



Prediction and predictability of Northern Hemisphere persistent maxima of 500-hPa geopotential height eddies in the GEFS

Bian He^{1,2,4} · Ping Liu² · Yuejian Zhu³ · Wenting Hu^{1,2,4}

Received: 22 October 2017 / Accepted: 8 July 2018 / Published online: 11 July 2018
© The Author(s) 2018

Abstract

This study analyzes the predictability of the persistent maxima of 500-hPa geopotential height (Z500; PMZ) zonal eddies over the Northern Hemisphere in the long-term forecast datasets of the Global Ensemble Forecast System (GEFS) version 10. PMZ patterns, which potentially extending the predictability of severe weather events, 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). The predicted PMZ occurrence frequencies in both the EAAS and the PNAS generally decrease with the lead time, which is consistent with classical blockings in early studies but different from the nearly invariant frequencies of blockings in a recent relevant diagnosis by Hamill and Kiladis. The Brier skill score associated with PMZ frequencies is generally higher in the PNAS than in the EAAS, indicating better predictions in the former. The forecast reliability decreases with the lead time in both sectors, particularly at the tails of probability distributions, suggesting some limitations of the GEFS. PMZ events longer than 1 week with anomaly correlation coefficients (ACCs) exceeding 0.6 in the Northern Hemisphere have a mean useful skill of nearly 10 lead days, which is approximately 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 25% extend the useful skill even further. A discussion is provided regarding how the better PMZ prediction skill in the PNAS can help improve 2 to 3-week predictions over North America.

Keywords Persistent atmospheric pattern · Prediction skills · 500-HPa geopotential height · Medium-range forecast · Blocking · Ensemble prediction · GEFS

1 Introduction

The Next Generation Global Prediction System (NGGPS), which is currently being developed by the National Weather Service (NWS) of the National Oceanic and Atmospheric Administration (NOAA) in collaboration with other

agencies, laboratories and universities in the U.S., aims to address growing service demands and extend weather forecasts up to 4 weeks, which is beyond the prediction limit of the first type of forecast system (Lorenz 1963, 1982). The NGGPS will adopt most of the packages representing physical processes in the current operational system known as the Global Ensemble Forecast System (GEFS; Zhu 2005; Zhu and Toth 2008; Zhou et al. 2016), which was recently upgraded to extend its operational forecasts to 16 days, while the ultimate goal of the NGGPS is to achieve a forecast length of 35 days (Zhu et al. 2017a, b). Calibrating and evaluating the ensemble forecasts of the GEFS at these new ranges is thus a valuable step toward 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 to provide short and medium-range forecasts. The anomaly correlation coefficient (ACC) for the 7th-day skill of the European Centre for Medium-Range Weather Forecasts (ECMWF) was improved from 0.4 in

✉ Bian He
heb@lasg.iap.ac.cn

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, NY 11794-5000, USA

³ Environmental Modeling Center, NCEP, NOAA, College Park, MD, USA

⁴ University of Chinese Academy of Sciences, Beijing 100029, China

1981 to 0.7 in 2017. The updated predictability of Z500 reaches up to 10 days when an ACC of 0.6 is considered the lower limit of a useful skill (<https://www.ecmwf.int/en/forecasts>). The compatible predictability in the ensemble mean of GEFS v10 was 8.9 days in 2014, and it improved to 10.5 days in version 11 in 2016; this predictability is better than the corresponding values of 7.9 days and 8.5 days in the deterministic forecasts by the Global Forecast System (GFS) in 2014 and 2016, respectively (Fig. 1 in Zhu et al. 2017b).

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

Blocking predictions have been improved with the development of models. In the early model versions of the ECMWF (e.g., Tibaldi and Molteni 1990; TM90 hereafter), the blocking frequency was severely underestimated, and the blocking onset was poorly predicted even a couple of days beforehand. Initializing those models with established blocking patterns, however, improved the associated weather predictions. In the early 2000s, the blocking prediction of the ECMWF model had been improved by as much as 50%, although the predicted frequency was still 30% smaller than that in the analysis (Mauritsen and Källén 2004). Similarly, a small 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 the more recent GEFS v10 was substantially improved over the Euro-Atlantic sector (Hamill and Kiladis 2014) with nearly constant occurrence frequencies that were only slightly smaller than those in the analysis even at lead times extending to 16 days. This occurrence frequency is nearly invariant in contrast to the frequencies that decreased with the lead time in early studies (e.g., TM90). The predicted blocking frequency remains very small in the Pacific sector, partly because blockings occur relatively rarely there (e.g., Pelly and Hoskins 2003).

Some persistent high pressure systems, in addition to closed blockings involving the reversal of meridional gradients, also induce disastrous weather events (IPCC 2013; Grotjahn et al. 2016; Liu et al. 2017). These systems include

omega-shape blockings, especially in their early stages, and persistent open ridges, which cannot be clearly identified as blockings by classical indices such as the TM90, which requires a reversal of meridional pressure gradients in reference to central blocking latitudes. These systems appear as closed anticyclones in the zonal eddy anomalies of Z500 after removing the zonal mean (Cheng and Wallace 1993; L'Heureux et al. 2008; Liu et al. 2017), but their time anomalies may not satisfy the criteria for blocking (e.g., Dole and Gordon 1983). In addition, persistent high pressure systems, some of which were classified as blockings, were identified according to Z500 time anomalies in the Southern Hemisphere, and they were mainly located over New Zealand, southeastern South America and the southern Indian Ocean (Trenberth and Mo 1985; Sinclair 1996). To date, however, the predictability of these persistent systems has not been investigated.

Predicting the persistent maxima of Z500 eddies (PMZs) in the Northern Hemisphere is particularly useful, as they effectively represent both the open ridges and the closed anticyclones of blocking patterns (Liu et al. 2017). For example, a PMZ event occurred over the northeastern Pacific during January 2013 (Appendix Table 2) and persisted for 17 days (Fig. 1a–h), leading to a cold surge over the western United States (Fig. 1i–p). In this case, most persistent flow patterns could not be identified as a blocking by a typical algorithm (e.g., TM90), because they were apparently characterized as strong open ridges before developing into a mature phase (Fig. 1g, h). However, this PMZ influenced the surface weather similar to blockings, and thus, PMZs need to be considered in forecasts.

The statistics of PMZs in observational data (Liu et al. 2017) also contrast with those of blocking events (Hamill and Kiladis 2014). Two substantial differences are evident in the climatology over the Pacific-North America sector (PNAS) during the wintertime. First, PMZs form near the U.S. West Coast (Fig. 9a in Liu et al. 2017), while typical blockings occur farther westward near the date line (Fig. 1 in Hamill and Kiladis 2014), suggesting a more direct impact of PMZ events on the weather over the U.S. Second, the U.S. West Coast exhibits a center of PMZ frequency of 33% (Fig. 9a in Liu et al. 2017), which is twice as large as the center of the blocking frequency (Fig. 2a in Hamill and Kiladis 2014) in the mid-Pacific. Therefore, it is worth examining the predictability of PMZs in the medium-range forecasts of the GEFS. Such an estimate would be helpful to extend predictions to 2–4 weeks.

In this study, we estimate the prediction skills and predictability of PMZs according to their occurrence frequencies and of individual PMZ events by employing GEFS v10 forecasts and several objective verification metrics. Section 2 summarizes the algorithm tracking PMZs in Liu et al. (2017) and introduces the datasets and methods. Section 3 estimates

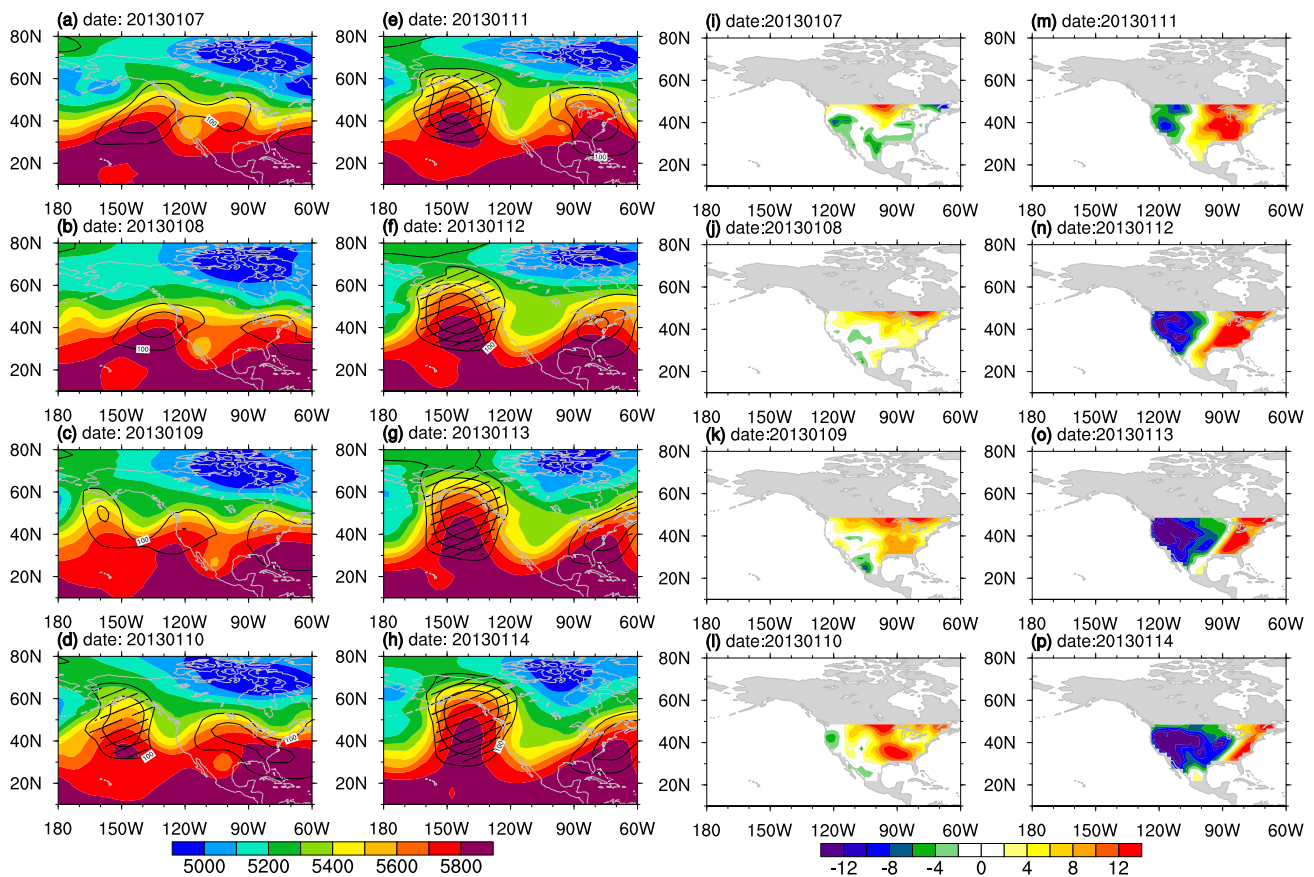


Fig. 1 a–h 500-hPa geopotential height (Z500, GPM, shaded) zonal eddies above 100 GPM (black contours) and impact areas of PMZs (hachured areas) in the ANL at 00Z UTC during 7–14 January 2013.

i–p Corresponding anomalies of the daily air temperature at 2 m (°C) over the continental U.S. (CONUS) in the NCEP-National Center for Atmospheric Research (NCAR) reanalysis

the prediction skills and predictability of PMZs in the occurrence frequency, Brier skill score (BSS), reliability diagram, probability of detection (POD), mean square error (MSE), and ACC. 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 these datasets were used to estimate the predictabilities of typical blockings by Hamill and Kiladis (2014). GEFS forecasts 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 from wavenumbers 254 (~55 km) to 190 (~70 km), 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 the analysis data for initializing the control run, and this initial condition was perturbed using the ensemble transform with the rescaling technique (Wei et al. 2008) to initialize other ensemble members. The uncertainty therein was estimated using the stochastic total tendency perturbation method (Hou et al. 2008).

The GEFS v10 forecasts between 1 January 1985 and 14 February 2012 were regenerated offline at the Earth System Research Laboratory (Hamill et al. 2013), and the forecasts to the present day have been made in real time. The offline 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 extended to 16 days. The forecasts prior to 31 December 2015 are combined in the present study after being bilinearly interpolated onto $2.5^\circ \times 2.5^\circ$ 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 PMZs

An objective algorithm was developed by Liu et al. (2017) to track the patterns of PMZs, including persistent open ridges, immature omega-shape blockings and mature blocking highs. The proposed algorithm identifies and connects the local maxima of zonal eddies of Z500, and it tracks PMZs in the GEFS analysis (ANL) at each 00Z UTC, which is slightly different from the daily mean data in Liu et al. (2017). It identifies a PMZ event as consecutive maxima lasting for 2 days or longer, which is notably shorter than the 4-day limit in Liu et al. (2017). As a result, more PMZs are tracked in the GEFS forecasts for verification. The tracking steps in consecutive order are summarized below.

- A core at each 00Z UTC is identified to include a local maximum of Z500 eddies and its neighboring grids, whose values are greater than 100 geopotential meters (GPMs) and decrease radially to 20 GPMs smaller than the maximum value.
- 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° longitude per day.
- The PMZ ends at the core without a successor.
- Each of the tracked cores is expanded to include an impact area consisting of more contiguous points whose zonal eddy values are above 100 GPMs. A non-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

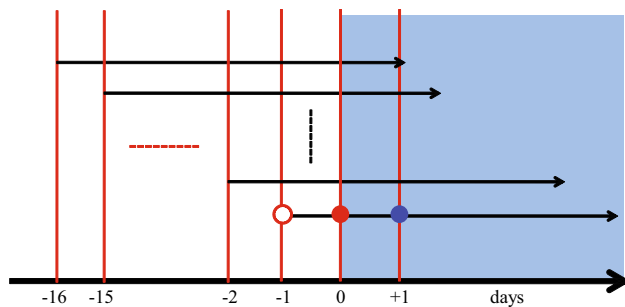


Fig. 2 Schematic diagram of time-lagged forecasting and PMZ tracking in the GEFS reforecasts. The bottom thick, black arrow denotes the time in days, where “0” represents the PMZ onset, and the thin, black arrows represent the forecast day (16 for each run) and PMZ tracking directions. The solid red 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 is represented by the red dot, and the day after onset is represented by the blue dot

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 the 17th day for one forecast. For the probabilistic forecast verification, the forecast datasets are regrouped into 16 time series ranging from the initial date on 1 January to 16 January 1985. Each time series contains 11,322 time slices for which PMZ events are tracked, and their impact areas are the 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. The prefixed observations guarantee that a tracked PMZ has an onset on or after the initial conditions, i.e., as early as on day -1 (open circle in Fig. 2), and that it extends to at least day $+1$ (blue dot in Fig. 2). As a result, the PMZs with an onset ranging from day $+1$ to day $+15$ will be used to estimate the deterministic prediction skills and predictability. It is noteworthy that the impact areas rather than the tracks of PMZs are verified in this study.

2.3 Evaluation metrics

The prediction skills of PMZs are evaluated using five objective metrics (Brankovic et al. 1990; Wilks 2006; Hamill and Juras 2006): the BSS, reliability diagram, POD, MSE, and ACC. Each metric is summarized below.

2.3.1 Brier skill score

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

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

where $I_i(j)$ is 1 if an event occurs and 0 otherwise, and n is the number of forecasts in the j th sample. The Brier score (BS_f) of the forecasts is defined as

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

where the subscript o denotes the observations, and m is the number of samples. The BSS is finally computed as

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

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 it is defined as

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

and

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

When an ensemble member is used as the reference (also known as a perfect model), the prediction skill above a threshold becomes the predictability of the PMZ. The ensemble mean is selected as the perfect model in this study.

2.3.2 Reliability diagram

The Brier score in Eq. (2) can be decomposed into three components

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

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 $p_f(j)$ (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 Eq. (4). The three terms on the right-hand side of Eq. (6) are successively known as the reliability, resolution, and uncertainty. The reliability diagram constructed with these three components comprehensively assesses the forecast quality by representing a joint distribution of forecasts and observations.

2.3.3 Probability of detection

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

$$POD = \frac{H}{H + M}, \tag{7}$$

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 decrease with later lead times (to be shown below), the false alarm rate of PMZs is not discussed here.

2.3.4 Mean square error

The MSE for an ensemble forecast of N members and the F_i for the i th member is expressed as

$$MSE = \frac{1}{N} \sum_{i=1}^N |F_i - X|^2 = |\bar{F} - X|^2 + \frac{1}{N} \sum_{i=1}^N |F_i - \bar{F}|^2, \tag{8}$$

where $\bar{F} = \frac{1}{N} \sum_{i=1}^N F_i$ 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, namely, the square errors from the ensemble mean ($|\bar{F} - X|^2$) and the variance from the ensemble mean ($\frac{1}{N} \sum_{i=1}^N |F_i - \bar{F}|^2$).

2.3.5 Anomaly correlation coefficient

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

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

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

3 Results

3.1 Statistical frequency verification

The climatological statistics of PMZ impact areas in the GEFS ANL are first presented. The frequency distributions in different seasons are shown in Fig. 3. PMZs mainly occur at latitudes of 30–80° in the Northern and Southern Hemispheres (Fig. 3a–d) with centers close to mid-latitude jet exit areas (Pelly and Hoskins 2003). The frequency has a clear annual cycle that is larger in the winter (Fig. 3a) than in the summer (Fig. 3c) in both hemispheres. In December–January–February (DJF), two maximum-frequency centers are located over the Northeast Pacific and Northeast Atlantic coasts. The Northeast Atlantic center expands from the Atlantic to the Euro-Asian continent with a size larger than that over the northeastern PNAS, similar to the blocking frequency distributions (TM90; Pelly and Hoskins 2003; Barriopedro et al. 2010; Hamill and Kiladis 2014). The locations of the frequency centers, however, are different: the maximum blocking frequency is located over the central Pacific close to 180°E (Fig. 1 in TM90), whereas the maximum PMZ frequency is situated along the western coastline of North America (Fig. 3a) corresponding to persistent northerly winds and

potentially colder weather in the western U.S. (Fig. 1). In addition, the maximum PMZ frequency near the U.S. West Coast reaches 40%, which is more than double that of the blocking frequency over the central Pacific, suggesting that the extreme events over the western U.S. are associated with PMZ patterns more than traditional blocking events. In June–July–August (JJA; Fig. 3c), the PMZ frequency center over North America shifts inland in contrast to the other three seasons. This eastward shift may lead to persistent high pressure systems over the western U.S., resulting in potential droughts and heat waves. In the Southern Hemisphere, the PMZ frequency is distributed quite similarly to the frequencies of blockings with large values primarily across the South Pacific (Trenberth and Mo 1985; Sinclair 1996; Renwick and Revell 1999). However, the PMZ frequency has a maximum of nearly 20%, which is higher than the 12% maximum for blockings (Fig. 4 in Sinclair 1996).

We next assess the skills of the GEFS in predicting PMZ occurrence frequencies in the Northern Hemisphere. The seasonality of PMZ frequencies (40–60°N mean) in the forecasts is compared with that in the ANL according to the lead time (different colored curves in Fig. 4). The predicted PMZ frequencies in all seasons have distributions that are overall similar to those in the ANL, and they decrease with the lead time. The decreasing rates, however, are different in each season. In DJF (Fig. 4a), two maxima in the ANL and forecasts are located near 120°E and 10°W, respectively. The predicted frequencies decrease from 40% on day +3 to 30% on day +15 in both the Pacific and the Euro-Atlantic sectors with a gap between days +9 and +12 decreasing

sharply from 35 to 30%. This decrease is consistent with blocking predictions in some studies (TM90; Mauritsen and Källén 2004; Jia et al. 2014) but is notably different from the predicted blockings whose frequencies do not change substantially with the lead time except for a slight decrease in the Euro-Atlantic sector (Hamill and Kiladis 2014). The detail is discussed in the last section. In March–April–May (MAM; Fig. 4b) and September–October–November (SON; Fig. 4d), the frequencies decrease from 25% on day +3 to less than 10% on day +15. In JJA (Fig. 4c), the frequencies are the smallest with a maximum of approximately 18% over the Pacific, and they decrease to less than 4% on day +15 when the rest of the frequencies become nearly zero at other longitudes (blue curve in Fig. 4c).

The BSS is then computed to assess the probabilistic forecasts of the PMZ frequencies in the Pacific and Euro-Atlantic sectors (TM90; Hamill and Kiladis 2014). Since the PMZ occurrence frequency is overall larger than the frequencies of blockings, the ranges of the two sectors are extended somewhat. The Pacific sector stretches to include the PNAS covering 180°E–60°W, and the Euro-Atlantic becomes the Euro-Atlantic-Asia sector (EAAS) at 60°W–120°E. The corresponding BSSs in DJF for both sectors are shown in Fig. 5, in which the dashed lines are the BSS of the perfect model. The BSS for the PNAS (red line) is overall higher than that for the EAAS (blue line) at all lead times, suggesting that PMZs are more predictable in the PNAS than in the EAAS. In both sectors, the BSS decreases rapidly from 0.9 to 0.5 in the first three days. This rapid decrease contrasts with that of the perfect model, where the BSS decreases more slowly and remains above 0.2 after day +9 (the dashed lines). Compared

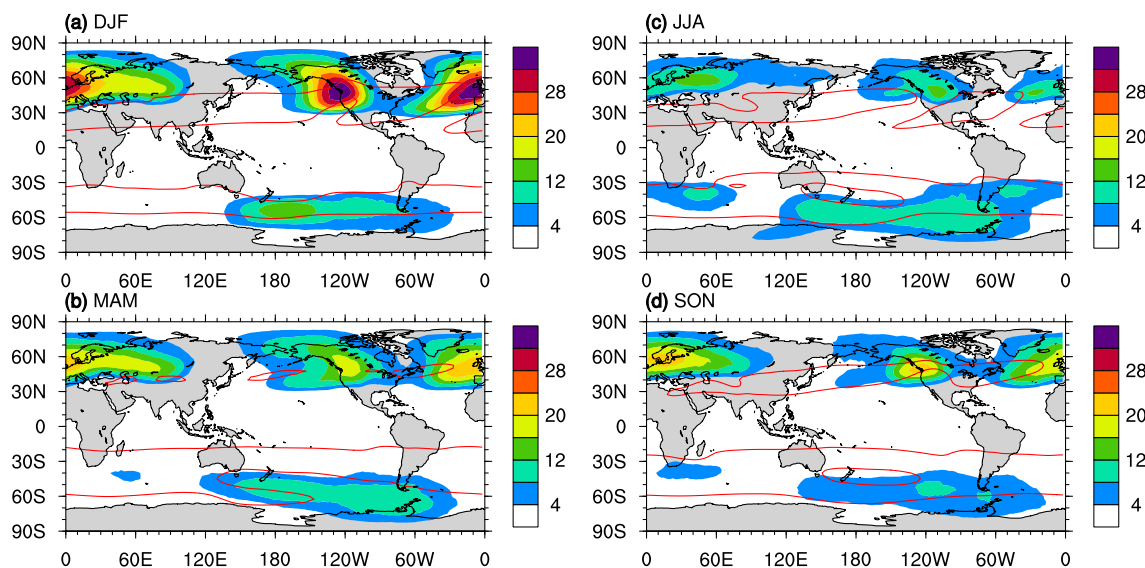


Fig. 3 Frequency distributions (%) of the PMZ impact areas in the GEFS ANL in **a** DJF, **b** MAM, **c** JJA, and **d** SON during 1979–2015. Red contours denote the climatological zonal wind speed of 20 m s^{-1} at 300 hPa

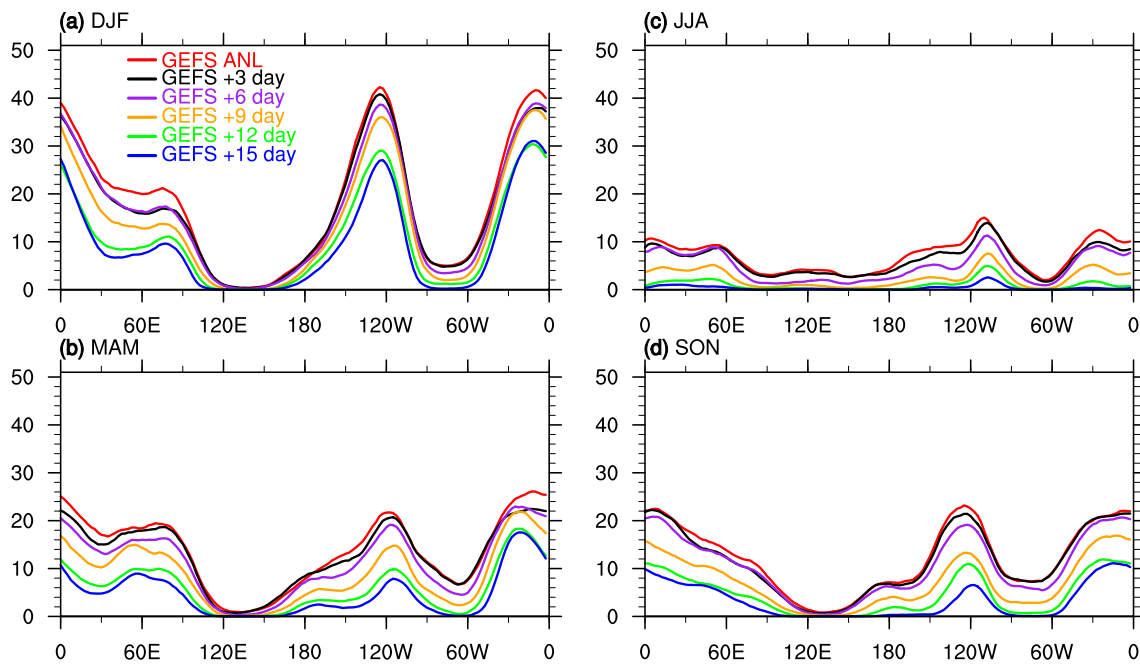


Fig. 4 Frequency distributions (%) of the PMZ impact areas averaged over 40–60°N for the ensemble mean of the GEFS reforecasts during a DJF, b MAM, c JJA, and d SON

with the BSS of the blocking frequency (Fig. 3 in Hamill and Kiladis 2014), the BSSs of the PMZs for both sectors are generally similar except for a faster decrease around day +5.

The reliability diagram is constructed for additional evaluations of the probabilistic skills. The 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 Eq. (6). The frequency distributions on the reliability diagram are shown in Fig. 6. On day +1 (Fig. 6a), the samples are concentrated in bin 0.0 with 14×10^5 and 24×10^5 grids for the PNAS and EAAS, respectively. These PMZ histograms are similar to those of blockings (Fig. 5 in Hamill and Kiladis 2014).

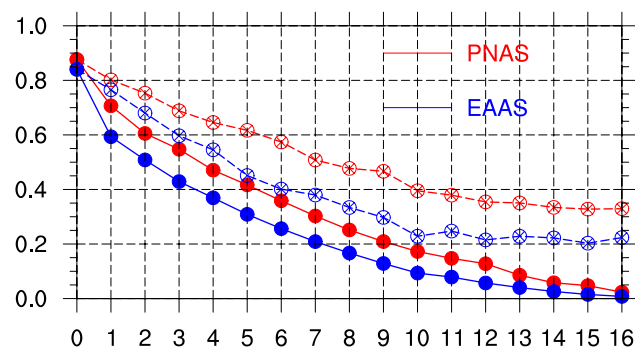


Fig. 5 Brier skill scores of PMZ probabilistic forecasts for the PNAS (dotted red) and EAAS (dotted blue) during DJF. Dashed lines are for the perfect model

Reliability diagrams for the PMZ probabilistic forecasts according to the lead time are shown in Fig. 7. The diagonal line denotes a perfect reliability, in which the frequencies predicted are identical to those observed. The green line indicates no skill, as it is located halfway between the observation and climatology. The thick, black dashed line represents no resolution, along which the forecast equals the climatology. Clearly, the forecasts for both the PNAS and the EAAS are quite reliable on day +1 (Fig. 7a) with a slight underestimation in bins 0.1–0.6 and a slight overestimation in bins 0.9–1.0; they become less reliable and more overestimated in the EAAS than in the PNAS in bins 0.6–1.0 at longer lead times (Fig. 7b–f). The underestimation is even more evident in bins 0.0–0.4 and on days +9 to +15 in both sectors (Fig. 7d–f). The reliability decreases with an increase in the lead time in larger frequency bins as well, but it remains situated between the no-skill and perfect-reliability lines, indicating that there is some skill. These results are very similar to those obtained for blockings (Fig. 5 of Hamill and Kiladis 2014).

3.2 Verification of the ensemble mean forecast

This section presents the skills of the GEFS ensemble mean in predicting individual PMZ events at different stages and estimates how these events extend the predictability of Z500 eddies. Individual PMZ events in the ANL are first counted in the PNAS and EAAS according to their durations of 4–7,

8–14 days, and greater than 14 days (Table 1). In the PNAS, there are 1255 events in total, 1002 of which have durations of 4–7 days, 226 have durations of 8–14 days and 27 have durations of longer than 14 days; in the EAAS, there are 1657 events in total, 1244 of which have durations of 4–7 days, 375 have durations of 8–14 days and 27 have durations of longer than 14 days. Events longer than 14 days will be of particular interest, and their dates, locations and intensities are listed in Appendix Table 2 for the PNAS and Appendix Table 3 for the EAAS.

The skills in predicting the three types of PMZ events are evaluated separately using the POD, MSE, and ACC;

greater focus is given to the events persisting for 8–14 days and longer than 14 days. The POD is derived by counting the PMZ impact areas in the forecasts and observations, and the MSE and ACC choose common regions enclosing all PMZ events that persist for 15 days (each color curve represents one event in Fig. 8). The common regions cover (25–85°N, 140–300°E) in the PNAS (Fig. 8a) and (25–85°N, 90°W–140°E) in the EAAS such that they include most of the PMZ events that persist for 8–14 days (not shown).

The verification results with the POD are summarized in Fig. 9. For the 15-day PMZs in the PNAS (Fig. 9a), the mean POD on the onset day (red line) decreases dramatically with

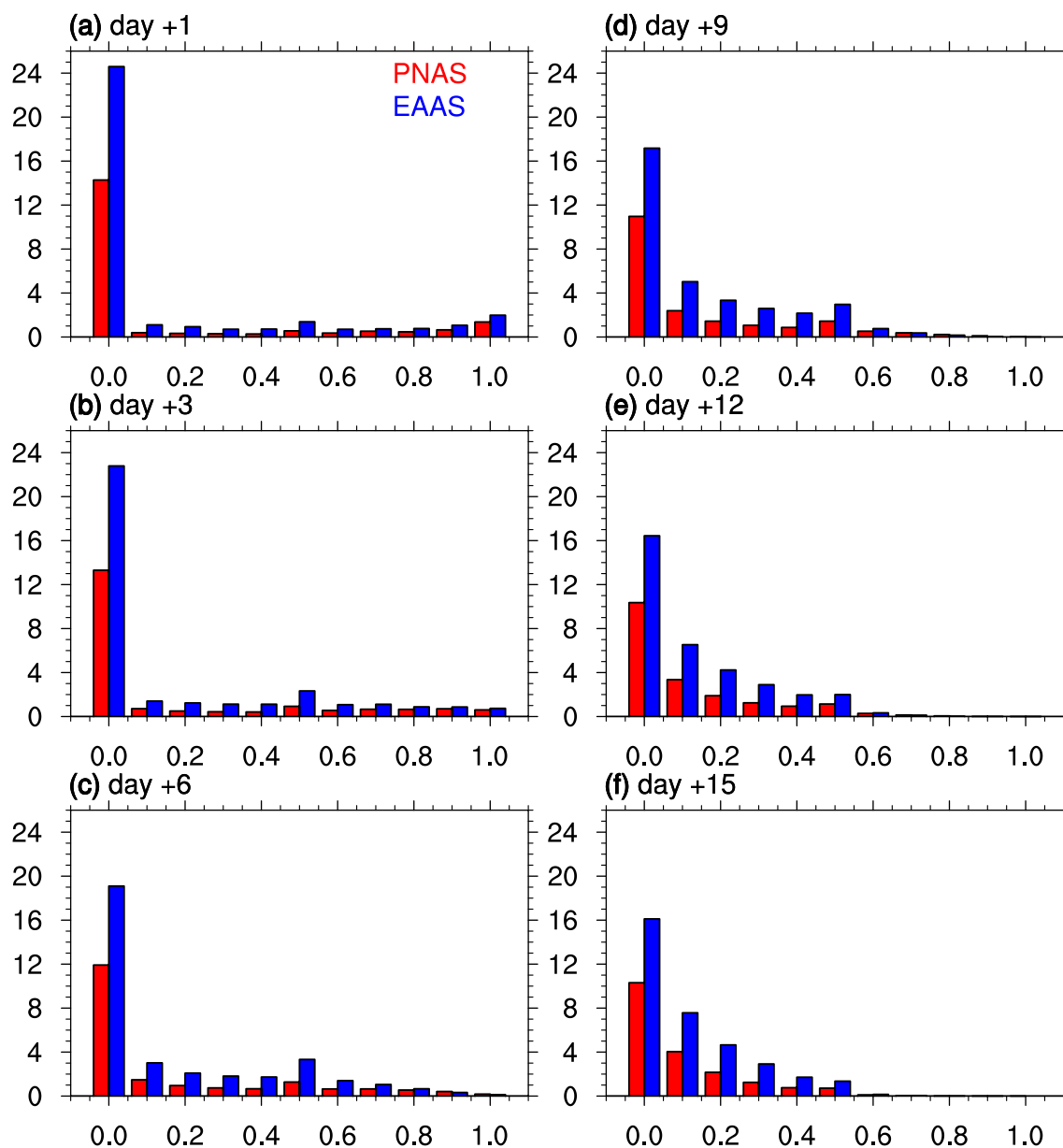


Fig. 6 Histograms for the numbers of grids ($\times 105$) inside PMZ impact areas in the PNAS (red) and EAAS (blue)

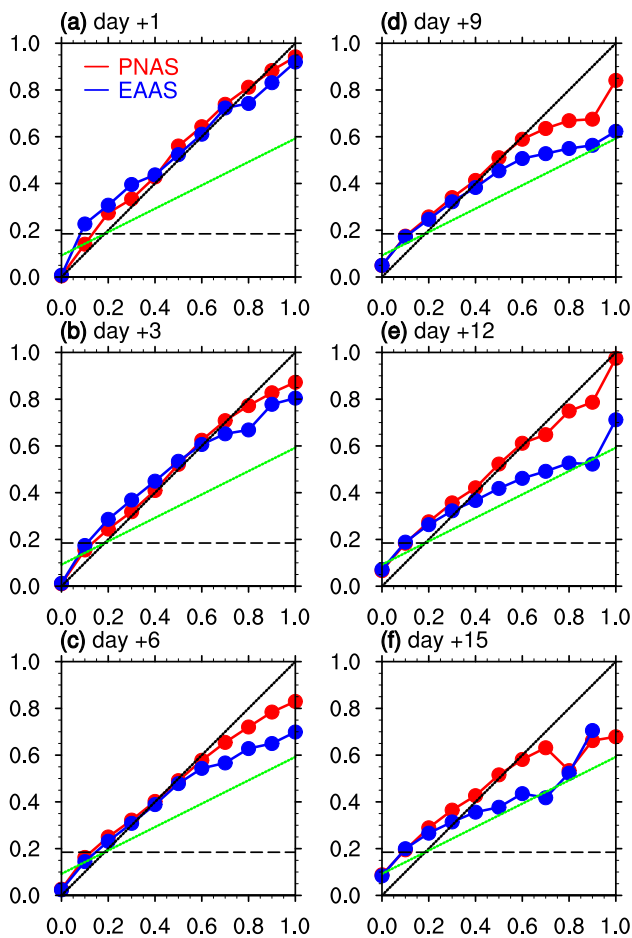


Fig. 7 Reliability diagrams of PMZ probabilistic forecasts for **a** day +1, **b** day +3, **c** day +6, **d** day +9, **e** day +12, and **f** day +15 in the Pacific (30–70°N, 180–280°E; dotted red) and Euro-Atlantic (30–70°N, 60°W–120°E; dotted blue) sectors. The black solid, green solid, and black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

increasing lead time from days +1 to +15. The POD after the onset (green lines) shows overall higher skills than those on the onset day, indicating that established PMZs in the initial conditions improve the prediction skill. In contrast,

Table 1 Statistics of PMZ events in the ANL for the Pacific (180°E–60°W; PNAS) and Euro-Atlantic-Asia (60°W–120°E; EAAS) sectors

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

The PMZs are grouped by durations of 4–7 days, 8–14 days, and longer than 15 days with maximum durations shown in the last row

the POD for day +15 shows a higher skill than the onset day until lead day +11, indicating better skills with regard to predicting the durations than the onsets of PMZs. In the EAAS (Fig. 9b), the mean POD on the onset day (red line) is 0.65 on lead day +1, which is slightly smaller than that in the PNAS (Fig. 9a). Meanwhile, the POD of day +15 (blue line in Fig. 9b) is 0.85 on lead day +1, which is higher than that in the PNAS (blue line in Fig. 9a). These results indicate that the GEFS exhibits a better skill in predicting the onsets of PMZs with a lifetime of longer than 15 days in the EAAS than in the PNAS. For the 8–14-day cases (Fig. 9c, d), the mean POD evolves more smoothly than those for the 15-day cases in both the PNAS and the EAAS; this is 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).

The POD analysis in Fig. 9 indicates that the errors in the prediction of individual PMZ events grow with an increase in the lead time. These errors can be random in nature, as they can originate either from the variability within the ensemble or from the model’s systematic bias; they are further measured by the MSE (Brankovic et al. 1990) of Z500 eddy anomalies (Fig. 10). Figure 10a shows the mean MSE for the 15-day cases in the PNAS, in which the red, black, green, and blue curves represent the onset, 1 day before onset, days +1 through +13, and day +14 after onset, respectively. Clearly, the MSE increases gradually with the lead time during the developing stages of PMZs, and it grows to 0.9×10^4 GPM² per grid on day +16. The MSEs grow more rapidly after lead day +4 and become notably larger during the developing stages of PMZs than those during days –1 and 0 (black and red curves) and after day +5. This result indicates that the forecast error increases when long-lived PMZ events develop into mature stages in the PNAS. For the 15-day cases in the EAAS (Fig. 10c), the

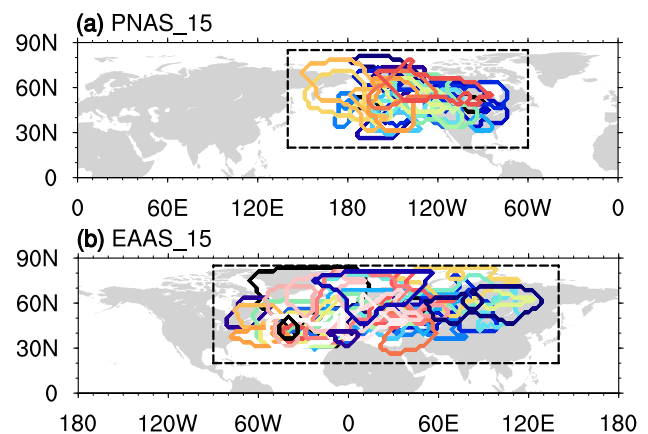


Fig. 8 Snapshots of the PMZ impact areas during the onset day in the PNAS (a) and EAAS (b) for the cases PNAS₁₅ and EAAS₁₅ in Table 1. The black rectangles define the two sectors

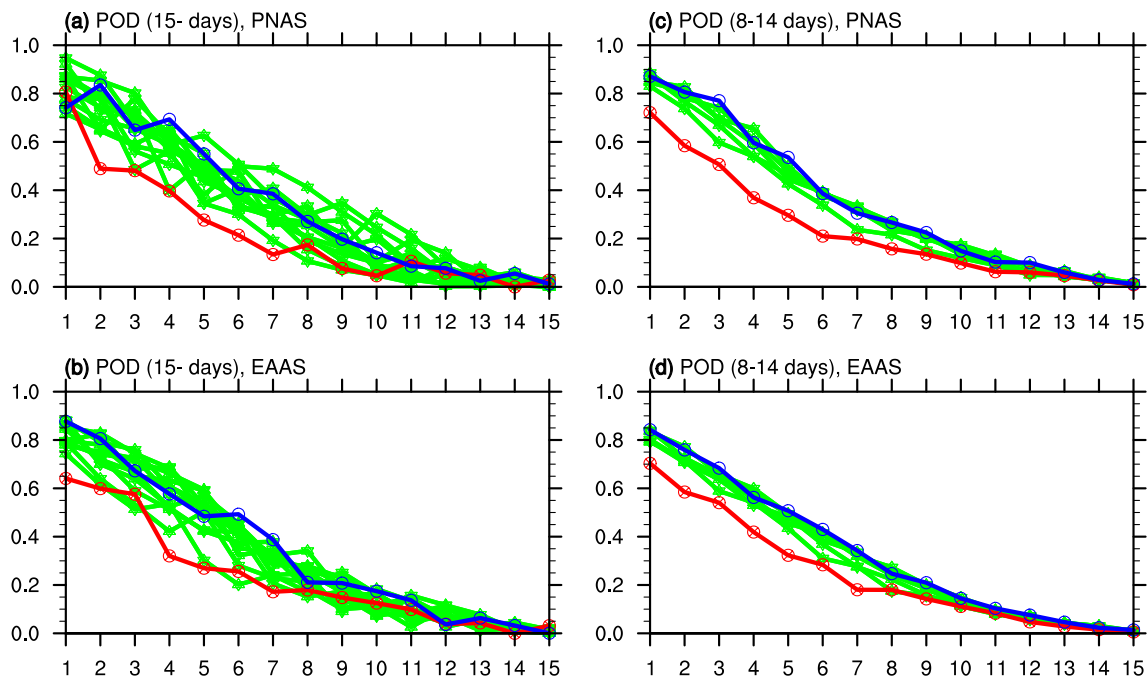


Fig. 9 **a** Probability of detection for the impact areas of the 15-day PMZ cases in the GEFS ensemble mean. The abscissa denotes the lead time (days). The red, blue, and green lines represent the PODs for the onset day (day+0), day+15 after the onset, and days+1

through +14, respectively. **b** The same as in **(a)** but for the PODs of the 15-day cases in the EAAS. **(c)** and **(d)** are the same as **(a)** and **(b)** but for the 8 to 14-day cases in the PNAS and EAAS, respectively

MSEs differ slightly at the PMZ developing stages: they are close to each other for days -1 through $+14$ (black, red, green, and blue curves). The mean MSE in the EAAS on day $+16$ reaches 0.8×10^4 GPM² per grid, which is roughly equivalent to that in the PNAS (Fig. 10a).

The MSEs on the onset day are shown as gray curves in Fig. 10a, c. Compared with the mean MSE (red), the MSEs among the cases exhibit large differences after day $+4$ with a range from 0.4 to 1.5×10^4 GPM² per grid on day $+16$. To identify the sources of the errors, the MSEs on the onset day (red) are decomposed into two components: one for the mean-squared error [first term on the right-hand side of Eq. (8); MSE_ens in black in Fig. 10b, and the other for the ensemble variance error [second term on the right-hand side of Eq. (8), MSE_spread in blue in Fig. 10b. The MSE_spread is overall smaller than the MSE_ens from day $+3$ to day $+16$ in both sectors (Fig. 10b, d). For the 8 to 14-day cases in the PNAS (Fig. 11a), the mean MSEs are close to each other at different PMZ development stages (black, red, green, and blue curves). For the random error and model bias in the 8 to 14-day cases, the MSE_ens is also larger than the MSE_spread in both the PNAS (Fig. 11b) and the EAAS (Fig. 11d). This suggests that the model bias is a dominant source of the forecasting errors in the GEFS. Meanwhile, the errors grow similarly

with an increase in the lead time when forecasting both the PMZ onset and the PMZ duration.

The ACC is a classic metric for quantifying the deterministic Z500 forecast skill. An ACC greater than 0.6 generally indicates a useful forecast with properly placed troughs and ridges at Z500 (Krishnamurti et al. 2003). The ACCs of the Z500 eddy fields for the 15-day PMZ cases in the PNAS with the forecast lead time are shown in Fig. 12a. The mean ACC of the ensemble mean for the PMZ onset day (represented by the red curve) is close to 1.0 from days 0 to $+2$ and decreases notably from days $+3$ to $+16$. This decrease is inherently associated with an increase in the MSE (cf. Fig. 10). The predictability of the PMZ onset is 8.5 days when an ACC of 0.6 is used as the threshold for a useful skill. The ACCs for the predictions starting on day -1 (black curve) show an evolution similar to that on the onset day, and the predictability is close to 9 days. However, the predictabilities during the development stages of PMZs (green and blue curves) are notably extended. The skill is extended to 10 days for the forecasts initialized on day $+14$ after PMZ onset (blue curve). It is worth noting that the capture of individual PMZ events is case dependent, especially after lead day $+5$. Similar to the MSEs, the uncertainties in the ACCs on the onset day (gray curves) increase with the lead time, and the ACCs for individual cases vary substantially from 0.8 to less than 0.2 on day $+16$.

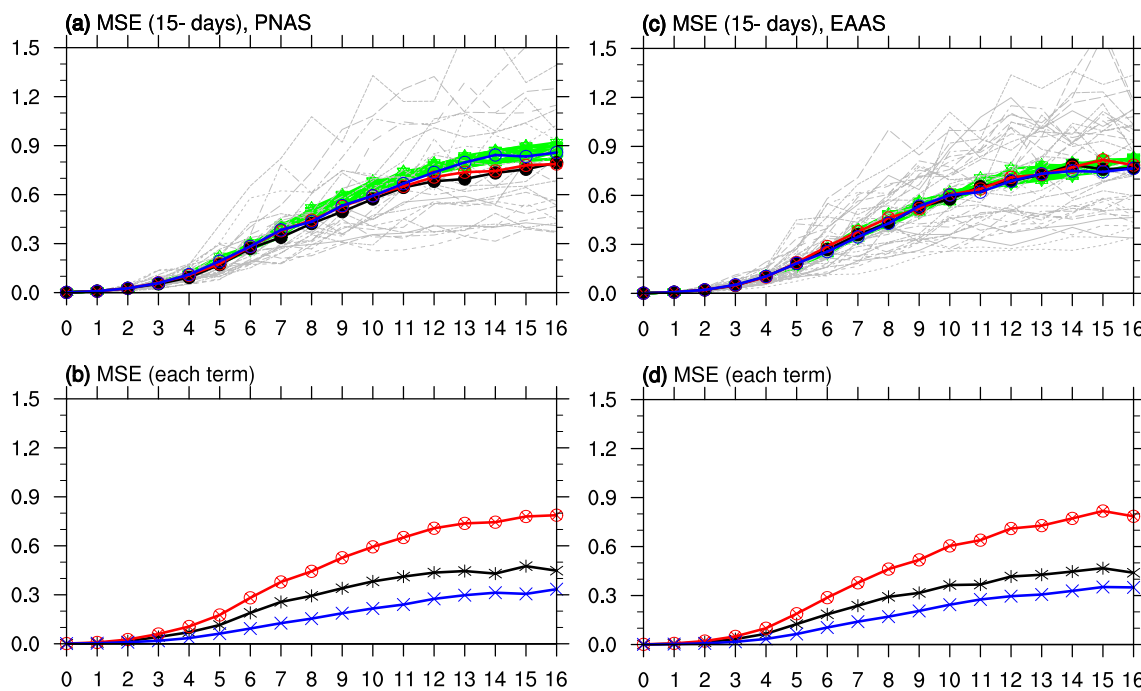


Fig. 10 **a** Averaged MSEs for the Z500 eddy anomalies ($\times 10^4$ GPM² per grid) during the 15-day PMZs in the PNAS in the GEFS ensemble mean. The red, black, blue and green curves denote the MSEs for the onset day (day 0), one day before onset (day -1), day +14 after onset, and days +1 through +13, respectively. **b** Each term of the

MSE for the onset day: the red curve is the same as in **(a)**, and the black and blue curves denote the ensemble mean and spread, respectively. **(c)** and **(d)** are the same as **(a)** and **(b)** but for the 15-days cases in the EAAS

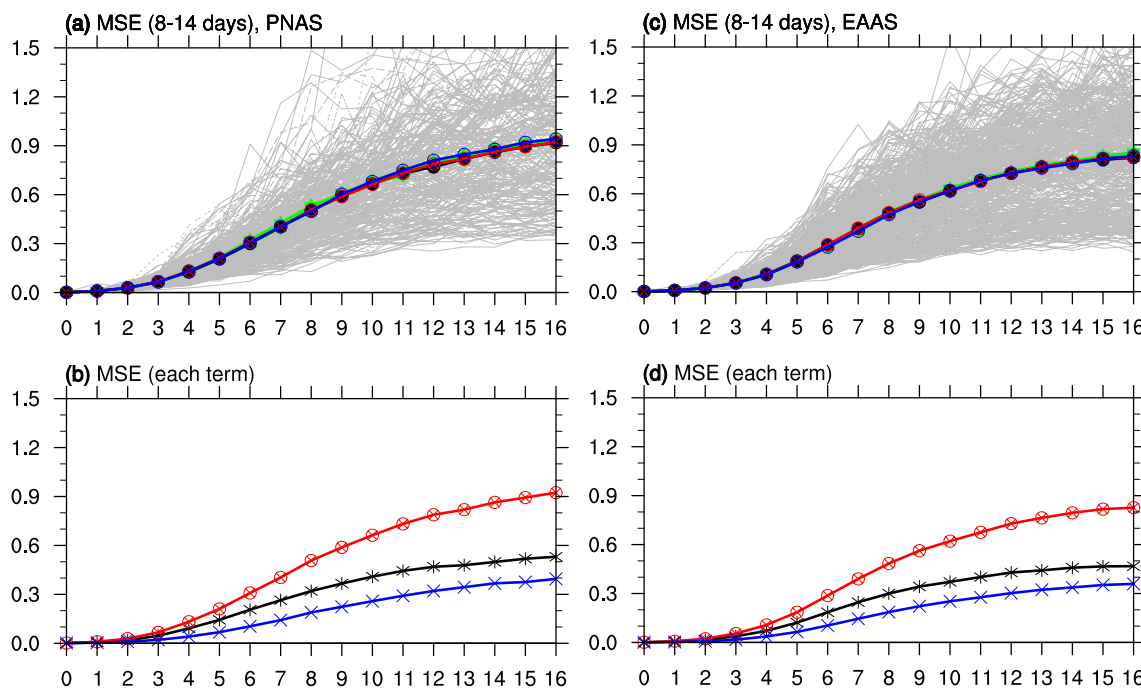


Fig. 11 Same as Fig. 10 but for the 8 to 14-day cases in the PNAS **(a, b)** and EAAS **(c, d)**. The blue line represents the MSEs for day +7 after onset

Next, the ACC skill conditioned with PMZs is compared with the averaged skill of Z500 eddy fields in the Northern Hemisphere to investigate the possible improvement attributable to the persistence of PMZs. The results are shown in Fig. 12b, in which the solid curves are the same as those in Fig. 12a, representing the ACCs in the PNAS for the day before PMZ onset (day -1 , black), the onset day (day 0 , red), and day $+14$ after onset (day $+14$, blue); moreover, the dashed curves use the same samples, but the calculated region extends to the Northern Hemisphere. In addition, the averaged ACC is computed from 1 January 1985 to 31 December 2015 (green dashed curve). The ACCs in the Northern Hemisphere are overall better than those in the PNAS by a half day. Meanwhile, 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 approximately one-half of a day shorter than the total (green). In contrast, the ACC skill for day $+14$ after PMZ onset (blue) is nearly 1.5 days better than that for the average, indicating that the PMZ persistence effectively extends the predictability of subseasonal signals. For the 15-day cases in the EAAS (Fig. 12c), the ACCs at different PMZ development stages are overall similar to those in the PNAS. The ACC on the onset day in the EAAS has a useful skill on lead day $+9$ (ACC > 0.6) that is approximately one-half of a day better than that in

the PNAS. Moreover, the ACC of day $+14$ after PMZ onset (blue) is close to that on the onset day (red) with an ACC of greater than 0.6, and it extends to day $+15$ in the EAAS with an ACC of greater than 0.5, which is approximately 3 days better than that in the PNAS at the same threshold. For the 15-day cases in the EAAS (Fig. 12d), the skills of days -1 , 0 , and $+14$ (black, red, and blue dashed curves, respectively) are all higher than the averaged skills for the Northern Hemisphere (green dashed lines), and they are over one-half of a day better than those in either sector.

The ACC skills for the 8 to 14-day PMZ cases are shown in Fig. 13. The mean ACCs at different PMZ development stages are close to each other in both sectors. The ACCs of days -1 , 0 , and $+7$ are overlaid with lead day $+9.5$ in the PNAS (Fig. 13a). In the EAAS (Fig. 13c), the ACCs of days -1 and 0 are overlaid with lead day $+9$, and the ACC of day $+7$ reaches that of lead day $+9.5$. We also compare the ACCs for the 8 to 14-day cases with the averaged ACC for all days from 1985 to 2015. In the PNAS (Fig. 13b), the ACCs in the Northern Hemisphere (dashed lines) are very close to the regional values (solid curves) for all PMZ development stages (days -1 , 0 , and $+7$). The skill reaches 9.5 days, which is slightly better than that of the total days (green dashed). In the EAAS (Fig. 13d), the skills of days -1 and 0 (black and red dashed lines, respectively) in the

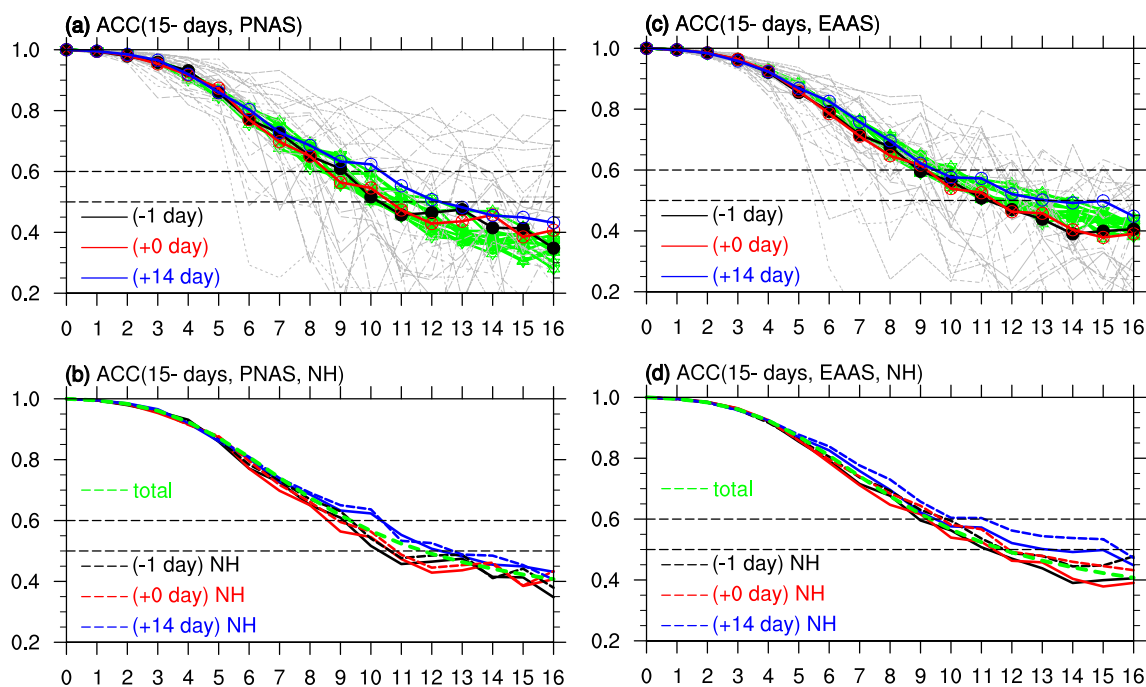


Fig. 12 **a** Averaged ACCs for the GEFS ensemble mean Z500 eddy anomalies of the 15-day cases in Appendix Table 2. The abscissa denotes the lead time (days). The black dashed lines represent ACCs of 0.6 and 0.5. The red, black, blue, green and gray curves denote the ACCs for the onset day (day $+0$), one day before the onset (day -1), day $+14$ after the onset, days $+1$ through $+13$, and the onset days for

the 26 individual cases, respectively. **b** The solid lines are the same as in (a), and the dashed lines denote the ACCs for the Northern Hemisphere. The green dashed line represents the averaged ACC 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 the EAAS

Northern Hemisphere are overall similar to the regional values (black and red solid lines), and the ACC skill of day + 7 for the Northern Hemisphere extends to lead day + 10. These results indicate that the GEFS has a better ACC score in the PNAS than in the EAAS for PMZ events that persist for longer than 1 week. The ACC skill in predicting the PMZ onset date is still lower than that in predicting the PMZ development, which is similar to that in predicting blockings (TM90).

The above verifications of the POD, MSE, and ACC indicate that the skills vary more dramatically in predicting PMZ events than in predicting different development stages of individual PMZs. We next quantify the uncertainty in the GEFS in predicting PMZ onsets by sorting the ACC scores. The ACC scores of the onset days for all four PMZ groups are shown in the corresponding boxplots (Fig. 14), in which the black dots denote the mean values for all cases, the upper and lower boundaries denote the upper and lower quartiles, and the horizontal line represents the median value. The horizontal line on the top of a box represents the maximum value, and the horizontal line on the bottom of a box stands for the minimum value. For PMZ events that persist for longer than 15 days in the PNAS (Fig. 14a), the prediction on lead day + 1 is the best. All predicted cases show consistently high ACC values, and the boxes become a horizontal line. The uncertainty increases with the lead time, and the boxes expand in the vertical direction. The ACCs are lower than 0.5 in less than 25% of the cases on lead day + 6

and above 0.6 until lead day + 9 in more than half of the cases, indicating that the prediction skill is approximately 9 days in the PNAS for PMZs that persist for longer than 2 weeks. From lead days + 12 to + 16, the ACCs exceed 0.6 for approximately 25% of the cases, and they are lower than -0.3 for another 25% of the cases. For the 15-day cases in the EAAS (Fig. 14b), the ACC skill extends to 11 days for more than half of the cases with a skill above 0.6. Similarly, the useful skill reaches 10 days for the 8 to 14-day cases in both the PNAS (Fig. 14c) and the EAAS (Fig. 14d).

It is worth noting that useful cases (ACCs above 0.6) still constitute approximately 25% of the total on lead days longer than 10 in all four groups. Analysis of the common features in these cases can potentially improve the model's prediction during weeks 2 through 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 day and day + 15 on lead days + 10 through + 16 (Appendix Tables 4, 5). The ACCs for these cases represent the prediction skills of the PMZ onset and duration. For the onset day (Appendix Table 4), four cases are selected with three in the PNAS (26 February 2011, 10 January 2013, and 15 May 2015) and one in the EAAS (24 November 2012). The first two cases are strong with an intensity above 300 GPMs. For the PMZ duration (Appendix Table 5), four cases are also selected with three in the PNAS (18 December 1993, 13 December 1999, and 19 February 2005) and one in the EAAS (24 February 1986). The eight cases are temporally different, suggesting that

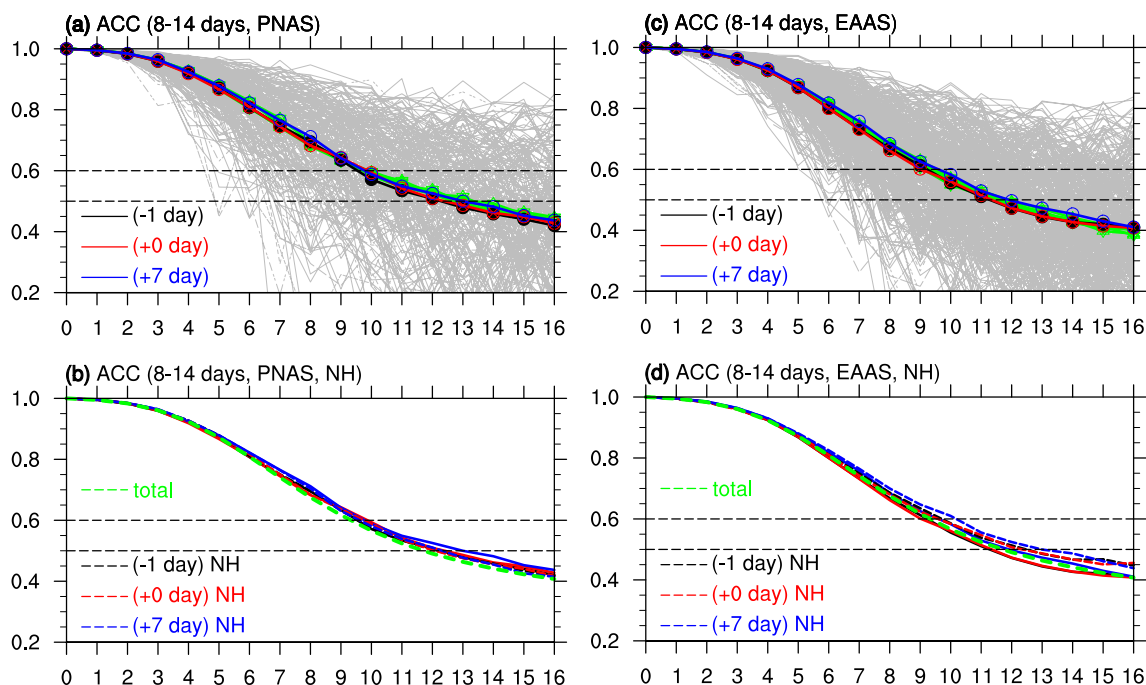


Fig. 13 Same as Fig. 12 but for the 8 to 14-day cases in Appendix Table 2 for the PNAS (a, b) and EAAS (c, d). The blue line denotes the ACCs for day + 7 after onset

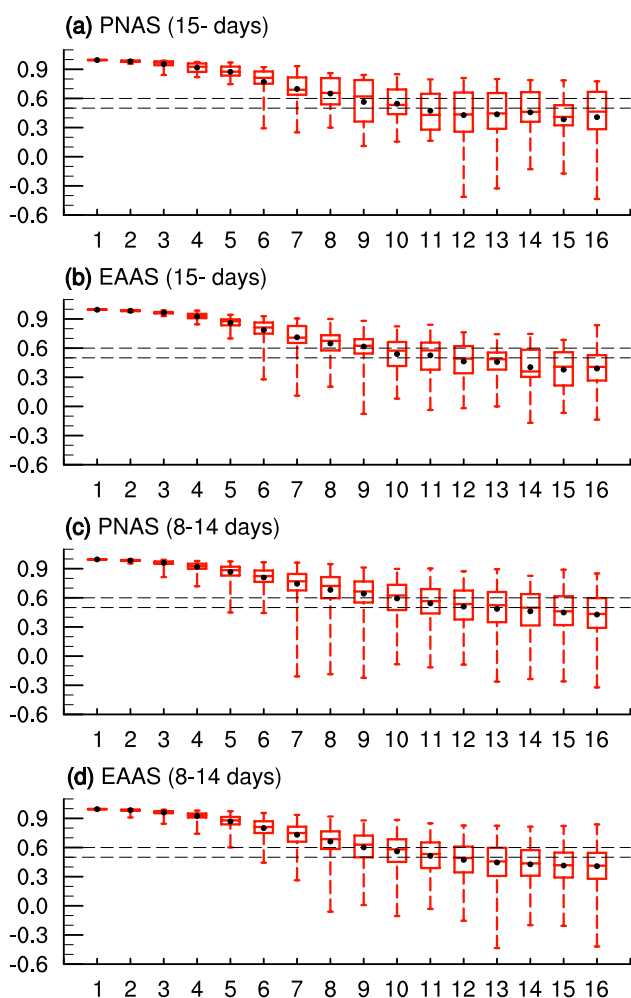


Fig. 14 **a** Box plots of the ACCs on the onset days of the 15-day PMZ cases in the PNAS (Appendix Table 3) with the lead time in the GEFS ensemble mean. The black dashed lines denote ACCs of 0.6 and 0.5. The black dots represent the mean values of the individual cases. **b** The same as (a) but for the 15-day cases in the EAAS. **c** The same as (a) but for the 8 to 14-day cases in the PNAS. **d** The same as (a) but for the 8 to 14-day cases in the EAAS

the ACC skills of the PMZ onset and duration are different. The prediction of the duration of long-lived PMZs is generally poor even when the onset is predicted well; cases that are predicted with relatively higher skill all occur in the wintertime. Although a total number of eight cases appears to be rather small, the long duration notably improves the predictions.

4 Summary and discussion

The daily forecast datasets of GEFS v10 from 1 January 1985 to 31 December 2015 are used in this study to evaluate the skills in predicting persistent Z500 patterns. The

persistent maxima of Z500 eddies (PMZ; Liu et al. 2017) in the PNAS and EAAS are tracked, and their impact areas serve as the targets for verification, because PMZs generally include persistent open ridges, omega-shape blockings, and closed blocking anticyclones. The skills in both probabilistic and deterministic predictions are evaluated, and the main results are summarized and discussed below.

- (1) The PMZ frequency is underestimated at longer lead times in both the PNAS and the EAAS. The predicted frequency on lead day +16 is underestimated by 30% in the winter and 90% in the summer. The BSS and reliability skill are both higher in the PNAS than in the EAAS, which is contrary to the skills of blockings (Figs. 3, 5 in Hamill and Kiladis 2014), partly because much more and shorter-lived PMZ events than traditional blockings are identified and evaluated here. In addition, the skills are overall better in predicting the PMZ duration than predicting the PMZ onset, which is similar to the prediction of blockings (TM90; Hamill and Kiladis 2014).
- (2) The predictabilities of the PMZ onset and PMZ duration differ with regard to the ACC skills, especially for PMZs with a duration of longer than 14 days. The onset has a useful skill reaching up to 8.5 days in the PNAS and 9 days in the EAAS, while the duration has a useful skill reaching up to 10 days in the PNAS and 9.5 days in the EAAS. The most uncertainty remains in predicting different PMZ cases. Half of the PMZ cases are predictable with an ACC skill exceeding 0.6 on lead day +9 through +10 in both sectors; approximately 25% cases are still predictable on lead days +10 through +16.
- (3) Compared with classical blocking patterns (Hamill and Kiladis 2014), PMZ events occur more frequently in the Northeast Pacific close to North America, and their frequencies are predictable with considerable skill by the GEFS v10. Notably, the predicted frequencies of PMZs decrease with the lead time, which is in contrast to the nearly invariant frequencies of predicted blockings, especially in the PNAS, reported by Hamill and Kiladis (2014, their Fig. 2) using the same GEFS forecast datasets. The decrease in the PMZ frequency with the lead time, however, agrees with the predicted blockings (identified by the same index as TM90) reported in the ECMWF models (TM90, their Fig. 2; Mauritsen and Källén 2004, their Fig. 6) and another NCEP model (Jia et al. 2014, their Fig. 1). Understanding these differences merits further study.

The largest difference between our study and previous works is that we use Z500 eddies to identify persistent anticyclone anomalies, including both open and closed

blockings. At 500-hPa, when the blocking is weak or appears as an open ridge, it is difficult to design an algorithm to isolate the system from the absolute field. Thus, the departure field from the mean flow becomes convenient for identifying such persistent high pressure systems. Thus, daily departures from the zonal mean field could correctly place ridges and troughs in the absolute field, as we used Z500 eddies in this study. Therefore, the eddy field could constitute an objective indicator as an alternative to time anomalies for modeling evaluations and forecasting verifications.

Improving the forecast skills of the GEFS at 2–4 weeks is a challenging task, and PMZs are one of the sources for potential improvement. Evaluations of GEFS v10 in predicting PMZ events have revealed some intriguing topics for future investigation. For instance, the prediction skill is more sensitive to PMZ cases (with an uncertainty of over 10 days) than to individual PMZ stages (with an uncertainty of approximately 2 days). The new NGGPS will probably reduce this uncertainty by reducing model biases, which appear to contribute largely to the overall uncertainty. In addition, errors grow similarly in the ensemble mean and ensemble spread, indicating an equivalent importance in improving the initial conditions and model physics. Understanding such sources of uncertainties will help develop and evaluate the NGGPS in the near future.

Acknowledgements We thank the three anonymous reviewers for their comments. B. He, W. Hu and P. Liu were supported by the National Weather Service under the Grants NA15NWS4680015 and NA15OAR4320064. B. He and W. T. Hu were partially supported by the National Key Basic Research Program of China (Grant 2014CB953904) and the National Natural Science Foundation of China under the Grants 41530426, 91737306, 41730963.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

See Tables 2, 3, 4 and 5.

Table 2 Onset dates, durations, central locations, and intensities of the PMZs with durations greater than 15 days over the 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 3 The same as Table 2 but over the 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 4 PMZ events with ACCs exceeding 0.6 on the onset day 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 5 The same as Table 4 but for day + 15 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

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. <https://doi.org/10.1007/200382-011-124306>
- Barriopedro D, Garcia-Herrera R, Gonzalez-Rouco JF, 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. *Q J Roy Meteorol Soc* 116:867–912
- Chen Y, Zhai P (2014) Precursor circulation features for persistent extreme precipitation in central-eastern China. *Weather Forecast* 29:226–240. <https://doi.org/10.1175/WAF-D-13-00065.1>
- Cheng X, Wallace JM (1993) Cluster analysis of the Northern Hemisphere wintertime 500-hPa height field: spatial patterns. *J Atmos Sci* 50:2674–2696
- Dole RM, Gordon ND (1983) Persistent anomalies of the extratropical Northern Hemisphere wintertime circulation: geographical distribution and regional persistence characteristics. *Mon Weather 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, et al (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. <https://doi.org/10.1007/s00382-015-2638-6>
- Hamil TM, Bates GT, Whitaker JS, Murray DR, Fiorino M, Galarneau TJ Jr, Zhu YJ, Lapenta W (2013) NOAA's second-generation global medium-range ensemble reforecast data set. *Bull Am Meteorol Soc* 94:1553–1565. <https://doi.org/10.1175/BAMS-D-12-00014.1>
- Hamil TM, Juras J (2006) Measuring forecast skill: is it real skill or is it the varying climatology? *Q J R Meteorol Soc* 132:2905–2923
- Hamil TM, Kiladis GN (2014) Skill of the MJO and Northern Hemisphere blocking in GEFS medium-range reforecasts. *Mon Weather Rev* 142:868–885
- Hamil TM, Whitaker JS (2007) Ensemble calibration of 500-hPa geopotential height and 850-hPa and 2-m temperatures using reforecasts. *Mon Weather 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. In: *Proceedings of 19th conference on probability and statistics*, New Orleans, LA, American Meteor Society. <https://ams.confex.com/ams/88Annual/webprogram/Paper134165.html>

- IPCC Climate Change (2013) Intergovernmental panel on climate change (IPCC), climate change 2007. Stocker TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds), The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, pp 3–32
- Jia XL, Yang S, Song W, He B (2014) Prediction of wintertime Northern Hemisphere blocking by the NCEP Climate Forecast System. *J Meteor Res* 28(1):76–90. <https://doi.org/10.1007/s13351-014-3085-8>
- 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, Cocks S, Ahlquist JE (2003) Improved skill for the anomaly correlation of geopotential heights at 500 hPa. *Mon Weather Rev* 131:1082–1102
- 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. <https://doi.org/10.1029/2008GL035205>
- Lejenäs H, Økland H (1983) Characteristics of Northern Hemisphere blocking as determined from a long time series of observational data. *Tellus* 35A:350–362
- Liu P et al (2017) Climatology of tracked persistent maxima of 500-hPa geopotential height. *Clim Dyn*. <https://doi.org/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
- 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, Petrolia T (1996) The ECMWF ensemble prediction system: methodology and validation. *Q J R Meteorol 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 Weather Rev* 112:1894–1912
- Renwick JA, Revell MJ (1999) Blocking over the South Pacific and Rossby wave propagation. *Mon Weather Rev* 127:2233–2247
- 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 Am Meteorol Soc* 91:1015–1057. <https://doi.org/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
- Sinclair MR (1996) A climatology of anticyclones and blocking for the Southern Hemisphere. *Mon Weather Rev* 124:245–263
- Tibaldi S, Molteni F (1990) On the operational predictability of blocking. *Tellus* 42:343–365
- Trenberth KE, Mo KC (1985) Blocking in the Southern Hemisphere. *Mon Weather Rev* 113:3–21
- Watson JS, Colucci SJ (2002) Evaluation of ensemble predictions of blocking in the NCEP global spectral model. *Mon Weather 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. <https://doi.org/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. 2 edn. Academic Press, Cambridge
- Zhou X, Zhu Y, Hou D, Luo Y, Peng J, Wobus D (2016) The NCEP global ensemble forecast system with the EnKF initialization. *Weather Forecast* (in press)
- Zhu Y (2005) Ensemble forecast: a new approach to uncertainty and predictability. *Adv Atmos Sci* 22:781–788
- Zhu Y, Wobus R, Wei M, Cui B, Toth Z (2007) March 2007 NAEFS upgrade. NOAA/NCEP/EMC, accessed 9 June 2015. [Available online at http://www.emc.ncep.noaa.gov/gmb/ens/ens_imp_news.html.]
- Zhu Y, Toth Z (2008) Ensemble based probabilistic verification. In: Preprints, 19th conf. on predictability and statistics, New Orleans, LA, American Meteorological Society 2.2. <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. *Weather Forecast* 32(6):2159–2174
- Zhu Y, Zhou X, Li W, Hou D, Melhauser C, Sinsky E, Pena M, Fu B, Guan H, Kolczynski W, Wobus R, Tallapragada V (2017b) An assessment of subseasonal forecast skill using an extended global ensemble forecast system (GEFS). *J Clim* (under review)