Tipping elements of the Indian monsoon: Prediction of onset and withdrawal

Article *in* Geophysical Research Letters · March 2016

DOI: 10.1002/2016GL068392

4 authors:

Veronika Stolbova

Potsdam Institute for Climate Impact Rese…

9 PUBLICATIONS **39** CITATIONS

SEE PROFILE

Elena Surovyatkina

Space Research Institute

28 PUBLICATIONS **96** CITATIONS

SEE PROFILE

Bodo Bookhagen

University of California, Santa Barbara

179 PUBLICATIONS **3,756** CITATIONS

SEE PROFILE

Juergen Kurths

Potsdam Institute for Climate Impact Rese…

1,130 PUBLICATIONS **38,438** CITATIONS

SEE PROFILE

Supporting Information for "Tipping elements of the Indian monsoon: prediction of onset and withdrawal"

Veronika Stolbova,[1,2,3] Elena Surovyatkina ,[1,4] Bodo Bookhagen[5,6], and Jürgen Kurths[1,2,3,7]

This supporting information contains the following sections:

1. Choice of the onset date (OD) of the Indian monsoon

2. Choice of data for prediction scheme

3. Details on the calculation of the variance of fluctuations

4. Figure S1: variance of fluctuations for the relative humidity

5. Determining the tipping elements of the Indian monsoon. Figures S2 and S3

6. Time series from the tipping elements of the monsoon for the case study in 2012. Figure S4

7. Parameters for calculation of predicted onset and withdrawal dates. Estimation of onset and withdrawal dates and Table S1

8. Identification of the optimal training period. Figure S5

Corresponding author: V. Stolbova, Potsdam Institute for Climate Impact Research, Potsdam, Germany, veronika.stolbova@pik-potsdam.de

[1]Potsdam Institute for Climate Impact

9. Robustness of the proposed method

10. Sensitivity of the prediction scheme for the onset and withdrawal dates on the starting and end points for the trend calculation. Figure S6

11. Estimation of the uncertainties of the predicted onset and withdrawal dates. Figure S7

12. Choice of a successful prediction

13. Comparison with a constant predictor

1. Choice of the onset date (OD) of the Indian monsoon

The Indian Meteorological Department (IMD) considers the onset of the monsoon date - as the onset date over the southern tip of the Indian subcontinent (Kerala) region (Figure 1, C), black dotted line). However, there are some difficulties in determining the onset of the monsoon date for this region, mainly due to "bogus" (false) monsoon onsets. In order to avoid this uncertainty in the definition of the correct day of monsoon onset over Kerala, in this study, we use as the onset date over the Eastern Ghats (EG) region as the monsoon onset date (OD) (Figure 1 C, red dotted line). Most importantly, arrival of the monsoon to the EG marks the full arrival of the monsoon onto the Indian subcontinent, as the northern position of the Intertropical convergence zone (ITCZ) during the monsoon is located near the EG (23.5N).

Research, P.O.Box 601203, 14412 Potsdam,

2. Choice of data for prediction scheme

Holloway and Neelin [2009] among other studies have shown that column water vapor w (vertically integrated specific humidity) is a good proxy for transition to the strong convection – onset of precipitation. In *Holloway and Neelin* [2009], and other parametric models of precipitation called convective parameterizations, precipitation occurs when w reaches a certain critical saturation value $w_c(T)$, which depends on the critical temperature T_{crit}. While vertically integrated climate variables are harder to measure with desired accuracy for prediction of the onset and withdrawal dates, values of the near-surface air temperature (T) and relative humidity (rh) are better represented in models, experiments, and observations. In this regard, we choose T and rh for the prediction of the onset and withdrawal dates of the monsoon. In addition, we also determine T_{crit} for our method - which is equal to T_{sat} (see Table S1).

3. Details on the calculation of the variance of fluctuations

We calculate the variance of fluctuations (σ^2) of T and rh for each grid point for the window of length w, d days before onset date of monsoon in a given year y as following:

$$\sigma^2(x,d,w,y) =< [x(t^*(y)-d-k) - \bar{x}(t^*(y)-d-k)]^2 >_w = \quad (1)$$

$$= \sum_{k=1}^{w} \frac{[x(t^*(y)-d-k) - \sum_{i=1}^{w} \frac{x(t^*(y)-d-i)}{w}]^2}{w} \quad (2)$$

where $x(t)$ is a time series, w the length of the time window, d the number of days prior to OD, y a given year, and $t^*(y)$ the OD in the given year y.

Germany.

For detection of most common positions of the RPs, we average σ^2 over the whole range of years:

$$\sigma^2 = \sigma^2(x,d,w) =< \sigma^2(x,d,w,y) >_y = \sum_{y=1}^{Y} \frac{\sigma^2(x,d,w,y)}{Y} \quad (3)$$

where Y is the total number of years.

4. Figure S1: variance of fluctuations for the relative humidity

The variance of fluctuations for the relative humidity is shown in Figure S1.

5. Determining the tipping elements of the Indian monsoon. Figures S2 and S3

Determining the geographic regions of critical behavior or the tipping elements of the Indian monsoon is supported by several additional facts. First, in *Stolbova et al.* [2014] using climate network analysis of extreme precipitation over the Indian subcontinent during the pre-monsoon, monsoon and post-monsoon seasons it was shown that only during the monsoon season the EG and NP experience the synchronization of the extreme rainfall. This is caused by the establishment of the monsoon trough, which connects these regions. Second, from a meteorological perspective, it is not so surprising that the EG and NP regions exhibit highly developed instability in observables on the eve of the monsoon onset. The maximum northern position of the ITCZ runs through the EG, and therefore, it coincides with hottest place on the Indian subcontinent (around latitude 23.4N) (see Figure S3). The topography of the EG creates a favorable condition for the formation of

---

[2]Department of Physics,

59 a low-pressure system, and when ITCZ reaches the EG, two branches of monsoon merge:
60 Arabian Sea and Bay of Bengal. At the same time, high pressure in NP and its intricate
61 topography bounded by the Himalaya, create favorable condition for the developing anti-

Humboldt-Universität zu Berlin, Newtonstr. 15, 12489 Berlin, Germany.

[3]Department of Control Theory, Nizhny Novgorod State University, Gagarin Avenue 23, 606950 Nizhny Novgorod, Russia

[4]Space Research Institute, Russian Academy of Sciences, Profsoyuznaya 84/32, GSP-7, Moscow 117997, Russia

[5]University of Potsdam, Institute of Earth and Environmental Science, Karl-Liebknecht-Str. 24-25, 14476 Potsdam-Golm, Germany

[6]Geography Department, University of California, Santa Barbara, Santa Barbara, CA 93106-4060

[7]Institute for Complex Systems and Mathematical Biology, University of Aberdeen, Aberdeen AB243UE, UK

cyclone. Altogether, it creates a general wind pattern, in which the center of the monsoon cyclone is near the EG, while the center of the anticyclone - is in the NP region. As a consequence, we observe the maximum growth of σ^2 in the centers of the cyclone and the anticyclone systems. Third, the collision of the anticyclone located in NP and the developing low pressure system associated with the monsoon, marks the onset of monsoon in the EG. Therefore, fluctuation in NP and the EG show indicators of this collision. The analysis of the SAT as a function of day of the year for different regions of the Indian subcontinent shows that the largest transition from the maximum temperature in a given region to its temperature during monsoon season is observed in the EG region (Figure S2). While, at the same time, the monsoon does not reach up to the NP region and this region does not exhibit the drop of the SAT similar to the EG and other regions of the Indian subcontinent which experience the transition to monsoon. This region is the northernmost geographical boundary of the monsoon. Therefore, time series analysis of the SAT points out on the fact that transition to monsoon can be marked by geographical boundaries starting with the largest transition on the EG to the disappearance of transition in the northern boundary of monsoon - NP region. All these factors combined allow us to identify EG and NP as tipping elements of the monsoon.

6. Time series from the tipping elements of the monsoon for the case study in 2012. Figure S4

As mentioned in the main text, intersection of the time series of air temperature (T) and relative humidity (rh) for each year are in the vicinity of actual onset and withdrawal dates (see Figure S4).

7. Parameters for calculation of predicted onset and withdrawal dates. Estimation of onset and withdrawal dates and Table S1

Here, we describe the parameters used for the estimation of the onset and withdrawal dates. The parameters are gathered into three rows, for each part of the predictability scheme: OD determination using T ($OD_p(T)$), OD determination using rh ($OD_p(rh)$) and WD determination using T ($WD_p(T)$) (see Table S1). Columns represent characteristics of the time series related to NP and EG for the mean and given year time series (trends for the given period of time), saturation values (connected with both NP and EG, calculated as values at which intersection of the mean time series from the tipping elements occur), as well as day of the saturation - determined as the day, when T in EG during the given year reached T_{sat}, which is temperature of intersection of the mean T in the EG and NP at the average onset date from the training period (Table S1).

Estimation of $OD_p(T)$. As the first step, trends are calculated for mean time series from NP ($< NP >$) and EG ($< EG >$) (from the training period) and for time series for the year when the prediction is made, for the periods 30–125 DOY (NP) and 30–100 DOY (EG). Second, T_{sat} is detected as intersection of the mean time series $< NP >$ and $< EG >$. If the difference between the slopes of the current year trend in NP and the mean trend in NP ($< NP >$) is less than 15% of the slope of the mean trend ($< NP >$), we estimate the $OD_p(T)$ as intersection of the trend in NP for the current year of the T_{sat} value. If the estimated value of $OD_p(T)$ lies in the boarders of the onset dates from the corresponding training period, we use this value as predicted OD. If the uncertainty of the trend estimation of either upper or lower bound of the estimate of onset date exceeds 16 days, we use estimation obtained from the training period of previous 14

years. Otherwise, if the deviation of the trend in NP for the current year is more than 15% or the estimated values is outside of the range determined by the corresponding training period, the $OD_p(T)$ is determined using the mean trend from the training period.

Estimation of $OD_p(rh)$. We detect rh_{sat} as a near surface relative humidity at which rh values intersect from the tipping elements of the monsoon. Second, we calculate the trend of rh in NP for the period 81–172 DOY. Third, from the time series of near surface air temperature we identify d_{sat} as the date when mean T in EG for the first time reaches T_{sat}, and we calculate the trend of rh in NP from this date untill 125 DOY. Then, $OD_p(rh)$ is estimated as an intersection of this trend with rh_{sat}. If the uncertainty of the trend estimation of either upper or lower bound of the estimate of onset date exceeds 16 days, we use estimation obtained from the training period of previous 14 years. In addition, when the estimated value lies outside the boundaries determined by the onset dates from the training period, we estimate $OD_p(rh)$ using the mean trend from NP from the training period and rh_{sat}.

Estimation of $WD_p(T)$. Estimation of the WD is based on the fact that on average, the slope of the linear trend of $< NP >$ for the period 30–125 DOY is equal with an opposite sign to the slope of the decrease of the average T in NP from the training period at the end of the monsoon season. We call this trend a *reversed trend*. We force the reversed trend to the mean time series from NP to intersect the average T in the EG at the $< WD >$ of the training period. We note, that this reversed trend intersects the normal mean trend close to the day when T of the mean T in NP reaches it's maximum during the year. Based on this observation, we estimate the WD for the given year as

follows: First, we identify the maximum value of T in NP for the period 195–205 DOY and mark the date when it happens. Second, we reconstruct the reversed trend in NP from the current year trend in NP including the condition that the obtained trend should intersect with the normal trend at the detected date of the maximum of T in NP. Third, we identify T_{mon} in the EG (monsoon temperature) from the training period by finding T in the EG at the $< WD >$. Finally, we estimate $WD_p(T)$ as the intersection of the reversed trend of NP and baseline for T_{mon}. When the predicted WD lies outside of the boundaries from the values from the training period for a given year, we estimate the WD using trends of the mean time series from the training period from the tipping elements of monsoon for this year.

8. Identification of the optimal training period. Figure S5

Skills of the proposed prediction scheme depend on the training period for the estimation of the onset and withdrawal dates. Here we define a correct prediction as a year when the difference in days between the predicted and real onset dates is equal or less than 7 days for the OD and 10 days for the WD. Therefore, the number of correctly predicted onset dates is what we call the true positive rate (TPR).

Similarly, we define a false positive predicted date when the predicted onset date is more than a week earlier than real onset date (FPR) and false negative when predicted onset date was more than a week later than the real onset date (FNR). We test our predictability scheme for the different training periods. The result is shown in the Figure S5.

9. Robustness of the proposed method

The prediction skill of the proposed method indeed is sensitive to the choice of geographic area. The NP and EG regions (NCEP/NCAR data set) used for the analysis are small and both considered regions include one grid cell with a size of 2.5°*2.5° (280 km* 280 km). We expanded the areas of the EG and NP regions and tested the prediction skill of our method. Extending each of the two regions by 4 additional grid cells (south, north, west and east) result in worse prediction skills of 60% (prediction of onset date based on near-surface air temperature), 50% (prediction of onset date based on the relative humidity), and 84% (prediction of withdrawal date), respectively.

Most importantly, we primarily rely on NCEP/NCAR data as it has the longest record of time series, which allow us to test the accuracy of the prediction on the results obtained for 50 years. There are no other observational or reanalysis data set to our knowledge that satisfies the condition of long time record and high resolution. We are thus limited to the extent of one grid cell because of the spatial resolution of the input data. The choice of a larger area for both regions, which would include more grid cells into the region leads to worse prediction, as including additional grid points to the data will also require additional information of the monsoon onset date in the added grid cell. Therefore, extending areas will influence the monsoon onset propagation date.

Another crucial point for the choice of NCEP/NCAR data for the prediction scheme is that the measured onset data for the EG region is provided specifically for this data set. It is reasonable to compare the predicted onset date on the same data set from which the

onset dates were measured. To our knowledge, there is no other existing long-term list of the onset dates for the EG region than provided in [*Singh and Ranade*, 2010].

10. Sensitivity of the prediction scheme for the onset and withdrawal dates on the starting and end points for the trend calculation. Figure S6

The choice of these two parameters can be explained by the following two factors: 1) the first date (30) is chosen because starting at 30 DOY (day of the year) the linear growth of the near-surface air temperature starts in both EG and NP regions based on our analysis of 65 years of surface air temperature data; 2) the second date (125) is chosen to fulfil two conditions: a) it is the date when we forecast the onset and it should be earlier than existing forecasting dates of monsoon onset, which is on May 15-th, 125 DOY is May 5-th, b) the date should not be too early, as the estimation of the linear trend will have larger uncertainty.

We performed our analysis of the method's sensitivity for the starting and end points of the linear trend estimation in the NP region (please see Figure S6). The results of the prediction scheme are robust for a certain range of starting and end points for the linear trend estimation: 30-42 DOY (starting) and 103-125 DOY (end point); 10-30 DOY (starting point) and 120-140 DOY (end point). These two ranges result in a successful prediction of more than 70% (marked with red on the Figure S5, see below), however, the second range can be used for the medium-range prediction (less than a month), while the first range allows to perform long-range prediction (more than 30 days in advance). In our study, we are particularly interested in the long-range prediction, therefore, we use starting and end points for the trend calculation from the first range. The parameters

used in our study (starting and end point for the trend calculation for the NP) are marked with black squares.

We choose the parameters of the starting and end point for the calculation of the trend in the NP as (30,125) to satisfy the high prediction success for both onset and withdrawal dates of the Indian Summer Monsoon.

11. Estimation of the uncertainties of the predicted onset and withdrawal dates. Figure S7

We used uncertainty of the slope of the linear regression for estimating the uncertainty of the onset and withdrawal dates for a chosen year as shown on Figure S7.

12. Choice of a successful prediction

A choice of a successful prediction ($\leq$ 7 days) is determined by the fact that the real onset date is also not given uniquely: the identified real onset date is indeed a range - as shown in Figure 4 A (blue range). It is caused by the difficulties in the determination of a real onset date for each year. This range of identification of the real onset date takes values between $\pm$4 and $\pm$7 days depending on the particular year (see *Singh and Ranade* [2010] for details). Therefore, smaller range for determination of the successful prediction does not make sense. Similar situation hold for the withdrawal, where the uncertainty of the determined WD is larger, therefore, we use the condition of difference between the predicted and real withdrawal dates of $\leq$ 10 days.

13. Comparison with a constant predictor

A comparison of the method proposed here with a constant predictor gives result 73% for our method and 75% for the constant predictor, and, although, on average for 50 years the constant predictor is on 2% better (one year more of successfully predicted years), this fact reflects the distribution of the monsoon onset dates over the long-term period. In addition, our method has some difficulties in predicting the onset dates in the 1985–1988, due to the high variability of the near-surface air temperature in the NP region. Note also, that forecasting of normal monsoon years is much more successful than anomalous years in the existing forecasting scheme (since they are distributed closely to June 19th). However, a great challenge is to forecast anomalous monsoon years associated with El-Niño and La-Niña and too early and too late monsoon onset dates. It is important to note that a constant predictor is not able to forecast correctly early and late monsoon onsets, while our method can successfully forecast them. Specifically, during anomalous La-Niña and El-Niño years, when it is difficult to forecast the monsoon onset date, our method improves the forecasting: e.g. 1972, 1989-1990, 1999-2001.

References

Holloway, C., and J. D. Neelin (2009), Moisture vertical structure, column water vapor, and tropical deep convection,*Journal of the Atmospheric Sciences*, pp. 1665–1683, *doi:10.1175/2008JAS2806.1*

Singh, N., and A. A. Ranade (2010), *Determination of Onset and Withdrawal Dates of Summer Monsoon across India using NCEP/NCAR Re-analysis data set*, Indian Institute of Tropical Meteorology.

Stolbova, V., P. Martin, B. Bookhagen, N. Marwan, and J. Kurths (2014), Topology and seasonal evolution of the network of extreme precipitation over the Indian subcontinent and Sri Lanka, *Nonlinear Processes in Geophysics*, *21*, 901–917.

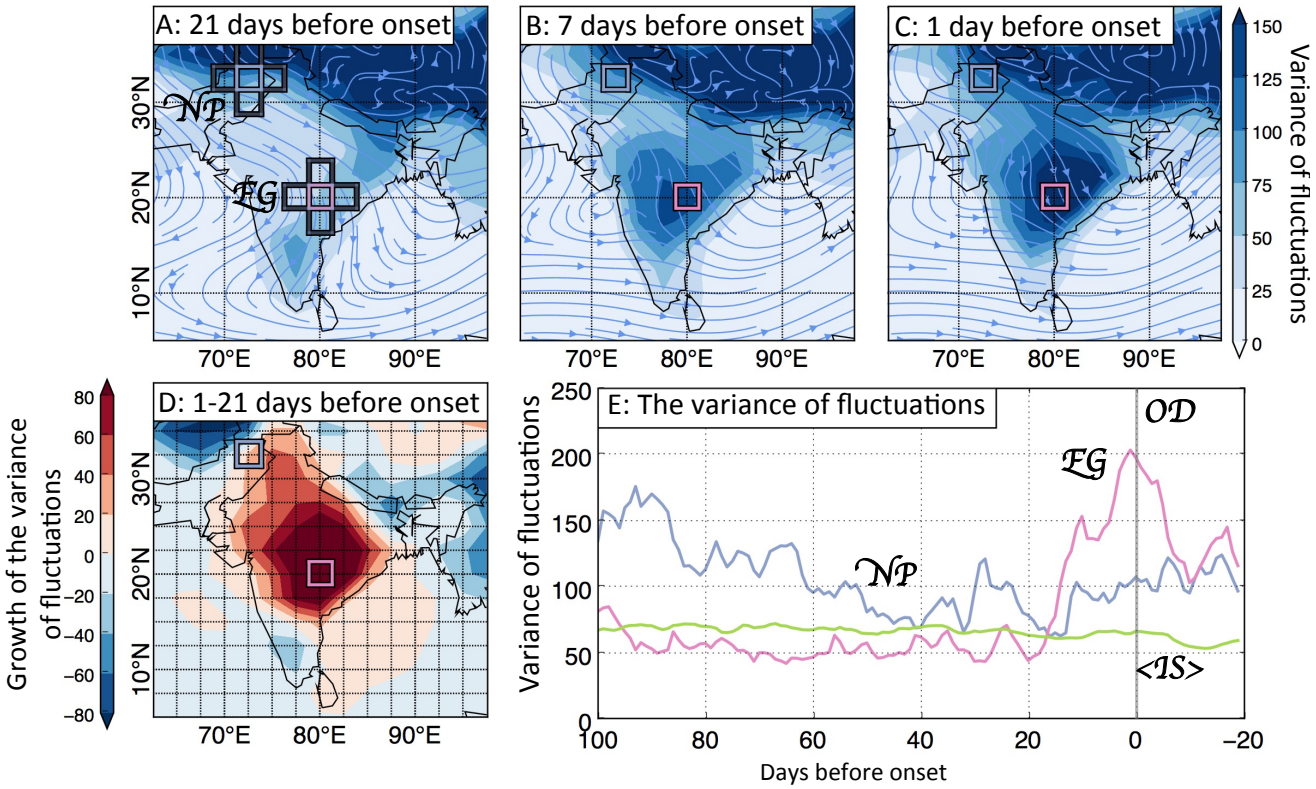

Figure S1. Pre-monsoon growth of the variance of fluctuations (σ^2) of the weekly mean values of near-surface relative humidity (rh) 21 days, 7 days and 1 day before the monsoon onset at the Eastern Ghats: A, B and C. D=C-A. Composites are for the period 1971-2001 and were calculated from the ERA40 reanalysis data set, 700 hPa winds are indicated by the blue lines. Two boxes refer to RPs: North Pakistan, NP (blue) and Eastern Ghats, EG (pink). E - growth of the variance of fluctuations in NP (blue), EG (pink), and averaged over the Indian subcontinent ($< IS >$) at approaching onset date of the monsoon (OD).

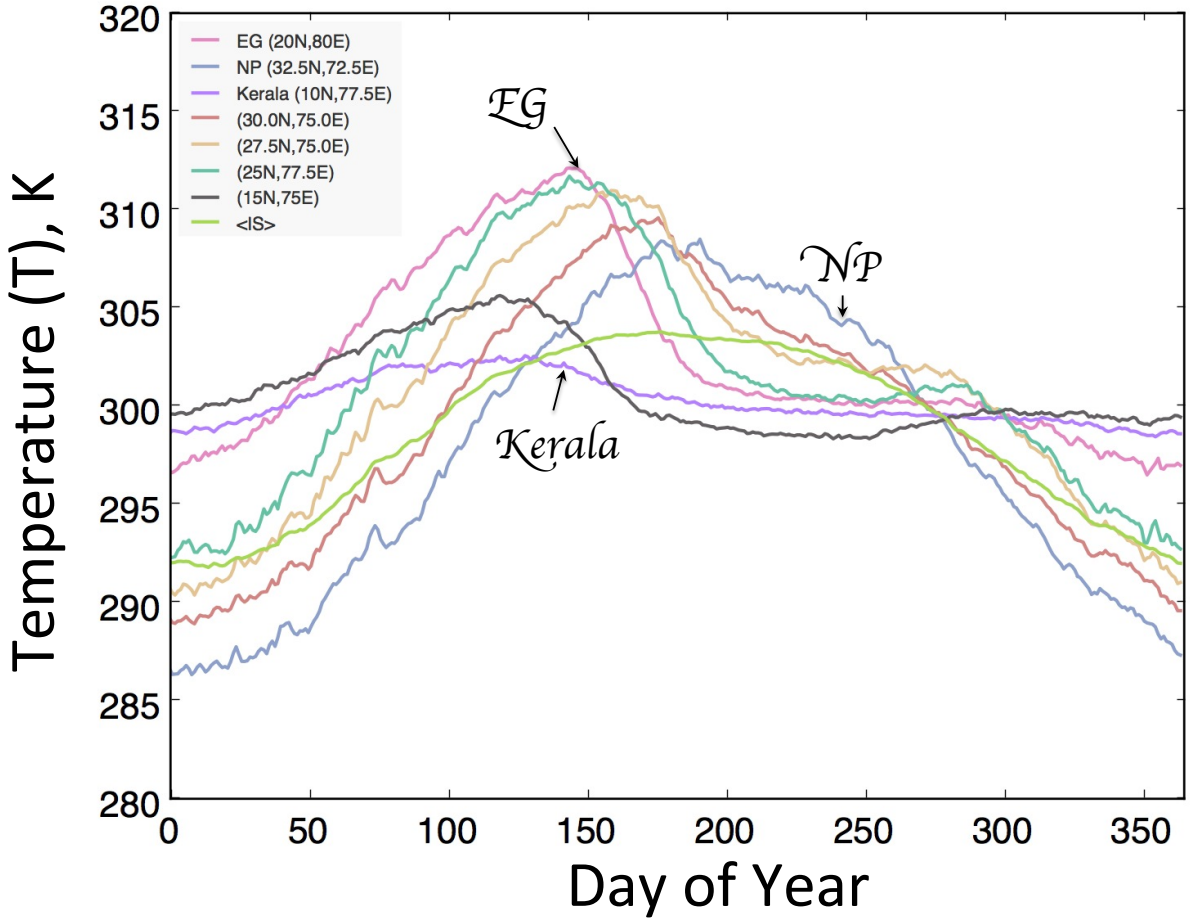

Figure S2. Time series of near-surface air temperature (T) depending on the day of the year (DOY) and geographical location. $< IS >$ - means average value over the whole considered region (62.5–97.5° E, 5.0–40.0° N).

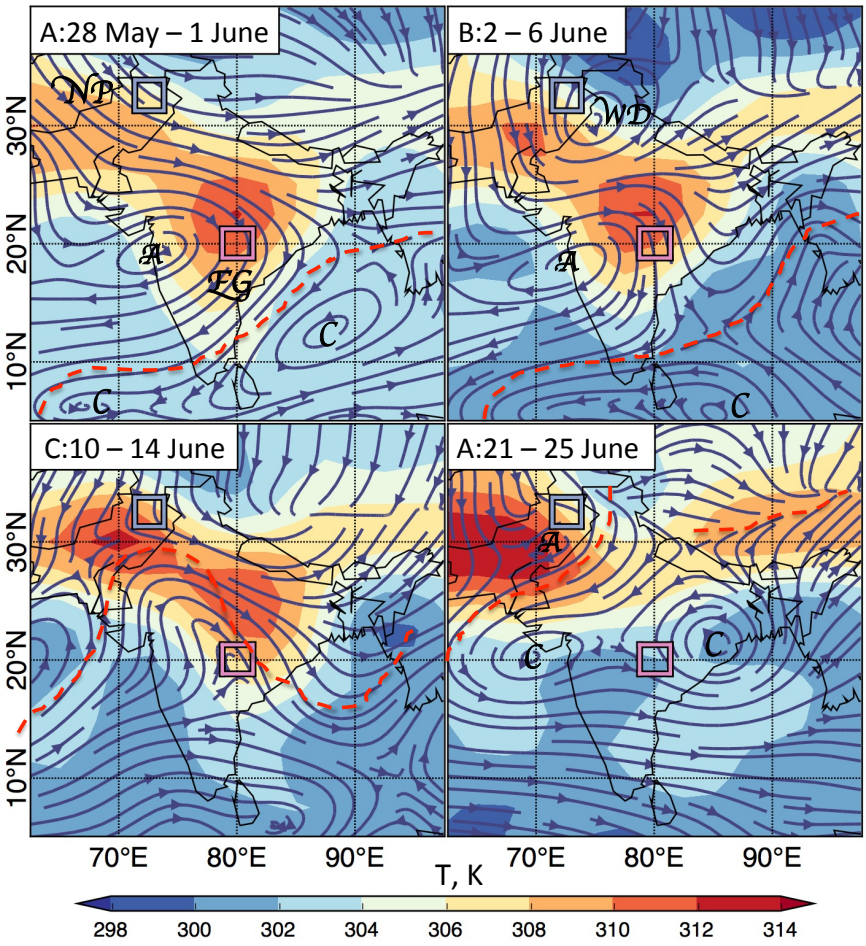

Figure S3. Streamline analysis of the mean resultant winds at 700 mb over the Indian region during the period: A) 28 May - 1 June 2015, B) 2-6 June 2015, C) 11-15 June, D) 21-25 June 2015. Onset of monsoon over the EG is on 13th June in 2015. C denotes cyclonic circulation associated with the equatorial trough of low pressure, and A - anticyclonic circulation associated with 'warm high' above surface 'heat low' over India, WD is referred to the Western Disturbances, associated with subtropical Westerlies. Northern Boundary of Monsoon is indicated by a red dashed line and is identified with the Intertropical Convergence Zone. Color shows near-surface air temperature (T), averaged over the mentioned periods, 2015.

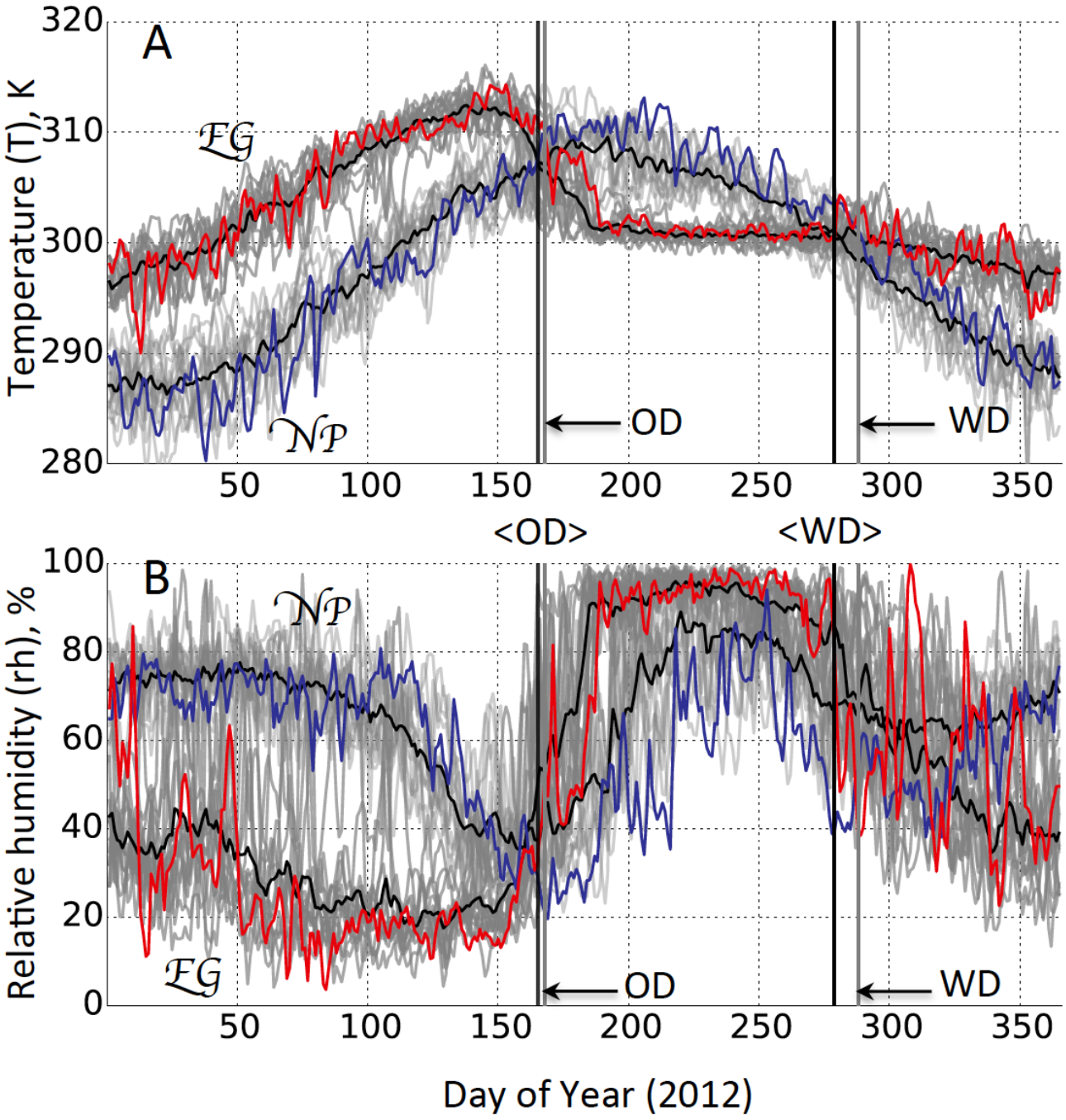

Figure S4. Time series from tipping elements of monsoon: air temperature at 1000 hPa (A) and relative humidity at 1000 hPa (B); 14-year mean (black) and 2012 values for NP (blue) and the EG (red). Grey lines on the background show time series from the NP and EG for the training period of 14 years. Black solid lines indicate mean values of the OD ($< OD >$) and WD ($< WD >$) for the training period, while solid grey lines - actual onset (OD) and withdrawal dates (WD) for 2012.

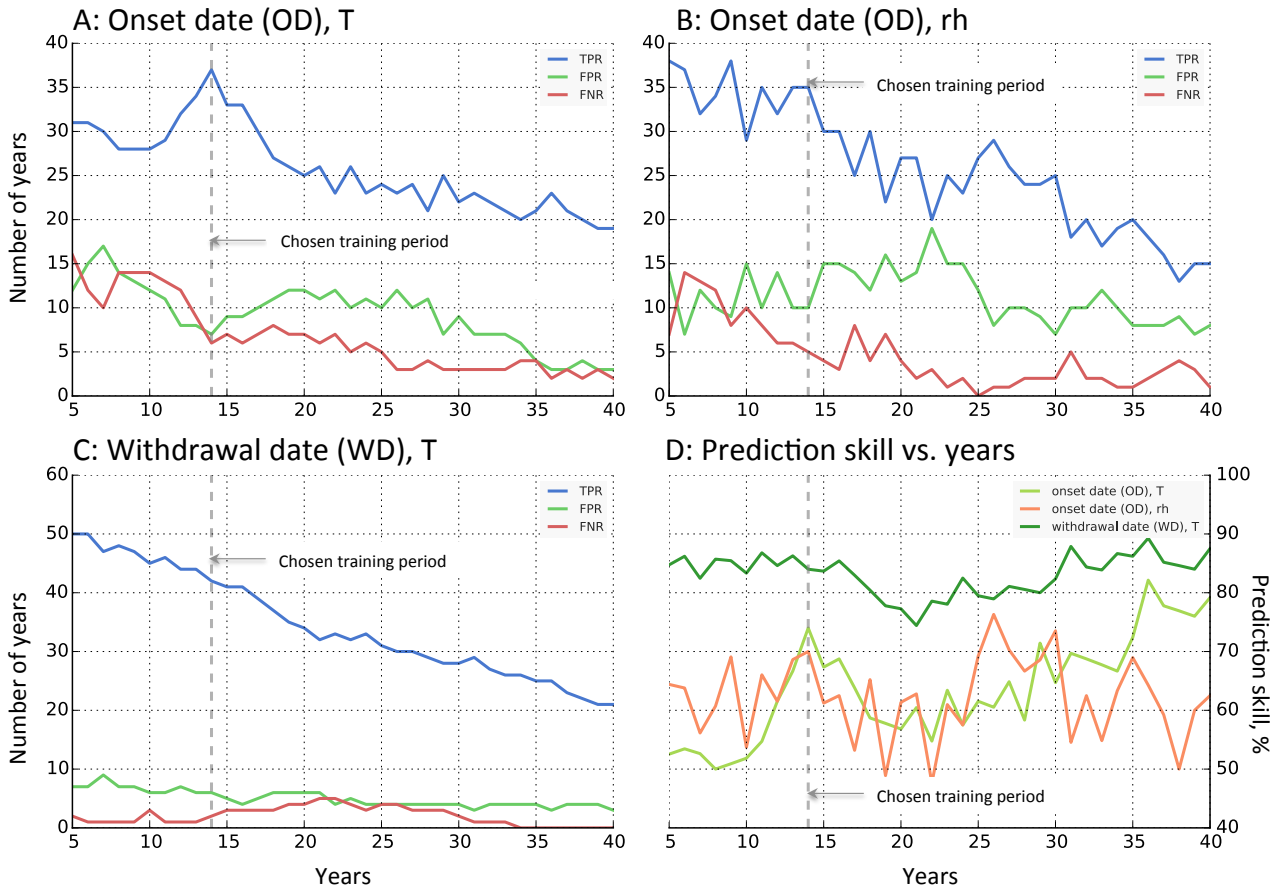

Figure S5. Sensitivity of the accuracy of prediction scheme to the length of the training period. A, B, C - Dependence of the TPR, FPR and FNR in years versus length of the training period. Number of the considered years for each value of training period=TPR+FPR+FNR. D - Dependence of prediction skill on the number of considered years.

Table S1. Parameters used for calculating monsoon onset and withdrawal date.

$< NP >$ and NP - characteristics related to the mean time series for the training period and for year for which prediction is made from the NP, $< EG >$ and EG - the same definition - for EG region, (x,y) - linear trend calculated for the period from x to y day of the year (DOY), T_{sat} and rh_{sat}- temperature and relative humidity of the intersection of the mean (during the training period) time series from NP and EG, d_{sat} - day of saturation, when $T_{EG}(d_{sat}) = T_{sat}$ for the first time during the pre-monsoon period, $T_{NP_{max}}$ - maximum temperature in the NP during the given period of time, $T_{mon} = T_{WD}$, where T_{mon} is temperature of monsoon, T_{WD} - is temperature of the mean withdrawal date, SV - saturation value, training period is 14 years.

	$< NP >$	$< EG >$	NP	EG	SV	DS
OD_T	(30,125) DOY	(30,100)	(30,125)	(30,100)	T_{sat}	$d_{sat} = f(T_{sat})$
OD_{rh}	(81,172) DOY	———	(d_{sat},125)	———	rh_{sat}	d_{sat}
WD_T	(30,125) DOY	———	$T_{NP_{max}}(195:205)$ (30,125)	———	T_{WD}	———

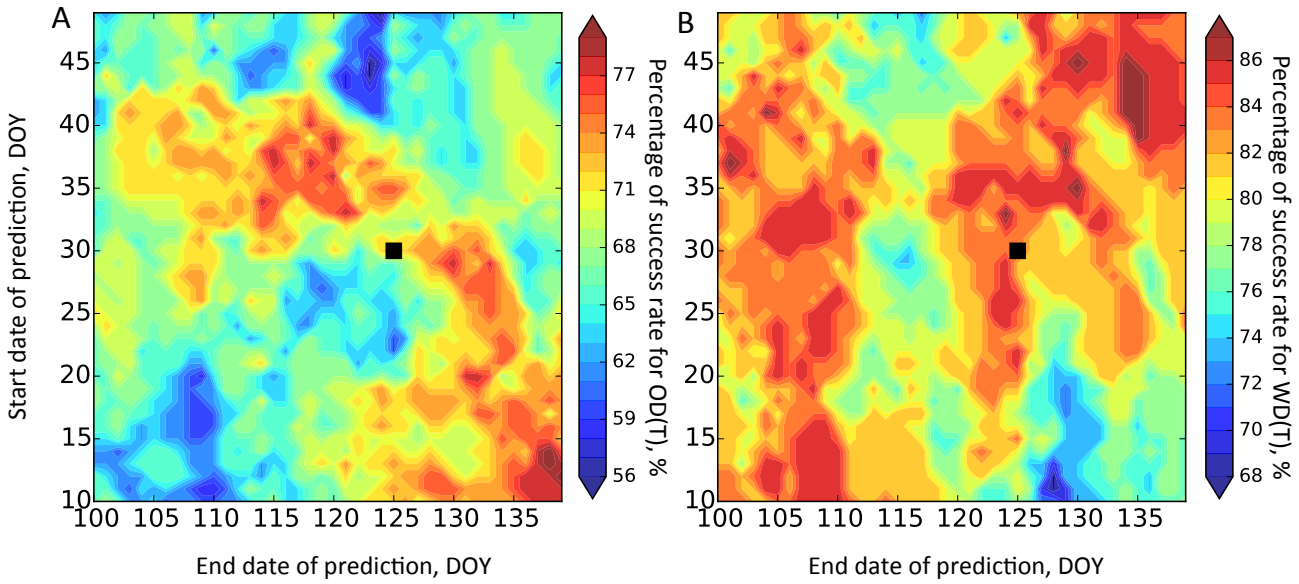

Figure S6. Sensitivity of the prediction scheme for the onset and withdrawal dates on the starting (x-axis) and end points (y-axis) for the trend calculation in the North Pakistan (NP) region. Colorbars represent values in percentage of the success rate while performing the prediction scheme for 50 years of data (1965-2015) with varying starting and end points for the trend estimation.

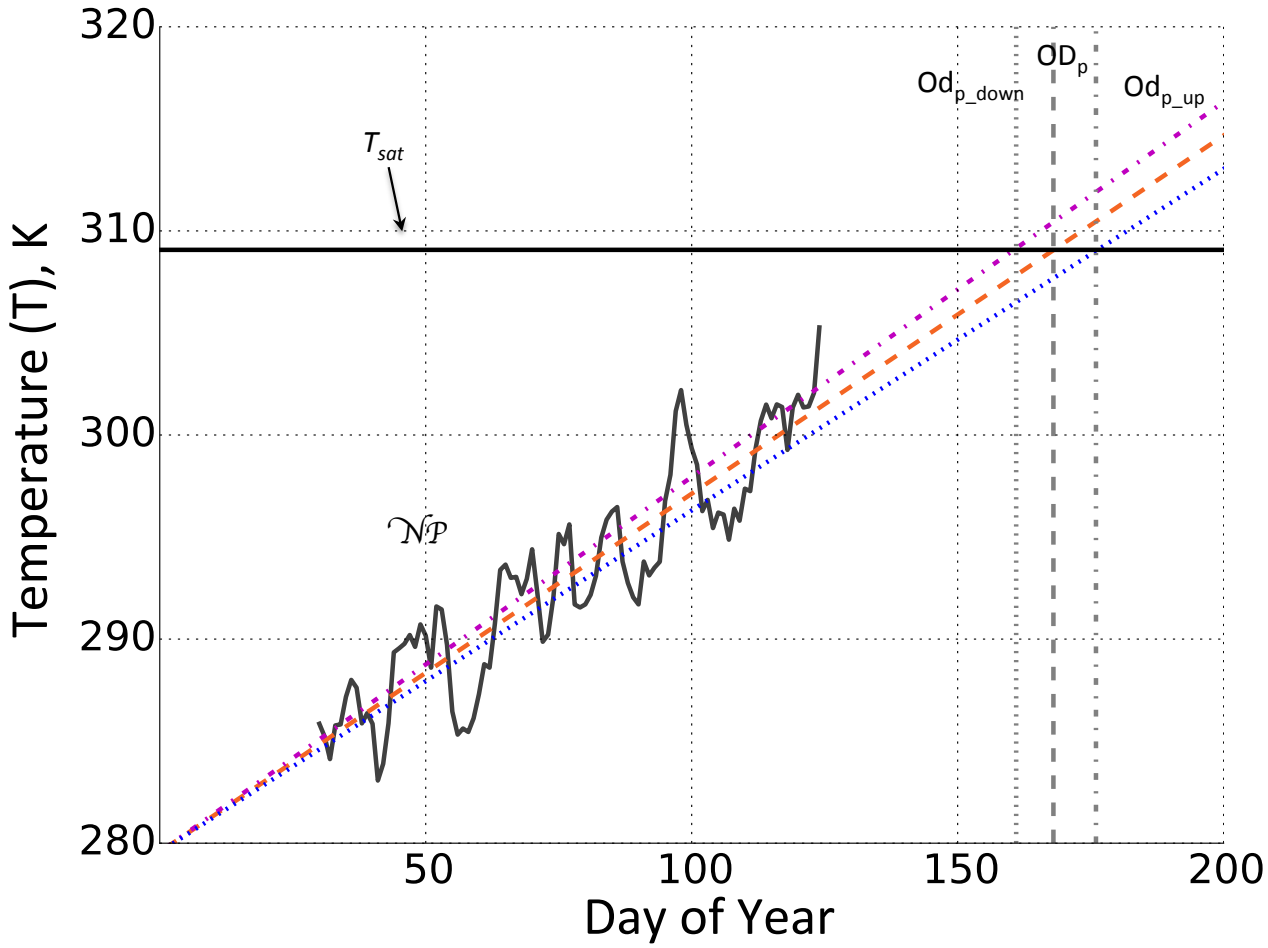

Figure S7. Illustration of the uncertainty estimation for the monsoon onset date in a given year for the NP region. Intersection of the dashed orange line (trend estimated in the NP from the near-surface air temperature) with horizontal black line (saturation temperature detected from the 14 previous years) gives an estimate of the onset date, while intersection of the dash-dotted magenta and dotted blue lines with the horizontal black line indicate upper and lower boundaries of the uncertainty of the onset date, taking into account the standard deviation uncertainty of the slope of the estimated linear trend in the North Pakistan. Uncertainties of the monsoon onset date estimated from the relative humidity time series as well as for the withdrawal dates are calculated following the same procedure.