

17. SEVERE PRECIPITATION IN NORTHERN INDIA IN JUNE 2013: CAUSES, HISTORICAL CONTEXT, AND CHANGES IN PROBABILITY

DEEPTI SINGH, DANIEL E. HORTON, MICHAEL TSIANG, MATZ HAUGEN, MOETASIM ASHFAQ, RUI MEI, DEEKSHA RASTOGI, NATHANIEL C. JOHNSON, ALLISON CHARLAND, BALA RAJARATNAM, AND NOAH S. DIFFENBAUGH

Cumulative precipitation in northern India in June 2013 was a century-scale event, and evidence for increased probability in the present climate compared to the preindustrial climate is equivocal.

The Event: June 2013 Flooding in Northern India. Parts of mountainous northern India—including Himachal Pradesh, Uttarakhand, and Uttar Pradesh—experienced extremely heavy precipitation during 14–17 June 2013 (Fig. 17.1a,b). Landslides, debris flows, and extensive flooding caused catastrophic damage to housing and infrastructure, impacted >100000 people, and resulted in >5800 deaths (Dobhal et al. 2013; Dube et al. 2013; Dubey et al. 2013; Joseph et al. 2014; Mishra and Srinivasan 2013). Subsequent

heavy rains on 24–25 June hampered rescue efforts, ultimately leaving thousands without food or shelter for >10 days (Prakash 2013).

Causes of the mid-June precipitation and associated flooding have been analyzed in detail (Dobhal et al. 2013; Dube et al. 2014; Mishra and Srinivasan 2013; Prakash 2013). Anomalously early arrival of monsoon-like atmospheric circulation over India (Fig. 17.1c, Supplementary Figure S17.1a) brought heavy rains to the mountainous regions where snow

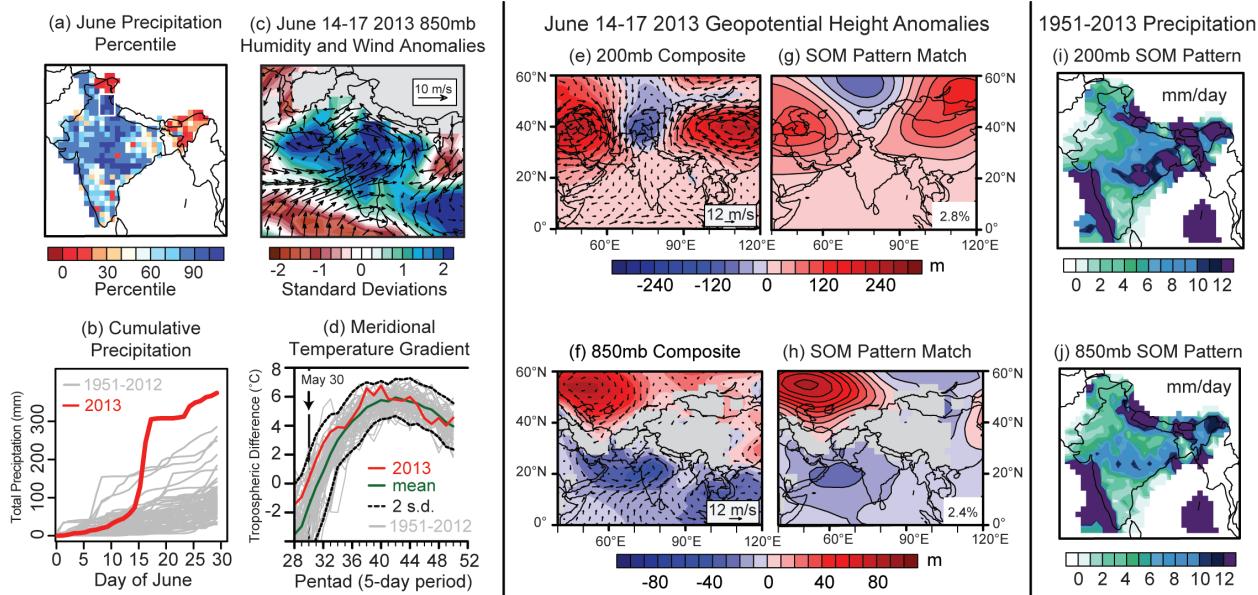


FIG. 17.1. Precipitation characteristics and synoptic environment. (a) June 2013 grid cell cumulative precipitation percentiles relative to June climatology (1951–2012). White box highlights the severe flooding domain (29° – 33° N, 77.5° – 80° E). (b) Daily cumulative precipitation distribution over the flood domain. (c) 14–17 June 2013 composite lower-level wind and specific humidity anomalies relative to 14–17 June climatology. (d) Climatological and 2013 meridional temperature gradient (MTG), defined as the zonally averaged (52° – 85° E) pentad mean tropospheric (200–500 mb) temperature difference between 30° N and 5° N. (e,f) 14–17 June 2013 composite upper- and lower-level wind and geopotential height anomalies relative to the 14–17 June climatology. (g,h) Upper- and lower-atmosphere self-organizing map (SOM) patterns that correspond to 14–17 June 2013. Pattern matches are autonomously selected from 35 SOM nodes, generated from an analysis of all 1951–2013 June days (see Supplemental Materials). (i,j) Composite precipitation for all June days during the 1951–2013 period that were associated with the upper- and lower-level SOM patterns shown in (g) and (h).

cover typically melts prior to monsoon onset (Dube et al. 2014; Joseph et al. 2014). Snow cover in local river basins was ~30% above normal in early June 2013 (Durga Rao et al. 2014). Heavy precipitation led to rapid snowmelt, overwhelming the regional hydrologic system, causing glacial lake outburst floods, and triggering catastrophic mass wastage events (Andermann et al. 2012; Dubey et al. 2013; Durga Rao et al. 2014; Prakash 2013; Siderius et al. 2013).

The upper- and lower-level synoptic conditions in early and mid-June supported the anomalously early monsoon-like circulation (Supplementary Fig. S17.1a) and excessive precipitation in northern India (Fig. 17.1a,b). In the upper atmosphere (200 mb), a persistent anticyclonic anomaly formed over Central Asia (Fig. 17.1e). This upper-level blocking pattern guided mid-to-high-latitude troughs southward, thereby facilitating the advection of relatively cold, dry, high-potential-vorticity air to the upper levels of the atmosphere over northern India (Joseph et al. 2014). In the lower atmosphere (850 mb), low-pressure systems formed over both the northern Bay of Bengal and the northern Arabian Sea (Joseph et al. 2014), with the Bay of Bengal system moving inland over central India and persisting for the duration of the event (Fig. 17.1f). Low-level convergence associated with these systems and a stronger-than-normal Somali Jet facilitated anomalous moisture advection to the Indian subcontinent (Fig. 17.1c). These co-occurring upper- and lower-level dynamics are consistent with a convectively unstable atmosphere (Hong et al. 2011; Ullah and Shouting 2013; Wang et al. 2011), which, when combined with orographic forcing from the surrounding northwestern Himalayan terrain, create an environment ripe for intense mesoscale convection (Houze et al. 2011).

In this study, we analyze the dynamics of this event within the context of the historical and preindustrial climates.

Historical context. We contextualize June 2013 precipitation using the Indian Meteorological Department (IMD) 1951–2013 $1^\circ \times 1^\circ$ gridded dataset (Rajeevan et al. 2010), with the caveat that the rain gauge network in the region could have changed over this period. Cumulative June precipitation exceeded the 80th percentile over much of central and northern India, and it exceeded the maximum quantile over a majority of the flood region (Fig. 17.1a). From 14 to 17 June, this domain (29° – 33° N, 77.5° – 80° E) received four-day total precipitation that was unprecedented in the observed record (Fig. 17.1b), with

the heaviest day (16 June) exceeding the previous one-day June maximum by 105% (Supplementary Fig. S17.2). Consequently, the flood region recorded the highest total accumulated June precipitation in the 1951–2013 record, with the previous maximum June total equaled by 17 June and exceeded by 31% by the end of the month (Fig. 17.1b).

Monsoon dynamics and thermodynamics were also unusual relative to June climatological norms. The monsoon onset date is closely associated with the reversal of the zonally averaged (52° – 85° E) meridional tropospheric (500–200 mb) ocean-to-continent (5° – 30° N) temperature gradient (Ashfaq et al. 2009; Webster et al. 1998), and with the vertical easterly zonal wind shear between 850 mb and 200 mb averaged over 0° – 30° N and 50° – 90° E (Li and Yanai 1996; Webster et al. 1998; G. Wu et al. 2012; Xavier et al. 2007). The 2013 meridional temperature gradient (MTG) reversal dates were among the earliest on record (1951–2013, Fig. 17.1d) and the vertical easterly wind shear was stronger than normal during early-June (Supplementary Fig. S17.1b). The early MTG reversal resulted from anomalously high land temperatures (~2 standard deviations; Supplementary Fig. S17.1c,d), which co-occurred with record-low Eurasian snow cover (NOAA 2013). In addition, as a result of the early monsoon-like circulation, low-level atmospheric humidity exceeded 2 standard deviations above the climatological 14–17 June mean (Fig. 17.1c).

Synoptic conditions were likewise extremely rare for mid-June. We categorize the occurrence of upper- and lower-level daily June atmospheric patterns in the National Centers for Environmental Prediction (NCEP) R1 reanalysis using self-organizing map (SOM) cluster analysis (Borah et al. 2013; Chattopadhyay et al. 2008; Hewitson and Crane 2002; Johnson 2013; Kohonen 2001; see Supplemental Materials). SOM analyses reveal persistent upper-level blocking patterns from 10 to 17 June and lower-level troughing patterns from 11 to 17 June (Supplementary Fig. S17.2). Additionally, the upper- and lower-level patterns (Fig. 17.1g,h) that persisted during the core of the event (14–17 June) are each historically associated with heavy precipitation over northern India (Fig. 17.1i,j). Although occurrence of the core-event upper-level pattern is not rare for June (median frequency of occurrence), the 850-mb pattern is much less common (<6 percentile frequency of occurrence). Further, mid-June 2013 was the only instance that the core-event upper- and lower-level patterns co-occurred in June during the 1951–2013 period. The atmospheric

configuration associated with the unprecedented mid-June extreme precipitation, therefore, appears to also have been unprecedented.

We note that this configuration is not necessarily unprecedented later in the monsoon season. For example, the co-occurrence of upper-level blocking with tropical moisture advection is similar to the conditions identified during the July 2010 Pakistan floods and during heavy precipitation events that occur during the core monsoon season (Hong et al. 2011; Houze et al. 2011; Lau and Kim 2011; Ullah and Shouting 2013; Webster et al. 2011).

Quantifying the probability of a 2013-magnitude event. In quantifying the probability of a 2013-magnitude event, we restrict our focus to the June 2013 total precipitation. We select the monthly scale extreme rather than the daily scale extreme because both the extreme magnitude of this event relative to the observed distribution of four-day June totals and the limited ability of climate models to accurately simulate the daily scale extremes make the problem intractable at the daily scale. Therefore, hereafter, “a 2013-magnitude event” refers to the total June rainfall, which in June 2013 was the most extreme on record (Fig. 17.1b).

Given the rarity of the event in the observed record (Fig. 17.2a), we fit a Pareto (heavy-tailed) distribution to the 1951–2012 observations of spatially averaged (area-weighted average) rainfall over the selected domain (Fig. 17.2a; Supplementary Fig. S17.3a). From the Pareto distribution, we estimate the sample quantile (Q_o) and return period (R_o) of the June 2013 total precipitation in the present climate (see Supplemental Materials). We find that the 2013 event exceeds the 99th percentile in the observed distribution ($Q_o = 99.1$ th quantile), yielding a return period of 111 years (Fig. 17.2a). Because the Pareto is a heavy-tailed distribution, extreme events are less likely to be found anomalous, and, thus, the corresponding return period can be considered a lower bound.

Next, we assess the influence of anthropogenic forcings on the likelihood of extreme June precipitation using the historical (20C) and preindustrial (PI) simulations from the CMIP5 climate model archive (Taylor et al. 2012). We use the Kolmogorov–Smirnov (K-S) goodness-of-fit test to identify the models that most closely simulate the observed distribution of the area-weighted average June total precipitation over the impacted region (Fig. 17.1a). (To control for the mean bias in the models, we first re-center each model’s distribution so that the model mean matches the observed mean.) Because the simulated change in

likelihood of extremes can be heavily influenced by biases in the simulated distribution, we restrict our analyses to 11 models whose K-S value exceeds 0.2 (Supplementary Fig. S17.3b), ensuring a comparatively good fit of the overall distribution, including in the tails. Among these 11 models that pass this goodness-of-fit criterion, 4 show greater mean and variability of June precipitation in the 20C simulations (Fig. 17.2b). However, 7 of the 11 show increased exceedance of the PI 99th percentile value (Fig. 17.2c), suggesting increased probability of extremely high June precipitation in the current climate. This result is consistent with studies that indicate an increase in extremes primarily from increased atmospheric-moisture availability (Allan and Soden 2008; O’Gorman and Schneider 2009).

Next, we use Pareto distributions to estimate the return period of the June 2013 total precipitation in the 20C and PI simulations. To control for the variability-bias in the models, we first determine the magnitude of the 111-year event ($Q_o = 99.1$ th quantile) in the fitted 20C distribution (Pr_H), and then determine the quantile (Q_{PI}) corresponding to Pr_H in the fitted PI distribution (see Supplemental Materials; Supplementary Fig. S17.3c). Further, we quantify the uncertainty in these likelihood estimates (Q_o/Q_{PI}) using the bootstrap (Fig. 17.2d). We find that 5 of the 11 models show >50% likelihood that the extreme June total precipitation has higher probability in the 20C climate. In addition, of the three models that have high p -values from the K-S test (> 0.8) and similar sample sizes in the 20C and PI populations (Fig. 17.2d), two suggest >50% likelihood that the extreme June total precipitation has higher probability in the 20C climate, and the third model suggests ~50% likelihood. Further, the model with the largest 20C ensemble (Centre National de Recherches Meteorologiques Coupled Global Climate Model; CNRM-CM5) demonstrates a ~50% likelihood that the probability of the extreme June total precipitation has at least doubled in the 20C climate. CNRM-CM5 also has the highest skill in simulating the summer monsoon precipitation and lower-level wind climatology (Sperber et al. 2013).

Conclusions. Our statistical analysis, combined with our diagnosis of the atmospheric environment, demonstrates that the extreme June 2013 total precipitation in northern India was at least a century-scale event. Precise quantification of the likelihood of the event in the current and preindustrial climates is limited by the relatively short observational record

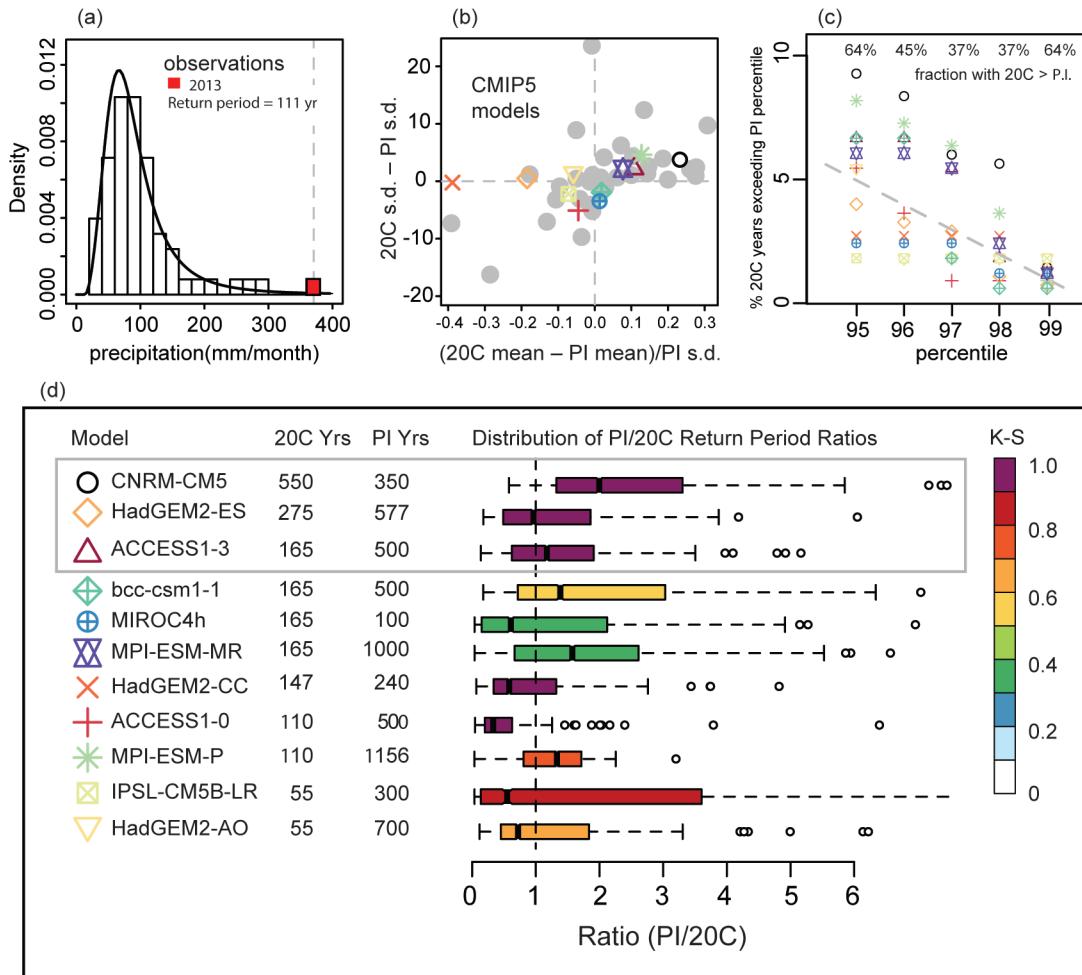


FIG. 17.2. Extreme precipitation statistics in the current and preindustrial climates. (a) Probability density function of the Pareto-fitted observed cumulative-June precipitation distribution (black line; 1951–2012), and probability of occurrence of the June 2013 cumulative precipitation magnitude in this distribution (red). The return period of the June 2013 magnitude in the observed distribution is indicated on the plot. (b) Change in mean and standard deviation of precipitation between the CMIP5 historical (20C) and preindustrial (PI) simulations. Gray dots represent all available CMIP5 models and colored symbols represent AI models that meet the Kolmogorov–Smirnov (K-S) goodness-of-fit test criteria (p value > 0.2). (c) Percent of years in the 20C simulations of AI models that exceed the respective PI quantiles of the AI models. The numbers on the plot indicate the fraction of AI models that exceed the PI quantiles in the 20C simulations. (d) Box plot representing the distribution of ratios of the return period of a June 2013 magnitude event in the PI and 20C simulations, calculated using the bootstrap. The lines in the boxes represent the median of the distribution for each model. The bounds of the boxes represent the 25th and 75th percentiles. The whiskers extend to the edges of 1.5×interquartile range, and points outside of those bounds are shown individually. The number of years indicated for the 20C Yrs and PI Yrs columns are the total years available from all realizations within each scenario. The color bar corresponding to the box plot indicates p values from the Kolmogorov–Smirnov test.

and by the resolution and ensemble size of the small subset of models that credibly simulate the seasonal rainfall distribution over northern India. Indeed, an attempt to quantify the probability of the unprecedented four-day precipitation total would present even greater analytical challenges. However, despite these limitations, our analyses of the observed and

simulated June precipitation provide evidence that anthropogenic forcing of the climate system has increased the likelihood of such an event, a result in agreement with previous studies of trends in rainfall extremes in India (Goswami et al. 2006; Krishnamurthy et al. 2009; Ghosh et al. 2012; Singh et al. 2014).

S17. SEVERE PRECIPITATION IN NORTHERN INDIA IN JUNE 2013: CAUSES, HISTORICAL CONTEXT, AND CHANGES IN PROBABILITY

DEEPTI SINGH, DANIEL E. HORTON, MICHAEL TSIANG, MATZ HAUGEN, MOETASIM ASHFAQ, RUI MEI, DEEKSHA RASTOGI, NATHANIEL C. JOHNSON, ALLISON CHARLAND, BALA RAJARATNAM, AND NOAH S. DIFFENBAUGH

Datasets. For the precipitation analysis, we use the $1^\circ \times 1^\circ$ gridded daily dataset from the Indian Meteorological Department (IMD; Rajeevan et al. 2006, 2010). This dataset is developed from approximately 2140 rain gauge stations over India, and it has been extensively used in literature to study characteristics of the Indian Monsoon (Dash et al. 2009; Ghosh et al. 2012; Singh 2013). As shown in Fig. 1a of Rajeevan et al. (2006), all stations included in the development of this dataset have at least 90% data availability within the observational period.

Daily atmospheric variables—including geopotential heights, atmospheric winds, and specific humidity—are obtained from the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis (R1) datasets at $2.5^\circ \times 2.5^\circ$ spatial resolution (Kalnay et al. 1996). In order to match the period of record of the IMD precipitation dataset, we analyze the 1951–2013 period in the reanalysis.

Self-organizing maps methodology. The self-organizing map (SOM) is a neural network-based cluster analysis, similar to k-means clustering, that partitions a high-dimensional dataset into a smaller number of representative clusters (Kohonen 2001). In contrast with conven-

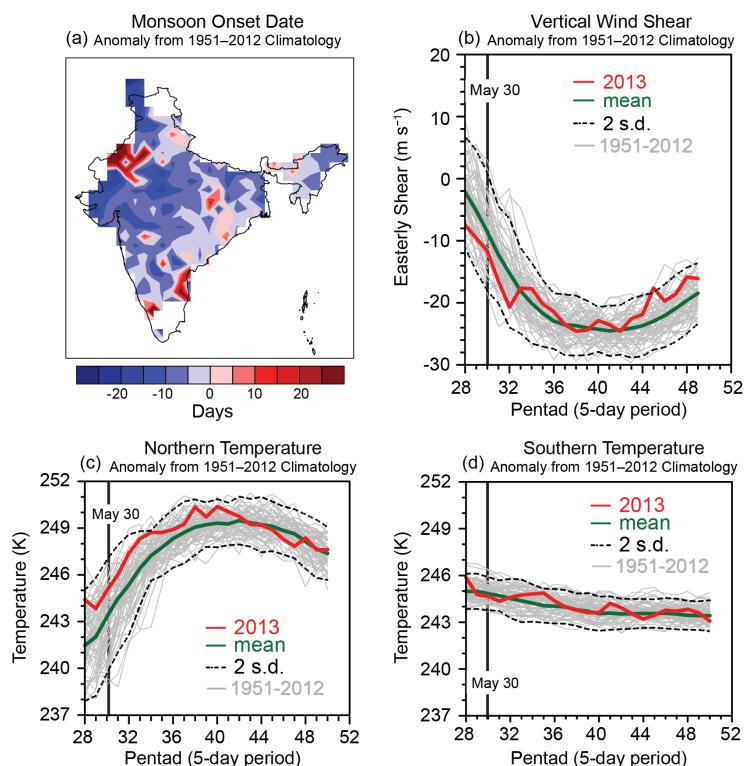


FIG. S17.1. June 2013 monsoon dynamics. (a) June 2013 monsoon onset anomalies from 1951–2012 climatology at each grid-cell. The onset date at each grid-cell is defined as the 5-day mean precipitation exceeding the January mean precipitation of that year and 5 mm day^{-1} . Negative anomalies over most of central and northwestern India suggest an early monsoon onset in June 2013. (b) June 2013 vertical easterly zonal wind shear. The easterly zonal wind shear is defined as the difference in winds between the upper (200-mb) and lower (850-mb) atmosphere averaged over the domain ($0^\circ\text{--}30^\circ\text{N}$, $50^\circ\text{--}90^\circ\text{E}$). The vertical wind shear was $\sim 2\sigma$ above the 1951–2012 climatological mean prior to the June 2013 event. (c,d) Zonally averaged ($52^\circ\text{--}85^\circ\text{E}$) tropospheric temperature (200–500-mb) above the land at 30°N (c) and the ocean at the equator (d). The temperature over land (“northern temperature”) also exceeded the 1951–2012 climatological mean by 2σ .

tional cluster analysis, these SOM clusters, each of which is associated with a component called a node or neuron, become topologically ordered on a lower-dimensional (typically two-dimensional) lattice so that similar clusters are located close together in the lattice and dissimilar clusters are located farther apart. This topological ordering occurs through the use of a neighborhood function, which acts like a kernel density smoother among a neighborhood of neurons within this low-dimensional lattice. As a result, neighboring neurons within this lattice influence each other to produce smoothly varying clusters that

represent the multidimensional distribution function of the data used to construct the SOM.

In the present study, we perform two separate SOM analysis to categorize daily June 1951–2013 upper-level (200 mb) and lower-level (850 mb) geopotential height anomalies in the domain centered over northern India (0°–60°N, 40°–120°E). In each SOM analysis, the daily geopotential height anomaly field is treated as an M -dimensional vector, where M is the number of grid points. The user specifies the number of clusters, K , and the final clustering is determined through an iterative procedure that approximately

minimizes the Euclidean distance between the daily geopotential height anomaly fields in M -dimensional space and their best-matching SOM patterns. During this iterative procedure, the SOM patterns also “self-organize” into a topologically ordered two-dimensional lattice or grid, as described above. Each daily height field is assigned to a best-matching SOM pattern on the basis of minimum Euclidean distance. Additional details of the SOM methodology can be found in the appendix of Johnson et al. (2008). See also Hewitson and Crane (2002) and Liu et al. (2006) for additional discussion of SOM applications in climate science, and see Chattopadhyay et al. (2008) and Borah et al. (2013) for SOM applications to Indian monsoon intraseasonal variability. Reusch et al. (2005) and Liu et al. (2006) also compare SOM analysis with empirical orthogonal function analysis for the purpose of pattern extraction, demonstrating several advantages of SOM analysis over the more conventional approach.

As mentioned above, the user must specify the choice of K prior to the iterative procedure that determines the SOM clusters. Although there are quite a few approaches for determining an appropriate value of K , there is no universally accepted method for determining the optimal K . In this study, we make the subjective choice of $K = 35$ (i.e., a SOM organized in a two-dimensional lattice with five rows and seven columns), which is a high enough value to resolve regional pattern detail

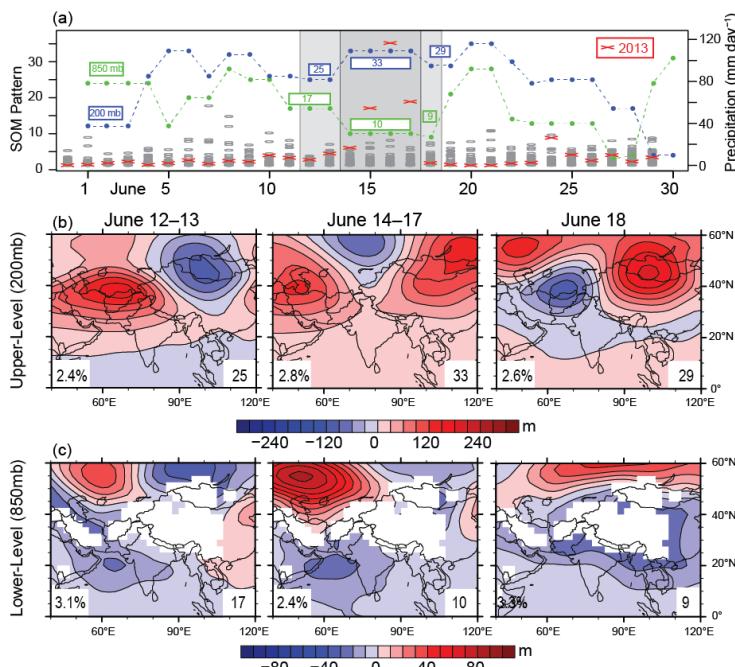


FIG. S17.2. Temporal progression of upper- and lower-level synoptic patterns. We use self-organizing map (SOM) cluster analysis to identify upper- and lower-level atmospheric patterns in June 2013, track their temporal progression, and assess their historical occurrence and co-occurrence (see text). (a) Time-series of the upper-level (blue) and lower-level (green) self-organizing map (SOM) patterns for each day of June 2013. Patterns relevant to the flooding event are highlighted in gray windows in the time-series panel, and their spatial patterns are displayed in panels (b) and (c). SOM pattern numbers display in the lower right of the maps. Daily precipitation accumulation over the flood region (Fig. 1a in the main report, white box) is indicated in the secondary y-axis of (a), with June 2013 values shown in red X's and each year in 1951–2012 shown in gray ovals. The relative frequency of occurrence of each SOM pattern to all historic June days is indicated in the lower left of each SOM map (b,c). From 10 June to 17 June, blocking patterns persisted in the upper atmosphere (200-mb, patterns 25 & 33), while low-pressure troughing strengthened in the lower atmosphere over the flood region (850-mb, patterns 10 & 17). On 18 June, the system lost upper-level support and weakened. 14–17 June 2013 was the first time upper-level pattern 33 co-occurred with lower-level pattern 10 during the month of June within the 1951–2013 historical record.

but small enough to maintain interpretability. We test that our results are robust with different choices for K and that all 35 SOM patterns are statistically distinguishable. For the test of statistical distinguishability, we follow the approach of Johnson (2013). This test is administered by evaluating whether or not all $K(K-1)/2$ SOM pattern pairs are statistically distinguishable on the basis of a field significance test, which in this case is based on the determination of the “false discovery rate” (Benjamini and Hochberg 1995; Wilks 2006). Additional details of the test are given in Johnson (2013).

In the present application, we add an additional step to account for the strong autocorrelation evident in daily data. Because a SOM pattern typically persists for several consecutive days, the number of daily occurrences of each SOM pattern would substantially overestimate the number of temporal degrees of freedom in the local t -tests that determine whether the SOM pattern composite anomalies are significantly different from each other. To correct for this potential source of bias, we perform the local t tests for subsets of the daily geopotential height fields assigned to each SOM pattern. Because the time scale of atmospheric teleconnection patterns is on the order of one to two weeks (e.g., Feldstein 2000), we set the condition that each geopotential height field assigned to the SOM pattern within a subset must be separated by all other geopotential height fields within the subset by at least 15 days. If this separation criterion is not met, then we only keep the daily field with the highest pattern amplitude, where the amplitude is defined as the projection of the daily geopotential height field onto the assigned SOM pattern.

The results of this distinguishability test reveal that all 35 SOM patterns in the 200-mb and 850-mb geopotential height SOMs are statistically distinguishable from each other at the 5% level. Furthermore, all SOM pattern pairs but one (SOM patterns 2 and 8 in the 850-mb geopotential height SOM) are statistically distinguishable from each other at the 1% level. These results suggest that the choice of $K = 35$ is reasonable in that this value of K is not too high such that the SOM patterns become statistically indistinguishable from each other. Furthermore, we also have performed the same analysis for $K = 20$ and found that the main features of the 20-pattern SOMs are quite similar to the main features of the 35-pattern SOMs. Therefore, these evaluations support the robustness of the SOM analysis in this study.

From these 35 distinct nodes, we identify the upper- and lower-level atmospheric patterns in June 2013 to track their temporal progression, and we

assess their historical occurrence and co-occurrence (Fig. 17.1g,h in the main report; Fig. S17.2). Additionally, we determine typical precipitation patterns associated with each geopotential height SOM pattern (e.g., Fig. 17.1i,j in the main report and Fig. S17.2) by creating composites of precipitation from each June day in 1951–2013 on which the atmospheric patterns best matched the corresponding node. This was performed on the basis of assignments of each day to a particular SOM pattern.

Determining a parametric fit for the observed distribution. The event being analyzed lies substantially outside the range of the observed distribution. Therefore, to quantify the probability of occurrence of such an event, we test different parametric distributions (gamma, weibull, lognormal, and pareto) to find a suitable fit for the observations (Fig. S17.3a). We use the maximum log-likelihood estimator to precisely determine the parameters of these potential distributions. Through evaluating the RMS errors (see Table S17.1) between the observed and fitted distributions, we conclude that the Pareto-III distribution most closely represents the observed distribution function.

Probability quantification. We use spatially averaged cumulative June precipitation from 1951 to 2013 over the flood-affected region (Fig. 17.1a in the main report) to generate an observed distribution to examine the probability of the June 2013 event. After fitting the Pareto-III distribution to the observed cumulative June precipitation, we determine the percentile or probability (p) of a specific event magnitude (Pr), and thereby quantify the return period ($T = 1/p$) of the event in the context of the observed climate. We find that the total June 2013 precipitation magnitude exceeds the 99.1th percentile (Fig. 17.1a in the main report) and has a return period of 111 years (Fig. 17.2a in the main report).

TABLE S17.1. Parametric fits of the observed June precipitation distribution.	
Parametric Model	RMS Error
Gamma	0.029
Weibull	0.049
Lognormal	0.021
Pareto	0.018
Root mean square error between fitted and empirical cumulative distribution functions of the observed June precipitation.	

We use the Coupled Model Intercomparison Project phase 5 (CMIP5) historical (20C) and preindustrial (PI) simulations (Table S17.2, S17.3) to quantify the influence of observed anthropogenic forcing on the likelihood of the June 2013 total cumulative precipitation (Fig. 17.2b–d in the main report). The method is illustrated in Fig. S17.3c. After shifting the 20C and PI distributions by the difference between the observed and 20C means, we use the Kolmogorov–Smirnov (K–S) goodness of fit test to measure

TABLE S17.2. Number of years in the historical (20C) and pre-industrial (PI) simulations of select models, and root mean square error between fitted and empirical cumulative distribution functions of model June precipitation.

CMIP5 Model	“20C” years	“PI” years	RMS Error
CNRM-CM5	550	350	0.017
ACCESSI-3	165	500	0.015
ACCESSI-0	110	500	0.025
HadGEM2-CC	147	240	0.019
HadGEM2-ES	275	577	0.016
IPSL-CM5B-LR	55	300	0.02

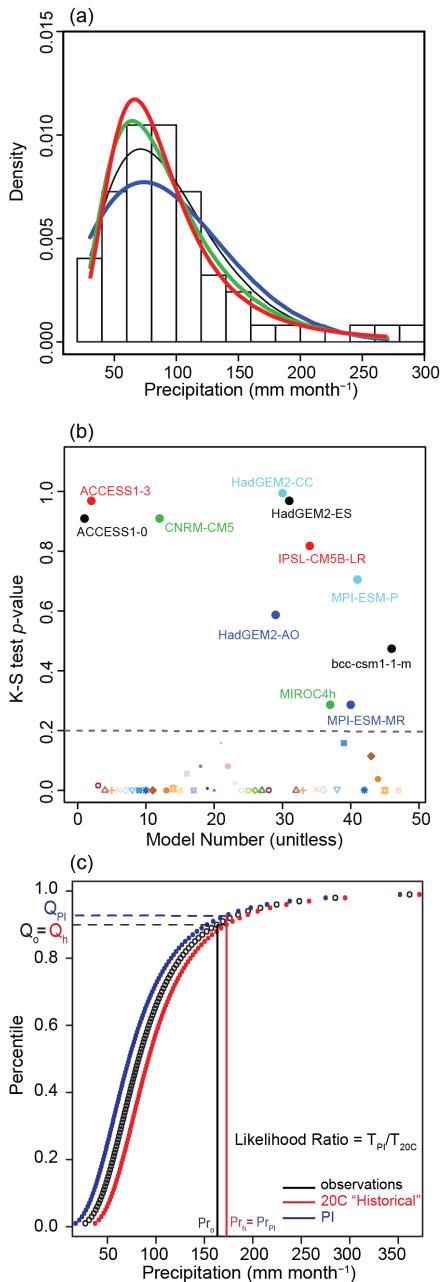


FIG. S17.3. Statistical framework for probability quantification. (a) Results from fitting the observed cumulative June precipitation (1951–2012) with 4 parametric models. The Pareto-III model shows the lowest root mean square error between the empirical and model cumulative distribution functions (shown in Table S17.1). (b) p-values of the Kolmogorov–Smirnov (“K–S”) goodness of fit test measuring the closeness of CMIP5-simulated historical (20C) distributions to the observed distribution. (See Table S17.2 for model names). Higher p-values indicate models that more closely simulate the observed distribution, including the tails. (c) Theoretical methodology for quantifying the probability of a specific event in the 20C and preindustrial PI climates. This involves determining the magnitude of the event (Pr_h) in the 20C distribution corresponding to the percentile of the observed event (Q_o = Q_h). The ratio (PI/20C) of the return periods is calculated by estimating the percentile (or probability) of this simulated event (Pr_i = Pr_h) in the pre-industrial (T_{PI}) and historical (T_{20C}) distributions.

the closeness of CMIP5 simulated historical (20C) distributions to the observed distribution. [Refer to Sperber et al. (2013) for an evaluation of the skill of these models to simulate other characteristics of Indian monsoon rainfall, circulation indices, and their teleconnections.] Because the simulated change in likelihood of extremes can be heavily influenced by biases in the simulated distribution, we restrict our analysis to 11 models whose K–S value exceeds 0.2 (A1; see Fig. S17.3b), ensuring a comparatively good fit of the overall distribution, including in the tails. We then fit the select A1 CMIP5 models with Pareto-III distributions.

We determine the magnitude of the 99.1th percentile event in the 20C distribution as representative of the simulated event magnitude (follow the vertical red line in Fig. S17.3c). By definition, the event has a return period of 111 years in the 20C distribution. Next, we find the corresponding percentile—and hence return period—of this 20C magnitude event in the preindustrial simulations (follow the horizontal blue line in Fig. S17.3c).

By comparing the likelihood of the event in the preindustrial climate to the historical climate, the ratio of the return periods (T_{PI}/T_{20C}) provides an estimate of the influence of 20C forcings on the probability of the June 2013 heavy precipitation.

However, the limited sample of precipitation creates uncertainty in the true fit of the Pareto-III distribution to the observed and CMIP5 simulated precipitation data, thereby introducing uncertainty in the PI/20C return period ratio. We therefore use a bootstrapping approach to quantify this uncertainty and provide confidence estimates for this ratio. We bootstrap the observed, 20C, and PI precipitation data

to generate many distributions to capture the uncertainty around the true parameters of the Pareto-III fits and thus the uncertainty in the return periods. By applying the probability quantification methodology described above to all these bootstrapped distributions, we generate a distribution of the PI/20C return period ratios, which are represented in box plots in Fig. 17.2d in the main report.

TABLE S17.3. Names of the 11 CMIP5 models that meet the K-S test criterion.

Model	Full Model name
CNRM-CM5 C	Centre National de Recherches Meteorologiques - Coupled Global Climate Model version 5
HadGEM2-ES	Hadley Global Environment Model 2 - Earth System
ACCESS1-0	Australian Community Climate and Earth-System Simulator (ACCESS) version 1.0
ACCESS1-3	Australian Community Climate and Earth-System Simulator (ACCESS) version 1.3
BCC-CSM1-I B	Beijing Climate Center - Climate System Model version 1.1
MIROC4h	Model for Interdisciplinary Research on Climate, version 4 High Resolution
MPI-ESM-MR	Max Planck Institute - Earth System Model - Medium Resolution
HadGEM2-CC	Hadley Global Environment Model 2 - Carbon Cycle
HadGEM2-AO	Hadley Global Environment Model 2 - Atmosphere Only
MPI-ESM-P	Max Planck Institute - Earth System Model - Low Resolution and Paleo mode
IPSL-CM5B-LR	Institut Pierre-Simon Laplace - Climate Model version 5B - Low Resolution