Hybrid assimilation on a parameter-calibrated model to improve the prediction of heavy rainfall events during the Indian summer monsoon

Sandeep Chinta[1,*], V. S. Prasad[2] and C. Balaji[1,3,4]

[1]Department of Mechanical Engineering, Indian Institute of Technology Madras, Chennai 600 036, India
[2]National Centre for Medium Range Weather Forecasting, A-50, Sector 62, Noida 201 309, India
[3]Centre of Excellence in Atmospheric and Climate Sciences, Indian Institute of Technology Madras, Chennai 600 036, India
[4]Divecha Centre for Climate Change, Indian Institute of Science, Bengaluru 560 012, India

Heavy rainfall events during the Indian summer monsoon cause landslides and flash floods resulting in a significant loss of life and property every year. The exactness of the model physics representation and initial conditions is critical for accurately predicting these events using a numerical weather model. The values of parameters in the physics schemes influence the accuracy of model prediction; hence, these parameters are calibrated with respect to observation data. The present study examines the influence of hybrid data assimilation on a parameter-calibrated WRF model. Twelve events during the period 2018–2020 were simulated in this study. Hybrid assimilation on the WRF model significantly reduced the model prediction error of the variables: rainfall (18.04%), surface air temperature (7.91%), surface air pressure (5.90%) and wind speed at 10 m (27.65%) compared to simulations with default parameters without assimilation.

Keywords: Heavy rainfall events, hybrid assimilation, numerical weather model, parameter calibration, summer monsoon.

THE Indian summer monsoon (ISM) is among the oldest global monsoon phenomena occurring with striking regularity every year. Heavy rainfall events during ISM cause landslides and flash floods resulting in a significant loss of life and property each year[1]. The number of low and moderate rainfall events averaged over the entire Indian region has substantially reduced during ISM, whereas heavy rainfall events have increased over the years[2]. Also, there has been a noticeable increase in both the average frequency of heavy rainfall events and the percentage of seasonal rainfall contributed by these events[3]. So, accurately simulating the heavy rainfall events during the ISM is crucial.

The accuracy of a numerical weather prediction (NWP) model depends upon both its ability to accurately represent the physics of the atmosphere and the precision of the initial conditions provided to the model[4]. In the NWP model, sub-grid-scale processes are parameterized based on reasonable physical or statistical representations. The parameterization schemes require information from the forecast variables about the process to be parameterized using a set of assumptions[5]. Multiple studies have been conducted to identify the optimal set of parameterization schemes for various regions and types of simulation events[6–12].

However, each parameterization scheme contains multiple parameters on which the scheme is formulated. Typically, the default values of these parameters are determined through theoretical or experimental studies by the developers of the scheme[13]. By calibrating the values of these parameters to observations, the accuracy of the prediction can be improved by increasing the ability of the model to accurately represent the physics of the atmosphere. Parameter calibration based on tuning to an objective can be classified into two categories[14]. The first category involves optimizing an objective function that evaluates the difference between the model simulation and a corresponding set of observations. To accelerate the optimization process, a model emulator or surrogate model of the actual physical model is constructed. Several studies have used this approach[15,16]. The second category quantifies critical uncertainty sources in the problem by employing a Bayesian approach. Some methods in this category use the actual model, while others use a statistical emulator to eliminate regions of the parameter space that are not physically possible and yield not-ruled-out-yet parameter space. Although the second category is robust in identifying the viable parameter space, it requires huge computational power compared to the first category. Many studies have used this approach for climate models[17,18]. As trial-and-error methods typically concentrate on tuning a limited number of parameters (usually one or two) at a time, the above two categories are more beneficial for calibrating a considerable number of parameters. Overtuning is a crucial factor to consider during parameter calibration. It refers to calibrating the values of the

*For correspondence. (e-mail: chprsandeep@gmail.com)

parameters to specific metrics resulting in the model performing well for those metrics, while the performance deteriorates for other processes or metrics. Therefore, conducting validation experiments for different metrics and simulating events not used for calibration can help check whether the parameter calibration resulted in overtuning.

It is computationally expensive to calibrate all the parameters. So, a sensitivity study is necessary to determine the parameters that profoundly impact model prediction[19]. Several studies have used various sensitivity analysis techniques to identify the sensitive parameters for different regions and types of events simulated[20–23]. Studies have also calibrated the sensitive parameters with respect to observations using advanced optimization techniques to improve model prediction[24–27]. The model parameters have been optimized to obtain a better forecast for various models such as watershed model[28], atmospheric general circulation model[29], and climate model[30,31] to reduce the prediction error.

Apart from the calibration of parameters, accuracy of the model prediction also depends on the exactness of the initial conditions. Data assimilation is used to improve the initial conditions utilizing observational data. Data assimilation algorithms require the background state of the atmosphere obtained from the short-range forecast of the previous cycle[32]. The background state, as it is a forecast, contains some uncertainty. Different assimilation algorithms handle this uncertainty differently in the form of an error covariance matrix. The assimilation algorithm corrects the background state using data from the observations and considering the error covariances of the background state and the observations. The background error covariance (BEC) matrix plays a critical role in data assimilation, particularly for weather systems such as heavy rainfall events that are intermittent and transient[33,34].

Three-dimensional variational (3DVar) assimilation method utilizes a static BEC and does not contain flow-dependent spatial covariance[35,36], which means that the errors pertaining to the flow of the day are assumed to be invariant. Ensemble Kalman filter data assimilation generates the ensembles using the Monte Carlo method[37], and flow-dependent BEC is obtained from the ensemble of forecasts. However, sampling error in the ensemble BEC is significant because of the relatively small sample size[38]. In hybrid assimilation, the weighted average of the static and flow-dependent BEC is used within a 3DVar framework[39,40]. Several studies have broadly concluded that hybrid data assimilation performs better than the variational assimilation techniques and is sometimes equivalent to or better than pure ensemble assimilation methods[40–45].

This study builds up on two previous studies. In the first study, sensitivity analysis was performed and the parameters that significantly influenced heavy rainfall events prediction were identified using the Morris one-at-a-time method[46,47]. In the second study, these sensitive parameters were calibrated with respect to observations using a multi-objective adaptive surrogate model-based optimization method[48,49]. These two previous studies only addressed one aspect of improving the NWP model prediction: increasing the accuracy with which the model represents the physics of the atmosphere. However, these studies did not consider the second aspect of improving the model prediction: increasing the accuracy of the initial conditions provided to the model. The present study addresses both these aspects by implementing hybrid assimilation on a parameter-calibrated model to improve prediction.

Data assimilation methodology

Two data assimilation algorithms, namely three-dimensional variational (3DVar) assimilation and three-dimensional ensemble variational (3DEnVar) hybrid assimilation have been used in this study. These algorithms are implemented using the Weather Research and Forecasting (WRF) model data assimilation system (WRFDA). In the 3D-Var algorithm, the analysis field X_a is estimated by minimizing a cost function (that calculates the distance from the background X_b and from the observations Y_o as formulated as

$$J(X) = (X - X_b)^T B^{-1} (X - X_b) + [Y_o - H(X)]^T R^{-1} [Y_o - H(X)], \quad (1)$$

where $J(X)$ is the cost function that needs to be minimized, B the BEC matrix estimated as the error (averaged over many forecasts) between two short-range forecasts valid at a particular time, H the forward operator that maps variables from the state space to the observation space and R is the observation error covariance matrix that contains the instrument error. In the National Meteorological Center method[35], BEC is formulated as

$$B_{3\mathrm{DVar}} \approx \alpha E[X_f(24\ \mathrm{h}) - X_f(12\ \mathrm{h})] \times [X_f(24\ \mathrm{h}) - X_f(12\ \mathrm{h})]^T, \quad (2)$$

where $X_f(24\ \mathrm{h})$ is the 24th hour and $X_f(12\ \mathrm{h})$ is the 12th hour regional forecast valid at a particular time. Different control variable (cv) transform options are available in WRFDA to evaluate $B_{3\mathrm{DVar}}$, where each cv transform focuses on different control variables. Some studies have observed that cv6 option performed better compared to other options, such as cv3, cv5 and cv7 in WRFDA[50,51]. The control variables used in cv6 are stream function (ψ), unbalanced temperature (T_u), unbalanced velocity potential (χ_u), unbalanced surface pressure ($P_{s.u}$) and unbalanced pseudo relative humidity ($RH_{s.u}$). As the assimilation is performed on a parameter-calibrated model, the value of $B_{3\mathrm{DVar}}$ is evaluated with the cv6 option using simulations from a parameter-calibrated model. Simulations were done for July–August

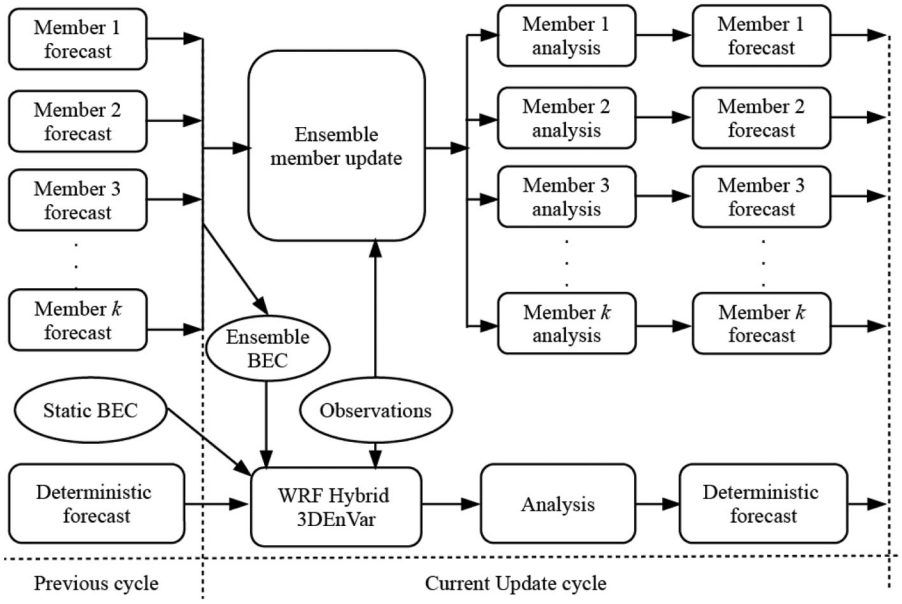

Figure 1. Flowchart of the hybrid assimilation methodology implemented in this study.

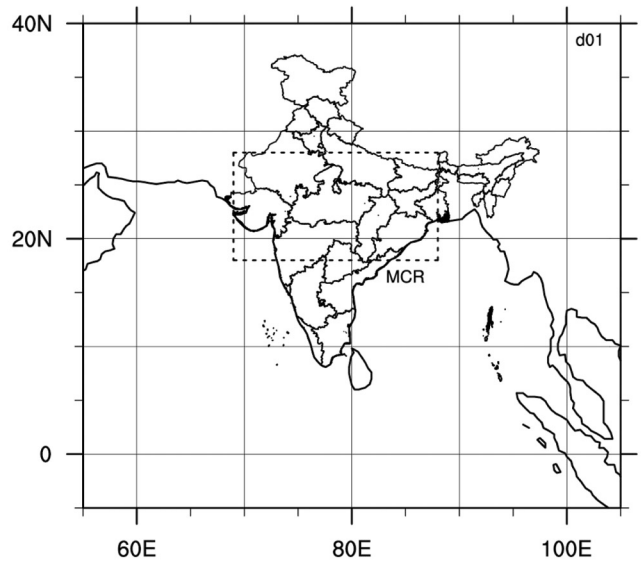

Figure 2. Configuration of the model domain used in this study. Monsoon core region (MCR) is also marked.

2016, and the 24th hour and 12th hour forecasts were evaluated at 12-h intervals for these two months. These forecasts were utilized to calculate B_{3DVar} using eq. (2).

In hybrid assimilation, BEC was estimated using a weighted average of the variational and ensemble BEC, as already mentioned. The background error covariance B_{hyb} was evaluated using eq. (3).

$$B_{hyb} = (1-\beta)B_{Ens} + \beta B_{3DVar}, \quad (3)$$

where β is the weighting factor and B_{Ens} is the ensemble BEC. Figure 1 presents a flowchart of the hybrid data assimilation methodology adopted in this study. The ensemble BEC was evaluated using ensemble forecasts. Global Ensemble Forecast System (GEFS) data[52] were used as the driving data for providing initial and boundary conditions for the ensembles. As GEFS data contain 21 ensemble members, the hybrid assimilation also contains 21 ensemble members ($k = 21$). A value of $\beta = 0.25$ was used[44] to evaluate B_{hyb} (ref. 44), i.e. 25% of B_{3DVar} and 75% of B_{Ens}. After the processing, quality control and thinning of observation data in WRFDA, the background forecast was updated using data observations and B_{hyb}. The updated background forecast is called analysis and contains the improved initial conditions. This was used as the background for the following forecast cycle. Ensemble members were also updated using the observation data. After ensemble assimilation, the ensemble background forecasts were updated to ensemble analysis for the next forecast cycle[53]. This procedure was repeated continuously for several cycles. Single-resolution assimilation was performed in this study, where both the deterministic and ensemble forecasts were simulated on a similar grid resolution. The ensemble analyses are recentred in some studies using the deterministic analysis for the next forecast cycle. However, recentring was not used in this study due to its small impact on the ensemble analyses[54].

Design of experiments

Domain and model configuration

The domain has a resolution of 12 km in the horizontal direction (Figure 2). WRF model version 4.0 was used in this study[55]. The domain covers the Indian subcontinent and its

Table 1. List of parameters and their default and calibrated values

Scheme	Parameter	Default value	Calibrated value	Description
Cumulus	pd	1	2	Multiplier for the downdraft mass flux rate
	pe	1	0.5693	Multiplier for the entrainment mass flux rate
	phusl	150	349.9996	Starting height of downdraft over USL (hPa)
	timec	2,700	2,770.13	Mean consumption time of CAPE (s)
Microphysics	ice stokes	14,900	18,147.36	Scaling factor applied to ice-fall velocity (s^{-1})
Shortwave	cssca fac	1e-5	5.73e-6	Scattering tuning parameter (m^2 kg^{-1})
Land surface	porsl	1	0.5	Multiplier for saturated soil water content
	bsw	1	1.168	Multiplier for Clapp and Hornberger *b* parameter
Planetary boundary layer	Brcr sb	0.25	0.4423	Critical Richardson number for the boundary layer of land

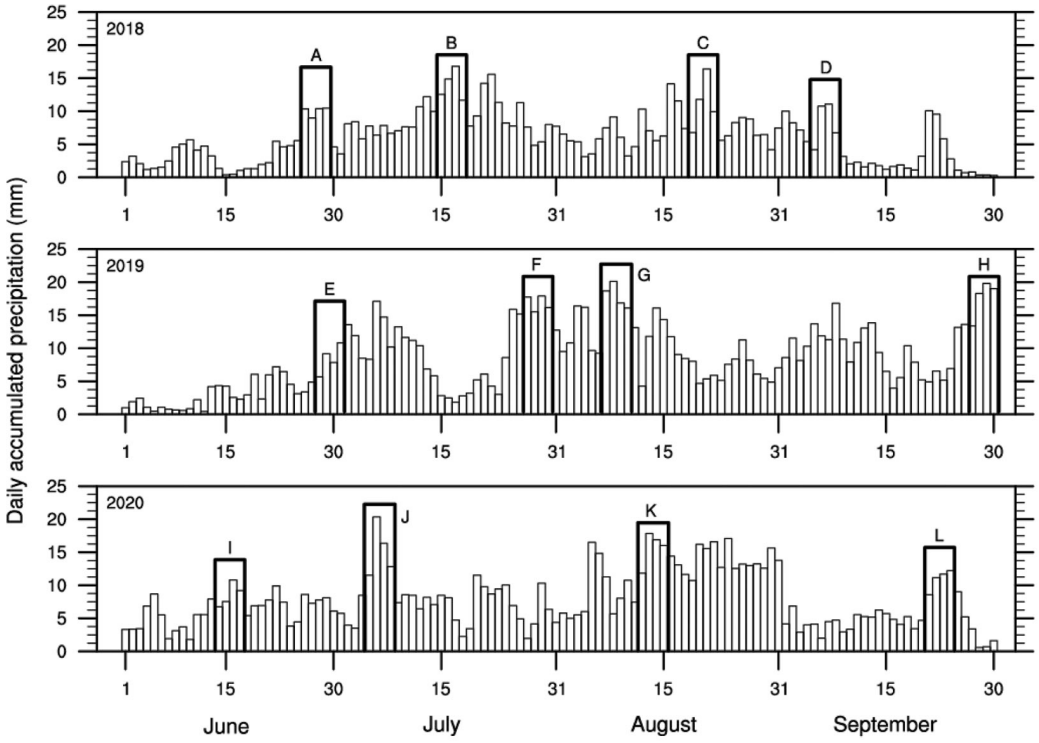

Figure 3. Daily regional average rainfall in the MCR during Indian summer monsoon for the period 2018–2020. Solid line boxes show the events that are simulated.

surrounding regions. It comprises 470 points in the zonal direction and 460 points in the meridional direction. The central point is 19°N, 80°E. The time-step used is 40 s. The model is discretized into 40 sigma (σ) layers vertically with the top layer at 50 hPa level in the atmosphere. The initial and lateral boundary conditions are obtained from the National Centers for Environmental Prediction Global Forecast System model six-hourly data at 0.5° × 0.5° resolution.

The parameterization schemes were the same as those used in parameter calibration[49]. The Kain–Fritsch eta scheme was used for cumulus parameterization[56], the WSM 6 single-class scheme for microphysics parameterization[57], the Dudhia scheme for shortwave radiation[58], the RRTM scheme for longwave radiation[59], the MM5 Monin–Obukhov scheme for the surface layer[60], the Yonsei University scheme for the planetary boundary layer[61], and the Noah scheme for land surface parameterization[62]. As mentioned earlier, each scheme has multiple parameters that can be calibrated. Table 1 lists nine sensitive parameters that influence the prediction[47]. The table also presents the default and calibrated values of these sensitive parameters. The calibrated parameter values minimized the prediction errors of daily accumulated rainfall (RAIN), surface air temperature (SAT), surface air pressure (SAP) and wind speed at 10 m (WS10)[49].

Events simulated

The monsoon core region (MCR) is a critical zone where the variation of ISM rainfall all over the country is similar

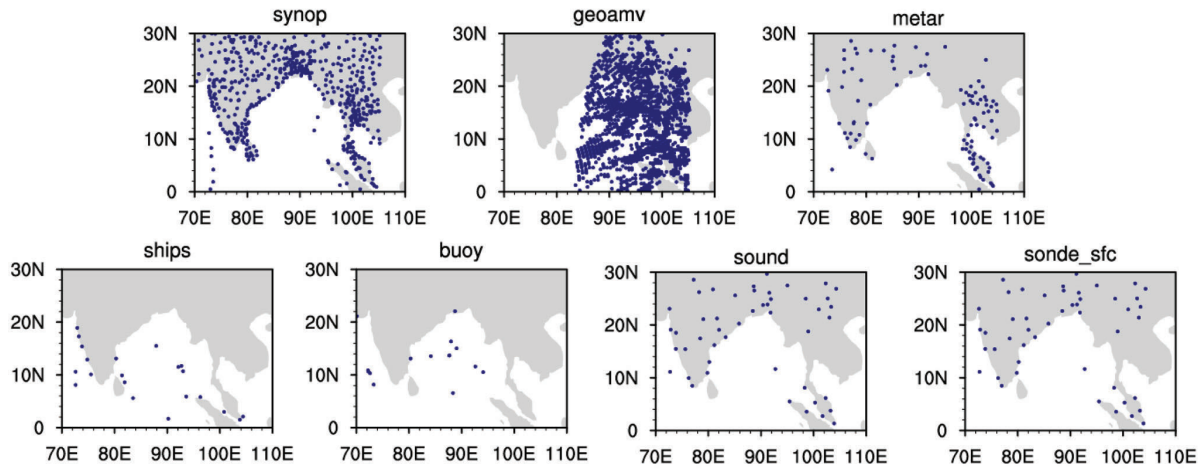

Figure 4. Distribution of conventional observations at 00 UTC on 7 July 2019 assimilated for event F.

Table 2. Overview of the observations assimilated

NCEP prepbufr observations	
Atmospheric winds	GEOAMV
Land surface	SYNOP, METAR, SONDE SFC
Marine surface	BUYO, SHIPS
Upper air	SOUND
Satellite radiance observations	
Instrument	Satellite
AMSU-A	NOAA 15,16,18,19; EOS-Aqua; METOP-A
AIRS	EOS-Aqua
HIRS-4	NOAA 18
IASI	METOP-A
MHS	METOP-A; NOAA 18,19
SSMI	DMSP 16

Table 3. Details of numerical experiments

Experiment	Description
DEF_NA	Simulations with default parameters without data assimilation
CAL_NA	Simulations with calibrated parameters without data assimilation
DEF_3DV	Simulations with default parameters and three-dimensional variational data assimilation
DEF_HYB	Simulations with default parameters and three-dimensional ensemble variational hybrid data assimilation
CAL_3DV	Simulations with calibrated parameters and three-dimensional variational data assimilation
CAL_HYB	Simulations with calibrated parameters and three-dimensional ensemble variational hybrid data assimilation

to that in MCR[63]. It ranges from 69°E to 88°E and 18°N to 28°N. So, the region of interest is MCR. The monsoon core region is also shown in Figure 2 for reference. Twelve four-day heavy rainfall events between June and September for the period 2018–20 over MCR are simulated in this study (Figure 3). Each event consists of the day with the highest accumulated rainfall, averaged over MCR, in the respective month and spans a period of four days[13]. The rainfall data are taken from the India Meteorological Department (IMD) daily accumulated gridded rainfall data at 0.25° × 0.25° resolution[64]. Figure 3 shows the dates of the 12 chosen events (A–H).

Observation data for assimilation and verification datasets

The observation data used for assimilation are summarized in Table 2, and include both conventional and satellite radiance data. Observations were assimilated at six-hourly intervals. Conventional observations comprise global surface and upper-air data collected by NCEP[65] and are provided in PREPBUFR format. The data contain land surface, marine surface, radiosonde, pibal and aircraft reports from the Global Telecommunications System, profiler and satellite wind data. Figure 4 presents the conventional data from various sources valid at 00 UTC on 7 July 2019, used in the assimilation for event F. Satellite radiance data (BUFR format) from various instruments on-board different satellites were used for assimilation[66]. The instruments include Advanced Microwave Sounding Unit-A, Atmospheric Infrared Sounder (AIRS), High-resolution Infrared Sounder-4, Infrared Atmospheric Sounding Interferometer, Microwave Humidity Sounder and Special Sensor Microwave/Imager.

The WRF model output variables were compared with the verification data to evaluate the accuracy of the simulations. RAIN was validated using IMD daily accumulated gridded rainfall data at 0.25° × 0.25° resolution. SAT, SAP and WS10 were verified with the Indian monsoon data assimilation and analysis regional reanalysis data at 0.12° × 0.12° resolution[67,68].

Experimental set-up

This study assesses the influence of hybrid assimilation on a parameter-calibrated WRF model on predicting heavy

Table 4. RMSE values of the variables rainfall (RAIN), surface air temperature (SAT), wind speed at 10 m (WS10), and surface air pressure (SAP), averaged over 48 days (12 four-day events) for different experiments

Experiments	RAIN (mm/day)	SAT (K)	WS10 (m/s)	SAP (hPa)
DEF_NA	28.93	1.72	2.69	874.21
CAL_NA	26.54 (8.24%)	1.42 (17.54%)	2.21 (17.77%)	825.93 (5.52%)
DEF_3DV	24.74 (14.47%)	1.86 (–8.00%)	2.16 (19.80%)	827.13 (5.39%)
DEF_HYB	24.82 (14.21%)	1.68 (2.71%)	2.07 (23.15%)	823.63 (5.79%)
CAL_3DV	24.09 (16.74%)	1.69 (1.88%)	1.99 (25.99%)	821.85 (5.99%)
CAL_HYB	23.71 (18.04%)	1.59 (7.91%)	1.95 (27.65%)	822.59 (5.90%)

Values within brackets represent the reduction in value of RMSE for the corresponding experiment with respect to the default experiment.

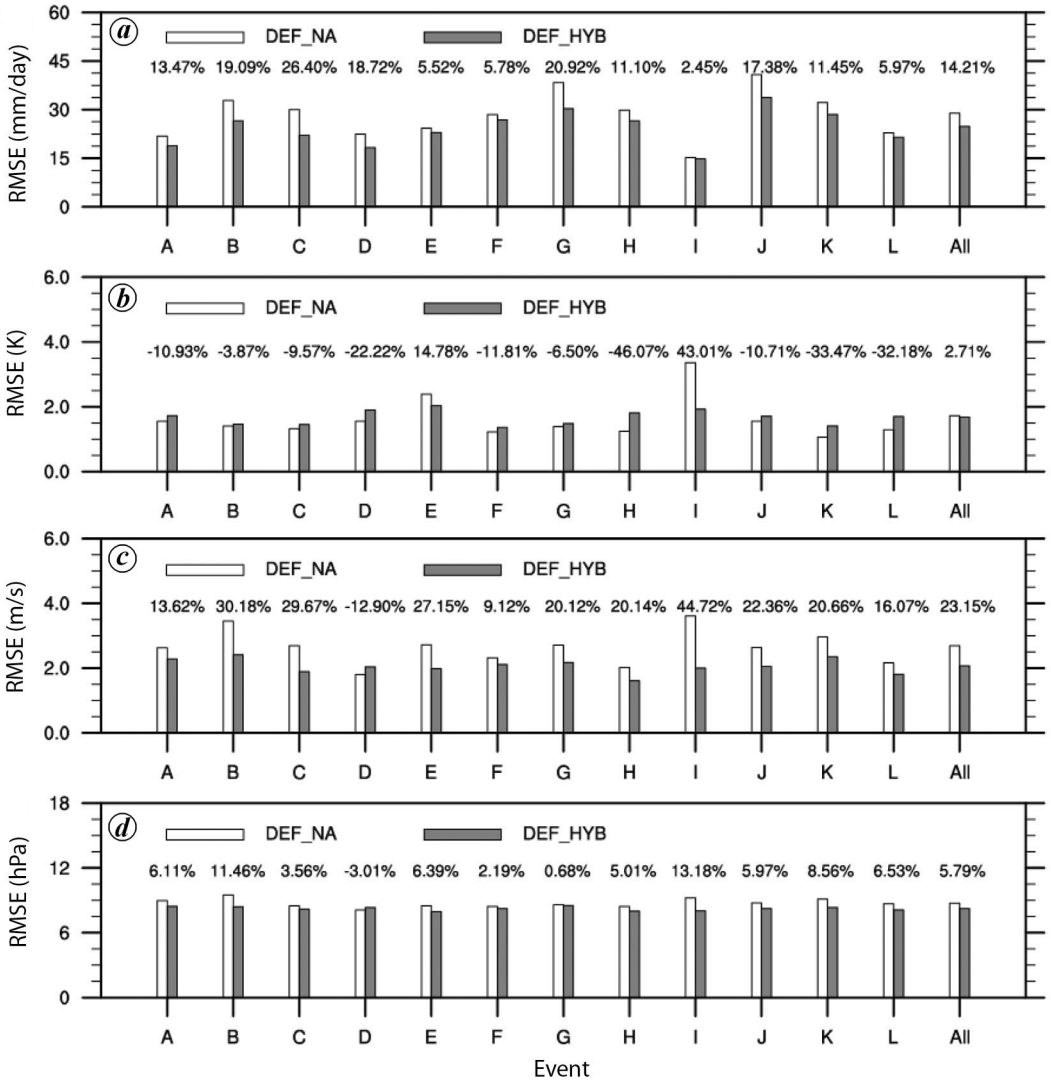

Figure 5. Comparison of RMSE values of the WRF model variables for DEF_NA and DEF_HYB experiments: ***a***, rainfall; ***b***, surface air temperature; ***c***, wind speed at 10 m; ***d***, surface air pressure. Values above the bars indicate reduction in RMSE of DEF_HYB compared to DEF_NA.

rainfall events during ISM. Six experiments (for each of the 12 events) were performed. Table 3 summarizes the details. The first experiment was performed with default parameters without data assimilation (DEF_NA) to set a benchmark for comparison. The second experiment was performed with calibrated parameters without data assimilation (CAL_NA) to evaluate the influence of calibration on the model output. The third and fourth experiments were performed with

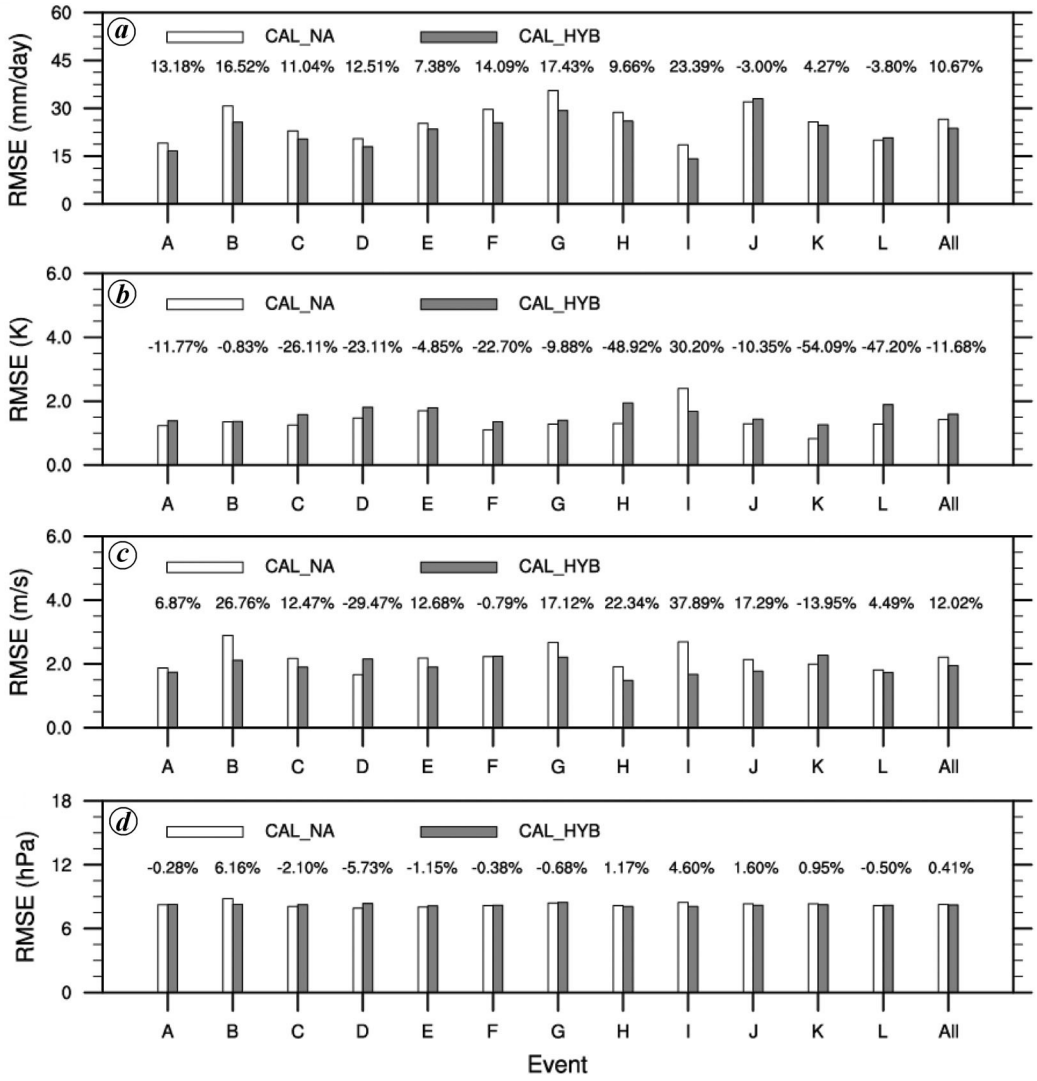

Figure 6. Comparison of RMSE values of the WRF model variables for CAL_NA and CAL_HYB experiments: ***a***, rainfall; ***b***, surface air temperature; ***c***, wind speed at 10 m; ***d***, surface air pressure. Values above the bars indicate reduction in RMSE of CAL_HYB compared to CAL_NA.

default parameters using 3DVar (DEF_3DV) and 3DEnVar assimilation (DEF_HYB) respectively, to evaluate the influence of assimilation on the default model parameters. The fifth and sixth experiments were performed with calibrated parameters using 3DVar (CAL_3DV) and 3DEnVar assimilation (CAL_HYB) respectively, to evaluate the influence of assimilation on the calibrated model parameters.

For experiments without data assimilation (DEF_NA and CAL_NA), the simulations were started at 00 UTC on the first day of each event and run continuously for 96 h. For experiments with data assimilation, simulations were started 24 h before the beginning of each event. A 6 h spin-up followed by four cycles of assimilation at six-hourly intervals was performed. After four cycles of assimilation, the model was continuously run for 96 h. Root mean square error (RMSE) as defined in eq. (4) was used for evaluating the simulations. It was also used to calibrate the parameters in the earlier study[49], as the objective was to determine the parameter values that minimize RMSE.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathrm{sim}_i^t - \mathrm{obs}_i^t)^2}{N \times T}}, \tag{4}$$

where obs_i^t and sim_i^t are the observed and simulated values at grid point i and time t respectively, N the total number of grid points in MCR and T is the number of simulation days.

Results and discussion

Table 4 summarizes the results obtained from the experiments. RMSE values of the variables RAIN, SAT, WS10

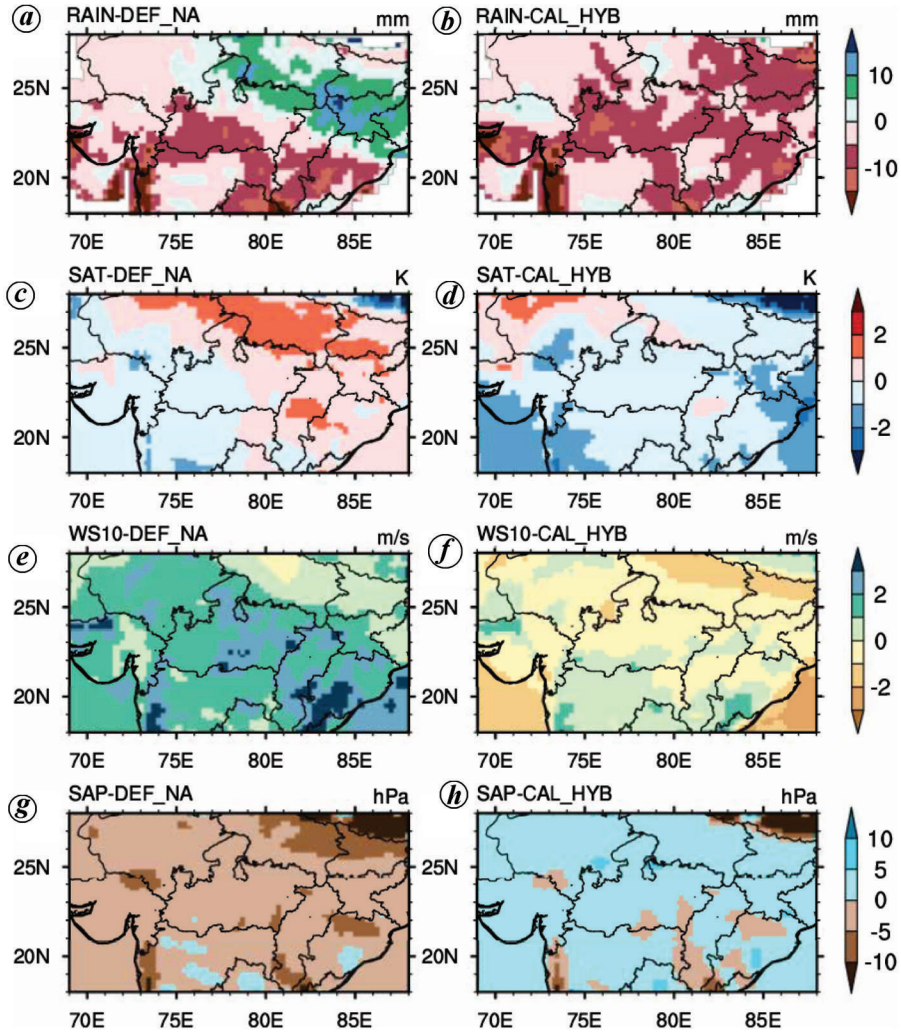

Figure 7. Spatial pattern comparison of bias for DEF_NA and CAL_HYB for the variables RAIN (***a***, ***b***), SAT (***c***, ***d***), WS10 (***e***, ***f***) and SAP (***g***, ***h***) averaged over 48 days (12 four-day events).

and SAP were evaluated for each experiment and averaged across 12 heavy rainfall events. The percentage values within brackets represent the reduction in RMSE for the corresponding experiment with respect to the DEF_NA. Calibrated parameters performed better than default parameters by reducing RMSE for all variables, viz. RAIN (8.24%), SAT (17.54%), WS10 (17.77%) and SAP (5.52%) without assimilation. Using assimilation on default parameters improved the overall prediction corresponding to both algorithms, with HYB performing similar to or better than 3DV for all variables compared to DEF_NA. The results were similar for DEF_3DV and DEF_HYB for RAIN (≈14%) and SAP (≈5%). However, DEF_HYB performed better than DEF_3DV for WS10 and SAT. Even for calibrated parameters, HYB performed similarly to or better than 3DV for all variables. CAL_HYB performed better than CAL_NA for all variables except SAT.

Figure 5 shows the event-wise RMSE values for DEF_HYB and DEF_NA experiments to assess the impact of data assimilation on the default parameters. These results show a general trend with a substantial reduction in RMSE of all variables for DEF_HYB, except SAT. RMSE values for RAIN were reduced for all events with an average decrease of 14.21% for DEF_HYB compared to DEF_NA. A similar trend was observed for WS10 (23.15%) and SAP (5.79%), where DEF_HYB performed better for almost all the events. However, for SAT, RMSE increased for most of the events. Although the average RMSE (2.71%) had reduced, it was skewed primarily because of event I, where it reduced (43%) from 3.36 K to 1.92 K. Figure 6 shows the event-wise RMSE values for CAL_HYB and CAL_NA experiments to assess the impact of data assimilation on the calibrated parameters. RMSE values of RAIN (10.67%) and WS10 (12.02%) had reduced after assimilation for most of the events. The RMSE of SAP were similar with and without assimilation, with no significant increase or decrease for almost all events after assimilation. However, the RMSE values of SAT increased after assimilation for

most of the events, with an overall increase of –11.68%. Although there was an increase in RMSE for SAT after assimilation, the overall effect of calibration and assimilation had reduced (7.91%) compared to the default parameters without assimilation.

Figure 7 compares the spatial patterns of bias (simulated minus observed) for the DEF_NA and CAL_HYB experiments for RAIN, SAT, WS10 and SAP, averaged over 48 days (12 four-day events). The bias of RAIN was similar for both DEF_NA and CAL_HYB in the western region. However, a strong positive bias in the eastern part for DEF_NA was replaced by a negative bias for CAL_HYB. Although the reduction in overall bias is not evident from this figure, the RMSE values (18.04%) indicate a significant decrease for CAL_HYB compared to DEF_NA. In the case of SAT, a strong positive bias in the northern region for DEF_NA was replaced by a weak negative bias for CAL_HYB. Also, a weak positive (negative) bias in the western (eastern) part was replaced by a strong negative bias. Overall, a reduction in RMSE (7.91%) was observed for CAL_HYB compared to DEF_NA. A strong positive bias in the entire region for DEF_NA was replaced by a weak negative or positive bias for CAL_HYB, which is consistent with a considerable reduction in RMSE (27.65%) for CAL_HYB compared to DEF_NA. In the case of SAP, a weak negative bias in DEF_NA was replaced by a weak positive bias for CAL_HYB in the entire region. The reduction in RMSE was also small (5.90%) for CAL_HYB compared to DEF_NA. Event-wise spatial comparison, presented in Supplementary Figures 1–12, also shows a general trend of improvement in the spatial pattern of all variables for CAL_HYB compared to DEF_NA.

Conclusion

The impact of assimilation on the parameter-calibrated WRF model was assessed. Twelve heavy rainfall events were simulated. The calibrated parameters performed better than the default parameters, reducing RMSE for all variables. There was considerable improvement in predicting all variables with assimilation on both the default and calibrated parameters, except for SAT. Also, 3DEnVar expectedly performed better than the 3DVar assimilation method. Overall, hybrid assimilation with calibrated parameters showed a significant improvement for the variables RAIN (18.04%), SAT (7.91%), WS10 (27.65%) and SAP (5.90%) compared to the default parameters without assimilation. A further improvement in prediction could be obtained using different methods. Performing an observing system simulation experiment can help identify observations that do not improve the initial conditions. These observations can be omitted, thereby improving the results from assimilation. Assimilating observations from other satellite instruments and radars could also help improve the predictions. Also, implementing advanced assimilation techniques such as four-dimensional ensemble variational (4DEnVar) hybrid assimilation and particle filter could further improve the model prediction.

1. Singh, D., Ghosh, S., Roxy, M. K. and McDermid, S., Indian summer monsoon: extreme events, historical changes, and role of anthropogenic forcings. *Wiley Interdiscip. Rev.: Climate Change*, 2019, **10**(2), e571.
2. Dash, S. K., Kulkarni, M. A., Mohanty, U. C. and Prasad, K., Changes in the characteristics of rain events in India. *J. Geophys. Res.: Atmosp.*, 2009, **114**(D10).
3. Pattanaik, D. R. and Rajeevan, M., Variability of extreme rainfall events over India during southwest monsoon season. *Meteorol. Appl.*, 2010, **17**(1), 88–104.
4. Bjerknes, V., Hesselberg, T. and Devik, O., Dynamic Meteorology and Hydrography2. Kinematics. Number v. 2 in Publication, Carnegie Inst., 1911.
5. Stensrud, D. J., *Parameterization Schemes: Keys to Understanding Numerical Weather Prediction Models*, Cambridge University Press, 2009.
6. Mukhopadhyay, P., Taraphdar, S., Goswami, B. N. and Krishnakumar, K., Indian summer monsoon precipitation climatology in a high-resolution regional climate model: impacts of convective parameterization on systematic biases. *Weather Forecast.*, 2010, **25**(2), 369–387.
7. Srinivas, C. V., Hariprasad, D., Bhaskar Rao, D. V., Anjaneyulu, Y., Baskaran, R. and Venkatraman, B., Simulation of the Indian summer monsoon regional climate using advanced research WRF model. *Int. J. Climatol.*, 2013, **33**(5), 1195–1210.
8. Attada Raju, Anant Parekh, Chowdary, J. S. and Gnanaseelan, C., Assessment of theIndian summer monsoon in the WRF regional climate model. *Climate Dyn.*, 2015, **44**(11–12), 3077–3100.
9. Ratnam, J. V., Behera, S. K., Krishnan, R., Doi, T. and Ratna, S. B., Sensitivity of Indian summer monsoon simulation to physical parameterization schemes in the WRF model. *Climate Res.*, 2017, **74**(1), 43–66.
10. Sandeep, C. P. R., Krishnamoorthy, C. and Balaji, C., Impact of cloud parameterization schemes on the simulation of cyclone Vardah using the WRF model. *Curr. Sci.*, 2018, **115**(6), 1143–1153.
11. Park Sojung and Park, S. K., A micro-genetic algorithm (GA v1. 7.1 a) for combinatorial optimization of physics parameterizations in the Weather Research and Forecasting model (v4. 0.3) for quantitative precipitation forecast in Korea. *Geosci. Model Dev.*, 2021, **14**(10), 6241–6255.
12. Baki, H., Chinta, S., Balaji, C. and Srinivasan, B., A sensitivity study of WRF model microphysics and cumulus parameterization schemes for the simulation of tropical cyclones using GPM radar data. *J. Earth Syst. Sci.*, 2021, **130**(4), 1–30.
13. Di, Z. *et al.*, Assessing WRF model parameter sensitivity: a case study with 5-day summer precipitation forecasting in the Greater Beijing area. *Geophys. Res. Lett.*, 2015, **42**(2), 579–587.
14. Hourdin, F. *et al.*, The art and science of climate model tuning. *Bull. Am. Meteorol. Soc.*, 2017, **98**(3), 589–602.
15. Bellprat, O., Kotlarski, S., Lüthi, D. and Schär, C., Exploring perturbed physics ensembles in a regional climate model. *J. Climate*, 2012, **25**(13), 4582–4599.
16. Di, Z., Duan, Q., Wang, C., Ye, A., Miao, C. and Gong, W., Assessing the applicability of WRF optimal parameters under the different precipitation simulations in the Greater Beijing Area. *Climate Dyn.*, 2018, **50**(56), 1927–1948.
17. Edwards, N. R., Cameron, D. and Rougier, J., Precalibrating an intermediate complexity climate model. *Climate Dyn.*, 2011, **37**(7–8), 1469–1482.
18. Williamson, D., Goldstein, M., Allison, L., Blaker, A., Challenor, P., Jackson, L. and Yamazaki, K., History matching for exploring and reducing climate model parameter space using observations

and a large perturbed physics ensemble. *Climate Dyn.*, 2013, **41**(7–8), 1703–1729.
19. Quan, J. *et al*., An evaluation of parametric sensitivities of different meteorological variables simulated by the WRF model. *Q. J. R. Meteorol. Soc.*, 2016, **142**(700), 2925–2934.
20. Hou, Z., Huang, M., Ruby Leung, L., Lin, G. and Ricciuto, D. M.. Sensitivity of surface flux simulations to hydrologic parameters based on an uncertainty quantification framework applied to the Community Land Model. *J. Geophys. Res.: Atmosp.*, 2012, **117**(D15).
21. Li, J. *et al.*, Assessing parameter importance of the Common Land Model based on qualitative and quantitative sensitivity analysis. *Hydrol. Earth Syst. Sci.*, 2013, **17**(8), 3279.
22. Wang, C. *et al.*, Assessing the sensitivity of land–atmosphere coupling strength to boundary and surface layer parameters in the WRF model over Amazon. *Atmosp. Res.*, 2020, **234**, 104738.
23. Baki, H., Chinta, S., Balaji, C. and Srinivasan, B., Determining the sensitive parameters of the Weather Research and Forecasting (WRF) model for the simulation of tropical cyclones in the Bay of Bengal using global sensitivity analysis and machine learning. *Geosci. Model Dev.*, 2022, **15**(5), 2133–2155.
24. Williamson, D., Blaker, A. T., Hampton, C. and Salter, J., Identifying and removing structural biases in climate models with history matching. *Climate Dyn.*, 2015, **45**(5–6), 1299–1324.
25. Duan, Q. *et al*., Automatic model calibration: a new way to improve numerical weather forecasting. *Bull. Am. Meteorol. Soc.*, 2017, **98**(5), 959–970.
26. Yang, B. *et al.*, Parametric and structural sensitivities of turbine–height wind speeds in the boundary layer parameterizations in the Weather Research and Forecasting model. *J. Geophys. Res.: Atmosp.*, 2019, **124**(12), 5951–5969.
27. Baki, H., Chinta, S., Balaji, C. and Srinivasan, B., Parameter calibration to improve the prediction of tropical cyclones over the Bay of Bengal using machine learning-based multi-objective optimization. *J. Appl. Meteorol. Climatol.*, 2022.
28. Duan, Q., Sorooshian, S. and Gupta, V. K., Optimal use of the SCE-UA global optimization method for calibrating watershed models. *J. Hydrol.*, 1994, **158**(3–4), 265–284.
29. Severijns, C. A. and Hazeleger, W., Optimizing parameters in an atmospheric general circulation model. *J. Climate*, 2005, **18**(17), 3527–3535.
30. Jackson, C. S., Sen, M. K., Huerta, G., Deng, Y. and Bowman, K. P., Error reduction and convergence in climate prediction. *J. Climate*, 2008, **21**(24), 6698–6709.
31. Ollinaho, P., Järvinen, H., Bauer, P., Laine, M., Bechtold, P., Susiluoto, J. and Haario, H., Optimization of NWP model closure parameters using total energy norm of forecast error as a target. *Geoscientific Model Dev.*, 2014, **7**(5), 1889–1900.
32. Bannister, R. N., A review of operational methods of variational and ensemble variational data assimilation. *Quart. J. R. Meteorol. Soc.*, 2017, **143**(703), 607–633.
33. Zhang, F., Meng, Z. and Aksoy, A., Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part I: Perfect model experiments. *Mon. Weather Rev.*, 2006, **134**(2), 722–736.
34. Liu, H. and Xue, M., Prediction of convective initiation and storm evolution on 12 June 2002 during IHOP-2002. Part I: Control simulation and sensitivity experiments. *Mon. Weather Rev.*, 2008, **136**(7), 2261–2282.
35. Parrish, D. F. and Derber, J. C., The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Weather Rev.*, 1992, **120**(8), 1747–1763.
36. Lorenc, A. C., Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, 1986, **112**(474), 1177–1194.
37. Evensen, G., Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.: Oceans*, 1994, **99**(C5), 10143–10162.
38. Hamill, T. M., Whitaker, J. S. and Snyder, C., Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Weather Rev.*, 2001, **129**(11), 2776–2790.
39. Hamill, T. M. and Snyder, C., A hybrid ensemble Kalman filter-3D variational analysis scheme. *Mon. Weather Rev.*, 2000, **128**(8), 2905–2919.
40. Lorenc, A. C., The potential of the ensemble Kalman filter for NWP – a comparison with 4D-Var. *Q. J. R. Meteorol. Soc.*, 2003, **129**(595), 3183–3203.
41. Hamill, T. M., Whitaker, J. S., Fiorino, M. and Benjamin, S. G., Global ensemble predictions of 2009s tropical cyclones initialized with an ensemble Kalman filter. *Mon. Weather Rev.*, 2011, **139**(2), 668–688.
42. Zhang, M. and Zhang, F., E4DVar: coupling an ensemble Kalman filter with four- dimensional variational data assimilation in a limited-area weather prediction model. *Mon. Weather Rev.*, 2012, **140**(2), 587–600.
43. Kleist, D. T. and Ide, K., An OSSE-based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS. Part I: system description and 3Dhybrid results. *Mon. Weather Rev.*, 2015, **143**(2), 433–451.
44. Prasad, V. S., Johny, C. J. and Sodhi, J. S., Impact of 3DVar GSI-ENKF hybrid data assimilation system. *J. Earth Syst. Sci.*, 2016, **125**(8), 1509–1521.
45. Singh, S. K. and Prasad, V. S., Evaluation of precipitation forecasts from 3DVar and hybrid GSI-based system during Indian summer monsoon 2015. *Meteorol. Atmosp. Phys.*, 2019, **131**(3), 455–465.
46. Morris, M. D., Factorial sampling plans for preliminary computational experiments. *Technometrics*, 1991, **33**(2), 161–174.
47. Chinta, S., Yaswanth Sai, J. and Balaji, C., Assessment of WRF model parameter sensitivity for high-intensity precipitation events during the Indian summer monsoon. *Earth Space Sci.*, 2021, **8**(6), e2020EA001471.
48. Wang, C., Duan, Q., Gong, W., Ye, A., Di, Z. and Miao, C., An evaluation of adaptive surrogate modeling based optimization with two benchmark problems. *Environ. Modell. Softw.*, 2014, **60**, 167–179.
49. Chinta, S. and Balaji, C., Calibration of WRF model parameters using multiobjective adaptive surrogate model-based optimization to improve the prediction of the Indian summer monsoon. *Climate Dyn.*, 2020, **55**(3), 631–650.
50. Dhanya, M. and Chandrasekar, A., Multivariate background error covariances in the assimilation of SAPHIR radiances in the simulation of three tropical cyclones over the Bay of Bengal using the WRF model. *Int. J. Remote Sensing*, 2018, **39**(1), 191–209.
51. Baki, H., Balaji, C. and Srinivasan, B., Impact of data assimilation on a calibrated WRF model for the prediction of tropical cyclones over the Bay of Bengal. *Curr. Sci.*, 2022, **122**(5), 569–583.
52. Cui, B., Toth, Z., Zhu, Y. and Hou, D., Bias correction for global ensemble forecast. *Weather Forecast.*, 2012, **27**(2), 396–410.
53. Gao, J., Fu, C., Stensrud, D. J. and Kain, J. S., OSSEs for an ensemble 3DVAR data assimilation system with radar observations of convective storms. *J. Atmosp. Sci.*, 2016, **73**(6), 2403–2426.
54. Wang, X., Parrish, D., Kleist, D. and Whitaker, J., GSI 3DVar-based ensemble-variational hybrid data assimilation for NCEP Global Forecast System: single-resolution experiments. *Mon. Weather Rev.*, 2013, **141**(11), 4098–4117.
55. Skamarock, W. C. *et al.*, A description of the advanced research WRF model version 4. National Center for Atmospheric Research, Boulder, CO, USA, 2019, p. 145.
56. Kain, J. S., The Kain–Fritsch convective parameterization: an update. *J. Appl. Meteorol.*, 2004, **43**(1), 170–181.
57. Hong, S.-Y. and Jade Lim, J.-O., The WRF single-moment 6-class microphysics scheme (WSM6). *Asia-Pac. J. Atmosp. Sci.*, 2006, **42**(2), 129–151.
58. Dudhia, J., Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *J. Atmosp. Sci.*, 1989, **46**(20), 3077–3107.

59. Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J. and Clough, S. A., Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-*k* model for the longwave. *J. Geophys. Res.: Atmosp.*, 1997, **102**(D14), 16663–16682.
60. Dudhia, J., *PSU/NCAR Mesoscale Modeling System, Tutorial Class Notes and Users' Guide*, MM5 Modeling System Version 3, 2005.
61. Hong, S.-Y., Noh, Y. and Dudhia, J., A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Weather Rev.*, 2006, **134**(9), 2318–2341.
62. Chen, F. and Dudhia, J., Coupling an advanced land surface–hydrology model with the Penn State-NCAR MM5 modeling system. Part I: model implementation and sensitivity. *Mon. Weather Rev.*, 2001, **129**(4), 569–585.
63. Rajeevan, M., Gadgil, S. and Bhate, J., Active and break spells of the Indian summer monsoon. *J. Earth Syst. Sci.*, 2010, **119**(3), 229–247.
64. Pai, D. S., Sridhar, L., Rajeevan, M., Sreejith, O. P., Satbhai, N. S. and Mukhopadhyay, B., 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*, 2014, **65**(1), 1–18.
65. NCSP, NCEP ADP global upper air and surface weather observations (prepbufr format), National Centers for Environmental Prediction, National Weather Service, NOAA, US Department of Commerce, 2008.
66. NCEP, NCEP GDAS satellite data 2004 – continuing, National Centers for Environmental Prediction, National Weather Service, 2009.
67. Ashrit, R. *et al.*, IMDAA regional reanalysis: performance evaluation during Indian summer monsoon season. *J. Geophys. Res.: Atmosp.*, 2020, e2019JD030973.
68. Indira Rani, S. *et al.*, IMDAA: high-resolution satellite-era reanalysis for the Indian monsoon region. *J. Climate*, 2021, **34**(12), 5109–5133.

ACKNOWLEDGEMENTS. This work was supported by the National Monsoon Mission (Project No. IITM/MM-II/IITM/2018/IND-8) of the Ministry of Earth Sciences (MoES), Government of India. Simulations were performed on the Aaditya high-performance computing system at the Indian Institute of Tropical Meteorology in Pune. Initial and boundary conditions as well as observation data were from the National Centers for Environmental Prediction, USA. We thank the National Centre for Medium Range Weather Forecasting, MoES, GoI for the IMDAA reanalysis.

Received 27 June 2022; revised accepted 24 December 2022

doi: 10.18520/cs/v124/i6/693-703