Nowcasting with Data Assimilation: A Case of Global Satellite Mapping of Precipitation

SHIGENORI OTSUKA AND SHUNJI KOTSUKI
RIKEN Advanced Institute for Computational Science, Kobe, Japan

TAKEMASA MIYOSHI
RIKEN Advanced Institute for Computational Science, Kobe, Japan, and Department of Atmospheric and Oceanic Science, University of Maryland, College Park, College Park, Maryland, and Japan Agency for Marine–Earth Science and Technology, Yokohama, Japan

(Manuscript received 1 March 2016, in final form 13 June 2016)

ABSTRACT

Space–time extrapolation is a key technique in precipitation nowcasting. Motions of patterns are estimated using two or more consecutive images, and the patterns are extrapolated in space and time to obtain their future patterns. Applying space–time extrapolation to satellite-based global precipitation data will provide valuable information for regions where ground-based precipitation nowcasts are not available. However, this technique is sensitive to the accuracy of the motion vectors, and over the past few decades, previous studies have investigated methods for obtaining reliable motion vectors such as variational techniques. In this paper, an alternative approach applying data assimilation to precipitation nowcasting is proposed. A prototype extrapolation system is implemented with the local ensemble transform Kalman filter and is tested with the Japan Aerospace Exploration Agency’s Global Satellite Mapping of Precipitation (GSMaP) product. Data assimilation successfully improved the global precipitation nowcasting with the real-case GSMaP data.

1. Introduction

Precipitation is one of the most important variables in real-time monitoring and forecasting of the weather, aiding disaster prevention, agriculture, the economy, and many other applications. Over the last decade, satellite-based global precipitation measurements have become available (e.g., Joyce et al. 2004; Huffman et al. 2007; Kubota et al. 2007), and they are particularly valuable in regions where ground-based precipitation observations are sparse.

In the case of ground-based weather radar networks, space–time extrapolation is widely used for nowcasting (e.g., Rinehart and Garvey 1978; Dixon and Wiener 1993; Bowler et al. 2006). Assuming that precipitation intensity does not change over time (Lagrangian persistence), motions of radar echoes are estimated either from consecutive radar images or from other data sources, such as Doppler radar winds, and future positions of radar echoes are forecasted. This is a simple yet powerful method for predicting the nearest future and has been used in operations at many centers around the world. The same framework can be applied to satellite-based precipitation estimates. In fact, satellite-based global precipitation mapping products also adopt space–time extrapolation in their estimation algorithms to fill in the observation gaps by using motion vectors derived from geostationary satellite cloud infrared images (Joyce et al. 2004; Kubota et al. 2007).

The advantage of satellite-based precipitation estimates is in its global availability. Even if no ground-based weather-radar network is available in a region, satellite-based global precipitation estimates can be used for precipitation nowcasting. The main purpose of the present study is to demonstrate global precipitation nowcasting using satellite-based global precipitation maps.
Noise reduction of motion vectors is a key to accurate space–time extrapolation. Although there are various techniques available to obtain more reliable motion vectors (e.g., Laroche and Zawadzki 1994; Li et al. 1995; Bowler et al. 2004), there is still room for improvement. Probabilistic nowcasting with perturbed motion vectors better handles the uncertainties of the motion vectors (e.g., Bowler et al. 2006). Here, we propose a sophisticated approach to dealing with the uncertainties of the motion vectors for better estimation: ensemble-based data assimilation.

Data assimilation optimally combines model forecasts and observations for better state estimation and has been used for numerical weather prediction (NWP) for decades. In particular, ensemble-based data assimilation methods such as the ensemble Kalman filter (EnKF; Evensen 1994) account for flow-dependent error covariance structures. Figure 1 shows a general flowchart of the data assimilation process. Ensemble spread represents forecast confidence and is used to optimally combine forecasts and observations. Probabilistic information may also be beneficial for end users. In this paper, we take advantage of the model-independent nature of EnKF and apply the same framework to the space–time extrapolation. We use a simple advection equation instead of the full set of equations used in physically based NWP. Input into the assimilation system (“observations”) is also limited to the motion vectors of precipitation patterns rather than meteorological variables such as temperature, moisture, and pressure.

This study uses the Global Satellite Mapping of Precipitation (GSMaP) dataset (Kubota et al. 2007) developed by the Japan Aerospace Exploration Agency (JAXA). GSMaP provides global precipitation estimates based on satellite microwave and infrared observations. The spatial resolution is 0.1° × 0.1° covering 60°S–60°N, and the temporal resolution is 1 h. The near-real-time (NRT) version of the GSMaP product (Ushio and Kachi 2009) becomes available 4 h after the observation time. By applying space–time extrapolation using motion vectors computed from consecutive GSMaP_NRT images, experimental short-lead-time global precipitation nowcasts are constructed.

This paper is organized as follows. Section 2 describes the methods used and section 3 explains the experimental setup. Section 4 presents the results and section 5 features a discussion of the findings. Finally, conclusions are drawn and presented in section 6.

2. Method
   a. Space–time extrapolation

   In this study, we take advantage of an existing space–time extrapolation system developed by Otsuka et al. (2016). The motion vector is computed by the Tracking Radar Echoes by Correlation (TREC; Rinehart and Garvey 1978) with the fractional motion vector technique of Otsuka et al. (2016). The precipitation field is advected by the fifth-order weighted essentially non-oscillatory scheme (WENO; Liu et al. 1994), and the motion vector field is advected by the first-order upwind scheme. The time integration is performed by the second-order Adams–Bashforth scheme. Although the original GSMaP employs a Kalman filtering approach to estimating Lagrangian growth and decay of the precipitation to fill in the gaps of microwave satellite observations available within an hour (Ushio et al. 2009), we assume Lagrangian persistence so that the impact of data assimilation on the accuracy of the motion vectors is highlighted.

   The existing extrapolation system is modified for use in a data assimilation system; the divergence damping term [Eq. (30) in Skamarock and Klemp (1992)] is added to the advection equations of the motion vectors. In a simple advection forecast, unrealistic shock-wave-like convergence lines appear if a spatially varying motion vector field is advected continuously in data assimilation cycling experiments. This leads to a failure of the system. However, the divergence component is also necessary to represent apparent motions of precipitation areas. Hence, in this study, the divergence damping term is used instead of the nondivergent constraint such as the Continuity of Tracking Radar Echoes by Correlation vectors (COTREC; Li et al. 1995).

   b. Data assimilation

   In this study, data assimilation is used to improve the TREC-estimated motion vectors, and the precipitation field is not directly used for data assimilation. Here, the local ensemble transform Kalman filter (LETKF; Hunt et al. 2007) is used. This study adopts the existing
LETKF code as outlined by Miyoshi (2005) as well as follow-on studies (e.g., Miyoshi and Yamane 2007; Miyoshi et al. 2007). The state vector is a complete, gridded set of motion vectors. The relaxation to prior spread (RTPS) method (Whitaker and Hamill 2012) is used to inflate the background error covariance. After applying RTPS, the ensemble spread is also relaxed to a prescribed constant $C$ in a similar manner as RTPS if the spread is smaller than $C$ [relaxation to constant spread (RTCS)]. Formally, ${x}_{0} = {x}_{0} + \lambda_p \sigma_{a} + 1$, where $x_0$ denotes the analysis ensemble perturbation of variable $x$; $\lambda_p$ and $\lambda_c$ are the inflation factors of RTPS and RTCS, respectively, and are given as

$$\lambda_p = \alpha_p \frac{\sigma_b - \sigma_a}{\sigma_a} + 1,$$

$$\lambda_c = \begin{cases} \frac{C - \lambda_p \sigma_a}{\lambda_p \sigma_a} + 1 & \text{(for } C > \lambda_p \sigma_a) \), \\ 1 & \text{(for } C \leq \lambda_p \sigma_a). \end{cases}$$

Here, $\sigma$ denotes the ensemble spread of $x$ and the subscripts $b$ and $a$ correspond to the first guess and the analysis, respectively; $\alpha_p$ and $\alpha_c$ represent the relaxation parameters for RTPS and RTCS, respectively. RTCS is introduced because the ensemble becomes underdispersive even if RTPS is applied. Related discussion is provided in section 5. Figure 2 shows the flowchart of the data assimilation system for the space–time extrapolation with examples of actual input/output data.

In principle, the TREC motion vectors are defined within the precipitating areas. Data assimilation can fill in the missing motion vectors in the no-precipitation areas using prior information from the previous forecasts. This is advantageous because motion vectors are needed at every grid point for space–time extrapolation. Data assimilation is also robust for noisy motion vectors because it optimally combines prior information and “observed” motions by considering error statistics.

3. Experimental setup

The input to the system is GSMaP_NRT (hereafter NRT). NRT data between 0000 UTC 13 July and 0000 UTC 13 August 2013 totaling 745 images are used, so that we have 744 initial conditions for nowcasting. NRT images between 0100 and 2000 UTC 13 August are additionally used for verification. The domain size is $3600 \times 1200$ grid points from 60°S to 60°N with a grid interval of
The TREC motion vectors are computed at each grid point using two consecutive NRT images with a 1-h interval, and the cross correlation is computed using a circular patch with a diameter of 46 grid points. The maximum feature velocity is limited to six grid points, roughly corresponding to 16.7 m s\(^{-1}\) at the equator. This might be too small in some cases, but optimization is beyond the scope of this study. Motion vectors with negative correlation coefficients are not used.

We set up two series of experiments: nowcasting with data assimilation (hereafter LETKF) and without data assimilation (NoDA). In the LETKF experiments, the TREC vectors are treated as observations and assimilated every hour. We simply assign the time of the latest NRT image as the observation time of the TREC vectors, rather than a time between the two NRT images. The number of TREC vectors is reduced by taking one out of 2 \(\times\) 2 grid points to avoid problems caused by correlated observation errors. The LETKF ensemble size is fixed at 20, and the covariance localization function is a Gaussian function with a standard deviation of 100 km. The zonal and meridional components of the TREC vectors are assimilated independently, and the observation error of each component is prescribed as four grid points per hour.

The relaxation parameter in RTPS is chosen to be \(\alpha_p = 0.95\), whereas RTCS uses the constant \(C = 2.0\) grid points per hour and the relaxation parameter \(\alpha_c = 0.005\). These covariance inflation parameters are manually optimized so that the ensemble spread stabilizes in a reasonable range. The system is not very sensitive to the choice of these parameters.

NoDA experiments use spatially smoothed TREC motion vectors derived at each grid point. Because the TREC vectors are defined only within the precipitating areas, for NoDA experiments and the first cycle of the LETKF experiments, missing motion vectors are filled by minimizing \(\nabla^2 \mathbf{u}\) outside the TREC-defined areas, where \(\mathbf{u}\) denotes the motion vector. After filling the no-precipitation areas, spatial smoothing is applied using Lanczos filtering with a critical wavelength of 40 grid points to manually optimize the forecast performance. The initial ensemble for the first LETKF cycle is prepared as follows; the initial conditions are the same as those of NoDA but with different TREC patch diameters ranging between 21 and 40 grid points, and the critical wavelength of Lanczos filtering is chosen to be the same as the TREC patch diameter.

We perform 20-h extended forecasts from the NRT precipitation distributions using the advection model.
driven by the LETKF analysis ensemble-mean motion vectors or the NoDA motion vectors. At the south and north boundaries, motion vectors are fixed to the mean motion at those latitudes. The boundary conditions for the rain field are the constant gradient condition for the outflow regions and zero rain for the inflow regions. The extended forecasts are compared with NRT itself for verification. As a reference, Eulerian persistence forecasts are also computed.

4. Results

Figure 3 shows an example of nowcasting by NoDA and LETKF initiated at 0300 UTC 13 July 2013. In the NoDA run, motion vectors around 33°S, 138°E and 38°S, 144°E have spurious northward components (Fig. 3b), leading to large errors in the precipitation forecast (Fig. 3e). In the LETKF run, the initial motion vector field does not have such erroneous vectors (Fig. 3c), resulting in better agreement with NRT during the 4-h forecast (Fig. 3f).

Figure 4 shows another example, initiated at 0900 UTC 21 July 2013. In the NoDA run, spurious southward motions around 35°S, 111°W and northeastward motions around 40°S, 107°W distort the precipitation pattern after the 4-h time integration (Figs. 4b,e). In contrast, the LETKF run does not show such a distortion (Figs. 4c,f), and the position of the precipitation area in the 4-h forecast agrees well with that in NRT at the same validation time (Figs. 4d,f).

To verify the performance statistically, the threat scores of the forecast rain rates are computed between 0000 UTC 14 July and 0000 UTC 13 August, after the 1-day spinup from 0100 UTC 13 July to 0000 UTC 14 July. Figure 5 shows the mean threat scores as a function of the lead time for LETKF, NoDA, and the Eulerian persistence forecast over the extratropics (20°–60°N and 20°–60°S; Figs. 5a,b) and the tropics (20°S–20°N; Figs. 5c,d). LETKF outperforms NoDA for the two threshold values of 1 and 5 mm h⁻¹ at lead times of up to 20 h, and the improvements are statistically significant at the confidence level of higher than 99.99%. In the extratropics, both LETKF and NoDA show higher threat scores compared to the Eulerian persistence forecast, indicating that the space–time extrapolation shows prediction skill. However, in the tropics, the threat scores of LETKF and NoDA are closer to that of the Eulerian persistence forecasts, and the prediction skill is lower than that in the extratropics. This seems to be due to the nature of the less-organized tropical convection; the lifetimes of individual cumulus convection events are comparable or shorter than the 1-h time interval of NRT, and the Lagrangian persistence assumption may not be appropriate.
Figure 6 shows the threat score differences (LETKF – NoDA) for the threshold value of 1 mm h\(^{-1}\) in the extratropics as a function of time (abscissa) and lead time (ordinate). During the spinup period of about 1 day, the extended forecasts show some degradation (blue colors in Fig. 6) at longer lead times, but after the spinup period, LETKF outperforms NoDA for most of the times and lead times; LETKF almost always outperforms NoDA for lead times up to about 5 h (red colors in Fig. 6). After 5–10-h lead times, the LETKF forecasts sometimes show slight degradation against NoDA, but the majority still shows some improvement. Overall, the LETKF system runs stably and shows the improved forecasts during the 1-month period.

5. Discussion

As noted in section 1, the real-time nowcasting of GSMaP will be valuable in regions with sparse ground-based weather radar observations. In terms of real-time operations, the current system can finish one assimilation cycle in about 5 min using an Intel Xeon E5–4650v2 2.4-GHz 10 cores/CPU × four CPUs, including 1-h ensemble forecasts with 20 members, LETKF updates, and a single 20-h deterministic
forecast. Although the optimization of the ensemble size and forecast length is needed, the current system is feasible for the real-time nowcasting of global precipitation. In addition, NRT is available with a 4-h delay from the real time, and Fig. 5 indicates skillful forecasts beyond 4 h.

In the EnKF with a simple advection–diffusion equation, there are several differences compared to the standard NWPs or conventional space–time extrapolation systems. First, the spurious convergence lines must be suppressed during time integration of the advection equation. This happens because of continuous advection of spatially varying motion vectors in the analysis–forecast cycle. If the motion vector field is not advected, as in many other space–time extrapolation systems (e.g., Bowler et al. 2006), the spurious convergence lines will not appear. However, in that case, motion vectors will become inconsistent with the precipitation areas. This is not the best choice for continuous cycling of data assimilation.

Second, the ensemble spread must be inflated largely by RTPS and RTCS for stable data assimilation. This is because the use of a simple advection–diffusion equation introduces much larger model errors compared to the physically based NWP models and because the nonchaotic nature of the advection equation suppresses the ensemble spread during time integration. Underinflation may result in filter divergence, whereas overinflation may result in overfitting to noisy observations and numerical instabilities during the time integration. As the first trial of data assimilation for space–time extrapolation, RTCS is, for simplicity, employed in this study. However, there are more sophisticated approaches to handling model errors, such as adding stochastic errors during the model time integration (e.g., Buizza et al. 1999). Improving the representation of the model errors in ensemble prediction may further improve the analysis motion vectors.

6. Conclusions

Space–time extrapolation is widely used in many applications such as precipitation nowcasting. Applying space–time extrapolation to satellite-based global precipitation data can be valuable in regions where ground-based precipitation nowcasts are not available. However, space–time extrapolation suffers from the inaccuracy of motion vectors. In this study, data assimilation was applied to improve the accuracy of motion vectors. LETKF was used to combine present TREC-derived motion vectors with the first guess from the previous extrapolation, which included information from the past motion vector estimates. As in the test case, the GSMaP_NRT product was used to make short-range global precipitation nowcasts. The system with data assimilation outperformed that without data assimilation in terms of precipitation forecasting as a result of improved motion vectors.

The framework proposed in this paper is applicable to various space–time extrapolation methods, as well as to various types of image sequences. Integrating multiple data sources into the single data assimilation system may enhance the capability of space–time extrapolation. As a next step, assimilating an image sequence directly, without retrieving motion vectors using conventional methods in advance of data assimilation, may also allow us to design more flexible and reliable systems.

Acknowledgments. The authors thank the two anonymous reviewers for their constructive comments. SO developed the software and conducted the experiments, SK contributed to the data handling of GSMaP, and TM is the PI and directed the research. The core library of LETKF is available online (https://github.com/takemasa-miyoshi/letkf/). The GSMaP data are open to the public and available online (http://sharaku.eorc.jaxa.jp/GSMaP/). This work has been partially supported by the JAXA Precipitation Measuring Mission. The figures were produced by the GFD-DENNOU Library.

REFERENCES

Bowler, N. E. H., C. E. Pierce, and A. Seed, 2004: Development of a precipitation nowcasting algorithm based upon optical


