On variational downscaling, fusion, and assimilation of hydrometeorological states: A unified framework via regularization

A. M. Ebtehaj[1,2] and E. Foufoula-Georgiou[1]

[1] Improved estimation of hydrometeorological states from down-sampled observations and background model forecasts in a noisy environment has been a subject of growing research in the past decades. Here we introduce a unified variational framework that ties together the problems of downscaling, data fusion, and data assimilation as ill-posed inverse problems. This framework seeks solutions beyond the classic least squares estimation paradigms by imposing a proper regularization, expressed as a constraint consistent with the degree of smoothness and/or probabilistic structure of the underlying state. We review relevant smoothing norm regularization methods in derivative space and extend classic formulations of the aforementioned problems with particular emphasis on land surface hydrometeorological applications. Our results demonstrate that proper regularization of downscaling, data fusion, and data assimilation problems can lead to more accurate and stable recovery of the underlying non-Gaussian state of interest with improved performance in capturing isolated and jump singularities. In particular, we show that the Huber regularization in the derivative space offers advantages, compared to the classic solution and the Tikhonov regularization, for spatial downscaling and fusion of non-Gaussian multisensor precipitation data. Furthermore, we explore the use of Huber regularization in a variational data assimilation experiment while the initial state of interest exhibits jump discontinuities and non-Gaussian probabilistic structure. To this end, we focus on the heat equation motivated by its fundamental application in the study of land surface heat and mass fluxes.

Citation: Ebtehaj, A. M., and E. Foufoula-Georgiou (2013), On variational downscaling, fusion, and assimilation of hydrometeorological states: A unified framework via regularization, *Water Resour. Res.*, *49*, 5944–5963, doi:10.1002/wrcr.20424.

1. Introduction

[2] In parallel to the growing technologies for earth remote sensing, we have witnessed an increasing interest to improve the accuracy of observations and integrate them with predictive models for enhancing our environmental forecast skills. Remote sensing observations are typically noisy and coarse-scale representations of a true state variable of interest, lacking sufficient details for fine-scale environmental modeling. In addition, environmental predictions are not perfect as models often suffer either from inadequate characterization of the underlying physics or inaccurate initialization. Given these limitations, several classes of estimation problems present themselves as continuous challenges for the atmospheric, hydrologic, and oceanic science communities. These include (1) downscaling (DS), which refers to the class of problems for enhancing the resolution of a measured or modeled state of interest by producing a fine-scale representation of that state with reduced uncertainty; (2) data fusion (DF), to produce an improved estimate from a suite of noisy observations at different scales; and (3) data assimilation (DA), which deals with estimating initial conditions in a predictive model consistent with the available observations and the underlying model dynamics. In this paper, we revisit the problems of downscaling, data fusion, and data assimilation focusing on a common thread between them as variational ill-posed inverse problems. Proper regularization and solution methods are proposed to efficiently handle large-scale data sets while preserving key statistical and geometrical properties of the underlying field of interest, namely, non-Gaussian and structured variability in real or transformed domains. Here, we only examine a few hydrometeorological inverse problems with particular emphasis on land-surface applications.

[3] In land-surface hydrologic studies, DS of precipitation and soil moisture observations has received considerable attention, using a relatively wide range of methodologies. DS methods in hydrometeorology and climate studies generally fall into three main categories, namely, dynamic downscaling, statistical downscaling, and variational downscaling. Dynamic downscaling often uses a regional physically based model to reproduce fine-scale details of the state of interest consistent with the large-scale

[1]Department of Civil Engineering, University of Minnesota, Minneapolis, Minnesota, USA.
[2]School of Mathematics, University of Minnesota, Minneapolis, Minnesota, USA.

Corresponding author: E. Foufoula-Georgiou, Department of Civil Engineering, Saint Anthony Falls Laboratory, University of Minnesota, Minneapolis, MN 55414, USA. (efi@umn.edu)

observations or outputs of a global circulation model [e.g., *Reichle et al.*, 2001a; *Castro et al.*, 2005; *Zupanski et al.*, 2010]. Statistical downscaling methods encompass a large group of methods that typically use empirical multiscale statistical relationships, parameterized by observations or other environmental predictors, to reproduce realizations of fine-scale fields. Precipitation and soil moisture statistical downscaling has been mainly approached via spectral and (multi)fractal interpolation methods, capitalizing on the presence of a power law spectrum and a statistical self-similarity/self-affinity in precipitation and soil moisture fields [*Lovejoy and Mandelbrot*, 1985; *Lovejoy and Schertzer*, 1990; *Gupta and Waymire*, 1993; *Kumar and Foufoula-Georgiou*, 1993; *Perica and Foufoula-Georgiou*, 1996; *Veneziano et al.*, 1996; *Wilby et al.*, 1998a, 1998b; *Deidda*, 2000; *Kim and Barros*, 2002; *Rebora et al.*, 2005; *Badas et al.*, 2006; *Merlin et al.*, 2006; among others]. In variational approaches, a direct cost function is defined whose optimal point is the desired fine-scale field which can be obtained via using an optimization method. Recently along this direction, *Ebtehaj et al.* [2012] cast the rainfall DS problem as an inverse problem using sparse regularization to address the intrinsic rainfall singularities and non-Gaussian statistics. This variational approach belongs to the class of methodologies presented and extended in this paper.

[4] The DF problem has also been a subject of continuous interest in the precipitation science community mainly due to the availability of rainfall measurements from multiple spaceborne (e.g., TRMM and GOES satellites) and ground-based sensors (e.g., the NEXRAD network and rain gauges). The accuracy and space-time coverage of remotely sensed rainfall are typically conjugate variables. In other words, more accurate observations are often available with lower space-time coverage and vice versa. For instance, low-orbit microwave sensors provide more accurate observations but with less space-time coverage compared to the high-orbit geo-stationary infrared (GOES-IR) sensors. Moreover, there are often multiple instruments on a single satellite (e.g., precipitation radar and microwave imager on TRMM), each of which measures rainfall with different footprints and resolutions. A wide range of methodologies, including weighted averaging, regression, filtering, and neural networks, has been applied to combine microwave and Geo-IR rainfall signals [e.g., *Adler et al.*, 2003; *Huffman et al.*, 1995; *Sorooshian et al.*, 2000; *Huffman et al.*, 2001; *Hong et al.*, 2004; *Huffman et al.*, 2007]. Furthermore, a few studies have addressed methodologies to optimally combine the products of the TRMM precipitation radar (PR) with the TRMM microwave imager (TMI) using Bayesian inversion and weighted least squares (WLS) approaches [e.g., *Masunaga and Kummerow*, 2005; *Kummerow et al.*, 2010]. From another direction, Gaussian filtering methods on Markovian tree-like structures, the so-called scale recursive estimation (SRE), have been proposed to merge spaceborne and ground-based rainfall observations at multiple scales [e.g., *Gorenburg et al.*, 2001; *Tustison et al.*, 2003; *Bocchiola*, 2007; *Van de Vyver and Roulin*, 2009; *Wang et al.*, 2011], see also *Kumar* [1999] for soil moisture applications. Recently, using the Gaussian-scale mixture probability model and an adaptive filtering approach, *Ebtehaj and Foufoula-*

Georgiou [2011a] proposed a fusion methodology in the wavelet domain to merge TRMM-PR and ground-based NEXRAD measurements, aiming to preserve the non-Gaussian structure and local extremes of precipitation fields.

[5] Data assimilation has played an important role in improving the skill of environmental forecasts and has become by now a necessary step in operational predictive models [see *Daley*, 1993]. Data assimilation amounts to integrating the underlying knowledge from the *observations* into the first guess or the *background* state, typically provided by a physical model from the previous forecast step. The goal is then to obtain an improved estimate of the current state of the system with reduced uncertainty, the so-called *analysis*. The analysis is then used to forecast the state at the next time step and so on (see *Daley* [1993] and *Kalnay* [2003] for a comprehensive review). One of the most common approaches to the data assimilation problem relies on variational techniques [e.g., *Sasaki*, 1958; *Lorenc*, 1986; *Talagrand and Courtier*, 1987; *Courtier and Talagrand*, 1990; *Parrish and Derber*, 1992; *Zupanski*, 1993; *Courtier et al.*, 1994; *Reichle et al.*, 2001b; *Margulis and Entekhabi*, 2003; among many others]. In these methods, one explicitly defines a cost function, typically quadratic, whose unique minimizer is the analysis state. On the other hand, very recently, *Freitag et al.* [2012] proposed a regularized variational data assimilation scheme to improve assimilation results in advection-dominated flow in the presence of sharp weather fronts.

[6] The common thread in the DS, DF, and DA problems is that, in all of them, we seek an improved estimate of the true state given a suite of noisy and down-sampled observations and/or uncertain model-predicted states. Specifically, let us suppose that the unknown *true* state in continuous space is denoted by $x(t)$ and its indirect observation (or model output), by $y(r)$. Let us also assume that $x(t)$ and $y(r)$ are related via a linear integral equation, called the Fredholm integral equation of the first kind, as follows:

$$\int_0^1 \mathcal{H}(r,t)x(t)dt = y(r), 0 \leq r \leq 1, \tag{1}$$

where $\mathcal{H}(r,t)$ is the known kernel relating $x(t)$ and $y(r)$. Recovery of $x(t)$ knowing $y(r)$ and $\mathcal{H}(r,t)$ is a classic linear inverse problem. Clearly, the deconvolution problem is a very special case with the kernel of the form $\mathcal{H}(r-t)$, which in its discrete form plays a central role in this paper. Linear inverse problems are by nature ill-posed, in the sense that they do not satisfy at least one of the following three conditions: (1) existence, (2) uniqueness, and (3) stability of the solution. For instance, when due to the kernel architecture, the dimension of the observation is smaller than that of the true signal, infinite choices of $x(t)$ may lead to the same $y(r)$ and there is no unique solution for the problem. For the case when $y(r)$ is noisy and has a larger dimension than the true state, the solution is typically very unstable because the high-frequency components in $y(r)$ are typically amplified and spoil the solution in the inversion process. A common approach to make an inverse problem well posed is via the so-called *regularization* methods [e.g., *Hansen*, 2010]. The goal of regularization is to properly constrain the inverse problem aiming to obtain a unique and sufficiently stable solution. The choice of regularization typically depends on the continuity and degree of

smoothness of the state variable of interest, often called the *regularity* condition. For instance, some state variables or environmental fluxes are very regular with high degree of smoothness and differentiability (e.g., pressure), while others might be more irregular and suffer from frequent and different sorts of discontinuities (e.g., rainfall). In fact, it can be shown that the proper choices of regularization not only yield unique and stable solutions but also reinforce the underlying regularity of the true state in the solution. It is important to note that different regularity conditions are theoretically consistent with different statistical signatures in the true state, a fact that may guide proper design of the regularization, as explored in this study.

[7] The central goal of this paper is to propose a unified framework for the class of DS, DF, and DA problems by recasting them as discrete linear inverse problems using a relevant regularization in the derivative space, aiming to solve them more accurately compared to the classic weighted least squares (WLS) formulations. From a statistical standpoint, the main motivation is to explicitly incorporate non-Gaussianity of the underlying state in the derivative domain as a prior knowledge to obtain an improved estimate of jump and isolated extreme variabilities in the time-space structure of the hydrometeorological state of interest. Note that the proposed framework relies on the seminal works by, for example, *Tibshirani* [1996], *Chen et al.* [2001], *Candes and Tao* [2006], and recent developments in mathematical formalisms of inverse problems [e.g., *Hansen*, 2010; *Elad*, 2010], which have received a great deal of attention in statistical regression and image processing, but are relatively new to the communities of hydrologic and atmospheric sciences. To the best of our knowledge, in these areas, the only studies that explore these methodologies are *Ebtehaj et al.* [2012] and *Freitag et al.* [2012] for rainfall downscaling and data assimilation of sharp fronts, respectively.

[8] The presented methodologies for the DS and DF problems are examined through downscaling and data fusion of remotely sensed rainfall observations, which have fundamental applications in flash flood predictions, especially in small watersheds [*Rebora et al.*, 2005; *Siccardi et al.*, 2005; *Rebora et al.*, 2006]. We show that the presented methodologies allow us to improve the quality of rainfall estimation and reduce estimation uncertainty by recovering the small-scale high-intensity rainfall extreme features, which have been lost in the low-resolution sampling of the sensor. For the DA family of problems, the promise of the presented framework is demonstrated via an elementary example using the heat equation, which plays a key role in the study of land surface heat and mass fluxes [e.g., *Peter-Lidard et al.*, 1997; *Liang et al.*, 1999]. The results demonstrate that the accuracy of the analysis and forecast cycles in a DA problem can be markedly improved, compared to the classic variational methods, especially when the initial state exhibits different forms of discontinuities.

[9] Section 2 provides conceptual insight into the discrete inverse problems. Section 3 describes the DS problem in detail, as a primitive building block for the other studied problems. Important classes of regularization methods are explained and their statistical interpretation is briefly discussed from the Bayesian point of view. Examples on rainfall downscaling are presented in this section by taking into account the specific regularity and statistical distribution of the rainfall fields in the derivative space. Section 4 is devoted to the regularized DF class of problems with examples and results on remotely sensed rainfall data. The regularized DA problem is discussed in section 5. Concluding remarks and future research perspectives are presented in section 6. The important duality between regularization and its statistical interpretation is further presented in Appendix A, while Appendix B is devoted to algorithmic details important for implementation of the proposed methodologies.

2. Discrete Inverse Problems: Conceptual Framework

[10] In this section, we briefly explain the conceptual key elements of discrete linear inverse estimation relevant to the problems at hand and leave further details for the next sections. Analogous to equation (1), linear discrete inverse problems typically amount to estimating the true high-resolution m-element state vector $\mathbf{x} \in \mathbb{R}^m$ from the following observation model:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v}, \qquad (2)$$

where $\mathbf{y} \in \mathbb{R}^n$ denotes the observations (e.g., output of a sensor), $\mathbf{H} \in \mathbb{R}^{n \times m}$ is an $n \times m$ observation operator which maps the state space onto the observation space, and $\mathbf{v} \sim \mathcal{N}(0, \mathbf{R})$ is the Gaussian error in $\mathbb{R}^n$. Note that the observation operator, which is a discrete representation of the kernel in equation (1), and the noise covariance are supposed to be known or properly calibrated. Depending on the relative dimension of $\mathbf{y}$ and $\mathbf{x}$, this linear system can be under determined ($m \gg n$) or overdetermined ($m \ll n$). In the under-determined case, there are infinite different $\mathbf{x}$'s that satisfy equation (2), while for the overdetermined case a unique solution may not exist. As is evident, the DS problem belongs to the class of under-determined systems because the sensor output is a coarse-scale and noisy representation of the true state. However, the class of DF and DA problems falls into the category of overdetermined systems, as the total size of the observations and background state exceeds the dimension of the true state.

[11] In each of the above cases, we may naturally try to obtain a solution with minimum error variance by solving a linear WLS problem. However, for the under-determined case the solution still does not exist, while for the overdetermined case it is commonly ill-conditioned and sensitive to the observation noise (see section 4). Therefore, the minimum variance WLS treatment cannot properly make the above inverse problems well posed. To obtain a unique and stable solution, the basic idea of regularization is to further constrain the solution. For instance, among many solutions that fit the observation model in equation (2), we can obtain the one with minimum energy, mean-squared curvature, or total variation. The choice of this constraint or regularization highly depends on a priori knowledge about the underlying regularity of $\mathbf{x}$. For sufficiently smooth $\mathbf{x}$, we naturally may promote a solution with minimum mean-squared curvature to impose the desired smoothness on the solution. However, if the state is nonsmooth and contains

frequent jumps and discontinuities, a solution with minimum total variation might be a better choice. In subsequent sections, we explain these concepts in more detail for the DS, DF, and DA problems with examples relevant to some land-surface hydrometeorological problems.

3. Regularized Downscaling

3.1. Problem Formulation

[12] To put the DS problem in a linear inverse estimation framework, we recognize that in the observation model of equation (2), the true high-resolution (HR) state $\mathbf{x} \in \mathbb{R}^m$ has a larger dimension than the low-resolution (LR) observation vector $\mathbf{y} \in \mathbb{R}^n$, that is, $m \gg n$. Throughout this work, a notation is adopted in which the vector $\mathbf{x} \in \mathbb{R}^m$ may also represent, for example, a 2-D field $X \in \mathbb{R}^{\sqrt{m} \times \sqrt{m}}$, which is vectorized in a fixed order (e.g., lexicographical).

[13] As explained in the previous section, the DS problem naturally amounts to obtaining the best WLS estimate $\hat{\mathbf{x}}$ of the HR or fine-scale true state as follows:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \frac{1}{2} ||\mathbf{y} - \mathbf{Hx}||_{\mathbf{R}^{-1}}^2 \right\}, \tag{3}$$

where $||\mathbf{x}||_{\mathbf{A}}^2 = \mathbf{x}^{\mathrm{T}} \mathbf{Ax}$ denotes the *quadratic norm*, while $\mathbf{A}$ is a positive definite matrix. Due to the ill-posed nature of the problem, this optimization does not have a unique solution, as setting the derivative of the cost function to zero, the Hessian $(\mathbf{H}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{H})$ is definitely singular. To narrow down all possible solutions to a stable and unique one, a common choice is to regularize the problem by constraining the squared Euclidean norm of the solution to be less than a certain constant, that is, $||\mathbf{Lx}||_2^2 \leq \text{const.}$, where $\mathbf{L}$ is an appropriately chosen transformation and $||\mathbf{x}||_2^2 = \sum_i |x_i|^2$ denotes the Euclidean ℓ_2-norm. Note that, by putting a constraint on the Euclidean norm of the state, we not only narrow down the solutions but also implicitly suppress the large components of the inverted noise and reduce their spoiling effect on the solution.

[14] Using the theory of Lagrange multipliers, the dual form of the constrained version of the optimization in equation (3) is

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \frac{1}{2} ||\mathbf{y} - \mathbf{Hx}||_{\mathbf{R}^{-1}}^2 + \lambda ||\mathbf{Lx}||_2^2 \right\}, \tag{4}$$

where $\lambda > 0$ is the Lagrange multiplier or the so-called *regularizer*. This problem is a *smooth* convex quadratic programming problem and is known as the *Tikhonov regularization* with the following unique analytical solution:

$$\hat{\mathbf{x}} = \left(\mathbf{H}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{H} + 2\lambda \mathbf{L}^{\mathrm{T}} \mathbf{L} \right)^{-1} \mathbf{H}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{y}, \tag{5}$$

provided that $\mathbf{L}^{\mathrm{T}} \mathbf{L}$ is positive definite [*Tikhonov et al.*, 1977; *Hansen*, 1998; *Golub et al.*, 1999; *Hansen*, 2010]. As is evident, the $\mathbf{L}$ transformation also plays a key role in the solution of the regularized DS problem. For instance, choosing an identity matrix in equation (4) implies that we are looking for a solution with the smallest Euclidean norm (energy), while if $\mathbf{L}$ represents a derivative operator, the

above regularization term minimizes the energy in the derivative space, which naturally imposes extra smoothness on the solution.

[15] Depending on the intrinsic regularity of the underlying state and the selected $\mathbf{L}$, other choices of the regularization term are also common. For example, in the case when the $\mathbf{L}$ projects a major part of the state vector onto (near) zero values, the preferred choice is the ℓ_1-norm regularization [e.g., *Tibshirani*, 1996; *Chen et al.*, 1998, 2001]. Such a property is often called *sparse representation* in the $\mathbf{L}$ space and gives rise to the following formulation of the regularized DS problem:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \frac{1}{2} ||\mathbf{y} - \mathbf{Hx}||_{\mathbf{R}^{-1}}^2 + \lambda ||\mathbf{Lx}||_1 \right\}, \tag{6}$$

where the ℓ_1-norm is $||\mathbf{x}||_1 = \sum_i |x_i|$. By choosing $\mathbf{L}$ as a derivative operator in equation (6), in effect we minimize a measure of total variation of the state of interest. It is well understood that in this case, we typically better recover discontinuities and local jump singularities compared to the ℓ_2-norm regularization in the derivative domain. Note that, contrary to the Tikhonov regularization in equation (4), the ℓ_1-norm regularization is a *nonsmooth* convex optimization as the regularization term is nondifferentiable and the conventional iterative gradient descent methods are no longer applicable in their standard forms.

[16] One of the common approaches to treat the nondifferentiability in equation (6) is to replace the ℓ_1-norm with a smooth approximation, the so-called Huber norm, $||\mathbf{x}||_{\mathrm{Hub}} = \sum_i \rho_T(x_i)$, where

$$\rho_T(x) = \begin{cases} x^2 & |x| \leq \tau \\ \tau(2|x| - \tau) & |x| > \tau \end{cases}, \tag{7}$$

and τ denotes a nonnegative threshold (Figure 1). The Huber norm is a hybrid norm that behaves similarly to the ℓ_1-norm for values greater than the threshold τ while for smaller values it is identical to the ℓ_2-norm. From the statistical regression point of view, the sensitivity of a norm as a penalty function to the outliers depends on the (relative) values of the norm for large residuals. If we restrict ourselves to convex norms, the least sensitive ones to the large residuals or say the outliers are those with linear behavior for large input arguments (i.e., ℓ_1 and Huber). Because of this property, these norms are often called *robust* norms [*Huber*, 1964, 1981; *Boyd and Vandenberghe*, 2004]. Throughout this paper, for solving equation (6), we use the Huber relaxation due to its simplicity, efficiency, and adaptivity to all of the concerning classes of DS, DF, and DA problems. This issue is further discussed in Appendix B.

[17] In general, the first term in equations (4) and (6) measures how well the solution approximates the given (noisy) data, while the second term imposes a specific regularity on the solution. In effect, the regularizer plays a trade-off role between making the fidelity to the observations sufficiently large, while not imposing too much regularity (degree of smoothness) on the solution. The smaller the value of λ, the more weight is given to fitting the (noisy) observations which typically results in solutions that are less regular and prone to overfitting. On the other hand, the larger the value of λ, the more weight is given to
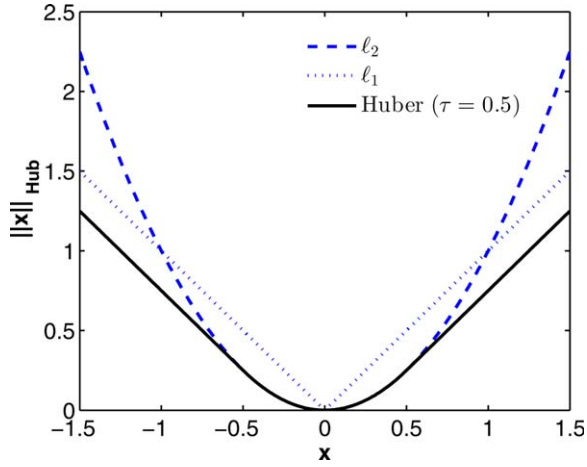
Figure 1. The Huber penalty is a smooth relaxation of the ℓ_1-norm which acts quadratically for input values smaller than the threshold τ, while it behaves linearly for larger inputs. For heavy-tailed inputs, linear penalization in the regularization term is advantageous compared to the quadratic penalization in which the overall cost function becomes dominated by a few large values in the tail of the distribution.

the regularization term which may result in a biased and overly smooth solution. Clearly, the goal is to find a balance between the two terms such that the solution is sufficiently close to the observations while obeying the underlying degree of regularity.

[18] It is important to note that, under the assumption of Gaussian error, the WLS problem (3) can be viewed as the maximum likelihood (ML) estimator of the HR field. On the other hand, the regularized problems (4) and (6) can be viewed as the Bayesian maximum a posteriori (MAP) estimator of the HR field. Indeed, the regularization terms refer to the prior knowledge about the probabilistic distribution of the state of interest. In other words, in equations (4) and (6), we implicitly assume that under the chosen transformation $\mathbf{L}$, the state of interest can be well explained by the family of multivariate Gaussian $p(\mathbf{x}) \propto \exp\left(-\lambda||\mathbf{Lx}||_2^2\right)$ and Laplace $p(\mathbf{x}) \propto \exp\left(-\lambda||\mathbf{Lx}||_1\right)$ densities, respectively. Similarly, selecting the Huber norm can also be interpreted as assuming that $\log p(\mathbf{x}) \propto \sum_i \rho_T(x_i)$, which is equivalent to considering the *Gibbs* density function as the prior probability model [*Geman and Geman*, 1984; *Schultz and Stevenson*, 1994] (see Appendix A for details). The equivalence between the regularization, which imposes constraints on the regularity of the solution, and its Bayesian interpretation, which takes into account the prior probabilistic knowledge about the state of interest, is very insightful. This relationship establishes an important duality which can guide the selection of the regularization method depending on the statistical properties of the state of interest in the real or derivative space.

### 3.2. Application in Rainfall Downscaling

3.2.1. Problem Formulation

[19] As is evident, to downscale a remotely sensed hydrometeorological state, using the explained discrete regularization methods, we need to have proper mathematical models for the downgrading operator and also a priori knowledge about the form of the regularization term. Clearly, in the presented framework, the downgrading operator needs to be a linear approximation of the sampling property of the sensor. If a sensor directly measures the state of interest while its maximum frequency channel is smaller than the maximum frequency content of the state (e.g., precipitation), the result of the sensing would be a smoothed and possibly down-sampled version of the true state. Thus, each element of the observed state in a grid scale might be considered as an LR representation of the true state, lacking the HR subgrid scale variability. To have a simple and tractable mathematical model, the downgrading matrix might be considered translation invariant and decomposed into $\mathbf{H} = \mathbf{DC}$, where $\mathbf{C}$ encodes the smoothing effect and $\mathbf{D}$ contains information about the sampling rate of the sensor. To this end, let us suppose that each grid point in the LR observation is a (weighted) average of a finite size neighborhood of the true HR state around the center of the grid. In this case, the sensor smoothing property in $\mathbf{C}$ can be encoded by the filtering and convolution operations, while $\mathbf{D}$ acts as a linear operator to simulate downsampling properties of the sensor (Figure 2). Note that these matrices can be formed explicitly, while direct matrix-vector multiplication (e.g., $\mathbf{Cx}$ and $\mathbf{C}^T\mathbf{x}$, $\mathbf{x} \in \mathbb{R}^m$) requires a computational cost in the order of $\mathcal{O}(m^2)$. However, for large-scale problems, we do not need to explicitly perform these matrix-vector multiplications as there are efficient algorithms such as the fast Fourier transformation [*Cooley and Tukey*, 1965] that can perform convolution operations with computational cost of $\mathcal{O}(m \log m)$.

[20] As is evident, the smoothing kernel needs to be estimated for each sensor, possibly by learning from a library of coincidental HR and LR observations or through a direct minimization of an associated cost [e.g., *Ebtehaj et al.*, 2012]. In the absence of prior knowledge, one possible choice is to assume that the sensor observes a coarse grained (i.e., nonoverlapping box averaging) and noisy version of the true state. In other words, to produce a field at the grid scale of $s_c \times s_c$ from a 1×1, this assumption is equivalent to selecting a uniform smoothing kernel of size $s_c \times s_c$, followed by a down-sampling operation with ratio s_c (Figure 3a).

[21] The error covariance matrix $\mathbf{R}$ in the observation model (2) plays a very important role on the results of the DS problem from both the mathematical and practical perspectives. Mathematically speaking, when the error is spatially white, the error covariance matrix is diagonal without any smoothing effect on the result [e.g., *Gaspari and Cohn*, 1999]; however, spatially correlated observation errors give rise to smoother results. Moreover, correlated errors with finite correlation length give rise to band error covariance matrices, which are prone to ill conditioning. This ill-conditioning is typically more severe in the case of ensemble error covariance estimation when the number of samples is typically much smaller than the observational dimension of the problem [e.g., *Ledoit and Wolf*, 2004]. Practically speaking, this error term captures the instrumental (e.g., ground-based NEXRAD radar) error. Although practical characterization of this error term is not in the scope of this study, for operational purposes this term needs
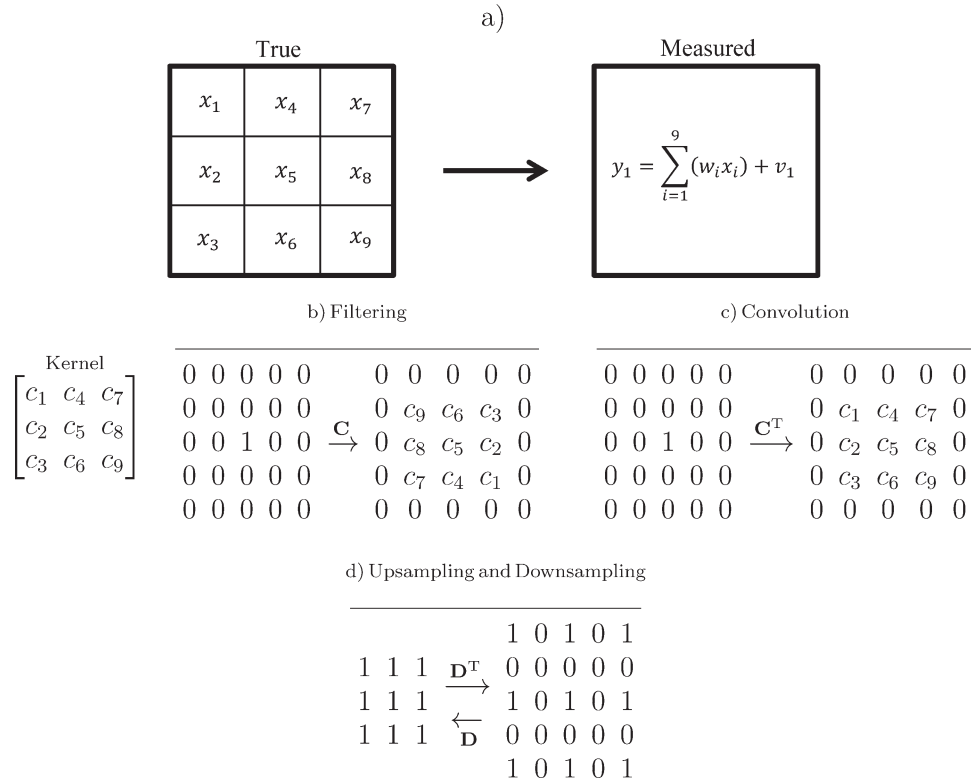
Figure 2. Two-dimensional mathematical models for the smoothing and down-sampling properties of an LR sensor via the convolution operation. (a) A simple representation of an observation model for a neighborhood of size 3×3 using a simple smoothing (averaging) observation operator. (b and c) A sample effect of the filtering operation ($\mathbf{C}$) and its transpose ($\mathbf{C}^{\mathrm{T}}$) on a discrete 2-D unit pulse, given the 3×3 kernel on the left. (d) A sample effect of the 2-D down-sampling operator ($\mathbf{D}$) and its transpose ($\mathbf{D}^{\mathrm{T}}$) with scaling ratio 2.

to be properly estimated and calibrated based on observational and theoretical studies [e.g., *Ciach and Krajewski*, 1999; *Hossain and Anagnostou*, 2005, 2006; *Krajewski et al.*, 2011; *Maggioni et al.*, 2012; *AghaKouchak et al.*, 2012].

[22] The choice of the regularization term also plays a very important role on the accuracy of the DS solution. Figure 4a demonstrates a NEXRAD reflectivity snapshot (resolution of 1×1 km) over the Texas TRMM satellite ground validation site, while Figure 4b displays the standardized histogram of the discrete Laplacian coefficients (second-order differences) and the fitted exponential of the

Figure 3. (a) A uniform smoothing (low pass) kernel of size $s_c \times s_c$. (b) The discrete (high pass) generalized Laplacian filter of size 3×3, where κ is a parameter ranging between 0 and 1. The Laplacian coefficients, obtained by filtering the 2-D state with the Laplacian kernel, are approximate measures of the second-order derivative. Throughout this paper, we choose $\kappa = 0.5$, which corresponds to the standard second-order differencing operation.

form $p(x) \propto \exp(-\lambda |x|)$. It is seen that the analyzed rainfall image exhibits (nearly) sparse representation in the derivative space with a large mass around zero and heavier tail than the Gaussian.

[23] This well-behaved non-Gaussian structure in the derivative space mainly arises due to the presence of spatial coherent and correlated patterns in the rainfall fields which contain sharp transitions (large gradients) and isolated singularities (high-intensity rain cells). In effect, over the large areas of almost uniform rainfall reflectivity values, a measure of derivative translates those values into a large number of (near) zero coefficients; however, over the less frequent jumps and isolated high-intensity rain cells, derivative coefficients are markedly larger than zero and form the tails. Note that this non-Gaussianity is due to the intrinsic spatial structure of rainfall fields and cannot be resolved by a logarithmic or power law transformation (e.g., *Z-R* relationship). It is seen that after applying a relevant *Z-R* relationship on the reflectivity fields, the shape of the rainfall histogram remains non-Gaussian and still can be approximated by the Laplace density (not shown here).

[24] The universality of this statistical structure in the distribution of derivative coefficients has been observed in many rainfall reflectivity fields [*Ebtehaj and Foufoula-Georgiou*, 2011b], denoting that the choice of the Laplace prior and ℓ_1-norm regularization is preferred in the rainfall DS problems rather than the choice of the Tikhonov
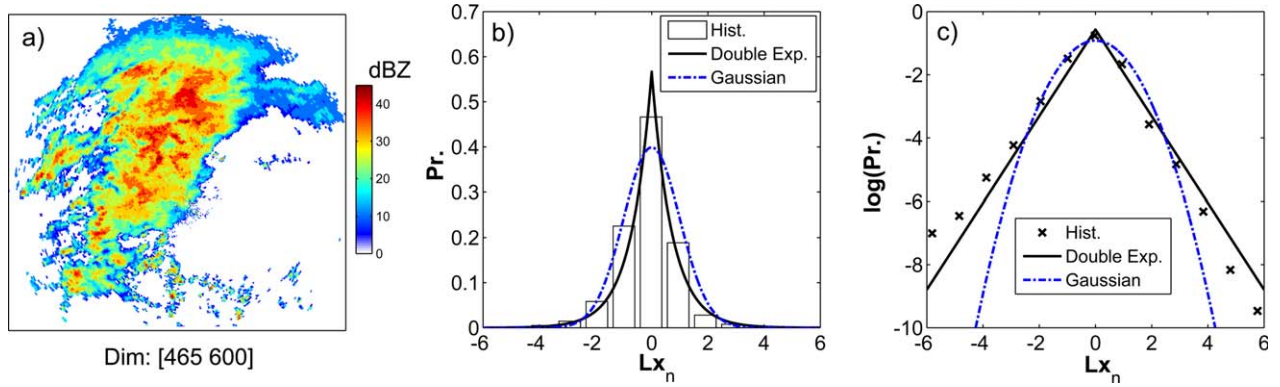
Figure 4. A rainfall reflectivity field and the distribution of its standardized Laplacian coefficients, $\mathbf{Lx}_n = \mathbf{Lx}/\text{std}(\mathbf{Lx})$, where $\text{std}(\cdot)$ is the standard deviation. (a) NEXRAD reflectivity snapshot at the TRMM GV-site in Houston, TX (HSTN) on 11/13/1998 (00:02:00 UTC) at scale 1×1 km. (b) The histogram of the standardized Laplacian coefficients, with $\kappa = 0.5$ (Figure 3b) and (c) their corresponding log histogram. Note that the zero coefficients over the nonrainy background have been excluded from the histogram analysis. The solid line in Figure 4b is the least squares fitted exponential of the form $p(x) \propto \exp(-\lambda|x|)$, and the dash-dot line shows a standard normal distribution for comparison. The log histogram in Figure 4c contrasts the heavy-tailed structure of the Laplacian coefficients versus the Gaussian distribution, clearer than the original histogram in Figure 4b.

regularization. Throughout this paper, we use the Laplacian for $\mathbf{L}$ not only for its sparsifying effect on rainfall fields but also because of our empirical evidence about its stabilizing role and computational adaptability for rainfall downscaling and data fusion problems.

[25] In practice, the histogram of the derivatives may exhibit a thicker tail than the Laplace density, requiring a heavier tail probability model, such as the Generalized Gaussian Density (GGD) of the form $p(x) \propto \exp(-\lambda|x|^p)$, where $p < 1$ [see *Ebtehaj and Foufoula-Georgiou*, 2011b]. However, using such a prior model gives rise to a nonconvex optimization problem in which convergence to the global minimum cannot be easily guaranteed. Therefore, the choice of the ℓ_1-norm (the Laplace prior) for rainfall downscaling is indeed the closest convex relaxation that can partially fulfill the strict statistical interpretation of the rainfall fields in derivative domains. Following our observations related to the distribution of the rainfall derivatives, here we direct our attention to the Huber penalty function as a smooth approximation of the ℓ_1 regularization, and cast the rainfall DS as the following constrained variational problem:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{Hx}\|_{\mathbf{R}^{-1}}^2 + \lambda \|\mathbf{Lx}\|_{\text{Hub}} \right\} \quad (8)$$
$$\text{s.t. } \mathbf{x} \geqslant 0.$$

[26] Obviously, the constraint is due to the nonnegativity of the rainfall fields. In this study, we adopted the gradient projection (GP) method [*Bertsekas*, 1999, p. 228], to solve the above variational problem (see Appendix B).
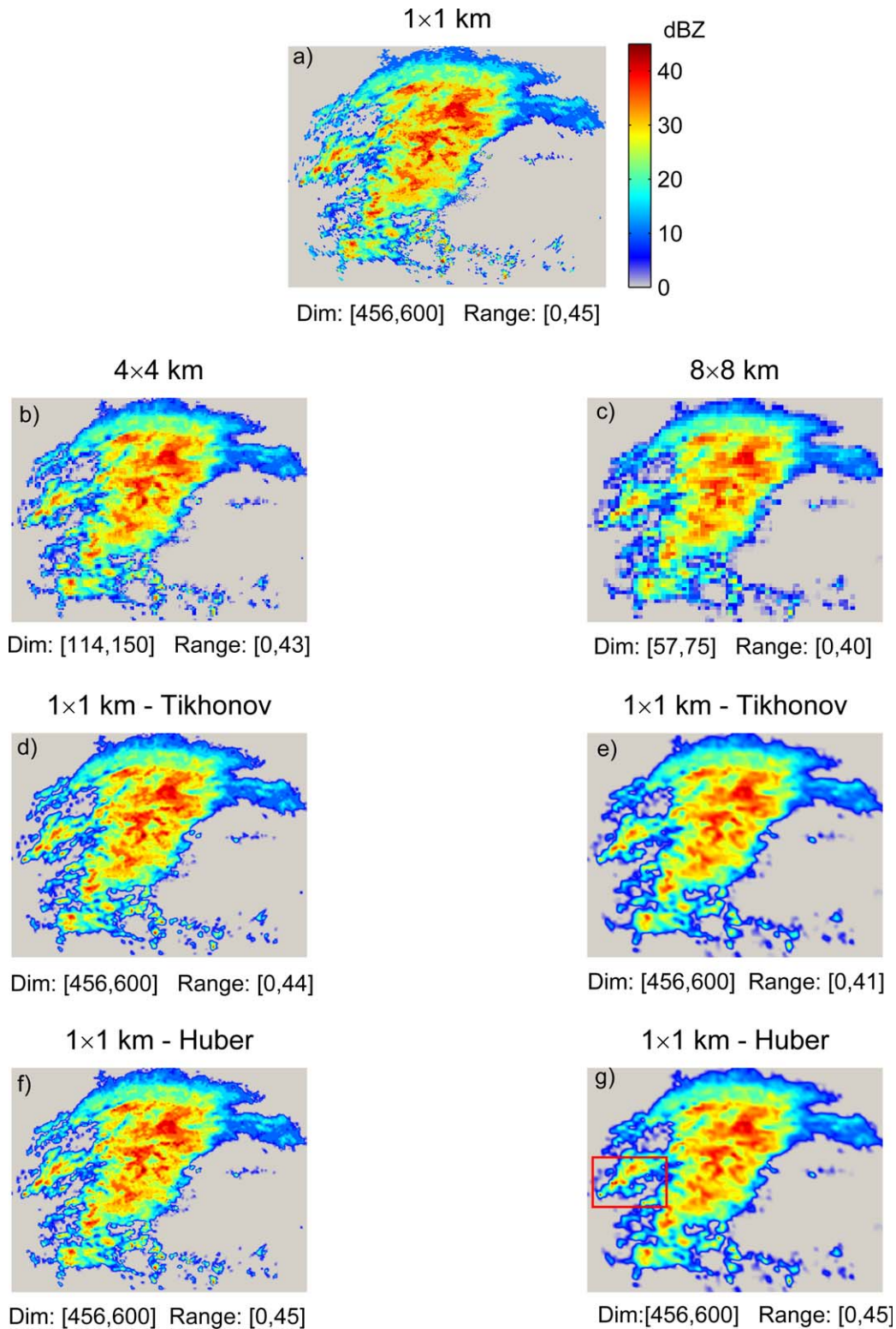
3.2.2. Rainfall Downscaling Results

[27] The same rainfall snapshot shown in Figure 4 has been used to examine the performance of the proposed regularized DS methodology. Throughout the paper, to make the reported parameters independent of the intensity range, the rainfall reflectivity fields are first scaled into the range between 0 and 1; however, the downscaling results are presented in the true range.

[28] To demonstrate the performance of the proposed regularized DS methodology, the NEXRAD HR observation $\mathbf{x}$ was assumed as the true state, while the LR observations $\mathbf{y}$ were obtained by smoothing $\mathbf{x}$ with an average filter of size $s_c \times s_c$, followed by a down-sampling operator with ratio s_c. Given the true state and constructed LR observations, we can quantitatively examine the effectiveness of the presented DS methodology by comparing the downscaled HR fields with the true HR field using some common quality metrics.

[29] Both the Huber and Tikhonov regularization methods were examined to downscale the observations from scales 4×4 and 8×8 km down to 1×1 km (Figure 5). A very small amount of white noise $\mathbf{v}$ with standard deviation of 1e-2 (5% of the standard deviation of the reference rainfall field only over the wetted areas) was added to the LR observations (equation (2)), giving rise to a diagonal error covariance matrix. In both of the regularization methods, for downscaling from 4-to-1 and 8-to-1 km in grid spacing, the regularization parameter λ was set to 5e-3 and 1e-2, respectively. These values were selected through trial and error; however, there are some formal methods for automatic estimation of this parameter, which are left for future work [e.g., *Hansen*, 2010, chap. 5]. In our experiments, it turned out that small values of the Huber threshold τ, typically less than 10% of the field maximum range of variability, led to a successful recovery of isolated singularities and local extreme rainfall cells (Figures 6 and 7).

[30] In the studied snapshot, coarse graining of the rainfall reflectivity fields to the scales of 4×4 and 8×8 km was equivalent to loosing almost 20% and 30% of the rainfall energy in the reflectivity domain in terms of the relative root-mean-square error (RMSE), RMSE $= \|\mathbf{x} - \hat{\mathbf{x}}\|_2 / \|\mathbf{x}\|_2$ (see Table 1). Note that to compute the RMSE of the LR observations, the size of those fields was extended to the size of the true field using the nearest neighborhood interpolation, that is, each LR pixel was replaced with $s_c \times s_c$ pixels with the same intensity value. In addition to the

Figure 5. Sample results of the rainfall regularized downscaling (DS). (a) True HR rainfall reflectivity: NEXRAD snapshot at the TRMM GV-site in Houston, TX (HSTN) on 11/13/1998 (00:02:00 UTC) at resolution 1×1 km. (b and c) The synthetically generated, 4×4 and 8×8 km, coarse-scale and noisy observations of the true rainfall reflectivity field. Left column: (d) Tikhonov and (f) Huber regularization results for downscaling from 4 to 1 km ($\tau = 0.02$). Right column: (e) Tikhonov and (g) Huber regularized DS for downscaling from 8 to 1 km ($\tau = 0.04$). Zooming views of the delineated box in Figure 5g are shown in Figure 6.
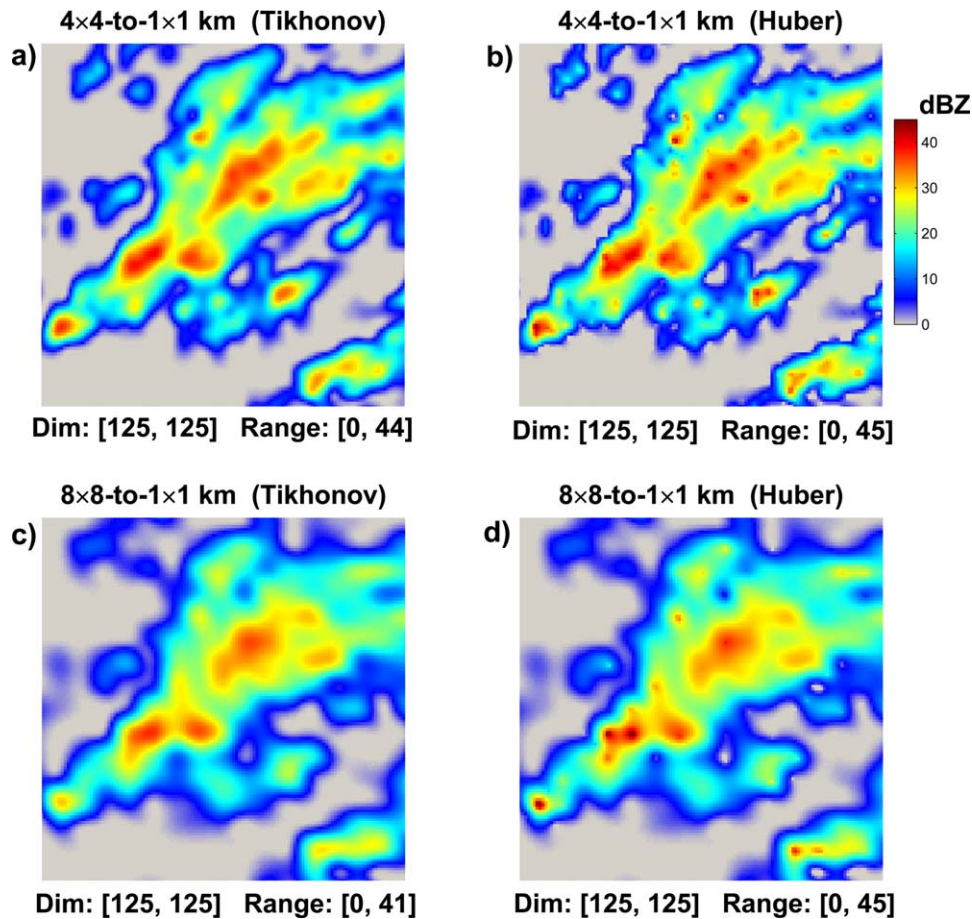
Figure 6. A zooming view for comparing qualitatively the Tikhonov (a and c) versus the Huber (b and d) regularization for the downscaling (DS) example in Figure 5. The results indicate a marginally improved performance by the Huber regularization, especially for smaller scaling ratio. The Huber regularization yields sharper results and is more capable to recover high-intensity rainfall cells and the correct range of variability; see Table 1 for quantitative comparison using a suit of metrics and Figure 7.

relative RMSE measure, we also used three other metrics: (1) relative mean absolute error (MAE), $\text{MAE} = \| \mathbf{x} - \hat{\mathbf{x}} \|_1 / \| \mathbf{x} \|_1$; (2) a logarithmic measure often called the peak signal-to-noise ratio (PSNR), $\text{PSNR} = 20 \log_{10}(\max(\hat{\mathbf{x}})/\text{std}(\mathbf{x} - \hat{\mathbf{x}}))$, where $\text{std}(\cdot)$ denotes the standard deviation; and (3) the structural similarity index (SSIM) by *Wang et al*. [2004]. The PSNR (in dB) represents a measure that not only contains RMSE information but also encodes the recovered range. The latter metric varies between -1 and 1 and the upper bound refers to the case where the estimated $\hat{\mathbf{x}}$ and reference (true) field $\mathbf{x}$ are perfectly matched. The SSIM metric is popular in the image processing community as it takes into account not only the marginal statistics such as the RMSE but also the correlation structure between the estimated and reference field. This metric seems very promising for analyzing the forecast mismatch with observations in hydrometeorological studies, especially when the large-scale systematic errors (e.g., displacement error) might be more dominant than the random errors; see *Ebtehaj et al*. [2012] for applications of SSIM in rainfall downscaling.

[31] On average, it is seen that almost 25% of the lost relative energy of the rainfall reflectivity fields can be restored via the regularized DS (Table 1). The ℓ_2-norm regularization led to smoother results, and as the scaling ratio grows, this regularization was almost incapable to recover the peaks and the correct variability range of the rainfall reflectivity field (Figure 6). Typically, as expected, the Huber-norm regularization results are slightly better than the Tikhonov ones, although not always significantly. For large scaling ratios (i.e., $s_c > 4$), the results of those methods tended to coincide in terms of the selected lump quality metrics such as the RMSE. However, using the Huber regularization, the recovered range was markedly better than that by the Tikhonov regularization, as reflected in the PSNR metric and recovered range. For example, in downscaling from 8-to-1 km $\times$ km via the Tikhonov regularization, the maximum recovered reflectivity values are approximately 41 dBZ, while using the Huber-norm regularization the maximum values are 45 dBZ (Figure 5). Employing the classic *Z-R* relationship for the NEXRAD products (i.e., $Z = 300R^{1.4}$), one can easily check that the
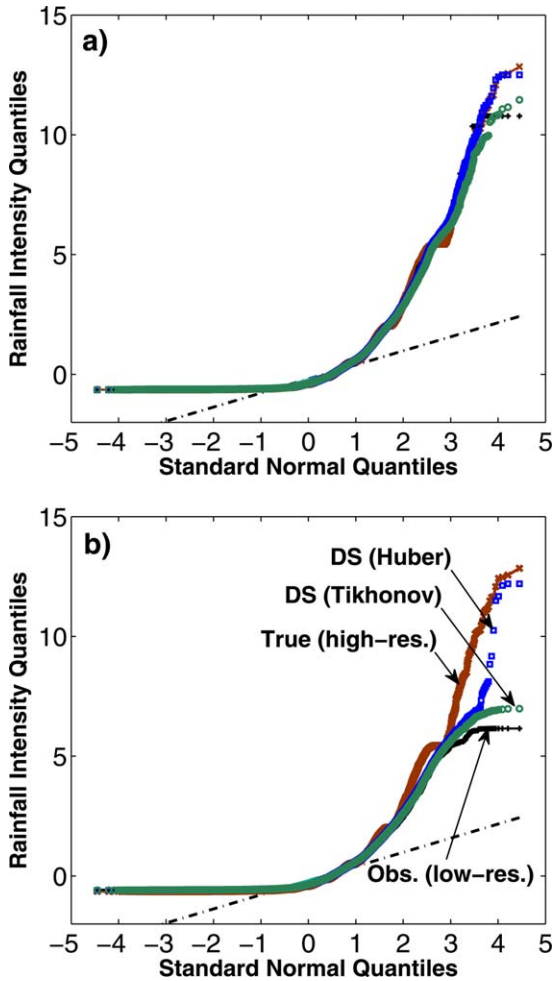
Figure 7. The quantiles of the standard normal density versus the standardized distribution of the recovered rain rates (mm/h), using $Z = 300R^{1.4}$ relationship, for the true HR field (red cross), the observed LR field (black plus), the downscaled HR fields via the Tikhonov regularization (green circle), and the Huber-norm regularization (blue square), respectively. (a and b) The quantile-quantile plots for the HR fields obtained by downscaling from 4×4 to 1×1 km and 8×8 to 1×1 km, respectively. The rainfall quantile values are only for the positive rainy part of the fields and are standardized by subtracting the mean and dividing by the standard deviation. The qq-plots signify that the Huber regularization performs better than the Tikhonov, especially over the tails, which represent the recovery of high-intensity and extreme rainfall values from the LR observations.

rain rates associated with the above reflectivity values are approximately 15 and 28 (mm/h), respectively. Therefore, although the lump quality metrics are comparable for the two methods in the reflectivity domain, the main advantage of the Huber norm over the ℓ_2-norm is the recovery of local extreme rain rates (Figure 7). It is clear from the quantile-quantile plots in Figures 7a and 7b that for a small scaling ratio, for example, $s_c = 4$, the Huber regularization can very well reproduce both the tail and the body of the true rainfall distribution. However, the tail of the recovered rainfall distribution falls below the true rainfall distribution

for larger scaling ratio, e.g., $s_c = 8$, indicating that in some high-intensity areas the method still underestimates the true field.

4. Regularized Data Fusion

4.1. Problem Formulation

[32] Analogous to the DS problem in the previous section, here we focus on the formulation of the DF problem. In the DF class of problems, typically an improved estimate of the true state is sought from a series of LR and noisy observations. Let $\mathbf{x} \in \mathbb{R}^m$ be the true state of interest while a set of N downgraded measurements, $\mathbf{y}^i \in \mathbb{R}^{n_i}$, $i = 1, \ldots, N$, are available through the following linear observation model:

$$\mathbf{y}^i = \mathbf{H}^i \mathbf{x} + \mathbf{v}, \qquad (9)$$

where $n_i \ll m$, $\mathbf{H}^i \in \mathbb{R}^{n_i \times m}$ and $\mathbf{v}^i \sim \mathcal{N}(0, \mathbf{R}^i)$ denote an uncorrelated Gaussian error in $\mathbb{R}^{n_i}$, $\mathbb{E}_{i \neq j}[\mathbf{v}^i(\mathbf{v}^j)^{\mathrm{T}}] = 0$. Compared to the DS family of problems, a DF problem is more constrained in the sense that usually there are more equations than the number of unknowns, $\Sigma_i^N n_i \gg m$, giving rise to an overdetermined linear system. As previously explained, naturally the linear WLS estimate of the true state, given the series of N observations, amounts to solving the following optimization problem:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\mathrm{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^{N} \left(\left\| \mathbf{y}^i - \mathbf{H}^i \mathbf{x} \right\|_{(\mathbf{R}^i)^{-1}}^2 \right) \right\}. \qquad (10)$$

[33] Note that the solution of the above problem not only contains information about all of the available observations (fusion) but also, with proper design of the observation operators, allows us to obtain an HR estimate of the state of interest (downscaling). Clearly, the inverse of each covariance matrix in equation (10) encodes the relative contribution or weight of each observation $\mathbf{y}^i$ in the cost function. In other words, if the elements of the covariance matrix of a particular observation vector are large compared to those of the other observation vectors, naturally the contribution of that observation to the obtained solution would be less significant.

[34] For notational convenience, the above system of equations can be augmented as follows:

$$\begin{bmatrix} \mathbf{y}^1 \\ \vdots \\ \mathbf{y}^N \end{bmatrix} = \begin{bmatrix} \mathbf{H}^1 \\ \vdots \\ \mathbf{H}^N \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{v}^1 \\ \vdots \\ \mathbf{v}^N \end{bmatrix},$$
$$\Rightarrow \underline{\mathbf{y}} = \underline{\mathbf{H}} \mathbf{x} + \underline{\mathbf{v}} \qquad (11)$$

where the concatenated error vector $\underline{\mathbf{v}}$ has the following block diagonal covariance matrix,

$$\underline{\mathbf{R}} = \mathbb{E}\left[\underline{\mathbf{v}}\,\underline{\mathbf{v}}^{\mathrm{T}}\right] = \begin{bmatrix} \mathbf{R}^1 & & 0 \\ & \ddots & \\ 0 & & \mathbf{R}^N \end{bmatrix}. \qquad (12)$$

[35] Therefore, the DF problem can be recast as the classic problem of estimating the true state from the augmented observation model of $\underline{\mathbf{y}} = \underline{\mathbf{H}} \mathbf{x} + \underline{\mathbf{v}}$. Thus, setting the gradient of the cost function in equation (10) to zero yields the following linear system:

Table 1. Results Showing the Effectiveness of the Proposed Regularized DS in Reducing the Estimation Error and Increasing the Accuracy of the Estimated Rainfall Fields[a]

Metric[b]	Observations Versus True		Tikhonov-DS Versus True		Huber-DS Versus True	
	4×4 km	8×8 km	4×4 km	8×8 km	4×4 km	8×8 km
RMSE	0.19	0.29	0.15	0.20	0.14	0.19
MAE	0.15	0.25	0.13	0.18	0.11	0.17
SSIM	0.71	0.56	0.78	0.66	0.80	0.66
PSNR	23.8	19.6	26.5	23.1	27.0	24.0

[a]The first two columns refer to the values of the quality metrics obtained by comparing the constructed LR observations with true 1×1 km reflectivity field. The other columns show the obtained metrics by comparing the downscaled fields with the true rainfall field. The performance of the Huber prior is slightly better than the Tikhonov regularization, especially for the small scaling ratios (i.e., 4×4 km).
[b]RMSE, relative root-mean-square error; MAE, relative maximum absolute error; SSIM, structural similarity; and PSNR, peak signal to noise ratio (see section 3.2.2 for definitions).

$$\left(\underline{\mathbf{H}}^{\mathrm{T}}\underline{\mathbf{R}}^{-1}\underline{\mathbf{H}}\right)\hat{\mathbf{x}} = \underline{\mathbf{H}}^{\mathrm{T}}\underline{\mathbf{R}}^{-1}\underline{\mathbf{y}}. \tag{13}$$

[36] This problem is overdetermined with a unique solution; however, the Hessian $\left(\underline{\mathbf{H}}^{\mathrm{T}}\underline{\mathbf{R}}^{-1}\underline{\mathbf{H}}\right)$ is likely to be very ill-conditioned. This ill-conditioning typically gives rise to an *unstable* solution with large estimation error [e.g., *Elad and Feuer*, 1997; *Hansen*, 2010]. Similar to the DS problem, one possible remedy for stabilizing the solution is the regularization. Recalling the formulation discussed in the previous section, a general regularized form of the rainfall DF problem can be written as

$$\hat{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x}} \left\{ \frac{1}{2} ||\underline{\mathbf{y}} - \underline{\mathbf{H}}\mathbf{x}||_{\underline{\mathbf{R}}^{-1}}^{2} + \lambda \psi_{\mathbf{L}}(\mathbf{x}) \right\}, \\ \text{s.t. } \mathbf{x} \geqslant 0 \tag{14}$$

where the convex regularization function $\psi_{\mathbf{L}}(\mathbf{x})$ can take different penalty norms, such as the Tikhonov $||\mathbf{Lx}||_{2}^{2}$, the ℓ_1-norm $||\mathbf{Lx}||_{1}$, or the Huber norm $||\mathbf{Lx}||_{\mathrm{Hub}}$. As is evident, similar to the DS problem, solution of equation (10) is equivalent to the frequentist ML estimator of the HR field while equation (14) is the Bayesian MAP estimator. For further explanations and statistical interpretations please see Appendix A.

4.2. Application in Rainfall Data Fusion and Results

[37] To quantitatively evaluate the effectiveness of the proposed regularized DF methodology for rainfall data, we constructed two synthetic LR and noisy observations from the original HR NEXRAD reflectivity snapshot. To resemble different sensing protocols and specifications, we chose different smoothing and down-sampling operations to construct each of the synthetic observation fields. The first observation field $\mathbf{y}^1$ was produced at resolution 6×6 km using a simple averaging filter of size 6×6, followed by a down-sampling ratio of $s_c = 6$. Analogously, the second field $\mathbf{y}^2$ was generated at scale 12×12 km using a Gaussian smoothing kernel of size 12×12 with a standard deviation of 4 km, followed by a down-sampling ratio of $s_c = 12$. To resemble the measurement random error, white Gaussian errors with standard deviations of 1e-2 and 2e-2 were also added respectively, which are equivalent to 5% and 10% of the standard deviation of the reference rainfall field only over the wetted areas. Roughly speaking, this selection of the error magnitudes implies that the degree of confidence (relative weight) on the observations at 6×6 km is twice that of the observations at

12×12 km. Here we only restrict our consideration to the Huber norm regularization because of its consistency with the underlying rainfall statistics and its better performance in recovering of the rainfall heavy-tailed structure (Figure 7). To solve the DF problem, we have used the same settings for the gradient projection (GP) method as explained in Appendix B.

[38] The solution of the ill-conditioned WLS formulation or the ML estimator in equation (10) is blocky, out of range, and severely affected by the amplified inverted noise (Figure 8c). On the other hand, the regularized DF can properly restore a fine-scale and coherent estimate of the rainfall field. The results show that more than 30% of the uncaptured subgrid energy of the examined rainfall reflectivity field can be restored through solving the proposed methodology (Table 2). As is evident, improvements of the selected fidelity measures in the DF problem are more pronounced compared to the results of the DS experiment (see Table 1). This naturally arises because more observations are available in the DF problem than the DS one, and thus the solution is better constrained. In terms of the selected lump metrics, analogous to the DS problem, we observed that the Huber-norm regularization is marginally better than the Tikhonov regularization, which is not reported here. However, as expected, in terms of recovery of the heavy-tailed structure of the rainfall, it is verified that the Huber-norm regularization can capture the lost extreme values much better than the Tikhonov regularization (see Figure 9). It is clear from Figure 9 that the Huber-norm regularization very well captures the local extreme rainfall intensity values while the Tikhonov regularization falls short and can only partially recover those extreme intensities.

5. Regularized Variational Data Assimilation

5.1. Problem Formulation

[39] Compared to the previously explained problems of downscaling and data fusion, the data assimilation problem is more involved in the sense that we also need to incorporate the evolution of a dynamical system in the estimation process. Despite the increased complexity, DA shares the same principles with the explained formulations of the DS and DF problems, from the estimation point of view. Here we briefly explain the classic linear three-dimensional variational (3D-VAR) data assimilation scheme and extend its formulation to a regularized format. Sample results of the
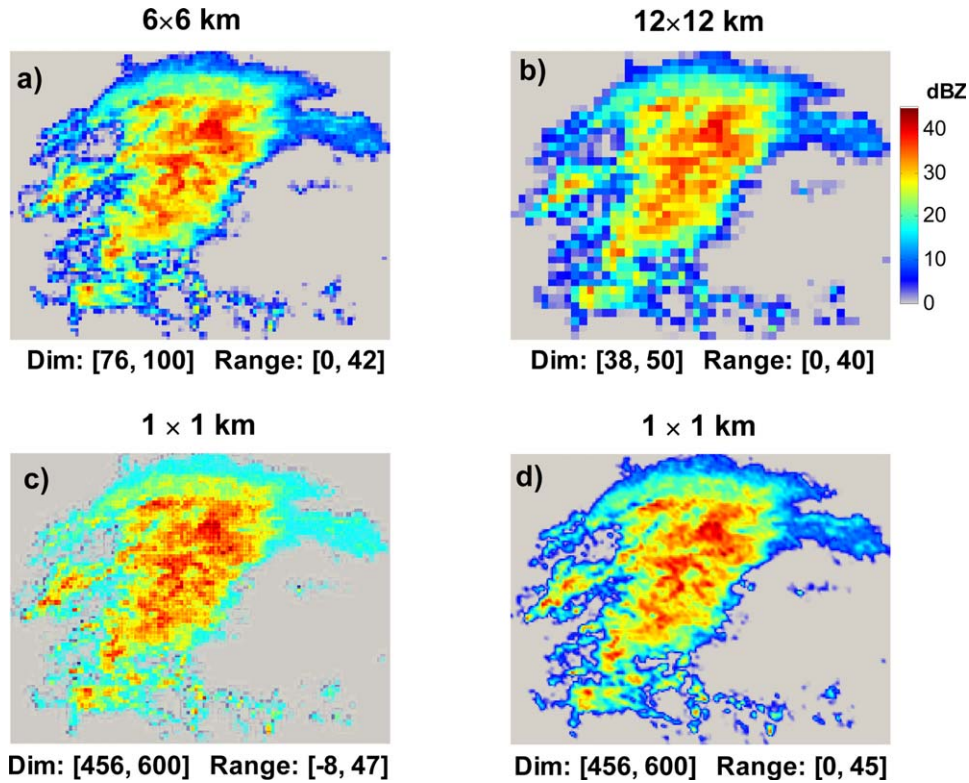
Figure 8. Data fusion and downscaling of multisensor remotely sensed rainfall reflectivity fields using the Huber regularization. (a and b) Reconstructed LR and noisy rainfall observations at scale 6 and 12 km in grid spacing. (c) The results of the WLS solution in equation (10) and (d) the solution of the regularized DF using the Huber norm with $\lambda = $ le-3 and $\tau = $ le-2.

regularized variational data assimilation problem are illustrated on the estimation of the initial conditions of the heat equation in a 3D-VAR setting.

[40] The 3D-VAR is a memoryless assimilation method. In other words, at each time step, the best estimate of the true initial state or analysis state is obtained based only on the present-time noisy observations and the background state. The analysis is then used for forecasting the state at the next time step and so on.

[41] Suppose that the true initial state of interest at discrete time t_0 is denoted by $\mathbf{x}_0 \in \mathbb{R}^m$, the observation is $\mathbf{y}_0 \in \mathbb{R}^n$, and $\mathbf{x}_0^b \in \mathbb{R}^m$ represents the background state. In the linear 3D-VAR data assimilation problem, obtaining the analysis state $\mathbf{x}_0^a \in \mathbb{R}^m$ amounts to finding the minimum point of the following cost function:

Table 2. Values of the Selected Fidelity Metrics in the Rainfall DF Experiment Using the Huber Regularization, see Section 3.2.2 for the Definitions[a]

Metric	Observations Versus True		Huber-DF Versus True
	6×6 km	12×12 km	1×1 km
RMSE	0.25	0.35	0.17
MAE	0.21	0.32	0.15
SSIM	0.60	0.50	0.72
PSNR	21.3	18.1	25.0

[a]Here the first two columns refer to comparison of the LR (6×6 and 12×12 km) observations with the true rainfall field, and the last column presents the metrics obtained by comparing the DF results with the true field.
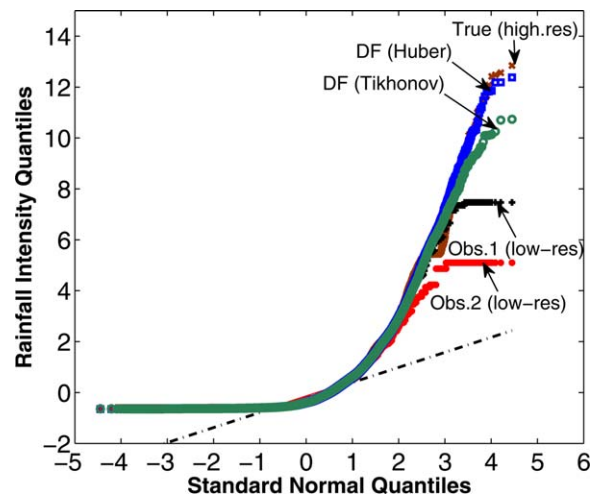
Figure 9. Quantiles of the standardized distribution of the recovered rain rates (mm/h), using $Z = 300R^{1.4}$ relationship, versus standard normal quantiles. It is clear that the Huber-norm regularization results in a better recovery of the rainfall extremes than the Tikhonov regularization. Evidently, because of extra information coming from multiple sensory data, the recovery of extreme rain rates is improved in the DF experiment compared to the DS results; see Figure 7.

$$\mathcal{J}_{3D}(\mathbf{x}_k) = \frac{1}{2}||\mathbf{x}_0^b - \mathbf{x}_0||_{\mathbf{B}^{-1}}^2 + \frac{1}{2}||\mathbf{y}_0 - \mathbf{H}\mathbf{x}_0||_{\mathbf{R}^{-1}}^2. \qquad (15)$$

[42] In the cost function (15), $\mathbf{B} \in \mathbb{R}^{m \times m}$ and $\mathbf{R} \in \mathbb{R}^{n \times n}$ are the background and observation error covariance matrices and $\mathbf{H}$ is the observation operator. The analysis is then defined as the minimizer of equation (15), denoted as $\mathbf{x}_0^a = \operatorname{argmin}_{\mathbf{x}_0} \{\mathcal{J}_{3D}(\mathbf{x}_0)\}$. Clearly, this 3D-VAR problem is a WLS problem, which has the following analytical solution:

$$\mathbf{x}_0^a = \left(\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\left(\mathbf{B}^{-1}\mathbf{x}_0^b + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}_0\right). \qquad (16)$$

[43] Because the error covariance matrices are positive definite, the matrix $\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$ is always positive definite and hence invertible. Thus, solution of the 3D-VAR requires no rank or dimension assumption on $\mathbf{H}$. However, this problem might be very ill-conditioned depending on the architecture of the covariance matrices and the measurement operator.

[44] Analogous to the previous discussions, the generic regularized form of the linear 3D-VAR under the predetermined transformation $\mathbf{L}$ might be considered as follows:

$$\mathbf{x}_0^a = \operatorname*{argmin}_{\mathbf{x}_k} \{\mathcal{J}_{3D}(\mathbf{x}_0) + \lambda\psi_{\mathbf{L}}(\mathbf{x}_0)\}, \qquad (17)$$

where $\psi_{\mathbf{L}}(\mathbf{x}_0)$ can take any of the explained regularization penalty functions, including the smooth Tikhonov $||\mathbf{L}\mathbf{x}_0||_2^2$, the nonsmooth ℓ_1-norm $||\mathbf{L}\mathbf{x}_0||_1$, and the smooth Huber norm $||\mathbf{L}\mathbf{x}_0||_{\mathrm{Hub}}$.

[45] In the above-regularized formulations, the analysis not only becomes close to the background and observations, in the weighted Euclidean sense, but it is also enforced to follow a regularity imposed by the $\psi_{\mathbf{L}}(\mathbf{x}_0)$. Here we emphasize that the regularized formulation in equation (17) typically yields a more stable and improved analysis than the classic formulation in equation (15). However, this gain comes at the price of introducing a bias in the solution whose magnitude can be kept small by proper selection of the regularization parameter λ [*Hansen*, 2010].

5.2. Application in the Study of Land Surface Heat and Mass Fluxes

[46] The promise of the proposed regularized 3D-VAR data assimilation methodology is shown via assimilating noisy and down-sampled observations into the dynamics of the heat equation. Diffusive transport of heat and moisture plays an important role in modeling of land surface water and energy balance processes [e.g., *Peter-Lidard et al.*, 1997; *Liang et al.*, 1999]. For example, in land surface energy balance, the ground heat flux is typically modeled by a 1-D heat diffusion equation for multiple layers of soil columns for which data assimilation has been the subject of special interest for improving hydrologic predictions [e.g., *Entekhabi et al.*, 1994; *Margulis et al.*, 2002; *Drusch and Viterbo*, 2007; *Bateni and Entekhabi*, 2012].

[47] Here we do not dwell into a detailed parameterization of the heat equation for a real case study of land surface heat and water budget. Rather, we only focus on a simple well-controlled assimilation experiment to demonstrate the promise of the regularized DA framework. More

specifically, we use a top-hat initial condition which is sparse in the derivative space and examine the results of the regularized DA while it evolves in time under the heat diffusion law. To this end, we construct an erroneous background state and LR noisy observations of the top-hat initial condition and then demonstrate the effectiveness of a proper regularization on the quality of the obtained analysis and forecast state. In the assimilation cycle, we obtain the analysis using the classic and regularized 3D-VAR assimilation methods and then examine those analysis states to obtain the forecast state at the next time step. The estimated analysis and forecast states are then compared with their available ground-truth counterparts.

[48] For a space-time representation of a 1-D scalar quantity $x(s, t)$, the well-known heat equation is

$$\frac{\partial x(s,t)}{\partial t} = \varepsilon\nabla^2 x(s,t) \\ x(s,0) = x_0(s), \qquad (18)$$

where $\infty < s < \infty, 0 < t < \infty$, and $\varepsilon(\mathrm{L}^2/\mathrm{T})$ denotes the diffusivity constant. In the rest of the paper for brevity and without loss of generality, we assume $\varepsilon = 1(\mathrm{L}^2/\mathrm{T})$.

[49] It is well understood that the general solution of the heat equation at time t is given by the convolution of the initial condition with the fundamental solution (kernel) as follows:

$$x(s,t) = \int K(s-r,t)x_0(r)dr, \qquad (19)$$

where

$$K(s,t) = (4\pi\varepsilon t)^{-1/2}\exp\left(\frac{-|s|^2}{4\varepsilon t}\right). \qquad (20)$$

[50] We can see that $x(s,t)$ is obtained via convolution of the initial condition with a Gaussian kernel of standard deviation $\sigma = \sqrt{2\varepsilon t}$. Clearly, estimation of the initial condition $x_0(s)$ only from the diffused observations $x(s,t)$ is an ill-posed deconvolution problem (see equation (1)).

[51] To reconstruct a 3D-VAR assimilation experiment, we assume that the true top-hat initial condition in discrete space is a vector of 256 elements ($\mathbf{x} \in \mathbb{R}^m$, where $m = 256$) as follows:

$$\mathbf{x}_0 = \begin{cases} 2 & 112 \leq x_i \leq 144 \\ 1 & \text{otherwise}. \end{cases} \qquad (21)$$

[52] We added a white Gaussian noise with $\sigma_w = 0.05$ (15% of the standard deviation of the initial state) to the true initial condition for defining the background state $\mathbf{x}_0^b$ for the assimilation experiment.

[53] We assume that the observation vector is a downgraded and noisy version of the true state, with the sensor only capturing the mean of every four neighbor elements of the true state. In other words, the observation is a noisy and LR version of the true state with one quarter of its size (Figure 10). To this end, using the linear model in equation (2), we employ the following architecture for the observation operator:
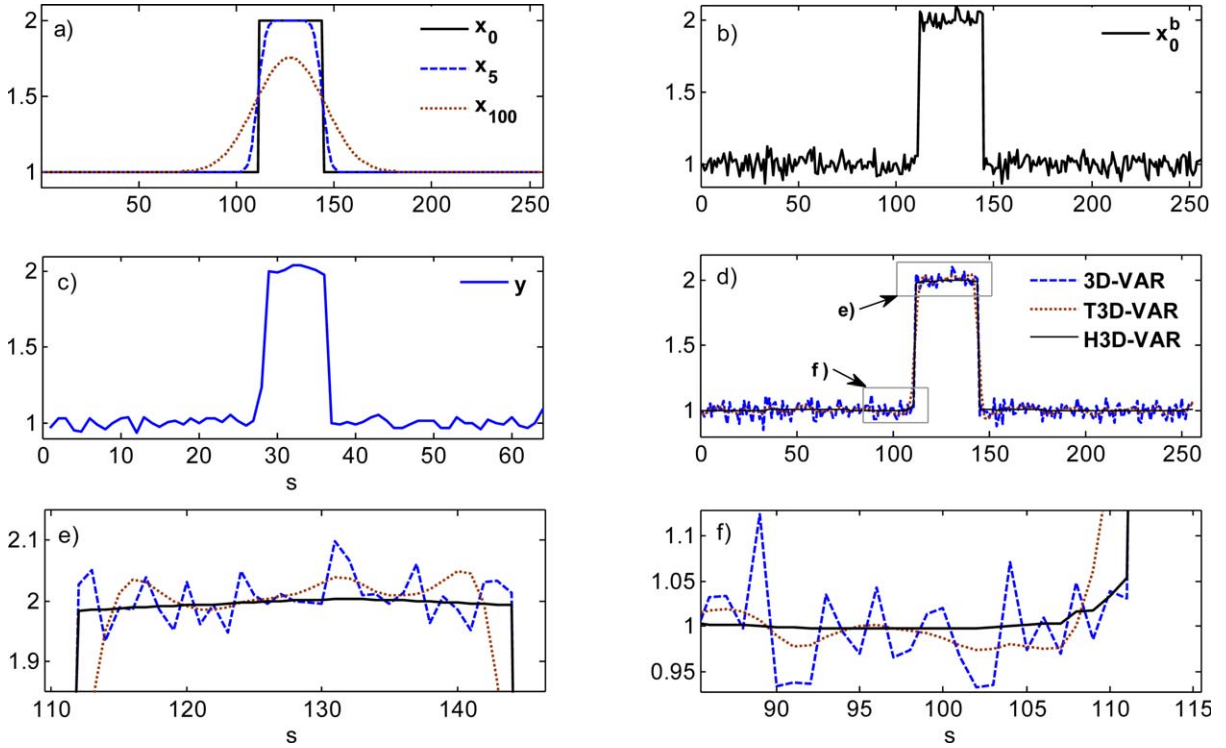
Figure 10. (a) The true initial condition $\mathbf{x}_0$ and the results of the heat equation at $t = 5$ and $t = 100$ (T) with $\varepsilon = 1$ (L^2/T). (b) The reconstructed background state by adding a white noise with $\sigma_w = 0.05$ to the true initial state and (c) the LR and noisy observations with $\sigma_v = 0.03$, respectively. (d) The results of the classic 3D-VAR and the regularized version using the explained Tikhonov (T3D-VAR) and the Huber (H3D-VAR) regularization methods (see equation (17)). (e and f) Magnified parts of the graphs in Figure 10d over the shown zooming windows.

$$\mathbf{H} = \frac{1}{4} \begin{bmatrix} 1111 & 0000 & \cdots & 0000 \\ 0000 & 1111 & \cdots & 0000 \\ \vdots & \vdots & \vdots & \vdots \\ 0000 & 0000 & \cdots & 1111 \end{bmatrix} \in \mathbb{R}^{n \times m}, \qquad (22)$$

and impose a white Gaussian error with $\sigma_v = 0.03$, equivalent to 10% of the standard deviation of the true signal.

[54] The top-hat initial condition is selected to emphasize the role of regularization, especially regularization resulting from linear penalization (i.e., the Huber or ℓ_1-norm). Clearly, the first-order derivative of the above initial condition is very sparse. In other words, the first-order derivative is zero everywhere on its domain except at the location of the two jumps, resembling a heavy tailed and sparse statistical distribution. This underlying structure prompts us to use a regularization norm with linear penalization and a first-order differencing operator for $\mathbf{L}$ in equation (17), as follows:

$$\mathbf{L} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \in \mathbb{R}^{(m-1) \times m}. \qquad (23)$$

[55] Figure 10 shows the inputs of the assimilation experiment and the results of the analysis cycle, using the classic versus the regularized 3D-VAR estimators. In this example, it is clear that the classic solution is subject to overfitting, while it slightly damps the noise. Indeed, the 3D-VAR is unable to effectively damp the high-frequency error components and recover the underlying true state. This overfitting may arise because the 3D-VAR cost function is a redundant WLS estimator and contains extra information (both observations and background) than needed for a proper estimation of the true state. On the other hand, in the regularized assimilation methods, not only the error term but also a cost associated with the regularity of the underlying state is also minimized. The Tikhonov regularization (T3D-VAR), i.e., $\psi_{\mathbf{L}}(\mathbf{x}) = ||\mathbf{Lx}||_2^2$, led to a smoother result compared to the classic one with slightly improved error statistics (Table 3). However, the result of the Huber regularization (H3D-VAR), i.e., $\psi_{\mathbf{L}}(\mathbf{x}) = ||\mathbf{Lx}||_{\text{Hub}}$, is the best. The rapidly varying noisy components are effectively damped in this regularization, while the sharp jump discontinuities have been preserved better than the T3D-VAR. The quantitative metrics in Table 3 indicate that in the analysis cycle, the RMSE and MAE metrics are improved dramatically, up to 85% in the H3D-VAR, compared to other assimilation schemes.

[56] As previously explained, there is no unique and universally accepted methodology for automated selection of the regularization parameters, namely, λ and τ. Here, to select the best parameters in the above assimilation examples, we performed a few trial and error experiments. In other words, over a feasible range of parameter values, we computed the analysis states and obtained the RMSE

Table 3. The Root-Mean-Square Error (RMSE) and the Mean Absolute Error (MAE) for the Studied Classic and Regularized 3D-VAR in the Analysis Cycle (A) and Forecast Step (F)

| Cycle	RMSE			MAE		
	3D-VAR	T3D-VAR	H3D-VAR	3D-VAR	T3D-VAR	H3D-VAR
A	0.0475	0.0397	0.0067	0.0376	0.0317	0.0043
F	0.0090	0.0088	0.0043	0.0071	0.0070	0.0033

metric by comparing them with the (known) true initial condition $\mathbf{x}_0$ (Figure 11). Note that the true initial condition is definitely not available in practice; however, here we used it to obtain the optimal values of the regularization parameters in the RMSE sense for comparison purposes and for demonstrating the importance of a proper regularization. In the T3D-VAR, as expected, larger values of the regularization parameter (λ_T) typically damp rapidly varying error components of the noisy background and observations; however, they may give rise to an overly smooth solution with larger bias and RMSE (Figures 10e and 10f). Here, for the T3D-VAR experiment, we used the value $\lambda_T = 0.05$ associated with the minimum RMSE (Figure 11a). In the H3D-VAR, in addition to the regularizer λ_H, we also need to choose the optimal threshold value τ of the Huber norm. A contour plot of the RMSE values for different choices of λ_H and τ is shown in Figure 11b. By inspection, we roughly chose $\lambda_H = 35$ and $\tau = 1.5e\text{-}3$ for the H3D-VAR assimilation experiment presented in Figure 10.

[57] The main purpose of the DA process is, indeed, to increase the quality of the forecast. Given the analysis state at initial time, we can forecast the profile of the scalar quantity, $x(s, t)$, at any future time step through the heat equation. One important property of the heat equation is its diffusivity. In other words, naturally noisy components and rapidly varying perturbations in the initial analysis are damped but become more correlated as the profile evolves in time. Thus, rapidly varying uncorrelated error components become low-varying and correlated features whose detection and removal is naturally more difficult than in the case of uncorrelated ones. Figure 12a shows the forecast profile at $t = 10(\text{T})$. The results indicate the importance of proper regularization on the quality of the forecast in the simple heat equation. The forecasts based on the classic 3D-VAR and the T3D-VAR almost coincide, while the T3D-VAR is marginally better. This behavior arises because neither of those methods could properly eliminate the noisy features in the analysis cycle; hence, low-varying error components appear in the forecast profile. However, the quality metrics in Table 3 indicate that using H3D-VAR, the RMSE and MAE of the forecast are improved by more than 50% compared to the other methods.

6. Conclusions

[58] In this paper, we presented a new direction in approaching hydrometeorological estimation problems by taking into account important intrinsic properties of the underlying state of interest, such as the presence of sharp jumps, isolated singularities (i.e., local extremes), and statistical sparsity in the derivative space. We started by explaining the concept of regularization and discussed the

common elements of the hydrometeorological problems of DS, DF, and DA as discrete linear inverse problems. We argued about the importance of proper regularization, which not only makes hydrometeorological inverse problems sufficiently well posed but also imposes the desired regularity and statistical property on the solution. Regularization methods were theoretically linked to the underlying statistical structure of the states and it was shown how information about the probability density of the state, or its derivative, can be used for proper selection of the regularization method. Specifically, we emphasized three types of regularization, namely, the Tikhonov, ℓ_1-norm, and Huber regularization methods. We argued that these methods are statistically equivalent to the maximum a posteriori (MAP) estimator while, respectively, assuming the Gaussian, Laplace, and Gibbs prior density for the state of interest in a derivative domain. It was argued that piecewise continuity of the state and the presence of frequent jumps are often translated into heavy-tailed distributions in the derivative space
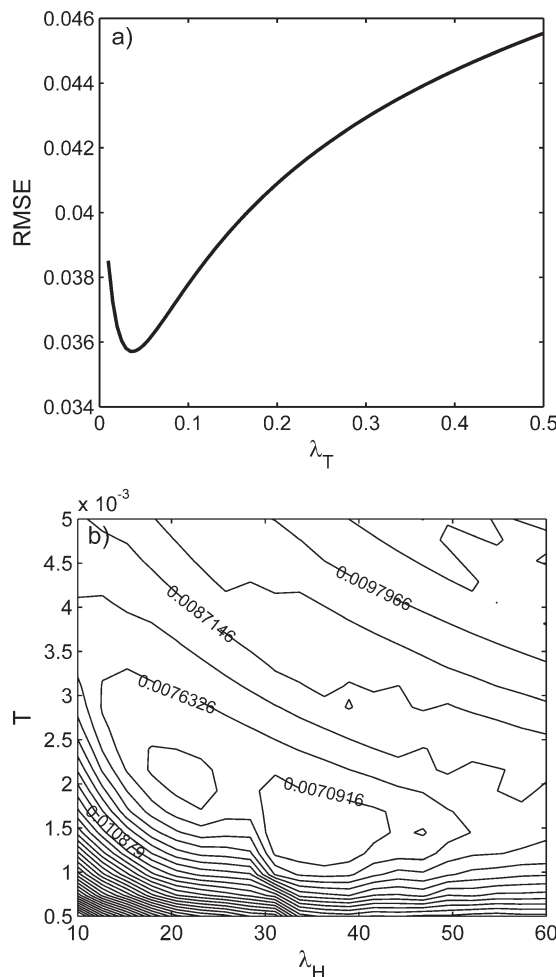
Figure 11. (a) Root-mean-square error (RMSE) of the implemented T3D-VAR as a function of the regularizer λ_T. (b) RMSE contour surface for the H3D-VAR experiment with different choices of the regularizer λ_H and the threshold value τ of the Huber norm. Clearly, depending on the choice of the regularization method, the magnitude of the regularizer might be markedly different.
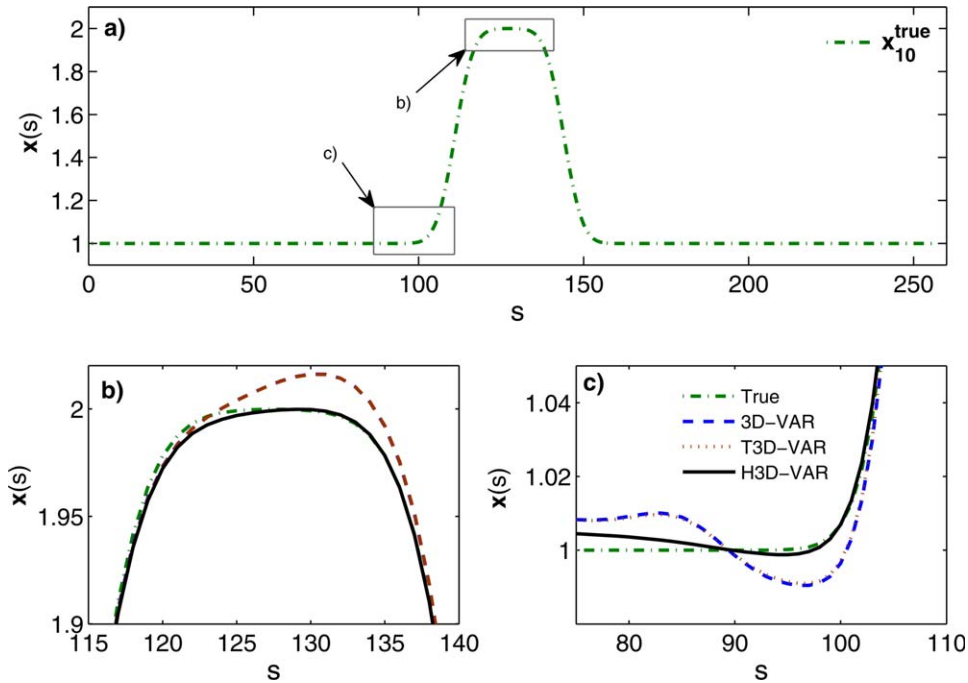
Figure 12. (a) True forecast state obtained by temporal evolution of the top-hat initial condition under the heat equation at $t = 10$ (T) (Figure 10a). (b and c) Magnified windows showing the forecast quality using classic and regularized 3D-VAR assimilation methods. It can be seen that, due to ineffective error removal by the classic 3D and T3D-VAR at the analysis cycle, large-scale correlated errors are propagated in the forecast profiles, while this problem is less substantial in the result of the H3D-VAR (see Table 3).

that favor the use of ℓ_1-norm or Huber-norm regularization methods.

[59] The effectiveness of the regularized DS and DF problems was tested via analysis of remotely sensed precipitation fields, and the superiority of the regularization with linear penalization was clearly demonstrated. The performance of the regularized DA was also studied via assimilating noisy observations into the evolution of the heat equation, which has fundamental applications in the study and data assimilation of land-surface heat and mass fluxes. We showed that adding a Huber regularization term in the variational assimilation methods outperforms the classic 3D-VAR method, especially for the case where the initial condition exhibits a sparse distribution in the derivative space (e.g., first-order derivative of the top-hat initial condition).

[60] The presented frameworks can be potentially applied to other hydrometeorological problems, such as soil moisture downscaling, fusion, and data assimilation. Clearly, proper selection of the regularization method requires careful statistical analysis of the underlying state of interest. Moreover, the problem of rainfall or soil moisture retrieval from satellite microwave radiance can be considered as a nonlinear inverse problem. This nonlinear inversion may be cast in the presented context, provided that the nonlinear kernel can be (locally) linearized with sufficient accuracy. Application of regularization in data assimilation is in its infancy (e.g., see *Freitag et al.* [2012] for a recent study) and is expected to play a significant role over the next decades, especially in the context of ensemble

methodologies for non-Gaussian and highly nonlinear dynamic systems.

Appendix A: Statistical Interpretation

[61] In this appendix, we discuss the statistical interpretation of the presented downscaling, data fusion, and data assimilation problems. We argue that the classic weighted least squares formulations can be interpreted as the frequentist maximum likelihood (ML) estimators, while the regularized formulations can be interpreted as the Bayesian maximum a posteriori (MAP) estimators. We also spell out the connection between the chosen regularization and the prior distribution of the state (or its derivative), which can guide proper selection of the regularization term in practical applications.

A1. Regularized Variational Downscaling and Data Fusion

[62] From the *frequentist* statistical point of view, it is easy to show that the WLS solution of equation (3) is equivalent to the maximum likelihood estimator (ML)

$$\hat{\mathbf{x}}_{ML} = \underset{\mathbf{x}}{\operatorname{argmax}}\; p(\mathbf{y}|\mathbf{x}), \qquad (A1)$$

given that the conditional density, $p(\mathbf{y}|\mathbf{x}) \propto \exp\left(-1/2(\mathbf{y} - \mathbf{Hx})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{Hx})\right)$, is Gaussian. Specifically, taking $-\log(\cdot)$, one can find the minimizer of the negative log-likelihood function $-\log\{p(\mathbf{y}|\mathbf{x})\}$ as follows:

$$\hat{\mathbf{x}}_{ML} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \frac{1}{2} (\mathbf{y} - \mathbf{Hx})^{\mathrm{T}} \mathbf{R}^{-1} (\mathbf{y} - \mathbf{Hx}) \right\}$$
$$= \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \frac{1}{2} ||\mathbf{y} - \mathbf{Hx}||^2_{\mathbf{R}^{-1}} \right\}, \qquad (A2)$$

which is identical to the WLS solution of problem (3).

[63] It is important to note that in the ML estimator, $\mathbf{x}$ is considered to be a deterministic variable (fixed), while $\mathbf{y}$ has a random nature. On the other hand, in the Bayesian perspective, a regularized solution of equations (4) or (6) is equivalent to the maximum a posteriori (MAP) estimator

$$\hat{\mathbf{x}}_{MAP} = \underset{\mathbf{x}}{\operatorname{argmax}} \, p(\mathbf{x}|\mathbf{y}), \qquad (A3)$$

where both $\mathbf{x}$ and $\mathbf{y}$ are considered of random nature. Specifically, using the Bayes theorem, ignoring the constant terms in $\mathbf{x}$ and applying $-\log(\cdot)$ on the posterior density $p(\mathbf{x}|\mathbf{y})$, we get

$$\hat{\mathbf{x}}_{MAP} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ -\log \left(\frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} \right) \right\}$$
$$= \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ -\log p(\mathbf{y}|\mathbf{x}) - \log p(\mathbf{x}) \right\}. \qquad (A4)$$

[64] The first term, $-\log p(\mathbf{y}|\mathbf{x})$, is just the negative log likelihood as appeared in the ML estimator and the second term is called the prior, which accounts for the a priori knowledge about the density of the state vector $\mathbf{x}$. Accordingly, the proposed Tikhonov regularization in equation (4) is equivalent to the MAP estimator assuming that the state, or the linearly transformed state $\mathbf{Lx}$, can be explained by a multivariate Gaussian of the following form:

$$\log p(\mathbf{x}) \propto \mathbf{x}^{\mathrm{T}} \mathbf{Qx}, \qquad (A5)$$

where the covariance is $\mathbf{Q} = \mathbf{L}^{\mathrm{T}} \mathbf{L}$ [e.g., *Tikhonov et al.*, 1977; *Elad and Feuer*, 1997; *Levy*, 2008]. Clearly, the choice of the ℓ_1-norm in equation (6) implies that $\log p(\mathbf{x}) \propto ||\mathbf{Lx}||_1$ or say the transformed state can be well explained by a multivariate *Laplace* density with heavier tail than the Gaussian case [e.g., *Tibshirani*, 1996; *Lewicki and Sejnowski*, 2000], while the Huber-norm regularization implies a *Gibbs* prior probability model $\log p(\mathbf{x}) \propto \sum_i \rho_T(x_i)$ for the state of interest [*Geman and Geman*, 1984; *Schultz and Stevenson*, 1994].

[65] Obviously, based on the selected type of regularization, statistical interpretation of the DF regularized class of problems is also similar to what was explained for the DS problem. In other words, given the augmented observation model in equation (11), it is easy to see that the solution of equation (10) is the ML estimator, while equation (14) can be interpreted as the MAP estimator with a prior density depending on the form of the regularization term.

A2. Regularized Variational Data Assimilation

[66] Statistical interpretation of the classic variational DA problems is a bit tricky compared to the DS and DF class of problems, mainly because of the involvement of the background information in the cost function. *Lorenc* [1986] derived the 3D-VAR cost function using Bayes theorem and called it the ML estimator [see, e.g., *Lorenc*, 1988; *Bouttier and Courtier*, 2002]. More recently, it has

been argued that the 4D-VAR, and thus as a special case the 3D-VAR cost function, can be interpreted via the Bayesian MAP estimator [*Johnson et al.*, 2005; *Freitag et al.*, 2010; *Nichols*, 2010]. For notational convenience, here we only explain the statistical interpretation of the 3D-VAR and its regularized version, which can be easily generalized for the case of the 4D-VAR problem.

[67] As discussed earlier, the ML estimator is basically a frequentist view to estimate the most likely value of an unknown deterministic variable $\mathbf{x}$ from (indirect) observations $\mathbf{y}$ of random nature. The ML estimator intuitively requires finding the state that maximizes the likelihood function as

$$\hat{\mathbf{x}}_{ML} = \underset{\mathbf{x}}{\operatorname{argmax}} \, p(\mathbf{y}|\mathbf{x}). \qquad (A6)$$

[68] Let us assume that, at the initial time step t_0, the background $\mathbf{x}_0^b$ is just a (random) realization of the true deterministic initial state $\mathbf{x}_0$. In other words, we consider $\mathbf{x}_0^b = \mathbf{x}_0 + \mathbf{w}$, where the error $\mathbf{w}$ can be well explained by a zero mean Gaussian density $\mathcal{N}(0, \mathbf{B})$, uncorrelated with the observation error, $\mathbb{E}[\mathbf{wv}^{\mathrm{T}}] = 0$. Here the background state is treated similarly to an observation that is of random nature. Thus, let us recast the problem of obtaining the analysis as a classic linear inverse problem by augmenting the available information in the form of

$$\underline{\mathbf{y}} = \underline{\mathbf{H}} \mathbf{x}_0 + \underline{\mathbf{v}}, \qquad (A7)$$

where $\underline{\mathbf{y}} = \left[(\mathbf{x}_0^b)^{\mathrm{T}}, \mathbf{y}_0^{\mathrm{T}} \right]^{\mathrm{T}}$, $\underline{\mathbf{H}} = \left[\mathbf{I}, \mathbf{H}^{\mathrm{T}} \right]^{\mathrm{T}}$, and $\underline{\mathbf{v}} \sim \mathcal{N}(0, \underline{\mathbf{R}})$, with the following block diagonal covariance matrix

$$\underline{\mathbf{R}} = \begin{bmatrix} \mathbf{B} & 0 \\ 0 & \mathbf{R} \end{bmatrix}. \qquad (A8)$$

[69] Note that $\underline{\mathbf{R}}$ is block diagonal because the background and observation errors are uncorrelated. Following the augmented representation and applying $-\log(\cdot)$, we have $\quad -\log p\left(\underline{\mathbf{y}}|\mathbf{x}_0 \right) \propto \frac{1}{2} \left(\underline{\mathbf{y}} - \underline{\mathbf{H}} \mathbf{x}_0 \right)^{\mathrm{T}} \underline{\mathbf{R}}^{-1} \left(\underline{\mathbf{y}} - \underline{\mathbf{H}} \mathbf{x}_0 \right)$; thus, it is easy to see that the ML estimator in terms of the augmented observations $\underline{\mathbf{y}}$,

$$\mathbf{x}_0^a = \underset{\mathbf{x}_0}{\operatorname{argmax}} \, p\left(\underline{\mathbf{y}}|\mathbf{x}_0 \right), \qquad (A9)$$

is equivalent to minimizing the 3D-VAR cost function in equation (15). Therefore, from the frequentist perspective, which considers the state deterministic and the observations random, the classic 3D-VAR solution is the ML estimator, assuming Gaussian observation error.

[70] On the other hand, from the Bayesian perspective, the state of interest and the available observations are considered to be random and the MAP estimator is the optimal point, which maximizes the posterior density as:

$$\hat{\mathbf{x}}_{MAP} = \underset{\mathbf{x}}{\operatorname{argmax}} \, p(\mathbf{x}|\mathbf{y}). \qquad (A10)$$

[71] Let us assume a priori that the (random) state of interest has a Gaussian density with mean $\mathbf{x}_b$ and covariance

B, that is, $p(\mathbf{x}_0) \sim \mathcal{N}(\mathbf{x}_0^b, \mathbf{B})$. More formally, this assumption implies that the deterministic background is the central (mean) forecast and is related to the random true state via $\mathbf{x}_0 = \mathbf{x}_0^b + \mathbf{w}$, where $\mathbf{w} \sim \mathcal{N}(0, \mathbf{B})$. Therefore, using Bayes theorem it immediately follows that the 3D-VAR is the MAP estimator, $\mathbf{x}_0^a = \text{argmax}_{\mathbf{x}_0} \, p(\mathbf{x}_0|\mathbf{y})$, assuming a Gaussian prior for the true state of interest.

[72] In conclusion, if we follow the frequentist approach to interpret the classic 3D-VAR in equation (15), the regularized 3D-VAR in equation (17) can be interpreted as the MAP estimator, where the prior density is characterized by the regularization term. On the other hand, taking the MAP interpretation for the classic 3D-VAR, the regularized version might be understood as the MAP estimator, which also accounts for an extra and independent prior on the distribution of the state under the **L** transformation.

Appendix B: Gradient Projection Method for the Huber Regularization

[73] Here we present the gradient projection (GP) method, using the Huber regularization, only for the downscaling (DS) problem, which can be easily generalized to the data fusion (DF) and data assimilation (DA) cases. In the case of the DS problem, the cost function and gradient of the Huber regularization with respect to the elements of the downscaled field are

$$\mathcal{J}(\mathbf{x}) = \frac{1}{2}||\mathbf{y} - \mathbf{Hx}||_{\mathbf{R}^{-1}}^2 + \lambda||\mathbf{Lx}||_{\text{Hub}} \qquad (B1)$$

$$\nabla\mathcal{J}(\mathbf{x}) = \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{Hx}) + \lambda\mathbf{L}^{\mathrm{T}}\rho_T'(\mathbf{Lx}), \qquad (B2)$$

where

$$\rho_T'(x) = \begin{cases} 2x & |x| \le \tau \\ 2\tau\,\text{sign}\,(x), & |x| > \tau. \end{cases} \qquad (B3)$$

[74] As is evident, the cost function in (B1) is a smooth and convex function. Thus, its minimum can be easily obtained using efficient first-order gradient descent methods in large dimensional problems. However, rainfall is a positive process and in order to obtain a feasible downscaled field $\hat{\mathbf{x}}$, the regularized DS problem needs to be solved on the nonnegative orthant $\{\mathbf{x}|x_i \ge 0, \forall i = 1, \ldots, m\}$,

$$\begin{aligned} \hat{\mathbf{x}} &= \text{argmin}\,\{\mathcal{J}(\mathbf{x})\} \\ &\text{s.t. } \mathbf{x} \geq 0. \end{aligned} \qquad (B4)$$

[75] We have used one of the primitive gradient projection (GP) methods to solve the above constrained DS problem [see *Bertsekas*, 1999, p. 228]. Accordingly, to obtain the solution of equation (B4) amounts to obtaining the fixed point of the following equation:

$$\mathbf{x}_* = [\mathbf{x}_* - \alpha\nabla\mathcal{J}(\mathbf{x}_*)]^+, \qquad (B5)$$

where α is a stepsize and

$$[x]^+ = \begin{cases} 0 & \text{if } x \le 0 \\ x & \text{otherwise}, \end{cases} \qquad (B6)$$

denotes the Euclidean projection operator onto the nonnegative orthant. As is evident, the fixed point can be obtained iteratively as

$$\mathbf{x}_{k+1} = [\mathbf{x}_k - \alpha_k\nabla\mathcal{J}(\mathbf{x}_k)]^+. \qquad (B7)$$

[76] Thus, if the descent at step k is feasible (i.e., $\mathbf{x}_k - \alpha_k\nabla\mathcal{J}(\mathbf{x}_k) \geq 0$), the GP iteration becomes an ordinary unconstrained steepest descent method; otherwise, the result is mapped back onto the feasible set by the projection operator in equation (B6).

[77] In our study, the stepsize (α_k) was selected using the *Armijo rule*, or the so-called *backtracking line search*, that is, a convergent and very effective stepsize rule and depends on two constants: $0 < \xi < 0.5, 0 < \varsigma < 1$. In this method, the stepsize is assumed $\alpha_k = \varsigma^{m_k}$, where m_k is the smallest nonnegative integer for which

$$\mathcal{J}(\mathbf{x}_k - \alpha_k\nabla\mathcal{J}(\mathbf{x}_k)) \le \mathcal{J}(\mathbf{x}_k) - \xi\alpha_k\nabla\mathcal{J}(\mathbf{x}_k)^{\mathrm{T}}\nabla\mathcal{J}(\mathbf{x}_k). \qquad (B8)$$

[78] In our DS examples, the above backtracking parameters are set to $\xi = 0.2$ and $\varsigma = 0.5$ (see *Boyd and Vandenberghe* [2004, p. 464] for more explanation). In our coding, the iterations terminate either if $\frac{||\mathbf{x}_k - \mathbf{x}_{k-1}||_2}{||\mathbf{x}_{k-1}||_2} \le 10^{-5}$ or the number of iterations exceeds 200.

[79] For the above-explained gradient projection algorithm and the employed parameters, the computational cost of the proposed framework is modest for a normal desktop machine at the present time. In particular, on a Windows operating system with an Intel(R)-i7 central processing unit (2.80 GHz clock rate), the process time of the presented downscaling and data fusion experiments was approximately 120 s.

References

Adler, R., et al. (2003), The version 2 global precipitation climatology project GPCP monthly precipitation analysis (1979-present), *J. Hydrometeorol.*, *4*(6), 1147–1167.

AghaKouchak, A., A. Mehran, H. Norouzi, and A. Behrangi (2012), Systematic and random error components in satellite precipitation data sets, *Geophys. Res. Lett.*, *39*, L09406, doi:10.1029/2012GL051592.

Badas, M. G., R. Deidda, and E. Piga (2006), Modulation of homogeneous space-time rainfall cascades to account for orographic influences, *Nat. Hazard Earth. Syst. Sci.*, *6*(3), 427–437, doi:10.5194/nhess-6-427-2006.

Bateni, S. M., and D. Entekhabi (2012), Surface heat flux estimation with the ensemble Kalman smoother: Joint estimation of state and parameters, *Water Resour. Res.*, *48*, W08521, doi:10.1029/2011WR011542.

Bertsekas, D. P. (1999), *Nonlinear Programming*, 2nd ed., 794 pp., Athena Scientific, Belmont, Mass.

Bocchiola, D. (2007), Use of scale recursive estimation for assimilation of precipitation data from TRMM (PR and TMI) and NEXRAD, *Adv. Water Resour.*, *30*(11), 2354–2372, doi:10.1016/j.advwatres.2007.05.012.

Bouttier, F., and P. Courtier (2002), Data assimilation concepts and methods, Meteorological Training Course Lecture Series, ECMWF, p. 59.

Boyd, S., and L. Vandenberghe (2004), *Convex Optimization*, 716 pp., Cambridge Univ. Press, New York.

Candes, E. J., and T. Tao (2006), Near-optimal signal recovery from random projections: Universal encoding strategies?, *IEEE Trans. Inf. Theory*, *50*(12), 5406–5425, doi:10.1109/TIT.2006.885507.

Castro, C. J., A. S. Pielke Roger, and G. Leoncini (2005), Dynamical downscaling: Assessment of value retained and added using the Regional Atmospheric Modeling System RAMS, *J. Geophys. Res.*, *110*, D05108, doi:10.1029/2004JD004721.

Chen, S. S., D. L. Donoho, and M. A. Saunders (1998), Atomic decomposition by basis pursuit, *SIAM J. Sci. Comput.*, *20*, 33–61.

Chen, S., D. Donoho, and M. Saunders (2001), Atomic decomposition by basis pursuit, *SIAM Rev.*, *43*(1), 129–159.

Ciach, G. J., and W. F. Krajewski (1999), On the estimation of radar rainfall error variance, *Adv. Water Resour.*, *22*(6), 585–595.

Cooley, J. W., and J. W. Tukey (1965), An algorithm for the machine calculation of complex Fourier series, *Math. Comput.*, *19*(90), 297–301.

Courtier, P., and O. Talagrand (1990), Variational assimilation of meteorological observations with the direct and adjoint shallow-water equations, *Tellus, Ser. A*, *42*(5), 531–549.

Courtier, P., J.-N. Thépaut, and A. Hollingsworth (1994), A strategy for operational implementation of 4D-VAR, using an incremental approach, *Q. J. R. Meteorol. Soc.*, *120*(519), 1367–1387, doi:10.1002/qj.49712051912.

Daley, R. (1993), *Atmospheric Data Analysis*, 472 pp., Cambridge Univ. Press, Cambridge, U.K.

Deidda, R. (2000), Rainfall downscaling in a space-time multifractal framework, *Water Resour. Res.*, *36*(7), 1779–1794.

Drusch, M., and P. Viterbo (2007), Assimilation of screen-level variables in ECMWF's integrated forecast system: A study on the impact on the forecast quality and analyzed soil moisture, *Mon. Weather Rev.*, *135*(2), 300–314, doi:10.1175/MWR3309.1.

Ebtehaj, A. M., and E. Foufoula-Georgiou (2011a), Adaptive fusion of multisensor precipitation using Gaussian-scale mixtures in the wavelet domain, *J. Geophys. Res.*, *116*, D22110, doi:10.1029/2011JD016219.

Ebtehaj, A. M., and E. Foufoula-Georgiou (2011b), Statistics of precipitation reflectivity images and cascade of Gaussian-scale mixtures in the wavelet domain: A formalism for reproducing extremes and coherent multiscale structures, *J. Geophys. Res.*, *116*, D14110, doi:10.1029/2010JD015177.

Ebtehaj, A. M., E. Foufoula-Georgiou, and G. Lerman (2012), Sparse regularization for precipitation downscaling, *J. Geophys. Res.*, *116*, D22110, doi:10.1029/2011JD017057.

Elad, M. (2010), *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, 376 pp., Springer, New York.

Elad, M., and A. Feuer (1997), Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images, *IEEE Trans. Image Process.*, *6*(12), 1646–1658, doi:10.1109/83.650118.

Entekhabi, D., H. Nakamura, and E. Njoku (1994), Solving the inverse problem for soil moisture and temperature profiles by sequential assimilation of multifrequency remotely sensed observations, *IEEE Trans. Geosci. Remote Sens.*, *32*(2), 438–448, doi:10.1109/36.295058.

Freitag, M. A., N. K. Nichols, and C. J. Budd (2010), L1-regularisation for ill-posed problems in variational data assimilation, *Proc. Appl. Math. Mech.*, *10*(1), 665–668, doi:10.1002/pamm.201010324.

Freitag, M. A., N. K. Nichols, and C. J. Budd (2012), Resolution of sharp fronts in the presence of model error in variational data assimilation, *Q. J. R. Meteorol. Soc.*, *139*, 742–757, doi:10.1002/qj.2002.

Gaspari, G., and S. E. Cohn (1999), Construction of correlation functions in two and three dimensions, *Q. J. R. Meteorol. Soc.*, *125*(554), 723–757, doi:10.1002/qj.49712555417.

Geman, S., and D. Geman (1984), Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intell.*, *6*(6), 721–741.

Golub, G., P. Hansen, and D. O'Leary (1999), Tikhonov regularization and total least squares, *SIAM J. Matrix Anal. Appl.*, *21*(1), 185–194.

Gorenburg, I. P., D. McLaughlin, and D. Entekhabi (2001), Scale-recursive assimilation of precipitation data, *Adv. Water Resour.*, *24*(9–10), 941–953, doi:10.1016/S0309-1708(01)00033-1.

Gupta, V., and E. Waymire (1993), A statistical analysis of mesoscale rainfall as a random cascade, *J. Appl. Meteorol.*, *32*, 251–251.

Hansen, P. (1998), *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*, vol. *4*, Soc. for Ind. Math., Philadelphia, Pa.

Hansen, P. (2010), *Discrete Inverse Problems: Insight and Algorithms*, vol. *7*, Soc. for Ind. and Appl. Math., Philadelphia, Pa.

Hong, Y., K. Hsu, S. Sorooshian, and X. Gao (2004), Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system, *J. Appl. Meteorol.*, *43*(12), 1834–1853.

Hossain, F., and E. N. Anagnostou (2005), Numerical investigation of the impact of uncertainties in satellite rainfall estimation and land surface model parameters on simulation of soil moisture, *Adv. Water Resour.*, *28*(12), 1336–1350.

Hossain, F., and E. N. Anagnostou (2006), A two-dimensional satellite rainfall error model, *IEEE Trans. Geosci. Remote Sens.*, *44*(6), 1511–1522.

Huber, P. (1964), Robust estimation of a location parameter, *Ann. Math. Stat.*, *35*(1), 73–101.

Huber, P. (1981), *Robust Statistics*, vol. 1, John Wiley, New York.

Huffman, G., R. Adler, B. Rudolf, U. Schneider, and P. Keehn (1995), Global precipitation estimates based on a technique for combining satellite-based estimates, rain gauge analysis, and NWP model precipitation information, *J. Clim.*, *8*(5), 1284–1295.

Huffman, G., R. Adler, M. Morrissey, D. Bolvin, S. Curtis, R. Joyce, B. McGavock, and J. Susskind (2001), Global precipitation at one-degree daily resolution from multisatellite observations, *J. Hydrometeorol.*, *2*(1), 36–50.

Huffman, G., D. Bolvin, E. Nelkin, D. Wolff, R. Adler, G. Gu, Y. Hong, K. Bowman, and E. Stocker (2007), The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, *J. Hydrometeorol.*, *8*(1), 38–55.

Johnson, C., N. K. Nichols, and B. J. Hoskins (2005), Very large inverse problems in atmosphere and ocean modelling, *Int. J. Numer. Methods Fluids*, *47*(8–9), 759–771, doi:10.1002/fld.869.

Kalnay, E. (2003), *Atmospheric Modeling, Data Assimilation, and Predictability*, 341 pp., Cambridge Univ. Press, New York.

Kim, G., and A. Barros (2002), Downscaling of remotely sensed soil moisture with a modified fractal interpolation method using contraction mapping and ancillary data, *Remote Sens. Environ.*, *83*(3), 400–413.

Krajewski, W. F., B. Vignal, B. C. Seo, and G. Villarini (2011), Statistical model of the range-dependent error in radar-rainfall estimates due to the vertical profile of reflectivity, *J. Hydrol.*, *402*, 306–316, doi:10.1016/j.jhydrol.2011.03.024.

Kumar, P. (1999), A multiple scale state-space model for characterizing subgrid scale variability of near-surface soil moisture, *IEEE Trans. Geosci. Remote Sens.*, *37*(1), 182–197, doi:10.1109/36.739153.

Kumar, P., and E. Foufoula-Georgiou (1993), A multicomponent decomposition of spatial rainfall fields. 2. Self-similarity in fluctuations, *Water Resour. Res.*, *29*(8), 2533–2544.

Kummerow, C. D., S. Ringerud, J. Crook, D. Randel, and W. Berg (2010), An observationally generated a priori database for microwave rainfall retrievals, *J. Atmos. Oceanic Technol.*, *28*(2), 113–130, doi:10.1175/2010JTECHA1468.1.

Ledoit, O., and M. Wolf (2004), A well-conditioned estimator for large-dimensional covariance matrices, *J. Multivariate Anal.*, *88*(2), 365–411.

Levy, B. C. (2008), *Principles of Signal Detection and Parameter Estimation*, 1st ed., 639 pp., Springer, New York, doi:10.1007/978-0-387–76544-0.

Lewicki, M., and T. Sejnowski (2000), Learning overcomplete representations, *Neural Comput.*, *12(2)*, 337–365.

Liang, X., E. F. Wood, and D. P. Lettenmaier (1999), Modeling ground heat flux in land surface parameterization schemes, *J. Geophys. Res.*, *104*(D8), 9581–9600.

Lorenc, A. (1988), Optimal nonlinear objective analysis, *Q. J. R. Meteorol. Soc.*, *114*(479), 205–240.

Lorenc, A. C. (1986), Analysis methods for numerical weather prediction, *Q. J. R. Meteorol. Soc.*, *112*(474), 1177–1194, doi:10.1002/qj.49711247414.

Lovejoy, S., and B. Mandelbrot (1985), Fractal properties of rain, and a fractal model, *Tellus, Ser. A*, *37*(3), 209–232.

Lovejoy, S., and D. Schertzer (1990), Multifractals, universality classes and satellite and radar, *J. Geophys. Res.*, *95*(D3), 2021–2034.

Maggioni, V., R. H. Reichle, and E. N. Anagnostou (2012), The impact of rainfall error characterization on the estimation of soil moisture fields in a land data assimilation system, *J. Hydrometeorol.*, *13*(3), 1107–1118, doi:10.1175/JHM-D-11–0115.1.

Margulis, S. A., and D. Entekhabi (2003), Variational assimilation of radiometric surface temperature and reference-level micrometeorology into a model of the atmospheric boundary layer and land surface, *Mon. Weather Rev.*, *131*(7), 1272–1288, doi:10.1175/1520-0493(2003)131<1272:VAORST>2.0.CO;2.

Margulis, S. A., D. McLaughlin, D. Entekhabi, and S. Dunne (2002), Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 Field Experiment, *Water Resour. Res.*, *38* (12), 1299, doi:10.1029/2001WR001114.

Masunaga, H., and C. Kummerow (2005), Combined radar and radiometer analysis of precipitation profiles for a parametric retrieval algorithm, *J. Atmos. Oceanic Technol.*, *22*(7), 909–929.

Merlin, O., A. Chehbouni, Y. Kerr, and D. Goodrich (2006), A downscaling method for distributing surface soil moisture within a microwave pixel: Application to the Monsoon'90 data, *Remote Sens. Environ.*, *101*(3), 379–389.

Nichols, N. K. (2010), Mathematical concepts of data assimilation, in *Data Assimilation: Making Sense of Observations, Part 1*, pp. 13–39, Springer, Berlin.

Parrish, D. F., and J. C. Derber (1992), The National Meteorological Center's spectral statistical-interpolation analysis system, *Mon. Weather Rev.*, *120*(8), 1747–1763, doi:10.1175/1520-0493(1992)120<1747: TNMCSS>2.0.CO;2.

Perica, S., and E. Foufoula-Georgiou (1996), Model for multiscale disaggregation of spatial rainfall based on coupling meteorological and scaling, *J. Geophys. Res.*, *101*(D21), 26,347–26,361.

Peter-Lidard, C. D., M. S. Zion, and E. F. Wood (1997), A soil-vegetation-atmosphere transfer scheme for modeling spatially variable water and energy balance processes, *J. Geophys. Res.*, *102*(D4), 4303–4324.

Rebora, N., L. Ferraris, J. Von Hardenberg, and A. Provenzale (2005), Stochastic downscaling of LAM predictions: An example in the Mediterranean area, *Adv. Geosci.*, *2*, 181–185.

Rebora, N., L. Ferraris, J. Von Hardenberg, and A. Provenzale (2006), Rainfall downscaling and flood forecasting: A case study in the Mediterranean area, *Nat. Hazards Earth Syst. Sci.*, *6*(4), 611–619.

Reichle, R., D. Entekhabi, and D. McLaughlin (2001a), Downscaling of radio brightness measurements for soil moisture estimation: A four-dimensional variational data assimilation approach, *Water Resour. Res.*, *37*(9), 2353–2364.

Reichle, R., D. McLaughlin, and D. Entekhabi (2001b), Variational data assimilation of microwave radio brightness observations for land surface hydrology applications, *IEEE Trans. Geosci. Remote Sens.*, *39*(8), 1708–1718, doi:10.1109/36.942549.

Sasaki, Y. (1958), An objective analysis based on variational method, *J. Meteorol. Soc. Jpn.*, *36*, 77–88.

Schultz, R., and R. Stevenson (1994), A Bayesian approach to image expansion for improved definition, *IEEE Trans. Image Process.*, *3*(3), 233–242.

Siccardi, F., G. Boni, L. Ferraris, and R. Rudari (2005), A hydrometeorological approach for probabilistic flood forecast, *J. Geophys. Res.*, *110*, D05101, doi:10.1029/2004JD005314.

Sorooshian, S., K. Hsu, G. Xiaogang, H. Gupta, B. Imam, and D. Braithwaite (2000), Evaluation of PERSIANN system satellite-based estimates of tropical rainfall, *Bull. Am. Meteorol. Soc.*, *81*(9), 2035–2046.

Talagrand, O., and P. Courtier (1987), Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory, *Q. J. R. Meteorol. Soc.*, *113*(478), 1311–1328.

Tibshirani, R. (1996), Regression shrinkage and selection via the lasso, *J. R. Stat. Soc. Ser. B*, *58*(1), 267–288.

Tikhonov, A., V. Arsenin, and F. John (1977), *Solutions of Ill-Posed Problems*, Winston. Washington, D. C.

Tustison, B., E. Foufoula-Georgiou, and D. Harris (2003), Scale-recursive estimation for multisensor Quantitative Precipitation Forecast verification: A preliminary assessment, *J. Geophys. Res*, *107*(D8), 8377, doi:10.1029/2001JD001073.

Van de Vyver, H., and E. Roulin (2009), Scale-recursive estimation for merging precipitation data from radar and microwave cross-track scanners, *J. Geophys. Res.*, *114*, D08104, doi:10.1029/2008JD010709.

Veneziano, D., R. L. Bras, and J. D. Niemann (1996), Nonlinearity and self-similarity of rainfall in time and a stochastic model, *J. Geophys. Res.*, *101*(D21), 26,371–26,392.

Wang, S., X. Liang, and Z. Nan (2011), How much improvement can precipitation data fusion achieve with a multiscale Kalman Smoother-based framework?, *Water Resour. Res.*, *47*, W00H12, doi:10.1029/2010WR009953.

Wang, Z., A. Bovik, H. Sheikh, and E. Simoncelli (2004), Image quality assessment: From error visibility to structural similarity, *IEEE Trans. Image Process.*, *13*(4), 600–612.

Wilby, R., T. Wigley, D. Conway, P. Jones, B. J. M. Hewitson, and D. S. Wilks (1998a), Statistical downscaling of general circulation model output: A comparison of methods, *Water Resour. Res.*, *34*(11), 2995–3008, doi:10.1029/98WR02577.

Wilby, R., H. Hassan, and K. Hanaki (1998b), Statistical downscaling of hydrometeorological variables using general circulation model output, *J. Hydrol.*, *205*(1), 1–19.

Zupanski, D., S. Q. Zhang, M. Zupanski, A. Y. Hou, and S. H. Cheung (2010), A prototype WRF-based ensemble data assimilation system for dynamically downscaling satellite precipitation observations, *J. Hydrometeorol.*, *12*(1), 118–134, doi:10.1175/2010JHM1271.1.

Zupanski, M. (1993), Regional four-dimensional variational data assimilation in a quasi-operational forecasting environment, *Mon. Weather Rev.*, *121*(8), 2396–2408.