New Tools for Comparing Beliefs about the Timing of Recurrent Events with Climate Time Series Datasets

MATHEW ALEXANDER STILLER-REEVE

Uni Research Climate, Uni Research AS, and Bjerknes Centre for Climate Research, Bergen, Norway

DAVID B. STEPHENSON

Department of Mathematics and Computer Science, University of Exeter, Exeter, United Kingdom

THOMAS SPENGLER

Geophysical Institute, University of Bergen, and Bjerknes Centre for Climate Research, Bergen, Norway

(Manuscript received 31 August 2015, in final form 22 July 2016)

ABSTRACT

For climate services to be relevant and informative for users, scientific data definitions need to match users' perceptions or beliefs. This study proposes and tests novel yet simple methods to compare beliefs of timing of recurrent climatic events with empirical evidence from multiple historical time series. The methods are tested by applying them to the onset date of the monsoon in Bangladesh, where several scientific monsoon definitions can be applied, yielding different results for monsoon onset dates. It is a challenge to know which monsoon definition compares best with people's beliefs. Time series from eight different scientific monsoon definitions in six regions are compared with respondent beliefs from a previously completed survey concerning the monsoon onset.

Beliefs about the timing of the monsoon onset are represented probabilistically for each respondent by constructing a probability mass function (PMF) from elicited responses about the earliest, normal, and latest dates for the event. A three-parameter circular modified triangular distribution (CMTD) is used to allow for the possibility (albeit small) of the onset at any time of the year. These distributions are then compared to the historical time series using two approaches: likelihood scores, and the mean and standard deviation of time series of dates simulated from each belief distribution.

The methods proposed give the basis for further iterative discussion with decision-makers in the development of eventual climate services. This study uses Jessore, Bangladesh, as an example and finds that a rainfall definition, applying a $10 \, \mathrm{mm \, day}^{-1}$ threshold to NCEP–NCAR reanalysis (Reanalyis-1) data, best matches the survey respondents' beliefs about monsoon onset.

1. Introduction

Of the several different functions climate services have (Miles et al. 2006), one is to translate climate research into practical applications useful for decision support (Brooks 2013; Visbeck 2008). If we want to communicate climate information as part of a climate service program, we must evaluate how appropriate this information is. We have to evaluate how well this information speaks to the stakeholders, as their decisions are based on their view of the world, and their experiences with weather and climate (Lemos 2008; Marx et al. 2007; Moss et al. 2013; Rosenzweig and Binswanger 1993; Weber 2006). It is therefore important to understand how the stakeholders define weather events or seasons that impact their lives, in order to understand the gap between producers and users of climate information (Buontempo et al. 2014). This gap can be particularly wide if stakeholders' definitions of an event differ from the wider research community (Pennesi 2007; Stiller-Reeve et al. 2015). We must ensure that the producers and the users of scientific information are

talking about the same phenomenon. For a climate phenomenon (e.g., the monsoon onset), the producers should define it in a way that yields timing and variability that the users can relate to. In this way, we increase the potential to develop science products that "fit" stakeholder needs (Ray et al. 2007). This paper attempts to make this fit by evaluating the timing of seasonal onsets given by different scientific definitions against the beliefs of stakeholders. Whereas in a previous paper (Stiller-Reeve et al. 2015), we made general comparisons between scientific definitions and belief, here we develop statistical tools to make quantitative comparisons.

One important climate phenomenon to millions of people around the world is the monsoon, in particular monsoon-related rainfall. Scientific information about a monsoon onset might be useful if it aligns with the beliefs of the intended stakeholders. This alignment is sometimes elusive. Pennesi (2007) showed that different ways of defining the wet season in Brazil led to distrust in the official forecasts. Monsoon definitions also diverge in Bangladesh, where the people define the monsoon differently in different locations (Stiller-Reeve et al. 2015). Furthermore, scientific narratives about when the monsoon starts also vary and depend on the scientific definition of the monsoon that is applied. There is still no consensus on how to define the monsoon in the research community (Wang and LinHo 2002; Yang et al. 2012; Zeng and Lu 2004). Over the years, research has postulated several monsoon definitions for Southeast Asia. These definitions use different variables and thresholds to designate the monsoon onset and withdrawal. While some use rainfall observations (Ahmed and Karmakar 1993; Matsumoto 1997) or gridded datasets (Ashfaq et al. 2009; Wang and LinHo 2002; Zeng and Lu 2004; Zhang 2010), others use proxies for rainfall like outgoing longwave radiation (Zhang et al. 2004) or those that are based on indices related to atmospheric circulation (Wang et al. 2009; Zhang 2010). Over Bangladesh, these monsoon definitions can give varying results that sometimes compare poorly with the beliefs of local people (Stiller-Reeve et al. 2015).

In a previous study, researchers asked the people of Bangladesh in the locations shown in Fig. 1 when they believed the monsoon normally started, as well as the earliest and latest onset dates over the past 20 years (Stiller-Reeve et al. 2015). The results from this survey showed that the respondents in Sylhet, northeast Bangladesh, believed the monsoon started particularly early. This perceived onset was roughly one month before onset dates from most scientific definitions. The challenge is how to best compare the onset dates given by the scientific definitions with the elicitations from the respondents.

In this study, we use a modified triangular distribution to model the beliefs of the respondents and to compare
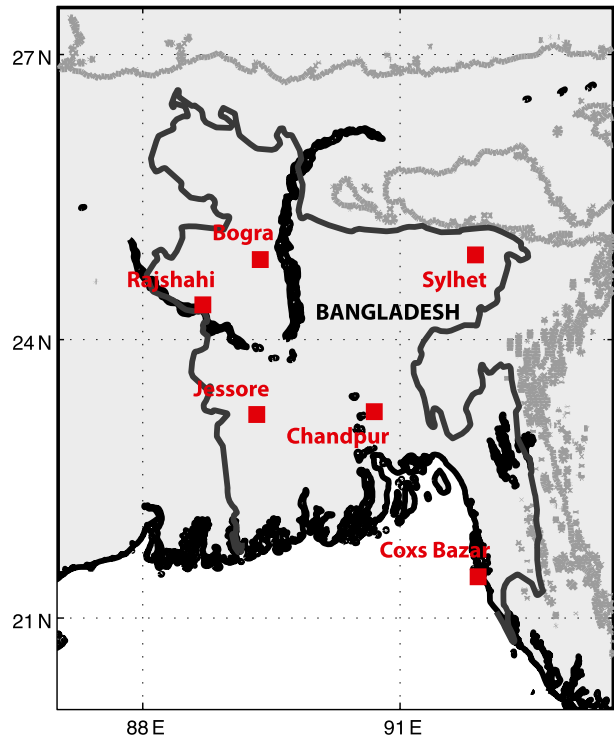

FIG. 1. Locations in Bangladesh where the questionnaire study was carried out by Stiller-Reeve et al. (2015).

them to historical time series of the monsoon onsets from empirical definitions (historical time series). Using these modified triangular distributions, we further develop two methods for comparison: one uses log-likelihoods and the other uses time series simulations.

In the next section, we present the data for comparison, including a closer look at the results from the questionnaire survey and a description of how we generate the historical climate time series of the monsoon onset, according to different scientific definitions. In section 3, we describe the statistical model and the method that we apply to compare the scientific and societal narratives about the monsoon onset. We explain why we use the triangular distribution and how we modify it for the present application. We illustrate the method and discuss the results in section 4, with a focus on the Jessore district in Bangladesh. Section 5 summarizes our results and presents some ideas for future research.

2. Data

a. Public perception

In 2011, colleagues at the Bangladesh Centre for Advanced Studies carried out a survey in six different locations around Bangladesh, as shown in Fig. 1 (Stiller-Reeve et al. 2015). In each location, around 200

TABLE 1. Overview of the different monsoon definitions and data to which we apply them.

Threshold type	Data source	Original resolution	Threshold	Historical time series tag
Rainfall static	APHRODITE	0.25°	5 mm day^{-1}	rain1_5_aphr
Rainfall static	NCEP–NCAR	T64	5 mm day^{-1}	rain1_5_ncep
Rainfall static	APHRODITE	0.25°	10 mm day^{-1}	rain1_10_aphr
Rainfall static	NCAR-NCEP	T64	10 mm day^{-1}	rain1_10_ncep
Rainfall variable	APHRODITE	0.25°	Pentad mean (mm day^{-1})	rain2_aphr
Rainfall variable	NCEP–NCAR	T64	Pentad mean (mm day^{-1})	rain2_ncep
OLR	NOAA OLR	2.5°	240 W m^{-2}	olr_noaa
Wind	NCEP–NCAR	2.5°	+ve v-wind 850 hPa −ve u-wind 200 hPa	wind_ncep

respondents participated. These respondents were chosen by the field teams by random sampling with several simple criteria. The respondents should be

1) over 40 years old,
2) permanent residents of the settlement concerned, and
3) agricultural workers.

Other professions were also represented, but the agricultural workers composed 93.0% of the entire survey. The respondents answered questions about monsoon onset timings, eliciting the respondents' beliefs of how the monsoon progresses across the country.

Questions were designed as straightforward as possible to leave little room for misunderstanding. Respondents stated when the monsoon normally started, followed by their opinions on the earliest and latest monsoon onsets during the previous 20 years. In the present study, these three answers (earliest, normal, and latest monsoon onsets) provide the foundation for the comparison with the historical time series. We do not delve any further into the backgrounds of the respondents in this paper and solely use their answers to help develop and test our proposed methods to compare their beliefs with historical climate time series.

b. Historical time series

To generate time series of monsoon onsets, we chose different monsoon definitions and different datasets to which to apply them. Each monsoon definition aims to pinpoint the transition between the premonsoon and monsoon seasons. To do this, a definition usually uses a specific parameter, such as rainfall, wind direction, or proxies for each of these. A definition usually applies a threshold to this parameter, which signifies the seasonal shift. Since these thresholds can be exceeded throughout the year, a definition also needs to identify a transition. This transition is commonly defined by how long a threshold must be exceeded before the monsoon is declared. In this study, we use three different parameters and five different thresholds on four different datasets, as described in Table 1. We designate the transitions using

the method developed by Stiller-Reeve et al. (2014). The parameters and thresholds that we use in this study are based on previously published studies on the monsoon onset. These studies were designed for different scientific objectives, without necessarily a climate service application in mind. Therefore, our results are in no way meant to undermine any of the previous work to which we refer. Instead, we use these definitions as our starting point to inform practices within climate services.

We apply the different monsoon definitions to the four datasets. Even though these data cover different periods, they all span the period 1978–2007. Hence, we choose this period to compile a climatology of onsets that is consistent for all the datasets. We keep the data resolution consistent for analysis by bilinearly interpolating all the results to 0.25°, which is identical to the highest-resolution dataset. We realize that the datasets may not be ideal at these higher resolutions, especially in regions of rapid topographic changes. Northeast Bangladesh is one such region where the flat plain is hugged by the Khasi Hills to the north. However, since we aim to compare local perceptions with data, it is favorable if the data themselves are also ''local.'' The highest resolution gives us the possibility to extract time series from closer to the survey locations. All the interpolated datasets are converted to Julian pentad values, which is commonly used in monsoon research (Matsumoto 1997; Wang and LinHo 2002). A Julian pentad is one of seventy-three 5-day periods that the year (if we exclude 29 February in leap years) can be separated into (e.g., 1–5 January). In the following section, we describe the different monsoon definitions that we apply in this study. The mean onsets over Bangladesh using the different definitions and datasets are shown in Fig. 2.

1) RAINFALL: SPATIALLY STATIC THRESHOLD

We apply a static threshold to the rainfall datasets. The threshold is static because it does not change spatially. The advantage of using a static value for the monsoon onset is that it facilitates the comparison of results between different regions. The problem is that it
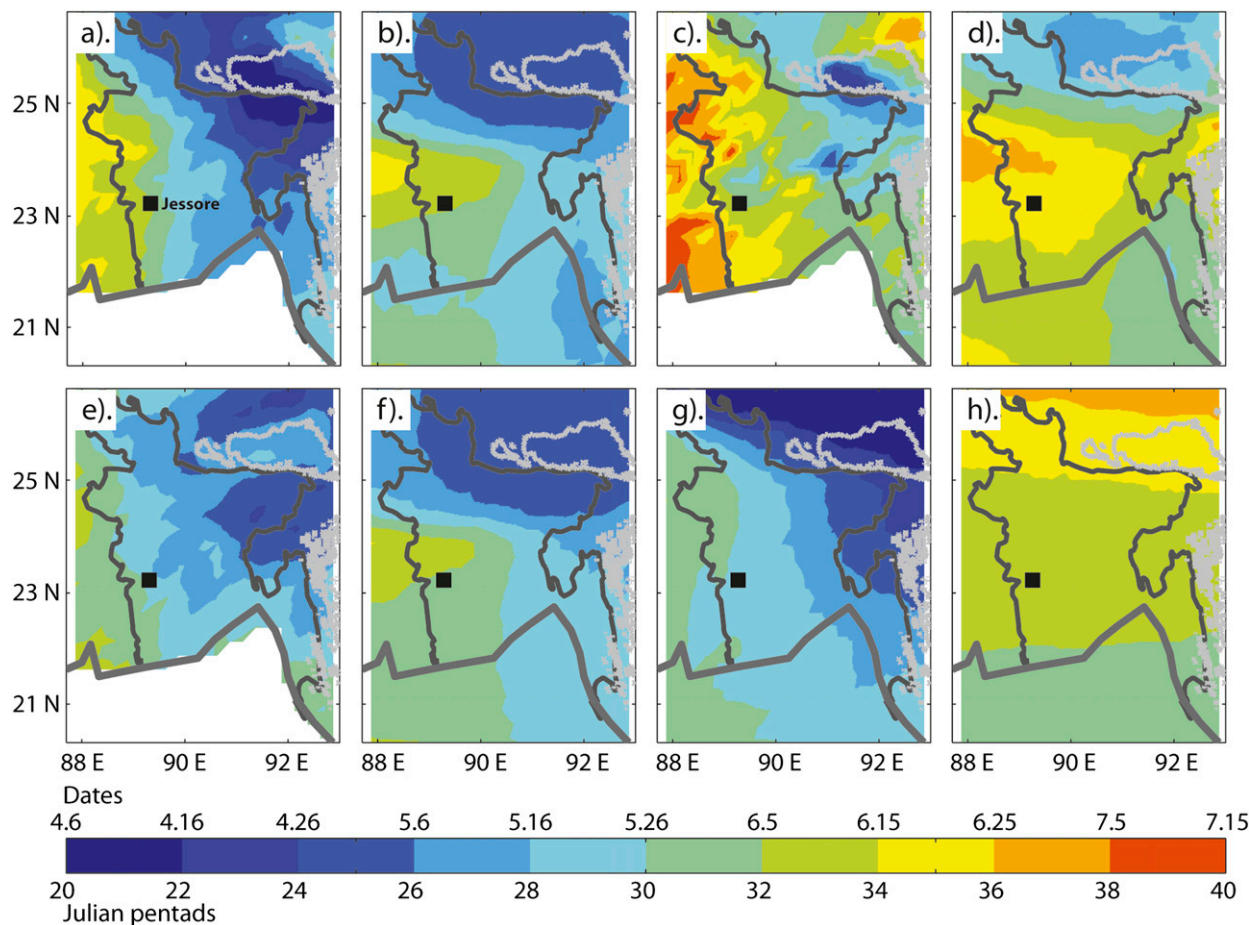
FIG. 2. The average monsoon onsets from the historical time series. The Bangladeshi border is shown in dark gray, and the 500-m topography contour is shown in light gray. The figures show the results (according to the time series tags in Table 1) from (a) rain1_5_aphr, (b) rain1_5_ncep, (c) rain1_10_aphr, (d) rain1_10_ncep, (e) rain2_aphr, (f) rain2_ncep, (g) olr_noaa, and (h) wind_ncep.

does not consider local effects and is usually chosen somewhat subjectively.

In this study, we use a 5 mm day^{-1} threshold (Ahmed 1994; Ahmed and Karmakar 1993). We also apply a 10 mm day^{-1} threshold, justified by the considerable convective activity in Bangladesh, especially in the northeast, between March and May (Yamane et al. 2010). These thresholds are applied to pentad rainfall values in two datasets. We chose two datasets so that we could also compare results from data with different resolutions. With this in mind we selected the APHRODITE rainfall dataset (Yatagai et al. 2012) and NCEP–NCAR reanalysis (Reanalysis-1) (Kalnay et al. 1996).

The APHRODITE dataset contains daily rainfall values (mm day^{-1}) over the Asian monsoon region at a 0.25° resolution. The dataset is created using observations from rain gauges over the Asian monsoon region and spans 1951–2007. The data are not very reliable over Bangladesh during the early 1970s, presumably because of domestic turmoil around the time of Bangladeshi independence. The Reanalysis-1 data for precipitation rate are resolved on a T62 Gaussian grid. Around Bangladesh, this implies a resolution of roughly 1.9°. The units are kg m^{-2} s^{-1}, which were converted to mm day^{-1}. The NCEP–NCAR data were linearly interpolated to the same grid as the APHRODITE data.

2) RAINFALL: SPATIALLY VARIABLE THRESHOLD

Spatially variable thresholds for determining the monsoon onset vary from grid point to grid point. Here, we determine the rainfall threshold by calculating the mean pentad rainfall (mm day^{-1}) for each grid point (Matsumoto 1997). For example, if a grid point features a 2500-mm average annual rainfall, the threshold will be set to 6.85 mm day^{-1}, equivalent to a *total* pentad amount of 34.25 mm. We apply theses thresholds to the interpolated Reanalysis-1 and APHRODITE datasets.

3) OUTGOING LONG WAVE RADIATION

Outgoing longwave radiation (OLR) datasets have the benefit of being available over land and ocean. OLR is also used because direct rainfall observations over many tropical regions are not very reliable. As OLR is a proxy for deep convection, OLR also acts as a proxy for monsoonal rainfall.

Previously, researchers applied static thresholds to identify the monsoon onset using OLR data. In this study, we use interpolated OLR NOAA data (Liebmann 1996), with a resolution of 2.5°, which we interpolated to 0.25°. The data were provided by the NOAA/OAR/ESRL/ Physical Sciences Division (PSD), Boulder, Colorado, from its website (http://www.esrl.noaa.gov/psd/). After converting the data to pentad values, we apply a threshold of $240 \, W \, m^{-2}$ as our monsoon criteria. This threshold has been used in previous studies concerning the Southeast Asian (Zhang et al. 2004) and West African monsoons (Fontaine et al. 2008).

4) WIND DIRECTION

As the advent of the monsoon is associated with a reversal in atmospheric circulation, we also include a definition based on wind direction. Definitions such as these are very dependent on the climate of the region in question. Around Bangladesh, the monsoon circulation can be characterized by a reversal in flow both at the surface and higher altitudes. We therefore chose a definition that takes into account the reversal of wind direction at two levels in the atmosphere. At the lower levels, we use the same threshold as Zhang (2010), that the winds must have a southerly component. This is an appropriate criterion, as the southerly winds advect moisture from the Bay of Bengal across Bangladesh. Ramage (1995) also defines the monsoon as the period when the winds are reversed compared with the winter months. As the winds are mainly from the north during the (boreal) winter, the southerly wind criterion seems justified (Ahmed 1994; Ahmed and Karmakar 1993). At higher levels (200 hPa) we use a criterion that takes into account the formation of the easterly jet, associated with the meridional temperature contrast related to the South Asian monsoon circulation. The second criterion therefore stipulates that the winds at 200 hPa must have an easterly component. We apply these criteria to data from Reanalysis-1, which have a resolution of 2.5°, interpolated to 0.25°.

c. Time series tags

Table 1 gives an overview of the different monsoon definitions we have applied in this study, along with the details of the data sources. Table 1 also presents the tags that we will use to refer to the results from the different monsoon definitions. The tags are also used in Fig. 2, which shows the average onsets for the different definitions across Bangladesh between 1978 and 2007. We extract historical time series from the closest grid points to the locations where the earlier questionnaire survey was completed (see Fig. 1). These historical time series and questionnaire results are the foundation of the analysis in this study. We will now describe our methodology and present a detailed analysis from Jessore as an example.

3. Methodology

a. Probabilistic representation of beliefs

To quantify their uncertain beliefs about the onset date, we assume that a survey respondent believes that the monsoon has probability $h(x)$ of starting on date x for $x = 1, 2, \ldots, T$, that is, all the calendar dates in a year. In other words, we consider the onset date to be a random discrete variable X having probability mass function (PMF) given by $\Pr(X = x) = h(x)$. Respondents with different beliefs have different PMFs. These PMFs are calculated by fitting appropriate parametric functions to each response from the previous survey.

Various parametric functions could be used to calculate these PMFs. The triangular distribution is a good starting point, as it is a continuous probability distribution used when the underlying probability is unknown (Kotz and van Dorp 2004). Recently, Marimo et al. (2015) used triangular distributions to generate "observations" of temperatures. These observations were used to assess how 289 undergraduate students interpreted uncertainty in forecasts. The triangular distribution is favorable, as it is simple and has been widely used in expert elicitation in risk analysis (Johnson 1997) and project management scenarios (Back et al. 2000; Briand et al. 2000; Galway 2007). Here, we consider the Bangladeshi people as experts regarding the timing of the monsoon in their local area. The respondents gave us answers for the earliest (x_E), the normal (x_N), and the latest (x_L) monsoon onset dates, which correspond to the three parameters used to construct a triangular distribution: minimum, maximum, and the most likely value, respectively. However, the triangular distribution is not suitable for this application because there is a nonzero chance of monsoon onsets in data falling outside of the ranges of the distribution. This unrealistic belief of *impossibility* leads to singularities in likelihood scores and so the triangular distribution has to be modified to allow for a small chance of onset dates lying outside the range. A further complication is that our date variable is periodic and so the modified triangular
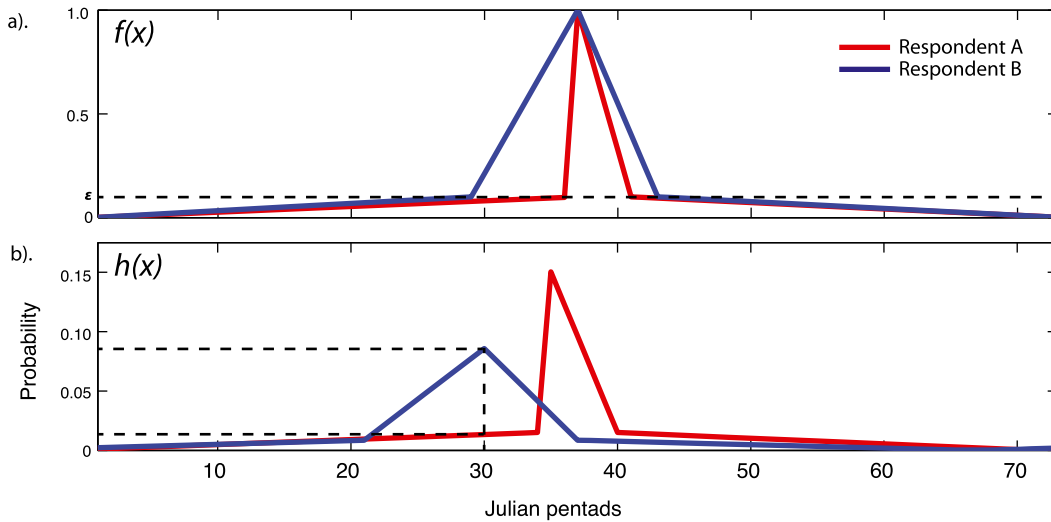
FIG. 3. Answers from two arbitrary respondents A and B to illustrate how we construct the PMFs. (a) First, we centralize each of the respondent's answers on Julian pentad 37 and construct the function with highest point equal to 1. Here we take $\varepsilon = 0.1$ as an example. (b) Then we shift the function back to the original answers given by the respondents and normalize. The figure also illustrates how we obtain likelihoods from an onset date of JP 30.

distribution has to be defined on a circular domain. Therefore, we propose a circular extension to the modified triangular distribution, that is, a triangular distribution modified to have upper and lower "wings" (Vally et al. 2014).

Our circular modified triangular distribution (CMTD), $h(x)$, is assumed to be a continuous periodic piecewise linear function with four segments (see Fig. 3b): an asymmetric triangle with maximum at $x = x_N$ that drops off to a low probability at $x = x_E$ and $x = x_L$, and then wings that drop off to zero at $x_N \pm T/2$. In other words, there are small nonzero probabilities outside the earliest and latest onset dates that linearly decline to zero half a year before and after the normal onset date, respectively. The PMF is easily constructed using shifted variables $y = x - x_N + (T + 1)/2$ to define the function shown in Fig. 3a:

$$g(y) = \begin{cases} \dfrac{\varepsilon(y - 0.5)}{y_E - 0.5}, & 1 \le y < y_E \\[2ex] 1 - \dfrac{(1 - \varepsilon)(y_N - y)}{y_N - y_E}, & y_E \le y < y_N \\[2ex] 1 - \dfrac{(1 - \varepsilon)(y - y_N)}{y_L - y_N}, & y_N \le y < y_L \\[2ex] \dfrac{\varepsilon(T + 0.5 - y)}{T + 0.5 - y_L}, & y_L \le y \le T \end{cases} \quad (1)$$

where $y_N = (T + 1)/2$, $y_E = x_E - x_N + (T + 1)/2$, and $y_L = x_L - x_N + (T + 1)/2$. The PMF is then given by $g(y)$

shifted back to x coordinates and scaled so that its sum over all x values is unity:

$$h(x) = \frac{g(y)}{\displaystyle\sum_{y'=1}^{T} g(y')}. \quad (2)$$

The regularization parameter ε controls the chance of the monsoon occurring outside of the earliest and latest onset dates and can be tuned toward zero to recover an unmodified triangular distribution.

The beliefs of two respondents are used in Fig. 3 to illustrate how the CMTD is constructed from the answers from each individual respondent. It also illustrates how the people's beliefs can vary in one geographical location. Respondent A believed that the monsoon normally begins at JP 34. This person gave an earliest onset of JP 34 and a latest onset at JP 39. Hence, the probability mass function is heavily skewed to the left since $x_N = x_E$. The probability mass function is also sharp, as respondent A believed the monsoon would always start within a period of 5 pentads (25 days). Respondent B's probability mass function is much flatter and more symmetric. This is due to the respondent's belief that the monsoon could start within a longer period of 14 pentads (70 days) with a normal onset at JP 30.

b. Simulation

Previous studies have shown that people in monsoon regions are dependent on the timing of the monsoon onset (Gadgil and Kumar 2006; Rosenzweig and

Binswanger 1993). These people may therefore be interested in when the monsoon starts on average and also by how much it varies from year to year. With the CMTD elicited from each respondent, we can simulate an artificial time series and calculate its mean and standard deviation.

We simulate the time series using the cumulative mass function of each CMTD and randomly generate numbers between 0 and 1. In our simulations, we generate 10 000 random numbers. The cumulative mass function is then used to map these values back onto Julian pentads between 1 and 73 and thereby simulating an artificial time series with 10 000 onset dates. From these simulations, we derive the mean and standard deviation and compare them with the equivalent values from the historical climatic time series.

c. Log-likelihood score

We can also consider the match between the respondents' beliefs and the historical time series by using log-likelihoods to construct a score. We have eight historical time series for each location in Bangladesh. Each time series has up to 30 monsoon onset dates for the period 1978–2007. By considering each CMTD in turn, we read the likelihoods for each of the dates in a historical time series. Figure 3b shows how we obtain the likelihood of a monsoon onset of JP 30 for respondents A and B. The likelihoods are much higher for dates that fall close to the respondent's answer for normal onset date (x_E). We take the log of these probabilities and construct a log-likelihood score (S) for each respondent's CMTD as follows:

$$S = \frac{l - l_{\min}}{l_{\max} - l_{\min}},$$ (3)

where

$$l = \log \prod_{i=1}^{N} h(x_i) = \sum_{i=1}^{N} \log h(x_i)$$ (4)

and

$$l_{\max} = n \log[\max_x h(x)] = n \log[h(x_n)]$$ (5)

and

$$l_{\min} = n \log[\min_x h(x)].$$ (6)

We repeat this process for each respondent and each historical time series. For each historical time series, we average the scores. This gives us an overall score indicating how well the historical time series reflects the respondents' beliefs. This overall score is what we refer to as the *log-likelihood score*.
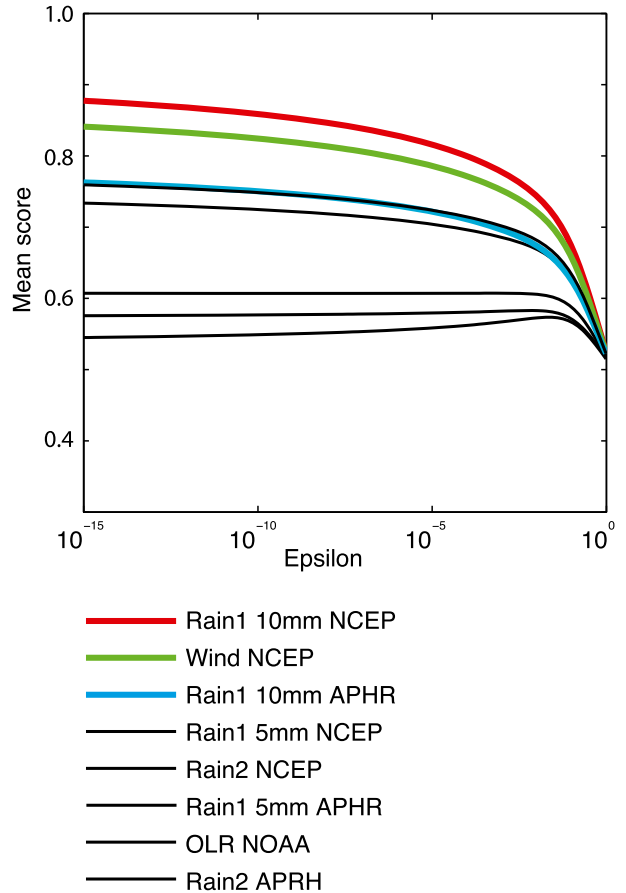

FIG. 4. Log-likelihood scores for each of the definition time series as ε is reduced toward zero. This analysis used the respondents' answers from Jessore. The top lines are color-coded for easier referencing.

Before we consider these scores in more detail, we should understand the impact of the size of the distribution's wings. In other words, we investigate how the choice of ε influences the analysis. To assess the effect of the value of ε, we calculate the log-likelihood scores for each historical time series at Jessore and different ε.

In Fig. 4, we see how the log-likelihood scores for each of the eight historical time series vary as we let ε go toward zero. The log-likelihood scores diverge and flatten for reduced ε, as we give progressively less weight to the dates that lie outside the beliefs of the respondents. Hence, the more we reduce ε, the easier it is to make a comparison between the final log-likelihood scores. Reducing ε toward zero also gives a probability mass function closer to a triangular distribution that best matches the respondents' beliefs. Following the arguments above, we set $\varepsilon = 10^{-15}$ in our time series simulations and log-likelihood scoring.

However, a time series containing 30 dates equal to x_E would get a perfect log-likelihood score of 1. Thus, even
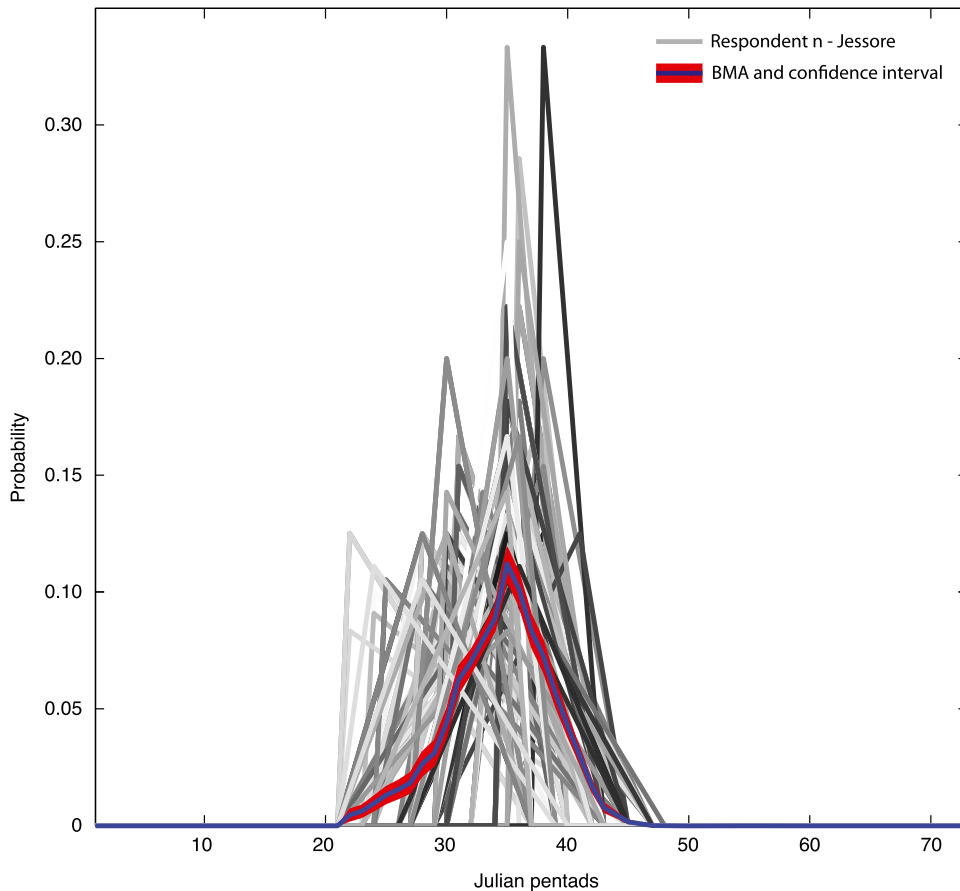
FIG. 5. The PMFs for each of the respondent's answers in Jessore (gray shading). The Bayesian model average (blue line) overlies with 95% confidence intervals shaded in red.

though a person might perceive variability in the monsoon onset date, the time series with a perfect score would have no variability. The log-likelihood score is therefore a good choice of comparison if we are most interested in how well the historical time series represents the respondent's perception of normal onset x_N rather than variability.

In the next section we present the results from the simulations and the log-likelihood score. We have eight historical time series to compare with up to 200 respondents at six different locations around Bangladesh. Presenting a detailed analysis of all locations will rapidly become convoluted, so we restrict our presentation arbitrarily to one example location: Jessore. However, we present and discuss the log-likelihood score results for all other locations.

4. Example: Jessore

Jessore is a district in western Bangladesh with a population of 2.8 million (2011 census). The district covers just over 2600 km^2 and nestles the Indian border to the west. The questionnaire survey (Stiller-Reeve et al. 2015) was carried out in the Monirampur Upazila in the eastern part of the Jessore District. Monirampur's main crops are paddy and wheat with its main export crop being dates.

a. Public perception

In the questionnaire, 190 people from Jessore answered questions about how they define the monsoon, as well as the timing of its onset. Figure 5 shows each respondent's CMTD covering their answers for the earliest, normal, and latest monsoon onset dates. We also calculated the Bayesian model average (Hoeting et al. 1999) from all the CMTDs to indicate overall perception. The Bayesian model average is calculated by the mean of the probabilities from all the respondents' CMTDs at each Julian pentad from 1 to 73. The Bayesian model average shows a peak at pentad 35. This corresponds to a date around 20 June. By comparing the Bayesian model average peak to the mean onsets in
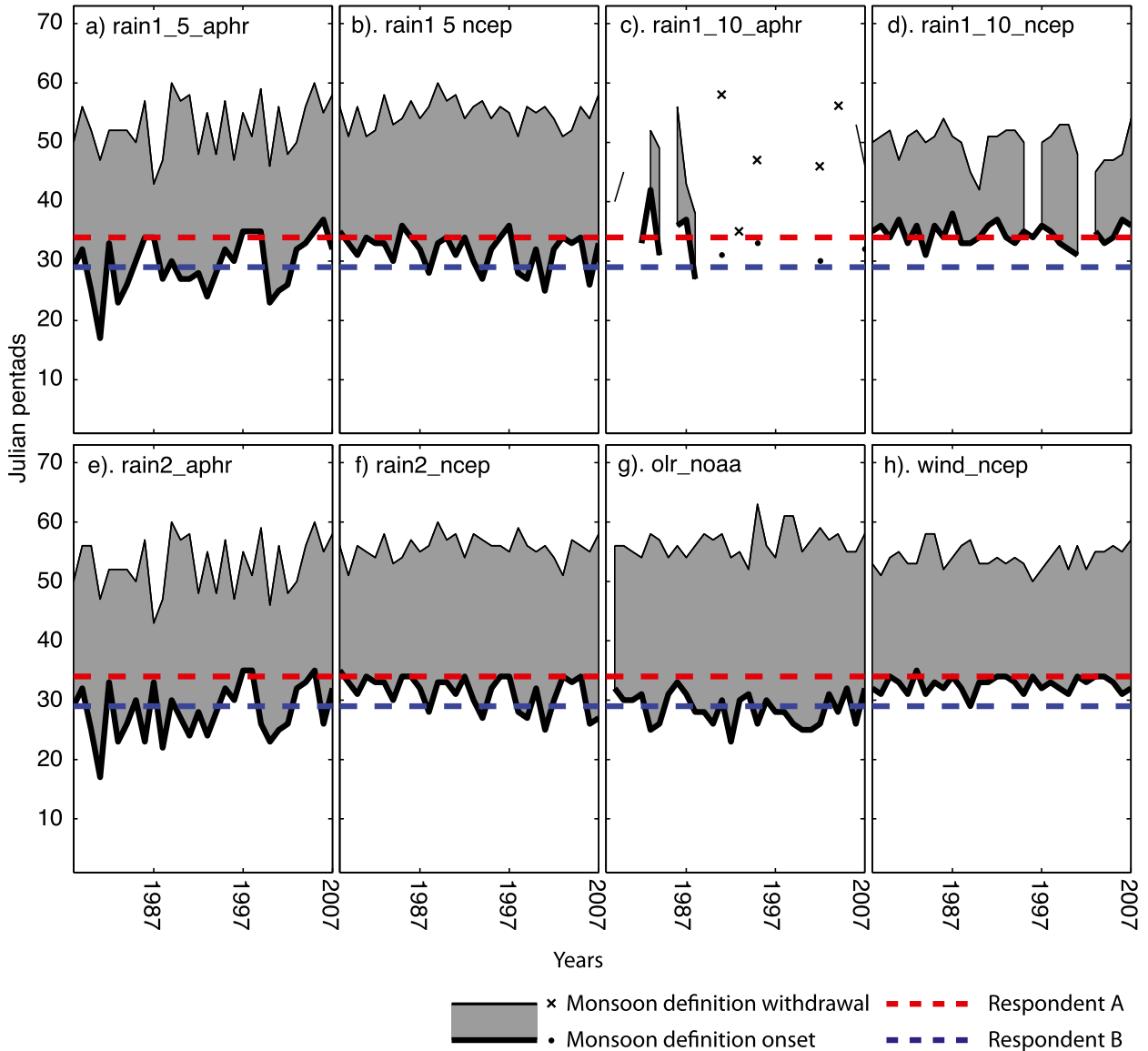
FIG. 6. Historical time series from the different monsoon definitions that we applied in this study. The graphs show both monsoon onset and withdrawal. The red and blue dashed lines show respondents A and B's answers for normal monsoon onset, respectively, to give a comparison. The graphs show the results from (a) rain1_5_aphr, (b) rain1_5_ncep, (c) rain1_10_aphr, (d) rain1_10_ncep, (e) rain2_aphr, (f) rain2_ncep, (g) olr_noaa, and (h) wind_ncep at the nearest grid point to Jessore.

Fig. 2, we see that the time series rain1_10_ncep initially gives the closest match.

b. Historical time series

Figure 6 shows the time series from each of the monsoon definitions (Table 1) from the closest grid point to Jessore. We present both the onset and withdrawal dates, as this shows that the definitions and the methodology can also be used for estimating monsoon withdrawal and therefore monsoon length. The rain1_10_aphr definition did not yield

satisfactory results under this framework, which could be due to an inappropriately high threshold for this dataset. Hence, we exclude the rain1_10_aphr results from further analysis in Jessore, because we are looking for time series to compare with the respondents' beliefs.

As an initial comparison we included the answers from respondents A and B for normal onset in Fig. 7. By visual comparison we see that rain1_10_ncep best matches the beliefs of respondent A, whereas olr_noaa seems to best reflect the beliefs of respondent B.
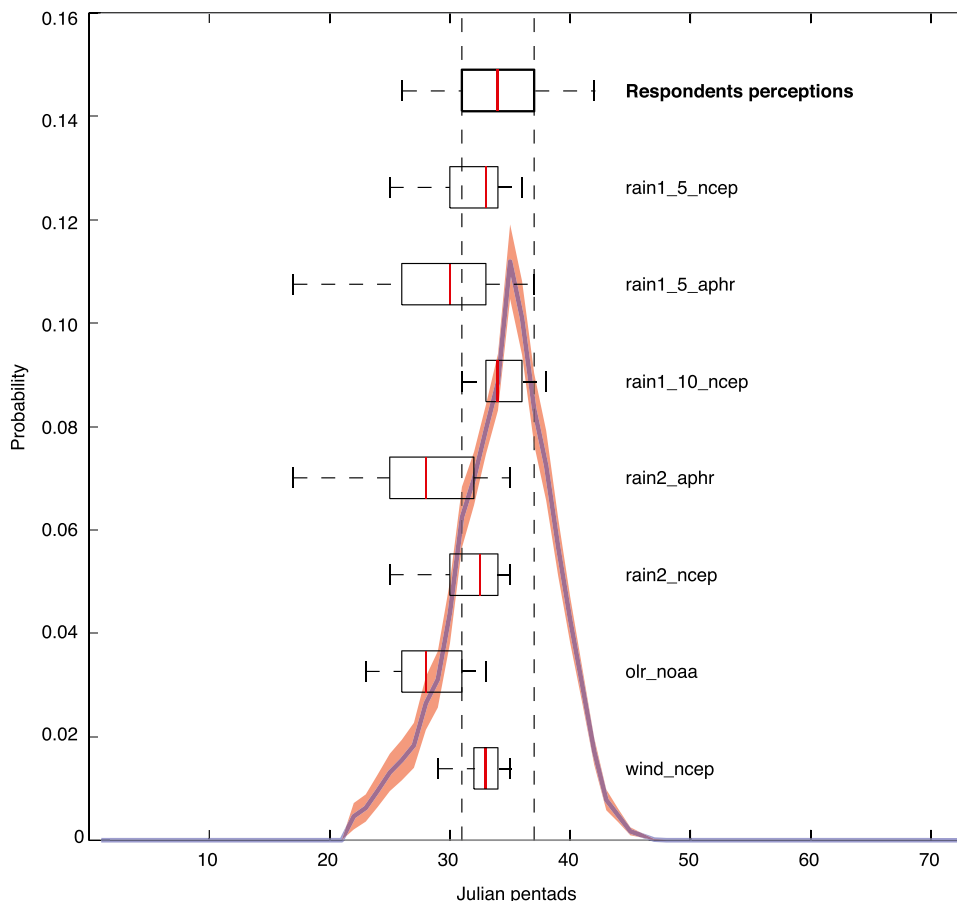
FIG. 7. Plot showing box-and-whisker plots from each of the historical time series (excluding rain1_10_aphr). For comparison, the box-and-whisker plot overlay the Bayesian model average of the PMFs of the respondents' beliefs. A simulated time series is generated from the Bayesian model average. The box-and-whisker plot from this time series is included at the top to facilitate the visual comparison.

c. Simulation

We simulate a time series from the Bayesian model average of the respondents' beliefs. With this time series, we construct a box-and-whisker plot that acts as a benchmark for the comparison with the historical time series. These plots, as shown in Fig. 7, give an initial impression of how the spread of the historical time series compare to the respondents' beliefs. We see large differences between the seven remaining historical time series. The rain1_10_ncep time series appears to give a good estimate of the mean, but it does not yield a large enough variability. To investigate this further, we analyze other values calculated from the time series simulations from each respondent.

By calculating the means and standard deviations from the respondents' simulated time series, we can directly compare to the equivalent values from the historical time series. Figure 8 shows these results for each

of the historical time series. Each of the black dots represents the simulated time series for each of the Jessore respondents. The large red dots represent the time series from the different historical time series. In the figure, the dots and lines resemble hedgehogs.

In the hedgehog plots, we notice the large spread in the respondents' beliefs with regard to the mean monsoon onset date and its variability. The respondents in Jessore perceive normal monsoon onsets anywhere between Julian pentad 26 and 39, with a standard deviation between 1 and 6 pentads. Despite this large spread, we notice a cluster around a mean onset of 34–36 Julian pentads and a standard deviation of 2–4 pentads. From visual inspection, the historical time series from rain1_5_ncep (Fig. 8a), rain2_aphr (Fig. 8d), and olr_noaa (Fig. 8f) have particularly early onsets and misrepresent the respondents' beliefs. Of the remaining historical time series, rain2_ncep (Fig. 9e) and rain1_5_ncep (Fig. 9b) give good estimates for standard deviation
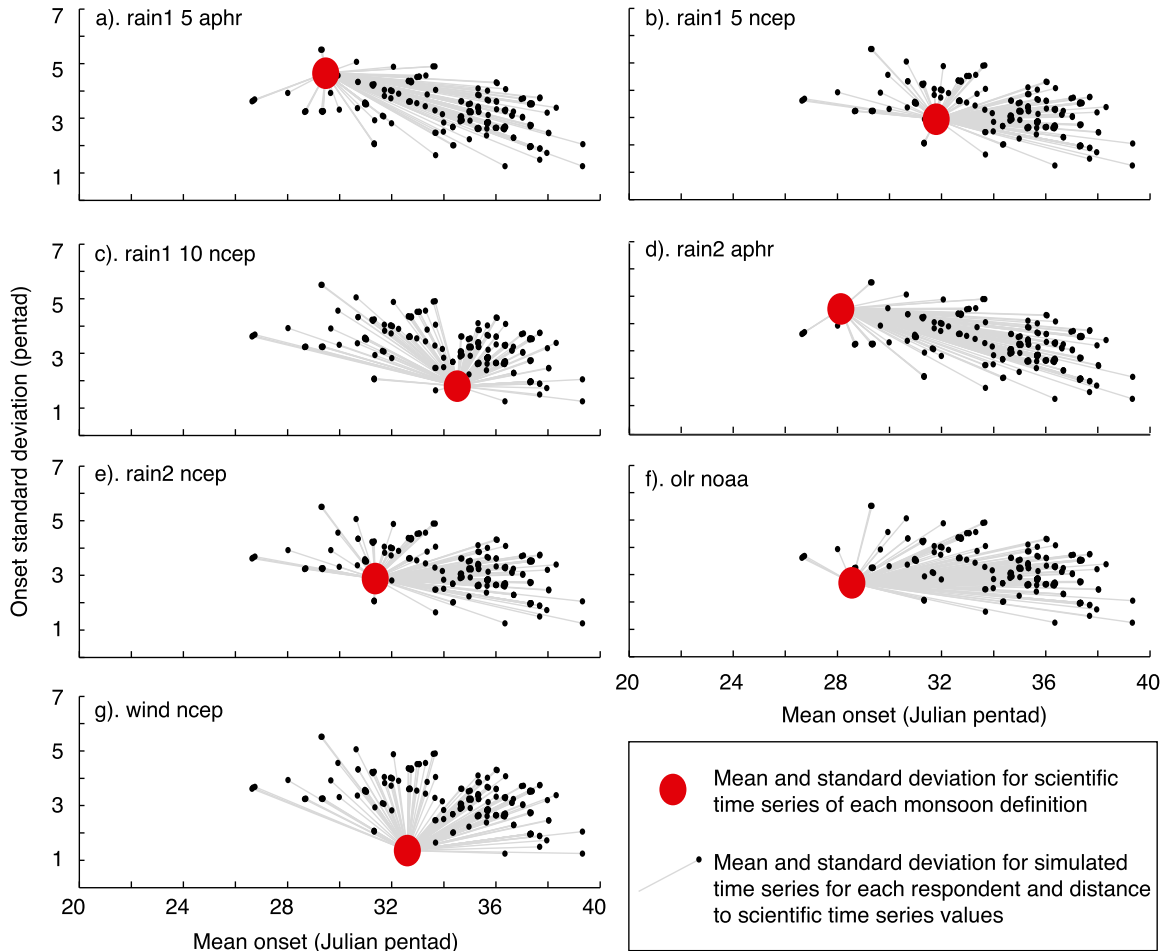
FIG. 8. Hedgehog plots showing how the mean and standard deviation of the historical time series relate to the same values from the time series simulated from the respondents' PMFs. The plots show the comparisons for (a) rain1_5_aphr, (b) rain1_5_ncep, (c) rain1_10_ncep, (d) rain2_aphr, (e) rain2_ncep, (f) olr_noaa, and (g) wind_ncep.

compared with the respondents' beliefs, but again the onsets seem too early. The rain1_10_ncep time series (Fig. 8c) gives a low estimate for standard deviation but a mean onset date comparable to the clustering of respondents' beliefs.

d. Log-likelihood score

1) JESSORE

All of the log-likelihood scores for the different historical time series are shown in Table 2 for Jessore. The scores quantify our speculation that the rain1_10_ncep time series compares well to the respondents' beliefs of normal onset date x_N. We previously acknowledged that the log-likelihood score rewarded time series that fluctuate least around the peak value of the probability mass functions. This is the case for the two historical time series that achieved the best log-likelihood scores:

rain1_10_ncep and wind_ncep. Both of these time series have means close to the Bayesian model average peak (see Fig. 7), but they also feature much less variability than the respondents' overall beliefs. As we descend in Table 2 and concomitantly compare with the box-and-whisker plots in Fig. 7, we realize that the mean values for the historical time series move progressively farther from the peak in the Bayesian model average. This happens despite some of the historical time series having standard deviations that compare well with the respondents' beliefs. The log-likelihood score is preferable to use if the mean onset is most important rather than the variability.

2) OTHER LOCATIONS

The highest log-likelihood scores for all six locations are displayed in Fig. 9. The figure reflects the diversity with respect to the monsoon onset in Bangladesh. No
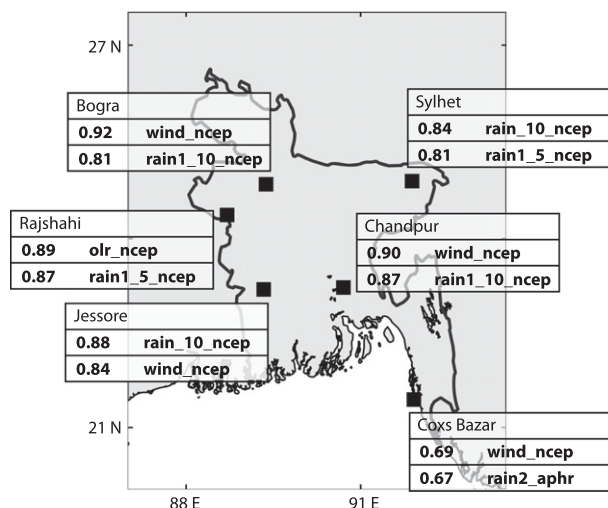
FIG. 9. Map showing the top two log-likelihood scores at each questionnaire location and the corresponding historical time series.

TABLE 2. Log-likelihood scores for each of the historical monsoon onset time series in Jessore. The higher the log-likelihood score (max 1), the better the comparison with the people's beliefs.

Historical time series	Log-likelihood score
rain1_10_ncep	0.88
wind_ncep	0.84
rain1_5_ncep	0.76
rain2_ncep	0.73
rain1_5_aphr	0.61
olr_noaa	0.58
rain2_aphr	0.54

single historical time series is noticeably better than the others. The rain1_10_ncep time series best matches the respondents' beliefs in Jessore and Sylhet, but it comes second in Bogra and Chandpur. This could indicate that a 10-mm rainfall threshold could be more appropriate for defining the monsoon in Bangladesh, at least when using Reanalysis-1 data. The wind-based monsoon definition (wind_ncep) comes out on top at three locations: Bogra, Chandpur, and Cox's Bazar. As we see from Figs. 7 and 8, the wind_ncep time series shows the least variability of all the historical time series. Therefore, if the wind_ncep time series fluctuates around the Bayesian model average peak for any location, then it would likely get the best log-likelihood score.

5. Conclusions

Our main objective was to develop tools to facilitate the comparison between climate time series datasets and local beliefs. We are motivated by climate services and the goal to transform climate research into practical applications and mitigation strategies. To achieve this, we have to evaluate how well science and the beliefs of stakeholders align. We have therefore developed methods to compare different scientific monsoon onset definitions with the beliefs of local people in Bangladesh, using Jessore as an example.

We generated eight historical time series of the monsoon onset, applying different scientific onset definitions to different datasets. Each of the time series covered the period 1978–2007. In addition to the historical time series, we also had data on the local people's beliefs of the earliest, normal, and latest monsoon onset

dates. These data were collected during a questionnaire survey in 2011 that covered six regions in Bangladesh. Our challenge was to find ways to compare the three answers respondents gave with historical time series of monsoon onsets.

We can find different ways to make this comparison if we first construct a probability mass function around the beliefs of the survey respondents. Here we argued that the triangular distribution was an appropriate point of departure. In particular, the distribution is based on three parameters that correspond with the three answers the Bangladeshi respondents gave. We had to adapt the distribution to make it circular and unable to take zero probabilities. We called the resulting distribution a circular modified triangular distribution (CMTD). The CMTDs represented the beliefs of the respondents and could be used to compare with the historical time series of monsoon onsets: 1) time series simulation and 2) log-likelihood scores.

From the CMTDs we simulated artificial time series. The properties of these simulations could be compared directly with equivalent properties from the different historical time series. We decided to calculate and compare the mean and standard deviation. By illustrating these values in a hedgehog plot (Fig. 8), we could qualitatively evaluate how the historical time series compare with the respondents' beliefs. For the Jessore example, most of the historical time series estimated monsoon onsets much earlier than most of the respondents' beliefs. The historical time series rain1_10_ncep used a threshold of $10\,\mathrm{mm\,day}^{-1}$ and seemed to match the people's beliefs best, at least with regard to mean monsoon onset.

Even though we evaluated the results from the simulations visually, we also aimed to develop a *measure* for the comparison between the historical time series and respondents' beliefs. To measure the comparison, we applied log-likelihoods and constructed a score. The score for each of the historical time series indicated how well they corresponded to the respondents' beliefs. The log-likelihood score supported the assumptions from the

simulations that the mean of rain1_10_ncep was particularly good, despite the standard deviation being too low.

The comparisons we have made in this study relied upon the application of CMTDs around the beliefs of the survey respondents. Using the CMTDs, we have demonstrated two different ways to evaluate how well different empirical monsoon definitions correspond to local belief. With regard to Bangladesh specifically, our results present some challenges. We showed that scientific definitions of the monsoon onset vary widely, as do the people's perceptions. In general, the subjective beliefs of individuals will never perfectly match scientific data definitions, and hence methods such as those proposed in this study are required to understand this gap. Applying the CMTD aided the comparison between science and belief, pinpointing stark contrasts among different definitions. However, if these methods are to be used in future climate services, it will also be crucial to gain a deeper understanding of the decision-making processes of the users in relation to the monsoon onset and to integrate the CMTD results back into a continuous dialogue with the stakeholders or users in question. In such a dialogue, it will likely emerge that some of the stakeholders' perceptions are incorrect, just as some of the scientific definitions are unusable. The CMTD methods and tools could be adapted to exclude some perceptions or to weigh others, allowing the dialogue to focus toward a robust result. The methods could also be used for other recurrent meteorological phenomena where scientific definitions and stakeholder perceptions may diverge.

REFERENCES

Ahmed, A., 1994: Variability of rainfall and duration of the summer monsoon, and their relationships with monsoon onset dates in Bangladesh. Proceedings of the International Conference on Monsoon Prediction and Variability, WMO/TD-619, WCRP-84, 320–327.

Ahmed, R., and S. Karmakar, 1993: Arrival and withdrawal dates of the summer monsoon on Bangladesh. *Int. J. Climatol.*, **13**, 727–740, doi:10.1002/joc.3370130703.

Ashfaq, M., Y. Shi, W.-W. Tung, R. J. Trapp, X. Gao, J. S. Pal, and N. S. Diffenbaugh, 2009: Suppression of south Asian summer monsoon precipitation in the 21st century. *Geophys. Res. Lett.*, **36**, L01704, doi:10.1029/2008GL036500.

Back, W. E., W. W. Boles, and G. T. Fry, 2000: Defining triangular probability distributions from historical cost data. *J. Constr. Eng. Manage.*, **126**, 29–37, doi:10.1061/(ASCE)0733-9364(2000)126:1(29).

Briand, L. C., B. Freimut, and F. Vollei, 2000: Assessing the cost-effectiveness of inspections by combining project data and expert opinion. *Proceedings: 11th International Symposium on Software Reliability Engineering; ISSRE 2000*, IEEE, 124–135, doi:10.1109/ISSRE.2000.885866.

Brooks, M. S., 2013: Accelerating innovation in climate services: The 3 E's for climate service providers. *Bull. Amer. Meteor. Soc.*, **94**, 807–819, doi:10.1175/BAMS-D-12-00087.1.

Buontempo, C., C. D. Hewitt, F. J. Doblas-Reyes, and S. Dessai, 2014: Climate service development, delivery and use in Europe at monthly to inter-annual timescales. *Climate Risk Manage.*, **6**, 1–5, doi:10.1016/j.crm.2014.10.002.

Fontaine, B., S. Louvet, and P. Roucou, 2008: Definition and predictability of an OLR-based West African monsoon onset. *Int. J. Climatol.*, **28**, 1787–1798, doi:10.1002/joc.1674.

Gadgil, S., and K. R. Kumar, 2006: The Asian monsoon—Agriculture and economy. *The Asian Monsoon*, Environmental Sciences, Springer, 651–683.

Galway, L. A., 2007: Subjective probability distribution elicitation in cost risk analysis: A review. Tech. Rep. 410, Rand Corporation, 21 pp.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky, 1999: Bayesian model averaging: A tutorial. *Stat. Sci.*, **14**, 382–417.

Johnson, D., 1997: The triangular distribution as a proxy for the beta distribution in risk analysis. *Statistician*, **46**, 387–398, doi:10.1111/1467-9884.00091.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kotz, S., and J. R. van Dorp, 2004: *Beyond Beta: Other Continuous Families of Distributions with Bounded Support and Applications*. World Scientific Publishing, 308 pp.

Lemos, M. C., 2008: What influences innovation adoption by water managers? Climate information use in Brazil and the United States. *J. Amer. Water Resour. Assoc.*, **44**, 1388–1396, doi:10.1111/j.1752-1688.2008.00231.x.

Liebmann, B. C., and A. Smith, 1996: Description of a complete (interpolated) outgoing longwave radiation dataset. *Bull. Amer. Meteor. Soc.*, **77**, 1275–1277.

Marimo, P., T. R. Kaplan, K. Mylne, and M. Sharpe, 2015: Communication of uncertainty in temperature forecasts. *Wea. Forecasting*, **30**, 5–22, doi:10.1175/WAF-D-14-00016.1.

Marx, S. M., E. U. Weber, B. S. Orlove, A. Leiserowitz, D. H. Krantz, C. Roncoli, and J. Phillips, 2007: Communication and mental processes: Experiential and analytic processing of uncertain climate information. *Global Environ. Change*, **17**, 47–58, doi:10.1016/j.gloenvcha.2006.10.004.

Matsumoto, J., 1997: Seasonal transition of summer rainy season over Indochina and adjacent monsoon region. *Adv. Atmos. Sci.*, **14**, 231–245, doi:10.1007/s00376-997-0022-0.

Miles, E. L., A. K. Snover, L. C. W. Binder, E. S. Sarachik, P. W. Mote, and N. Mantua, 2006: An approach to designing a national climate service. *Proc. Natl. Acad. Sci. USA*, **103**, 19 616–19 623, doi:10.1073/pnas.0609090103.

Moss, R. H., and Coauthors, 2013: Hell and high water: Practice-relevant adaptation science. *Science*, **342**, 696–698, doi:10.1126/science.1239569.

Pennesi, K., 2007: Improving forecast communication: Linguistic and cultural considerations. *Bull. Amer. Meteor. Soc.*, **88**, 1033, doi:10.1175/BAMS-88-7-1033.

Ramage, C. S., 1995: Forecasters guide to tropical meteorology. AWS TR 240 Updated, Tech. Rep., Air Weather Service Scott AFB, 493 pp.

Ray, A. J., G. M. Garfin, M. Wilder, M. Vásquez-León, M. Lenart, and A. C. Comrie, 2007: Applications of monsoon research: Opportunities to inform decision making and reduce regional vulnerability. *J. Climate*, **20**, 1608–1627, doi:10.1175/JCLI4098.1.

Rosenzweig, M. R., and H. P. Binswanger, 1993: Wealth, weather risk and the composition and profitability of agricultural investments. *Econ. J.*, **103**, 56–78, doi:10.2307/2234337.

Stiller-Reeve, M. A., T. Spengler, and P.-S. Chu, 2014: Testing a flexible method to reduce false monsoon onsets. *PLoS One*, **9**, e104386, doi:10.1371/journal.pone.0104386.

——, A. Syed, T. Spengler, J. A. Spinney, and R. Hossain, 2015: Complementing scientific monsoon definitions with social perception in Bangladesh. *Bull. Amer. Meteor. Soc.*, **96**, 49–57, doi:10.1175/BAMS-D-13-00144.1.

Vally, H., and Coauthors, 2014: Proportion of illness acquired by foodborne transmission for nine enteric pathogens in Australia: An expert elicitation. *Foodborne Pathog. Dis.*, **11**, 727–733, doi:10.1089/fpd.2014.1746.

Visbeck, M., 2008: From climate assessment to climate services. *Nat. Geosci.*, **1**, 2–3, doi:10.1038/ngeo.2007.55.

Wang, B., and LinHo, 2002: Rainy season of the Asian–Pacific summer monsoon. *J. Climate*, **15**, 386–398, doi:10.1175/1520-0442(2002)015<0386:RSOTAP>2.0.CO;2.

——, Q. Ding, and P. V. Joseph, 2009: Objective definition of the Indian summer monsoon onset. *J. Climate*, **22**, 3303–3316, doi:10.1175/2008JCLI2675.1.

Weber, E. U., 2006: Experience-based and description-based perceptions of long-term risk: Why global warming does not scare us (yet). *Climatic Change*, **77**, 103–120, doi:10.1007/s10584-006-9060-3.

Yamane, Y., T. Hayashi, A. M. Dewan, and F. Akter, 2010: Severe local convective storms in Bangladesh: Part I. Climatology. *Atmos. Res.*, **95**, 400–406, doi:10.1016/j.atmosres.2009.11.004.

Yang, X., T. Yao, W. Yang, B. Xu, Y. He, and D. Qu, 2012: Isotopic signal of earlier summer monsoon onset in the Bay of Bengal. *J. Climate*, **25**, 2509–2516, doi:10.1175/JCLI-D-11-00180.1.

Yatagai, A., K. Kamiguchi, O. Arakawa, A. Hamada, N. Yasutomi, and A. Kitoh, 2012: APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Bull. Amer. Meteor. Soc.*, **93**, 1401–1415, doi:10.1175/BAMS-D-11-00122.1.

Zeng, X., and E. Lu, 2004: Globally unified monsoon onset and retreat indexes. *J. Climate*, **17**, 2241–2248, doi:10.1175/1520-0442(2004)017<2241:GUMOAR>2.0.CO;2.

Zhang, H., 2010: Diagnosing Australia-Asian monsoon onset/retreat using large-scale wind and moisture indices. *Climate Dyn.*, **35**, 601–618, doi:10.1007/s00382-009-0620-x.

Zhang, Z., J. C. L. Chan, and Y. Ding, 2004: Characteristics, evolution and mechanisms of the summer monsoon onset over Southeast Asia. *Int. J. Climatol.*, **24**, 1461–1482, doi:10.1002/joc.1082.