Robust Data Assimilation Using L_1 and Huber Norms

Rao, Vishwas; Sandu, Adrian; Ng, Michael; Nino-Ruiz, Elias D.

Published in:
SIAM Journal on Scientific Computing

DOI:
10.1137/15M1045910

Published: 22/06/2017

Document Version:
Publisher's PDF, also known as Version of record

SIAM J. SCI. COMPUT.
Vol. 39, No. 3, pp. B548–B570

ROBUST DATA ASSIMILATION USING L_1 AND HUBER NORMS*

VISHWAS RAO†, ADRIAN SANDU‡, MICHAEL NG§, AND ELIAS D. NINO-RUIZ¶

Abstract. Data assimilation is a process used to fuse information from priors, observations of nature, and numerical models, in order to obtain best estimates of the parameters or state of a physical system of interest. Presence of large errors in some observational data, e.g., data collected from a faulty instrument, negatively affect the quality of the overall assimilation results. This work develops a systematic framework for robust data assimilation. The new algorithms continue to produce good estimates of parameters or state in the presence of observation outliers. The approach is based on replacing the traditional L_2 norm formulation of data assimilation problems with formulations based on L_1 and Huber norms. Numerical experiments using the Lorenz-96 and the shallow water on the sphere models illustrate how the new algorithms outperform traditional data assimilation approaches in the presence of data outliers.

Key words. data assimilation, 4D-var, ADMM, Huber norm

AMS subject classification. 65

DOI. 10.1137/15M1045910

1. Introduction.

Dynamic data-driven application systems (DDDAS) [4] integrate computational simulations and physical measurements in symbiotic and dynamic feedback control systems. Within the DDDAS paradigm, data assimilation defines a class of inverse problems that fuses information from an imperfect computational model based on differential equations (which encapsulates our knowledge of the physical laws that govern the evolution of the real system), from noisy observations (sparse snapshots of reality), and from an uncertain prior (which encapsulates our current knowledge of reality). Data assimilation integrates these three sources of information and the associated uncertainties into a Bayesian framework to provide the posterior, i.e., the probability distribution conditioned on the uncertainties in the model and observations.

Two approaches to data assimilation have gained widespread popularity: ensemble-based estimation and variational methods. The ensemble-based methods are rooted in statistical theory, whereas the variational approach is derived from optimal control theory. The variational approach formulates data assimilation as a nonlinear optimization problem constrained by a numerical model. The initial conditions

*Submitted to the journal's Computational Methods in Science and Engineering section October 29, 2015; accepted for publication (in revised form) December 13, 2016; published electronically June 22, 2017.

http://www.siam.org/journals/sisc/39-3/M104591.html

Funding: This work was supported by AFOSR DDDAS program through the award AFOSR FA9550–12–1–0293–DEF managed by Dr. Frederica Darema, and by the Computational Science laboratory at Virginia Tech. The third author's research was supported in part by HKRGC GRFs 202013 and 12301214, and HKBU FRG2/14-15/087.

†Institute for Computational Engineering and Sciences, University of Texas, Austin (visrao@vt.edu).

‡Department of Computer Science, Computational Science Laboratory, Virginia Tech., Blacksburg, VA 24060 (sandu@cs.vt.edu).

§Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong (mng@math.hkbu.edu.hk).

¶Department of Computer Science, Universidad del Norte, Barranquilla, Atlantico, Colombia (enino@vt.edu).

(as well as boundary conditions, forcing, or model parameters) are adjusted to minimize the discrepancy between the model trajectory and a set of time-distributed observations. In real-time operational settings the data assimilation process is performed in cycles: observations within an assimilation window are used to obtain an optimal trajectory, which provides the initial condition for the next time window, and the process is repeated in the subsequent cycles.

Large errors in some observations can adversely impact the overall solution to the data assimilation system, e.g., can lead to spurious features in the analysis [27]. Various factors contribute to uncertainties in observations. Faulty and malfunctioning sensors are a major source of large uncertainties in observations, and data quality control is an important component of any weather forecasting data assimilation system [29]. The goal of quality control is to ensure that only correct observations are used in the data assimilation system. Erroneous observations can lead to spurious features in the resulting solution [27]. under the assumption that observation errors have a Gaussian distribution. This assumption makes the large innovation values very unlikely, and the analysis is pulled towards the corresponding observations

Departure of the observation from the background forecast is usually used as a measure for accepting or rejecting observations: observations with large deviations from the background forecast are rejected in the quality control step [22]. However, this process has important drawbacks. A large departure from the background can be due to a large observation error, in which case not using that data point helps the analysis. On the other hand, a large distance between the data and the background can also be due to a large (local) error in the background, in which case the (relatively) accurate observation carries a lot of information, and not using it hurts the analysis. The fundamental issue with using the disagreement between the observations and the model forecasted background as the basis for data quality control is that this metric cannot discriminate between erroneous data and highly informative data. Indeed, Tavolato and Isaksen present several case studies [43, 44] that demonstrate that rejecting observations on the basis of background departure statistics leads to the inability to capture small scales in the analysis.

An alternative to rejecting the observations that fail the quality control tests is to revisit the standard assumption that observation errors have a Gaussian distribution. Under the Gaussian assumption large observation departures are highly unlikely, and the corresponding data points have a large impact on the analysis. More realistic probability distributions accommodate large observation departures from the model forecasted background (by assigning them higher probabilities), and thereby decrease the sensitivity of the analysis with respect to the data points that are in disagreement with the model. This leads to a (more) robust formulation of the data assimilation problem.

The early "variational quality control" algorithm [1, 25] makes the assumption that data error distribution is a convex combination of a Gaussian (representing the random errors) and a constant distribution over a bounded interval of values (representing the gross errors). The observation term of the variational cost function is modified so as to account for this data error distribution. In this approach the probability density of very large observation errors falls back to Gaussian, which is precisely the assumption that one seeks to avoid using robust statistics.

The Huber norm has been used in the context of robust methods for seismic inversion [20]. Tavolato and Isaksen [43, 44] provide a rigorous formulation of the Huber norm four-dimensional (4D)-Var. The parameters of the Huber norm, describing the

observation errors, are learned from real data sets. The 4D-Var problem is solved in the incremental approach: at each inner iteration a linearized cost function is optimized, subject to linearized model dynamics constraints. At the outer iteration level the optimal increment is added to the current solution to obtain the next solution, a new linearization of the model and cost function is computed, and the cycle is repeated. Their Huber norm approach [43, 44] adjusts error covariances of observations based on their departures from the current solution. No observations are discarded, but smaller weights are given to low quality observations (that show large departures from the current solution). The weights are kept constant during the inner loop, where a regular incremental 4D-Var is run. The weights are recomputed for each new solution at each outer iteration.

A Huber-type approach has been employed to construct a robust ensemble Kalman filter by Roh et al. [37]. A modified analysis is obtained by clipping those innovations (discrepancies between observations and background) that are either too small or too large with respect to the clipping threshold which is selected dynamically. This algorithm is based on data departures form the background, and therefore faces the fundamental issue discussed above. Moreover, it is not clear what statistical problem is solved by the clipped innovations algorithm; a rigorous formulation of the underlying problem is not specified.

Previous work has also formulated data assimilation problems using robust norms for the background term. Freitag, Nichols, and Budd [18] use an L_1 norm on the background term, and show that the resulting 4D-Var analyses are better able to capture sharp gradients and discontinuities in the solution. Freitag, Nichols, and Budd [19] extend this work to using a mixed norm, where an L_1 or a total variation norm penalty for the background is added as an additional regularization term to the traditional L_2 problem formulation. A framework for using general background regularization terms is discussed by Ebtehaj et al. [13]. Here we are concerned only with robustness with respect to data errors.

This paper develops a systematic mathematical framework for data assimilation approaches that are robust with respect to outliers in the data. Different data assimilation algorithms are reformulated as optimization problems using L_1 and Huber norms, and are solved via the alternating direction method of multipliers (ADMM) developed by Boyd [7], and via the half-quadratic minimization method presented in [33]. To the best of our knowledge, no prior work has fully addressed the construction of robust data assimilation methods using both the L_1 and the Huber norms in a unified framework.

The remainder of the paper is organized as follows: section 2 reviews the data assimilation problem and the traditional solution algorithms. Section 3 presents the new robust algorithms for three-dimensional (3D)-Var, section 4 for 4D-Var, and section 5 develops new robust algorithms for ensemble-based approaches. Numerical results are presented in section 6. Concluding remarks and future directions are discussed in section 7.

2. Data assimilation. Data assimilation is the process of fusing information from priors, imperfect model predictions, and noisy data, to obtain a consistent description of the true state $\mathbf{x}^{\text{true}}$ of a physical system [12, 26, 39, 40]. The resulting best estimate is called the analysis $\mathbf{x}^{\text{a}}$.

The prior information encapsulates our current knowledge of the system. Prior information typically consists of a background estimate of the state $\mathbf{x}^{\text{b}}$, and the corresponding background error covariance matrix $\mathbf{B}$.

The model captures our knowledge about the physical laws that govern the evolution of the system. The model evolves an initial state $\mathbf{x}_0 \in \mathbb{R}^n$ at the initial time t_0 to states $\mathbf{x}_i \in \mathbb{R}^n$ at future times t_i. A general model equation is

$$\mathbf{x}_i = \mathcal{M}_{t_0 \to t_i}(\mathbf{x}_0), \quad i = 1, \ldots, N. \tag{2.1}$$

Observations are noisy snapshots of reality available at discrete time instances. Specifically, measurements $\mathbf{y}_i \in \mathbb{R}^m$ of the physical state $\mathbf{x}^{\text{true}}(t_i)$ are taken at times t_i,

$$\mathbf{y}_i = \mathcal{H}(\mathbf{x}_i) + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_i), \quad i = 1, \ldots, N, \tag{2.2}$$

where the observation operator $\mathcal{H} : \mathbb{R}^n \to \mathbb{R}^m$ maps the model state space onto the observation space. The random observation errors ε_i are assumed to have normal distributions. In general, both the model and the observation operator are nonlinear.

Variational methods solve the data assimilation problem in an optimal control framework. Specifically, one adjusts a control variable (e.g., model parameters) in order to minimize the discrepancy between model forecasts and observations. Ensemble-based methods are sequential in nature. Generally, the error statistics of the model estimate is represented by an ensemble of model states and ensembles are propagated to predict the error statistics forward in time. When the observations are assimilated, the analysis is obtained by operating directly on the model states. It is important to note that all the approaches discussed in the following subsections aim to minimize the L_2 norm of the discrepancy between observations and model predictions.

Ensemble-based methods employ multiple model realizations in order to approximate the background error distribution of any model state. The empirical moments of the ensemble are used in order to estimate the background state and the background error covariance matrix. Since the model space is several times larger than the ensemble size, localization methods are commonly used in order to mitigate the impact of sampling errors (spurious correlations). For example, the local ensemble transform Kalman filter [24] performs the assimilation of observations locally in order to avoid the impact of analysis innovations coming from distant components.

We next review these traditional approaches.

2.1. Three-dimensional variational data assimilation. The 3D-Var data assimilation approach processes the observations successively at times $t_1, t_2, \ldots, t_N$. The background state (i.e., the prior state estimate at time t_i) is given by the model forecast started from the previous analysis (i.e., the best state estimate at t_{i-1}):

$$\mathbf{x}_i^{\text{b}} = \mathcal{M}_{t_{i-1} \to t_i}\left(\mathbf{x}_{i-1}^{\text{a}}\right). \tag{2.3}$$

The discrepancy between the model state $\mathbf{x}_i$ and observations at time t_i, together with the departure of the state from the model forecast $\mathbf{x}_i^{\text{b}}$, are measured by the 3D-Var cost function

$$\mathcal{J}(\mathbf{x}_i) = \frac{1}{2}\|\mathbf{x}_i - \mathbf{x}_i^{\text{b}}\|^2_{\mathbf{B}_i^{-1}} + \frac{1}{2}\|\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i\|^2_{\mathbf{R}_i^{-1}}. \tag{2.4a}$$

When both the background and observation errors are Gaussian the function (2.4a) equals the negative logarithm of the posterior probability density function given by Bayes' theorem. While in principle a different background covariance matrix should be used at each assimilation time, in practice the same matrix is reused throughout the assimilation window, $\mathbf{B}_i = \mathbf{B}$, $i = 1, \ldots, N$. The 3D-Var analysis is the maximum a posteriori (MAP) estimator, and is computed as the state that minimizes (2.4a)

$$\mathbf{x}_i^{\text{a}} = \arg\min_{\mathbf{x}_i} \mathcal{J}(\mathbf{x}_i). \tag{2.4b}$$

2.2. Four-dimensional variational data assimilation. Strong-constraint 4D-Var processes all the observations within the assimilation window simultaneously. The control parameters are typically the initial conditions $\mathbf{x}_0$, which uniquely determine the state of the system at all future times under the assumption that the model (2.1) perfectly represents reality. The background state is the prior best estimate of the initial conditions $\mathbf{x}_0^{\mathrm{b}}$, and has an associated initial background error covariance matrix $\mathbf{B}_0$. The 4D-Var problem provides the estimate $\mathbf{x}_0^{\mathrm{a}}$ of the true initial conditions as the solution of the following optimization problem

$$\mathbf{x}_0^{\mathrm{a}} = \arg\min_{\mathbf{x}_0} \; \mathcal{J}(\mathbf{x}_0) \qquad \text{subject to (2.1)}, \tag{2.5a}$$

$$\mathcal{J}(\mathbf{x}_0) = \frac{1}{2}\|\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\|^2_{\mathbf{B}_0^{-1}} + \frac{1}{2}\sum_{i=1}^{N}\|\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i\|^2_{\mathbf{R}_i^{-1}}. \tag{2.5b}$$

The first term of the sum (2.5b) quantifies the departure of the solution $\mathbf{x}_0$ from the background state $\mathbf{x}_0^{\mathrm{b}}$ at the initial time t_0. The second term measures the mismatch between the forecast trajectory (model solutions $\mathbf{x}_i$) and observations $\mathbf{y}_i$ at all times t_i in the assimilation window. The covariance matrices $\mathbf{B}_0$ and $\mathbf{R}_i$ need to be predefined, and their quality influences the accuracy of the resulting analysis.

In this paper we focus on the strong-constraint 4D-Var formulation (2.5a). The minimizer of (2.5a) is computed iteratively using gradient-based numerical optimization methods. First-order adjoint models provide the gradient of the cost function [8], while second-order adjoint models provide the Hessian vector product (e.g., for Newton-type methods) [41]. The methodology for building and using various adjoint models for optimization, sensitivity analysis, and uncertainty quantification is discussed in [9, 40]. Various strategies to improve the the 4D-Var data assimilation system are described in [10]. The procedure to estimate the impact of observation and model errors is developed in [35, 36]. A framework to perform derivative-free variational data assimilation using the trust-region framework is given in [38].

2.3. The ensemble Kalman filter. The uncertain background state at time t_i is assumed to be normally distributed, $\mathbf{x}_i \sim \mathcal{N}\left(\overline{\mathbf{x}}_i^{\mathrm{b}}, \mathbf{P}_i^{\mathrm{b}}\right)$. Given the new observations $\mathbf{y}_i$ at time t_i, the Kalman filter computes the Kalman gain matrix, the analysis state (i.e., the posterior mean), and the analysis covariance matrix (i.e., the posterior covariance matrix) using the following formulas, respectively:

$$\mathbf{K}_i = \mathbf{P}_i^{\mathrm{b}}\,\mathbf{H}_i^{\mathrm{T}}\left(\mathbf{H}_i\,\mathbf{P}_i^{\mathrm{b}}\,\mathbf{H}_i^{\mathrm{T}} + \mathbf{R}_i\right)^{-1}, \tag{2.6a}$$

$$\overline{\mathbf{x}}_i^{\mathrm{a}} = \overline{\mathbf{x}}_i^{\mathrm{b}} + \mathbf{K}_i\,\left(\mathbf{y}_i - \mathcal{H}(\overline{\mathbf{x}}_i^{\mathrm{b}})\right), \tag{2.6b}$$

$$\mathbf{P}_i^{\mathrm{a}} = \left(\mathbf{I} - \mathbf{K}_i\,\mathbf{H}_i\right)\,\mathbf{P}_i^{\mathrm{b}}, \tag{2.6c}$$

where $\mathbf{H}$ is the Jacobian of the observation operator.

The ensemble Kalman filter describes the background probability density at time t_i by an ensemble of $\mathrm{N_{ens}}$ forecast states $\{\mathbf{x}_i^{\mathrm{b}\langle\ell\rangle}\}_{\ell=1,\ldots,\mathrm{N_{ens}}}$. The background mean and covariance are estimated from the ensemble:

$$\overline{\mathbf{x}}_i^{\mathrm{b}} \approx \frac{1}{\mathrm{N_{ens}}}\sum_{\ell=1}^{\mathrm{N_{ens}}}\mathbf{x}_i^{\mathrm{b}\langle\ell\rangle}, \qquad \mathbf{P}_i^{\mathrm{b}} \approx \frac{1}{\mathrm{N_{ens}}-1}\,\mathbf{X}_i^{\mathrm{b}}\,(\mathbf{X}_i^{\mathrm{b}})^T, \tag{2.7}$$

where the matrix of state deviations from the mean is

$$\mathbf{X}_i^{\mathrm{b}} = \left[\mathbf{x}_i^{\mathrm{b}\{1\}} - \overline{\mathbf{x}}_i^{\mathrm{b}}, \ldots, \mathbf{x}_i^{\mathrm{b}\{\mathrm{N_{ens}}\}} - \overline{\mathbf{x}}_i^{\mathrm{b}}\right] \approx \frac{1}{\sqrt{\mathrm{N_{ens}}-1}}\left(\mathbf{P}_i^{\mathrm{b}}\right)^{1/2}.$$

The background observations and their deviations from the mean are defined as

$$\overline{\mathbf{h}}_i^{\mathrm{b}} = \frac{1}{\mathrm{N_{ens}}} \sum_{\ell=1}^{\mathrm{N_{ens}}} \mathcal{H}\left(\mathbf{x}_i^{\mathrm{b}\langle\ell\rangle}\right), \qquad \mathbf{Y}_i^{\mathrm{b}} = \left[\mathcal{H}\left(\mathbf{x}_i^{\mathrm{b}\{1\}}\right) - \overline{\mathbf{h}}_i^{\mathrm{b}}, \ldots, \mathcal{H}\left(\mathbf{x}_i^{\mathrm{b}\{\mathrm{N_{ens}}\}}\right) - \overline{\mathbf{h}}_i^{\mathrm{b}}\right].$$

Any vector in the subspace spanned by the ensemble members can be represented as

$$\mathbf{x} = \overline{\mathbf{x}}_i^{\mathrm{b}} + \mathbf{X}_i^{\mathrm{b}}\, w.$$

The ensemble Kalman filter makes the approximations (2.7) and computes the Kalman gain matrix (2.6a) as follows:

$$\mathbf{K}_i = \mathbf{X}_i^{\mathrm{b}}\, \mathbf{S}_i\, (\mathbf{Y}_i^{\mathrm{b}})^{\mathrm{T}}\, \mathbf{R}_i^{-1}, \tag{2.8a}$$

$$\mathbf{S}_i := \left((\mathrm{N_{ens}}-1)\mathbf{I} + \left(\mathbf{Y}_i^{\mathrm{b}}\right)^T\, \mathbf{R}_i^{-1}\, \mathbf{Y}_i^{\mathrm{b}}\right)^{-1}. \tag{2.8b}$$

The mean analysis (2.6b) is

$$\overline{\mathbf{x}}_i^{\mathrm{a}} = \overline{\mathbf{x}}_i^{\mathrm{b}} + \mathbf{X}_i^{\mathrm{b}}\, \mathbf{S}_i\, (\mathbf{Y}_i^{\mathrm{b}})^{\mathrm{T}}\, \mathbf{R}_i^{-1}\, \left(\mathbf{y}_i - \mathcal{H}(\overline{\mathbf{x}}_i^{\mathrm{b}})\right). \tag{2.9}$$

Using $\overline{\mathbf{x}}_i^{\mathrm{a}} = \overline{\mathbf{x}}_i^{\mathrm{b}} + \mathbf{X}_i^{\mathrm{b}}\, \overline{w}_i^{\mathrm{a}}$ we have the mean analysis in ensemble space:

$$\overline{w}_i^{\mathrm{a}} = \mathbf{S}_i\, (\mathbf{Y}_i^{\mathrm{b}})^{\mathrm{T}}\, \mathbf{R}_i^{-1}\, \left(\mathbf{y}_i - \mathcal{H}(\overline{\mathbf{x}}_i^{\mathrm{b}})\right). \tag{2.10}$$

The analysis error covariance (2.6c) becomes

$$\mathbf{P}_i^{\mathrm{a}} = \mathbf{X}_i^{\mathrm{b}}\, \mathbf{S}_i\, (\mathbf{X}_i^{\mathrm{b}})^T. \tag{2.11}$$

Ensemble square root filter (EnSRF). The 3D-Var cost function (2.4a) is formulated with the background error covariance $\mathbf{B}_i$ replaced by the forecast ensemble covariance matrix $\mathbf{P}_i^{\mathrm{b}}$ (2.7). It can be shown that the analysis mean weights (2.10) are the minimizer of [24]:

$$\overline{w}_i^{\mathrm{a}} := \arg\min_w\ \mathcal{J}(w), \quad \mathcal{J}(w) = (\mathrm{N_{ens}}-1)\, \|w\|_2^2 + \|\mathcal{H}\left(\overline{\mathbf{x}}_i^{\mathrm{b}} + \mathbf{X}_i^{\mathrm{b}}\, w\right) - \mathbf{y}_i\|_{\mathbf{R}_i^{-1}}^2. \tag{2.12}$$

The local transform ensemble Kalman filter [24] computes the symmetric square root of the matrix (2.8b)

$$\mathbf{S}_i = (\mathrm{N_{ens}}-1)^{-1}\, \mathbf{W}_i\, \mathbf{W}_i^T, \tag{2.13}$$

and obtains the analysis ensemble weights by adding columns of the factors to the mean analysis weight:

$$w_i^{\mathrm{a}\langle\ell\rangle} = \overline{w}_i^{\mathrm{a}} + \mathbf{W}_i(:,\ell), \qquad \mathbf{x}_i^{\mathrm{a}\langle\ell\rangle} = \overline{\mathbf{x}}_i^{\mathrm{b}} + \mathbf{X}_i^{\mathrm{b}}\, w_i^{\mathrm{a}\langle\ell\rangle}. \tag{2.14}$$

Traditional ensemble Kalman filter (EnKF). In the traditional version of EnKF the filter (2.9) is applied to each ensemble member:

$$\mathbf{x}_i^{\mathrm{a}\langle\ell\rangle} = \mathbf{x}_i^{\mathrm{b}\langle\ell\rangle} + \mathbf{X}_i^{\mathrm{b}}\, \mathbf{S}_i\, (\mathbf{Y}_i^{\mathrm{b}})^{\mathrm{T}}\, \mathbf{R}_i^{-1}\, \left(\mathbf{y}_i^{\langle\ell\rangle} - \mathcal{H}(\mathbf{x}_i^{\mathrm{b}\langle\ell\rangle})\right), \quad \ell = 1, \ldots, \mathrm{N_{ens}},$$

where $\mathbf{y}_i^{\langle\ell\rangle}$ are perturbed observations. This leads to the analysis weights

$$\begin{aligned} w_i^{\mathrm{a}\langle\ell\rangle} &= w_i^{\mathrm{b}\langle\ell\rangle} + \mathbf{S}_i\, (\mathbf{Y}_i^{\mathrm{b}})^{\mathrm{T}}\, \mathbf{R}_i^{-1}\, \left(\mathbf{y}_i^{\langle\ell\rangle} - \mathcal{H}(\mathbf{x}_i^{\mathrm{b}\langle\ell\rangle})\right) \\ &\approx \mathbf{S}_i\, \left((\mathrm{N_{ens}}-1)\, w_i^{\mathrm{b}\langle\ell\rangle} + (\mathbf{Y}_i^{\mathrm{b}})^{\mathrm{T}}\, \mathbf{R}_i^{-1}\, \left(\mathbf{y}_i^{\langle\ell\rangle} - \mathcal{H}(\overline{\mathbf{x}}_i^{\mathrm{b}})\right)\right), \end{aligned} \tag{2.15}$$

where the second relation comes from linearizing the observation operator about the background mean state.

A comparison of (2.15) with (2.10) reveals that the classical EnKF solution (2.15) obtains the weights of the ℓth ensemble member by minimizing the cost function:

$$
\begin{aligned}
w_i^{\mathrm{a}\langle\ell\rangle} &= \arg\min_{w} \; \mathcal{J}^{\langle\ell\rangle}(w), \\
\mathcal{J}^{\langle\ell\rangle}(w) &= (\mathrm{N_{ens}}-1)\,\|w - w_i^{\mathrm{b}\langle\ell\rangle}\|_2^2 + \|\mathcal{H}\left(\overline{\mathbf{x}}_i^{\mathrm{b}} + \mathbf{X}_i^{\mathrm{b}}\, w\right) - \mathbf{y}_i^{\langle\ell\rangle}\|_{\mathbf{R}_i^{-1}}^2.
\end{aligned}
\tag{2.16}
$$

The paper develops next a systematic framework to perform robust data assimilation using Huber and L_1 norms. This will ensure that important information coming from outliers is not rejected but used during the quality control.

3. Robust 3D-Var data assimilation.

3.1. L_1 norm 3D-Var. Consider the 3D-Var cost function in (2.4a) that penalizes the discrepancy between the model state $\mathbf{x}_i$ and observations at time t_i, together with the departure of the state from the model forecast $\mathbf{x}_i^{\mathrm{b}}$. Function (2.4a) represents the negative log-likelihood posterior probability density under the assumption that both the background and the observation errors are Gaussian. In particular, the scaled innovation vector

$$
\mathbf{z}_i := \mathbf{R}_i^{-1/2}\,\left[\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i\right],
\tag{3.1}
$$

which represents scaled observation errors, is assumed to have a normal distribution $\mathbf{z}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ (each observation error $(\mathbf{z}_i)_\ell$ is independent and has a standard normal distribution). We are interested in the case where some observation errors have large deviations. To model this we assume that the scaled observation errors (3.1) are independent, and each of them has a univariate Laplace distribution:

$$
\mathcal{P}(z) = (2\,\xi)^{-1}\,\exp\left(-\xi^{-1}\,|z-\zeta|\right), \quad \mathtt{E}\,[z] = 0, \quad \mathtt{Var}\,[z] = 2\,\xi^2.
$$

The univariate Laplace distribution models an exponential decay on both sides away from the mean ζ. Assuming that observation errors are unbiased ($\zeta = 0$) and independent the probability distribution of the scaled innovation vector $\mathbf{z}_i$ is

$$
\mathcal{P}(\mathbf{z}_i) \propto \exp\left(-\xi^{-1}\,\|\mathbf{z}_i\|_1\right), \quad \mathtt{E}\,[\mathbf{z_i}] = \mathbf{0}, \quad \mathtt{Cov}\,[\mathbf{z_i}] = 2\,\xi^2\,\mathbf{I}.
\tag{3.2}
$$

Under these assumptions the negative log-likelihood posterior function (2.4a) measures the discrepancy of the model state with observations in the L_1 norm. This leads to the following revised cost function:

$$
\mathcal{J}(\mathbf{x}_i) = \frac{1}{2}\|\mathbf{x}_i - \mathbf{x}_i^{\mathrm{b}}\|_{\mathbf{B}_i^{-1}}^2 + \frac{1}{\xi}\left\|\mathbf{R}_i^{-1/2}\,\left[\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i\right]\right\|_1 .
\tag{3.3}
$$

In this paper we choose $\xi = 2$, which leads to a cost function (3.3) that is similar to (2.4a). This translates to an assumed variance of each observation error component equal to eight. This is no restriction of generality: the solution process discussed here can be applied to any value of ξ. For example, one can choose $\xi = 1/\sqrt{2}$ for an error component variance of one.

Following the ADMM [7] we use the scaled innovation vector (3.1) to obtain the L_1-3D-Var problem:

$$
\begin{aligned}
&\min\; \mathcal{J}(\mathbf{x}_i, \mathbf{z}_i) = \frac{1}{2}\|\mathbf{x}_i - \mathbf{x}_i^{\mathrm{b}}\|_{\mathbf{B}_i^{-1}}^2 + \frac{1}{2}\|\mathbf{z}_i\|_1 \\
&\text{subject to} \quad \mathbf{z}_i = \mathbf{R}_i^{-1/2}\,[\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i].
\end{aligned}
\tag{3.4}
$$

The augmented Lagrangian equation for (3.4) is given by

$$\mathcal{L} = \frac{1}{2}\|\mathbf{x}_i - \mathbf{x}_i^{\mathrm{b}}\|^2_{\mathbf{B}_i^{-1}} + \frac{1}{2}\|\mathbf{z}_i\|_1 - \boldsymbol{\lambda}_i^T\left[\mathbf{R}_i^{-1/2}[\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i] - \mathbf{z}_i\right] + \frac{\mu}{2}\|\mathbf{R}_i^{-1/2}[\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i] - \mathbf{z}_i\|_2^2, \tag{3.5}$$

where $\boldsymbol{\lambda}_i$ is the Lagrange multiplier and μ is the penalty parameter (a positive number). Using the basic relation between the norm and the dot product

$$\frac{\mu}{2}\left\|\mathbf{R}_i^{-1/2}[\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i] - \mathbf{z}_i - \frac{\boldsymbol{\lambda}_i}{\mu}\right\|_2^2 = \frac{\mu}{2}\left\|\mathbf{R}_i^{-1/2}[\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i] - \mathbf{z}_i\right\|_2^2 + \frac{1}{2\mu}\|\boldsymbol{\lambda}_i\|_2^2 - \boldsymbol{\lambda}_i^T\left(\mathbf{R}_i^{-1/2}[\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i] - \mathbf{z}_i\right), \tag{3.6}$$

the augmented Lagrangian (3.5) can be written as

$$\mathcal{L} = \frac{1}{2}\|\mathbf{x}_i - \mathbf{x}_i^{\mathrm{b}}\|^2_{\mathbf{B}_i^{-1}} + \frac{1}{2}\|\mathbf{z}_i\|_1 - \frac{1}{2\mu}\|\boldsymbol{\lambda}\|_2^2 + \frac{\mu}{2}\left\|\mathbf{R}_i^{-1/2}[\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i] - \mathbf{z}_i - \frac{\boldsymbol{\lambda}_i}{\mu}\right\|_2^2. \tag{3.7}$$

The cost function (3.7) can be iteratively minimized as follows:

Step 1: Initialize $\mathbf{x}_i^{\{0\}} = \mathbf{x}_i^{\mathrm{b}}$, $\mathbf{z}_i^{\{0\}} = \mathbf{R}_i^{-1/2}\left[\mathcal{H}(\mathbf{x}_i^{\mathrm{b}}) - \mathbf{y}_i\right]$, $\boldsymbol{\lambda}_i^{\{0\}} = \mathbf{0}$, $\mu^{\{0\}} = 1$

Step 2: Start with $\mathbf{x}_i^{\{0\}}$, $\mathbf{z}_i^{\{0\}}$, $\boldsymbol{\lambda}_i^{\{0\}}$, and $\mu^{\{0\}}$.

Step 3: Fix $\mathbf{z}_i^{\{k\}}$, $\boldsymbol{\lambda}_i^{\{k\}}$, and $\mu^{\{k\}}$, and solve

$$\mathbf{x}_i^{\{k+1\}} := \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{x} - \mathbf{x}_i^{\mathrm{b}}\|^2_{\mathbf{B}_i^{-1}} + \frac{\mu^{\{k\}}}{2}\left\|\mathbf{R}_i^{-1/2}[\mathcal{H}(\mathbf{x}) - \mathbf{y}_i] - \mathbf{z}_i^{\{k\}} - \frac{\boldsymbol{\lambda}_i^{\{k\}}}{\mu^{\{k\}}}\right\|_2^2. \tag{3.8a}$$

To solve (3.8a) we carry out a regular L_2-3D-Var minimization of the form (2.4)

$$\mathbf{x}_i^{\{k+1\}} := \arg\min_{\mathbf{x}} \frac{1}{2}\left\|\mathbf{x} - \mathbf{x}_i^{\mathrm{b}}\right\|^2_{\mathbf{B}_i^{-1}} + \frac{1}{2}\left\|\mathcal{H}(\mathbf{x}) - \mathbf{y}_i^{\{k\}}\right\|^2_{\mathbf{R}_i^{\{k\}-1}}, \tag{3.8b}$$

but with modified observations and a scaled covariance matrix.

$$\mathbf{y}_i^{\{k\}} := \mathbf{y}_i + \mathbf{R}_i^{1/2}\left(\mathbf{z}_i^{\{k\}} + \frac{\boldsymbol{\lambda}_i^{\{k\}}}{\mu^{\{k\}}}\right), \quad \mathbf{R}_i^{\{k\}} := \mathbf{R}_i/\mu^{\{k\}}, \tag{3.8c}$$

Step 4: Fix $\mathbf{x}_i^{\{k+1\}}$, $\boldsymbol{\lambda}_i^{\{k\}}$, and $\mu^{\{k\}}$, and solve

$$\mathbf{z}_i^{\{k+1\}} := \arg\min_{\mathbf{z}} \quad \|\mathbf{z}\|_1 + \frac{\mu^{\{k\}}}{2}\left\|\mathbf{d}_i^{\{k+1\}} - \mathbf{z} - \frac{\boldsymbol{\lambda}_i^{\{k\}}}{\mu^{\{k\}}}\right\|_2^2, \tag{3.9a}$$

where

$$\mathbf{d}_i^{\{k+1\}} := \mathbf{R}_i^{-1/2}\left[\mathcal{H}\left(\mathbf{x}_i^{\{k+1\}}\right) - \mathbf{y}_i\right]. \tag{3.9b}$$

The above minimization subproblem can be solved by using the shrinkage procedure defined in Algorithm 1:

$$\mathbf{z}_i^{\{k+1\}} := \textsc{LONEShrinkage}\left(\mu^{\{k\}}; \mathbf{d}_i^{\{k+1\}}; \boldsymbol{\lambda}_i^{\{k\}}\right). \tag{3.9c}$$

Step 5: Update $\boldsymbol{\lambda}_i$:

$$\boldsymbol{\lambda}_i^{\{k+1\}} := \boldsymbol{\lambda}_i^{\{k\}} - \mathbf{d}_i^{\{k+1\}} + \mathbf{z}_i^{\{k+1\}}.$$

Step 6: Update μ:

$$\mu^{\{k+1\}} := \rho\, \mu^{\{k\}}, \qquad \rho > 1.$$

Algorithm 1. L_1_Shrinkage.

procedure $\mathbf{z}$=LONESHRINKAGE($\mu; \mathbf{d}; \boldsymbol{\lambda}$)
  Input: $[\mu \in \mathbb{R};\ \mathbf{d} \in \mathbb{R}^m;\ \boldsymbol{\lambda} \in \mathbb{R}^m]$
  Output: $[\mathbf{z} \in \mathbb{R}^m]$
  $\mathbf{z} := \max\{\, \|\mathbf{d} - \boldsymbol{\lambda}/\mu\|_2 - 1/\mu\,,\, 0\,\} \cdot \dfrac{\mathbf{d} - \boldsymbol{\lambda}/\mu}{\|\mathbf{d} - \boldsymbol{\lambda}/\mu\|_2}.$
end procedure

3.2. Hubernorm 3D-Var. Using the L_1 norm throughout spoils the smoothness properties near the mean. The smoothness property of the L_2 norm near the mean is highly desirable and can be retained by using the Huber norm. The Huber norm treats the errors using the L_2 norm in the vicinity of the mean, and using the L_1 norm far from the mean [28]. From a statistical perspective, small scaled observation errors (3.1) are assumed to have independent standard normal distributions, while large scaled observation errors are assumed to have independent Laplace distributions (3.2). The good properties of the traditional 3D-Var are preserved when the data are reasonable, but outliers are treated with the more robust L_1 norm. The Huber norm formulation gives larger weights to outliers as compared to other robust estimators [23] .

The 3D-Var problem in the Huber norm reads

$$\begin{aligned} &\min\ \mathcal{J}(\mathbf{x}_i, \mathbf{z}_i) = \frac{1}{2}\|\mathbf{x}_i - \mathbf{x}_i^{\mathrm{b}}\|^2_{\mathbf{B}_i^{-1}} + \frac{1}{2}\|\mathbf{z}_i\|_{\mathrm{HUB}} \\ &\text{subject to} \quad \mathbf{z}_i = \mathbf{R}_i^{-1/2}\,[\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i], \end{aligned} \tag{3.10}$$

where

$$\|\mathbf{z}_i\|_{\mathrm{HUB}} = \sum_{\ell=1}^{m} g_{\mathrm{HUB}}(\mathbf{z}_{i,\ell}), \qquad g_{\mathrm{HUB}}(a) = \begin{cases} a^2/2, & |a| \le \tau, \\ \tau|a| - \tau^2/2, & |a| > \tau. \end{cases} \tag{3.11}$$

Here $\mathbf{z}_{i,\ell}$ is the ℓth entry of the vector $\mathbf{z}_i$. The threshold τ represents the number of standard deviations after which we switch from the L_2 to L_1 norm.

3.2.1. ADMM solution for Huber-3D-Var. The Huber norm minimization problem (3.10) can be solved using the ADMM framework in a manner similar to the one described in section 3.1. The only difference is in Step 3: after fixing $\mathbf{x}_i^{\{k+1\}}$, $\boldsymbol{\lambda}_i^{\{k\}}$, and $\mu^{\{k\}}$ we do not solve the L_1 norm problem (3.9a). Rather, we solve the Huber norm problem

$$\mathbf{z}_i^{\{k+1\}} := \arg\min_{\mathbf{z}} \quad \|\mathbf{z}\|_{\mathrm{HUB}} + \frac{\mu^{\{k\}}}{2}\left\| \mathbf{d}_i^{\{k+1\}} - \mathbf{z} - \frac{\boldsymbol{\lambda}_i^{\{k\}}}{\mu^{\{k\}}} \right\|_2^2, \tag{3.12}$$

Algorithm 2. Huber_Shrinkage.

procedure $\mathbf{z}$=HUBERSHRINKAGE($\mu; \mathbf{d}; \boldsymbol{\lambda}$)
  Input: $[\mu \in \mathbb{R};\ \mathbf{d} \in \mathbb{R}^m;\ \boldsymbol{\lambda} \in \mathbb{R}^m]$
  Output: $[\mathbf{z} \in \mathbb{R}^m]$
  for $\ell = 1, 2, \ldots, m$ **do**
    if $|\mathbf{d}_\ell| \geq \tau$ **then**
      $\mathbf{z}_\ell = \max\{\|\mathbf{d} - \boldsymbol{\lambda}/\mu\|_2 - 1/\mu, 0\} \cdot \dfrac{\mathbf{d}_\ell - \boldsymbol{\lambda}_\ell/\mu}{\|\mathbf{d} - \boldsymbol{\lambda}/\mu\|_2}.$
    else
      $\mathbf{z}_\ell = \dfrac{\mu}{1+\mu}(\mathbf{d}_\ell - \boldsymbol{\lambda}_\ell/\mu)$
    end if
  end for
end procedure

where $\mathbf{d}_i^{\{k+1\}}$ is defined in (3.9b). The closed-form solution of (3.12) is given by the Huber shrinkage procedure defined in Algorithm 2:

$$\mathbf{z}_i^{\{k+1\}} := \text{HUBERSHRINKAGE}\left(\mu^{\{k\}};\, \mathbf{d}_i^{\{k+1\}};\, \boldsymbol{\lambda}_i^{\{k\}}\right).$$

Algorithms 1 *and* 2 *describe the iterative shrinkage procedure to minimize nonsmooth functions (e.g.,* L_1 *norms). This has been independently proposed by* [5, 11, 14, 15, 16, 17, 21, 34]. *Various modifications and enhancements have been applied to Algorithms* 1 *and* 2 [6, 45, 47] *and for ease of exposition, we use a basic version in this paper.*

3.2.2. Half-quadratic optimization solution for Huber-3D-Var. An alternative approach to carry out the minimization (3.10) is to reformulate it as a multiplicative half-quadratic optimization problem [33]. We construct an augmented cost function which involves the auxiliary variable $\mathbf{u}_i \in \mathbb{R}^m$,

$$\mathcal{A}(\mathbf{x}_i, \mathbf{u}_i) = \frac{1}{2}\|\mathbf{x}_i - \mathbf{x}_i^{\mathrm{b}}\|^2_{\mathbf{B}_i^{-1}} + \frac{1}{2}\sum_{\ell=1}^{m}\big(\mathbf{u}_{i,\ell}\,\mathbf{z}_{i,\ell}^2/2 + \psi(\mathbf{u}_{i,\ell})\big), \tag{3.13a}$$

where ψ is a dual potential function determined by using the theory of convex conjugacy such that

$$\mathcal{J}(\mathbf{x}_i) = \min_{\mathbf{u}_i} \mathcal{A}(\mathbf{x}_i, \mathbf{u}_i). \tag{3.13b}$$

The minimizer of $\mathcal{A}$ is calculated using alternate minimization. Let the solution at the end of iteration $\{k-1\}$ read $(\mathbf{x}_i^{\{k\}}, \mathbf{u}_i^{\{k\}})$. At iteration k we proceed as follows.

Step 1: First calculate $\mathbf{u}_i^{\{k+1\}}$, where $\mathbf{x}_i^{\{k\}}$ is fixed, such that

$$\mathcal{A}\left(\mathbf{x}_i^{\{k\}}, \mathbf{u}_i^{\{k+1\}}\right) \leq \mathcal{A}\left(\mathbf{x}_i^{\{k\}}, \mathbf{u}\right) \quad \forall\, \mathbf{u}\,.$$

This amounts to finding $\mathbf{u}_i^{\{k+1\}}$ according to

$$\mathbf{u}_i^{\{k+1\}} = \sigma\left(\mathbf{R}_i^{-1/2}\,\left[\mathcal{H}(\mathbf{x}_i^{\{k\}}) - \mathbf{y}_i\right]\right) \tag{3.14a}$$

(where the function σ is applied to each element of the argument vector). For Huber regularization

$$\sigma(a) = \begin{cases} 1, & |a| \leq \tau, \\ \tau/|a|, & |a| > \tau. \end{cases} \tag{3.14b}$$

Step 2: Next, calculate $\mathbf{x}_i^{\{k+1\}}$, where $\mathbf{u}_i^{\{k+1\}}$ is fixed, such that

$$\mathcal{A}\left(\mathbf{x}_i^{\{k+1\}}, \mathbf{u}_i^{\{k+1\}}\right) \leq \mathcal{A}\left(\mathbf{x}, \mathbf{u}_i^{\{k+1\}}\right) \quad \forall\, \mathbf{x}\,.$$

This is achieved by solving the following minimization problem:

$$\begin{aligned} \mathbf{x}_i^{\{k+1\}} = \arg\min_{\mathbf{x}_i} \; & \frac{1}{2}\|\mathbf{x}_i - \mathbf{x}_i^{\mathrm{b}}\|^2_{\mathbf{B}_i^{-1}} \\ & + \frac{1}{2}\left(\mathbf{R}_i^{-1/2}\left[\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i\right]\right)^T \operatorname{diag}\left(\mathbf{u}_i^{\{k+1\}}/2\right)\left(\mathbf{R}_i^{-1/2}\left[\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i\right]\right). \end{aligned}$$

This minimization problem is equivalent to solving a regular L_2-3D-Var problem

$$\mathbf{x}_i^{\{k+1\}} = \arg\min_{\mathbf{x}_i} \; \frac{1}{2}\|\mathbf{x}_i - \mathbf{x}_i^{\mathrm{b}}\|^2_{\mathbf{B}_i^{-1}} + \|\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i\|^2_{\mathbf{R}_i^{\{k+1\}\,-1}} \tag{3.15}$$

with a modified observation error covariance

$$\mathbf{R}_i^{\{k+1\}} := \mathbf{R}_i^{1/2} \cdot \operatorname{diag}\left(2/\mathbf{u}_i^{\{k+1\}}\right) \cdot \mathbf{R}_i^{1/2}. \tag{3.16}$$

For the case where $\mathbf{R}_i$ is diagonal, (3.16) divides each variance by $\mathbf{u}_i^{\{k+1\}}$. The smaller the entry $\mathbf{u}_{i,\ell}^{\{k+1\}}$ is the lower the weight given to observation ℓ in the 3D-Var data assimilation procedure becomes.

4. Robust 4D-Var.

4.1. L_1-4D-Var. Given the background value of the initial state $\mathbf{x}_0^{\mathrm{b}}$, the covariance of the initial background errors $\mathbf{B}_0$, the observations $\mathbf{y}_i$ at t_i, and the corresponding observation error covariances $\mathbf{R}_i$ $(i = 1, 2, \ldots, N)$, the L_1-4D-Var problem looks for the MAP estimate $\mathbf{x}_0$ of the true initial conditions by solving the following constrained optimization problem:

$$\begin{aligned} \min_{\mathbf{x}_0} \; \mathcal{J}(\mathbf{x}_0) \; &:= \frac{1}{2}\|\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\|^2_{\mathbf{B}_0^{-1}} + \frac{1}{2}\sum_{i=1}^{N}\|\mathbf{z}_i\|_1 \\ \text{subject to} \; & \mathbf{z}_i = \mathbf{R}_i^{-1/2}[\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i], \quad i = 1, 2, \ldots, N, \\ & \mathbf{x}_i = \mathcal{M}_{i-1,i}(\mathbf{x}_{i-1}), \quad i = 1, 2, \ldots, N. \end{aligned} \tag{4.1}$$

The augmented Lagrangian associated with (4.1) reads

$$\begin{aligned} \mathcal{L} = & \frac{1}{2}\left\|\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right\|^2_{\mathbf{B}_0^{-1}} + \frac{1}{2}\sum_{\mathrm{i}=1}^{\mathrm{N}}\|\mathbf{z}_i\|_1 \\ & - \sum_{\mathrm{i}=1}^{\mathrm{N}}\theta_i^T\left(\mathbf{x}_i - \mathcal{M}_{i-1,i}(\mathbf{x}_{i-1})\right) + \frac{\nu}{2}\sum_{i=1}^{\mathrm{N}}\|\mathbf{x}_i - \mathcal{M}_{i-1,i}(\mathbf{x}_{i-1})\|^2_{\mathbf{P}_i^{-1}} \\ & - \sum_{\mathrm{i}=1}^{\mathrm{N}}\boldsymbol{\lambda}_i^T\left(\mathbf{R}_i^{-1/2}[\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i] - \mathbf{z}_i\right) + \frac{\mu}{2}\sum_{i=1}^{\mathrm{N}}\left\|\mathbf{R}_i^{-1/2}[\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i] - \mathbf{z}_i\right\|_2^2, \end{aligned} \tag{4.2}$$

where $\mathbf{P}_i$'s are model error scaling matrices.

Using the identity (3.6) the Lagrangian (4.2) becomes

$$\begin{aligned}(4.3)\qquad \mathcal{L} = {} & \frac{1}{2}\left\|\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right\|^2_{\mathbf{B}_0^{-1}} + \frac{1}{2}\sum_{\mathrm{i}=1}^{\mathrm{N}}\|\mathbf{z}_i\|_1 \\ & - \sum_{\mathrm{i}=1}^{\mathrm{N}}\theta_i^T\left(\mathbf{x}_i - \mathcal{M}_{i-1,i}(\mathbf{x}_{i-1})\right) + \frac{\nu}{2}\sum_{i=1}^{\mathrm{N}}\|\mathbf{x}_i - \mathcal{M}_{i-1,i}(\mathbf{x}_{i-1})\|^2_{\mathbf{P}_i^{-1}} \\ & + \frac{\mu}{2}\sum_{\mathrm{i}=1}^{\mathrm{N}}\left\|\mathbf{R}_i^{-1/2}[\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i] - \mathbf{z}_i - \boldsymbol{\lambda}_i/\mu\right\|_2^2 - \frac{1}{2\mu}\sum_{\mathrm{i}=1}^{\mathrm{N}}\|\boldsymbol{\lambda}_i\|_2^2.\end{aligned}$$

The problem (4.1) can be solved by alternating minimization of the Lagrangian (4.3), as follows:

Step 1: Initialize $\mathbf{x}_0^{\{0\}} = \mathbf{x}_0^{\mathrm{b}}$.

Step 2: Run the forward model starting from $\mathbf{x}_0^{\{0\}}$ to obtain $\mathbf{x}_i^{\{0\}}$, $i = 1, \ldots, N$.

Step 3: Set $\mathbf{z}_i^{\{0\}} = \mathbf{R}_i^{-1/2}\left[\mathcal{H}(\mathbf{x}_i^{\{0\}}) - \mathbf{y}_i\right]$, $\boldsymbol{\lambda}_i^{\{0\}} = \mathbf{0}$, for $i = 1, \ldots, N$, $\mu^{\{0\}} = 1$.

Step 4: Fix $\mathbf{z}_i^{\{k\}}$ and $\boldsymbol{\lambda}_i^{\{k\}}$, and minimize the Lagrangian to solve for $\mathbf{x}_0^{\{k+1\}}$:

$$\begin{aligned}\mathbf{x}_0^{\{k+1\}} &:= \arg\min\ \mathcal{L}, \\ \mathcal{L} = {} & \frac{1}{2}\left\|\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right\|^2_{\mathbf{B}_0^{-1}} - \sum_{\mathrm{i}=1}^{\mathrm{N}}\theta_i^T\left(\mathbf{x}_i - \mathcal{M}_{i-1,i}(\mathbf{x}_{i-1})\right) \\ & + \frac{\nu}{2}\sum_{i=1}^{\mathrm{N}}\|\mathbf{x}_i - \mathcal{M}_{i-1,i}(\mathbf{x}_{i-1})\|^2_{\mathbf{P}_i^{-1}} \\ & + \frac{\mu^{\{k\}}}{2}\sum_{\mathrm{i}=1}^{\mathrm{N}}\left\|\mathbf{R}_i^{-1/2}[\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i] - \mathbf{z}_i^{\{k\}} - \boldsymbol{\lambda}_i^{\{k\}}/\mu^{\{k\}}\right\|_2^2.\end{aligned}$$

This is equivalent to solving the traditional L_2-4D-Var problem

$$\begin{aligned}\mathbf{x}_0^{\{k+1\}} &:= \arg\min\ \mathcal{J}(\mathbf{x}_0) = \frac{1}{2}\|\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\|^2_{\mathbf{B}_0^{-1}} + \frac{1}{2}\sum_{i=1}^{N}\|\mathcal{H}(\mathbf{x}_i) - \widetilde{\mathbf{y}}_i\|^2_{\widetilde{\mathbf{R}}_i^{-1}} \\ & \text{subject to} \quad \mathbf{x}_{i+1} = \mathcal{M}_{i,i+1}(\mathbf{x}_i), \quad i = 0, 1, 2, \ldots, N-1,\end{aligned}$$

with the modified observation vectors and observation error covariances (3.8c).

Step 5: Run the forward model starting from $\mathbf{x}_0^{\{k+1\}}$ to obtain $\mathbf{x}_i^{\{k+1\}}$, $i = 1, \ldots, N$.

Step 6: Fix $\mathbf{x}_i^{\{k+1\}}$ and $\boldsymbol{\lambda}_i^{\{k\}}$, and find $\mathbf{z}_i^{\{k+1\}}$, $i = 1, \ldots, N$ as

$$\begin{aligned}\mathbf{z}^{\{k+1\}} &:= \arg\min\ \mathcal{L}, \\ (4.4)\quad \mathcal{L} &= \frac{1}{2}\sum_{\mathrm{i}=1}^{\mathrm{N}}\|\mathbf{z}_i\|_1 + \frac{\mu}{2}\sum_{\mathrm{i}=1}^{\mathrm{N}}\left\|\mathbf{R}_i^{-1/2}\mathcal{H}(\mathbf{x}_i^{\{k+1\}}) - \mathbf{y}_i] - \mathbf{z}_i - \boldsymbol{\lambda}_i^{\{k\}}/\mu^{\{k\}}\right\|_2^2 \\ &= \frac{1}{2}\|\mathbf{z}\|_1 + \frac{\mu}{2}\left\|\boldsymbol{d}^{\{k+1\}} - \boldsymbol{\lambda}^{\{k\}}/\mu^{\{k\}} - \mathbf{z}\right\|_2^2,\end{aligned}$$

where we denote

$$\mathbf{z} = \begin{bmatrix}\mathbf{z}_1 \\ \vdots \\ \mathbf{z}_N\end{bmatrix}, \quad \boldsymbol{\lambda}^{\{k\}} = \begin{bmatrix}\boldsymbol{\lambda}_1^{\{k\}} \\ \vdots \\ \boldsymbol{\lambda}_N^{\{k\}}\end{bmatrix}, \quad \boldsymbol{d}^{\{k+1\}} = \begin{bmatrix}\mathbf{R}_1^{-1/2}[\mathcal{H}(\mathbf{x}_1^{\{k+1\}}) - \mathbf{y}_1] \\ \vdots \\ \mathbf{R}_N^{-1/2}[\mathcal{H}(\mathbf{x}_N^{\{k+1\}}) - \mathbf{y}_N]\end{bmatrix}.$$

Problem (4.4) is solved using Algorithm 1:

$$\mathbf{z}^{\{k+1\}} := \textsc{LoneShrinkage}\left(\mu^{\{k\}};\, \boldsymbol{d}^{\{k+1\}};\, \boldsymbol{\lambda}^{\{k\}}\right). \tag{4.5}$$

Step 7: Update $\boldsymbol{\lambda}$:

$$\boldsymbol{\lambda}_i^{\{k+1\}} := \boldsymbol{\lambda}_i^{\{k\}} - \mathbf{R}_i^{-1/2}\left[\mathcal{H}\left(\mathbf{x}_i^{\{k+1\}}\right) - \mathbf{y}_i\right] + \mathbf{z}_i^{\{k+1\}}, \quad i = 1, \ldots, N.$$

4.2. Huber 4D-Var. The Huber 4D-Var problem reads

$$\begin{aligned} \min_{\mathbf{x}_0}\ & \mathcal{J}(\mathbf{x}_0, \mathbf{z}) := \frac{1}{2}\|\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\|^2_{\mathbf{B}_0^{-1}} + \frac{1}{2}\sum_{i=1}^{N}\|\mathbf{z}_i\|_{\mathrm{HUB}} \\ \text{subject to}\ & \mathbf{z}_i = \mathbf{R}_i^{-1/2}[\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i], \quad i = 1, 2, \ldots, N, \\ & \mathbf{x}_i = \mathcal{M}_{i-1,i}(\mathbf{x}_{i-1}), \quad i = 1, 2, \ldots, N. \end{aligned} \tag{4.6}$$

4.2.1. ADMM solution of Huber-4D-Var. The Huber norm minimization problem (4.6) can be solved using the ADMM framework in a manner similar to the one described in section 4.1. The only difference is in Step 3, where we find $\mathbf{z}$ by solving the Huber norm problem

$$\mathbf{z}^{\{k+1\}} := \arg\min_{\mathbf{z}}\ \mathcal{L} = \frac{1}{2}\left\|\mathbf{z}\right\|_{\mathrm{HUB}} + \frac{\mu^{\{k\}}}{2}\left\|\boldsymbol{d}^{\{k+1\}} - \boldsymbol{\lambda}^{\{k\}}/\mu^{\{k\}} - \mathbf{z}\right\|_2^2. \tag{4.7}$$

The closed-form solution of (4.7) is obtained by the Huber shrinkage procedure defined in Algorithm 2:

$$\mathbf{z}^{\{k+1\}} := \textsc{HuberShrinkage}\left(\mu^{\{k\}};\, \mathbf{d}^{\{k+1\}};\, \boldsymbol{\lambda}^{\{k\}}\right).$$

4.2.2. Half-quadratic optimization solution for Huber-4D-Var. An alternative procedure to solve the 4D-Var problem described in (4.6) is to use the steps described in section 3.2.2. Use (3.14) to update the scaling factors $\mathbf{u}_i^{\{k+1\}}$ for each $i = 1, \ldots, N$. Use (3.16) to compute the scaled observation error covariance matrices $\mathbf{R}_i^{\{k+1\}}$ for each $i = 1, \ldots, N$. Solve a regular L_2-4D-Var problem (2.5) with the modified observation error covariance matrices to obtain the analysis $\mathbf{x}_0^{\{k+1\}}$. Propagate the initial solution in time to obtain the forward solution $\mathbf{x}_i^{\{k+1\}}$, $i = 1, \ldots, N$. Iterate by computing again the new scaling factors $\mathbf{u}_i^{\{k+2\}}$ and so on.

5. Robust data assimilation using the EnKF. Robust EnKF algorithms are obtained by reformulating the equivalent optimization problems (2.12) or (2.16) using $\|\mathcal{H}\left(\overline{\mathbf{x}}_i^{\mathrm{b}} + \mathbf{X}_i^{\mathrm{b}}\, w\right) - \mathbf{y}_i\|_*$, where $\|\cdot\|_*$ stands for either L_1 or Huber norms. The solution of the modified optimization problems relies on repeated applications of the standard EnKF procedure. For traditional EnKF (2.16) a different modified optimization problem has to be solved for each ensemble member. This is computationally expensive. For this reason we discuss below only the robust EnSRF.

5.1. L_1-EnSRF. The L_1-EnSRF problem reads

$$\begin{aligned} \min\ & \mathcal{J}(w_i, \mathbf{z}_i) = (\mathrm{N_{ens}} - 1)\,\|w\|_2^2 + \|\mathbf{z}_i\|_1 \\ \text{subject to}\ & \mathbf{z}_i = \mathbf{R}_i^{-1/2}\,\left[\mathcal{H}\left(\overline{\mathbf{x}}_i^{\mathrm{b}} + \mathbf{X}_i^{\mathrm{b}}\, w_i\right) - \mathbf{y}_i\right]. \end{aligned}$$

The augmented Lagrangian for problem (5.1) is given by

$$\mathcal{L} = (\mathrm{N_{ens}} - 1)\,\|w_i\|_2^2 + \|\mathbf{z}_i\|_1 - \frac{1}{2\mu}\,\|\boldsymbol{\lambda}\|_2^2 + \frac{\mu}{2}\,\left\|\mathbf{R}_i^{-1/2}\,\left[\mathcal{H}\left(\overline{\mathbf{x}}_i^{\mathrm{b}} + \mathbf{X}_i^{\mathrm{b}}\,w_i\right) - \mathbf{y}_i\right] - \mathbf{z}_i - \boldsymbol{\lambda}/\mu\right\|_2^2. \tag{5.1}$$

The problem (5.1) is solved by ADMM as follows:

Step 1: Fix $\mu^{\{k\}}$, $\mathbf{z}_i^{\{k\}}$, and $\boldsymbol{\lambda}_i^{\{k\}}$, and solve

$$w_i^{\{k+1\}} := \arg\min_{w}\ \mathcal{J}(w) = (\mathrm{N_{ens}} - 1)\,\|w\|_2^2 + \frac{\mu^{\{k\}}}{2}\,\left\|\mathbf{R}_i^{-1/2}\,\left[\mathcal{H}\left(\overline{\mathbf{x}}_i^{\mathrm{b}} + \mathbf{X}_i^{\mathrm{b}}\,w\right) - \mathbf{y}_i\right] - \mathbf{z}_i^{\{k\}} - \frac{\boldsymbol{\lambda}_i^{\{k\}}}{\mu^{\{k\}}}\right\|_2^2. \tag{5.2}$$

For a general nonlinear observation operator, (5.2) can be solved by a regular nonlinear minimization in the ensemble space. It is useful to compare (5.2) with the L_2-EnSRF equation (2.12). The weights update, (5.2), can be obtained by applying the EnSRF methodology to a data assimilation problem with the modified observations $\mathbf{y}_i^{\{k\}}$ and the scaled observation error covariance matrix $\mathbf{R}_i^{\{k\}}$ given by (3.8c).

Step 2: Fix $\mu^{\{k\}}$, $w_i^{\{k+1\}}$, and $\boldsymbol{\lambda}_i^{\{k\}}$, and solve for $\mathbf{z}_i$:

$$\mathbf{z}_i^{\{k+1\}} := \arg\min_{\mathbf{z}}\ \mathcal{J}(\mathbf{z}),$$
$$\mathcal{J}(\mathbf{z}) = \|\mathbf{z}\|_1 + \frac{\mu^{\{k\}}}{2}\,\left\|\mathbf{R}_i^{-1/2}\,\left[\mathcal{H}\left(\overline{\mathbf{x}}_i^{\mathrm{b}} + \mathbf{X}_i^{\mathrm{b}}\,w_i^{\{k+1\}}\right) - \mathbf{y}_i\right] - \mathbf{z} - \frac{\boldsymbol{\lambda}^{\{k\}}}{\mu^{\{k\}}}\right\|_2^2. \tag{5.3}$$

The above minimization subproblem is solved by using the shrinkage procedure defined in Algorithm 1:

$$\begin{aligned}\mathbf{d}_i^{\{k+1\}} &:= \mathbf{R}_i^{-1/2}\,\left[\mathcal{H}\left(\overline{\mathbf{x}}_i^{\mathrm{b}} + \mathbf{X}_i^{\mathrm{b}}\,w_i^{\{k+1\}}\right) - \mathbf{y}_i\right],\\ \mathbf{z}^{\{k+1\}} &:= \textsc{LoneShrinkage}\left(\mu^{\{k\}};\,\mathbf{d}_i^{\{k+1\}};\,\boldsymbol{\lambda}^{\{k\}}\right).\end{aligned} \tag{5.4}$$

Step 3: Update $\boldsymbol{\lambda}_i$:

$$\boldsymbol{\lambda}_i^{\{k+1\}} := \boldsymbol{\lambda}_i^{\{k\}} - \mathbf{R}_i^{-1/2}\,\left[\mathcal{H}\left(\overline{\mathbf{x}}_i^{\mathrm{b}} + \mathbf{X}_i^{\mathrm{b}}\,w_i^{\{k+1\}}\right) - \mathbf{y}_i\right] + \mathbf{z}_i^{\{k+1\}}.$$

Assume that after M iterations we achieve satisfactory convergence and stop. The mean analysis provided by the robust EnSRF is given by the optimal weights

$$\overline{w}_i^{\mathrm{a}} := w_i^{\{M\}}. \tag{5.5}$$

The weights of individual analysis ensembles are obtained by applying (2.14) with the modified observation covariance of the last iteration:

$$\begin{aligned}&\left((\mathrm{N_{ens}} - 1)\mathbf{I} + \mu^{\{M\}}\,\left(\mathbf{Y}_i^{\mathrm{b}}\right)^T\,\mathbf{R}_i^{-1}\,\mathbf{Y}_i^{\mathrm{b}}\right)^{-1} = (\mathrm{N_{ens}} - 1)^{-1}\,\mathbf{W}_i^{\{M\}}\,\mathbf{W}_i^{\{M\}\,T},\\ &w_i^{\mathrm{a}\langle\ell\rangle} := \overline{w}_i^{\mathrm{a}} + \mathbf{W}_i^{\{M\}}(:,\ell), \quad \mathbf{x}_i^{\mathrm{a}\langle\ell\rangle} = \overline{\mathbf{x}}_i^{\mathrm{b}} + \mathbf{X}_i^{\mathrm{b}}\,w_i^{\mathrm{a}\langle\ell\rangle}, \quad \ell = 1, \ldots, \mathrm{N_{ens}}.\end{aligned} \tag{5.6}$$

5.2. Huber EnSRF. The Huber EnSRF problem reads

$$\begin{aligned} &\min\ J(w,\mathbf{z}_i) = (\mathrm{N_{ens}}-1)\,\|w\|_2^2 + \|\mathbf{z}_i\|_{\mathrm{HUB}} \\ &\text{subject to}\quad \mathbf{z}_i = \mathbf{R}_i^{-1/2}\,\left[\mathcal{H}\left(\overline{\mathbf{x}}_i^{\mathrm{b}} + \mathbf{X}_i^{\mathrm{b}}\,w\right) - \mathbf{y}_i\right]. \end{aligned} \tag{5.7}$$

The problem (5.7) is solved iteratively by half-quadratic minimization. The minimizer of $\mathcal{A}$ is calculated using alternate minimization. Let the solution at iteration $\{k\}$ be $w_i^{\{k\}}, \mathbf{u}_i^{\{k\}}$. We proceed as follows.

Step 1: First calculate

$$\mathbf{u}_i^{\{k+1\}} = \sigma\left(\mathbf{R}_i^{-1/2}\,\left[\mathcal{H}\left(\overline{\mathbf{x}}_i^{\mathrm{b}} + \mathbf{X}_i^{\mathrm{b}}\,w_i^{\{k\}}\right) - \mathbf{y}_i\right]\right),$$

where the function σ is given by (3.14b).

Step 2: Next, update the state $\mathbf{x}_i^{\{k+1\}} = \overline{\mathbf{x}}_i^{\mathrm{b}} + \mathbf{X}_i^{\mathrm{b}}\,w_i^{\{k+1\}}$ as follows:

$$\begin{aligned} &w_i^{\{k+1\}} = \arg\min_{w}\ \ (\mathrm{N_{ens}}-1)\,\|w\|_2^2 + \|\mathcal{H}\left(\overline{\mathbf{x}}_i^{\mathrm{b}} + \mathbf{X}_i^{\mathrm{b}}\,w\right) - \mathbf{y}_i\|^2_{\mathbf{R}_i^{\{k+1\}\,-1}}, \\ &\text{where}\quad \mathbf{R}_i^{\{k+1\}\,-1} = \mathbf{R}_i^{-1/2}\,\mathrm{diag}\left(\mathbf{u}^{\{k\}}/2\right)\,\mathbf{R}_i^{-1/2}. \end{aligned}$$

This second step requires the application of EnSRF (2.12) with a modified observation covariance matrix. The analysis weights provide the new iterate $w_i^{\{k+1\}} \equiv w_i^{\mathrm{a}}$.

Assume that after M iterations we achieve satisfactory convergence and stop. The mean analysis provided by the robust EnSRF is given by the optimal weights (5.5). The weights of individual analysis ensemble members are obtained from (5.6) with the modified observation covariance of the last iteration:

$$\begin{aligned} &\left((\mathrm{N_{ens}}-1)\mathbf{I} + \left(\mathbf{Y}_i^{\mathrm{b}}\right)^T\,\mathbf{R}_i^{-1/2}\,\mathrm{diag}\left(\mathbf{u}^{\{M\}}/2\right)\,\mathbf{R}_i^{-1/2}\,\mathbf{Y}_i^{\mathrm{b}}\right)^{-1} \\ &\qquad = (\mathrm{N_{ens}}-1)^{-1}\,\mathbf{W}_i^{\{M\}}\,\mathbf{W}_i^{\{M\}\,T}. \end{aligned}$$

5.3. Convergence of ADMM. There are many convergence results discussed in the literature. A set of basic convergence results are discussed in [7]. Although ADMM tends to converge quite slowly, for most engineering applications the method converges to acceptable results in a few tens of iterations. In the context of numerical weather prediction, ADMM converges fastest with the Huber half-quadratic formulation. A detailed convergence analysis for residuals, objective, and dual variable convergence can be found in [7].

6. Numerical experiments. We compare the performance of variational and ensemble data assimilation techniques using L_1, L_2, and Huber norms using the Lorenz-96 model [30] and a shallow water model in spherical coordinates (∼8,000 variables) [42]. The error statistic used to compare analyses against the reference solution is the root mean squared error (RMSE) metric:

$$\mathbf{RMSE}_i = \mathrm{N}_{\mathrm{var}}^{-1/2}\,\left\|\mathbf{x}_i - \mathbf{x}_i^{\mathrm{true}}\right\|_2, \tag{6.1}$$

where $\mathbf{x}^{\mathrm{true}}$ is the reference state of the system and $\mathrm{N_{var}}$ is the size of state space. The RMSE is calculated for each time within the assimilation window. We describe next the two testing models.

6.1. Lorenz-96 model. The Lorenz-96 model [30] is given by the equations

$$\frac{\mathrm{d}\mathbf{x}_k}{\mathrm{d}t} = \mathbf{x}_{\mathrm{k}-1}\left(\mathbf{x}_{\mathrm{k}+1} - \mathbf{x}_{\mathrm{k}-2}\right) - \mathbf{x}_k + F\,, \quad k = 1, \ldots, 40, \tag{6.2}$$

with periodic boundary conditions and the forcing term $F = 8$ [30]. A vector of equidistant components ranging from -2 to 2 is integrated for one time unit, and the result is taken as the reference initial state for the experiments. The background errors at initial time are Gaussian, with a diagonal background error covariance matrix, and each component's standard deviation is equal to 8% of the time-averaged magnitude of the reference solution. The same covariance matrix is used for all times in the 3D-Var calculations. Synthetic observations are generated by perturbing the reference trajectory with normal noise with mean zero, diagonal observation error covariance, and standard deviation of each component equal to 5% of the time-averaged magnitude of the reference solution.

6.2. Shallow water model on the sphere. The shallow water equations have been used extensively as a simple model of the atmosphere since they contain the essential wave propagation mechanisms found in general circulation models [42]. The shallow water equations in spherical coordinates are

$$\frac{\partial u}{\partial t} + \frac{1}{a\cos\theta}\left(u\frac{\partial u}{\partial \boldsymbol{\phi}} + v\cos\theta\frac{\partial u}{\partial \theta}\right) - \left(f + \frac{u\tan\theta}{a}\right)v + \frac{g}{a\cos\theta}\frac{\partial h}{\partial \boldsymbol{\phi}} = 0, \tag{6.3a}$$

$$\frac{\partial v}{\partial t} + \frac{1}{a\cos\theta}\left(u\frac{\partial v}{\partial \boldsymbol{\phi}} + v\cos\theta\frac{\partial v}{\partial \theta}\right) + \left(f + \frac{u\tan\theta}{a}\right)u + \frac{g}{a}\frac{\partial h}{\partial \theta} = 0, \tag{6.3b}$$

$$\frac{\partial h}{\partial t} + \frac{1}{a\cos\theta}\left(\frac{\partial\,(hu)}{\partial \boldsymbol{\phi}} + \frac{\partial (hv\cos\theta)}{\partial \theta}\right) = 0. \tag{6.3c}$$

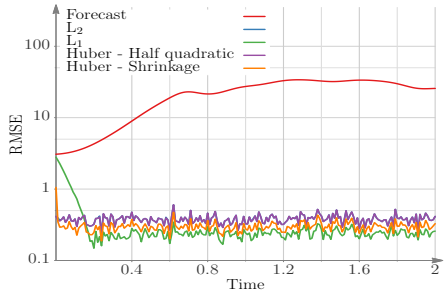
Here f is the Coriolis parameter given by $f = 2\Omega\sin\theta$, where Ω is the angular speed of the rotation of the Earth, h is the height of the homogeneous atmosphere, u and v are the zonal and meridional wind components, respectively, θ and $\boldsymbol{\phi}$ are the latitudinal and longitudinal directions, respectively, a is the radius of the earth, and g is the gravitational constant. The space discretization is performed using the unstaggered Turkel–Zwas scheme [31, 32]. The discretized spherical grid has 36 nodes along the longitude and 72 nodes along the latitude. The semi-discretization in space leads to a discrete model of the form (2.1). In (2.1) the zonal wind, meridional wind, and the height variables are combined into the vector $\mathbf{x} \in \mathbb{R}^n$ with $n = 3 \times \mathrm{nlat} \times \mathrm{nlon}$. We perform the time integration using an adaptive time-stepping algorithm. For a tolerance of 10^{-8} the average time-step size of the time integrator is 180 seconds. A reference initial condition is used to generate a reference trajectory.

Synthetic observation errors at various times t_i are normally distributed with mean zero and a diagonal observation error covariance matrix with entries equal to $(\mathbf{R}_i)_{k,k} = 1\,\mathrm{m^2 s^{-2}}$ for u and v components, and $(\mathbf{R}_i)_{k,k} = 10^6\,\mathrm{m}^2$ for the h components. These values correspond (approximately) to a standard deviation of 5% of the time-averaged values for u and v components, and 2% of the time-averaged values for the h component. A flow dependent background error covariance matrix is constructed as described in [2, 3]. The standard deviation of the background errors for the height component is 2% of the average magnitude of the height component in the reference initial condition. The standard deviation of the background errors for the wind components is 15% of the average magnitude of the wind component in the reference initial condition.

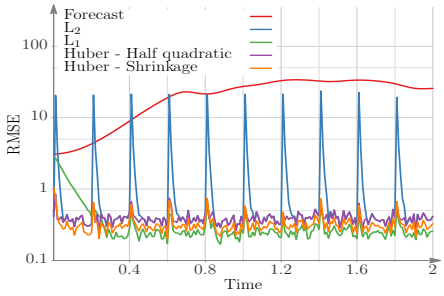
6.3. 3D-Var experiments with the Lorenz-96 model. The assimilation window is 2 units long. All components of the state space are observed. Experiments are performed for two observation frequencies, namely, one observation for every 0.1 units and 0.01 units, respectively. The tests are carried out with both good and erroneous data. The good data contain no outliers. The erroneous data contain one outlier (measurement of one state) which lies $\sim$ 100 standard deviations away from the mean. This outlier occurs every 0.2 time units.

The optimization proceeds in stages of inner and outer iterations. The inner iterations aim to find an iterate which minimizes the cost function for a particular value of the penalty parameter (μ). The outer iterations involve increasing the penalty parameter and updating the Lagrange multipliers. For experimental uniformity, we use 15 outer iterations with 3D-Var, 4D-Var, and local ensemble transform Kalman Filter (LETKF) for all formulations: L_1, Huber with ADMM, and Huber with half-quadratic solutions. Numerical results are presented in Figure 1.

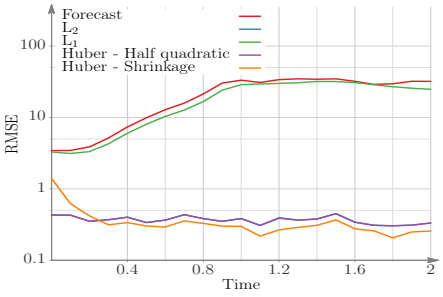
Figure 1(a) shows that all formulations perform very well at high observation frequencies (one observation every 0.01 units) and using good data. At a low observation frequency with good data the L_1 formulation does not perform as well as the two Huber formulations, as seen in Figure 1(c). Since the data are good, the Huber formulations perform nearly as well as the L_2 formulation. For example in Figure 1(c), the half-quadratic Huber formulation curve overlaps with the L_2 formulation curve.



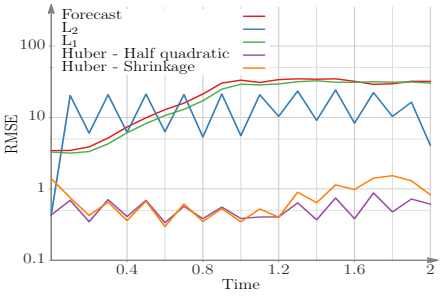
(a) Observations with only small random errors. The frequency of observations is 0.01 units. The Huber norm uses $\tau = 1$.



(b) Observations with outliers. The frequency of observations is 0.01 units. The Huber norm uses $\tau = 1$.

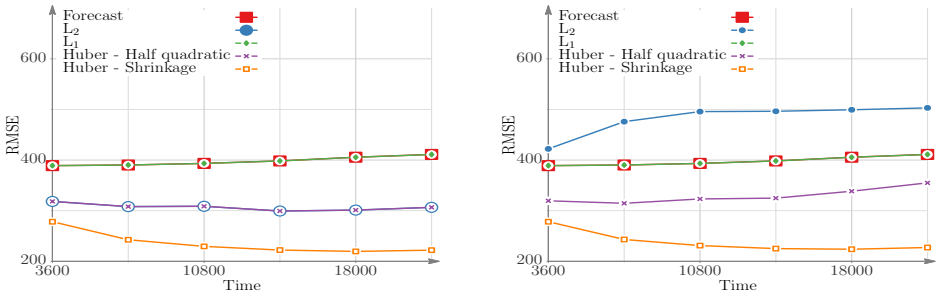


(c) Observations with only small random errors. The frequency of observations is 0.1 units. The Huber norm uses $\tau = 1$. The Huber - half quadratic curve overlaps with the L_2 curve.



(d) Observations with outliers. The frequency of observations is 0.1 units. The Huber norm uses $\tau = 1$.

FIG. 1. *3D-Var results for the Lorenz-96 model* (6.2). *Erroneous observations occur every* 0.2 *time units.*



(a) Observations with small random errors. L_1 curve overlaps with the forecast, hence to differentiate, we use bigger boxes for forecast. L_2 overlaps with Huber half-quadratic formulation, hence we use larger circles for L_2 for the sake of clarity.

(b) Observations with outliers. L_1 curve overlaps with the forecast curve, hence, for clarity, we use bigger boxes to represent the forecast curve.

FIG. 2. *3D-Var results for the shallow water model* (6.3). *The Huber threshold is* $\tau = 2$.

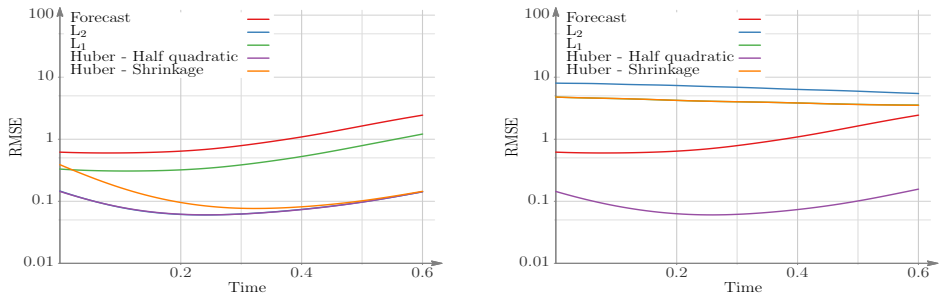
For frequent (good) observations, both L_1 and L_2 formulations perform well, even when outliers are present in the data, as seen in Figure 1(b). However, at a low frequency of measurements and when outliers are present the L_2 formulation fails to produce good results, while the Huber formulations demonstrate their robustness, as seen in Figure 1(d).

6.4. 3D-Var experiments with the shallow water on the sphere model. Experiments with the shallow water model are performed for an assimilation window of six hours, with all variables being observed hourly. The good data contain random noise but no outliers. The erroneous data contain one outlier, namely, the height component at latitude 32° south, longitude 120° east and it has a value which is ~ 50 standard deviations away from the mean. This outlier occurs at all observation times and, hence, simulates one faulty sensor periodically providing incorrect observation values.

Figures 2(a) and 2(b) present the results of assimilating good and erroneous observations, respectively. RMSEs are shown for trajectories corresponding to forecast, L_1, L_2, and two implementations of the Huber norm. The Huber norm implementations perform as well as the L_2 norm analysis when only good observations are used. When the observations contain outliers, the L_2 norm analysis is inaccurate. However, the Huber norm analysis is robust and remains unaffected by the quality of observations. The L_1 norm analysis also remains unaffected by the quality of observations, however, its errors are similar to those of the forecast; the L_1 formulation does not effectively use the information from the good observations. Consequently, the Huber norm offers the best 3D-Var formulation, leading to analyses that are both accurate and robust.

6.5. 4D-Var experiments with the Lorenz-96 model. 4D-Var experiments with the Lorenz model are performed for one assimilation window of 0.6 units. All components are observed every 0.1 units. The good data contain random noise but no outliers. The erroneous data contain one outlier at all observation times and its value is ~ 100 standard deviations away from the mean.

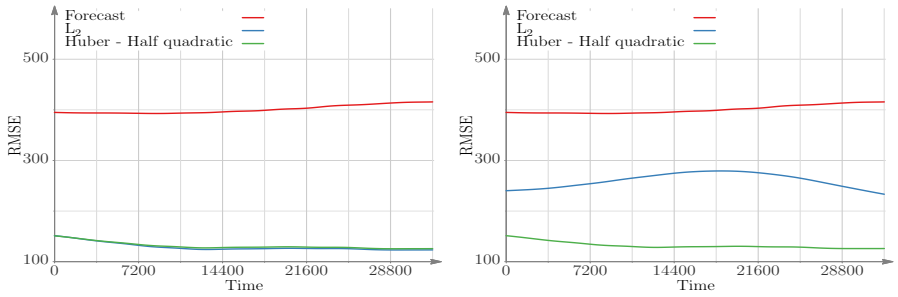
Figure 3(a) presents the 4D-Var results with good data. The performance of the Huber 4D-Var with half-quadratic formulation matches the performance of the L_2



(a) Observations with small random errors. Huber half-quadratic curve overlaps with the L_2 curve.

(b) Observations with outliers. L_1 curve overlaps with Huber shrinkage curve.

FIG. 3. *4D-Var results for the Lorenz-96 model* (6.2). *The Huber threshold is* $\tau = 2$.



(a) Observations with small random errors.

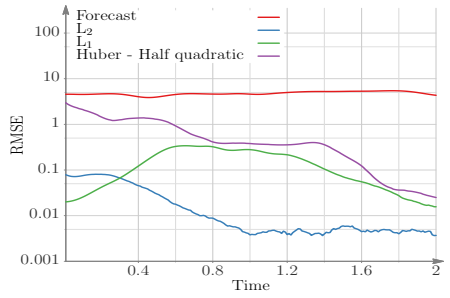
(b) Observations with outliers.

FIG. 4. *4D-Var results for the shallow water model* (6.3). *The Huber threshold is* $\tau = 2$.

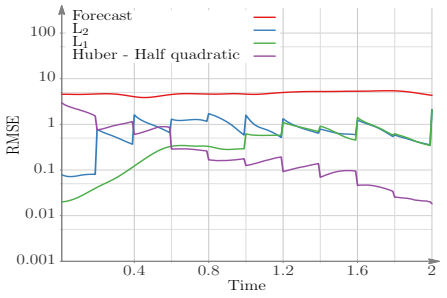
formulation. The analyses provided by L_1 and Huber formulations with shrinkage are less accurate. Figure 3(b) presents the results of assimilating data with outliers. The 4D-Var Huber analysis using the half-quadratic formulation remains unaffected by the data errors. All other formulations (L_1, L_2, and Huber with shrinkage) are negatively impacted by the presence of outliers.

6.6. 4D-Var experiments with the shallow water on the sphere model. We consider one nine hours long assimilation window, with all components of the system observed hourly. The good data contain only small random errors. The bad data contain one large outlier with an error of $\sim$50 standard deviations away from the mean. This simulates a faulty sensor that provides erroneous data at all observation times.

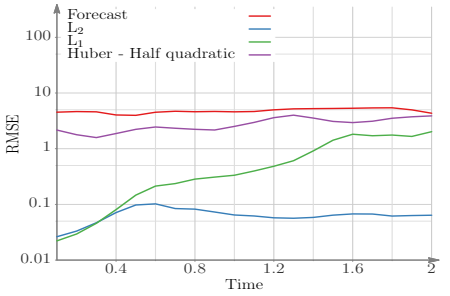
The L_1 and Huber norm with shrinkage formulations of 4D-Var did not perform well for the Lorenz model, and we do not test them on the shallow water model. Figures 4(a) and 4(b) show the 4D-Var results for the shallow water model using good and faulty observations, respectively. We compare the performance of Huber with half-quadratic formulation with the L_2 norm formulation. When the quality of the observations is good the Huber norm and the L_2 norm formulations perform equally well. When the observations contain outliers the L_2 formulation gives inaccurate analyses, while the Huber norm formulation is robust and continues to provide good results.



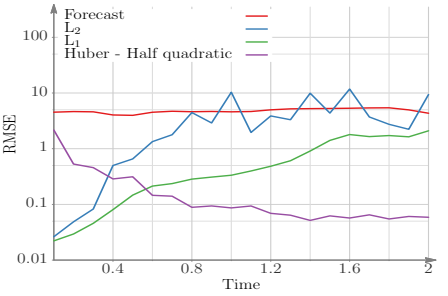
(a) Observations with small random errors. The frequency of observations is 0.01 time units. The Huber threshold is $\tau = 1$.



(b) Observations with outliers. The frequency of observations is 0.01 time units. The Huber threshold is $\tau = 1$.



(c) Observations with small random errors. The frequency of observations is 0.1 time units. The Huber threshold is $\tau = 1$.



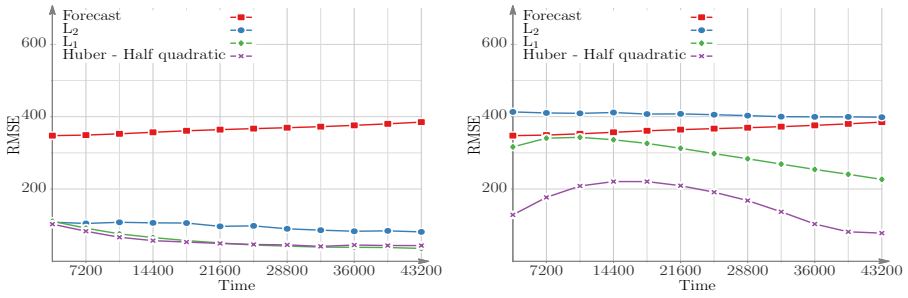
(d) Observations with outliers. The frequency of observations is 0.1 time units. The Huber threshold is $\tau = 1$.

FIG. 5. *LETKF results for the Lorenz-96 model* (6.2). *Erroneous observations occur every* 0.2 *time units.*

6.7. EnSRF experiments with the Lorenz-96 model. In this section we discuss the results of the LETKF implementations based on the L_1, L_2, and Huber norms. We note that, due to localization, the data outliers in LETFK will impact the states in only small regions of the domain. The number of ensemble members for the experiments is 20.

Assimilation experiments are carried out with different frequencies for outliers in the observations. Figures 5(a) and 5(b) show results with observations taken every 0.01 time units; for the bad data outliers are present every 0.2 time units. The L_2 formulation converges the fastest when only good data are used, as seen in Figure 5(a). The L_1 and Huber formulations converge more slowly, and give similar results at the end of the assimilation window. When bad data are used LETKF-Huber outperforms the other formulations, as illustrated in Figure 5(b). The L_2 LETKF formulation still gives improvements over the forecast, since the frequency of outliers is small (every 20 observation times), and injection of good observations at other times avoids filter divergence.

Figure 5(c) shows results with observations taken every 0.1 time units; for the bad data outliers are present every 0.2 time units. The relatively frequent outliers (every other observation) affect the analysis obtained with the L_2 EnKF formulation. In this case the L_1 formulation performs better that the L_2 one. Huber LETKF gives the best analyses overall. When $\tau = 1$, the method can treat good observations as outliers, and therefore the contributions of some good observations to the analyses



(a) Observations with small random errors. (b) Observations with outliers.

FIG. 6. *LETKF results for the shallow water model* (6.3). *The Huber threshold is* $\tau = 1$.

can be attenuated; this can be seen in Figure 5(c), where the Huber formulation is less accurate than the L_2 formulation for good data. When τ is increased, for instance, $\tau = 3$, the trajectories of the Huber LETKF for good and bad observations are almost identical, as can be seen in Figure 5(a).

6.8. EnSRF experiments with the shallow water on the sphere model. The shallow water model (6.3) experiments use ensembles with 30 members. The initial ensemble perturbation is normal, with a standard deviation of 5% of the background state for each component. Figure 6 shows the analysis errors for the L_1, L_2, and Huber LETKF formulations. The Huber half-quadratic LETKF provides analyses that are as accurate as the L_2 formulation analyses when only good data are used. When outliers are present, however, the L_2 results are very inaccurate. The Huber formulation is almost completely unaffected by the data outliers, and provides equally better analyses with and without data outliers. The L_1 formulation results are also unaffected by outliers, but their overall accuracy is quite low.

7. Conclusions. This paper develops a systematic framework for performing robust 3D-Var, 4D-Var, and ensemble-based filtering data assimilation. The traditional algorithms are formulated as optimization problems where the cost functions are L_2 norms of background and observation residuals; the L_2 norm choice corresponds to Gaussianity assumptions for background and observation errors, respectively. The L_2 norm formulation has to make large corrections to the optimal solution in order to accommodate outliers in the data. Consequently, a few observations containing large errors can considerably deteriorate the overall quality of the analysis. In order to avoid this effect traditional data quality control rejects certain observations, which may result in loss of useful information. A more recent approach performs a one-time, offline adjustment of the data weights based on the estimated data quality.

The robust data assimilation framework described herein reformulates the underlying optimization problems using L_1 and Huber norms. The resulting robust data assimilation algorithms do not reject any observations. Moreover, the relative importance weights of the data are not decided offline, but rather are adjusted iteratively for each observation based on their deviation from the mean forecast. Numerical results show that the L_1 norm formulation is very robust, being the least affected by the presence of outliers. However, the resulting analyses are inaccurate: the L_1 norm rejects not only outliers, but useful information as well. The Huber norm formulation is able to fully use the information from "good data" while remaining robust (rejecting the influence of outliers). We consider two solution methods for Huber

norm optimization. The shrinkage operator solution displays slow convergence and can become impractical in real-life problems. The Huber norm solution using the half-quadratic formulation seems to be the most suitable approach for large scale data assimilation applications. It is only slightly more expensive than the traditional L_2 4D-Var approach, and yields good results both in the absence and in the presence of data outliers. Future work will apply robust data assimilation algorithms using a Huber norm formulation with a half-quadratic solution to real problems using the weather research and Forecasting model [46].

REFERENCES

[1] E. Anderson and H. Jarvinen, *Variational quality control*, Q. J. Roy. Meteorol. Soc., 125 (1999).

[2] A. Attia, V. Rao, and A. Sandu, *A hybrid Monte-Carlo sampling smoother for four dimensional data assimilation*, Internat. J. Numer. Methods Fluids, 83 (2014), pp. 90–112.

[3] A. Attia, V. Rao, and A. Sandu, *A sampling approach for four dimensional data assimilation*, in Proceedings of the Dynamic Data Driven Environmental System Science Conference, Lecture Notes in Comput. Sci. 8964, Springer, Cham, 2014, pp. 215–226.

[4] A. Aved, F. Darema, and E. Blasch, *Dynamic Data Driven Application Systems*, www.1dddas.org, 2014.

[5] J. Bect, L. Blanc-Féraud, G. Aubert, and A. Chambolle, *A L_1-unified variational framework for image restoration*, in European Conference on Computer Vision, Lecture Notes in Comput. Sci. 3024, Springer, Berlin, 2004, pp. 1–13.

[6] J. Bioucas-Dias and M. Figueiredo, *A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration*, IEEE Trans. Image Process., 16 (2007), pp. 2992–3004.

[7] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn., 3 (2011), pp. 1–122.

[8] D. G. Cacuci, M. Ionescu-Bujor, and I. M. Navon, *Sensitivity and Uncertainty Analysis, Vol.* II: *Applications to Large-Scale Systems*, CRC Press, Boca Raton, FL, 2005.

[9] A. Cioaca, M. Alexe, and A. Sandu, *Second-order adjoints for solving PDE-constrained optimization problems*, Optim. Methods Softw., 27 (2012), pp. 625–653.

[10] A. Cioaca and A. Sandu, *An optimization framework to improve 4D-Var data assimilation system performance*, J. Comput. Phys., 275 (2014), pp. 377–389.

[11] P. L. Combettes and V. R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. & Simul., 4 (2005), pp. 1168–1200.

[12] R. Daley, *Atmospheric Data Analysis*, Vol. 2, Cambridge University Press, Cambridge, 1993.

[13] A. Ebtehaj, M. Zupanski, G. Lerman, and E. Foufoula-Georgiou, *Variational data assimilation via sparse regularization*, Tellus A, 66 (2014), 21789.

[14] M. Elad, *Why simple shrinkage is still relevant for redundant representations?*, IEEE Trans. Inform. Theory, 52 (2006), pp. 5559–5569.

[15] M. Elad, B. Matalon, J. Shtok, and M. Zibulevsky, *A wide-angle view at iterated shrinkage algorithms*, in Wavelets XII, SPIE Conf. 6701, 2007, SPIE, Bellingham, WA, 670102.

[16] M. Elad, B. Matalon, and M. Zibulevsky, *Coordinate and subspace optimization methods for linear least squares with non-quadratic regularization*, Appl. Comput. Harmon. Anal., 23 (2007), pp. 346–367.

[17] M. Figueiredo and R. Nowak, *An EM algorithm for wavelet-based image restoration*, IEEE Trans. Image Process., 12 (2003), pp. 906–916.

[18] M. Freitag, N. Nichols, and C. Budd, *L_1-regularisation for ill-posed problems in variational data assimilation*, PAMM, 10 (2010), pp. 665–668, https://doi.org/10.1002/pamm.201010324.

[19] M. Freitag, N. Nichols, and C. Budd, *Resolution of sharp fronts in the presence of model error in variational data assimilation*, Q. J. Roy. Meteorol. Soc., 139 (2013), pp. 742–757.

[20] A. Guitton and W. Symes, *Robust inversion of seismic data using the Huber norm*, Geophysics, 68 (2003), pp. 1310–1319.

[21] E. Hale, W. Yin, and Y. Zhang, *A Fixed-Point Continuation Method for L_1-Regularized Minimization with Applications to Compressed Sensing*, Technical report, CAAM TRO7-07, Rice University, Houston, Tx, (2007).

[22] A. Hollingsworth, D. Shaw, P. Lönnberg, L. Illari, K. Arpe, and A. Simmons, *Monitoring of observation and analysis quality by a data assimilation system*, Monthly Weather Rev., 114 (1986), pp. 861–879.

[23] P. Huber, *Robust statistics: A review*, Ann. Math. Statis., 43 (1972), pp.1041–1067.
[24] B. Hunt, E. Kostelich, and I. Szunyogh, *Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter*, Phys. D, 230 (2007), pp. 112–126.
[25] N. Ingleby and A. Lortenc, *Bayesian quality control using multivariate normal distributions*, Q. J. Roy. Meteorol. Soc., 119 (1993), pp. 1195–1225.
[26] E. Kalnay, *Atmospheric Modeling, Data Assimilation, and Predictability*, Cambridge University Press, Cambridge, 2003.
[27] A. Lorenc, *Analysis methods for numerical weather prediction*, Q. J. Roy. Meteorol. Soc., 112 (1986), pp. 1177–1194.
[28] A. Lorenc, *A Bayesian Approach to Observation Quality Control in Variational and Statistical Assimilation*, Technical report, Defense Technical Information Center, Fort Belvoir, VA, 1993.
[29] A. Lorenc and O. Hammon, *Objective quality control of observations using Bayesian methods. Theory, and a practical implementation*, Q. J. Roy. Meteorol. Soc., 114 (1988), pp. 515–543.
[30] E. N. Lorenz, *Predictabilty: A problem partly solved*, in Proceedings of Seminar on Predictability, ECMWF, Rending, England, 1996, pp. 40–58.
[31] I. M. Navon and R. De Villiers, *The application of the Turkel-Zwas explicit large time-step scheme to a hemispheric barotropic model with constraint restoration*, Monthly Weather Rev., 115 (1987), pp. 1036–1052.
[32] I. M. Navon and J. Yu, *Exshall: A Turkel-Zwas explicit large time-step FORTRAN program for solving the shallow-water equations in spherical coordinates*, Comput. Geosci., 17 (1991), pp. 1311–1343.
[33] M. Nikolova and M. K. Ng, *Analysis of half-quadratic minimization methods for signal and image recovery*, SIAM J. Sci. Comput., 27 (2005), pp. 937–966.
[34] R. Nowak and M. Figueiredo, *Fast wavelet-based image deconvolution using the EM algorithm*, in Proceedings of the 2001 Conference Record of the Thirty-Fifth Asilomar Conference on Signals, Systems and Computers, Vol. 1, IEEE, Piscateway, NJ, 2001, pp. 371–375.
[35] V. Rao and A. Sandu, *A posteriori error estimates for DDDAS inference problems*, Procedia Comput. Sci., 29 (2014), pp. 1256–1265.
[36] V. Rao and A. Sandu, *A posteriori error estimates for the solution of variational inverse problems*, SIAM/ASA Journal on Uncertainty Quantification, 3 (2015), pp. 737–761.
[37] S. Roh, M. Genton, M. Jun, I. Szunyogh, and I. Hoteit, *Observation quality control with a robust ensemble Kalman filter*, Monthly Weather Rev., 141 (2013), pp. 4414–4428.
[38] E. D. N. Ruiz and A. Sandu, *A derivative-free trust region framework for variational data assimilation*, J. Comput. Appl. Math., 293 (2016), pp. 164–179.
[39] A. Sandu and T. Chai, *Chemical data assimilation—An overview*, Atmosphere, 2 (2011), pp. 426–463.
[40] A. Sandu, D. N. Daescu, G. R. Carmichael, and T. Chai, *Adjoint sensitivity analysis of regional air quality models*, J. Comput. Phys., 204 (2005), pp. 222–252.
[41] A. Sandu and L. Zhang, *Discrete second order adjoints in atmospheric chemical transport modeling*, J. Comput. Phys., 227 (2008), pp. 5949–5983, https://doi.org/10.1016/j.jcp.2008.02.011.
[42] A. St-Cyr, C. Jablonowski, J. Dennis, H. Tufo, and S. Thomas, *A comparison of two shallow water models with nonconforming adaptive grids*, Monthly Weather Rev., 136 (2008), pp. 1898–1922.
[43] C. Tavolato and L. Isaksen, *Huber Norm Quality Control in the IFS*, ECMWF Newsletter No. 122 (Winter 2009/10), pp.27–31.
[44] C. Tavolato and L. Isaksen, *On the use of a Huber norm for observation quality control in the ECMWF 4D-Var*, Q. J. Roy. Meteorol. Soc., 141 (2015), pp. 1514–1527.
[45] J. Tropp and A. Gilbert, *Signal recovery from random measurements via orthogonal matching pursuit*, IEEE Trans. Inform. Theory, 53 (2007), pp. 4655–4666.
[46] WRF, *The Weather Research & Forecasting Model*, http://wrf-model.org.
[47] S. Wright, R. Nowak, and M. Figueiredo, *Sparse reconstruction by separable approximation*, IEEE Trans. Signal Process., 57 (2009), pp. 2479–2493.