A New Forecast Model Based on the Analog Method for Persistent Extreme Precipitation

BAIQUAN ZHOU AND PANMAO ZHAI
State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, China Meteorological Administration, Beijing, China

(Manuscript received 11 December 2015, in final form 25 May 2016)

ABSTRACT

This study aims to establish an analog prediction model for forecasting daily persistent extreme precipitation (PEP) during a PEP event (PEPE) using the temporal sequences of predictors with different weights applied in the atmospheric spatial field. The predictors are atmospheric variables in areas where the key influential systems of a PEPE are active in the THORPEX Interactive Grand Global Ensemble (TIGGE) dataset. By means of the cosine similarity measure and the cuckoo search technique, a forecast model was established and named the Key Influential Systems Based Analog Model (KISAM). Validations through threat scores (TSs) and root-mean-square errors for PEP during 17–25 June 2010 indicate that KISAM is able to identify the approaching PEP earlier and yield a more accurate forecast for the location and intensity of PEP than direct model output (DMO) at 3-day and longer lead times in the Yangtze–Huai River valley. For the independent PEPE case on 17–19 June 2010, KISAM is able to predict the PEPE about 8 days in advance. That is much earlier than with DMO. In addition, KISAM produces better intensity forecasts and predicts the extent of the PEPE better than DMO at the same lead time of 5 days. In terms of the forecast experiments during June 2010 and 2015, KISAM shows relatively stronger capacity than DMO in predicting the occurrence and intensity of extreme precipitation (EP) and PEP events at lead times of 1 week or even longer. Through validation of more EP, better performance of KISAM compared to DMO on average is further confirmed at 3-day and longer lead times.

1. Introduction

Persistent extreme precipitation events (PEPEs) in summer are characterized by long duration and high intensity, which often result in catastrophic consequences such as flooding and landslides. Therefore, there is an urgent need for developing several skilled forecasting approaches to enhance the forecast capacity for daily extreme precipitation (EP) during a PEPE. The definition of regional PEPE introduced by Chen and Zhai (2013) is that the daily precipitation amount must exceed 50 mm for at least three consecutive days in a region. The event comes to an end when the daily precipitation amount is <50 mm for the following two consecutive days. The extreme precipitation is then allowed to break for at most 1 day before continuing. The daily occurrences of EP in a row composing a PEPE can be called persistent extreme precipitation (PEP); otherwise they are ordinary EP. However, it is a real challenge to extend the forecasting valid time of extreme weather events to 1 day–2 weeks, which is of significant value for governmental decision-making. Such a challenge is consistent with the objective of The Observing System Research and Predictability Experiment (THORPEX) initiated by the WMO to accelerate improvements in the accuracy of 1-day–2-week high-impact weather forecasts (Bougeault et al. 2010; Zhai et al. 2013). The THORPEX Interactive Grand Global Ensemble (TIGGE) collects 1-day–2-week ensemble prediction data from the leading global numerical weather prediction (NWP) centers, the resolution of which can reach 0.5° × 0.5°. Nevertheless, because of the distinct discontinuity both in time and space for precipitation distribution, the resolution of these NWP models cannot satisfy the demand of PEP forecasters, as these events often occur in small local regions (Krishnamurti et al. 2009; Zhi et al. 2013). Among the several evaluated NWP models in

Corresponding author address: Panmao Zhai, Chinese Academy of Meteorological Sciences, No. 46 Zhong-guan-cun South Street, Haidian District, Beijing 100081, China.
E-mail: pmzhai@cma.gov.cn

DOI: 10.1175/WAF-D-15-0174.1

© 2016 American Meteorological Society
TIGGE, all tend to underforecast heavier precipitation exceeding 25 mm day$^{-1}$ (Su et al. 2014). Compared with precipitation, TIGGE models appear to show better performance in the forecasting of large-scale variables and key synoptic systems (Park et al. 2008; Pelly and Hoskins 2003; Froude 2010; Niu and Zhai 2013; Niu et al. 2015; Zhou et al. 2015). Considering the above aspects, using a statistical approach to downscale the more precisely predicted large-scale variables or key synoptic systems in TIGGE to obtain reliable small-scale information could be an appropriate choice to improve the forecast accuracy of PEP.

The analog method (AM), the simplest statistical downscaling technique available, can establish non-linear relationships between variables straightforwardly (Fernández and Sáenz 2003). AM has two main advantages when predicting precipitation: it is favorable for hydrological studies on account of using the observed weather patterns and retaining the spatial covariance structure of the local-scale weather and, when constructing forecast scenarios, AM does not need to assume the form of the probability distribution of the downscaled variables; thus, it is easy for downscaling nonnormally distributed variables such as daily precipitation (Matulla et al. 2008). AM was first used in the field of weather forecasting (Lorenz 1969; Kruizinga and Murphy 1983) and short-term climate prediction applications (Livezey and Barnston 1988; Van den Dool 1989, 1994). Analogs were searched for in the past history of the predictand variable, and the forecast was achieved through the similar evolution of the present weather or climate situation and the past analog. Forecasts produced by AM are based on the assumption that when large-scale meteorological conditions are analogous, the corresponding observed local atmospheric variables such as precipitation or temperature should also be similar. Zorita et al. (1995) and Zorita and Von Storch (1999) initiated the use of AM in their downscaling scheme. Zorita and Von Storch (1999) compared AM with other more complicated statistical downscaling techniques and found that AM performed generally as well as other more complicated methods. Based on the research that tested the canonical correlation analysis (CCA) approach for searching analogs (Zorita et al. 1992; Von Storch et al. 1993; Ulbrich et al. 1999), Fernández and Sáenz (2003) presented an improvement of a new AM by finding the analogs in the space of the CCA temporal expansion coefficients to the commonly used principal component analysis approach. The improvement seemed to be that the CCA identified the areas in the predictor field most connected to the predictand. Matulla et al. (2008) made an advance when including large-scale information of bygone days in an analog matching scheme, particularly for the simulation of wet and dry spells when examining the performance of several downscaling schemes. Wang and Fan (2009) proposed a prediction scheme considering both dynamical model prediction and the observed spatial pattern of historical analog years to improve the summer rainfall over East Asia. An analog forecast system was developed by the Hong Kong Observatory (HKO) with special focus put on the forecasting of heavy rainfall events that occurred in the vicinity of Hong Kong. The system demonstrated superior performance when compared with both direct model output and the subjective forecasts issued by HKO (Chan et al. 2014). Even though studies of AM have been abundant, they have usually concentrated on downscaling large-scale information outputted by general circulation models instead of NWP models in TIGGE. Other new statistical forecasting schemes are also emerging (Fan et al. 2008, 2012; Harpham and Wilby 2005; Hu et al. 2013). However, both the studies of AM and these new statistical forecast schemes seldom get involved in the application of PEP. The present study aims to propose a new downscaling model (DM) based on AM to forecast PEP. In the model, we develop a new similarity measure with prominence given to the large-scale weather systems highly correlated to PEPs. When matching analogs, the large-scale information from bygone days before the occurrence of PEPs is also included. It is hoped that the accuracy of the PEP forecasts will be improved by using the large-scale weather patterns outputted by those NWP models in TIGGE and that the results will outperform the direct outputs of PEP forecasts by the NWP models.

This paper continues with a description of the study region and the data analyzed in section 2, followed by the introduction of the methodology adopted to establish the forecast model in section 3. The validation of the DMs during the calibration period is exhibited in section 4. In section 5, evaluation and comparison of the forecast performance of the established forecast model and direct NWP model output are presented, and the paper finishes with our conclusions and discussion in section 6.

2. Study region and data

The Yangtze–Huai River valley (26°–34°N, 112°–121°E) is chosen as the study region where the PEPs are prone to occur, as shown in Fig. 1a. According to the definition of regional PEPs, Chen and Zhai (2013) identified 25 PEPs in the selected area with details listed in Table 1. We choose the first 24 events for training and left the last one for independent verification. The 24 events took
place over 22 yr with a total duration of 144 days. Those 144 days are treated as “event days” for the daily AM model, and the other 1880 days in those 22 summers (1 June–31 August) are considered to be “nonevent days.” Event days refer to days when a PEPE was observed; on the other hand, nonevent days refer to days when no PEP occurred. The average precipitation amount of the 144 event days is presented in Fig. 1b. It reaches as high as 150 mm, with the heaviest precipitation located in the eastern part of the study region.

**Table 1.** PEPEs in central–eastern China (26°–34°N, 112°–121°E) between 1951 and 2010. The detailed information for each case includes the start and end dates, the duration in unit of days, the number of affected stations, the affected area (10^4 km^2), the boundaries in latitude (°N) and longitude (°E), and the maximum and minimum precipitation amounts (mm) observed by affected stations over the duration of the event.

<table>
<thead>
<tr>
<th>Year</th>
<th>Start date</th>
<th>End date</th>
<th>Duration (days)</th>
<th>No. of affected stations</th>
<th>Affected area (10^4 km^2)</th>
<th>North (°)</th>
<th>South (°)</th>
<th>West (°)</th>
<th>East (°)</th>
<th>Max precipitation (mm)</th>
<th>Min precipitation (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1954</td>
<td>4 Jul</td>
<td>7 Jul</td>
<td>4</td>
<td>3</td>
<td>3.23</td>
<td>32.55</td>
<td>32.1</td>
<td>115.37</td>
<td>117.23</td>
<td>430.2</td>
<td>265.4</td>
</tr>
<tr>
<td>1955</td>
<td>18 Jun 23 Jun</td>
<td>6</td>
<td>6</td>
<td>5.7</td>
<td>29.44</td>
<td>28.41</td>
<td>115.59</td>
<td>119.39</td>
<td>516.8</td>
<td>172</td>
<td>265.4</td>
</tr>
<tr>
<td>1961</td>
<td>7 Jun 11 Jun</td>
<td>5</td>
<td>3</td>
<td>2.59</td>
<td>31.26</td>
<td>28.18</td>
<td>117.13</td>
<td>119.29</td>
<td>264.9</td>
<td>172</td>
<td>265.4</td>
</tr>
<tr>
<td>1964</td>
<td>24 Jun 29 Jun</td>
<td>6</td>
<td>5</td>
<td>5.29</td>
<td>30.44</td>
<td>29.24</td>
<td>110.1</td>
<td>115.4</td>
<td>574.9</td>
<td>197.6</td>
<td>262.2</td>
</tr>
<tr>
<td>1967</td>
<td>17 Jun 22 Jun</td>
<td>6</td>
<td>3</td>
<td>2.59</td>
<td>29</td>
<td>27.03</td>
<td>114.55</td>
<td>118.54</td>
<td>358.9</td>
<td>176.5</td>
<td>305.5</td>
</tr>
<tr>
<td>1968</td>
<td>16 Jun 19 Jun</td>
<td>4</td>
<td>3</td>
<td>1.82</td>
<td>27.03</td>
<td>26.39</td>
<td>118.1</td>
<td>118.59</td>
<td>424.5</td>
<td>303.9</td>
<td>305.5</td>
</tr>
<tr>
<td>1968</td>
<td>13 Jul 20 Jul</td>
<td>8</td>
<td>5</td>
<td>4.84</td>
<td>33.36</td>
<td>30.4</td>
<td>113.1</td>
<td>119.02</td>
<td>565.9</td>
<td>238.7</td>
<td>206.8</td>
</tr>
<tr>
<td>1970</td>
<td>8 Jul 14 Jul</td>
<td>7</td>
<td>4</td>
<td>3.47</td>
<td>30.08</td>
<td>27.55</td>
<td>115.59</td>
<td>118.32</td>
<td>295.6</td>
<td>177.6</td>
<td>238.7</td>
</tr>
<tr>
<td>1974</td>
<td>14 Jul 17 Jul</td>
<td>4</td>
<td>3</td>
<td>2.59</td>
<td>30.08</td>
<td>29.18</td>
<td>117.12</td>
<td>118.17</td>
<td>279.5</td>
<td>238.7</td>
<td>177.6</td>
</tr>
<tr>
<td>1982</td>
<td>13 Jun 19 Jun</td>
<td>7</td>
<td>9</td>
<td>9.16</td>
<td>28.04</td>
<td>27.03</td>
<td>111.28</td>
<td>118.32</td>
<td>551.5</td>
<td>240.8</td>
<td>302.8</td>
</tr>
<tr>
<td>1989</td>
<td>29 Jun 3 Jul</td>
<td>5</td>
<td>4</td>
<td>3.8</td>
<td>29</td>
<td>27.48</td>
<td>114.23</td>
<td>118.54</td>
<td>379.3</td>
<td>231.6</td>
<td>302.8</td>
</tr>
<tr>
<td>1991</td>
<td>12 Jun 15 Jun</td>
<td>4</td>
<td>4</td>
<td>3.8</td>
<td>32.33</td>
<td>31.53</td>
<td>115.37</td>
<td>120.53</td>
<td>370.4</td>
<td>231.6</td>
<td>302.8</td>
</tr>
<tr>
<td>1991</td>
<td>1 Jul 11 Jul</td>
<td>11</td>
<td>9</td>
<td>9.05</td>
<td>32.52</td>
<td>30.21</td>
<td>112.09</td>
<td>120.19</td>
<td>742.2</td>
<td>236.9</td>
<td>302.8</td>
</tr>
<tr>
<td>1992</td>
<td>4 Jul 8 Jul</td>
<td>5</td>
<td>4</td>
<td>3.47</td>
<td>28.04</td>
<td>25.31</td>
<td>117.28</td>
<td>119.47</td>
<td>447.9</td>
<td>215.3</td>
<td>236.9</td>
</tr>
<tr>
<td>1995</td>
<td>21 Jun 26 Jun</td>
<td>6</td>
<td>3</td>
<td>2.07</td>
<td>30.08</td>
<td>28.41</td>
<td>118.09</td>
<td>118.54</td>
<td>469</td>
<td>286</td>
<td>215.3</td>
</tr>
<tr>
<td>1996</td>
<td>29 Jun 2 Jul</td>
<td>4</td>
<td>4</td>
<td>2.77</td>
<td>30.21</td>
<td>29.43</td>
<td>118.09</td>
<td>120.1</td>
<td>619.9</td>
<td>291.2</td>
<td>286</td>
</tr>
<tr>
<td>1997</td>
<td>7 Jul 12 Jul</td>
<td>6</td>
<td>5</td>
<td>4.66</td>
<td>30.44</td>
<td>26.51</td>
<td>116.2</td>
<td>122.27</td>
<td>389.1</td>
<td>275.1</td>
<td>286</td>
</tr>
<tr>
<td>1998</td>
<td>12 Jun 27 Jun</td>
<td>16</td>
<td>12</td>
<td>10.78</td>
<td>30.37</td>
<td>23.48</td>
<td>113.32</td>
<td>118.59</td>
<td>1053.9</td>
<td>283.6</td>
<td>286</td>
</tr>
<tr>
<td>1999</td>
<td>24 Jun 1 Jul</td>
<td>8</td>
<td>7</td>
<td>4.84</td>
<td>31.09</td>
<td>29.37</td>
<td>113.55</td>
<td>120.1</td>
<td>813.5</td>
<td>313.8</td>
<td>286</td>
</tr>
<tr>
<td>2000</td>
<td>9 Jun 12 Jun</td>
<td>4</td>
<td>6</td>
<td>4.68</td>
<td>28.04</td>
<td>25.31</td>
<td>118.02</td>
<td>120.12</td>
<td>376.5</td>
<td>204.2</td>
<td>313.8</td>
</tr>
<tr>
<td>2002</td>
<td>14 Jun 17 Jun</td>
<td>4</td>
<td>3</td>
<td>2.54</td>
<td>26.54</td>
<td>26.39</td>
<td>116.2</td>
<td>118.1</td>
<td>551.3</td>
<td>378.8</td>
<td>204.2</td>
</tr>
<tr>
<td>2003</td>
<td>8 Jul 10 Jul</td>
<td>3</td>
<td>5</td>
<td>5.57</td>
<td>31.11</td>
<td>28.5</td>
<td>108.46</td>
<td>115.01</td>
<td>481.7</td>
<td>197.2</td>
<td>378.8</td>
</tr>
<tr>
<td>2005</td>
<td>18 Jun 24 Jun</td>
<td>7</td>
<td>9</td>
<td>7.26</td>
<td>27.55</td>
<td>23.48</td>
<td>114.44</td>
<td>120.12</td>
<td>706.8</td>
<td>295.1</td>
<td>378.8</td>
</tr>
<tr>
<td>2010</td>
<td>17 Jun 25 Jun</td>
<td>9</td>
<td>6</td>
<td>5.08</td>
<td>27.55</td>
<td>26.54</td>
<td>116.39</td>
<td>118.32</td>
<td>754.4</td>
<td>441.1</td>
<td>295.1</td>
</tr>
</tbody>
</table>
The remaining PEPE in 2010, which lasted for 9 days, was used as the independent verification sample. The forecast model was operated in practice to yield forecasts in June 2015 to further investigate its skill.

When training the DM during the calibration period (144 event days), the daily reanalysis dataset provided by the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) was used to get the large-scale weather information. The data for the period from 1951 to 2009 were used with horizontal resolution of 2.5° × 2.5° (Kalnay et al. 1996), and the detailed variables in the specific area that we used to match for analogs are presented in the next section. When running independent forecasts, the predicted 2.5° × 2.5° large-scale information used is the model output of the European Centre for Medium-Range Weather Forecasts (ECMWF) at 1200 UTC for 1–15-day lead times during 2010 and 2015. To compare with the forecasts made by the DM, we also used the direct model output (DMO) of total precipitation from ECMWF at 1200 UTC for the same lead times at 1° × 1° resolution in 2010 and 2015. In addition, the daily precipitation amounts from 756 stations during the period of 1951–2010 used in the training and independent tests were obtained from the National Meteorological Information Center, China Meteorological Administration. Because precipitation data were often missing from several stations before 1961, for convenience of comparison with the total precipitation directly outputted by ECMWF, the station data were interpolated to an identical grid of 1° × 1° through the Cressman interpolation method. In total, there are 90 grid points in the selected study area on the 1° × 1° grid.

3. Methodology

a. Large-scale meteorological variables used

According to the study carried out by Chen and Zhai (2014), PEP in the study area is mainly influenced by several key systems such as the South Asian high and the high-level jet stream at 200 hPa, the western Pacific subtropical high (WPSH) and blocking high at higher latitudes at 500 hPa, and water vapor transport systems at 850 hPa. The long-lived blocking pattern at high latitudes continuously steers cold/dry air from the mid-high latitudes to central–eastern China. An anomalous anticyclone forms with a westward shift of the WPSH. At the lower and middle levels, warm/moist air from the lower latitudes is also anomalously conveyed to central–eastern China by intensified southeasterlies of the anomalous anticyclone. Cold/dry air encountering anomalous warm/moist air in the same region is the most important condition for the formation of PEP. In the upper troposphere, the displacement of the South Asian high and jets favors divergence for both typical circulation patterns. More importantly, they revealed that the concurrent combination of anomalies in the aforementioned key systems results in the occurrence and maintenance of PEP. Therefore, we should take account of all variables of these key systems in the troposphere when matching two large-scale atmospheric states in the AM scheme. Correspondingly, we selected the zonal wind at 200 hPa in the area of 25°–55°N and 70°–160°E and where the subtropical jet stream was usually active and the geopotential height at 500 hPa in the region of 0°–70°N and 30°E–180°, which contained areas where blocking highs and WPSH were active. We also chose the zonal and meridional water vapor transport at 700 hPa over 0°–35°N and 70°–160°E, which were the products of zonal and meridional wind with specific humidity. The selection of water vapor transport at 700 hPa instead of at 850 hPa contributes to the better performance of specific humidity at 700 hPa for downsampling daily precipitation during the summer (Cavazos and Hewitson 2002). Basically, the forecast skills of these atmospheric variables in TIGGE have been well investigated and validated. Zhou et al. (2015) assessed the predictability of the East Asian subtropical westerly jet based on TIGGE and drew the conclusion that ECMWF showed the highest level of skill. Park et al. (2008) found that ECMWF provided the best predictions of 500-hPa geopotential height. Pelly and Hoskins (2003) validated the skill of ECMWF when predicting blocking and found that the forecasts were shown to be skillful out to day 10 in terms of sector blocking. Niu and Zhai (2013) and Niu et al. (2015) verified the performance of four TIGGE models for predicting the WPSH and the low-level winds of the Asian summer monsoon. Based on the above studies, the lead times with forecast skill for the related variables in ECMWF used as predictors in this study can reach up to 10 ~ 12 days. Thus, these variables may reliably be taken into account in our forecast model up to a lead time of 10 ~ 12 days. Furthermore, the anomalies of these variables were used to match the large-scale atmospheric patterns, which were drawn from the differences between the values of the variables on the sample days and their respective climatological mean from 1951 to 2009 on the same day.

b. Analog search

1) SIMILARITY METRICS

One of the key aspects in searching past analogs is the similitude measure. The similarity measure used was the cosine of the angle between two patterns, which was
convinced to show the best performance in the simulation of the lengths of wet spells (Matulla et al. 2008). It is calculated as

\[ \text{similarity} = \frac{A \times B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^{n} G(i)A_i G(i)B_i}{\sqrt{\sum_{i=1}^{n} |G(i)A_i|^2} \sqrt{\sum_{i=1}^{n} |G(i)B_i|^2}}, \]

where \( A \) and \( B \) are two atmospheric patterns, and \( n \) is the total number of grid points between them. The similarity was measured between atmospheric patterns of the selected four predictors individually, and then the obtained similarities are combined together to get one similarity measure, the details of which are described later. An obvious difference from the original cosine angular formulation is that a weight function is assigned to each grid point to emphasize the role in matching analogs of those grid points that have larger correlations with PEP. The weight function is as follows:

\[ G(i) = \frac{|r_i|}{\sum_{i=1}^{n} |r_i|}; \]

in the formula, \( r_i \) is the correlation coefficient between each simultaneous or bygone large-scale variable at each grid point and the regional averaged precipitation in the study area for the 144 event days of the calibration period. Figure 2 lists the correlations between the mean precipitation in the study area during the calibration period (144 event days) and the simultaneous large-scale atmospheric conditions at each grid point: (a) 200-hPa zonal wind, (b) 500-hPa geopotential height, (c) 700-hPa zonal water vapor transport, and (d) 700-hPa meridional water vapor transport. Dotted areas indicate significant correlation coefficients at the 90% confidence level, estimated by a local Student’s \( t \) test.
(2014), including the jet stream in the upper troposphere of eastern China and Japan, blocking highs at higher latitudes, and the WPSH in the midtroposphere and the water vapor transport related to the WPSH at low levels. Thus, the applied weight function is able to help identify the key influential systems in the area that we have selected to match analogs and to enhance their effects in the AM scheme.

Moreover, for the convenience of processing and calculating the similarity hereafter, the similarity measure was transformed into a normalized positive form, which ranges from 0 to 1 with 1 for perfect similarity and 0 for the opposite:

\[ S = 1 - \frac{\cos^{-1} \text{(similarity)}}{\pi}. \]  

(3)

2) ANALOG SELECTION

Considering that PEPEs are caused by the persistent circulation anomalies, PEP on a given target day with a circulation pattern denoted as \( \phi^0 \) is supposed to be determined by the state of the circulation on that day and the preceding 7 days. This pattern forms an 8-day sequence \( \phi = (\phi^5, \phi^6, \ldots, \phi^0) \) called the target sequence. The possible analogous sequence \( \varphi = (\phi^7, \phi^8, \ldots, \phi^0) \) is searched to match the target sequence among all of the event days of the PEPEs, except for the event year that the target day \( \phi^0 \) is in. When searching analogs of the target day, it is obvious that analogs can be easily found among the event days in the same PEPE with the target day. So the event year that the target day \( \phi^0 \) is in must be ruled out in case the training process loses its validity when establishing the model. The similarity of the two sequences can be evaluated by the sum of the weight-applied similarity of each day, which can be shown as

\[ S_{y,p}(\phi, \varphi) = \sum_{k=0}^{7} \gamma_p(k) s(\phi^k, \varphi^k) \]  

(4)

\[ \gamma_p(k) = (8-k)/36, \]  

(5)

where \( \phi^k \) and \( \varphi^k \) are the circulation patterns of each day in the sequence and \( k \) stands for the number of days of each day in the sequence away from the target day. The lowercase \( s \) represents the aforementioned cosine angular similarity whose form is transformed in Eq. (3). The corresponding weight function is calculated with the correlations between large-scale variables for those \( k \) days ahead of the target day and the mean precipitation in the study area during the calibration period (figures omitted). Here, \( \gamma_p(k) \) is the weight applied to each day, which decreases linearly with the increase in distance from each day to the target day with a sum that amounts to 1.

The chosen large-scale meteorological variables constitute four different atmospheric state sequences, 200-hPa zonal wind, 500-hPa geopotential height, 700-hPa zonal water vapor transport, and 700-hPa meridional water vapor transport; thus, their similarity scores can be achieved and referred to as \( P_{200}, P_{500}, P_{700qv}, \) and \( P_{700qv} \), respectively. For the sake of combining the anomalies of the variables, an integral similarity score was defined to combine their similarity scores:

\[ S_{\text{integral}} = P_{200}S_{200} + P_{500}S_{500} + P_{700qv}S_{700qv} + P_{700qv}S_{700qv}, \]  

(6)

where \( P_{200}, P_{500}, P_{700qv}, \) and \( P_{700qv} \) are the normalized weights of the four different atmospheric state sequences above, and they also sum to 1. The integral similarity score also ranges from 0 to 1 with 1 being the perfect score, reflecting to what extent the whole atmospheric state is similar to the other. An optimization method was used to determine the four parameters to enable the DM to achieve its best performance.

c. Forecast output and optimization procedure

The larger the integral similarity score is, the closer the corresponding atmospheric states are. Of all the historical PEPE days, some can be chosen as analogs of the target day because of their identical quality. However, under the analogous atmospheric states, there may exist different precipitation situations. To solve this problem, we introduced the weighted average method for the analogs found (Fernández and Sáenz 2003). The weight is the similarity scores of the analogs found; thus, more weights are applied to analogs with higher similarity scores. With the main components of the forecast model presented in Fig. 3, there are two ways to yield the precipitation output after searching the analogs matches for the four large-scale predictors. One way is shown by the left branch of the schematic graph used to construct the precipitation output directly by employing the precipitation of the closest large-scale analog record; the other way takes into account the precipitation from all of the analog records with nearly equal quality to construct the precipitation of the target day (right branch of the schematic graph). Furthermore, a critical value parameter \( P_{cv} \) needs to be defined to judge which sequences of the historical record are sufficiently similar to the sequence of the target day. It is supposed that all of the selected historical records are of equal quality, provided that their integral similarity scores are larger than \( P_{cv} \). The cuckoo search technique (Yang and Deb 2009) is also used to determine \( P_{cv} \). In this study, the number of analog records chosen is \( n = 1, 3, \) or 6. The downscaling skills of these options were verified.
during the calibration period to help decide which number should be chosen for superior forecast output.

The NCEP–NCAR reanalysis dataset during the first 24 PEPEs (144 event days in total) and the rest of the days of the 22 summers when no PEP occurred (1880 nonevent days in total) were utilized in the training scheme. The optimal parameters we applied in the similarity score were determined by means of the cuckoo search technique. The cuckoo search technique is a bionic metaheuristic algorithm for searching for the optimal solution to equations or problems with some advantages over other algorithms (Yang and Deb 2009). The training was performed using the reanalysis data for the 144 event days and the 1880 nonevent days. For the event days, each day was taken as the target day, and then the relatively closest analogs were searched from the event days of other event years with similarity scores above $P_{cv}$. For the nonevent days, we searched the closest analogs with similarity scores above $P_{cv}$ from all 144 of the event days. If the closest analogs with similarity scores above $P_{cv}$ were enough to satisfy the demands of searching the 1, 3, or 6 relatively closest analogs, we got the weight-averaged output with member numbers of 1, 3, or 6 as the forecast of the target day. Otherwise, the day was considered to be a day with no PEP, and the forecast was set to be a constant field of missing value. However, the parameters $P_{200}, P_{500}, \ldots, P_{cv}$ were not determined. As a result, the closest analogs cannot be decided. Therefore, the cuckoo search technique is introduced into the training simulation, and the subjective function for the optimization problem is

$$ SF = \frac{H}{H + M + F} - \frac{NFR_1 + NFR_2}{NFR_1 + NFR_2}; $$

on the right-hand side of the equation, the first item is the critical success index also called the threat score (TS) of 50-mm precipitation for the 144 event days. The score indicates that the skill of a “yes” forecast corresponds to an observed yes event. For the first item, $H, M,$ and $F$ represent hits, misses, and false alarms, respectively. They are all obtained from the four combinations of forecasts (yes or no) and observations (yes or no) called the joint distribution including hits, misses, false alarms, and correct negatives. A hit indicates that an event is forecasted to occur and does occur. Misses denote that an event is not forecasted to occur but the event does occur. The rest can be deduced in the same way. The second and third items, both considered to be not found ratios (NFRs), are the ratio of the cases where no analogs can be found in the historical PEP records for the 144 event days and 1880 nonevent days, respectively. The metaheuristic algorithm cuckoo search was
designed to seek one set of optimal parameters \( (P_{200}, P_{500}, \ldots, P_{cv}) \) to enable the subjective function to reach its largest value. Therefore, when the training scheme ends, the TS score of the 50-mm precipitation forecast for the 144 event days should reach its largest value, but the NFR of the 144 event days (NFR1) should reach its minimum and the NFR of the 1880 nonevent days (NFR2) should reach its maximum. Hence, the optimized training scheme helps to maximize the skill of the DM to accurately simulate the observed 50-mm precipitation in PEPEs and to enable as many as possible of the 144 event days to find corresponding analogs in the historical PEPE records, at the same time minimizing the number of nonevent days that have matching analogs in the historical PEPE records. It is worth noting that in the calibrating procedure we also guaranteed that the similarity score of the 500-hPa anomalous atmospheric state was above 0.5 and the NFR of the 144 event days was less than 0.05.

The optimal parameters and the values of the subjective function determined in the training procedure are shown in Table 2. With the decrease in the geopotential height the optimization parameters generally show an increase, which suggests that the lower layer has a crucial influence on the AM scheme.

<table>
<thead>
<tr>
<th></th>
<th>( P_{200} )</th>
<th>( P_{500} )</th>
<th>( P_{700\text{gpa}} )</th>
<th>( P_{700\text{gpa}} )</th>
<th>( P_{cv} )</th>
<th>SF</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n = 1 )</td>
<td>0.1731</td>
<td>0.1578</td>
<td>0.3583</td>
<td>0.3108</td>
<td>0.5826</td>
<td>0.5927</td>
</tr>
<tr>
<td>( n = 3 )</td>
<td>0.0951</td>
<td>0.2234</td>
<td>0.2633</td>
<td>0.4182</td>
<td>0.5740</td>
<td>0.6277</td>
</tr>
<tr>
<td>( n = 6 )</td>
<td>0.1278</td>
<td>0.2292</td>
<td>0.2325</td>
<td>0.4105</td>
<td>0.5634</td>
<td>0.6001</td>
</tr>
</tbody>
</table>

4. Validation of the DMs during the calibration period

The performance of the DMs searching the \( n = 1, 3, \) and 6 analogs during the 144 event days is evaluated through the root-mean-square error (RMSE) and the variance skill; the results are shown in Fig. 4. The variance skill, which indicates the capacity of the DMs to reproduce the variability of the original series, is defined as the ratio of the variance of series outputted by the DMs to the variance of the original series during the calibration period. As can be seen from Fig. 4, the RMSE decreases with the number of analogs searched (\( n \)). It is remarkable that the RMSE decreases little for the averages of the six nearest analogs compared to the three analogs averaged. However, the variance skill shows an opposite tendency with the increase of \( n \). The
more analogs used to compute the average, the lower the variance skill. So the performance of DMs does not improve in all aspects when increasing $n$. The result is consistent with the study by Fernández and Sáenz (2003), and the intermediate value of $n = 3$ is a good choice for the final forecast model.

To sum up, after all of the optimal parameters and the number of nearest analogs are determined, a DM is established that is called the Key Influential Systems Based Analog Model (KISAM). In addition, the performance of KISAM for the average intensity of the precipitation in the study area is examined during the calibration period (Fig. 5). When calculating the average intensity of the day with no analogs found, it was taken as a constant field of 0 instead of being treated as a missing value. Results show that KISAM is a reliable model that can produce precipitation intensity that is basically close to the observed value. In Fig. 5, the correlation coefficient between the predicted values and the observations is 0.47, which is significant at the 99% confidence level, estimated by a local Student’s $t$ test.

5. Evaluation of the forecast performance of KISAM

a. Assessment of daily PEP forecasts by KISAM

KISAM was assessed for the hindcast of the last PEPE we retained, which is for 17–25 June 2010 using the forecast data at 1200 UTC for the four large-scale variables retrieved from the ensemble mean with a weight equal to the ECMWF simulations at 1–15-day lead times. The hindcast was verified and compared with the DMO produced by ECMWF for the same lead time. The verification was conducted against interpolated precipitation station data. For the accuracy of the forecast we used TS scores to conduct the verification and the comparison of KISAM and DMO. Figure 6 shows the TS scores for 40-mm precipitation predicted by KISAM and DMO during 17–25 June 2010 at 1-, 3-, 6-, 9-, 12-, and 15-day lead times. A TS score of 0 indicates that the weak intensity of the precipitation forecast or the location of the precipitation forecast is inaccurate, and a TS score of $\approx 0.1$ stands for a day that was forecasted by KISAM as a nonevent day with a constant field of missing value. We can see from Fig. 6 that KISAM shows better performance at 3-day lead time, with TS scores much higher than DMO, except for on 17, 19, and 20 June. At 6-day lead time, the advantages of KISAM over DMO are more evident, especially on 17–18 and 21–25 June. In addition, when the lead time extends to 1 week or longer, DMO can hardly show any capacity to predict PEP. KISAM mistook these event days as nonevent days mostly at the lead time of 12 and 15 days, but the model still has certain skill on some days like 18 and 23–24 June, particularly for 9-day lead time. DMO provides better predictions than does KISAM only at 1-day lead time. DMO usually provides better predictions at shorter lead times for precipitation (Su et al. 2014) and can take the short-term mesoscale and small-scale systems into consideration. However, KISAM...
is mainly based on the downscaling of large-scale variables. Thus, the performance of DMO is so good that KISAM cannot surpass DMO at 1-day lead time.

In addition, we further investigated the performance of DMO and KISAM in predicting the precipitation above 40 mm through the analysis of RMSEs. Considering the fact that the forecast by KISAM for 23 June at 1-day lead time is a nonevent day, which is a constant field with a missing value, it is set as 0 when calculating the RMSE for this day. For convenience of comparison, differences between the RMSEs of DMO and those of KISAM are calculated. Close attention is paid to the lead times of 3, 6, 9, and 12 days, when KISAM has advantages in TS scores, as shown in Fig. 7a. It is found that higher and lower RMSEs of KISAM than DMO are both discovered on different days during 17–25 June 2010. On the whole, the RMSEs of DMO appear to be higher than those of KISAM, especially on 18, 21, 23, and 24 June. From the average values of RMSEs at 3-, 6-, 9-, and 12-day lead times in Fig. 7b, lower-average RMSEs of KISAM than DMO can also be found on 18 and 21–25 June. Thus, overall low RMSEs also verify that KISAM performs better than DMO when predicting the intensity of PEP. Another aspect that needs to be recognized is that the RMSEs of DMO are all much higher than KISAM when predicting 18 and 23 June at 3-, 6-, 9-, and 12-day lead times (Figs. 7a, b), and the precipitation on 18 and 23 June is also much more intense than on other event days, with the amount reaching around 100 mm. This is attributed to a better prediction for strong PEP by KISAM while there is an underestimation of heavy precipitation by DMO, which is a general tendency of NWP models (Su et al. 2014). In summary, with regard to predicting the location and intensity of PEP, KISAM is superior to DMO at 3-day lead time and longer. For the early warning of PEP, KISAM is also more helpful than DMO to the forecast in advance of 1 week.

b. Performance in predicting PEPE

According to the definition used by Chen and Zhai (2013), a PEPE should meet the criterion that the
precipitation with an amount exceeding 50 mm must persist for at least three consecutive days in a region. While conducting forecasts of the PEPE, the duration is not known in advance, so we take a PEPE with the shortest duration, which is 3 days, as an example for verifying the performance of KISAM. When producing forecasts of KISAM, once three consecutive occurrences of EP are found, it can be forecasted as a PEPE. The duration of the PEPE is then added if more EP following the 3-day event are found. Otherwise, the days with analogs found will be predicted as an independent EP. Figure 8 shows the comparison of the observations and the hindcast of the PEPE during 17–19 June; the observed precipitation events with intensity exceeding 50 mm are mainly located in the southern part of central–eastern China (Figs. 8a1–a3). For the DMO produced at 11 June, there was no PEPE captured with weak precipitation (Figs. 8b1–b3). After 1 day, the amount of precipitation above 50 mm appeared in DMO to be in generally the right location and of the correct intensity (Fig. 8c1–c3), so the lead time of this day can be taken as the forecasting valid time of DMO. In other words, the forecasting valid time of DMO is 5 days. Comparatively speaking, as shown in Figs. 8d1–d3, KISAM captured the persistent 50-mm precipitation at 9 June, which was 8 days in advance of the PEPE, and the location of the precipitation predicted was relatively more accurate than for DMO at 5-day lead time. Moreover, we compared the hindcasts of KISAM and DMO at the same 5-day lead time (Figs. 8e1–e3) and KISAM possessed an obviously higher level of skill than DMO. Therefore, KISAM is able to capture the PEPE on 17–19 June 2010 ahead of DMO and to provide more accurate forecasts of the location and intensity of precipitation than DMO at the same 5-day lead time.

c. Skill of KISAM to forecast the occurrence of EP and PEPE

All of the investigations involving KISAM mentioned above are performed under the condition that the PEPE during 17–25 June 2010 has already been known beforehand. So verification needs to be done across a relatively long time, assuming that when the PEPE occurred is unknown, to investigate the skill of KISAM to separate event days from nonevent days. In this study, the hindcasts derived by KISAM during June 2010 for 1-, 3-, 6-, 9-, 12-, and 15-day lead times are verified through the joint distribution of forecasts and observations, which consists of hits, misses, false alarms, and correct negatives. As the results show in Table 3, for all lead
Fig. 8. Observed and forecasted precipitation distributions during 17–19 Jun 2010: (a) observation and forecast by (b) DMO at 6-day lead time, (c) DMO at 5-day lead time, (d) KISAM at 8-day lead time, and (e) KISAM at 5-day lead time.
times KISAM issued the right forecast of the PEPE event days (17–25 June) at the lead times of 1, 3, and 6 days, except for the miss in the forecast for 23 June at 1-day lead time. Furthermore, for the nonevent days far from the event days like 1–14 June and 27–30 June, KISAM also generally provides the correct negative forecasts. This suggests that KISAM performs well in the discrimination of event days and nonevent days. However, KISAM performs poorly during the nonevent days near PEPE about 1–2 days before the event and 1 day after the event. False alarms appear in the forecasts by KISAM for this period at 1-, 3-, and 6-day lead times. In other words, KISAM mistakes these nonevent days as event days at almost all of these lead times. As a consequence, when utilizing KISAM to identify a PEPE in advance of 1, 3, and 6 days, the model tends to yield a PEPE forecast prematurely, which is about 2 days before the PEPE actually occurs, and to incorrectly lengthen the duration of the PEPE by about 3 days in the forecast. When predicting the PEPE at longer lead times like 9, 12, and 15 days, the premature forecast and the duration are lengthened more substantially. This phenomenon in the KISAM forecasts appears to be related to the key influential systems remaining stable ahead of the PEPE and retreating behind the PEPE. KISAM yielded forecasts of PEP based on the key influential systems established while in actuality the PEP did not occur. Another circumstance to bear in mind is that the key influential systems still exist while PEPE has terminated. This will result in incorrect KISAM forecasts. We expect to calibrate this tendency of the false alarm through verifying more PEPE samples.

To verify the skill of KISAM in predicting EP and PEP, forecasts in practice were further carried out by KISAM for June 2015 using the large-scale information produced by ECMWF at 1–15-day lead times. As a result, KISAM could yield forecasts of the future 1–15 days each day to identify whether EP or PEP will occur in the future 1–15 days. Taking 18 June for example, forecasts for the first 1–6 future days corresponding to 19–24 June were all predicted as nonevent days by KISAM. From Figs. 9a1–a6, we can draw that EP occurred from 25 to 28 June as KISAM’s 18 June prediction indicated. In other words, KISAM yielded a PEPE forecast that lasted for 4 days in advance of 7 days. However, the corresponding observed precipitation from 25–28 June (Figs. 9c1–c6) did not meet the PEPE definition proposed by Chen and Zhai (2013) because the precipitation on 26 June was lower than 50 mm. In addition, after a 1-day break on 29 June, EP occurred again on 30 June. During this complicated precipitation process, KISAM mistook the event as a 4-day PEPE and missed the EP that occurred on 30 June. But it was reasonable to expect that KISAM might have missed the EP on 30 June as the forecast time reached out to 12 days. By comparison, DMO forecasts, which also started on 18 June for 25–30 June, are shown in Figs. 9b1–b6. DMO yielded obviously weaker precipitation than was observed and failed to predict the EP on 25, 27–28, and 30 June. The performance of KISAM was much better than DMO during this complicated precipitation process. It is worth noting that KISAM correctly identified 25 and 27–28 June as EP in advance of 1 week. Particularly, with regard to the location and intensity of the precipitation, KISAM showed high skill levels in the forecast for 27–28 June, which can provide great help for forecasters.

Moreover, better performance of KISAM is further confirmed from the average TS scores for predicting the 40-mm precipitations on the days when EP occurred in June 2010 and 2015 (Fig. 10). Because the atmospheric variables used in KISAM are reliable merely up to lead

<table>
<thead>
<tr>
<th>FLD: 1</th>
<th>FLD: 3</th>
<th>FLD: 6</th>
<th>FLD: 9</th>
<th>FLD: 12</th>
<th>FLD: 15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–11 Jun</td>
<td>Cn</td>
<td>Cn</td>
<td>Cn</td>
<td>Cn</td>
<td>Cn (1–3 Jun F)</td>
</tr>
<tr>
<td>12 Jun</td>
<td>Cn</td>
<td>Cn</td>
<td>Cn</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>13 Jun</td>
<td>F</td>
<td>F</td>
<td>Cn</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>14 Jun</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>15 Jun</td>
<td>Cn</td>
<td>F</td>
<td>F</td>
<td>Cn</td>
<td>F</td>
</tr>
<tr>
<td>16 Jun</td>
<td>Cn</td>
<td>F</td>
<td>F</td>
<td>Cn</td>
<td>F</td>
</tr>
<tr>
<td>26 Jun</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>27 Jun</td>
<td>Cn</td>
<td>Cn</td>
<td>Cn</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>28 Jun</td>
<td>Cn</td>
<td>Cn</td>
<td>F</td>
<td>Cn</td>
<td>F</td>
</tr>
<tr>
<td>29–30 Jun</td>
<td>Cn</td>
<td>Cn</td>
<td>Cn (30 Jun F)</td>
<td>Cn</td>
<td>Cn</td>
</tr>
</tbody>
</table>
FIG. 9. Observed and forecasted precipitation distributions during 25–30 Jun 2015: (a1)–(a6) KISAM prediction on 18 June for 25–30 June (blank represents forecast as nonevent days), (b1)–(b6) DMO prediction on 18 June for 25–30 June, and (c1)–(c6) the corresponding observations.
times of 10 ~ 12 days and TS scores of DMO decrease to 0 at lead times longer than 12 days, the performance of KISAM and DMO is only compared at 1–12-day lead times. The TS score of DMO decreases as the lead time is extended, dropping to 0 with no skill at 10-day lead time. The TS scores for KISAM are stable during the first 6 days and start to show a decline at day 7. At lead times of 3 days and longer, except for day 4, KISAM shows an advantage over DMO on average in predicting EPs according to its higher TS scores. The advantages of KISAM become obvious in the medium range of 7–12-day lead times, which verifies the results achieved above.

To sum up, KISAM has a relatively high level of skill when predicting occurrences of EP and PEP even at lead times of 1 week or longer, and the location and intensity of precipitation provided by KISAM match closely the observations. At 3-day leads and longer, KISAM has better performance than DMO on average in the forecast accuracy of EP. Additionally, KISAM is skillful in separating nonevent days from event days at relatively shorter lead times. However, KISAM is prone to mistaking nonevent days near PEPE for event days and for prolonging the duration of PEPEs. This weakness of KISAM may be revised through verification of more PEPE samples.

6. Conclusions and discussion

In this study, training of an analog forecast model is constructed with characteristic features including the use of the large-scale atmospheric variables in the active area of key influential systems, the similarity method of cosine angular with weight applied, and the cuckoo search optimal algorithm. The model established is the Key Influential System Based Analog Model (KISAM) for EP and PEP forecasts. Verification of the performance of KISAM in independent forecasts was carried out during the reserved PEPE, which occurred on 17–25 June 2010. Furthermore, KISAM was operational during June 2015 to assess its performance in actual forecasts. The forecasts derived by KISAM are analyzed through comparison with the DMO by ECMWF. The main conclusions are as follow.

On the aspect of predicting daily PEP during the PEPE on 17–25 June 2010, DMO shows higher skill than KISAM merely at 1-day lead time. At 3 days and longer, KISAM displays an advantage over DMO for its generally high TS scores. The RMSEs of KISAM are overall lower than those of DMO, especially for the event days with more intense PEP. All of the above results indicate that KISAM is able to identify the approaching PEP earlier and yield a more accurate forecast for location and intensity of PEP than DMO at 3 days and longer. When forecasting an independent PEPE case, like that of 17–19 June 2010, KISAM can capture the PEPE at a lead of 8 days, which is 3 days earlier than DMO. So KISAM can give a warning about PEPE much earlier and that is of great importance. In addition, through comparison with forecasts by DMO at the same lead time of 5 days, it is shown clearly that the KISAM forecasts are obviously closer to the observations.

In the end, we carried out verification of KISAM during the whole of June 2010 to investigate the
capability of KISAM to separate event days from nonevent days. It is encouraging to see that KISAM could yield generally correct forecasts at shorter lead times for event and nonevent days that were not close to a PEPE. As a real-case application, we verified forecasts carried out by KISAM for June 2015 at lead times of 1–15 days. Results show that KISAM was relatively more skillful than DMO in predicting the occurrences of EP and provides close location and intensity estimates of precipitation compared to observations at lead times of 1 week or even longer. The forecasts of location and intensity of precipitation provided by KISAM match the observations closely. Through validation of more EP, KISAM shows better performance than DMO on average at 3 days and longer.

PEP is characterized by its persistence, extremity, and regional features. Many other investigations have focused on one of these features for their precipitation forecasts. The analog forecast system developed by the Hong Kong Observatory (HKO) specially focuses on the forecasting of heavy rainfall above 25 mm and records the rainfall category results daily for Hong Kong (Chan et al. 2014). Other new statistical methods were usually used for predicting the amounts and diagnostic tests of daily precipitation, which are scarcely concerned with the persistence and extremity of precipitation (Harpham and Wilby 2005; Hu et al. 2013). KISAM gives consideration to all of the features of PEP concurrently. In general, KISAM improves the forecasts of PEP over the study region compared to DMO. However, KISAM also exposes weaknesses when predicting nonevent days near PEPEs. As a result, the model is prone to prolonging the duration of PEPEs. This weakness is probably due to the key influential systems such as blocking highs, the northwestern Pacific subtropical high, and the high-level jet that is established before PEP initiation and retreats afterward. Such mismatches cause KISAM to yield a PEP duration prediction that is longer than that observed. Improvements are expected to be made through verifying more PEPE samples to add more appropriate parameters in the future.

Acknowledgments. This study was supported by the National Key Basic Research Program of China (Grant 2012CB417205), National Key Technology Support Program (Grant 2015BAC03B02), and Basic Research Fund of the Chinese Academy of Meteorological Sciences (2015Z001). The authors are grateful to the editor, Dr. Yuqing Wang, and three anonymous reviewers for their invaluable and constructive suggestions and comments that helped improve the manuscript.

REFERENCES


