**Geophysical Research Letters**

**RESEARCH LETTER**

10.1029/2021GL094962

Tijana Janjić and Yuefei Zeng contributed equally to this study.

**Key Points:**
- A weakly constrained LETKF for mass conservation and non-negativity is introduced and examined in convective-scale data assimilation
- Combining both constraints results in the least biases in total mass of hydrometeors and in mass-flux divergence and vorticity in analyses
- Best forecasts are also achieved by the combination

**Supporting Information:**
Supporting Information may be found in the online version of this article.

**Correspondence to:**
Y. Zeng and T. Janjić, 
yuefei.zeng@lmu.de; 
tijana.pfander@lmu.de

**Citation:**
Janjić, T., & Zeng, Y. (2021). Weakly constrained LETKF for estimation of hydrometeor variables in convective-scale data assimilation. *Geophysical Research Letters*, 48, e2021GL094962. https://doi.org/10.1029/2021GL094962

Received 23 JUL 2021
Accepted 30 NOV 2021

**Author Contributions:**
- Conceptualization: Tijana Janjić
- Formal analysis: Yuefei Zeng
- Funding acquisition: Tijana Janjić
- Investigation: Yuefei Zeng
- Methodology: Tijana Janjić
- Project Administration: Tijana Janjić
- Supervision: Tijana Janjić
- Validation: Yuefei Zeng
- Visualization: Yuefei Zeng
- Writing – review & editing: Tijana Janjić, Yuefei Zeng

© 2021. The Authors. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

**Abstract**

The ensemble Kalman filter algorithm can produce negative values for non-negative variables. To mitigate this sign problem and to simultaneously maintain the mass conservation, a new concept of combining weak constraints on mass conservation and non-negativity has been introduced in this work, with a focus on hydrometeor variables in convective-scale data assimilation. We modify the local ensemble transform Kalman filter with weak constraints on mass conservation for each hydrometeor variable and adopt the assimilation of clear-air reflectivity data as a weak constraint on non-negativity. We examine the concept by a series of sensitivity experiments using an idealized setup. Results show that both weak constraints successfully improve the mass conservation property in analyses and both reduce the biased increase in integrated mass-flux divergence and vorticity. Furthermore, the least biased increase is obtained by combining both constraints, and the best forecasts are also achieved by the combination.

**Plain Language Summary**

Often physical properties of a system that we are modeling dictate plausible values of the initial conditions of our numerical models. Unfortunately, by using modern data assimilation techniques to obtain these initial conditions, physical property of non-negativity is frequently violated. On the other hand, algorithms that are able to preserve the non-negativity usually would break mass conservation. Here, we propose a fast, easy to implement modification of the existing algorithm (local ensemble transform Kalman filter) that is able to weakly preserve both properties of mass conservation and non-negativity. In idealized experiments that assimilate radar data in non-hydrostatic, convection-permitting numerical model and update hydrometeor values, we show the benefit of the proposed approach on prediction of atmospheric water variables.

1. Introduction

Weather radar observations are available in very high spatial and temporal resolution. In severe convective weather situations, they provide us valuable information on developments of convective systems. To obtain accurate initial conditions for prediction of convective events, data assimilation algorithms are usually applied, which combine those observations with our knowledge of atmospheric physics and dynamics described by the numerical weather prediction (NWP) models. Although the NWP models can reproduce highly nonlinear physics and dynamics, in short runs, they are very sensitive to initial and boundary conditions (Kühnlein et al., 2014). Therefore, the performance of data assimilation algorithms is key for making skillful predictions of convective storms.

Pioneering work on the use of ensemble Kalman filters (EnKFs, Evensen, 2003) in idealized settings for radar data assimilation has started with the work of Snyder and Zhang (2003), Zhang et al. (2004), Tong and Xue (2005), and real data experiments were performed by Dowell et al. (2004). These authors quickly recognized the potential of EnKF for convective scale data assimilation. The EnKF technique can handle complex, nonlinear physical processes without a need for the adjoint of the model or simplification of the model physics, with the possibility of including hydrometeors in the state vector and easily initializing probabilistic forecasts. Since then, the EnKF has been widely applied in convective-scale data assimilation using actual observations and updating cloud and precipitating hydrometeors (Dowell et al., 2011; Minamide & Zhang, 2019; Zeng et al., 2018, 2019, 2020). Gaussian assumptions in EnKF, however, lead to non-optimal filters, for example, resulting in significant unrealistic changes in mass, divergence, and vorticity (Janjić et al., 2014; Zeng & Janjić, 2016; Zeng, de Lozar, et al., 2021), especially when assimilating radar data frequently. Proper estimates of hydrometeor variables are important for the prediction of convective storms (Sun et al., 2014), they are crucial in the evolution of model atmosphere and can also influence many of the simulated observations (Bannister et al., 2020). Therefore, it is essential to update them in state vector jointly with other prognostic variables. But, there are still challenges...
that remain unresolved in case prognostic variables of the microphysics scheme are updated with EnKF (Janjić et al., 2021). For instance, due to sparse observations, Gaussian assumption, and other approximations in these methods, the analysis ensemble members are prone to be associated with negative hydrometeor variables, which is physically impossible. In current practice, one way to deal with this issue is to execute physical consistency checks after the analysis step, which clips the negative values to zero for each ensemble member (e.g., Gustafsson et al., 2018). However, this concept inevitably causes a biased increase in the total mass of hydrometeor variables, as shown in Janjić et al. (2014) and Zeng, Janjić, et al. (2021). An alternative approach is to transform the variables by assuming that they follow pre-specified non-Gaussian distributions (Cohn, 1997; Lauvernet et al., 2009; Simon & Bertino, 2009); for example, Lien et al. (2013) applied Gaussian anamorphosis on the assimilation of precipitation with the local ensemble transform Kalman filter (LETKF, Hunt et al., 2007). These methods may avoid the sign problem but do not account for mass conservation, as shown in Janjić et al. (2014) with a simple model. Yet another alternative approach, proposed by Janjic et al. (2014), incorporates physical constraints in the data assimilation algorithm for convective scale data assimilation. To this end, a Quadratic Programming Ensemble Kalman filter (QPEns, Janjić et al., 2014), can be used since it incorporates mass conservation and non-negativity as strong constraints (i.e., constraints are perfectly satisfied) in the analysis update, so that the analysis ensemble mean and each ensemble member meet all of these requirements. The QPEns was further evaluated by Ruckstuhl and Janjić (2018) for convective-scale data assimilation, using a modified 1D shallow water model (Würsch & Craig, 2014), and it was shown that the mass conservation helps suppress the spurious convection. Furthermore, Zeng et al. (2017) extended the QPEns to include nonlinear constraints, such as conservation of total energy and enstrophy, and examined the new algorithm with a 2D shallow water model. It was concluded that the enstrophy conservation effectively reduces the unphysical energy cascade to smaller scales for global-scale data assimilation. Currently, the QPEns is still computationally expensive since the strongly constrained optimization problem that may require numerous iterations to converge has to be solved for each ensemble member. The high expense prevents its application to high-dimensional NWP models. However, recent developments (e.g., Janjić et al., 2021; Ruckstuhl et al., 2021) show that computational aspects for QPEns might be overcome in the near future.

In practice, weak constraints (i.e., constraints that are approximately satisfied) are much more established for application with high-dimensional systems. For instance, Liu and Xue (2006) imposed a weak non-negative water vapor constraint in the cost function of the three-dimensional variational data assimilation (3DVAR) algorithm for a global model. Gao et al. (1999) incorporated an anelastic mass continuity equation as a weak constraint in the storm-scale 3DVAR system of the Advanced Regional Prediction System (ARPS; Xue et al., 2000) and showed that this constraint prevents severe error accumulation in the vertical velocity. Subsequently, Ge et al. (2012) added a diagnostic pressure equation as a weak constraint to the 3DVAR of the ARPS to achieve a more dynamically consistent analysis. Li et al. (2015) implemented a dynamic weak constraint based on the steady momentum equations in the 3DVAR system of the Weather Research and Forecasting (WRF) model, and the constraint improves the accuracy of analyses in pressure, temperature, and moisture within the typhoon.

Although weak constraints have often been adopted in variational approaches, only rarely constraints have been used in ensemble methods. Yilmaz et al. (2011) implemented a weak constraint in the EnKF to mitigate the water balance residuals in the land data assimilation, and it was demonstrated that the analyses of the constrained EnKF are more accurate than those of the unconstrained filter. In Barth et al. (2016), a local assimilation scheme that satisfies conservation of mass was tested in an idealized ocean model and compared to the traditional covariance localization with an ad-hoc step enforcing conservation, and it was shown that the inclusion of the mass conservation property reduces the analysis error. In this study, we will modify the existing LETKF algorithm, which is implemented in the Kilometer-scale ENsemble Data Assimilation (KENDA, Schraff et al., 2016) system for the model of COntsortium for Small-scale Modeling (COSMO, Baldlau et al., 2010), by imposing weak constraints on the mass conservation of hydrometeor variables. Additionally, to preserve the non-negativity of hydrometeor variables, both reflectivity and in particular clear-air reflectivity data are assimilated. It is well recognized (e.g., Aksoy et al., 2009) that the assimilation of clear-air reflectivity data can effectively suppresses spurious convection. As experimentally shown in this work, it actually reduces the number of grid points with negative values in the analysis. Janjić et al. (2014) employed a toy model to demonstrate that when the QPEns strongly preserves both mass and non-negativity, it yields better estimates than algorithms that do not preserve or preserve only mass or non-negativity. In this study, we will explore this idea with the modified LETKF, using the idealized setup for radar data assimilation (Zeng, Janjić, et al., 2021).
The study is organized as follows: Section 2 presents the concept of the weakly constrained LETKF. Section 3 gives a brief introduction to the COSMO model and a detailed description of the twin experiments setup. Section 4 provides the experimental results. Section 5 gives concluding remarks.

2. Weakly Constrained LETKF

2.1. Weakly Constrained LETKF for Mass Conservation

Janjić et al. (2014) showed that ensemble-type Kalman filters are mass conservative if no localization is applied. Since localization is necessary for NWP to reduce spurious long-range correlation, the LETKF is not mass conservative by its nature (e.g., Barth et al., 2016; Zeng et al., 2017; Zeng & Janjić, 2016). In this study, we modify the LETKF algorithm by adding a term for domain-wide mass conservation to the cost function and accordingly also modify the update equation. The result is an approximately mass-conserving analysis error covariance as well as analysis ensemble mean and members. Mass conservation is imposed for each type of mixing ratios (i.e., cloud water $q_c$, cloud ice $q_i$, rain $q_r$, snow $q_s$, and graupel $q_g$), respectively. Details of the algorithm are given in Section S1 in Supporting Information S1.

2.2. Weak Constraint on Non-Negativity

Since the hydrometeor variables are non-negative in nature, assimilation of observations that also have the non-negative property, such as radar reflectivity and cloud products, can be utilized to weakly constrain the non-negativity in the analysis step. As shown in Cohn (1997), on a simple example, if observation and background are positive at a grid point, analysis at the grid point will be positive; however, negative values can be obtained in grid points that are updated using covariances and where there are no observations.

The availability of radar reflectivity measurements indicates non-negative values of hydrometeor variables; however, due to the logarithmic unit, reflectivity values are negative for very small reflectivities. To avoid spurious convection (Aksoy et al., 2009) and overestimated analysis increments (Zeng, Janjić, et al., 2021), a non-negative threshold value is set for very small reflectivities. In this work, the threshold value is 5 dBZ, which means all reflectivity values lower than 5 dBZ are set to 5 dBZ. We call those data “clear-air reflectivity data.” Radar reflectivity data depend nonlinearly on hydrometeors. Further, reflectivity data (including clear-air reflectivity data) are available at radar observation locations, therefore not in every grid point of the model; thus, the example of Cohn (1997) does not apply directly. However, as we will show later, by assimilating additional clear-air reflectivity data, we are reducing negative values that could be obtained through covariance updates and asking in the approximate weak sense that non-negativity is preserved in the analysis of hydrometeors.

Note that in the LETKF, the analysis ensemble is generated around the mean that could have zero values, which will yield negative values in some ensemble members. In stochastic EnKF formulation, when assimilating clear-air reflectivity for each ensemble member, the non-negativity could be approximately satisfied, in a sense discussed above, in each ensemble member. But unless we impose non-negativity strictly as in Janjić et al. (2014), we cannot completely avoid unphysical values in the analysis ensemble either.

3. Description of Model and Experimental Design

The COSMO model is run with a 2-km horizontal resolution and a grid size of $200 \times 200 \times 65$, coupled with an Efficient Modular VOlume scanning RADar Operator (EMVORADO, Zeng et al., 2014, 2016). Due to the high horizontal resolution, the deep convection is explicitly resolved, whereas the shallow convection is parametrized by the Tiedtke scheme (Tiedtke, 1989). The microphysics scheme is a Lin-type one-moment bulk microphysics scheme (Lin et al., 1983; Reinhardt & Seifert, 2006). Turbulence parameterization is performed with the prognostic turbulent kinetic energy equation of Raschendorfer (2001). Periodic boundary conditions are applied. The convection is triggered by a warm bubble approach (Weisman & Klemp, 1982).

The nature run is described as follows: the supercell is triggered by warm bubble perturbations (Zeng et al., 2020) with an analytical profile (Weisman & Klemp, 1982). The nature run starts at 12:00 UTC, warm bubble perturbations are added at 12:30 UTC, and a supercell is triggered. The supercell splits into two cells at 15:00 UTC. The
left-moving cell remains stable in size and shape, while the right-moving cell varies with time and further splits. An illustration of the nature run is given in Figure S1 in Supporting Information S1.

A radar network consisting of seven stations is positioned such that their scanning area covers the propagation path of supercells and the domain of interest. The synthetic radar observations are generated by adding random Gaussian noise with a standard deviation of 5.0 dBZ and 1.0 m/s to the true reflectivity and radial wind data from the nature run, respectively. The radar observations are available in 5-min intervals and are spatially superobbed to 5 km. Both radial wind and reflectivity data are assimilated. If clear-air reflectivity data are assimilated, a threshold value of 5 dBZ is set, that is, all reflectivity values smaller than 5 dBZ are set to 5 dBZ. If clear-air reflectivity data are not assimilated, all reflectivity values smaller than 5 dBZ are set to missing values. The horizontal localization radius is 8 km, and the vertical localization radius changes from 0.0075 to 0.5 (in logarithm of pressure) with the height. The observation error covariance matrix is diagonal.

The ensemble size is 80. The initial ensemble is created by the first perturbing time, location, and intensity of the warm bubble at 12:30 UTC (more details about the perturbations can be found in Zeng, Janjić, et al., 2021) and then evolving ensemble in time freely until 15:00 UTC. An illustration of the initial ensemble is given in Figure S2 in Supporting Information S1. Data assimilation starts at 15:00 and ends at 17:00 UTC with an assimilation window of 6 min. The multiplicative inflation (Anderson & Anderson, 1999) is employed with a coefficient of 1.1. In all experiments, after the analysis update, a physical consistency check is applied to each ensemble member, which seeks negative values of hydrometeor variables and clips them to zero. We refer to the analysis after the analysis update but before the physical consistency check as “pre-analysis.” The ensemble of pre-analysis could be associated with negative values for hydrometeor variables. Note that the model states at the lateral boundaries are not updated. At 16:00 UTC and 17:00 UTC, 2-h ensemble forecasts are run (forecast data are outputted every 15 min).

To explore the performance of weak constraints on mass conservation and non-negativity, four sensitivity experiments are executed:

1. E_Control: Clear-air reflectivity data are not assimilated and the mass conservation constraints are not imposed;
2. E_M (“M” indicates mass conservation constraints): Clear-air reflectivity data are not assimilated and the mass conservation constraints are imposed;
3. E_P (“P” indicates positivity constraint): Clear-air reflectivity data are assimilated and the mass conservation constraints are not imposed;
4. E_MP: Clear-air reflectivity data are assimilated and the mass conservation constraints are imposed.

All experiments update model state variables including 3D wind, temperature, pressure, and hydrometeor variables. In addition, in E_M and E_MP, the ensemble means of hydrometeor variables are corrected by Equation S6 in Supporting Supporting Information S1.

4. Experimental Results

To evaluate the performance during data assimilation, the root mean square errors (RMSEs) of the ensemble means are calculated for the aggregate of hydrometers (i.e., $q_c + q_i + q_r + q_s + q_g$) and for the meridional wind $v$. Furthermore, as in many previous studies in idealized setups (e.g., Tong & Xue, 2005), to obtain a more subtle and accurate measure for assessing the storm, the RMSEs are also calculated at points at which reflectivities in the nature run are greater than 5 dBZ. The relative changes (compared to true values in the nature run) in the analysis ensembles for the total mass of each hydrometeor variable and the total amount of divergence and vorticity, as well as the number of model grid points with negative values of hydrometeor variables in pre-analysis ensembles, are provided. Furthermore, the analysis ensembles are compared by grid-point-based ensemble probabilities for the reflectivity composite (i.e., for each grid point calculating the percentage of ensemble members that exceed a given threshold value of reflectivity). To evaluate the forecast skill, the fractions skill score (FSS, Robert & Lean, 2008) for the reflectivity composite is computed.

Figure 1 shows the RMSEs of the ensemble means for the aggregate of hydrometeors and for the meridional wind $v$. From the vertical profiles in Figure 1a, E_P and E_MP produce smaller RMSEs than E_Control and E_M.
for the aggregate in case of considering the whole domain, however, if considering only the area of storms, the RMSEs of E_M (E_MP) are larger than E_Control (E_P). This is because the mass conservation constraints act globally, not just trying to fit radar data inside storms. If breaking down the RMSEs into cycles as in Figure 1b, no clear differences can be seen except that the RMSEs of E_Control become considerably larger at the end of cycles. For the $v$, it is noticed in Figure 1c that the RMSEs in E_Control and E_M are smaller than in E_P and E_MP for both the whole domain and the area of storms. This is attributed to less accurate results of E_P and E_MP in the first half of the cycles as shown in Figure 1d. Since radial wind data are directly correlated to $v$, in case of not assimilating clear-air reflectivity data, radial wind data have more impacts on $v$ and more accurate $v$ can be achieved. Otherwise, some spin-up time is required to obtain an appropriate covariance between clear-air

Figure 1. (a) The vertical profiles of the RMSEs of the analysis ensemble means for the aggregate of hydrometeors in E_Control, E_M, E_P, and E_MP, over all grid points (dashed) and over points at which the reflectivities in the nature run are greater than 5 dBZ (solid), averaged over all cycles. The model levels are labeled from the model top. (b) The RMSEs of the background and analysis ensemble means (sawtooth pattern) for the aggregate of hydrometeors at each cycle between 15 and 17 UTC. (c) and (d) are for the meridional wind $v$. 
reflectivity data and $v$. Overall, imposing mass conservation constraints reduces the RMSEs of hydrometeors in global but not necessarily in area of storms, and in case of assimilating clear-air reflectivity data, a longer spin-up time is required for dynamical components of storms.

Figure 2a shows the relative changes in the total mass for each hydrometeor variable and for the aggregate of them (indicated by “all”) in E_Control, E_M, E_P, and E_MP compared to those of the nature run. Positive bar values mean greater than the nature run and vice versa. (b) Numbers of model grid points in analysis with negative $q_c$, $q_i$, $q_r$, $q_s$, and $q_g$ before the physical consistency check. Bars indicate numbers of points averaged over all analysis ensemble members and all cycles before the physical consistency check, and error bars indicate standard deviations. (c) Relative changes [%] of total amount for integrated mass-flux divergence (sd) and vorticity (sv) compared to those of the nature run.

Figure 2. (a) Relative changes [%] of total mass for each hydrometeor variable (i.e., $q_c$, $q_i$, $q_r$, $q_s$, and $q_g$) and the aggregate of them (indicated by “all”) in E_Control, E_M, E_P, and E_MP compared to those of the nature run. Bars indicate relative changes averaged over all analysis ensemble members and all cycles, and error bars indicate standard deviations. Positive bar values mean greater than the nature run and vice versa. (b) Numbers of model grid points in analysis with negative $q_c$, $q_i$, $q_r$, $q_s$, and $q_g$ before the physical consistency check. Bars indicate numbers of points averaged over all analysis ensemble members and all cycles before the physical consistency check, and error bars indicate standard deviations. (c) Relative changes [%] of total amount for integrated mass-flux divergence (sd) and vorticity (sv) compared to those of the nature run.
Figure 2b provides the number of the model grid points with negative $q_c$, $q$, $q_r$, and $q_s$ in the pre-analysis ensembles. Taking $q_r$ as an example, $E_{\text{Control}}$ produces approximately 90,000 grid points with negative values ($\sim 3.5\%$ of total grid points), and $E_M$ produces slightly fewer grid points. $E_P$ is associated with much fewer negative points ($\sim 50,000$) and $E_{\text{MP}}$ further reduces the number of negative points to 35,000. For the other variables, it also holds that $E_{\text{Control}}$ has the largest number of negative points, followed by $E_M$, $E_P$, and $E_{\text{MP}}$. Furthermore, Figure 2c illustrates the relative changes in total amount for integrated mass-flux divergence (denoted by “sdiv”) and vorticity (denoted by “svor”). We regard changes in these quantities as an indicator for dynamical imbalance for the entire domain, in particular, since sdiv is proportional to the surface pressure tendency in the case of hydrostatic pressure (Zeng, de Lozar, et al., 2021). Positive biases can be seen for both sdiv and svor in all experiments. For both sdiv and svor, $E_{\text{Control}}$ produces the greatest biased increase by average and $E_{\text{MP}}$ produces the least biased increase by average. For sdiv, $E_M$ produces slightly more than $E_P$, whereas the opposite results are achieved for svor. Assimilation of clear-air reflectivity as a weak constraint on non-negativity greatly contributes towards reducing the number of grid points with negative values of hydrometeor variables in pre-analyses and towards mitigating the biased increase and uncertainties in the total mass. In this regard, imposing weak constraints on mass conservation proves to be effective (especially in reducing mass bias) but not as effective as assimilating clear-air reflectivity. The best results are obtained by a combination of both constraints. Moreover, imposing these two constraints also mitigates the biased increase in divergence and vorticity, therefore, a slow error growth in forecasts is expected as shown below.

Furthermore, Figure 3 compares the analysis ensembles of $E_{\text{Control}}$, $E_M$, $E_P$, and $E_{\text{MP}}$ at 16:00 UTC by the reflectivity composite at elevation 0.5° (in terms of the gridpoint-based ensemble probability, i.e., the number of ensemble members exceeding the threshold value of 25 dBZ divided by the ensemble size). After 10 cycles, $E_{\text{Control}}$ captures the correct cells with very high probabilities; however, it produces spurious convection in a widespread area in the rear, with fairly high probabilities at some locations. $E_M$ also produces spurious convection, but with areal coverage much smaller than that of $E_{\text{Control}}$. The false signals are essentially removed in $E_P$ and even less present in $E_{\text{MP}}$. This finding is consistent with Figure 2.

Figure 4a provides the FSSs of reflectivity composite at 0.5° for a threshold of 25 dBZ and a box length of 16 km as a function of the forecast lead time. The FSS values of $E_P$ and $E_{\text{MP}}$ are very close at the initial time (both are greater than 0.95), but the skills of $E_P$ decrease faster with the lead time than those of $E_{\text{MP}}$. $E_M$ is considerably worse than $E_P$ and $E_{\text{MP}}$, especially at the initial time (with FSS values of $\sim 0.72$). However, $E_M$ is still much better than $E_{\text{Control}}$ throughout the forecast lead time. Figure 4b shows the FSS values of reflectivity composite as a function of the box length (from 4 to 128 km), where the values are averaged over all members and over all forecast lead times. It is clear that the FSS values of $E_{\text{MP}}$ are consistently higher (approximately 0.05) than those of $E_P$ for all box lengths. $E_P$ outperforms $E_M$ by approximately 0.1; the latter outperforms $E_{\text{Control}}$ by approximately 0.1 for a box length of 4 km and the advantage increases slightly with increasing box lengths. The fact that experiments with weak constraints enable improved forecast skills may be, on one hand, due to less biased hydrometeor variables in the analysis, and on the other hand, due to more dynamically balanced analysis state as seen through a smaller increase in divergence and vorticity.

5. Conclusion and Outlook

Despite the significant benefits of EnKFT for convective scale data assimilation, updates of hydrometeors with observations can produce negative values in the analysis ensemble and add mass bias to the model. To mitigate this sign problem and to simultaneously maintain mass conservation, a new concept of combining weak constraints on mass conservation and non-negativity has been introduced in this work with a focus on the hydrometeor variables. We modify the LETKFT with weak constraints on the mass conservation of each hydrometeor variable and adopt the assimilation of clear-air reflectivity data as a weak constraint on non-negativity.

We examine the concept by a series of sensitivity experiments, using an idealized setup for radar data assimilation. The results show that both weak constraints may not always reduce the errors of model state estimates; however, they both successfully improve the mass conservation property of hydrometeor variables of analyses. In detail, weak constraints on both mass conservation and non-negativity considerably reduce the number of grid points with negative values of hydrometeor variables during analysis update and mitigate the biased mass increase and uncertainties in total mass, while the latter constraint is more effective. It is seen that the reduction in
biased mass increase causes less spurious convection. In addition, both constraints also reduce the biased increase in integrated mass-flux divergence and vorticity. Overall, the least biased increase is obtained by combining both constraints. Weak constraints on mass conservation and non-negativity also enhance the forecast skills, which can be attributed to the less biased hydrometeor variables, integrated mass-flux divergence, and vorticity.

How the impacts of constraints vary for different assimilation windows is worth of further investigation. Based on the results in more idealized setups (Ruckstuhl & Janjić, 2018), there are two effects to consider. On one hand, if there is more time between analysis, errors grow in mass as well. On the other hand, the noise and spurious convection are more extreme for more frequent updates (Lange et al., 2017). In both cases, it seems to be beneficial to include these constraints, but in which case it is more beneficial it is not clear. Further, this method can also be combined with the adjustment filter (Zeng, de Lozar, et al., 2021) which works more on the dynamical balance.

As the proof of concept illustrates, method proposed here is well suited for the assimilation of radar data in convection-permitting models. It is appropriate for estimation of high-dimensional state and requires only minor changes to the already existing implementation of LETKF. In the next steps, we plan to apply it to the convective-scale data assimilation with real data. The total mass is unknown in real cases; therefore, it should be treated as uncertain and estimated during data assimilation (Janjić et al., 2014). As a choice, the value of total mass of each analysis ensemble member can also be restricted to that of the background ensemble member as in Ruckstuhl and Janjić (2018). Furthermore, this concept should also be of interest for the other research fields as chemistry, ecosystem, and ocean data assimilation (Bertino et al., 2003; Bocquet et al., 2010; Buehner et al., 2013; Simon & Bertino, 2012), in which mass conservation and non-negativity could also be important.

Figure 3. Comparison of analysis ensembles at 16:00 UTC in E_Control, E_M, E_P, and E_MP, in terms of the gridpoint-based ensemble probability of reflectivity composite at elevation 0.5° exceeding the threshold value 25 dBZ. Reflectivity ≥25 dBZ in the nature run is denoted with contour solid lines.
Data Availability Statement

Access to the source code of the COSMO model is restricted to COSMO licenses. A free license can be obtained for research if following the procedure described at http://www.cosmo-model.org/content/consortium/licencing.htm. The processed data and plotting scripts are freely accessible under https://doi.org/10.5281/zenodo.5129256.

References

Aksoy, A., Dowell, D. C., & Snyder, C. (2009). A multiscale comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analyses. Monthly Weather Review, 137, 1805–1824. https://doi.org/10.1175/2008mwr2691.1

Anderson, J. L., & Anderson, S. L. (1999). A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. Monthly Weather Review, 127, 2741–2758. https://doi.org/10.1175/1520-0493(1999)127<2741:amciot>2.0.co;2

Baldaluf, M., Förstner, J., Klink, S., Reinhardt, T., Schraff, C., Seifert, A., & Stephan, K. (2010). Kurze Beschreibung des Lokal-Modells Kurzfrist COSMO-DE (LMK) und seiner Datenbanken auf dem Datenserver des DWD (Tech. Rep.). Deutscher Wetterdienst.

Bannister, R. N., Chipilski, H. G., & Martinez-Alvarado, O. (2020). Techniques and challenges in the assimilation of atmospheric water observations for numerical weather prediction towards convective scales. Quarterly Journal of the Royal Meteorological Society, 146, 1–48. https://doi.org/10.1002/qj.3652

Barth, A., Yan, Y., Alvera-Azcárate, A., & Beckers, J.-M. (2016). Local ensemble assimilation scheme with global constraints and conservation. Ocean Dynamics, 66, 1651–1664. https://doi.org/10.1007/s10236-016-0999-y

Bertino, L., Evensen, G., & Wackernagel, H. (2003). Sequential data assimilation techniques in oceanography. International Statistical Reviews, 71, 223–241. https://doi.org/10.1111/1.1751-5823.2003.tb00194.x

Bocquet, M., Pires, C. A., & Wu, L. (2010). Beyond Gaussian statistical modeling in geophysical data assimilation. Monthly Weather Review, 138, 2997–3023. https://doi.org/10.1175/2010mwr3164.1

Buchner, M., Caya, A., Pogson, L., Carrières, T., & Peticœur, P. (2013). A new environment canada regional ice analysis system. Atmosphere-Ocean, 51(1), 18–34. https://doi.org/10.1080/07055900.2012.747171

Cohn, S. E. (1997). An introduction to estimation theory. Journal of the Meteorological Society of Japan, 75, 257–288. https://doi.org/10.2151/jmsj1965.75.1b_257

Dowell, D. C., Wicker, L. J., & Snyder, C. (2011). Ensemble Kalman Filter assimilation of radar observations of the 8 May 2003 Oklahoma City Supercell: Influences of reflectivity observations on storm-scale analyses. Monthly Weather Review, 139, 272–294. https://doi.org/10.1175/2010mwr3438.1

Dowell, D. C., Zhang, F., Wicker, L. J., Snyder, C., & Crook, N. A. (2004). Wind and temperature retrievals in the 17 May 1981 Arcadia, Oklahoma, supercell: Ensemble Kalman filter experiments. Monthly Weather Review, 132. https://doi.org/10.1175/1520-0493(2004)132<1982:wtret>2.0.co;2

Evensen, G. (2003). The Ensemble Kalman Filter: Theoretical formulation and practical implementation. Ocean Dynamics, 53, 343–367. https://doi.org/10.1007/s10236-003-0036-9

Gao, J., Xue, M., Shapiro, A., & Droegemeier, K. K. (1999). A variational analysis for the retrieval of three-dimensional mesoscale wind fields from two doppler radars. Monthly Weather Review, 127, 2128–2142. https://doi.org/10.1175/1520-0493(1999)127<2128:avfasd>2.0.co;2

Ge, G., Gao, J., & Xue, M. (2012). Diagnostic pressure equations as a weak constraint in a storm-scale three dimensional variational radar data assimilation system. Journal of Atmospheric and Oceanic Technology, 29, 1075–1092. https://doi.org/10.1175/jtech-d-11-00201.1
Gustafsson, N., Janjić, T., Schraff, C., Leuenberger, D., Weissman, M., Reich, H., et al. (2018). Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. Quarterly Journal of the Royal Meteorological Society, 144, 1218–1256. https://doi.org/10.1002/qj.3179

Hunt, B. R., Kostelich, E. J., & Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman Filter. Physica D: Nonlinear Phenomena, 230, 112–126. https://doi.org/10.1016/j.physd.2006.11.008

Janjić, T., McLaughlin, D., Cohn, S. E., & Verlaan, M. (2014). Conservation of mass and preservation of positivity with ensemble-type Kalman Filter algorithms. Monthly Weather Review, 142, 755–773. https://doi.org/10.1175/MWR-D-13-00056.1

Janjić, T., Ruckstuhl, Y., & Toint, P. L. (2021). A data assimilation algorithm for predicting rain. Quarterly Journal of the Royal Meteorological Society, 147(736), 1949–1963. https://doi.org/10.1002/qj.4004

Kühnelin, C., Keil, C., Craig, G. C., & Gebhardt, C. (2014). The impact of downscaled initial condition perturbations on convective-scale ensemble forecasts of precipitation. Quarterly Journal of the Royal Meteorological Society, 140, 1552–1562. https://doi.org/10.1002/qj.2238

Langer, H., Craig, G. C., & Janjić, T. (2017). Characterizing noise and spurious convection in convective data assimilation. Quarterly Journal of the Royal Meteorological Society, 143, 3060–3069. https://doi.org/10.1002/qj.3162

Lauvaret, C., Brandt, J. M., Castruccio, F., Bouquet, G., Brasseur, P., & Verron, J. (2009). A truncated Gaussian filter for data assimilation with inequality constraints: Application to the hydrostatic stability in ocean models. Ocean Modelling, 27, 1–17. https://doi.org/10.1016/j.ocemod.2008.10.007

Li, X., Ming, J., Xue, M., Wang, Y., & Zhao, K. (2015). Implementation of a dynamic equation constraint based on the steady state momentum equations within the WRF hybrid ensemble-3DVar data assimilation system and test with radar T-TREC wind assimilation for tropical Cyclone Chandu (2010). Journal of Geophysical Research: Atmospheres, 120, 4017–4039. https://doi.org/10.1002/2014JD022706

Lien, G., Kalnay, E., & Miyoshi, T. (2013). Effective assimilation of global precipitation: Simulation experiments. Tellus, 65, 1915. https://doi.org/10.3402/tellusa.v65i6.19115

Lin, Y., Farley, R. D., & Orville, H. D. (1983). Bulk parameterization of the snow field in a cloud model. Journal of Applied Meteorology and Climatology, 22, 1065–1082. https://doi.org/10.1175/1520-0450(1983)022<1065:bpotsf>2.0.co;2

Liu, H., & Xue, M. (2006). Retrieval of moisture from slant-path water vapor observations of a hypothetical GPS network using a three-dimensional variational scheme with anisotropic background error. Monthly Weather Review, 134, 933–949. https://doi.org/10.1175/mwr3105.1

Minamide, M., & Zhang, F. (2019). An adaptive background error inflation method for assimilating all-sky radiances. Quarterly Journal of the Royal Meteorological Society, 145, 805–823. https://doi.org/10.1002/qj.3466

Raschendorfer, M. (2001). The new turbulence parametrization of LM. COSMO-Newsletter, 1, 89–97.

Reinhardt, T., & Seifert, A. (2006). A three-category ice scheme for LMK. COSMO Newsletter, 6, 115–120.

Roh, N., & Lean, H. (2008). Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. Monthly Weather Review, 136, 78–97. https://doi.org/10.1175/2007mwr2123.1

Ruckstuhl, Y., & Janjić, T. (2018). Parameter and state estimation with ensemble Kalman filter based algorithms for convective-scale applications. Quarterly Journal of the Royal Meteorological Society, 144, 826–841. https://doi.org/10.1002/qj.3257

Ruckstuhl, Y., Janjić, T., & Rasp, S. (2021). Training a convolutional neural network to conserve mass in data assimilation. Nonlinear Processes in Geophysics, 28(1), 111–119. https://doi.org/10.5194/npg-28-111-2021

Schraff, C., Reich, H., Rhodin, A., Schomburg, A., Stephan, K., Perlaídez, A., & Potthast, R. (2016). Kilometre-scale ensemble data assimilation for the cosmo-model (VENDA). Quarterly Journal of the Royal Meteorological Society, 142, 1453–1472. https://doi.org/10.1002/qj.2748

Simon, E., & Bertino, L. (2009). Application of the Gaussian anamorphosis to assimilation in a 3-D coupled physical-ecosystem model of the North Atlantic with the ENKF: A twin experiment. Ocean Science, 5, 495–510. https://doi.org/10.5194/os-5-495-2009

Simon, E., & Bertino, L. (2012). Gaussian anamorphosis extension of the DNEKF for combined state parameter estimation: Application to a 1D ocean ecosystem model. Journal of Marine Systems, 89, 1–18. https://doi.org/10.1016/j.jmarsys.2011.07.007

Snyder, C., & Zhang, F. (2003). Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. Monthly Weather Review, 131, 1663–1677. https://doi.org/10.1175/1520-0493(2003)131<1663:ASOSDRO>2.0.CO;2

Sun, J., Xue, M., Wilson, J. W., Zawadzki, I., Ballard, S. P., Onvlee-Hooimeyer, J., et al. (2014). Use of nwp for nowcasting convective precipitation: Additive noise, relaxation methods and combinations. Quarterly Journal of the Royal Meteorological Society, 140, 1065–1092. https://doi.org/10.1002/qj.2238

Tellus, 50, 1949–1963. https://doi.org/10.1175/1520-0450(1989)022<1779:acmfsf>2.0.co;2

Wang, Z., & Xue, M. (2016). Use of nwp for nowcasting convective precipitation: Additive noise, relaxation methods and combinations. Quarterly Journal of the Royal Meteorological Society, 140, 1552–1562. https://doi.org/10.1002/qj.3162

Xue, M., Droegemeier, K. K., & Wong, V. (2000). The advanced regional prediction system (ARPS) – A multi-scale nonhydrostatic atmospheric simulation and prediction tool. Part I: Model dynamics and verification. Meteorology and Atmospheric Physics, 75, 161–193. https://doi.org/10.1007/s007030050003

Yilmaz, M. Y., Delsole, T., & Houser, P. R. (2011). Improving land data assimilation performance with a water budget constraint. Journal of Hydrometeorology, 12, 1040–1055. https://doi.org/10.1175/2011jhm1346.1

Zeng, Y., Blahak, U., & Jerger, D. (2016). An efficient modular volume-scanng radar forward operator for NWP models: Description and coupling to the COSMO model. Quarterly Journal of the Royal Meteorological Society, 142, 3234–3256. https://doi.org/10.1002/qj.2904

Zeng, Y., Blahak, U., Neuper, M., & Jerger, D. (2014). Radar beam tracing methods based on atmospheric refractive index. Journal of Atmospheric and Oceanic Technology, 31, 2650–2670. https://doi.org/10.1175/jtech-d-13-00152.1

Zeng, Y., de Lozar, A., Janjić, T., & Seifert, A. (2021). Applying a new integrated mass-flux adjustment filter in rapid update cycling of convective-scale data assimilation for the cosmo-model (V5.07). Geoscientific Model Development, 14, 1295–1307. https://doi.org/10.5194/gmd-14-1295-2021

Zeng, Y., & Janjić, T. (2016). Study of conservation laws with the local ensemble transform Kalman Filter. Quarterly Journal of the Royal Meteorological Society, 699, 2359–2372. https://doi.org/10.1002/qj.3289

Zeng, Y., Janjić, T., de Lozar, A., Blahak, U., Reich, H., Keil, C., & Seifert, A. (2018). Representation of model error in convective-scale data assimilation: Additive noise, relaxation methods and combinations. Journal of Advances in Modeling Earth Systems, 10, 2889–2911. https://doi.org/10.1002/2018MS001375

Zeng, Y., Janjić, T., de Lozar, A., Rasp, S., Blahak, U., Seifert, A., & Craig, G. C. (2020). Comparison of methods accounting for subgrid-scale model error in convective-scale data assimilation. Monthly Weather Review, 148, 2457–2477. https://doi.org/10.1175/mwr-d-19-0064.1
Zeng, Y., Janjić, T., de Lozar, A., Welzbacher, C., Blahak, U., & Seifert, A. (2021). Assimilating radar radial wind and reflectivity data in an idealized setup of the COSMO-KENDA system. *Atmospheric Research*, 249, 105282. https://doi.org/10.1016/j.atmosres.2020.105282

Zeng, Y., Janjić, T., Ruckstuhl, Y., & Verlaan, M. (2017). Ensemble-type Kalman filter algorithm conserving mass, total energy and enstrophy. *Quarterly Journal of the Royal Meteorological Society*, 143, 2902–2914. https://doi.org/10.1002/qj.3142

Zeng, Y., Janjić, T., Sommer, M., de Lozar, A., Blahak, U., & Seifert, A. (2019). Representation of model error in convective-scale data assimilation: Additive noise based on model truncation error. *Journal of Advances in Modeling Earth Systems*, 11, 752–770. https://doi.org/10.1029/2018ms001546

Zhang, F., Snyder, C., & Sun, J. (2004). Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Monthly Weather Review*, 132(5), 1238–1253. https://doi.org/10.1175/1520-0493(2004)132<1238:ioieao>2.0.co;2

**Reference From the Supporting Information**

Bishop, C. H., Etherton, B. J., & Majumdar, S. J. (2001). Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Monthly Weather Review*, 129, 420–436. https://doi.org/10.1175/1520-0493(2001)129<0420:aswtet>2.0.co;2