Assimilation of the Rain Gauge Measurements Using Particle Filter

Prashant Kumar

1Atmospheric Sciences Division, Atmospheric and Oceanic Sciences Group, Space Applications Centre, ISRO, Ahmedabad, India

Abstract The well-recognized constraint of nonlinear and non-Gaussian distribution of rainfall observation limits its assimilation in the high-dimensional numerical weather prediction (NWP) model. In this study, rainfall observed from Indian Meteorological Department (IMD) rain gauges over Indian landmass is assimilated in the Weather Research and Forecasting (WRF) model using particle filter. In the framework of imperfect weather models, particles (or ensembles) for rainfall predictions are created with various combinations of model physics (viz., cumulus parameterization, microphysics and planetary boundary layer schemes). The multiple hypotheses are used to determine the weights for different particles, and this is the step where IMD rainfall data are used for assimilation. Further, a resampling step is performed to generate new particles from high weight particles using stochastic kinetic-energy backscatter scheme (SKEBS) method in which dynamical variables are perturbed into the model physics. Results, based on rainfall verification scores, suggest that the assimilation of the rainfall using particle filter could improve the prediction of rainfall over CNT runs (unweighted particles; without assimilation). Moreover, surface and vertical profile of temperature, water vapor mixing ratio (WVMR), and wind speed are also improved in 24-hr forecasts.

1. Introduction

Accurate rainfall prediction from the numerical weather prediction (NWP) model is one of the most challenging concerns for weather modeling society in general and during Indian summer monsoon in particular. Attada et al. (2015, 2018) demonstrated the importance of the temperature and moisture profiles assimilation on the Indian summer monsoon rainfall prediction. The rainfall assimilation has large impact on weather forecast mainly over the tropics, in which moist convection plays a prominent role, and links directly or indirectly to humidity, cloud cover, latent heating, and the divergent component of the large-scale circulation (Hou et al., 2004; Marecal & Mahfouf, 2003). Various efforts have been attempted to assimilate rainfall information in the NWP model using nudging method, variational assimilation, and Kalman filter in recent decades (Kumar et al., 2014; Kumar & Kishtawal, 2017; Kumar & Varma, 2016 and references therein). It is well studied that assimilation of satellite derived rainfall in the NWP model helps to improve the analyses and subsequent short range weather forecasts (Bauer et al., 2011; Kumar et al., 2014; Kumar & Kishtawal, 2017; Kumar & Varma, 2016; Lien et al., 2013; Lien, Kalnay et al., 2016; Lien, Miyoshi et al., 2016; Lopez, 2011 and references therein). Kumar and Varma (2016) found that assimilation of satellite retrieved rainfall in the NWP model improved the forecast for unprecedented heavy rainfall, which is not able to predict from operational centers. Moreover, Kumar and Kishtawal (2017) showed that variational assimilation of both rain and no-rain information from satellites has a positive impact on short-range weather forecasts.

Errico et al. (2007) suggested that rainfall assimilation is a more complex problem compared to assimilation of conventional observations or clear-sky satellite radiance. Due to nonlinear and non-Gaussian characteristics of rainfall, still assimilation of rainfall in the NWP model is a challenging problem. Few studies (Bauer et al., 2011; Kotsuki et al., 2017; Kumar & Varma, 2016) mentioned common difficulties in the rainfall assimilation mainly due to the strong nonlinearity of moist process and non-Gaussian characteristics of precipitation. Kumar et al. (2014) discussed the importance of quality control on rainfall assimilation and showed that with strict quality control generally difficult to improve forecasts beyond a few hours due to the non-Gaussian nature of rainfall data. Additionally, due to limitations of the realistic representation of the nonlinear model physics as a tangent linear model, numerical models are not able to assimilate rainfall
precisely using the four-dimensional variational (4D-Var) data assimilation method. Most of the previous studies based on variational method and ensemble Kalman filter (EnKF) assume(convert non-Gaussian distribution of rainfall to Gaussian error statistics which lead to suboptimal analysis (e.g., Posselt et al., 2014; Posselt & Bishop, 2012; Van Leeuwen, 2009, 2010). One well-known advantage of the particle filter over EnKF is that the particle filter can work for non-Gaussian distribution (Kumar & Shukla, 2019; Mattern et al., 2013; Ratheesh et al., 2016).

The objective of this study is to assimilate the Indian Meteorological Department (IMD)'s rain gauge observed rainfall in the Weather Research and Forecasting (WRF) model using a particle filter during 1–10 August 2015. Details of the particle filter and design of experiment are given in section 2, and section 3 is discussing data used. Results and discussions are provided in section 4 and are concluded in the last section.

2. Methodology

Particle filter can be used in various fields of geosciences to estimate the optimal state of a system from an imperfect model with noisy and inadequate data (Chorin et al., 2013; Kumar & Shukla, 2019). The particle filter method is computationally more expensive compared to optimal interpolation (OI), 3D/4D-Var, and EnKF data assimilation methods. These later discussed methods are highly based on accuracy of the first guess and accurate representation of the error covariances (Van Leeuwen, 2009). Moreover, particle filter do not require tangent linear and adjoint models of the nonlinear model, which have a large uncertainty over the tropical region due to linear assumption of the nonlinear dynamical model (Kumar & Shukla, 2019). Moreover, estimation of background error covariance matrices is also not needed which also contributed large errors in other data assimilation methods. However, the issue of flow-dependent background error is resolved to some extent in EnKF method, in which artificial tricks like covariance inflation and localization are needed to get good results in high-dimensional systems (Van Leeuwen, 2009). Particle filter do not have these difficulties, and particles (or ensembles) are not adjusted, which do not destroy the dynamical balances in the analysis. The issue in the particle filter implementation is that the particles are not modified, so that after a few analysis steps, only one particle has all the weights against observations, and all other particles are moved away from the observations. It means that the statistical information in the ensemble becomes too low to be meaningful and known as filter degeneracy (Ades & Van Leeuwen, 2015). Details of the mathematical formulation of particle filters are discussed in previous studies (Ades & Van Leeuwen, 2015; Chen, 2003; Doucet et al., 2001; Kumar & Shukla, 2019; Maskell & Gordon, 2001; Ristic et al., 2004; Van Leeuwen, 2009, 2010 and references therein).

In the present study, initially rainfall forecasts are simulated from the WRF model with a total 90 different combinations of model physics, viz., cumulus physics, microphysics, and planetary boundary layer (PBL) schemes (Figure 1). Nine different cumulus parameterization schemes available in the WRF model are selected named as new Kain-Fritsch (CP1), Betts-Miller-Janjic (CP2), Grell-Freitas (CP3), Simplified Arakawa Schubert (CP4), Grell-3D ensemble (CP5), Tiedtke (CP6), New SAS (CP14), Grell-Devenyi (CP93), and old Kain-Fritsch (CP99) schemes. Five different microphysics schemes available in the WRF model named as Lin (Purdue; MP2), WSM3 (WRF Single Moment; MP3), Eta (Ferrier; MP5), WSM6 (MP6), and Thompson (MP8) are used to generate particles. These microphysics are selected based on complexity from simple WSM3 scheme, which considers only rain and cloud as hydrometer to WSM6 or Thompson scheme in which other hydrometers, e.g., ice, snow, and graupel, are also considered. The YSU (PBL1) and MYNN2 (PBL5) two different PBL schemes are selected to generate particles. Details of these physics schemes are given in the WRF user guide (Skamarock et al., 2008). Each cumulus physics (9), microphysics (5), and PBL (2) schemes are used to generate 10, 18, and 45 particles, respectively (Figure 1) on 1 August 2015. These choices of physics options are considered to perform reference experiments (unweighted particles or CNT runs). This is the forecast step of the algorithm.

In general, the particle filter considers a probability distribution function (pdf) of a state, and the pdf is approximated by particles consisting of a large number of discrete samples (here choice of physics options) to represent and approximate posteriori by a weighted sample (Kumar & Shukla, 2019). The selected particles based on different physics options (p) represent a sample form of priori pdf given as

$$x_{p,k} = f_{k-1}^{p} (x_{k-1}, v_{k-1}) \text{ for } k > 0$$

(1)
Here, \( x_{p,k} \) is the set of state vector with \( p \) different physics options to be estimated at time step \( k \), \( f_{k-1} \) is a known imperfect nonlinear model (here WRF model) with \( p \) different physics options, and \( x_{k-1} \), is the best state taken from global model analysis having noise of \( v_{k-1} \) at time step \( k-1 \). The idea is to represent the prior pdf by a set of particles \( x_{p,k} \), which are delta functions centered around state vectors. If one represents the prior pdf by a number of particles, like in the EnKF, so

\[
p_{x}(x) = \sum_{p=1}^{N} \delta(x - x_{p,k})
\]

where \( N \) is number of particles (which are 90 here). Then, from Bayes theorem

\[
p_{x/y}(x) = \sum_{p=1}^{N} w_{p} \delta(x - x_{p,k})
\]

in which weights \( w_{p} \) are given by

\[
w_{p} = \frac{p(y/x_{p})}{\sum_{q=1}^{N} p(y/x_{q})}
\]

where the density \( p(y/x_{p}) \) is the probability density of the observations (\( y \)) given the model state \( x_{p} \) in forecast time step at which IMD-observed rainfall is available over Indian landmass grids (Kumar & Shukla, 2019).

In the analysis step, observed rainfall from IMD is used to determine weight (\( w_{p} \)) for each particle. This step involves weighting to each particle and subsequent weight-based resampling. The weights for each particle depend on IMD rainfall observation, and this is the step in which rainfall observation is used for the assimilation process. Assimilation in particle filtering amounts to sequential importance resampling (SIR) weighting of particles. The weights, in principle, should be proportional to likelihoods or conditional probabilities (Kumar & Shukla, 2019). In this study, the likelihood is depending on multiple hypotheses (Dubuisson, 2015). The first hypothesis is the value of variance in rainfall forecast should be less, and the next hypothesis is the value of mean equitable threat score (ETS; section 4) should be high for different rainfall thresholds (like 5 mm/24 h and 10 mm/24 h). In the first hypothesis, likelihood is inversely related to a suitable distance between model-simulated rainfall and IMD-observed rainfall. The distance is taken to be

### Figure 1.
Schematic of rainfall assimilation using particle filter to estimate target pdf from an imperfect model using initial state from the NCEP analysis in the WRF model with different model physics and dynamic variable perturbation in the physical parametrization during 1–9 August 2015. The numbers in the boxes represent the number of particles using that particular scheme.

| Model Physics   | PBL1 | PBL5 | MP2 | MP3 | MP5 | MP6 | MP8 | CP1 | CP2 | CP3 | CP4 | CP5 | CP6 | CP14 | CP93 | CP99 |
|-----------------|------|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 01 August 2015  | 45   | 45   | 18  | 18  | 18  | 18  | 18  | 10  | 10  | 10  | 10  | 10  | 10  | 10  | 10  |
| 02 August 2015  | 40   | 50   | 15  | 20  | 18  | 17  | 20  | 17  | 00  | 09  | 11  | 14  | 16  | 10  | 16  |
| 03 August 2015  | 38   | 52   | 12  | 22  | 19  | 15  | 22  | 19  | 00  | 00  | 13  | 8   | 10  | 28  | 04  |
| 04 August 2015  | 42   | 48   | 10  | 20  | 21  | 11  | 28  | 29  | 00  | 00  | 17  | 00  | 10  | 34  | 00  |
| 05 August 2015  | 28   | 62   | 08  | 25  | 17  | 11  | 29  | 49  | 00  | 00  | 18  | 00  | 10  | 13  | 00  |
| 06 August 2015  | 20   | 70   | 05  | 41  | 15  | 29  | 20  | 57  | 00  | 00  | 17  | 00  | 05  | 11  | 00  |
| 07 August 2015  | 15   | 75   | 04  | 46  | 15  | 07  | 18  | 57  | 00  | 00  | 17  | 00  | 11  | 00  | 00  |
| 08 August 2015  | 26   | 64   | 05  | 33  | 17  | 06  | 29  | 42  | 00  | 00  | 32  | 00  | 05  | 11  | 00  |
| 09 August 2015  | 47   | 43   | 05  | 08  | 20  | 07  | 50  | 15  | 00  | 59  | 00  | 05  | 11  | 00  | 00  |

**PBL** | **Micro Physics** | **Cumulus Physics**
the usual variance between simulations and observations for daily-accumulated rainfall at observation grids. First, the distances $d_p$ is computed between the accumulated rainfall from $p$ particle at time $k$ and IMD-observed rainfall valid for same period, with $p$ varying from 1 to $N$, which is the total number of particles. Now, we calculate the raw weights as inverses of these distances. Intermediate weights are calculated by dividing the raw weights by the maximum weight and raising the ratio to a power $a$, which is an adjustable parameter. The intermediate weights are then normalized so that their sum is unity. Finally, a median filter is used to select the first 45 particles having higher weights. Next 45 particles are selected for which likelihood is directly related to a suitable mean ETS between model-simulated rainfall and IMD-observed rainfall for different rainfall thresholds. In this way, total numbers of particles (here 90 particles) are the same for all time steps, whereas, few combinations of model physics receive high weights (duplicate). In this process, observation uncertainty is not considering which may be a scope for future. Particles having large variance and less mean ETS values are rejected which contribute very little to the approximation of the target pdf.

Due to imperfect model physics (no scheme is perfect), diverse combinations of model physics are matched well with IMD-observed rainfall at different time steps, whereas, few combinations of model physics are completely rejected which are not performed well over the tropical regions in this study (Figure 1). Moreover, rainfall prediction from the NWP model is a complicated process based on different meteorological parameters (e.g., temperature, moisture, winds, fluxes, and cloud), surface characteristics (e.g., vegetation, roughness, albedo, and land type), and model physics (e.g., cumulus and microphysics). It is important to note that the dimension of the NWP model is very high ($\sim 10^9$). So a subspace (here rainfall; dimension of $\sim 10^3$) from high-dimensional model space is used for particle filter implementation, and changes in subspace modify the full state of the high-dimensional model.

Finally, in the resampling step, particles having higher weights are resampled at the observation time, whose distribution forms a weak approximation of the target pdf. In this step, new particles are generated from large weight particles (selected physics option) using stochastic kinetic-energy backscatter scheme (SKEBS; Berner et al., 2009, 2011). The advantage of SKEBS scheme is that it perturbs the dynamic state directly, and the perturbed dynamical variables are then fed into the physical parameterizations (model physics). The SKEBS scheme is very different from perturbing the physical tendencies directly, which can introduce inconsistencies between the physics and dynamics. Therefore, the tendency of the model might be to read just any such inconsistencies, possibly leading to erroneous phenomena (e.g., spurious gravity waves) (Berner et al., 2011). In this way, the total numbers of particles are the same at all observation time steps. The idea is to focus the particles toward high probability regions of the target pdf, so that the number of particles required for a good approximation of target pdf remains manageable within subspace having very less dimension as compared to actual model space.

2.1. Design of Experiment

In this study, two different sets of experiments are performed with (EXP; weighted particles) and without (CNT; unweighted particles) assimilation of IMD-observed rainfall using particle filter during 1–9 August 2015. The summer monsoon 2015 was the fourth case of two consecutive all India deficit monsoon years during the last 115 years. In the first (June and July) and second (August and September) half of the monsoon season, the country received around 95% and 77% rainfall of long period average. Author considers the first 10 days of the second half of the monsoon season to assess performance during deficit monsoon. For the comparison purpose, accumulated rainfall forecasts predicted from the WRF model are interpolated to IMD observation grids using bilinear interpolation for the same time period (here 24 hr). The National Centers for Environmental Prediction (NCEP) Global Data Assimilation System (GDAS) global model analysis at 1.0° × 1.0° spatial resolution is used to prepare initial and lateral boundary conditions for the WRF model. The WRF model simulations are performed at 25-km spatial resolution using different physics options and SKEBS perturbations. The details of the selected WRF model configurations are given in Kumar and Shukla (2019). The NCEP GDAS analysis is the final analysis from NCEP that assimilated various kinds of observations available from ground and satellites, and including late arriving observations. Here, the objective is to estimate target pdf from an imperfect model with different model physics and dynamic variable perturbation in the physical parameterization using the best initial state (assume here) available from the NCEP analysis. The selection of different model physics after rainfall assimilation during 1–9 August 2015 is shown in Figure 1, which shows that few model physics which are not very appropriate
over this part of the world are rejected, and provide the higher weight to model physics which are more suitable over this region. To avoid rapid filter degeneracy where it approaches single model physics, a dynamic variable perturbation in model physics is included using the SKEBS method. The choice of the NCEP GDAS analysis as input in CNT runs is mainly to assess the superiority of rainfall assimilation using particle filter.

3. Data Used

In this study, the IMD-observed rainfall, the Tropical Rainfall Measuring Mission (TRMM) 3B42 merged rainfall product, and the NCEP GDAS global model analysis are used in different stages. The NCEP GDAS global model analysis is used to create initial and lateral boundary conditions for the WRF model during 1–10 August 2015. Furthermore, forecasts of surface and vertical profile of temperature, moisture and wind speed are compared with the NCEP GDAS analysis on 10 August 2015. The IMD-observed rainfall is used majorly for assimilation by particle filter (EXP runs) during 1–9 August 2015, and TRMM 3B42 merged rainfall products are used to assess the skill of rainfall prediction on 10 August 2015. Details of these data sets are given below:

3.1. IMD Rainfall

Daily gridded rainfall data over the Indian landmass is available since January 1901 at a spatial resolution of 0.25° latitude/longitude. This data set is prepared from daily-recorded information from about 7,000 Surface Rain Gauge (SRG) stations well spread across the country after incorporating the necessary quality control (Pai et al., 2014). The quality control test involves verification of the location information of the gauge station, eliminating the missing data, eliminating the coding errors, extreme value check, etc. These data are interpolated using Shepard interpolation method into a regular grid (Pai et al., 2014). The distribution of gauges over India is satisfactory in terms of number and regional distribution, except some small regions of Jammu and Kashmir (J&K) and extreme northwest parts of India.

3.2. NCEP GDAS Analysis

The NCEP implemented operationally a series of numerical models for the generation of global model analyses and forecasts. One of the operational systems is GDAS (Kanamitsu, 1989), which uses the spectral Medium Range Forecast (MRF) model. The GDAS analysis is the final run in a series of the NCEP operational model; therefore, it is also known as the final run at the NCEP that also includes the late arriving conventional and satellite observations. It is run four times a day, i.e., at 0000, 0600, 1200, and 1800 UTC. Model output at analysis time and a 6-hr forecast are available from the National Oceanic and Atmospheric Administration (NOAA) National Operational Model Archive & Distribution System (NOMADS; http://nomads.ncdc.noaa.gov/) server. After postprocessing of the NCEP GDAS, data from spectral coefficient form convert to 1° latitude-longitude (360 × 181) grids and from sigma levels to mandatory pressure levels. It uses three-dimensional variational (3D-Var) data assimilation method for data assimilation. Details of the GDAS analysis are described by Kalnay et al. (1996).

3.3. TRMM 3B42 Rainfall

The TRMM is a joint US-Japan satellite mission to monitor tropical and subtropical precipitation. It was launched in year 1997 into a near circular orbit approximately at 350 km altitude (raised to 403 km since 2001) at 35° inclinations from the equatorial plane. The complete description of the sensor package of TRMM is given by Kummerow et al. (1998). The operational TRMM data set used in the present study is TRMM 3B42, which is a merged product from Geostationary InfraRed (IR) and Microwave data (Huffman et al., 2003, 2007). The TRMM 3B42 estimates are produced in four stages: (1) the microwave precipitation estimates are calibrated and combined; (2) infrared precipitation estimates are created using the calibrated microwave precipitation; (3) the microwave and IR estimates are combined; and (4) rescaling to monthly data is applied. This rainfall product has been downloaded from TRMM Online Visualization and Archive System (TOVAS) at spatial resolution of 0.25° latitude/longitude.

4. Results and Discussions

The two different sets of experiments are performed in this study during 1–10 August 2015. The collection of unweighted particles is considered as “CNT runs”, and collection of particles with SIR is considered as “EXP runs”, in which IMD-observed rainfall is used to select appropriate model physics and resample high weight
particles using dynamic variable perturbations using SKEBS method. These particle-filtering steps are performed during 1–9 August 2015, and selected model physics options on 9 August 2015 are used for forecast verification on 10 August 2015. The choice of the NCEP GDAS analysis as input is mainly to assess the superiority of rainfall assimilation using particle filter over CNT runs, where initial conditions are taken from the NCEP GDAS analysis, which assimilated various kind of observations and one of the final analysis available from the NCEP. In verification step, the WRF model predicted accumulated rainfall, initialized from selected model physics, and validated against TRMM 3B42 rainfall valid for same time. The surface and vertical profiles of temperature, moisture, and wind speed forecasts are verified against NCEP final analysis.

The mean difference (bias), RMSD, and rainfall verification scores are used as standard parameters for statistical evaluation. Various rainfall verification scores based on contingency table (Table 1), viz., ETS, extreme dependency score (EDS), probability of detection (POD), and false alarm rate (FAR) over a wide range of rainfall thresholds (1–80 mm/day), are used to measure the impact of rainfall assimilation on rainfall predictions for grid wise evaluation (Bhomia et al., 2019). The POD measures the fraction of observed events that were correctly diagnosed and is sometimes called the “hit rate.” The FAR gives the fraction of diagnosed events that were actually nonevents. Perfect values for these scores are POD = 1 and FAR = 0. The ETS was formulated to account for the hits that would occur purely due to random chance. The ETS, though not a true skill score, is often interpreted that way since it has a value of 1 for perfect correspondence and 0 for no skill. It penalizes misses and false alarms equally, and for this reason, it is commonly used in the NWP rainfall verification (Bhomia et al., 2019). The new score EDS is used mainly for determining skill at higher value of rainfall thresholds. This score has the advantage that it can converge to different values for different forecasting systems and furthermore it does not explicitly depend upon the bias of the forecasting system (Bhomia et al., 2019; Stephenson et al., 2008).

![Equation 5](image)

![Equation 6](image)

![Equation 7](image)

![Equation 8](image)

where $e = \frac{(a + b)(a + c)}{(a + b + c + d)}$ refers to the expected number of correct forecasts above a rain threshold with a random forecast.

Figure 2 shows mean (line) and median (dash line) values of POD (Equation 7) and FAR (Equation 8) for CNT (blue) and EXP (red) runs. The POD for CNT runs is shown as light blue lines valid on 10 August 2015. The POD for EXP runs that assimilate IMD-observed rainfall using particle filter is shown as light red lines. Figure 2a shows that slightly more mean POD is found in EXP compared to CNT for less rainfall threshold. This positive impact of rainfall assimilation is more for high rainfall threshold (>35 mm/day). It suggests that rainfall assimilation using particle filter improves the skill of rainfall forecast for heavy rainfall. It is also important to mention here that slightly less value of median POD is found for high rainfall.

| Observation ≥ threshold | Observation < threshold |
|------------------------|------------------------|
| Forecast ≥ threshold   | $a = \text{Hits}$       | $b = \text{False alarms}$ |
| Forecast < threshold   | $c = \text{Misses}$     | $d = \text{Correct rejections}$ |
threshold (>40 mm/day), which indicates that forecasts from CNT show reduced skill for heavy rainfall. This large positive impact in rainfall prediction is due to resampling steps in the particle filter. In the resampling step, only those particles are selected that are able to capture the target pdf with the help of minimum variance and highest ETS values. Therefore, model predictions are approaches to particles that are closer to observations and able to capture high rainfall values more accurately. Moreover, the median value of POD for EXP runs is slightly higher than mean POD for high rainfall threshold. It is also found that different model physics are predicting similar values of low rainfall (less spread for low rainfall threshold), whereas this distribution is more for high rainfall threshold. Moreover, most of the particles are predicting better POD values after rainfall assimilation (light red lines) compared to CNT run (light blue lines). Similar to POD, mean and median values of FAR show less number of false alarms in rainfall assimilation experiments (EXP) compared to CNT experiments. Better mean FAR score is seen for high rainfall threshold compared to low rainfall threshold which shows that all particles are able to predict low rainfall precisely, and large uncertainty are seen for heavy rainfall. Overall, these results show that assimilation of IMD-observed rainfall using particle filter improves rainfall predictions for higher rainfall values. The accurate prediction of heavy rainfall has large societal benefits that obtained with the help of particle filter over CNT runs.

The ETS (Equation 5) is one of the most widely used skill scores for rainfall verification, and EDS (Equation 6) score is normally used for high rainfall threshold (Stephenson et al., 2008). Figure 3 shows mean and median value of ETS and EDS rainfall verification score for CNT and EXP runs. Figure 3a shows that the skill of rainfall prediction is improved for low and high rainfall thresholds after rainfall assimilation. It is also seen that ETS predicted from the WRF model is ~0.35 for low rainfall threshold that represents a very high skill of prediction. This high skill score is mainly due to initialization of the WRF model from the NCEP final analysis (best state) which is generally not a situation in operational weather forecasts. Moreover, it is noticed from Figure 3a that model predictions have more uncertainty for high rainfall values, and after rainfall assimilation skill of the rainfall forecasts are improved for high rainfall threshold (>40 mm/day). Similar to ETS, the value of EDS rainfall score is also improved after rainfall assimilation. Around 0.2 value of EDS is found for >40 mm/day rainfall threshold, whereas value of ETS is less than 0.1 for the same rainfall threshold. It is important to note here that these improvements in rainfall prediction are over CNT runs that are performed using NCEP GDAS analysis as initial condition. Large number of satellite and conventional observations are assimilated in this analysis. Therefore, observed advances in EXP runs after rainfall assimilation have noteworthy improvement over CNT runs.

Results discussed above support that assimilation of IMD-observed rainfall using particle filter improved rainfall forecasts compared to CNT runs. The further interest is to evaluate the impact of assimilation in subspace (here rainfall) on the prediction of other model subspaces (like temperature, moisture, and winds) due to nonlinear coupling of rainfall with these meteorological parameters. The WRF model-predicted temperature, moisture, and wind speed are verified against the NCEP GDAS analysis valid at same time. Figure 4 shows RMSD in 2-m air temperature and water vapor mixing ratio (WVMR), and 10-m wind speed for CNT and EXP runs. Figure 4a shows that assimilation of rainfall improved temperature forecasts for all forecast lengths (at 3-hr intervals on 10
August 2015), except 9-hr (a local maximum temperature occurred at this time) forecasts. Moreover, some particles having large RMSD in CNT runs are rejected in EXP runs after rainfall assimilation. Similar kinds of positive impact can be seen in WVMR (Figure 4b) and wind speed (Figure 4c) forecasts. It is important to mention here that assimilation of IMD-observed rainfall improved other basic meteorological parameters. These findings are similar to variational methods in which due to multivariate nature of data assimilation; assimilation of particular control parameters also modifies other control parameters (Kumar et al., 2014; Kumar & Shukla, 2019). Overall, we found that rainfall assimilation using particle filter improved surface temperature, WVMR, and wind speed forecasts. Further, we want to focus on these improvements over CNT runs where the WRF model is initialized from the NCEP final analysis that assimilated all kinds of observations including late arriving observations to prepare final analysis.

Further, vertical profiles of 24-h temperature, WVMR, and wind speed forecasts valid on 10 August 2015 are verified against NCEP GDAS analysis valid at same time (Figure 5). Results suggest that assimilation of IMD rainfall in EXP runs improves temperature profile (Figure 5a) at different vertical levels compared to CNT runs. Slightly higher positive impact can be seen at upper levels (above 300 hPa) in EXP runs. Mixed impact is found in WVMR profile from surface to 900 hPa (Figure 5b) and depicts improved prediction of WVMR above 900 hPa in EXP runs. Assimilation of IMD rainfall also improves vertical profile of wind speed with maximum improvements at mid and upper vertical levels.

Figures 6a–6c show the spatial distribution of rainfall from TRMM 3B42 merge-rainfall product, accumulated 24-h rainfall predictions from the CNT and EXP runs, respectively. Figure shows that both CNT (Figure 6b) and EXP (Figure 6c) runs are able to capture rainfall distribution over the Equatorial Indian Ocean (EIO) region, with less in magnitude as compared to TRMM rainfall (Figure 6a) product. Large
distribution of rainfall is seen over the Western Ghats, Bay of Bengal regions in CNT predictions that improved after rainfall assimilation in the EXP runs and more closer to TRMM rainfall. Both the model simulations are not able to capture the spatial distribution of rainfall over central India and in the foothills of the Himalaya with little improvement in EXP runs. To understand further the impact of rainfall assimilation, an impact parameter has been used to depict improvement or degradation in the rainfall prediction from EXP runs over CNT runs. The positive (negative) value of impact parameter shows improvement (degradation) of the rainfall assimilation using particle filter in the EXP runs over CNT runs. The spatial distribution of the impact parameter is defined as

\[ \text{Impact Parameter} = |\text{Rain}_{\text{CNT}} - \text{Rain}_{\text{TRMM}}| - |\text{Rain}_{\text{EXP}} - \text{Rain}_{\text{TRMM}}| \]

where \( \text{Rain}_{\text{CNT}}, \text{Rain}_{\text{EXP}}, \) and \( \text{Rain}_{\text{TRMM}} \) represent the rainfall from CNT runs, EXP runs, and TRMM 3B42 rainfall, respectively valid for the same period on 10 August 2015. Figure 6d suggests large improvement over the Western Ghats, Bay of Bengal, and EIO regions. However, few pockets of degradation are also seen that represent negative impact of rainfall assimilation using particle filter. However, the domain average value of impact parameter is positive that represents positive impact of rainfall assimilation using particle filter method.

Overall, these preliminary results suggest that assimilation of rainfall using particle filter improves prediction of basic meteorological parameters (like temperature, moisture, and winds) at surface and vertically. These improvements in basic meteorological parameters are mainly due to rainfall that is indirectly coupled with these basic parameters. Generally, in most of the previous rainfall assimilation studies (Kumar & Varma, 2016 and references therein), major objectives are to improve initial model states (like temperature, moisture, and winds) using rainfall observation either through indirect (like latent heat nudging and 1D + 4D-Var) or direct (4D-Var, LETKF) assimilation of rainfall. But in this study, the particle filter method is used to select appropriate model physics with perturbation in dynamic variables in model physics using
IMD rainfall observations, whereas no changes are performed in the initial model state like traditional methods. Moreover, rainfall subspace is indirectly coupled with other model subspaces (like temperature, moisture, and winds), so any modification in rainfall subspace changes other subspaces also in forecasts. Another important point to note that less distribution is observed in EXP runs compared to CNT runs in short forecasts (mainly 3-hr forecasts; Figure 4). Since all particles are initialized from the same model state (here NCEP GDAS for initial state), the differences are mainly due to selection of model physics and dynamic variable perturbation in model physics using SKEBS. Therefore, in short-range forecasts (3 hr), not all selected particles are able to represent true pdf, and it is the step where this filter may not be able to produce appropriate pdf. The possible solution may be to use local particle filter (Poterjoy, 2015) or equivalent-weight particle filter (Ades & Van Leeuwen, 2015; Browne, 2016), which considers proposal density to generate distribution of initial state in place of deterministic initial state opt in present study.

5. Conclusion

In this study, IMD-observed rainfall is assimilated using particle filter, a nonlinear filter that takes care of non-Gaussian nature of rainfall observations. Two different sets of experiments are performed with and without rainfall assimilation using different model physics options during 1–10 August 2015. Particle filter is implemented in rainfall subspace (having less dimension) compared to full high-dimensional model space with multiple hypotheses (based on less variance and large value of mean ETS) to produce new particles in resampling steps. Rainfall is one of diagnostic parameters from the weather model that non-linearly depends on various parameters (like initial model state, terrestrial data, model physics, and dynamical variables into parameterizations). Furthermore, the dynamic variable perturbation through the SKEBS method is used to generate new particles from high weight particles such that the total number of particles (here 90) should be the same. Kumar and Shukla (2019) also used this approach to change two dynamical parameters in guided particle filter to assimilate satellite measurements. The use of the SKEBS method to generate new particles in resampling steps provides additional guidance to the particle toward future observations. Results based on different rainfall verification scores suggest that skill of rainfall forecast is improved with the assimilation of rainfall using particle filter compared to CNT runs. Moreover, rainfall assimilation also improves temperature, WVMR, and wind speed forecasts at surface and different vertical levels. These results support that implementation of rainfall assimilation using particle filter, which considers nonlinear and non-Gaussian distribution, improves prediction from the WRF model. In the case of the EnKF, the same configuration of physics parameterizations is kept for each ensemble member and all that changes at analysis time is the model state x_{p,k} itself. In this study, particles are targeting best-suited model physics with assimilation of rainfall.

In present study, all particles are initialized from the same model state (here NCEP GDAS analysis), and the differences are mainly due to selection of model physics and dynamic variable perturbation in model physics using SKEBS. Therefore, in short-range forecasts (3-hr forecast; Figure 4), no selected particles are able to represent true pdf, and it is the step where this filter may not be able to produce appropriate pdf. The possible solution may be to use local particle filter (Poterjoy, 2015) or equivalent-weight particle filter (Ades & Van Leeuwen, 2015; Browne, 2016), which considers proposal density to generate distribution of initial state in place of deterministic initial state. The objective of this study is to understand the role of model physics in an imperfect model using initial state (best) from the NCEP GDAS analysis. This work motivates to use equivalent-weight particle filter proposed by Ades and Van Leeuwen (2015) for a high-dimensional nonlinear weather model to produce distribution of initial model states and further select the appropriate model physics for imperfect (weak) model using particle filter and develop an “efficient particle filter” for the NWP model. This may be a scope for future research in the fast developing field of nonlinear data assimilation.

Data Availability Statement

IMD-gridded daily rainfall observations are available online (http://www.imd.gov.in/ncdc_new/Request.html). The NCEP GFS 0.25 Degree Global Forecast are downloaded from CISL-RDA (https://rda.ucar.edu/datasets/ds084.1/), and three-hourly rainfall product from TRMM Multi-satellite Precipitation Analysis (TRMM Multi-satellite Precipitation Analysis) are available online (https://gpm.nasa.gov/data-access/downloads/trmm).
Earth and Space Science

References

Aides, M., & Van Leeuwen, P. J. (2015). The equivalent-weights particle filter in a high dimensional system. Quarterly Journal of the Royal Meteorological Society, 141(687), 484–503. https://doi.org/10.1002/qj.2370

Attada, R., Parekh, A., Chowdary, J. S., & Gnanaseelan, C. (2018). Reanalysis of the Indian summer monsoon: Four dimensional data assimilation of AIRS retrievals in a regional data assimilation and modeling framework. Climate Dynamics, 50(7–8), 2905–2923. https://doi.org/10.1007/s00382-017-3781-z

Attada, R., Parekh, A., Kumar, P., & Gnanaseelan, C. (2015). Evaluation of the impact of AIRS profiles on prediction of Indian summer monsoon using WRF variational data assimilation system. Journal of Geophysical Research: Atmospheres, 120, 8112–8131. https://doi.org/10.1002/2014JD023024

Bauer, P., Auligne, T., Bell, W., Geer, A., Guichard, V., Heilliette, S., et al. (2011). Satellite cloud and precipitation assimilation at operational NWP centres. Quarterly Journal of the Royal Meteorological Society, 137(661), 1934–1951. https://doi.org/10.1002/qj.905

Berner, J., Ha, S. Y., Hacker, J. F., Fournier, A., & Snyder, C. (2011). Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. Monthly Weather Review, 139(5), 1972–1995. https://doi.org/10.1175/2010MWR3595.1

Berner, J., Shutts, G. J., Leutbecher, M., & Palmer, T. N. (2009). A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. Journal of Atmospheric and Oceanic Technology, 26(5), 630–643. https://doi.org/10.1175/2008JTEchafl.1

Bhomia, S., Kumar, P., & Kishitawal, C. M. (2019). Evaluation of the Weather Research and Forecasting model forecasts for Indian summer monsoon rainfall of 2014 using ground based observations. Asia-Pacific Journal of Atmospheric Sciences, 51, 124–134.

Brown, P. A. (2016). A comparison of the equivalent weights particle filter and the local ensemble transform Kalman filter in application to the barotropic vorticity equation. Tellus A: Dynamic Meteorology and Oceanography, 68(1), 3046.

Chen, Z. (2003). Bayesian line nonlinear/3D-4D variational analysis and forecast. Monthly Weather Review, 131(8), 2094–2109. https://doi.org/10.1175/1520-0493(2003)131<2094:VCAPCT>2.0.CO;2

Doucet, A., De Freitas, N., & Gordon, N. (2001). Sequential Monte-Carlo methods in practice. New York: Springer-Verlag. https://doi.org/10.1007/978-1-4757-3437-9

Dubuisson, S. (2015). Tracking with particle filter for high-dimensional observation and state spaces. Hoboken, NJ: John Wiley & Sons. https://doi.org/10.1002/9781119004868

Errico, R. M., Bauer, P., & Mahfouf, J. F. (2007). Issues regarding the assimilation of cloud and precipitation data. Journal of the Atmospheric Sciences, 64(11), 3785–3798. https://doi.org/10.1175/2006JAS2044.1

Hou, A. Y., Zhang, S. Q., & Reale, O. (2004). Variational continuous assimilation of TMI and SSM/I rain rates: Impact on GEOS–3 hurricane analysis and forecast. Monthly Weather Review, 132(8), 2094–2109. https://doi.org/10.1175/1520-0493(2004)132<2094:VCAOTA>2.0.CO;2

Huffman, G. J., Adler, R. F., Stocker, E. F., Bolvin, D. T., Nelkin, E. J. (2003). Analysis of TRMM 3-hourly multi-satellite precipitation estimates computed in both real and post-real time. 83rd AMS Annual Meeting. Paper P4.11 in 12th Conf. on Sat. Meteor. and Oceanogr., 9–13 Feb. 2003, Long Beach, CA, pp. 6.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., et al. (2007). The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. Journal of Hydrometeorology, 8(1), 38–55. https://doi.org/10.1175/JHM560.1

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The NCEP/NCAR 40 year reanalysis project. Bulletin of the American Meteorological Society, 77(3), 437–471. https://doi.org/10.1175/1520-0477(1996)077<0437:TYNRP2.0.CO;2

Kalnay, M. (1989). Description of the NMC global data assimilation and forecast system. Weather and Forecasting, 4(3), 335–342. https://doi.org/10.1175/1520-0434(1989)004<0335:DOTNGD>2.0.CO;2

Kotsuki, S., Miyoshi, T., Terasaki, K., Lien, G. Y., & Kalnay, E. (2017). Assimilating the global satellite mapping of precipitation data with the Nonhydrostatic Icosahedral Atmospheric Model (NIMAM). Journal of Geophysical Research: Atmospheres, 122, 631–650. https://doi.org/10.1002/2016JD025535

Kumar, P., & Kishitawal, C. M. (2017). Importance of satellite-retrieved rain/no-rain information on short-range weather predictions. International Journal of Remote Sensing, 38(13), 3851–3864. https://doi.org/10.1080/01431161.2017.1306140

Kumar, P., Kishitawal, C. M., & Pal, P. K. (2014). Impact of satellite rainfall assimilation on Weather Research and Forecasting model predictions over the Indian region. Journal of Geophysical Research: Atmospheres, 119, 2017–2031. https://doi.org/10.1002/2013JD020005

Kumar, P., & Shukla, M. V. (2019). Assimilating INSAT-3D thermal infrared window imager observation with the particle filter: A case study for Vardah Cyclone. Journal of Geophysical Research: Atmospheres, 124, 1897–1911. https://doi.org/10.1029/2018JD028827

Kumar, P., & Varma, A. K. (2016). Assimilation of INSAT-3D hydro-estimator method retrieved rainfall for short-range weather prediction. Quarterly Journal of the Royal Meteorological Society, 143(702), 384–394. https://doi.org/10.1002/qj.2929

Kummerow, C., Barnes, W., Kozu, T., Shiue, J., & Simpson, J. (1998). The Tropical Rainfall Measuring Mission (TRMM) sensor package. Journal of Atmospheric and Oceanic Technology, 15(3), 809–817. https://doi.org/10.1175/1520-0426(1998)015<0809:TTRMMT>2.0.CO;2

Lien, G. Y., Kalnay, E., & Miyoshi, T. (2013). Effective assimilation of global precipitation: Simulation experiments. Tellus A, 65(1), 19915. https://doi.org/10.3402/tellusa.v65.19915

Lien, G. Y., Kalnay, E., Miyoshi, T., & Huffman, G. J. (2016). Statistical properties of global precipitation in the NCEP GFS model and TMPA observations for data assimilation. Monthly Weather Review, 144(2), 663–679. https://doi.org/10.1175/MWR-D-15-0150.1

Lien, G. Y., Miyoshi, T., & Kalnay, E. (2016). Assimilation of TRMM multisatellite precipitation analysis with a low-resolution NCEP global forecasting system. Monthly Weather Review, 144(2), 643–661. https://doi.org/10.1175/MWR-D-15-0149.1

Lopez, P. (2011). Direct 4D-Var assimilation of NWP stage IV radar and gauge precipitation data at ECMWF. Monthly Weather Review, 139(7), 2098–2116. https://doi.org/10.1175/2010MWR3565.1

Marceal, V., & Mahfouf, J. F. (2003). Experiments on 4D-Var assimilation of rainfall data using an incremental formulation. Quarterly Journal of the Royal Meteorological Society, 129, 3137–3160.

Maskell, S., & Gordon, N. (2001). A tutorial on particle filters for on-line nonlinear/non-Gaussian Bayesian tracking. IEEE Transactions on Signal Processing, 50(2), 174–188.

Mattern, J. P., Dowd, M., & Fennel, K. (2013). Particle filter-based data assimilation for a three-dimensional biological ocean model and satellite observations. Journal of Geophysical Research: Oceans, 118, 2746–2760. https://doi.org/10.1002/jgrc.20213

KUMAR

Acknowledgments

The author is thankful to the Director, Space Applications Centre, ISRO, India. Author acknowledges IMD-gridded daily rainfall observations.
Pai, D. S., Sridhar, L., Rajeevan, M., Sreejith, O. P., Sarthai, N. S., & Mukhopadhyay, B. (2014). 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.

Posselt, D. J., & Bishop, C. H. (2012). Nonlinear parameter estimation: Comparison of an ensemble Kalman smoother with a Markov chain Monte Carlo algorithm. *Monthly Weather Review*, 140(6), 1957–1974. https://doi.org/10.1175/MWR-D-11-0042.1

Posselt, D. J., Hodyss, D., & Bishop, C. H. (2014). Errors in ensemble Kalman smoother estimates of cloud microphysical parameters. *Monthly Weather Review*, 142(4), 1631–1654. https://doi.org/10.1175/MWR-D-13-00290.1

Poterjoy, J. (2015). A localized particle filter for high-dimensional nonlinear systems. *Monthly Weather Review*, 144, 59–76.

Ratheesh, S., Chakraborty, A., Sharma, R., & Basu, S. (2016). Assimilation of satellite chlorophyll measurements into a coupled biophysical model of the Indian Ocean with a guided particle filter. *Remote Sensing Letters*, 7(5), 446–455. https://doi.org/10.1080/2150704X.2016.1143985

Ristic, B., Arulampalam, S., & Gordon, N. (2004). *Beyond the Kalman filter: Particle filters for tracking applications*. London, UK: Artech House.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., et al. (2008). A description of the advanced research WRF version 3, Tech. Note NCAR/TN-475 STR. pp. 113, Mesoscale and Microscale Meteorology Division, National Center of Atmospheric Research, June 2008.

Stephenson, D., Casati, B., Ferro, C. A. T., & Wilson, C. A. (2008). The extreme dependency score: A non-vanishing measure for forecasts of rare events. *Meteorological Applications*, 15(1), 41–50. https://doi.org/10.1002/met.53

Van Leeuwen, P. J. (2009). Particle filtering in geophysical systems. *Monthly Weather Review*, 137, 4089–4114.

Van Leeuwen, P. J. (2010). Nonlinear data assimilation in geosciences: An extremely efficient particle filter. *Quarterly Journal of the Royal Meteorological Society*, 136, 1991–1999.