteor Soc 116:867-912

Cheng X, Wallace JM (1993) Cluster analysis of the Northern Hemisphere wintertime 500-hPa height field: Spatial patterns. J Atmos Sci 50:2674-2696

Chen Y, Zhai P (2014) Precursor circulation features for persistent extreme precipitation in central-eastern China. Wea Forecasting 29:226-240 doi:10.1175/WAF-D-13-00065.1

Dole RM, Gordon ND (1983) Persistent anomalies of the extratropical Northern Hemisphere wintertime circulation: Geographical distribution and regional persistence characteristics. Mon Wea Rev 111:1567-1586

Dole RM, Hoerling M, Perlwitz J, Eischeid J, Pegion P, Zhang T, Quan XW, Xu TY, Murray D (2011) Was there a basis for anticipating the 2010 Russian heat wave?

Geophys Res Lett 38:L06702

Grotjahn R, Coauthors (2016) North American extreme temperature events and related large-scale meteorological patterns: A review of statistical methods, dynamics, modeling, and trends. Clim Dyn 46:1151-1184 doi:10.1007/s00382-015-2638-6

Hamil TM, Bates GT, Whitaker JS, Murray DR, Fiorino M, Galarneau Jr TJ, Zhu YJ, Lapenta W (2013) NOAA’s second-generation global medium-range ensemble reforecast data set. Bull Amer Meteor Soc 94:1553-1565 doi:10.1175/ BAMS-D-12-00014.1

Hamill TM, Kiladis GN (2014) Skill of the MJO and Northern Hemisphere blocking in GEFS medium-range reforecasts. Mon Wea Rev 142:868-885

Hamill TM, Juras J (2006) Measuring forecast skill: is it real skill or is it the varying climatology? Quart J Roy Meteor Soc 132:2905-2923

Hamill TM, Whitaker JS (2007) Ensemble Calibration of 500-hPa Geopotential Height and 850-hPa and 2-m Temperatures Using Reforecasts. Mon Wea Rev 135:3273-3280.

Hou D, Toth Z, Zhu Y, Yang W (2008) Evaluation of the impact of the stochastic perturbation schemes on global ensemble forecast. Proc. 19th Conf. on Probability and Statistics, New Orleans, LA, Amer Meteor Soc [Available online at https://ams.confex.com/ams/88Annual/webprogram/Paper134165.html.]

IPCC Climate Change (2013) The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Stocker TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM, Ed., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 3-32.

Kaas E, Branstator G (1993) The relationship between a zonal index and blocking activity. J Atmos Sci 50:3061-3077

Krishnamurti TN, Rajendran K, Vijaya Kumar TSV, Lord S, Toth Z, Zou XL, Cocke S, Ahlquist JE (2003) Improved skill for the anomaly correlation of geopotential heights at 500 hPa. Mon Wea Rev 131:1082-1102

Lejenäs H, Økland H (1983) Characteristics of Northern Hemisphere blocking as determined from a long time series of observational data. Tellus 35A:350-362

L'Heureux ML, Kumar A, Bell GD, Halpert MS, Higgins RW (2008) Role of the Pacific-North American (PNA) pattern in the 2007 Arctic sea ice decline, Geophys Res Lett 35:L20701 doi:10.1029/2008GL035205

Liu P, Coauthors (2017) Climatology of tracked persistent maxima of 500-hPa geopotential height. Clim Dyn doi: 10.1007/s00382-017-3950-0

Lorenz EN (1963) Deterministic nonperiodic flow. J Atmos Sci 20:130-141

Lorenz EN (1982) Atmospheric predictability experiments with a large numerical model. Tellus 34:505-513

Matsueda M (2011) Predictability of Euro-Russian blocking in summer of 2010. Geophys Res Lett 38:L06801 doi:10.1029/2010GL046557

Mauritsen T, Källén E (2004) Blocking prediction in an ensemble forecasting system. Tellus 56A:218-228

Metz W (1986) Transient cyclone-scale vorticity forcing of blocking highs. J Atmos Sci 43:1467-1483

Molteni F, Buizza R, Palmer TN, Petroliagis T (1996) The ECMWF ensemble prediction system: Methodology and validation. Quart J Roy Meteor Soc 122:73-119

Pelly JL, Hoskins BJ (2003) A new perspective on blocking. J Atmos Sci 60:743-755

Quiroz RS (1984) The climate of the 1983-1984 winter–A season of strong blocking and severe cold in North America. Mon Wea Rev 112:1894–1912

Rex DF (1950) Blocking action in the middle troposphere and its effect upon regional climate. I. An aerological study of blocking action. Tellus 2:196-211

Saha S, Coauthors (2010) The NCEP Climate Forecast System Reanalysis. Bull Amer Meteor Soc 91:1015-1057 doi:10.1175/2010BAMS3001.1

Schwierz C, Croci-Maspoli M, Davies HC (2004) Perspicacious indicators of atmospheric blocking. Geophys Res Lett 31:6125-6128

Sillmann J, Croci-Maspoli M, Kallache M, Katz RW (2011) Extreme cold winter temperatures in Europe under the influence of North Atlantic atmospheric blocking. J Clim 24:5899-5913

Tibaldi S, Molteni F (1990) On the operational predictability of blocking. Tellus 42:343-365

Watson JS, Colucci SJ (2002) Evaluation of ensemble predictions of blocking in the NCEP global spectral model. Mon Wea Rev 130:3008-3021

Wei M, Toth Z, Wobus R, Zhu Y (2008) Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system. Tellus 60A:62-79 doi:10.1111/j.1600-0870.2007.00273.x

Whan K, Zwiers FW, Sillmann J (2016) The influence of atmospheric blocking on extreme winter minimum temperatures in North America. J Clim 29:4361-4381

Wilks DS (2006) Statistical Methods in the Atmospheric Sciences. 2nd ed. Academic Press, 627 pp.

Zhou X, Zhu Y, Hou D, Luo Y, Peng J, Wobus D (2016) The NCEP Global Ensemble Forecast System with the EnKF Initialization. Wea Forecasting: in press.

Zhu Y (2005) Ensemble forecast: A new approach to uncertainty and predictability. Adv Atmos Sci 22:781-788

Zhu Y, Toth Z (2008) Ensemble based probabilistic verification. Preprints, 19th Conf. on Predictability and Statistics, New Orleans, LA, Amer Meteor Soc 2.2. [Available online at https://ams.confex.com/ams/pdfpapers/131645.pdf.]

Zhu Y, Zhou X, Pena M, Li W, Melhauser C, Hou D (2017a) Impact of Sea Surface Temperature Forcing on Weeks 3 & 4 Forecast Skill in the NCEP Global Ensemble Forecasting System. Submitted to Wea Forecasting (conditionally accepted)

Zhu Y., Zhou X, Li W, Hou D, Melhauser C, Sinsky E, Pena M, Fu B, Guan H, Kolczynski W, Wobus R, Tallapragadaand V (2017b) An Assessment of Subseasonal Forecast Skill Using an Extended Global Ensemble Forecast System (GEFS). Submitted to J Clim (under review)

Table. 1 Statistics of PMZ events in ANL for the Pacific (180^{o}E-60^{o}W; PNAS) and Euro-Atlantic-Asia (60^{o}W-120^{o}E; EAAS) sectors. The PMZs are grouped by durations into 4~7, 8~14, and above 15 days. The maximum durations are shown in the “Max (days)” row.

PMZ events	PNAS	EAAS
Total	1255	1657
4~7	1002	1244
8~14	226	375
15~	27	38
Max (days)	24	32

Figure captions

Fig. 1 (a-h) 500-hPa geopotential height (Z500, gpm, shaded), zonal eddies above 100 gpm (black contours) and impact areas of a PMZ (slash) in ANL at 00Z UTC from 7-14 Jan 2013. (i-p) Corresponding anomalies of daily air temperature at 2 meter (^{o}C) over CONUS in the NCEP-NCAR reanalysis.

Fig. 2 Schematic of time-lagged forecasting and PMZ tracking in GEFS reforecasts. The bottom thick black arrow denotes time in days with “0” for the PMZ onset and black thin arrows for forecasting days (16 for each run) and PMZ tracking directions. The solid red lines denote the initial time for each ensemble member. For a single run, the onset day of a PMZ is represented by the red solid circle, the day before onset by the red dot, and the day after by the blue dot.

Fig. 3 Frequency distributions (%) of the PMZ impact areas in GEFS-ANL in (a) DJF, (b) MAM, (c) JJA, and (d) SON seasons during 1979-2015.

Fig. 4 Frequency distributions (%) in longitude of the PMZ impact areas averaged in 40-60^{o}N for the ensemble mean of GEFS reforecast during (a) DJF, (b) MAM, (c) JJA, and (d) SON seasons.

Fig. 5 Brier skill scores of PMZ probability forecasts for the PNAS (dotted red) and EAAS (dotted blue) sectors during DJF. Dashed lines are from the “perfect model”.

Fig. 6 Histograms for the numbers ($\times 10^5$ grids) of samples in the PNAS (red) and EAAS (blue) sectors.

Fig. 7 Reliability diagrams of PMZ probability forecasts for (a) +1 day, (b) +3 day, (c) +6 day, (d) +9 day, (e) +12 day, and (e) +15 day in the Pacific (30-70^{o}N, 180-280^{o}E; dotted red) and Euro-Atlantic (30-70^{o}N, 60 ^{o}W-120^{o}E; dotted blue) sectors. Black-solid, green-solid, and black-dashed lines denote the perfect skill, no skill, and climatology probability, respectively.

Fig. 8 Snapshots of PMZ impact areas during the onset day in PNAS (a) and EAAS (b) for the cases PNAS_15 and EAAS_15 in Table.1. The black rectangles correspond to the regions of 25-85^{o}N, 140-300^{o}E (a) and 25-85^{o}N, 90^{o}W-140^{o}E (b).

Fig. 9 (a) Probability of detection (POD) for the impact areas of 15-day PMZ cases in the GEFS ensemble mean. The abscissa denotes the lead time (days). The red, blue, and green lines denote the POD for the onset day (+0 day), +15 day after the onset, from +1 to +14 days, respectively. (b) The same as in (a) but for the POD of the 15-day cases in EAAS. (c) and (d) are the same as (a) and (b) but for the 7-14 day cases in PNAS and EAAS, respectively.

Fig. 10 (a) Averaged MSE for the GEFS ensemble mean of 500-hPa geopotential height anomaly (×10^4 gpm^2 per grid) for the 15-day cases in PNAS. The red, black, blue and green curves denote the MSE for the onset day (0 day), one day before the onset (-1 day) and +14 day after the onset, and +1 to +13 days, respectively. (b) Each term of MSE for the onset day: the red curve is the same as in (a); the black and blue curves denote the MSE from the ensemble mean and spread, respectively. (c) and (d) are the same as (a) and (b) but for the 15-days cases in EAAS.

Fig. 11 The same as Fig.10 but for the 8-14-day cases in PNAS (a and b) and EAAS (c and d), respectively. The blue line denotes the MSE for +7 days after the onset.

Fig. 12 (a) Averaged ACC for the GEFS ensemble mean 500-hPa geopotential height anomaly of the 15-day cases in Appendix Table 1. The abscissa denotes lead times (days). The black dashed lines denote the 0.6 and 0.5, respectively. The red, black, blue, green and gray curves denote the ACC for the onset day (+0 day), one day before the onset (-1 day), +14 day after the onset, +1 to +13 days, and the onset day for the 26 individual cases, respectively. (b) The solid lines are the same as in (a) and the dashed lines denote the ACC for the Northern Hemisphere. The green dashed line denotes averaged ACCs from 1 January 1985 to 31 December 2015. (c) and (d) are the same as (a) and (b) but for the 15-day cases in EAAS.

Fig. 13 The same as Fig.12 but for the 8-14 days cases in Appendix Table 1 for PNAS (a, b) and EAAS (c, d). The blue line denotes the ACC for +7 day after the onset.

Fig. 14 (a) Box plots of ACCs for the onset day of the 15-day cases in PNAS (Appendix Table.1) with lead days in the GEFS ensemble mean. The black dashed lines denote 0.6 and 0.5, respectively. The black dots denote the mean value of the cases. (b) The same as (a) but for the 15-day cases in EAAS. (c) The same as (a) but for the 8-14-day cases in PNAS. (d) The same as (a) but for the 8-14-day cases in EAAS.

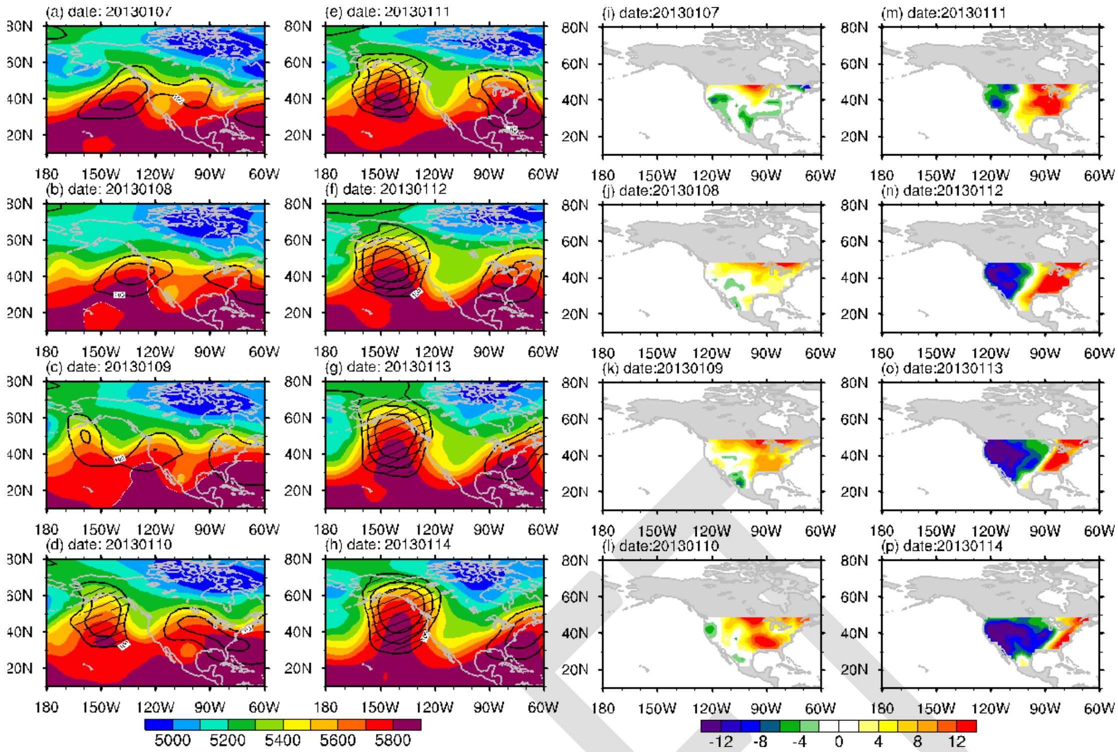


Fig. 1 (a-h) 500-hPa geopotential height (Z500, gpm, shaded), zonal eddies above 100 gpm (black contours) and impact areas of a PMZ (slash) in ANL at 00Z UTC from 7-14 Jan 2013. (i-p) Corresponding anomalies of daily air temperature at 2 meter (^{o}C) over CONUS in the NCEP-NCAR reanalysis.

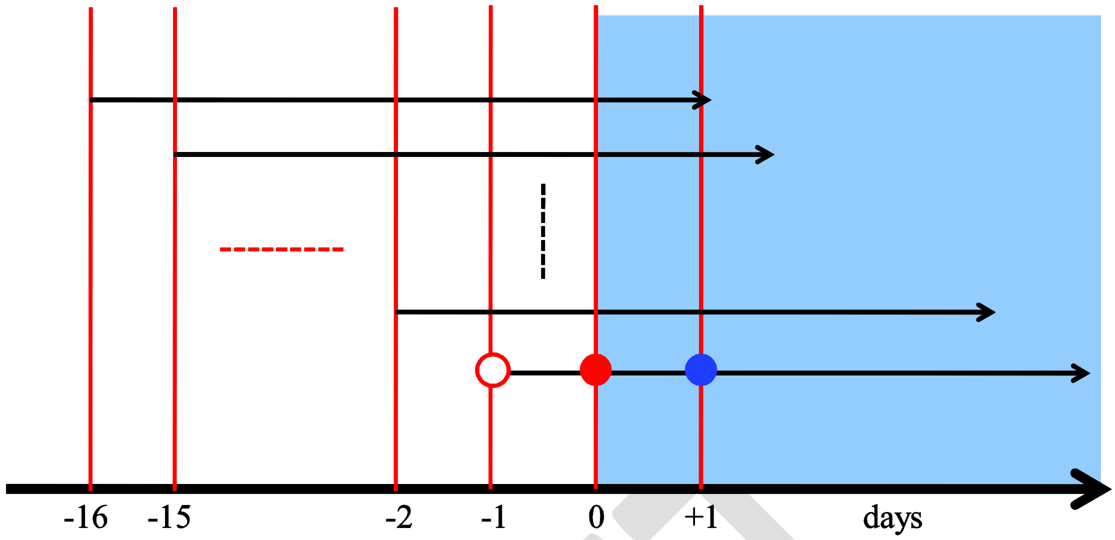


Fig. 2 Schematic of time-lagged forecasting and PMZ tracking in GEFS reforecasts. The bottom thick black arrow denotes time in days with "0" for the PMZ onset and black thin arrows for forecasting days (16 for each run) and PMZ tracking directions. The solid red lines denote the initial time for each ensemble member. For a single run, the onset day of a PMZ is represented by the red solid circle, the day before onset by the red dot, and the day after by the blue dot.

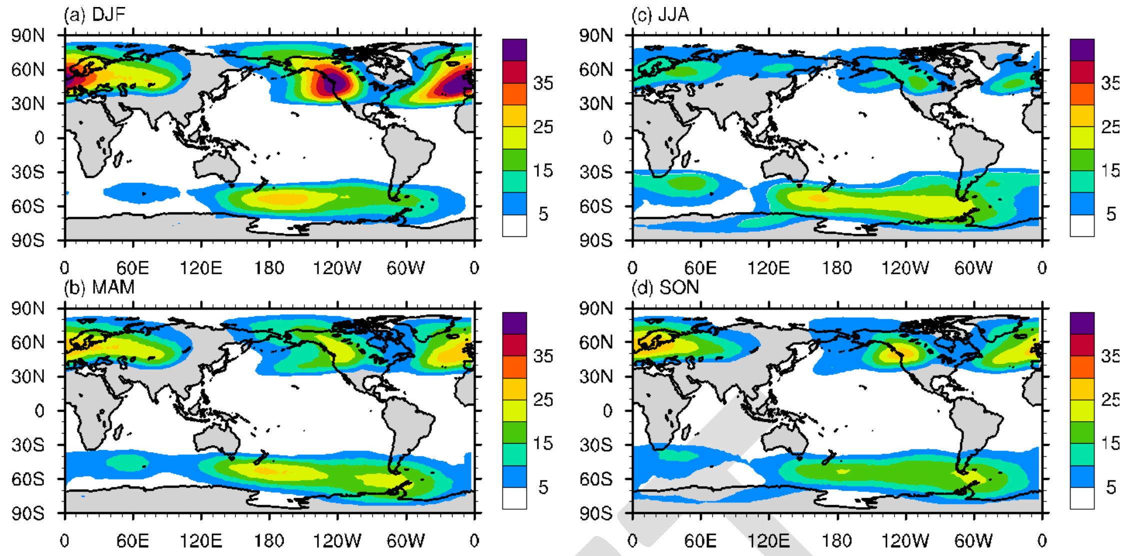


Fig. 3 Frequency distributions (%) of the PMZ impact areas in GEFS-ANL in (a) DJF, (b) MAM, (c) JJA, and (d) SON seasons during 1979-2015.

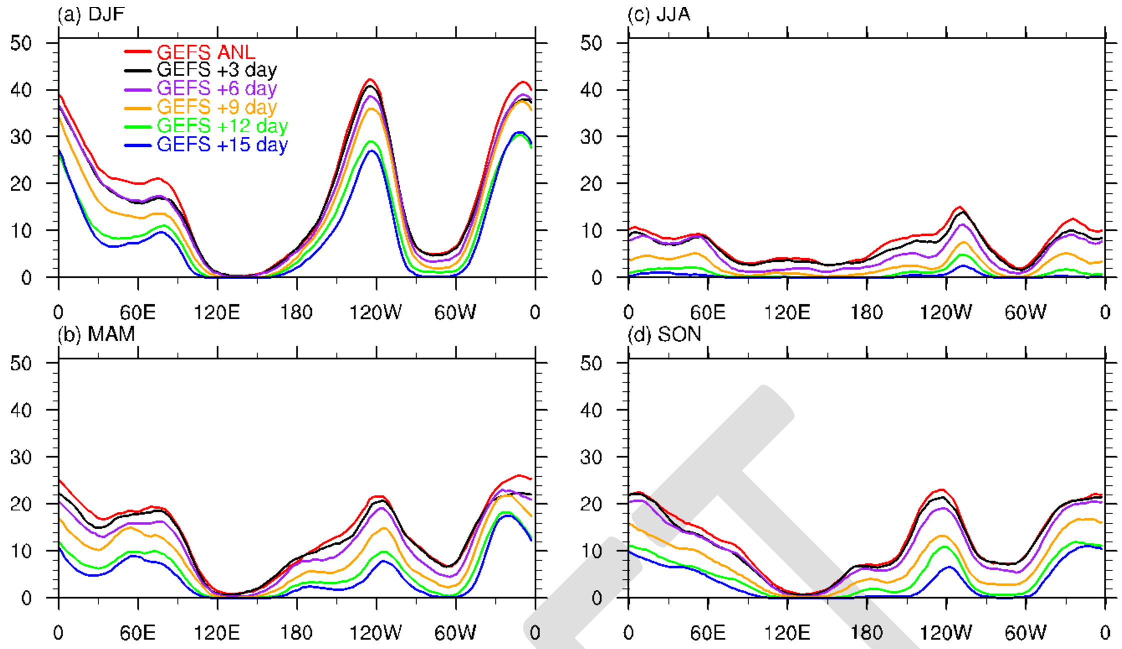


Fig. 4 Frequency distributions (%) in longitude of the PMZ impact areas averaged in 40-60^{o}N for the ensemble mean of GEFS reforecast during (a) DJF, (b) MAM, (c) JJA, and (d) SON seasons.

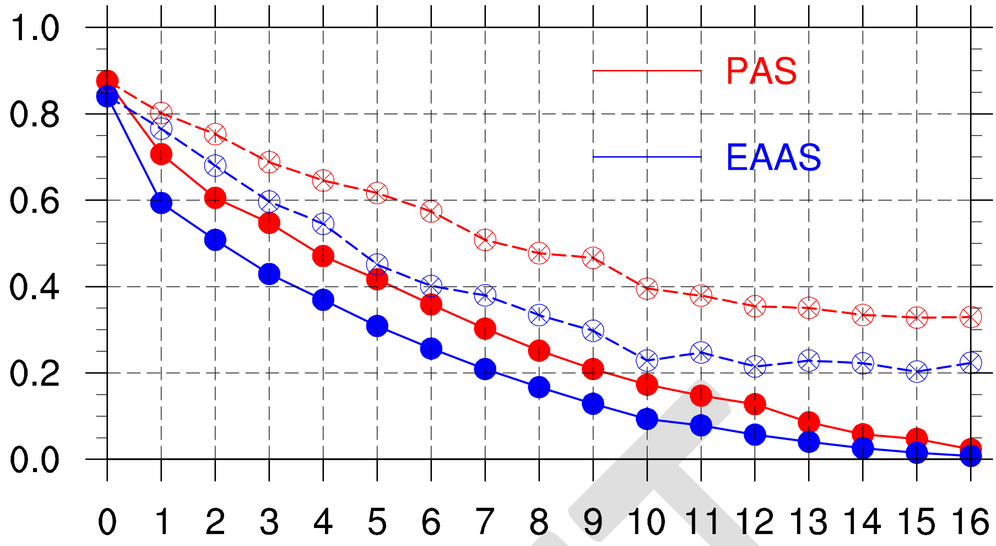


Fig. 5 Brier skill scores of PMZ probability forecasts for the PNAS (dotted red) and EAAS (dotted blue) sectors during DJF. Dashed lines are from the "perfect model".

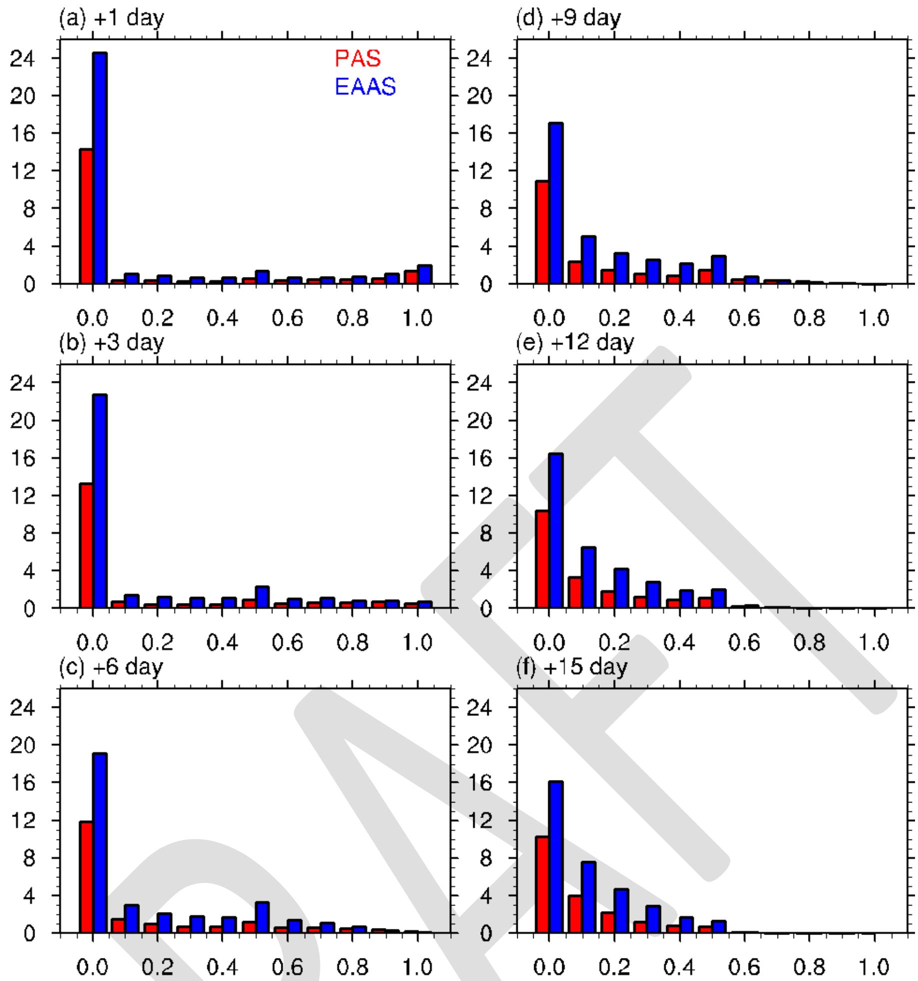


Fig. 6 Histograms for the numbers ($\times 10^5$ grids) of samples in the PNAS (red) and EAAS (blue) sectors.

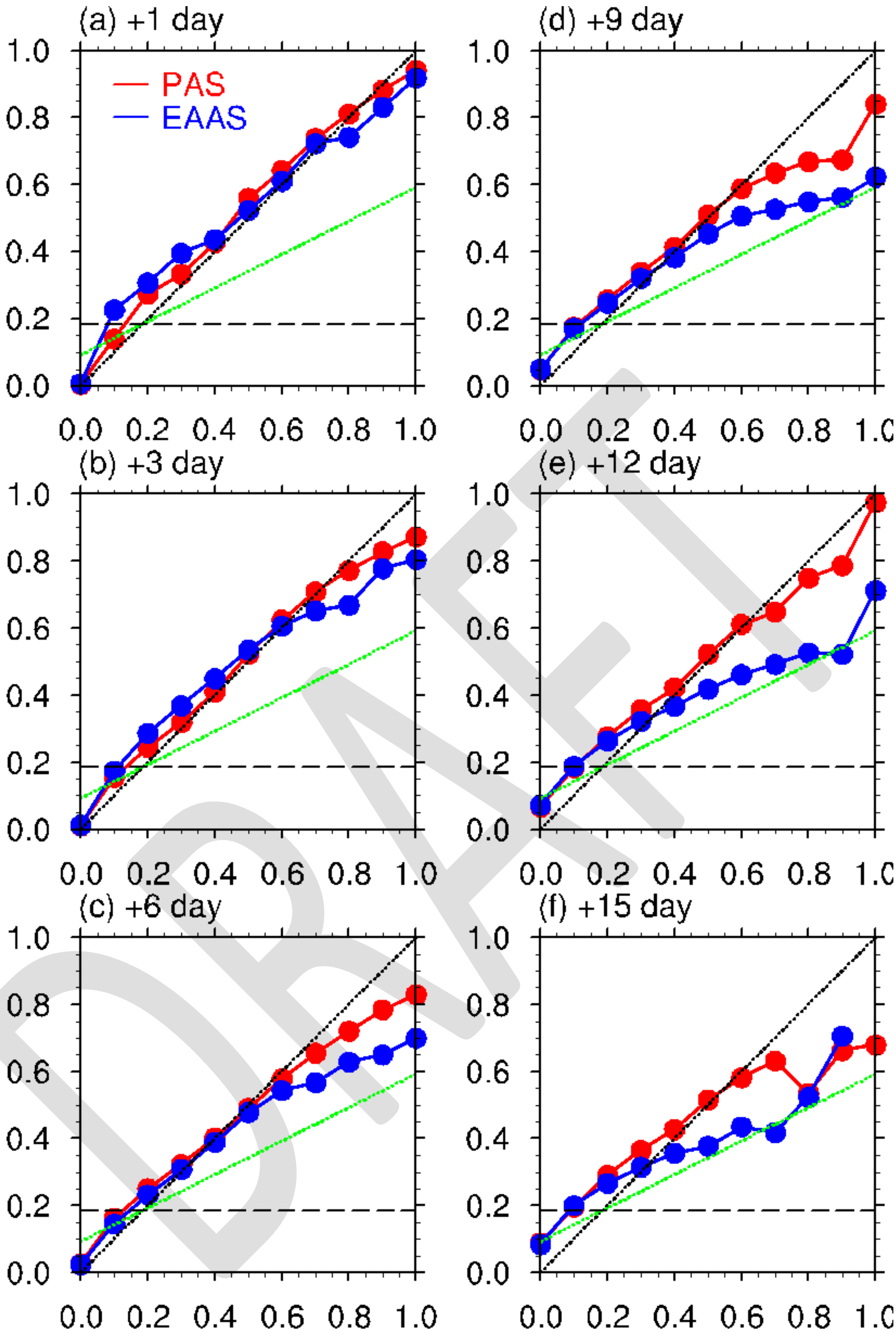
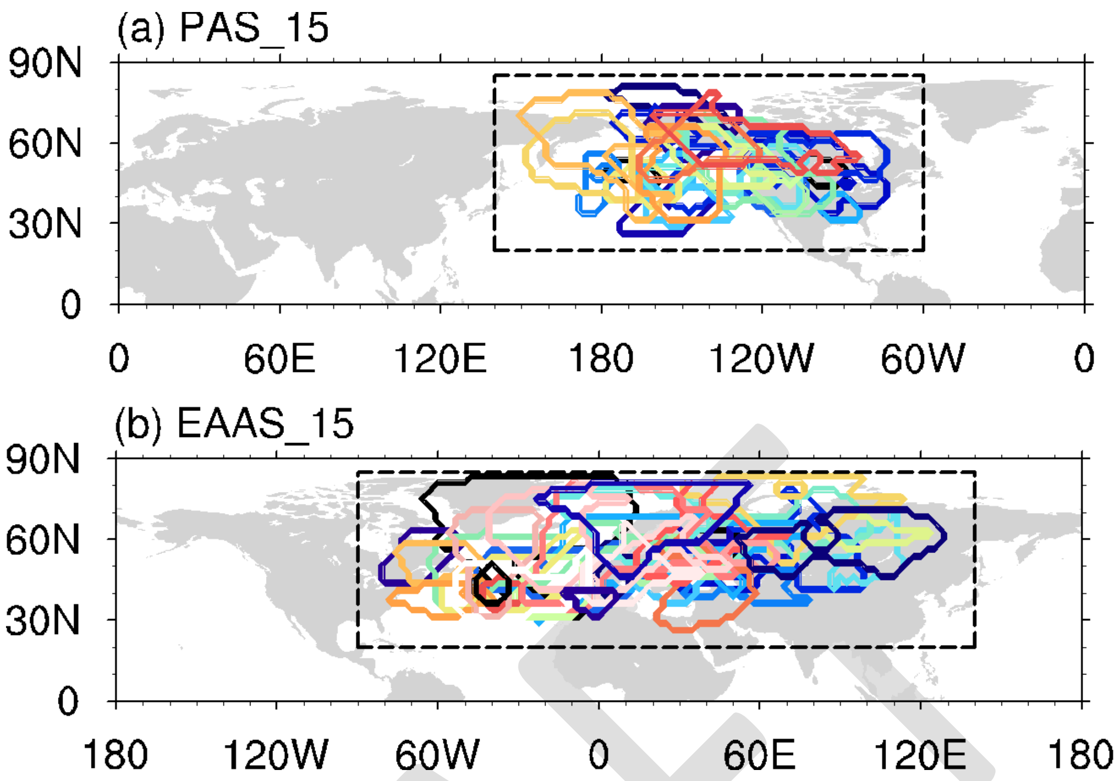


Fig. 7 Reliability diagrams of PMZ probability forecasts for (a) +1 day, (b) +3 day, (c) +6 day, (d) +9 day, (e) +12 day, and (e) +15 day in the Pacific (30-70^{o}N, 180-280^{o}E; dotted red) and Euro-Atlantic (30-70^{o}N, 60 ^{o}W-120^{o}E; dotted blue) sectors. Black-solid, green-solid, and black-dashed lines denote the perfect skill, no skill, and climatology probability, respectively.



855
856

857 **Fig. 8** Snapshots of PMZ impact areas during the onset day in PNAS (a) and EAAS
858 (b) for the cases PNAS_15 and EAAS_15 in Table.1. The black rectangles correspond
859 to the regions of 25-85^{o}N, 140-300^{o}E (a) and 25-85^{o}N, 90^{o}W-140^{o}E (b).

860

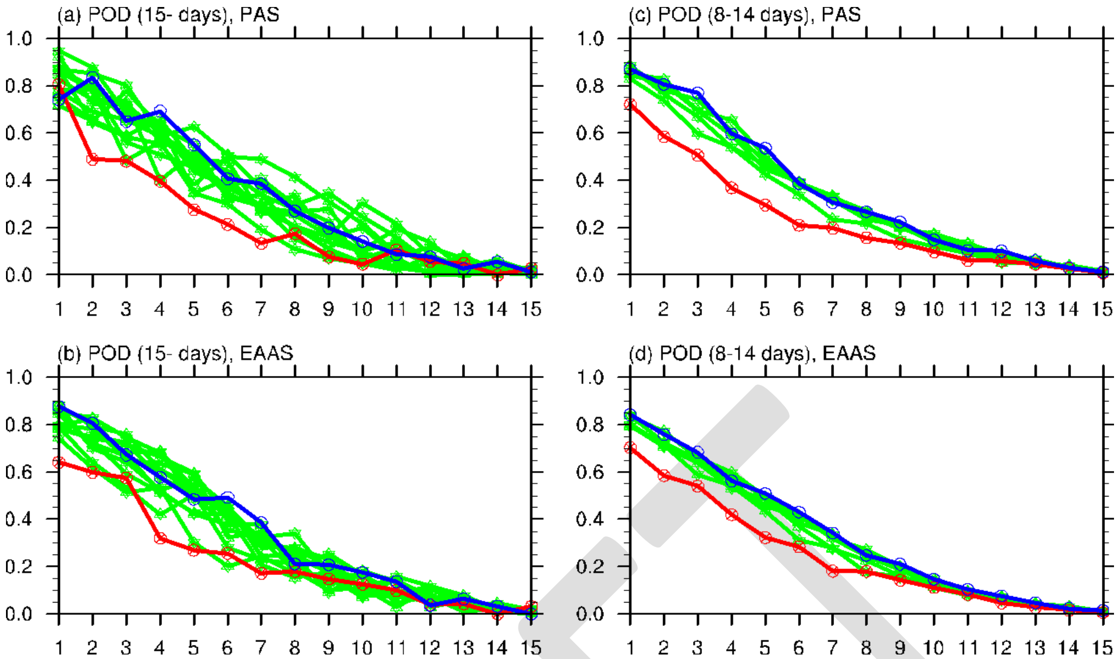


Fig. 9 (a) Probability of detection (POD) for the impact areas of 15-day PMZ cases in the GEFS ensemble mean. The abscissa denotes the lead time (days). The red, blue, and green lines denote the POD for the onset day (+0 day), +15 day after the onset, from +1 to +14 days, respectively. (b) The same as in (a) but for the POD of the 15-day cases in EAAS. (c) and (d) are the same as (a) and (b) but for the 7-14 day cases in PNAS and EAAS, respectively.

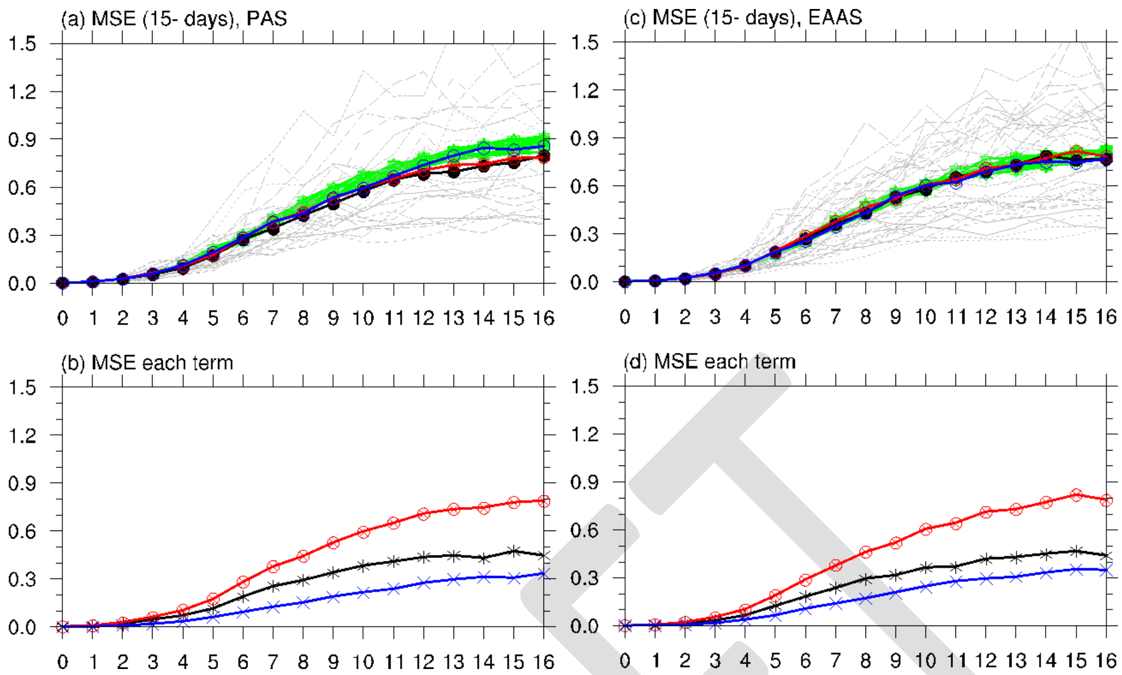


Fig. 10 (a) Averaged MSE for the GEFS ensemble mean of 500-hPa geopotential height anomaly ($\times 10^4$ gpm^2 per grid) for the 15-day cases in PNAS. The red, black, blue and green curves denote the MSE for the onset day (0 day), one day before the onset (-1 day) and +14 day after the onset, and +1 to +13 days, respectively. (b) Each term of MSE for the onset day: the red curve is the same as in (a); the black and blue curves denote the MSE from the ensemble mean and spread, respectively. (c) and (d) are the same as (a) and (b) but for the 15-days cases in EAAS.

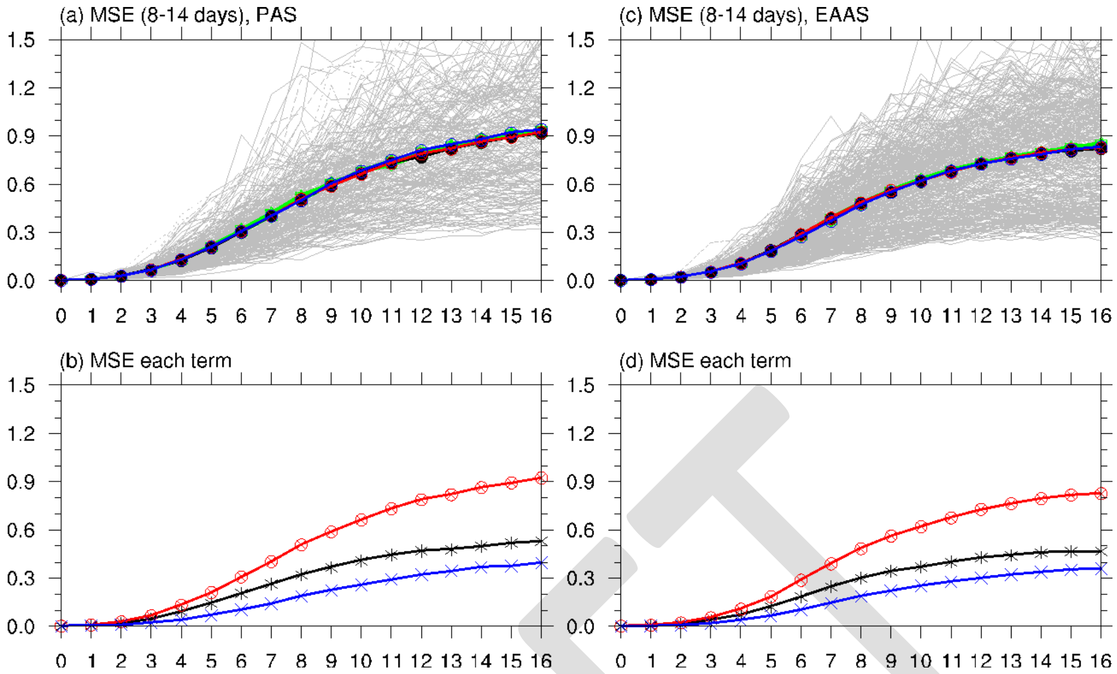


Fig. 11 The same as Fig.10 but for the 8-14-day cases in PNAS (a and b) and EAAS (c and d), respectively. The blue line denotes the MSE for +7 days after the onset.

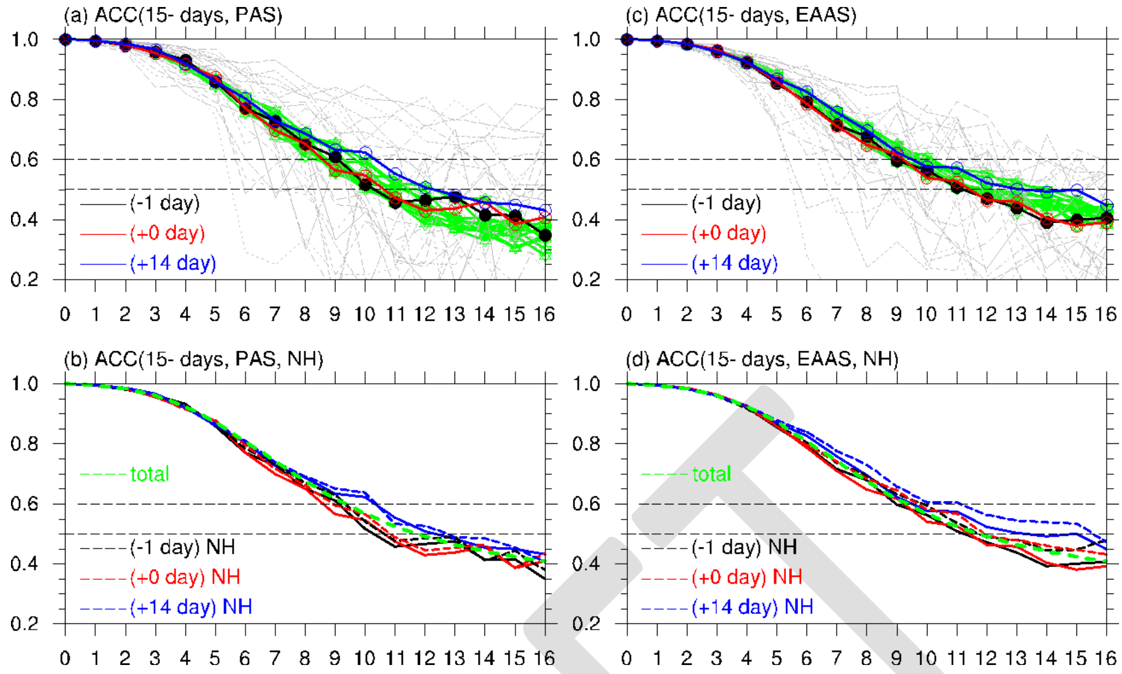


Fig. 12 (a) Averaged ACC for the GEFS ensemble mean 500-hPa geopotential height anomaly of the 15-day cases in Appendix Table 1. The abscissa denotes lead times (days). The black horizontal dashed lines denote the 0.6 and 0.5, respectively. The red, black, blue, green and gray curves denote the ACC for the onset day (+0 day), one day before the onset (-1 day), +14 day after the onset, +1 to +13 days, and the onset day for the 26 individual cases, respectively. (b) The solid lines are the same as in (a) and the dashed lines denote the ACC for the Northern Hemisphere. The green dashed line denotes averaged ACCs from 1 January 1985 to 31 December 2015. (c) and (d) are the same as (a) and (b) but for the 15-day cases in EAAS.

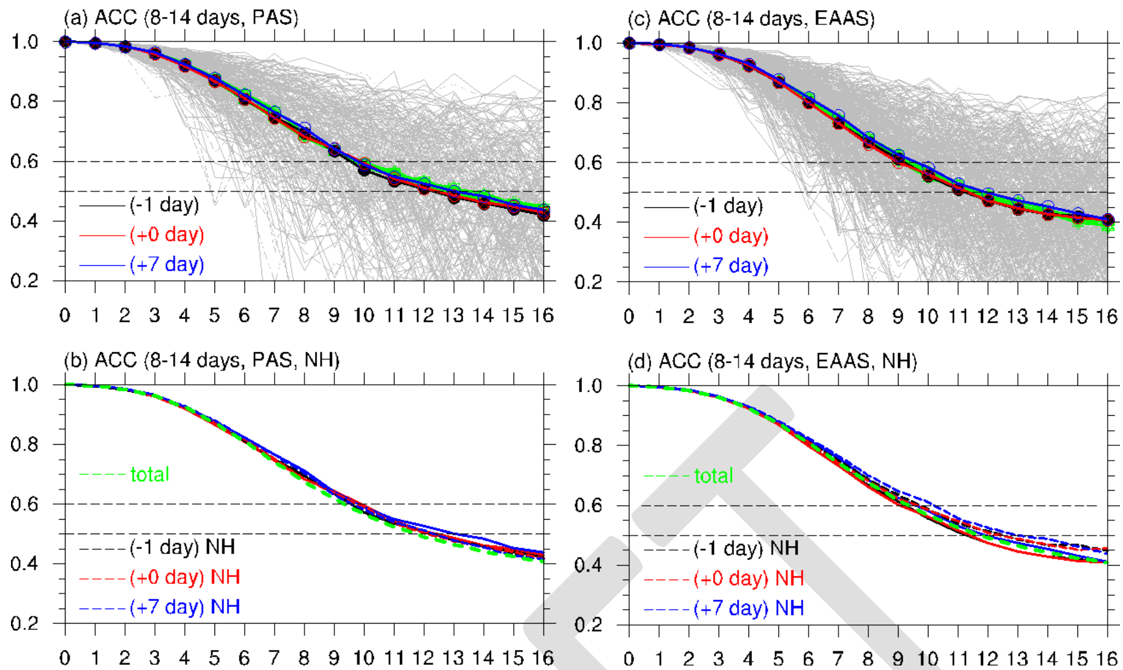


Fig. 13 The same as Fig.12 but for the 8-14 days cases in Appendix Table 1 for PNAS (a, b) and EAAS (c, d). The blue line denotes the ACC for +7 day after the onset.

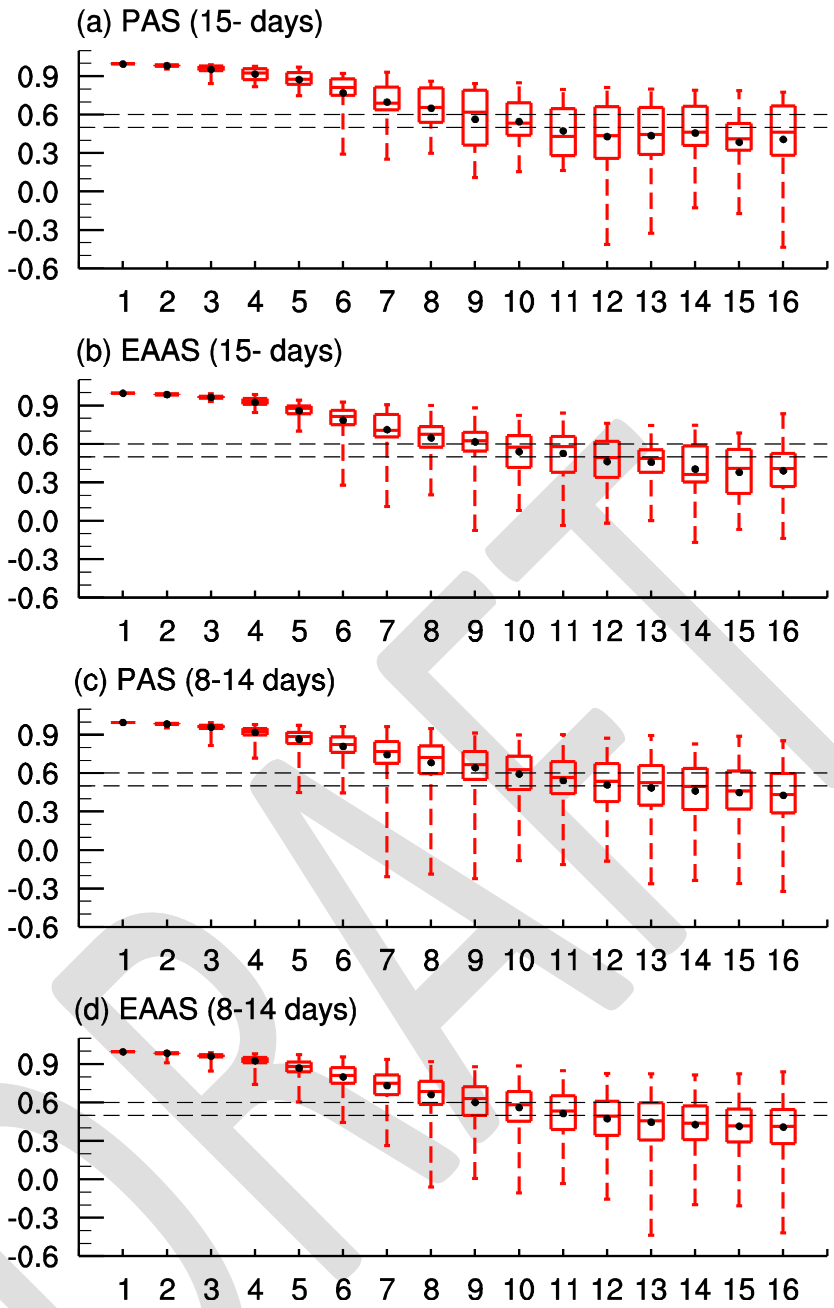


Fig. 14 (a) Box plots of ACCs for the onset day of the 15-day cases in PNAS (Appendix Table.1) with lead days in the GEFS ensemble mean. The black dashed lines denote 0.6 and 0.5, respectively. The black dots denote the mean value of the cases. (b) The same as (a) but for the 15-day cases in EAAS. (c) The same as (a) but for the 8-14-day cases in PNAS. (d) The same as (a) but for the 8-14-day cases in EAAS.

Appendix

Table 1 Onset dates, durations, central locations, and intensity for the PMZs longer than 15 days over PNAS in the GEFS ANL.

Date	Duration	Ave lat	Ave lon	intensity
19850101	20	48.9	237.1	365.3
19850911	24	50.1	216.9	197.4
19880528	19	51.4	265.6	192.7
19890522	15	62.1	223.2	166.6
19890706	19	62.4	241.9	163.3
19900421	16	40.6	223.5	171.6
19910512	19	53.9	265.7	188.1
19930827	16	55.1	222.8	216.4
19931218	17	52.4	232.6	279.2
19940416	16	43.4	266.8	177.8
19960429	15	55.9	199.6	190.4
19990424	20	54.1	285.3	199.5
19990726	15	58.2	221.3	178.9
19991213	19	47.7	229.3	348.1
20010621	16	46.3	255.6	153.2
20010714	15	59.6	231.6	151.4
20021012	16	56.3	231.7	245.7
20021123	20	55.3	233.8	316.7
20030710	15	45.0	249.0	151.0
20030902	17	52.2	272.7	194.3
20050219	22	49.4	237.5	279.5
20090928	18	53.8	214.9	214.0
20101208	15	53.1	182.0	313.0
20110226	18	68.1	186.2	316.8
20130110	17	48.9	230.2	355.2
20130717	16	56.7	203.7	171.8
20150515	18	64.3	224.4	259.2

Table 2 The same as Table 1 but over EAAS.

Date	Duration	Ave lat	Ave lon	intensity
19850805	15	57.9	38.6	200.8
19860224	16	42.4	328.0	237.1
19860625	20	53.6	28.9	162.9
19870603	17	63.1	323.9	189.0
19870605	26	56.6	52.2	203.3
19880318	16	58.4	55.7	253.2
19910617	17	54.9	47.9	197.4
19930130	16	40.7	83.4	151.7
19950324	18	49.8	346.2	308.1
19950513	32	57.3	42.1	238.7
19960407	19	51.7	4.8	230.1
19960430	16	54.7	49.0	202.2
19961111	19	52.9	51.7	297.9
19970410	16	48.0	85.6	213.4
19980904	16	58.1	40.4	225.5
19990930	21	56.2	62.2	246.2
20000410	16	65.2	322.1	204.1
20001030	15	56.3	50.8	256.0
20010408	15	48.7	337.3	236.0
20010818	17	61.0	107.5	160.3
20020217	16	42.9	351.9	224.1
20030815	15	65.4	81.4	171.6
20030923	25	52.9	332.2	263.5
20031223	16	46.1	331.7	273.4
20051203	17	49.7	60.6	190.1
20060523	15	48.8	344.4	164.5
20061120	17	54.9	31.9	290.1
20071212	15	60.5	11.6	435.4
20080617	16	63.2	57.6	133.9
20091230	20	62.8	347.2	332.3
20100505	20	53.6	339.3	212.8
20100709	30	58.7	39.3	189.6
20121124	16	49.9	47.6	199.2
20130210	18	56.1	351.0	315.5
20130616	15	45.1	333.9	163.5
20130621	15	64.1	49.7	234.5
20130730	20	52.5	32.2	152.3
20140123	15	58.6	28.5	318.6

Table 3 PMZ events with ACCs for the onset day above 0.6 from lead days +10 to +16 in the GEFS ensemble mean

Date	Duration	Ave lat	Ave lon	intensity
20110226	18	68.1	186.2	316.8
20130110	17	48.9	230.2	355.2
20150515	18	64.3	224.4	259.2
20121124	16	49.9	47.6	199.2

Table 4 The same as Table 3 but for +15 days after the onset

Date	Duration	Ave lat	Ave lon	intensity
19931218	17	52.4	232.6	279.2
19991213	19	47.7	229.3	348.1
20050219	22	49.4	237.5	279.5
19860224	16	42.4	328.0	237.1