PNAS RESEARCH ARTICLE | EARTH, ATMOSPHERIC, AND PLANETARY SCIENCES OPEN ACCESS

Improved subseasonal prediction of South Asian monsoon rainfall using data-driven forecasts of oscillatory modes

Eviatar Bach[a,b,1], V. Krishnamurthy[c], Safa Mote[d,e,1], Jagadish Shukla[f], A. Surjalal Sharma[g,2], Eugenia Kalnay[e], and Michael Ghil[h,i,j]

Edited by Timothy Palmer, University of Oxford, Oxford, United Kingdom; received July 23, 2023; accepted February 1, 2024

Predicting the temporal and spatial patterns of South Asian monsoon rainfall within a season is of critical importance due to its impact on agriculture, water availability, and flooding. The monsoon intraseasonal oscillation (MISO) is a robust northward-propagating mode that determines the active and break phases of the monsoon and much of the regional distribution of rainfall. However, dynamical atmospheric forecast models predict this mode poorly. Data-driven methods for MISO prediction have shown more skill, but only predict the portion of the rainfall corresponding to MISO rather than the full rainfall signal. Here, we combine state-of-the-art ensemble precipitation forecasts from a high-resolution atmospheric model with data-driven forecasts of MISO. The ensemble members of the detailed atmospheric model are projected onto a lower-dimensional subspace corresponding to the MISO dynamics and are then weighted according to their distance from the data-driven MISO forecast in this subspace. We thereby achieve improvements in rainfall forecasts over India, as well as the broader monsoon region, at 10- to 30-d lead times, an interval that is generally considered to be a predictability gap. The temporal correlation of rainfall forecasts is improved by up to 0.28 in this time range. Our results demonstrate the potential of leveraging the predictability of intraseasonal oscillations to improve extended-range forecasts; more generally, they point toward a future of combining dynamical and data-driven forecasts for Earth system prediction.

South Asian monsoon | ensemble forecasting | data-driven forecasting | subseasonal-to-seasonal prediction

Significance

The South Asian monsoon affects more than a billion people in the Indian subcontinent. The monsoon intraseasonal oscillation (MISO) determines the spatial structure of the monsoon rainfall on subseasonal timescales, and its accurate prediction is therefore key for agricultural and hydrological planning. Here, we combine data-driven forecasts of MISO with an ensemble of dynamical forecasts of the full system, leveraging the predictability of MISO to improve monsoon forecasts. Our results show significant improvement compared to state-of-the-art dynamical model forecasts, demonstrating the potential of data-driven forecasts to improve subseasonal monsoon prediction.

The South Asian monsoon is one of the most important seasonal features of the global climate system. Despite modest improvements in recent years, the forecast skill of monsoon prediction has lagged behind improvements in numerical weather prediction (1–3). Moreover, accurate forecasts of regional rainfall on intraseasonal timescales, rather than just the seasonal mean rainfall, are crucial for the agricultural and hydrological sectors (4).

Monsoon Intraseasonal Variability

It has long been known that the monsoon possesses intraseasonal variability in the form of active and break phases within the summer monsoon season (5–7). The active phase is associated with high rainfall over central India, while the break phase is associated with low rainfall over central India, but high rainfall over northern and southeastern India (6).

These intraseasonal variations are dominant at two spectral peaks, roughly at 45 d and 20 d (8). They consist of northward-propagating rainfall anomalies, referred to as monsoon intraseasonal oscillations (MISOs), closely tied to the boreal summer intraseasonal oscillation (BSISO) (9). MISOs characterize the active and break phases of the monsoon and much of the regional rainfall patterns (8).

Current state-of-the-art dynamical models poorly predict MISOs (2, 10, 11). Improving MISO prediction is recognized as a crucial part of improving monsoon forecasts on intraseasonal timescales (1).

Data-Driven Forecasts

Oscillations in the climate system, due to their near-regularity and low frequency compared to synoptic-scale weather variability, are recognized as an important source of predictability beyond the weather timescale (12). This recognition has led to work in predicting these oscillatory modes using data-driven methods.*

*Here, we use "data-driven" to mean that the forecasts are accomplished entirely by learning from past data. Data-driven forecasts are also referred to in the literature as statistical, empirical, or machine learning forecasts.

Author contributions: E.B., V.K., S.M., J.S., A.S.S., E.K., and M.G. designed research; E.B. performed research; E.B., V.K., S.M., J.S., A.S.S., E.K., and M.G. contributed new analytic tools; E.B., V.K., S.M., J.S., A.S.S., E.K., and M.G. analyzed data; and E.B., V.K., S.M., J.S., A.S.S., E.K., and M.G. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: eviatarbach@protonmail.com or Mote@pdx.edu.

[2]Deceased January 5, 2024.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2312573121/-/DCSupplemental.

Published April 1, 2024.

There is an extensive literature on data-driven forecasting of climate oscillations (13), and several works have developed data-driven (10, 14, 15) or low-order model (9, 16) forecasts for MISO. Some of these methods demonstrate skill in MISO prediction up to 50 d, while dynamical models only demonstrate skill for up to 30 d, and often less (17–19). The superiority of data-driven forecasts may be due to model error in the dynamical forecasts, but can also be associated with initial condition error (20). This gap demonstrates the potential for improved intraseasonal prediction of monsoon rainfall.

Since the intraseasonal oscillations only comprise a fraction of the total variance of the field of interest, their prediction is not directly useful (21–24). For instance, MISOs comprise about 14% of the variance in daily rainfall anomalies over India (8), or about 23% of the variance of 15-d rainfall over the monsoon region (25). Moreover, there is no way to faithfully infer the state of the full field from a forecast of an oscillatory mode (13). This suggests, therefore, the need for a method that combines the information from a full dynamical model with that of the data-driven forecasts.

Prediction Using Dynamical and Data-Driven Forecasts

The idea of combining full-field dynamical forecasts with data-driven forecasts of intraseasonal oscillations was previously suggested by various studies (15, 21, 22, 26, 27). While there have been a number of works that correct dynamical forecasts of an oscillatory mode using either data-driven forecasts (28–30) or post-processing (31–33), these did not attempt to correct the full-field forecasts. In this work, we correct a full-field dynamical forecast using data-driven forecasts of specific modes.

Strong (21) was perhaps the first to propose a method for correcting the physical model forecasts, by adding a term nudging the dynamical equation to the statistical forecast, or by using a Kalman filter.[†] However, to our knowledge, this was never implemented. Recent work (34–36) suggested a multi-model data assimilation–based approach for similar problems.

In Bach et al. (13) we introduced Ensemble Oscillation Correction (EnOC), to beneficially combine data-driven forecasts of oscillations with an ensemble of dynamical forecasts of the full system. EnOC works by projecting a dynamical ensemble into a subspace corresponding to the mode of interest, and weighting the ensemble members by their distance from a data-driven oscillation forecast in that subspace. See Fig. 1 for a basic schematic, ref. 13, and *SI Appendix*, section 1A for more details.

Here, we apply EnOC to the South Asian monsoon by leveraging the predictability of MISO. Our approach possesses two crucial features of machine learning applied to weather and climate problems (37, 38): 1) It is interpretable, in that the forecast improvements can be attributed to a specific physical mode; and 2) it does not introduce spurious unphysical features, since the forecast is always the mean of a subset of the dynamical ensemble. An additional advantage of EnOC is that in the current implementation, it is carried out offline, meaning that it is applied as a post-processing step to model outputs, avoiding the need for access to the computational and data resources of operational forecasting centers.

Results

Data-Driven MISO Forecast Skill. First, we quantify the prediction skill of the data-driven MISO forecasts by comparing these to the MISO mode extracted from observations. We use the bivariate correlation coefficient, defined in *SI Appendix*, section 5A.

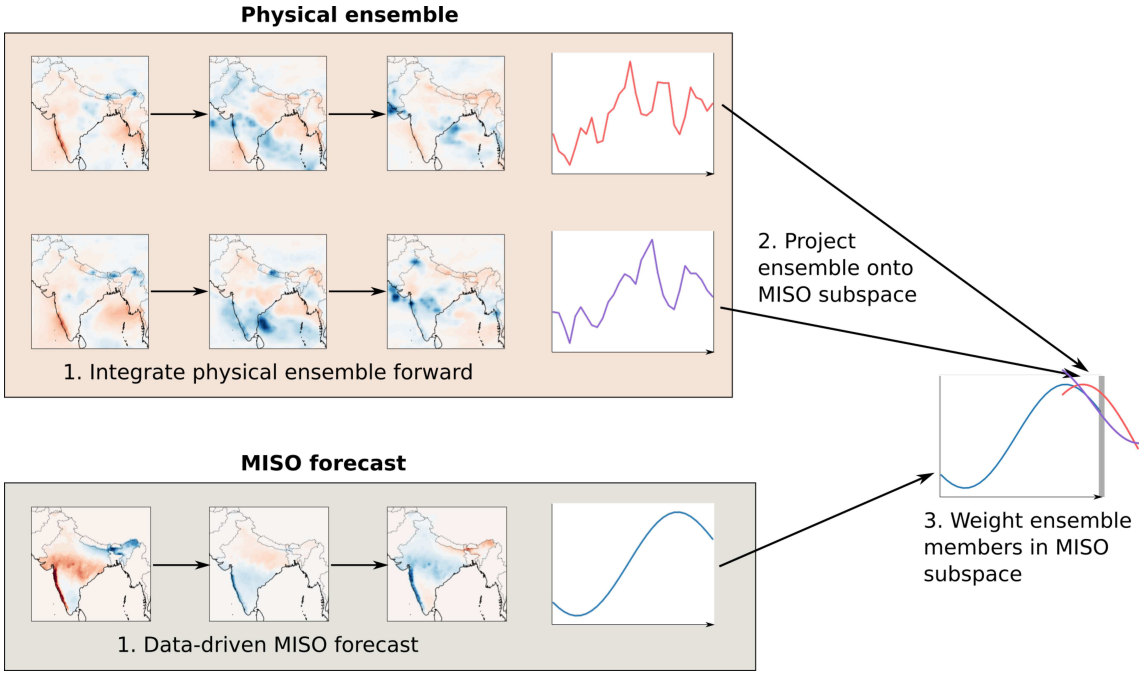

Fig. 1. A simplified diagram of the EnOC algorithm, with two dynamical ensemble members for simplicity. Here, the second (purple) ensemble member will receive a higher weight, since it is closer to the MISO forecast in the subspace. Note that in the real implementation, we reduce the dynamics in the MISO subspace to the first two principal components of the MISO mode. See *SI Appendix*, section 1A for more details.

[†]Ideas attributed to Eugenia Kalnay and Robert Vautard, respectively.

As per Fig. 2, these forecasts maintain a bivariate correlation above 0.5 for over 46 d when initialized in August and September, and for about 28 d when initialized in July. The higher forecast skill in the later months is consistent with (9, 39). The reason for this phenomenon should be investigated in future research.

Some of the previous works on MISO prediction (10, 15) did not test the real-time forecasting context, in which one has to deal with filtering end effects. This filtering is the reason why the bivariate correlation coefficient in Fig. 2 does not begin at 1 at a lead time of 0. These results can be compared to ref. 9 and achieve similar skill.

Improvements in Precipitation Forecasts. We now focus on the improvements in the full-field precipitation forecasts made through application of EnOC. In the following results, we use the 15-d mean rainfall for the skill calculations: That is, at lead time i, we compare the mean rainfall forecasted over leads $i-7, i-6, \ldots, i+7$ to the observations averaged over the same interval. Due to the highly intermittent nature of rainfall, such smoothing is often used in the literature (40, 41). We also include the results with 7-d averaging, which are qualitatively similar, in *SI Appendix*, section 6E.

We show results both over India and the wider monsoon region, which we define as 6–39°N and 66–100°E. The latter includes India, Bangladesh, Bhutan, Kyrgyzstan, Myanmar, Nepal, Sri Lanka, Tajikistan, most of Pakistan, southwestern China, and part of the northern Indian Ocean. This domain is close to the extended Indian monsoon rainfall region (42) widely used in studies on the South Asian monsoon.

We use the temporal correlation to quantify the skill of rainfall forecasts. For a given initialization time and lead time, we compute the correlation of the predicted to observed total rainfall over a given spatial domain, over the verification period of 2008 to 2016.

All-India and Monsoon Region. We consider the temporal correlation of rainfall averaged over India and the wider monsoon region. We achieve a remarkable improvement in this correlation using the data-driven MISO forecast for July and August, as shown in Fig. 3. In July, the improvement in the temporal correlation reaches over 0.3 in the 23 to 28 d range over the monsoon region. Over India, the improvement reaches over 0.15 in the 22 to 30 d range. In August, the improvement over the monsoon region reaches 0.09, and over India reaches over 0.1.

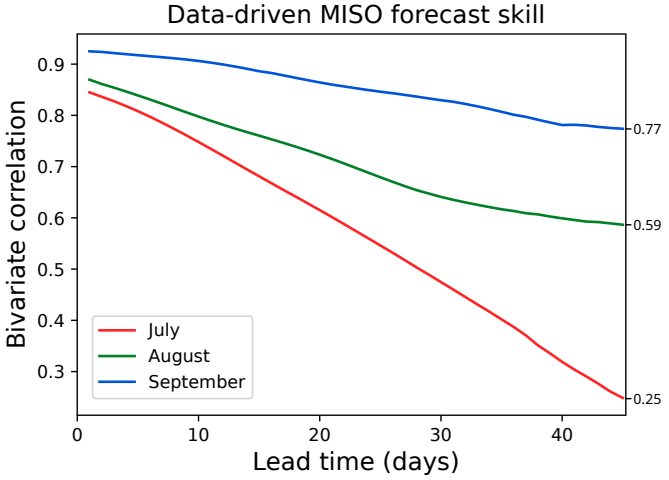

Fig. 2. Correlation coefficient between predicted and observed MISO mode as a function of lead time, for forecasts starting on the 1st of July, August, and September, from 2008 to 2016.

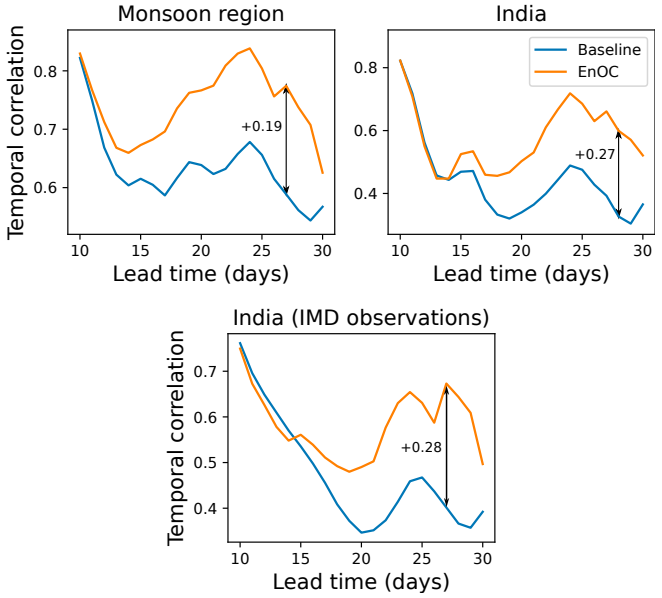

Fig. 3. *Top*: Temporal correlation of monsoon region and India rainfall averages over forecasts initialized in July and August, verified against ERA5 reanalysis. *Bottom*: Same as *Top Right*, but verified against India Meteorological Department (IMD) observations.

Although there is no apparent improvement in the temporal correlation for September, there is improvement in other skill metrics discussed below. The average improvement for July, August, and September forecasts over the monsoon region is statistically significant at the 95% confidence level for the total 10 to 30 d rainfall, using the bootstrap method discussed in *SI Appendix*, section 7.

Regional results. We now look at the temporal correlation of rainfall in the homogeneous rainfall regions of India; see Fig. 4. *SI Appendix*, section 3 defines the regions.

For 14- to 20-d lead time forecasts, there are statistically significant skill improvements (at the 85% confidence level) in all regions except for Hilly Regions, with the largest improvements in the West Central (+0.22), Northwest (+0.12), and Northeast (+0.12) regions. For 21- to 27-d lead time forecasts, the skill improvements are statistically significant in the Northwest (+0.19), Central Northeast (+0.16), and South Peninsular (+0.10) regions.

Other skill metrics. In *SI Appendix*, section 6, we evaluate the forecasts with respect to three other skill metrics: the anomaly correlation, RMSE skill score, and the skill in predicting the MISO index defined in ref. 25. The EnOC forecasts are also improved with respect to these skill metrics.

Consistency of improvements in precipitation forecast skill. To verify that the improvements in precipitation prediction are indeed due to improved prediction of MISO, we compute the correlation between 1) the difference between the errors in the dynamical and data-driven predictions of the MISO mode and 2) the difference in the RMSE of uncorrected and corrected precipitation forecasts. If the improvements in the EnOC-corrected forecasts of precipitation were due to the skill of the data-driven MISO forecast, we would expect a positive correlation. Indeed, over forecasts initialized in July, August, and September, and lead times from 8 to 39 d, the correlation is 0.37. For the individual months, we have July (0.29), August (0.52),

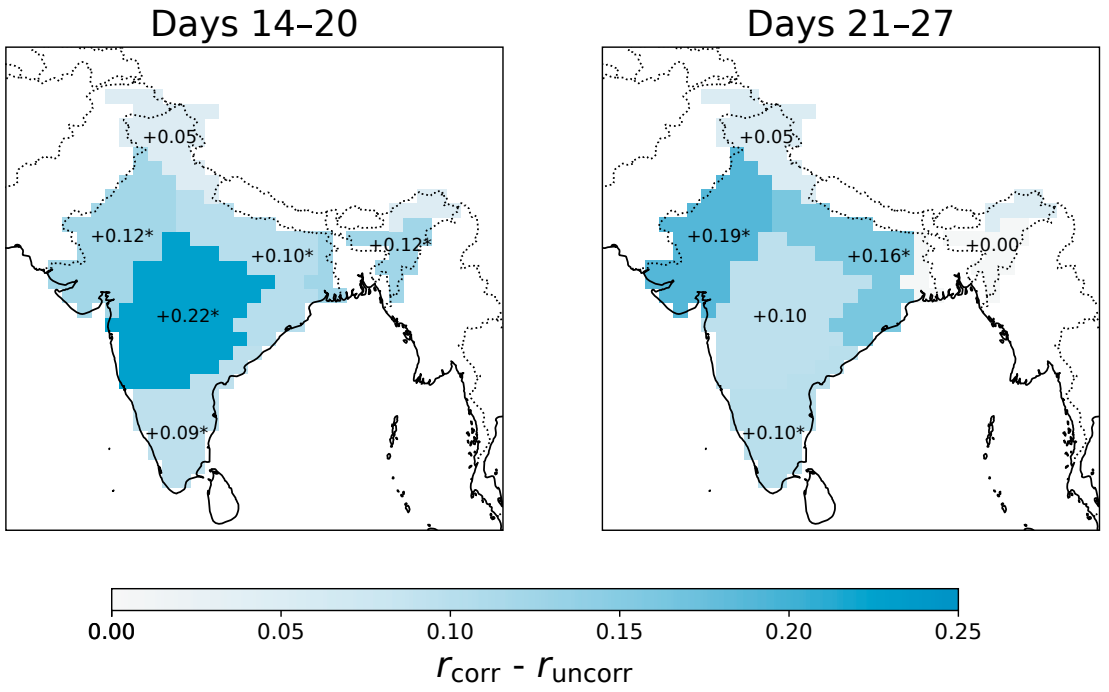

Fig. 4. The difference between the corrected and uncorrected temporal correlation of forecasted rainfall by region. The asterisks indicate statistical significance at the 85% confidence level using a bootstrap methodology described in *SI Appendix*, section 7.

and September (0.45). For all these correlations, the two-sided P-value is < 0.001, with a null hypothesis of no correlation.

As another verification of the consistency of the forecast skill improvement, we examine the forecasts of low-level (850 hPa) relative vorticity. MISO modulates relative vorticity over the monsoon region: in active phases, the vorticity is increased, which in turn increases the genesis of strong low-pressure systems and precipitation (43–45). Due to this physical link, one would expect the EnOC-corrected ensemble to also have better prediction of relative vorticity, despite the fact that the data-driven forecast was based only on precipitation. *SI Appendix*, Section 6F shows that this is indeed the case: the EnOC ensembles have improved temporal correlation over India as well as improved anomaly correlation over the monsoon region.

Conclusion

The MISO is predictable beyond the predictability limit of chaotic weather variability, due to its regularity and low frequency (10). Harnessing this predictability is an important part of improving monsoon prediction on intraseasonal timescales. We show that data-driven MISO forecasts can significantly improve state-of-the-art dynamical precipitation forecasts in the South Asian monsoon region. The improvements in the temporal correlation reach 0.15 to over 0.25 on 10- to 30-d lead times.

This methodology could also be applied to other intraseasonal oscillations, such as the Madden–Julian oscillation, for which data-driven forecasts are quite skillful (46–48). More generally, this work demonstrates the power of combining dynamical and data-driven models for Earth system prediction. Other recent work on such hybrid forecasting has also been promising (49, 50), as has the growing field of purely data-driven full-field forecasts on subseasonal-to-seasonal timescales (50–52). As machine learning forecasts of weather and climate continue to improve, we envision the integration of dynamical and data-driven forecasts for both real-time prediction and data assimilation. Recent work on multi-model ensemble Kalman filters (35) provides a versatile method for this application.

Materials and Methods

Data. We use the IMD gridded rainfall dataset (53) for extraction and prediction of MISO. It is based on interpolation of rain gauge data and is provided at $0.25° \times 0.25°$ resolution with daily coverage from 1901 to 2016.

Unless otherwise stated, we use the ERA5 reanalysis as verification for forecasts. ERA5, and its predecessor ERA-Interim, is generally the most accurate in precipitation among reanalyses produced by major operational centers and is accurate in the South Asian monsoon region in particular (54–56). We also verify against the IMD rainfall observations in Fig. 3.

Extraction of MISO. We extract MISO from the IMD rainfall data using multichannel singular spectrum analysis (M-SSA: 57, 58). M-SSA applies principal component analysis to multivariate delay-embedded time-series data, in order to identify and extract spatiotemporal modes. More information is provided in *SI Appendix*, section 1B.

M-SSA has previously been used to extract MISOs in a number of studies, from precipitation, outgoing longwave radiation, and wind fields (8, 10, 44, 59–63). Krishnamurthy and Shukla (8) demonstrated the existence of two MISO modes, with periods of about 45 d and 20 d, extracted from precipitation data. Here, we focus on the 45-d northeastward-propagating mode, which is most prominent, and shown to be statistically distinguishable from red noise in ref. 10. We closely follow refs. 8 and 10 in the extraction of the mode, and *SI Appendix*, section 4 provides some of its characteristics.

Following ref. 10, we furthermore take the two leading spatial empirical orthogonal functions (EOFs) of the MISO reconstructed component, and henceforth consider only the dynamics of the corresponding two-dimensional principal component (PC) time-series.

Projection onto the MISO subspace and data-driven forecasts. The process of reconstructing a mode using M-SSA involves forward and reverse filtering in time, posing a problem for real-time forecasting of the mode (9, 13, 64). In this real-time forecasting context, we must approximate the

projection onto the MISO subspace, necessarily incurring an error in the initial conditions used for the data-driven forecast.

Here, we use a simple neural network architecture (described in *SI Appendix*, section 1C) in order to project from the full phase space onto the two-dimensional MISO space. To obtain the initial condition for the data-driven MISO forecasts (step 1 in Fig. 1), we use the current day and past 28 d of IMD rainfall data as inputs to the network. To project the dynamical ensemble members onto the MISO space (step 3), we use the current day, past 14 d, and 14 future days of forecasts for each ensemble member. For lead times less than 14 d, we use the current day and 28 future days.

We follow refs. 10 and 65 in the data-driven prediction of the MISO mode. We look for the closest historical analogs in the two-dimensional MISO space and average over their trajectories at a given lead time. See *SI Appendix*, section 1D for more details.

Note that EnOC does not depend on a specific data-driven forecasting method. Future work could use other methods for approximating the projection (13, 64), or use a method designed for real-time forecasting, where an approximate projection does not need to be used.

Dynamical Forecast System. We apply EnOC to hindcasts from SEAS5, the European Centre for Medium-Range Weather Forecasts (ECMWF) seasonal forecasting system (66). The atmospheric component is the Integrated Forecasting System (IFS) model, with a configuration similar to ECMWF's operational medium- and extended-range forecasts, and NEMO is used for the ocean. We use hindcasts from 1993 to 2016, initialized on the first of July, August, and September, which are provided with 25 ensemble members. Note that the first 2 wk of June are an interval of rapid change during the monsoon onset, which provides a plausible explanation for the fact that data-driven forecasts do not exhibit satisfactory skill when initialized on June 1st. We have not included them, therefore, in the forecasts reported herein. The ECMWF forecasts have been shown to outperform those of other operational centers in predicting Asian monsoon intraseasonal variability (17, 18, 67), making them a state-of-the-art baseline.

We perform a simple bias correction of SEAS5 by subtracting its own climatology dependent on lead and initialization time from 1993 to 2007. This practice is common in subseasonal-to-seasonal forecasting (68). The years for bias correction did not overlap with the verification period of 2008 to 2016, avoiding the problem of artificial skill (69). The years 1993 to 2007 are also used for tuning the m' parameter of EnOC for each initialization time and lead; see *SI Appendix*, section 1A.

Data, Materials, and Software Availability. The raw and processed forecasts, as well as the code for the method have been deposited in Zenodo (70).

ACKNOWLEDGMENTS. The work of E.B., S.M., and E.K. on this paper was funded through the Monsoon Mission II (Grant IITMMMIIUNIVMARYLANDUSA2018INT1) provided by the Ministry of Earth Science, Government of India. Additionally, the work of S.M. and E.K. on this paper was funded by the NASA grant IRET-QRS-22-0001. S.M. and E.K. acknowledge the support from Eugenia and Michael Brin. S.M. also acknowledges the NSF RTG grant DMS-2136228. We thank two anonymous referees for helpful comments. We thank A.S.S., who passed away on January 5, 2024, for his humility, kindness, and lasting contributions to science.

Author affiliations: [a]Department of Environmental Science and Engineering, California Institute of Technology, Pasadena, CA 91125; [b]Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA 91125; [c]Center for Ocean-Land-Atmosphere Studies, George Mason University, Fairfax, VA 22030; [d]Fariborz Maseeh Department of Mathematics and Statistics, Portland State University, Portland, OR 97201; [e]Department of Atmospheric and Oceanic Sciences and Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742; [f]Department of Atmospheric, Oceanic and Earth Sciences, George Mason University, Fairfax, VA 22030; [g]Department of Astronomy, University of Maryland, College Park, MD 20742; [h]Geosciences Department and Laboratoire de Météorologie Dynamique (CNRS and Institut Pierre-Simon Laplace), École Normale Supérieure and Paris Sciences et Lettres University, Paris, France 75005; [i]Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, CA 90095; and [j]Department of Mathematics, Imperial College London, London SW7 2BX, United Kingdom

1. J. Shukla, Monsoon mysteries. *Science* **318**, 204–205 (2007).
2. V. Krishnamurthy, Predictability of CFSv2 in the tropical Indo-Pacific region, at daily and subseasonal time scales. *Clim. Dyn.* **50**, 3931–3948 (2018).
3. S. A. Rao *et al*., Monsoon Mission: A targeted activity to improve monsoon prediction across scales. *Bull. Am. Meteorol. Soc.* **100**, 2509–2532 (2019).
4. P. K. Xavier, B. N. Goswami, An analog method for real-time forecasting of summer monsoon subseasonal variability. *Mon. Weather Rev.* **135**, 4149–4160 (2007).
5. S. Gadgil, The Indian monsoon and its variability. *Annu. Rev. Earth Planet. Sci.* **31**, 429–467 (2003).
6. V. Krishnamurthy, J. L. Kinter III, "The Indian monsoon and its relation to global climate variability" in *Global Climate: Current Research and Uncertainties in the Climate System*, X. Rodó, F. A. Comín, Eds. (Springer, Berlin/Heidelberg, Germany, 2003), pp. 186–236.
7. B. N. Goswami, "South Asian monsoon" in *Intraseasonal Variability in the Atmosphere-Ocean Climate System*, W. K. M. Lau, D. E. Waliser, Eds. (Springer-Verlag, Berlin/Heidelberg, Germany, 2nd ed., 2012), pp. 21–72.
8. V. Krishnamurthy, J. Shukla, Intraseasonal and seasonally persisting patterns of Indian monsoon rainfall. *J. Clim.* **20**, 3–20 (2007).
9. N. Chen, A. J. Majda, C. T. Sabeerali, R. S. Ajayamohan, Predicting monsoon intraseasonal precipitation using a low-order nonlinear stochastic model. *J. Clim.* **31**, 4403–4427 (2018).
10. V. Krishnamurthy, A. S. Sharma, Predictability at intraseasonal time scale. *Geophys. Res. Lett.* **44**, 8530–8537 (2017).
11. V. Krishnamurthy, Tropical intraseasonal oscillations in CFSv2 during Boreal summer and winter. *Int. J. Climatol.* **37**, 3674–3693 (2017).
12. M. Ghil, A. W. Robertson, "Waves" vs. "particles" in the atmosphere's phase space: A pathway to long-range forecasting? *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2493–2500 (2002).
13. E. Bach *et al*., Ensemble Oscillation Correction (EnOC): Leveraging oscillatory modes to improve forecasts of chaotic systems. *J. Clim.* **34**, 5673–5686 (2021).
14. R. Chattopadhyay, A. K. Sahai, B. N. Goswami, Objective identification of nonlinear convectively coupled phases of monsoon intraseasonal oscillation: Implications for prediction. *J. Atmos. Sci.* **65**, 1549–1569 (2008).
15. R. Alexander, Z. Zhao, E. Székely, D. Giannakis, Kernel analog forecasting of tropical intraseasonal oscillations. *J. Atmos. Sci.* **74**, 1321–1342 (2017).
16. N. Chen, A. J. Majda, Predicting the cloud patterns for the boreal summer intraseasonal oscillation through a low-order stochastic model. *Math. Clim. Weather Forecast.* **1**, 1–20 (2015).
17. H. J. Kim, Y. R. Jung, H. I. Jeong, "Skill assessment of real-time BSISO forecasts: APCC's ceaseless efforts to improve subseasonal monsoon prediction" in *Workshop on Sub-seasonal to Seasonal Predictability of Monsoons* (Sub-seasonal to Seasonal Prediction Project, National Institute of Meteorological Research, Jeju, South Korea, 2015).
18. W. Jie, F. Vitart, T. Wu, X. Liu, Simulations of the Asian summer monsoon in the sub-seasonal to seasonal prediction project (S2S) database. *Q. J. R. Meteorol. Soc.* **143**, 2282–2295 (2017).
19. A. K. Sahai *et al*., Multi-model multi-physics ensemble: A futuristic way to extended range prediction system. *Front. Clim.* **3**, 655919 (2021).
20. R. Vautard, C. Pires, G. Plaut, J. Sarda, "Statistical, dynamical and hybrid long-range forecast approaches using predictable components" in *Proceedings of a Seminar Held at ECMWF on Predictability* (Shinfield Park, Reading, UK, 1996), vol. 1, pp. 275–304.
21. C. M. Strong, "Intraseasonal Variability and Extended-Range Prediction in the Northern Hemisphere Extratropics," Ph.D. thesis, University of California, Los Angeles, Los Angeles, CA (1994).
22. C. Strong, F. F. Jin, M. Ghil, Intraseasonal oscillations in a barotropic model with annual cycle, and their predictability. *J. Atmos. Sci.* **52**, 2627–2642 (1995).
23. K. C. Mo, Adaptive filtering and prediction of intraseasonal oscillations. *Mon. Weather Rev.* **129**, 802–817 (2001).
24. S. Abhilash *et al*., Prediction and monitoring of monsoon intraseasonal oscillations over Indian monsoon region in an ensemble prediction system using CFSv2. *Clim. Dyn.* **42**, 2801–2815 (2014).
25. E. Suhas, J. M. Neena, B. N. Goswami, An Indian monsoon intraseasonal oscillations (MISO) index for real time monitoring and forecast verification. *Clim. Dyn.* **40**, 2605–2616 (2013).
26. R. Vautard, P. Yiou, M. Ghil, Singular-spectrum analysis: A toolkit for short, noisy chaotic signals. *Physica D* **58**, 95–126 (1992).
27. M. Ghil, D. Kondrashov, F. Lott, A. W. Robertson, "Intraseasonal oscillations in the mid-latitudes: Observations, theory and GCM results" in *ECMWF/CLIVAR Workshop on Simulation and Prediction of Intra-Seasonal Variability with Emphasis on the MJO* (Shinfield Park, Reading, UK, 2004), pp. 35–54.
28. J. Sarda, G. Plaut, C. Pires, R. Vautard, Statistical and dynamical long-range atmospheric forecasts: Experimental comparison and hybridization. *Tellus A* **48**, 518–537 (1996).
29. I. S. Kang, H. M. Kim, Assessment of MJO predictability for boreal winter with various statistical and dynamical models. *J. Clim.* **23**, 2368–2378 (2010).
30. M. Kim, C. Yoo, J. Choi, Enhancing subseasonal temperature prediction by bridging a statistical model with dynamical Arctic Oscillation forecasting. *Geophys. Res. Lett.* **48**, e2021GL093447 (2021).
31. J. Wu, F. F. Jin, Improving the MJO forecast of S2S operation models by correcting their biases in linear dynamics. *Geophys. Res. Lett.* **48**, e2020GL091930 (2021).
32. H. Kim, Y. G. Ham, Y. S. Joo, S. W. Son, Deep learning for bias correction of MJO prediction. *Nat. Commun.* **12**, 3087 (2021).
33. R. Silini *et al*., Improving the prediction of the Madden–Julian Oscillation of the ECMWF model by post-processing. *Earth Syst. Dyn.* **13**, 1157–1165 (2022).
34. Y. Chen, S. N. Stechmann, Multi-model communication and data assimilation for mitigating model error and improving forecasts. *Chin. Ann. Math. Ser. B* **40**, 689–720 (2019).
35. E. Bach, M. Ghil, A multi-model ensemble Kalman filter for data assimilation and forecasting. *J. Adv. Model. Earth Syst.* **15**, e2022MS003123 (2023).
36. J. L. Torchinsky, S. Stechmann, Mitigating model error via a multimodel method and application to tropical intraseasonal oscillations. *SIAM J. Appl. Dyn. Syst.* **22**, 3025–3058 (2023).

37. K. Kashinath *et al*., Physics-informed machine learning: Case studies for weather and climate modelling. *Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci.* **379**, 20200093 (2021).
38. C. Irrgang *et al*., Towards neural Earth system modelling by integrating artificial intelligence in Earth system science. *Nat. Mach. Intell.* **3**, 667–674 (2021).
39. S. S. Lee, B. Wang, D. E. Waliser, J. M. Neena, J. Y. Lee, Predictability and prediction skill of the boreal summer intraseasonal oscillation in the Intraseasonal Variability Hindcast Experiment. *Clim. Dyn.* **45**, 2123–2135 (2015).
40. H. Zhu, M. C. Wheeler, A. H. Sobel, D. Hudson, Seamless precipitation prediction skill in the tropics and extratropics from a global model. *Mon. Weather Rev.* **142**, 1556–1569 (2014).
41. V. Moron, A. W. Robertson, L. Wang, "Weather within climate: Sub-seasonal predictability of tropical daily rainfall characteristics" in *Sub-Seasonal to Seasonal Prediction*, A. W. Robertson, F. Vitart, Eds. (Elsevier, 2019), pp. 47–64.
42. B. N. Goswami, V. Krishnamurthy, H. Annmalai, A broad-scale circulation index for the interannual variability of the Indian summer monsoon. *Q. J. R. Meteorol. Soc.* **125**, 611–633 (1999).
43. B. N. Goswami, R. S. Ajayamohan, P. K. Xavier, D. Sengupta, Clustering of synoptic activity by Indian summer monsoon intraseasonal oscillations. *Geophys. Res. Lett.* **30** (2003).
44. N. Karmakar, W. R. Boos, V. Misra, Influence of intraseasonal variability on the development of monsoon depressions. *Geophys. Res. Lett.* **48**, e2020GL090425 (2021).
45. K. M. R. Hunt, A. G. Turner, Non-linear intensification of monsoon low-pressure systems by the BSISO. *Weather Clim. Dyn.* **3**, 1341–1358 (2022).
46. D. Kondrashov, M. D. Chekroun, A. W. Robertson, M. Ghil, Low-order stochastic model and "past-noise forecasting" of the Madden–Julian Oscillation. *Geophys. Res. Lett.* **40**, 5305–5310 (2013).
47. R. Silini, M. Barreiro, C. Masoller, Machine learning prediction of the Madden–Julian oscillation. *npj Clim. Atmos. Sci.* **4**, 1–7 (2021).
48. A. Delaunay, H. M. Christensen, Interpretable deep learning for probabilistic MJO prediction. *Geophys. Res. Lett.* **49**, e2022GL098566 (2022).
49. T. Arcomano *et al*., A hybrid approach to atmospheric modeling that combines machine learning with a physics-based numerical model. *J. Adv. Mod. Earth Syst.* **14**, e2021MS002712 (2022).
50. S. He *et al*., Learning and dynamical models for sub-seasonal climate forecasting: Comparison and collaboration. *Proc. AAAI Conf. Artif. Intell.* **36**, 4495–4503 (2022).
51. J. A. Weyn, D. R. Durran, R. Caruana, N. Cresswell-Clay, Sub-seasonal forecasting with a large ensemble of deep-learning weather prediction models. *J. Adv. Model. Earth Syst.* **13**, e2021MS002502 (2021).
52. J. Xie, P. C. Hsu, Y. Hu, M. Ye, J. Yu, Skillful extended-range forecast of rainfall and extreme events in east China based on deep learning. *Weather Forecast.* **38**, 467–486 (2023).
53. D. S. Pai *et al*., Development of a new high spatial resolution (0.25° × 0.25°) long period (1901–2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. *MAUSAM* **65**, 1–18 (2014).
54. P. Kishore *et al*., Precipitation climatology over India: Validation with observations and reanalysis datasets and spatial trends. *Clim. Dyn.* **46**, 541–556 (2016).
55. H. Hersbach *et al*., The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).
56. B. Hassler, A. Lauer, Comparison of reanalysis and observational precipitation datasets including ERA5 and WFDE5. *Atmosphere* **12**, 1462 (2021).
57. M. Ghil *et al*., Advanced spectral methods for climatic time series. *Rev. Geophys.* **40**, 3-1-3-41 (2002).
58. G. Plaut, R. Vautard, Spells of low-frequency oscillations and weather regimes in the Northern Hemisphere. *J. Atmos. Sci.* **51**, 210–236 (1994).
59. V. Krishnamurthy, J. Shukla, Seasonal persistence and propagation of intraseasonal patterns over the Indian monsoon region. *Clim. Dyn.* **30**, 353–369 (2008).
60. V. Moron, A. W. Robertson, M. Ghil, Impact of the modulated annual cycle and intraseasonal oscillation on daily-to-interannual rainfall variability across monsoonal India. *Clim. Dyn.* **38**, 2409–2435 (2012).
61. V. Krishnamurthy, D. Achuthavarier, Intraseasonal oscillations of the monsoon circulation over South Asia. *Clim. Dyn.* **38**, 2335–2353 (2012).
62. A. Hazra, V. Krishnamurthy, Space–time structure of diabatic heating in monsoon intraseasonal oscillation. *J. Clim.* **28**, 2234–2255 (2015).
63. N. Karmakar, A. Chakraborty, R. S. Nanjundiah, Decreasing intensity of monsoon low-frequency intraseasonal variability over India. *Environ. Res. Lett.* **10**, 054018 (2015).
64. H. R. Ogrosky, S. N. Stechmann, N. Chen, A. J. Majda, Singular spectrum analysis with conditional predictions for real-time state estimation and forecasting. *Geophys. Res. Lett.* **46**, 1851–1860 (2019).
65. J. Chen, A. S. Sharma, Modeling and prediction of the magnetospheric dynamics during intense geospace storms. *J. Geophys. Res.: Space Phys.* **111** (2006).
66. S. J. Johnson *et al*., SEAS5: The new ECMWF seasonal forecast system. *Geosci. Model Dev.* **12**, 1087–1117 (2019).
67. N. Vigaud, A. W. Robertson, M. K. Tippett, N. Acharya, Subseasonal predictability of boreal summer monsoon rainfall from ensemble forecasts. *Front. Environ. Sci.* **5** (2017).
68. K. Pegion *et al*., The Subseasonal Experiment (SubX): A multimodel subseasonal prediction experiment. *Bull. Am. Meteorol. Soc.* **100**, 2043–2060 (2019).
69. J. S. Risbey *et al*., Standard assessments of climate forecast skill can be misleading. *Nat. Commun.* **12**, 4346 (2021).
70. E. Bach, enoc_monsoon. Zenodo. https://doi.org/10.5281/zenodo.10795900. Deposited 7 March 2024.