

A predictable prospect of the South Asian summer monsoon

Tuantuan Zhang

School of Atmospheric Sciences, Sun Yat-sen University

Xingwen Jiang (✉ xingwen.jiang@yahoo.com)

China Meteorological Administration

Song Yang

Sun Yat-sen University <https://orcid.org/0000-0003-1840-8429>

Junwen Chen

Shenzhen Wiselec Technology Co., Ltd.

Zhenning Li

Hong Kong University of Science and Technology

Article

Keywords:

Posted Date: June 3rd, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1662244/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Version of Record: A version of this preprint was published at Nature Communications on November 18th, 2022. See the published version at <https://doi.org/10.1038/s41467-022-34881-7>.

1 **A predictable prospect of the South Asian summer monsoon**

2

3

4 Tuantuan Zhang^{1,2}, Xingwen Jiang^{3*}, Song Yang^{1,2}, Junwen Chen⁴, and Zhenning Li⁵

5

6

7 ¹ School of Atmospheric Sciences, Sun Yat-sen University, Southern Laboratory of
8 Ocean Science and Engineering (Zhuhai), Zhuhai, Guangdong 519082, China

9 ² Guangdong Province Key Laboratory for Climate Change and Natural Disaster
10 Studies, Sun Yat-sen University, Zhuhai, Guangdong 519082, China

11 ³ Institute of Plateau Meteorology, China Meteorological Administration, Chengdu,
12 Sichuan 610072, China

13 ⁴ Shenzhen Wiselec Technology Co., Ltd., Shenzhen, Guangdong 518048, China

14 ⁵ Division of Environment and Sustainability, The Hong Kong University of Science
15 and Technology, Hong Kong, China

16

17 Submitted to *Nature Communications* in May 2022

18

19

20 **Corresponding authors & addresses:* Dr. Xingwen Jiang, Institute of Plateau
21 Meteorology, China Meteorological Administration, Chengdu, Sichuan 610072, China,
22 e-mail: xingwen.jiang@yahoo.com. ORCID: 0000-0003-1727-9266

23 Prediction of the South Asian summer monsoon (SASM) has remained a challenge for
24 both scientific research and operational climate prediction for decades. By identifying
25 two dominant modes of the SASM, we show that the unsatisfactory prediction may be
26 due to the fact that the existing SASM indices are mostly related to the less predictable
27 second mode. The first mode, in fact, is highly predictable. It is physically linked to the
28 variation of the Indian monsoon trough coupled with large rainfall anomalies over core
29 monsoon zone and the northern Bay of Bengal. An index is constructed as a physical
30 proxy of this first mode, which can be well predicted one season in advance, with an
31 overall skill of 0.698 for 1979-2020. This result suggests a predictable prospect of the
32 SASM, and we recommend the new index for real-time monitoring and prediction of
33 the SASM.
34

35 The South Asian summer monsoon (SASM) exhibits significant variations in rainfall
36 and atmospheric circulation on a wide range of time scales^{1,2}. The interannual
37 variability of SASM rainfall not only affects the lives of more than one billion people³,
38 but also regulates local and remote atmospheric circulations by releasing diabatic
39 heating⁴⁻⁶. Thus, substantial effort has been devoted to understanding the mechanisms
40 and prediction of the SASM⁴⁻⁸. However, the operational seasonal prediction skill of
41 the interannual variability of the SASM is quite low in recent decades, in terms of the
42 popular All India Rainfall Index (AIRI)⁸. The AIRI is constructed based only on rainfall
43 over the Indian subcontinent, ignoring the fact that a large amount of the SASM rainfall
44 appears over the Bay of Bengal (BOB), which is the location of the maximum center
45 of global atmospheric heat source⁹.

46 In addition to the AIRI, other indices have been proposed to quantify the interannual
47 variability of the SASM based on rainfall and wind shear over South Asia. Dynamic
48 indices are constructed by using the convection-circulation relationship in the tropics¹⁰
49 and the key features of the SASM, which include the lower-level westerly and upper-
50 level easterly jet streams and local Hadley circulation^{11,12}. One of the salient features of
51 the SASM is the Indian monsoon trough (IMT), which extends from the northern BOB
52 to western India¹³. The mesoscale convective systems embedded in the monsoon trough
53 contribute a large proportion of rainfall to South Asia^{14,15}). For example, most monsoon
54 depressions, which account for almost all extreme rainfall events (rainfall >100 mm
55 day⁻¹) over central India, form over the warm waters of the northern BOB and move
56 westward or northwestward along the IMT^{16,17}. The dynamic and thermodynamic

57 conditions differ over the eastern and western flanks of the IMT¹⁸. Nevertheless, the
58 variation of the IMT over the BOB is not well considered by the existing monsoon
59 indices.

60 The interannual variability of the SASM is forced by the sea-surface temperature
61 (SST) anomalies related to El Niño-Southern Oscillation (ENSO), and their relationship
62 has undergone apparent interdecadal changes^{19,20}. The AIRI mostly represents the first
63 leading mode of the Indian summer monsoon rainfall, which is concurrently associated
64 with equatorial Pacific SST anomalies. The second mode is related to the IMT over
65 northern India, which exhibits different rainfall anomalies and can be partially
66 explained as a lagged response to ENSO²¹. It is not easy to predict the SASM
67 successfully based on the AIRI due to the spring predictability barrier related to ENSO¹¹.
68 Nevertheless, because of the relatively persistent SST anomalies during the ENSO
69 decaying phase²², the variability of the SASM related to the IMT may be better
70 predicted compared to the AIRI.

71 This study is aimed at investigating the interannual variability of the SASM by
72 considering the coupling variability between rainfall and atmospheric circulation over
73 the entire South Asia rather than just the rainfall over the Indian subcontinent or large-
74 scale circulation. Our focus is on the roles of convection and monsoon trough over the
75 northern BOB in the interannual variability of the SASM. An index is proposed to
76 represent the variability of the first mode, which reflects the tight coupled features of
77 the SASM rainfall and low-level wind and is distinct from the anomalous patterns
78 associated with the previous well-known monsoon indices. This monsoon index is

79 closely related to tropical SST anomalies from the previous winter to the simultaneous
80 summer, and can be well predicted one season ahead, exhibiting a predictable prospect
81 of the SASM.

82

83 **Results**

84 **Physical interpretations of the dominant SASM modes.** Singular value
85 decomposition (SVD) analysis is performed based on normalized and detrended 850-
86 hPa wind and rainfall over South Asia in June-August (JJA) during 1979-2020, to
87 reflect the unique coupling characteristics between rainfall and low-level circulation of
88 the SASM system. We focus on the two leading modes that account for 75.89% of the
89 total squared covariance fraction and are statistically distinguished from the rest of the
90 eigenvectors according to the Monte Carlo approach. The first SVD (SVD1 hereafter)
91 mode, accounting for 51.91% of the total squared covariance, is depicted by an
92 anomalous low-level cyclonic circulation over the northern BOB and region south of
93 the Tibetan Plateau (TP) (Fig. 1a) and by a nearly northwest-southeast elongated
94 sandwich-like pattern of rainfall anomalies, with positive anomalies along the central
95 India-northeastern BOB region and negative anomalies on its two sides (Fig. 1c).
96 Overall, positive centers of rainfall anomalies overlap the heavy climatic rainfall
97 centers along central India-northeastern BOB, and negative centers of rainfall
98 anomalies overlap the heavy climatic rainfalls over the adjacent seas along the west
99 coast of India and the southeastern edge of the TP (Fig. 1c, Supplementary Fig. 1). The
100 anomalous cyclonic circulation is located to the central-east of the climatic IMT (Fig.

101 1a, Supplementary Fig. 1), signifying a strengthening and eastward extension of the
102 IMT. The SVD1 mode reveals a highly coupled feature of atmospheric circulation and
103 rainfall for the SASM, and the corresponding time series (PC1 hereafter) of SVD1 for
104 850-hPa wind and rainfall yields a significant correlation coefficient of 0.937 (Fig. 1e).

105 The second SVD (SVD2 hereafter) mode is responsible for 23.98% of the total
106 squared covariance (left panels of Fig. 1). The wind field associated with SVD2
107 displays an anomalous low-level cyclonic circulation over the Indian subcontinent, with
108 westerly anomalies stretching from the Arabian Sea to India south of 20°N and easterly
109 anomalies along the southern edge of the TP (Fig. 1b). Negative anomalies of rainfall
110 appear over the southeastern edge of the TP and positive anomalies over the other parts
111 of South Asia (Fig. 1d). The corresponding time series (PC2 hereafter) of SVD2 for
112 850-hPa wind and rainfall yield a correlation coefficient of 0.870. Note that the positive
113 center covering the eastern TP, Pakistan, central-northern India, and the interior of the
114 Indochina subcontinent does not well overlap with heavy climatic rainfall (Fig. 1d,
115 Supplementary Fig. 1). Therefore, the overwhelmingly dominant contribution of SVD1
116 mode in the coupling of the SASM system is proposed, in the concept of its explained
117 covariance and a high degree of coincidence with the climatological features.
118 Broadening or narrowing the domain, or focusing on June-September (JJAS) for the
119 SVD analysis yields almost identical results (Supplementary Figs. 2-4), suggesting the
120 robustness of the two leading modes discussed above.

121 To identify the physical processes associated with the two leading modes, we
122 investigate the composite results with respect to PC1 and PC2 (Fig. 2). The positive

123 years are picked when both the PCs of 850-hPa wind and rainfall are above 0.8 standard
124 deviations, and the negative years, below -0.8 standard deviations. The results are not
125 significantly different when the threshold is changed to 1.0 standard deviation (figure
126 not shown). In summer, there is a northwest-southeast elongated belt of low pressure
127 extending from Rajasthan, India to the BOB (Figs. 2a-d), namely, the IMT. Such a low-
128 pressure system brings sufficient rainfall to the central India-northeastern BOB region.
129 Interaction between low-level westerly winds and terrain forms another heavy rainfall
130 zone along the west coast of India and its adjacent seas. During the positive phase of
131 SVD1, the IMT strengthens and extends eastward, with stronger westerlies on its
132 southern flank and anomalous easterlies over region south of the TP (Figs. 2a, 2c, and
133 2e). Correspondingly, rainfall increases over central India-northeastern BOB, and
134 decreases over the southeastern edge of the TP possibly due to a weaker moisture
135 transport by the anomalous easterlies. The decreased rainfall over southern India and
136 its adjacent seas is accompanied by increased geopotential height over the region (Figs.
137 2a, 2c, and 2e). On the other hand, the IMT is more likely to shift southwestward during
138 the positive phase of SVD2, producing an anomalous cyclonic circulation and above-
139 normal rainfall over the Indian subcontinent (Figs. 2b, 2d, and 2f). The decreased
140 rainfall over the southeastern edge of the TP is consistent with the weakened
141 southwesterlies (or northeasterly anomalies).

142 It has long been recognized that ENSO can remotely modulate the SASM, through
143 the inter-seasonal thermal storage in the Indian Ocean, or the simultaneous atmospheric
144 forcing^{21,23-28}. We find that the SVD1 mode is significantly correlated with the SSTs in

145 the central-eastern tropical Pacific and the Indian Ocean in the preceding winter and
146 spring seasons, reminiscent of an ENSO decaying process (left panels of
147 Supplementary Fig. 5). On the other hand, significant SST signals appear over the
148 central-eastern tropical Pacific in the simultaneous summer, but significant signals can
149 hardly be found in the previous seasons for the SVD2 mode (right panels of
150 Supplementary Fig. 5). The correlation coefficients between PC1 of rainfall and the
151 Niño-3 (Niño-3.4) index in preceding winter and simultaneous summer are -0.581 (-
152 0.542) and 0.199 (0.340), respectively; they are 0.001 (0.021) and -0.267 (-0.343) for
153 the PC2 of rainfall, respectively. Note that the 95% confidence level by the Student-*t*
154 test is 0.304. Similar results are obtained when the PCs of 850-hPa winds are applied.
155 The above analysis suggests that the SVD1 mode exhibits a lagged response to ENSO
156 whereas the SVD2 mode is connected simultaneously with ENSO, illustrating the
157 superiority in monsoon predictability of the first mode of the SASM.

158

159 **Mining an index capable of proxying the predictable dominant mode.** Although the
160 PC1 provides a robust index, it is not easily applicable for real-time monitoring and
161 prediction of the SASM. Mining a simple index with a strong physical connection to
162 this predictable dominant mode is much needed.

163 Table 1 lists nine existing widely used indices for the SASM, which can be
164 dynamically classified into four categories. The first category interprets the north-south
165 thermal contrast, i.e., the Webster-Yang index (WYI)¹¹ defined by the vertical shear of
166 zonal winds, and the monsoon Hadley circulation index (hereafter MHI)¹² defined as

167 the area-averaged meridional wind shear over South Asia. The second category
168 represents the overall convection (rainfall) intensity for South Asia or the Indian
169 subcontinent, including the AIRI²⁹, the extended Indian monsoon rainfall index (EIMRI)
170 proposed by Goswami¹², and the convection index (CI) proposed by Wang and Fan¹⁰.
171 The third category is a shear vorticity index termed the Indian monsoon index (IMI)³⁰.
172 Finally, some indices are based on the seasonality of wind field, i.e., the three SASM
173 indices (briefly denoted as the SASMI, SASMI1, and SASMI2) defined as the area-
174 averaged dynamically normalized seasonality (DNS) within different domains of South
175 Asia^{31,32}. Given that the first mode is actually a reflection of the intensity and zonal
176 position of the IMT (mostly its eastern flank), which is tightly coupled with rainfall
177 variation, these indices can hardly represent its physical proxy perfectly. A new
178 dynamic index needs to be constructed to fill this gap for monsoon research.

179 Motivated by the physical interpretation of the SVD1 mode, we define a new index
180 using the area-averaged 850-hPa zonal wind over (85°-100°E, 10°-18°N) minus that
181 over (83°-95°E, 23°-27°N). The two domains well overlap with the southern and
182 northern flanks of the low-level cyclonic anomalies associated with the SVD1 (Fig. 1a),
183 as outlined in Figs. 3a and 3b. Since the basic idea behind this new index is to measure
184 the intensity and zonal position of the IMT, we refer to the index as the IMT index
185 (IMTI). Both the regressions of 850-hPa wind upon the PC1 of rainfall and the IMTI
186 present an anomalous cyclonic circulation from the northern BOB to the south of the
187 TP, and the features are almost identical to each other (Figs. 3a and 3b). The rainfall
188 anomalies associated with the IMTI also show a northwest-southeast elongated

189 sandwich-like pattern like those associated with the SVD1 (Figs. 1b, 4a). The time
190 series of standardized IMTI, shown in Fig. 3c, exhibits consistent positive and negative
191 years with the PCs of the SVD1 (Fig. 1e). The correlation coefficients between the IMTI
192 and PC1s of 850-hPa wind and rainfall are 0.896 and 0.938, respectively, both
193 exceeding the 99.9% confidence level. Therefore, the IMTI is an appropriate physical
194 proxy for the dominant predictable mode of the SASM.

195

196 **Merits of the new index.** To objectively evaluate the performance of multiple indices
197 in measuring the SASM, correlations between each index and each of the two leading
198 modes are examined (Table 2). Most of the existing indices (i.e., the AIRI, EIMR, MHI,
199 CI, and IMI) are significantly correlated with the SVD2 mode, instead of the SVD1
200 mode, while both the SASMI and SASMI2 exhibit reverse relationship with the two
201 modes (Table 2). The other two indices (i.e., the WYI and SASMI1), however, are not
202 well correlated with either SVD1 or SVD2. This is because the WYI primarily measures
203 a broad-scale SASM rather than the regional features and the SASMI1 targets
204 Southwest Asia, which is located west of the study domain.

205 Figure 4 shows the relationships of the indices with rainfall and atmospheric
206 circulation anomalies over South Asia. The regression patterns associated with the AIRI,
207 EIMR, MHI, CI, and IMI present positive anomalies of rainfall over most parts of South
208 Asia and low-level cyclonic circulation anomalies over India, similar to those for the
209 SVD2 mode. Nevertheless, none of these indices reflect rainfall anomalies over the
210 southeastern edge of the TP, with the exception of the IMI. Rainfall and circulation

211 anomalies associated with the SASMI and SASMI2 are similar to those associated with
212 the SVD1 mode. Compared with the IMTI, the SASMI reflects overall weaker rainfall
213 and circulation anomalies, and the SASMI2 is associated with weaker rainfall
214 anomalies over central India and the southeastern edge of the TP. The anomalous pattern
215 associated with the WYI is similar to that with the SVD1 mode, displaying a quite
216 weaker magnitude. On the other hand, significant anomalies only appear over the
217 northeastern corner of India for the SASMI1. Rainfall anomalies derived from the
218 Global Precipitation Climatology Centre (GPCC) and Asian Precipitation-Highly
219 Resolved Observational Data Integration toward Evaluation of Water Resources
220 (APHRODITE) replicate the patterns similar to those from the Global Precipitation
221 Climatology Project (GPCP) (Supplementary Figs. 6-7).

222 Relationships of each index with the SST signals in preceding and simultaneous
223 seasons are examined next. Consistently, the indices that are closely connected with the
224 SVD1 mode (i.e., IMTI, SASMI, and SASMI2) are significantly related to the Niño-3
225 index in the preceding winter, whereas those associated with the SVD2 mode (i.e., AIRI,
226 CI, MHI, and IMI) are more likely correlated with the Niño-3 index in simultaneous
227 summer significantly (Table 2). In particular, the correlation coefficients of the
228 preceding Niño-3 index with the IMTI and SASMI2 are statistically significant at the
229 99.9% confidence level. The WYI is correlated with the simultaneous Niño-3 index
230 highly (-0.632) but with the preceding winter Niño-3 index moderately (-0.330); and
231 the SASMI1 shows a moderate significant correlation with the simultaneous Niño-3
232 index. These results suggest that the IMTI, SASMI, SASMI2, and WYI have a higher

233 predictability than the others, in particular for the IMTI and SASM2. Similar results
234 can be obtained when the Niño-3.4 index is used instead.

235 We thus conclude that the new index has several obvious merits: (1) a strong physical
236 connection and identical representation of the most dominant mode of the SASM, (2)
237 its simplicity for real-time monitoring, and (3) high predictability. Here, a statistical
238 prediction of the IMTI is provided based on the physical linkage between the index and
239 the preceding SST signals. As shown in Fig. 5, the correlation pattern between the IMTI
240 and the SST from preceding winter to simultaneous summer exhibits a mature-decaying
241 process of ENSO, bearing a large resemblance to that for the SVD1 mode
242 (Supplementary Fig. 5). The SSTs in the central-eastern tropical Pacific, tropical Indian
243 Ocean, and tropical Atlantic show high correlations with the IMTI. While the SST
244 signals in the central-eastern tropical Pacific attenuate rapidly after winter season, those
245 in the tropical Indian Ocean and tropical Atlantic persist across the three seasons. We
246 select two predictors in preceding spring for one-season lead prediction: the area-
247 averaged SSTs over the eastern tropical Indian Ocean (90°-130°E, 0°-20°N) and tropical
248 Atlantic (305°-345°E, 10°S-20°N), referred to as the TEIO_SST and TA_SST,
249 respectively. Both the TEIO_SST and TA_SST persist well from spring to summer (see
250 Figs. 5b-c), and are highly correlated with the IMTI (-0.567 and -0.597, respectively).
251 The two predictors are also significantly correlated with the preceding winter ENSO,
252 with correlation coefficients exceeding the 99.9% confidence level (0.681 and 0.399,
253 respectively). On the other hand, only a moderate correlation is found between the two
254 predictors (0.372). Afterwards, the least absolute shrinkage and selection operator

255 (LASSO) is adopted to construct a regression function with the two predictors. The
256 simulation equation derived from the training period of 1979-2010 is $IMTI = -1.642 \times$
257 $TEIO_SST - 1.210 \times TA_SST + 79.319$. The overall skill is quite satisfactory for the
258 entire 42 years (1979-2020), which reaches a significant correlation coefficient of 0.698
259 (exceeding the 99.9% confidence level) between the observed and inferred IMTI (Fig.
260 3c). Specifically, the correlation coefficient between the observed and simulated IMTI
261 is 0.719 for the training period, and there are only two false hits (years 2013 and 2016)
262 during the validation period of 2011-2020 (Fig. 3c).

263

264 **Discussion**

265 The two most dominant modes of the interannual variability of the SASM, which
266 account for the majority of the total squared covariance, have clear implications on both
267 rainfall intensity and atmospheric circulation anomalies. The first mode is physically
268 linked to the intensity and zonal extension (or contraction) of the IMT, which is coupled
269 with rainfall anomalies over the hardest hit area in terms of climatology (i.e., central
270 India-northeastern BOB regions and the southeastern edge of the TP). The second mode
271 is associated with a meridional shift of the IMT, which is coupled with the opposite
272 anomalies of rainfall over the southeastern edge of the TP and the rest of South Asia.
273 The first mode is highly robust and captures more than half of the total squared
274 covariance, and thus is recommended to be an objective measure of the SASM variation.
275 Importantly, this mode is highly predictable, while the second mode shows a low
276 predictability.

277 With elaborations on the dynamic features of different modes, it is not surprising to
278 see disparate measures of the SASM strength given by different indices (e.g., the WYI
279 and AIRI)^{33,34} An explanation for this discrepancy is that the indices are respectively
280 connected to different modes of monsoon variation. Accordingly, the multiple SASM
281 indices can be basically re-classified into two groups: the first mode-related including
282 the IMTI, WYI, SASMI, and SASMI2, and the second mode-related including the AIRI,
283 EIMR, MHI, CI, and IMI. The SASMI1, however, is a reflection of the monsoon
284 features over west of India rather than the core region of South Asia, and thus is
285 excluded from both groups. Most of the existing indices cannot reflect rainfall
286 anomalies over central India or the southeastern edge of the TP. Although the SASMI2
287 shows a comparable correlation with the first mode and the previous tropical SST
288 anomalies as the new index (the IMTI), it has a different physical meaning from the
289 first mode. On the other hand, the simple new index is connected to the same dynamic
290 system (namely, the IMT) and represents the first mode with a high fidelity, showing
291 advantages of representativeness and objectiveness of the coupling SASM system.

292 Most of the existing indices tend to be associated with the second mode (such as the
293 AIRI, EIMR, MHI, CI, and IMI) or moderately correlated with the first mode (such as
294 the WYI and SASMI), which seem to illustrate that the SASM is unpredictable. We
295 point out this is not true by providing compelling evidence in this paper. The new index
296 defined according to the first mode is well predicted by a statistical model one-season
297 ahead, with a skill of $R=0.698$ for the recent 42 years, suggesting that at least 48.7% of
298 the SASM variance is predictable. The statistical equation is constructed solely based

299 on the physical connection between the new index and previous SST anomalies, and
300 we are working hard to reach a higher prediction skill in the near future.

301 The present study points out the excellent predictable prospect of the SASM by
302 highlighting the role of the first mode, different from the conventional wisdom. Here,
303 we provide a further physical understanding of the superior predictability of the first
304 mode than the second mode. The latter is associated with the developing phase of ENSO,
305 and hence its prediction may be limited by the well-known spring predictability barrier.
306 On the other hand, the former is closely linked to persistent SST anomalies from spring
307 to summer during the decaying phase of ENSO. In comparison, the first mode is more
308 closely associated with atmospheric anomalies at the lower latitudes than the second
309 mode (Fig. 1), and is thus more susceptible to the tropical forcing. Moreover, given that
310 large anomalies associated with the first mode appear on the eastern part of South Asia,
311 the TP blocks the influences from the mid-to-high latitudes for the first mode. The
312 second mode, however, is closely linked to the atmospheric forcing at the mid-to-high
313 latitudes and determined largely by the internal atmospheric variability^{35,36} (i.e., the
314 wave train propagating along the westerly jet stream). Another possible reason is that
315 the year-to-year variability of the first mode may be largely due to sampling fluctuations
316 associated with the intra-seasonal variability, regarding to its tight connection with the
317 IMT^{21,37,38}. The 10-20-day and 30-60-day oscillations will introduce considerable
318 tropical forcing for the first mode of interannual variability. This raises another merit
319 of the new index, that is, measuring the SASM on various time scales. As an example,
320 the time series of the standardized 5-day running averaged IMTI in 1994 is shown

321 (Supplementary Fig. 8), which exhibits apparent intra-seasonal oscillation. Patterns of
322 rainfall and circulation anomalies similar to those associated with the first mode are
323 obtained from the composite differences between the active and inactive periods
324 (Supplementary Fig. 8). The multi-scale real-time monitoring and prediction of using
325 the new index can be anticipated.
326

327 **Methods**

328 **Data sets.** The monthly data sets of 850-hPa wind and geopotential height are obtained
329 from the European Centre for Medium-range Weather Forecasts Reanalysis v5³⁹ (ERA5,
330 <https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5>), which have a
331 high resolution of $0.25^{\circ} \times 0.25^{\circ}$. The monthly SST with a resolution of $1^{\circ} \times 1^{\circ}$ is from the
332 Hadley Centre Sea Ice and Sea Surface Temperature data set (HadISST;
333 <https://www.metoffice.gov.uk/hadobs/hadisst>) Version 1.1⁴⁰. The monthly rainfall from
334 the GPCP (<https://psl.noaa.gov/data/gridded/data.gpcp.html>) Version 2.3 with a
335 resolution of $2.5^{\circ} \times 2.5^{\circ}$ is utilized in this study⁴¹. The monthly rainfall from the
336 GPCP^{42,43} (<https://www.dwd.de/EN/ourservices/gpcp/gpcp.html>) Version 2020 with a
337 resolution of $1^{\circ} \times 1^{\circ}$ and the daily rainfall from the APHRODITE
338 (<http://aphrodite.st.hirosaki-u.ac.jp/products.html>) V1101 and V1101EX_R1 are used
339 for validation⁴⁴. All the data sets cover the period of 1979-2020, except for the
340 APHRODITE with a shorter period of 1979-2015. The AIRI for 1979-2019 is obtained
341 from the National Informatics Centre, Ministry of Electronics and Information
342 Technology, Government of India (<https://data.gov.in/catalog/rainfall-india>). The
343 SASMI, SASMI1, and SASMI2 for 1979-2020 are obtained from the author's website
344 (<http://lijianping.cn/dct/page/65576>).

345

346 **SVD analysis.** To identify the major coupled features of atmospheric circulation and
347 rainfall during the SASM, SVD analysis was performed on 850-hPa wind and rainfall
348 over South Asia. The SVD isolates the dominant modes of cross-covariance between

349 the data sets. The output consists of the left matrix and right matrix, and each mode has
 350 a unique eigenvalue representing the percentage of squared covariance explained by the
 351 specific mode, which decreases with the order of the modes. The corresponding
 352 temporal variations (i.e., the PCs) for the two matrices are similar to each other with
 353 high correlation coefficients, signifying their highly coupled features.

354

355 **Correlation and significance test.** Here, correlation coefficient r is calculated as
 356 follows:

$$357 \quad r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

358 where \bar{x} and \bar{y} represent the mean values of variables x_i and y_i from $i=1$ to $i=n$ (n
 359 is the sample size).

360 The statistical significance level is calculated based on the Student's t -test, using the
 361 equation:

$$362 \quad t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{(m-1)s_1^2 + (n-1)s_2^2}{m+n-2} \left(\frac{1}{m} + \frac{1}{n} \right)}} \quad (2)$$

363 where m and n are the sample sizes for variables x_i and y_i , respectively. \bar{x} and \bar{y}
 364 are the mean values of these variables. s_1^2 and s_2^2 are the variances of x_i and y_i ,
 365 while $m + n - 2$ is the degree of freedom.

366 In this study, the values of correlation coefficient (r) above 0.304, 0.393, and 0.490
 367 are used to estimate the 95%, 99%, and 99.9% confidence levels, respectively, for a 42-
 368 year record length of 1979-2020, according to the Student's t -test. The values should
 369 be 0.308, 0.398, and 0.495 for the AIRI correlation with the 41-year record length of

370 1979-2019.

371

372 **LASSO.** The LASSO is one of the extensively used linear regression analysis method,
373 with a regularization term expressed as the L_1 norm of the coefficients in order to
374 enhance the interpretability and prediction skill^{45,46}. In this study, the training period
375 and prediction period are 1979-2010 and 2011-2020, respectively. The 5-fold cross
376 validation is performed in the training period to construct an optimal regression
377 equation. The explained variances of the training period and prediction period are 44.60%
378 and 69.55%, respectively. The root mean square errors of the two periods are both 0.71.

379

380 **Data availability.** All raw data can be accessed via the links provided above.

381

382 **References**

- 383 1. Webster, P. J. et al. Monsoons: Processes, predictability, and the prospects for
384 prediction. *J. Geophys. Res. Oceans* **103**, 14451-14510 (1998).
- 385 2. Gadgil, S. The Indian monsoon and its variability. *Annu. Rev. Earth Planet. Sci.* **31**,
386 429-467 (2003).
- 387 3. Turner, A. G. & Annamalai, H. Climate change and the South Asian summer
388 monsoon. *Nat. Clim. Chang.* **2**, 587-595 (2012).
- 389 4. Rodwell, M. J. & Hoskins, B. J. Monsoons and the dynamics of deserts. *Q. J. R.*
390 *Meteorol. Soc.* **122**, 1385-1404 (1996).
- 391 5. Rodwell, M. J. & Hoskins, B. J. Subtropical anticyclones and summer monsoons. *J.*
392 *Clim.* **14**, 3192-3211 (2001).
- 393 6. Walker, J. M., Bordoni, S. & Schneider, T. Interannual variability in the large-scale
394 dynamics of the South Asian summer monsoon. *J. Clim.* **28**, 3731-3750 (2015).
- 395 7. Jiang, X. et al. Seasonal-to-interannual prediction of the Asian summer monsoon in
396 the NCEP Climate Forecast System version 2. *J. Clim.* **26**, 3708-3727 (2013).
- 397 8. Wang, B. et al. Rethinking Indian monsoon rainfall prediction in the context of recent
398 global warming. *Nat. Commun.* **6**, 7154 (2015).
- 399 9. Yanai, M. & Tomita, T. Seasonal and interannual variability of atmospheric heat
400 sources and moisture sinks as determined from NCEP–NCAR reanalysis. *J. Clim.*
401 **11**, 463-482 (1998).
- 402 10. Wang, B. & Fan, Z. Choice of South Asian summer monsoon indices. *Bull. Am.*
403 *Meteorol. Soc.* **80**, 629-638 (1999).

- 404 11. Webster, P. J. & Yang, S. Monsoon and ENSO: Selectively interactive systems. *Q. J. R. Meteorol. Soc.* **118**, 877-926 (1992).
- 405
- 406 12. Goswami, B. N., Krishnamurthy, V. & Annmalai, H. A broad-scale circulation index
407 for the interannual variability of the Indian summer monsoon. *Q. J. R. Meteorol. Soc.*
408 **125**, 611-633 (1999).
- 409 13. Rao, Y. P. *Southwest monsoon*. (Indian Meteorol. Dep., New Delhi, 1976).
- 410 14. You, Y. & Ting, M. Observed trends in the South Asian monsoon low-pressure
411 systems and rainfall extremes since the late 1970s. *Geophys. Res. Lett.* **48**,
412 e2021GL092378 (2021).
- 413 15. Das, P. K. *Monsoons*. (World Meteorol. Organ., Geneva, 1986).
- 414 16. Dhar, O. N. & Nandargi, S. On some characteristics of severe rainstorms of India.
415 *Theor. Appl. Climatol.* **50**, 205-212 (1995).
- 416 17. Ajayamohan, R. S., Merryfield, W. J. & Kharin, V. V. Increasing trend of synoptic
417 activity and its relationship with extreme rain events over central India. *J. Clim.* **23**,
418 1004-1013 (2010).
- 419 18. Sikka, D. R. & Narasimha, R. Genesis of the monsoon trough boundary layer
420 experiment (MONTBLEX). *Proc. Indian Acad. Sci. (Earth Planet. Sci.)* **104**, 157-
421 187 (1995).
- 422 19. Kumar, K. K., Rajagopalan, B., Hoerling, M., Bates, G. & Cane, M. A. Unraveling
423 the mystery of Indian monsoon failure during El Niño. *Science* **314**, 115-119 (2006).
- 424 20. Kumar, K. K., Rajagopalan, B. & Cane, M. A. On the weakening relationship
425 between the Indian monsoon and ENSO. *Science* **284**, 2156-2159 (1999).

- 426 21. Mishra, V., Smoliak, B. V., Lettenmaier, D. P. & Wallace, J. M. A prominent pattern
427 of year-to-year variability in Indian Summer monsoon rainfall. *Proc. Natl. Acad. Sci.*
428 *USA* **109**, 7213-7217 (2012).
- 429 22. Xie, S.-P. et al. Indian Ocean capacitor effect on Indo-western Pacific climate during
430 the summer following El Niño. *J. Clim.* **22**, 730-747 (2009).
- 431 23. Shukla, J. *Interannual variability of monsoons*. (Wiley, New York, 1987).
- 432 24. Wainer, I. & Webster, P. J. Monsoon/El Niño-Southern Oscillation relationships in
433 a simple coupled ocean-atmosphere model. *J. Geophys. Res. Oceans* **101**, 25599-
434 25614 (1996).
- 435 25. Annamalai, H., Hamilton, K. & Sperber, K. R. The South Asian summer monsoon
436 and its relationship with ENSO in the IPCC AR4 simulations. *J. Clim.* **20**, 1071-
437 1092 (2007).
- 438 26. Fan, F. et al. Revisiting the relationship between the South Asian summer monsoon
439 drought and El Niño warming pattern. *Atmos. Sci. Lett.* **18**, 175-182 (2017).
- 440 27. Miyakoda, K., Kinter III, J. L. & Yang, S. The role of ENSO in the South Asian
441 monsoon and pre-monsoon signals over the Tibetan Plateau. *J. Meteorol. Soc. Jpn.*
442 *Ser. II* **81**, 1015-1039 (2003).
- 443 28. Chakraborty, A. & Singhai, P. Asymmetric response of the Indian summer monsoon
444 to positive and negative phases of major tropical climate patterns. *Sci. Rep.* **11**, 22561
445 (2021).
- 446 29. Parthasarathy, B., Kumar, K. & Kothawale. Indian summer monsoon rainfall indices:
447 1871-1990. *Meteorol. Mag.* **121**, 174-186 (1992).

- 448 30. Wang, B., Wu, R. & Lau, K.-M. Interannual variability of the Asian summer
449 monsoon: Contrasts between the Indian and the western North Pacific–East Asian
450 monsoons. *J. Clim.* **14**, 4073-4090 (2001).
- 451 31. Li, J. & Zeng, Q. A unified monsoon index. *Geophys. Res. Lett.* **29**, 115111-115114
452 (2002).
- 453 32. Li, J. & Zeng, Q. A new monsoon index and the geographical distribution of the
454 global monsoons. *Adv. Atmos. Sci.* **20**, 299-302 (2003).
- 455 33. Krishnamurti, T. N., Bedi, H. S. & Subramaniam, M. The summer monsoon of 1988.
456 *Meteorol. Atmos. Phys.* **42**, 19-37 (1990).
- 457 34. Parthasarathy, B., Munot, A. A. & Kothawale, D. R. All-India monthly and seasonal
458 rainfall series: 1871-1993. *Theor. Appl. Climatol.* **49**, 217-224 (1994).
- 459 35. Ding, Q. & Wang, B. Circumglobal teleconnection in the northern Hemisphere
460 summer. *J. Clim.* **18**, 3483-3505 (2005).
- 461 36. Krishnan, R., Kumar, V., Sugi, M. & Yoshimura, J. Internal feedbacks from
462 monsoon–midlatitude interactions during droughts in the Indian summer monsoon.
463 *J. Atmos. Sci.* **66**, 553-578 (2009).
- 464 37. Krishnamurthy, V. & Shukla, J. Intraseasonal and interannual variability of rainfall
465 over India. *J. Clim.* **13**, 4366-4377 (2000).
- 466 38. Krishnamurthy, V. & Shukla, J. Seasonal persistence and propagation of
467 intraseasonal patterns over the Indian monsoon region. *Clim. Dyn.* **30**, 353-369
468 (2008).
- 469 39. Hersbach, H. et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999-

- 470 2049 (2020).
- 471 40. Rayner, N. A. et al. Global analyses of sea surface temperature, sea ice, and night
472 marine air temperature since the late nineteenth century. *J. Geophys. Res. Atmos.* **108**
473 (2003).
- 474 41. Adler, R. F. et al. The Global Precipitation Climatology Project (GPCP) monthly
475 analysis (New Version 2.3) and a review of 2017 global precipitation. *Atmos.* **9**, 138
476 (2018).
- 477 42. Schneider, U., Becker, A., Finger, P., Rustemeier, E.& Ziese, M. GPCP monitoring
478 product: Near real-time monthly land-surface precipitation from rain-gauges based
479 on SYNOP and CLIMAT data.
480 http://dx.doi.org/10.5676/DWD_GPCP/MP_M_V2020_100 (2020a).
- 481 43. Schneider, U., Becker, A., Finger, P., Rustemeier, E.& Ziese, M. GPCP full data
482 monthly product version 2020 at 1.0°: Monthly land-surface precipitation from rain-
483 gauges built on GTS-based and historical data.
484 http://dx.doi.org/10.5676/DWD_GPCP/FD_M_V2020_100 (2020b).
- 485 44. Yatagai, A. et al. APHRODITE: Constructing a long-term daily gridded
486 precipitation dataset for Asia based on a dense network of rain gauges. *Bull. Am.*
487 *Meteorol. Soc.* **93**, 1401-1415 (2012).
- 488 45. Tibshirani, R. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc., Ser.*
489 *B (Methodol.)* **58**, 267-288 (1996).
- 490 46. Tibshirani, R. Regression shrinkage and selection via the Lasso: A retrospective. *J.*
491 *R. Stat. Soc., Ser. B (Methodol.)* **73**, 273-282 (2011).

492 **Acknowledgements**

493 This study was jointly supported by the National Key R&D Program of China
494 (2019YFC1510400), the National Natural Science Foundation of China (Grants
495 42088101, 42105015, and 42175023), and the Guangdong Province Key Laboratory
496 for Climate Change and Natural Disaster Studies (Grant 2020B1212060025).

497

498 **Author Contributions**

499 X. J. conceived the idea for the study, and T. Z. downloaded the data and performed the
500 calculations. All authors discussed the results throughout the whole process. X. J. and
501 T. Z. drafted the manuscript, and S.Y. contributed to the revision.

502

503 **Competing interests**

504 The authors declare no competing interests.

505

506

507 **Figure legends**

508 **Figure 1 | The first and second SVD modes of SASM.** Spatial patterns of (a-b) 850-
509 hPa wind (m s^{-1} ; vector) and (c-d) rainfall (mm day^{-1} ; shading), and (e-f) corresponding
510 standardized PCs for the first (left) and second (right) SVD modes in summer (JJA).
511 The explained squared covariance of each mode is presented at the top right of (a-b).
512 The magenta and blue curves in (e-f) represent the PCs for 850-hPa wind and rainfall,
513 respectively, and their correlation coefficient is presented at the top right. Values within
514 the gray zone (absolute values ≤ 0.8) represent the normal years; and those below and
515 above the gray zone are for the negative and positive years, respectively.

516

517 **Figure 2 | Atmospheric processes associated with the two leading SVD modes.**

518 Composite patterns of rainfall (mm day^{-1} ; shading), 850-hPa wind (m s^{-1} ; vector), and
519 850-hPa geopotential height (m; contour) in (a-b) positive and (c-d) negative years with
520 respect to the PCs of rainfall. Composite differences (positive minus negative) in 850-
521 hPa wind (m s^{-1} ; vector) and geopotential height (m; shading) with respect to (e) PC1
522 and (f) PC2. Significant values exceeding the 95% confidence level are marked by
523 magenta dots or black vectors in (e-f).

524

525 **Figure 3 | Definition and prediction of the IMTI.** Regressions of 850-hPa wind (m s^{-1} ;

526 m s^{-1} ; vector) against (a) PC1 of rainfall and (b) IMTI. (c) Time series of the standardized

527 IMTI in observation (black curve) and prediction (red curve). In (a-b), shading denotes

528 zonal wind speed. Significant values exceeding the 95% confidence level are marked

529 by magenta dots or black vectors. The blue boxes outline the regions for the definition
530 of the IMTI. In (c), “R” and “R (1979-2010)” represent the correlation coefficients of
531 observed IMTA with inferred IMTI for 1979-2010 and 1979-2020, respectively. The
532 vertical dashed line denotes the line of demarcation between the training period (1979-
533 2010) and prediction period (2011-2020). Observed values within the gray zone
534 (absolute values ≤ 0.8) represent the normal years; and those below and above the gray
535 zone are for the negative and positive years, respectively.

536

537 **Figure 4 | Spatial patterns of 850-hPa wind and rainfall associated with SASM**
538 **indices.** Regressions of 850-hPa wind (m s^{-1} ; vector) and rainfall (mm day^{-1} ; shading)
539 against (a) IMTI, (b) WYI, (c) AIRI, (d) EIMR, (e) MHI, (f) CI, (g) IMI, (h) SASMI,
540 (i) SASMI1, and (j) SASMI2. Significant values exceeding the 95% confidence level
541 are marked by magenta dots or black vectors.

542

543 **Figure 5 | IMTI-predictor correlations.** Correlation of SST in (a) preceding winter,
544 (b) preceding spring, and (c) simultaneous summer with the IMTI. The absolute values
545 above 0.25, 0.30, 0.39, and 0.49 indicate that they significantly exceed the 90%, 95%,
546 99%, and 99.9% confidence levels, respectively. The red boxes outline the regions for
547 the predictors.

548

Table 1	Description of the SASM indices.
Define variables, levels (hPa), and regions	
WYI ¹¹	u, 850-200, (40°-110°E, 0°-20°N)
AIRI ²⁹	Rainfall, all the Indian sub-divisions
EIMRI ¹²	Rainfall, (70°-110°E, 10°-30°N,)
MHI ¹²	v, 850-200, (70°-110°E, 10°-30°N,)
CI ¹⁰	OLR, (70-100°E, 10°-25°N)
IMI ³⁰	u, 850, (40°-80°E, 5°-15°N) - (70°-90°E, 20°-30°N)
SASMI ³¹	DNS, 850, (35°-97.5°E, 5°-22.5°N)
SASMI1 ³¹	DNS, 850, (35°-70°E, 2.5°-20°N)
SASMI2 ³¹	DNS, 850, (70°-110°E, 2.5°-20°N)
<p>Note: u and v represent zonal and meridional wind components, respectively. OLR denotes the out-going longwave radiation. The DNS^{31,32} index is given by $\sigma_{m,n} = \frac{\ \bar{V}_1 - V_{m,n}\ }{\ \bar{V}\ } - 2$, where \bar{V}_1, \bar{V}, and $V_{m,n}$ represent climatological January wind vector, the mean of climatological January and July wind vectors, and the monthly wind vector in the year n and month m, respectively. For a specific variable A, $\ A\ = (\iint_S A ^2 dS)^{1/2}$, where S denotes the integration domain.</p>	

549

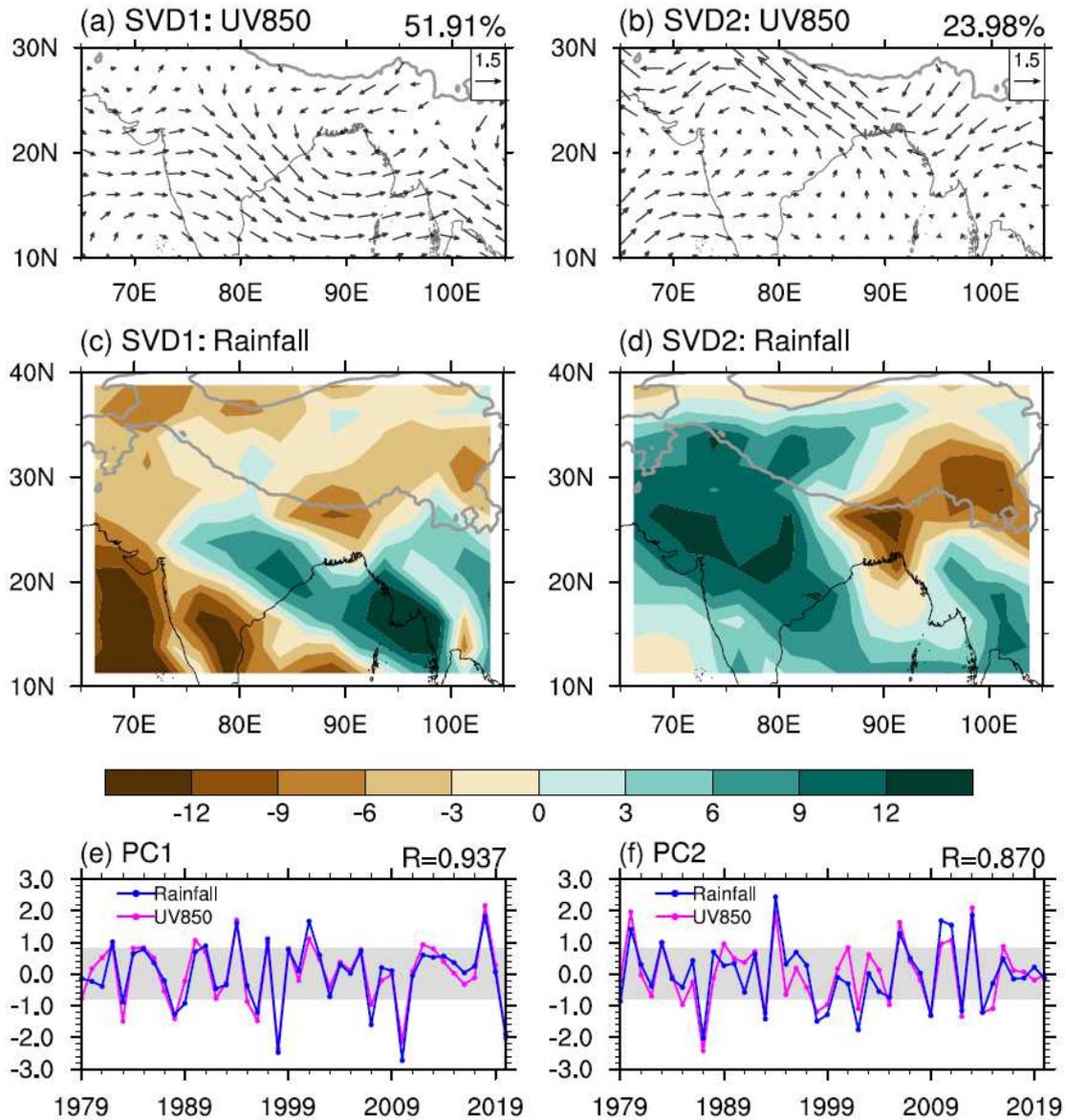
550

Table 2	Correlation coefficients between the SASM indices and PCs, and the preceding winter and simultaneous summer Niño-3 indices.			
	PC1	PC2	Niño3_DJF	Niño3_JJA
IMTI	0.896	0.246	-0.472	0.190
WYI	0.268	0.148	<i>-0.330</i>	-0.632
AIRI	-0.136	0.790	0.057	-0.442
EIMR	0.063	0.616	-0.001	-0.253
MHI	-0.129	<i>0.373</i>	-0.012	<i>-0.392</i>
CI	-0.025	0.480	0.025	-0.491
IMI	0.099	0.686	0.103	-0.486
SASMI	0.439	0.177	<i>-0.335</i>	-0.235
SASMI1	0.059	0.158	-0.095	<i>-0.336</i>
SASMI2	-0.878	-0.070	-0.476	0.119

Column 1 lists the multiple SASM indices, and columns 2-5 present correlation coefficients for the first and second PCs of rainfall, and the Niño-3 index in the preceding winter and simultaneous summer, respectively. Boldface and italics denote the values exceeding the 99% and 95% confidence levels, respectively.

551

552



553

554 **Figure 1 | The first and second SVD modes of SASM.** Spatial patterns of (a-b) 850-

555 hPa wind (m s^{-1} ; vector) and (c-d) rainfall (mm day^{-1} ; shading), and (e-f) corresponding

556 standardized PCs for the first (left) and second (right) SVD modes in summer (JJA).

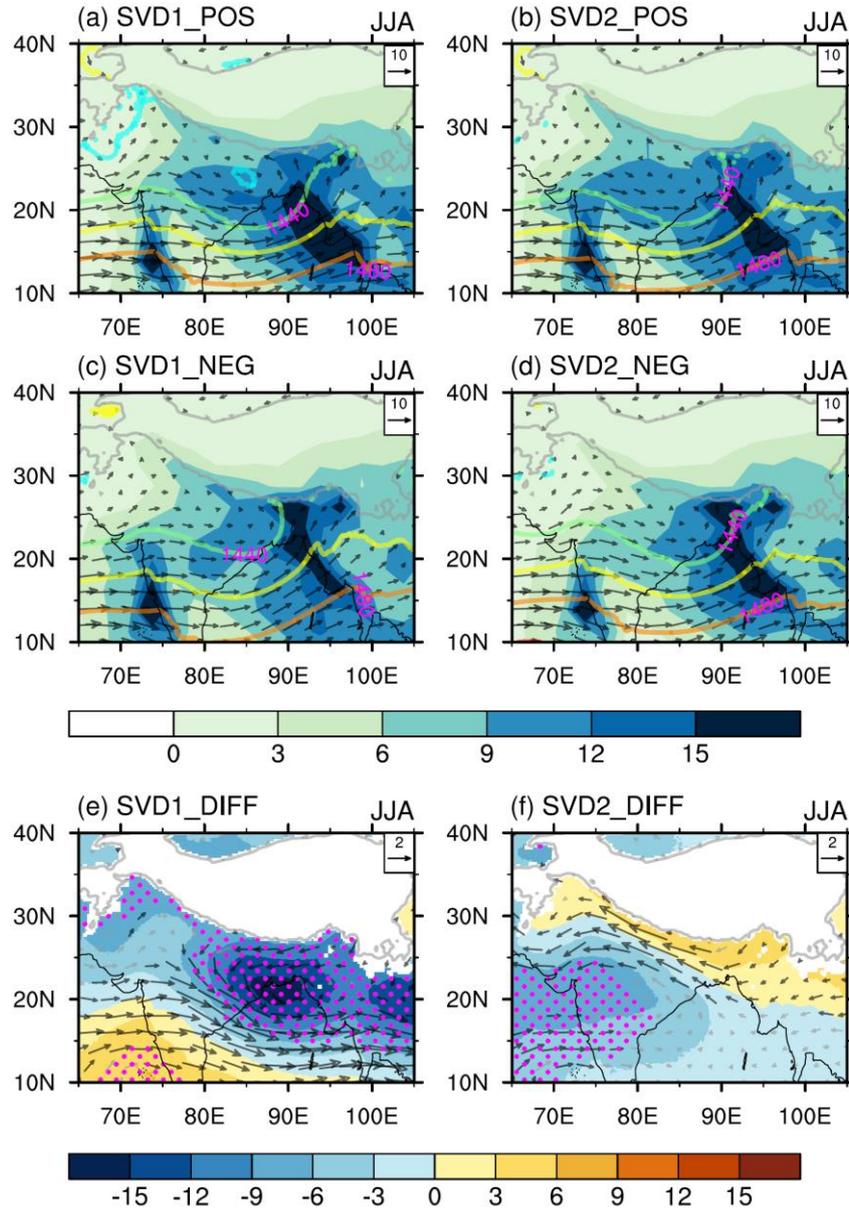
557 The explained squared covariance of each mode is presented at the top right of (a-b).

558 The magenta and blue curves in (e-f) represent the PCs for 850-hPa wind and rainfall,

559 respectively, and their correlation coefficient is presented at the top right. Values within

560 the gray zone (absolute values ≤ 0.8) represent the normal years; and those below and

561 above the gray zone are for the negative and positive years, respectively.



562

563 **Figure 2 | Atmospheric processes associated with the two leading SVD modes.**

564 Composite patterns of rainfall (mm day⁻¹; shading), 850-hPa wind (m s⁻¹; vector), and

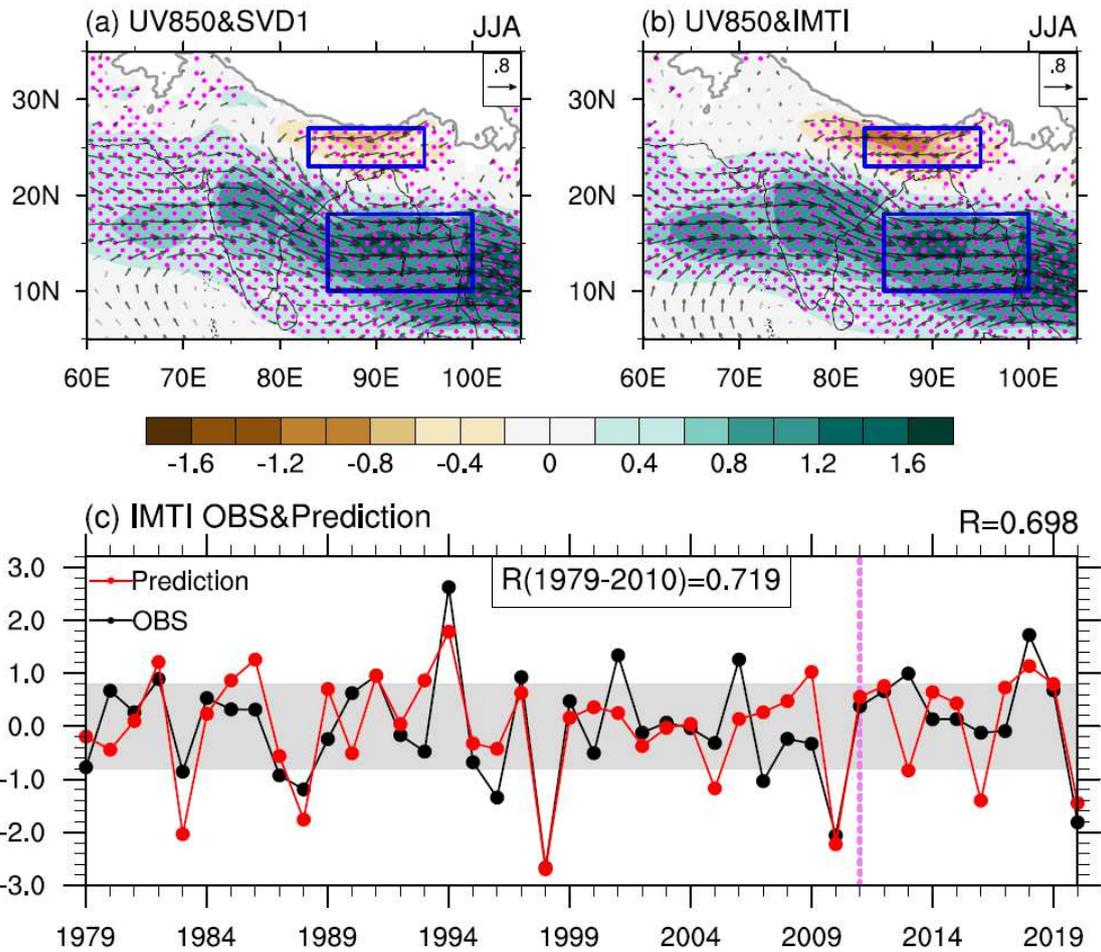
565 850-hPa geopotential height (m; contour) in (a-b) positive and (c-d) negative years with

566 respect to the PCs of rainfall. Composite differences (positive minus negative) in 850-

567 hPa wind (m s⁻¹; vector) and geopotential height (m; shading) with respect to (e) PC1

568 and (f) PC2. Significant values exceeding the 95% confidence level are marked by

569 magenta dots or black vectors in (e-f).



570

571 **Figure 3 | Definition and prediction of the IMTI.** Regressions of 850-hPa wind (m s^{-1} ;

572 \vec{v} ; vector) against (a) PC1 of rainfall and (b) IMTI. (c) Time series of the standardized

573 IMTI in observation (black curve) and prediction (red curve). In (a-b), shading denotes

574 zonal wind speed. Significant values exceeding the 95% confidence level are marked

575 by magenta dots or black vectors. The blue boxes outline the regions for the definition

576 of the IMTI. In (c), “R” and “R (1979-2010)” represent the correlation coefficients of

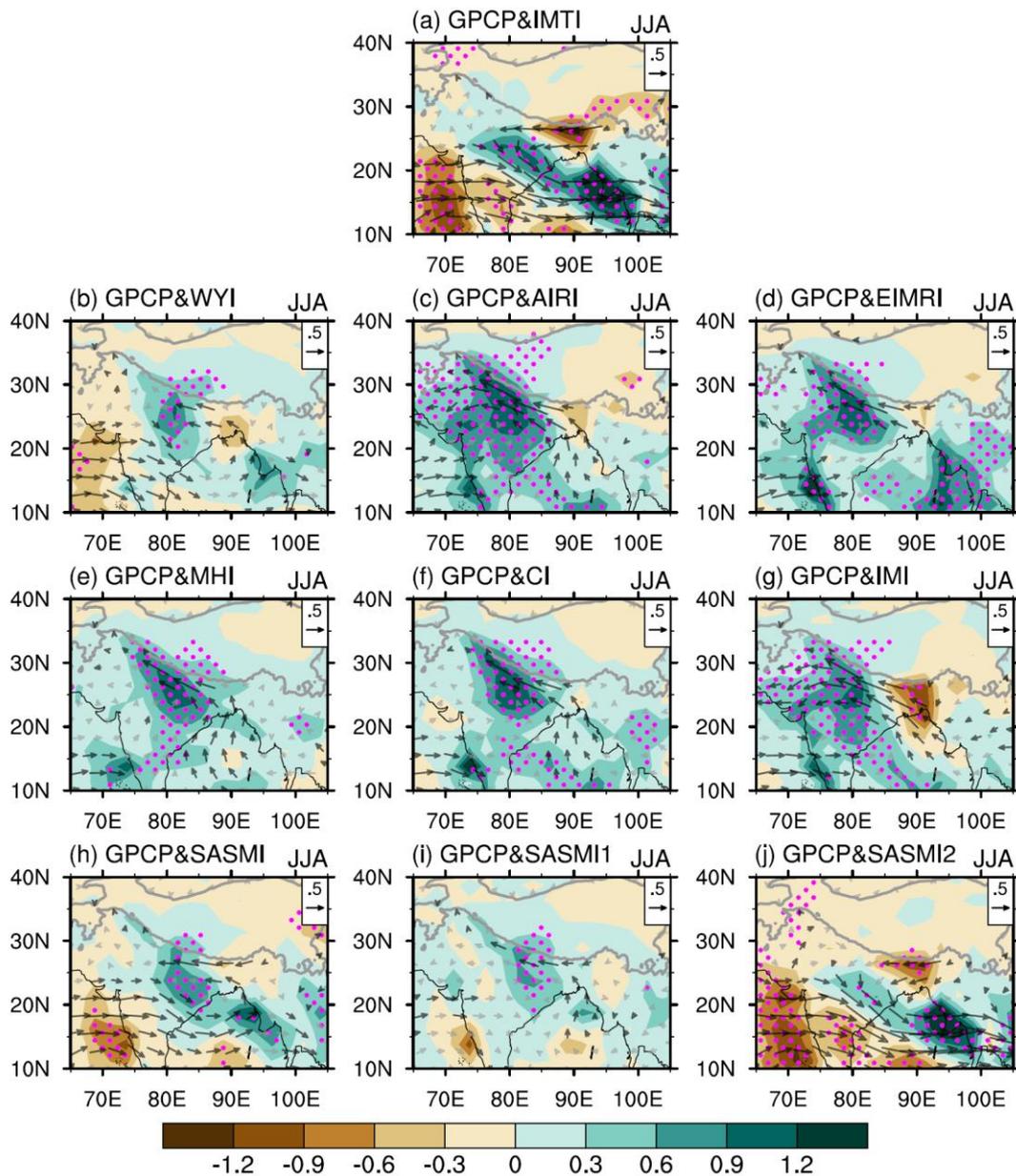
577 observed IMTI with inferred IMTI for 1979-2010 and 1979-2020, respectively. The

578 vertical dashed line denotes the line of demarcation between the training period (1979-

579 2010) and prediction period (2011-2020). Observed values within the gray zone

580 (absolute values ≤ 0.8) represent the normal years; and those below and above the gray

581 zone are for the negative and positive years, respectively.



582

583 **Figure 4 | Spatial patterns of 850-hPa wind and rainfall associated with SASM**

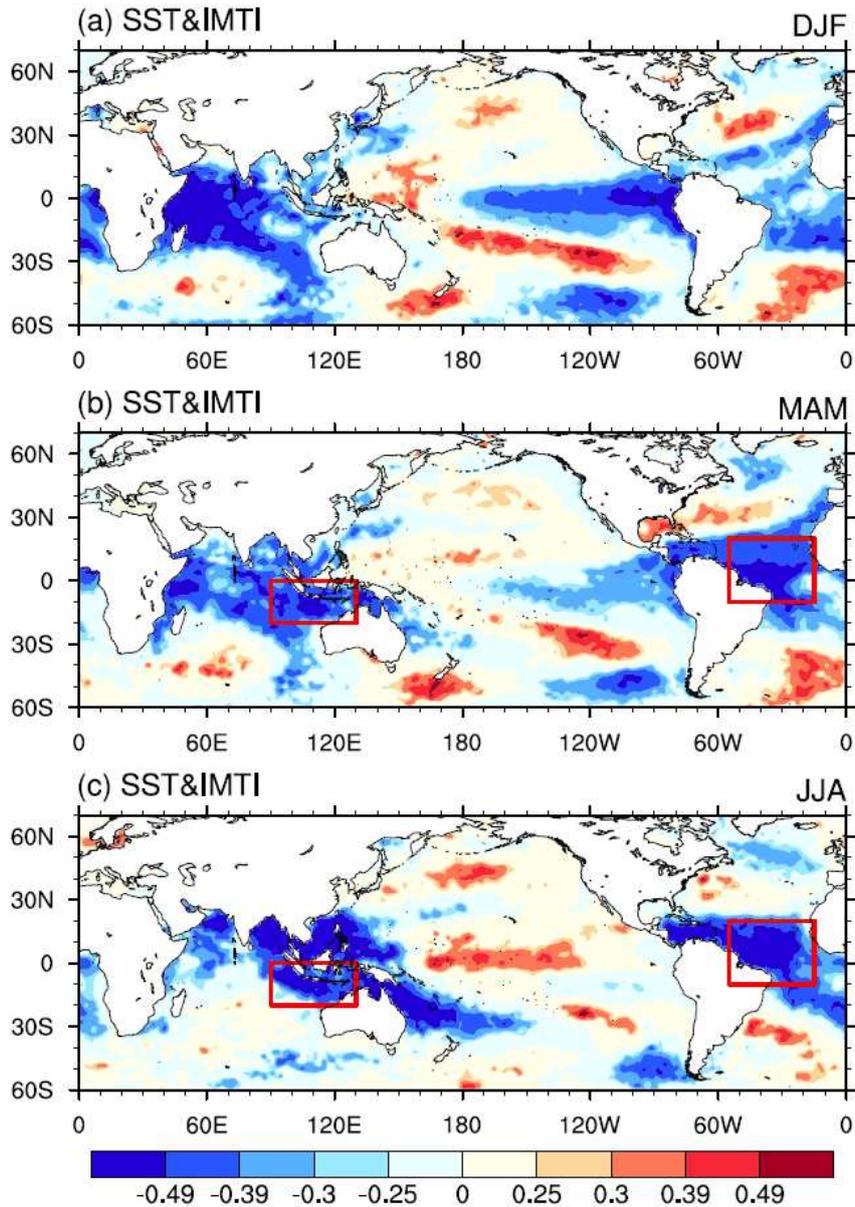
584 **indices.** Regressions of 850-hPa wind (m s^{-1} ; vector) and rainfall (mm day^{-1} ; shading)

585 against (a) IMTI, (b) WYI, (c) AIRI, (d) EIMR, (e) MHI, (f) CI, (g) IMI, (h) SASMI,

586 (i) SASMI1, and (j) SASMI2. Significant values exceeding the 95% confidence level

587 are marked by magenta dots or black vectors.

588



589

590 **Figure 5 | IMTI-predictor correlations.** Correlation of SST in (a) preceding winter,
 591 (b) preceding spring, and (c) simultaneous summer with the IMTI. The absolute values
 592 above 0.25, 0.30, 0.39, and 0.49 indicate that they significantly exceed the 90%, 95%,
 593 99%, and 99.9% confidence levels, respectively. The red boxes outline the regions for
 594 the predictors.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementarySAindex1.pdf](#)