Sparse regularization for precipitation downscaling

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

Received 26 October 2011; revised 15 February 2012; accepted 13 March 2012; published 21 April 2012.

[1] Downscaling of remotely sensed precipitation images and outputs of general circulation models has been a subject of intense interest in hydrometeorology. The problem of downscaling is basically one of resolution enhancement, that is, appropriately adding details or high frequency features onto a low-resolution observation or simulated rainfall field. Invoking the property of rainfall self similarity, this mathematically ill-posed problem has been approached in the past within a stochastic framework resulting in ensemble of possible high-resolution realizations. In this work, we recast the rainfall downscaling into an ill-posed inverse problem and introduce a class of nonlinear estimators to properly regularize it and obtain the best high-resolution estimate in an optimal sense. This regularization capitalizes on two main observations: (1) precipitation fields are sparse when transformed into an appropriately chosen domain (e.g., wavelet), and (2) small-scale organized precipitation features tend to recur within and across different storm environments. We demonstrate the promise of the proposed methodology through downscaling and error analysis of level III precipitation reflectivity snapshots provided by the ground-based next generation Doppler weather radars in a ground validation sites of the Tropical Rainfall Measuring Mission.

Citation: Ebtehaj, A. M., E. Foufoula-Georgiou, and G. Lerman (2012), Sparse regularization for precipitation downscaling, *J. Geophys. Res.*, *117*, D08107, doi:10.1029/2011JD017057.

1. Introduction

[2] The coarseness of the native spatiotemporal scales of some remotely sensed atmospheric state variables and also General Circulation Models (GCM) of the climate system is often restrictive for many terrestrial applications [e.g., *Wilby and Wigley*, 1997]. *Hostetler* [1994] reported that the largest prediction and parameter estimation error of both GCMs and hydrologic models occurs at the interfacial scales at which terrestrial and climate models need to be linked.

[3] Detailed hydrologic modeling and parameter estimation often involve sub-watershed hillslope scale transport processes. Several studies have highlighted the effect of small-scale rainfall variability on watershed hydrologic response [e.g., *Woods and Sivapalan*, 1999; *Smith et al.*, 2004; *Rebora et al.*, 2006a; *Schuurmans and Bierkens*, 2007; *Vivoni et al.*, 2007; *Younger et al.*, 2009]. These studies typically report marked changes on the shape and peak of the hydrograph in small-scale hydrologic units (basins <1000 km^2), particularly during fast evolving extreme events. For instance, using a distributed hydrologic model in a 135 km^2 basin, it was shown by *Younger et al.* [2009] that under different perturbation scenarios of the rainfall patterns, not only the hydrograph but also the estimated parameters of the hydrologic model may drastically change. As a primary treatment, computationally efficient mathematical downscaling models have received considerable attention in hydrologic applications. For instance, *Rebora et al.* [2006b] coupled a downscaling model with a semi-distributed hydrologic model and quantitatively verified the significance of rainfall downscaling for flood ensemble forecasting of small watersheds and urban areas in the order of hundreds of square kilometers. A similar analysis was performed by *Nykanen et al.* [2001] in which a precipitation downscaling model was dynamically coupled with a coarse-scale numerical weather prediction model to show improvement on the computation of land-atmosphere fluxes at a watershed scale.

[4] During the past decades, special attention has been devoted to developing different classes of stochastic models, $\mathcal{M}(\mathcal{Y},\theta,\zeta)$, to reproduce consistently the missing high-frequency part of the rainfall fields, from a low-resolution input ($\mathcal{Y}$) using a set of parameters (θ) and a random generator (ζ) [e.g., *Lovejoy*, 1981; *Lovejoy and Mandelbrot*, 1985; *Gupta and Waymire*, 1993; *Perica and Foufoula-Georgiou*, 1996; *Menabde et al.*, 1997; *Badas et al.*, 2006; *Deidda*, 2000; *Rebora et al.*, 2006a]. To parameterize the random generator, a large number of these models often relies on the observed statistical regularities of the precipitation fields such as power law Fourier spectrum and scaling of the higher-order statistical moments of rainfall fluctuations, often referred to as multifractality. Although, these stochastic generators can be designed to be statistically consistent with the observed small-scale rainfall variability, the spatial coherency of the generated high-resolution fields is often not satisfying.

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

0148-0227/12/2011JD017057

Blockiness and non-uniqueness of the downscaled fields are typical drawbacks of this sort of stochastic generators, enforcing their interpretation in an ensemble mode; see *Ferraris et al.* [2003] for a review of available statistical/mathematical downscaling models.

[5] Apart from hydrometeorological applications, enhancing the resolution and quality of low-resolution images, often referred to as super-resolution in the image processing community, has been a subject of interest for many years. The state-of-the-art approaches in this area are currently relying on sparse approximation [e.g., *Mallat and Zhang*, 1993; *Chen et al.*, 1999] and statistical learning [e.g., *Donoho and Stark*, 1989; *Tibshirani*, 1996], which interestingly have some roots in subsurface geophysics [*Claerbout and Muir*, 1973]. The central results suggest that a high-resolution signal can be well recovered from a low-resolution one via solving a constrained optimization, provided that the signal of interest exhibits a sufficiently sparse representation (a large fraction of close to zero coefficients) in an appropriately chosen domain (e.g., wavelet) [e.g., *Yang et al.*, 2010; *Mallat and Yu*, 2010; *Elad*, 2010; *Zeyde et al.*, 2010].

[6] In this paper we suggest a new framework for downscaling of rainfall images motivated by the sparsity of precipitation images in the wavelet domain as well as the recurrence of small-scale organized precipitation features within and across different storm environments. The proposed framework follows recent progress in sparse approximation by recasting the rainfall downscaling problem into solving an inverse ill-posed problem to obtain the best high-resolution estimate via an appropriately selected optimality criterion. The main advantages of the proposed method compared to the traditional stochastic downscaling models can be summarized as follows: (1) given the low-resolution rainfall image, the downscaled field is unique with reduced estimation error; (2) the method is robust to measurement noise; and (3) the solution is smooth enough and free of the blockiness commonly observed in stochastic models.

[7] Section 2 is devoted to explain notation and terminology. In section 3, a sparse inverse estimator and its connection with other estimation paradigms is explained. Sparsity of the spatial rainfall images is elaborated in the wavelet domain in section 4. We provide quantitative evidence in section 5, that the organized small-scale features of a storm snapshot may recur within the larger storm domain or also across different storm environments. This indicates that the small-scale rainfall features which might have been lost in the low-resolution sensing and/or modeling, might be able to be reconstructed using a database of appropriately collected coincidental high and low-resolution precipitation images. In section 6, the implementation details of the sparse inverse estimator for rainfall downscaling are explained. Section 7, points out concluding remarks and future research directions.

2. Notation

[8] We adopt the following notation and nomenclature throughout. For any vector $\mathbf{x}\in\mathbb{R}^m, x_i$ refers to the ith element and the standard l_p-norm of $\mathbf{x}$ is denoted by $\|\mathbf{x}\|_p = \left(\sum_i^m |x_i|^p\right)^{1/p}$, where $p \geq 1$. The zero norm of a vector $\|\mathbf{x}\|_0$, denotes the number of non-zero elements of $\mathbf{x}$, while the infinity norm is $\|\mathbf{x}\|_\infty = \max_i|x_i|$. A redundant dictionary $\mathbf{\Phi} = [\phi_1, \phi_2, \ldots, \phi_M]$ in $\mathbb{R}^{m\times M}$ refers to a typically "fat matrix" with more columns than rows ($M \geq m$), where a linear combination of the column vectors (called "atoms"), $\phi_i\in\mathbb{R}^m$, can well approximate a class of signals of interest $\mathbf{x}\cong\sum_i^M \phi_i c_i = \mathbf{\Phi c}$, where $\mathbf{c}$ contains the "representation" coefficients. When the atoms are obtained through samples of a particular process, we refer to $\mathbf{\Phi}$ as an empirical dictionary. We say that $\mathbf{x}\in\mathbb{R}^m$ has a "sparse representation" on $\mathbf{\Phi}$, if $\|\mathbf{c}\|_0 \ll m$. By convention, we also may refer to an image $\mathcal{X}$ as a vector $\mathbf{x}$ by staking all the pixel values in a predefined fixed order.

3. Sparse Inverse Estimator

[9] Estimation of a high-resolution signal or image $\mathbf{x} = [x_1, \ldots, x_m]^T\in\mathbb{R}^m$, from its low-resolution counterpart $\mathbf{y}\in\mathbb{R}^n$, where $n \leq m$, can be recast as an inverse problem. It amounts to estimating $\mathbf{x}$ from $\mathbf{y}$, while $\mathbf{y}$ may relate to $\mathbf{x}$ through a linear structured degradation operator $\mathbf{H}\in\mathbb{R}^{n\times m}$ (e.g., blurring and downsampling) and an intrinsic additive noise $\mathbf{e}$:

$$\mathbf{y} = \mathbf{Hx} + \mathbf{e}. \tag{1}$$

The degradation operator $\mathbf{H}$ is obviously a rectangular matrix with more columns than rows giving rise to a linear underdetermined system of equations with many solutions, making this an ill-posed inverse problem. To narrow down the solutions to a well-defined one, this problem needs to be *regularized* by taking into account additional constraints and a priori assumptions about the desired signal of interest $\mathbf{x}$.

[10] As explained earlier, sparsity of $\mathbf{x}$ implies that it can be well approximated by its orthogonal projection $\mathbf{x}_S$ onto a subspace generated by a few atoms $\{\phi_i\}_{i=1}^M$ of a suitable redundant dictionary $\mathbf{\Phi}\in\mathbb{R}^{m\times M}$

$$\mathbf{x}_S = \mathbf{\Phi c},$$

where $m \leq M$ and the vector of representation coefficients $\mathbf{c}$ is fairly sparse with the number of non-zero elements much smaller than the signal length; i.e., $\|\mathbf{c}\|_0 \ll m$.

[11] Substitution of $\mathbf{x}_S$ into equation (1) results in

$$\mathbf{y} = \mathbf{H\Phi c} + \mathbf{e}', \tag{2}$$

where $\mathbf{e}' = \mathbf{H}(\mathbf{x} - \mathbf{x}_S) + \mathbf{e}$. This implies that the low-resolution observation $\mathbf{y}$ has similar representation in a transformed dictionary $\mathbf{\Psi} = \mathbf{H\Phi}\in\mathbb{R}^{n\times M}$ which is naturally more redundant than $\mathbf{\Phi}$ ($n \leq m$). Thus, by finding an appropriate sparse representation $\hat{\mathbf{c}}$ of $\mathbf{y}$ in the so-called *low-resolution dictionary* $\mathbf{\Psi}$, the same coefficients $\hat{\mathbf{c}}$ can be used to reconstruct the desired high-resolution signal $\mathbf{x}$ using the corresponding atoms of the *high-resolution dictionary* $\mathbf{\Phi}$ that is, $\hat{\mathbf{x}} = \mathbf{\Phi}\hat{\mathbf{c}}$. Given only equation (2), due to the redundancy of $\mathbf{\Psi}$ and possible observation noise, the representation is not obviously unique. The a priori assumption on sparsity of the signal, naturally implies that among many solutions, those with the minimum number of non-zero elements are the optimal ones. Following that the estimated representation coefficients shall obey the fidelity constraint

imposed by equation (2), the optimal coefficients can be obtained by solving

$$\mathbf{c} = \arg\min_{c} \|\mathbf{c}\|_0 \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{\Psi c}\|_2 \le \epsilon', \tag{3}$$

where $\epsilon' = \|\mathbf{e}'\|_2$. However, this gives rise to a Non-deterministic Polynomial-time hard (NP-hard) problem for which a direct solution is still hopeless.

[12] Greedy methods such as Orthogonal Matching Pursuit (OMP) by *Mallat and Zhang* [1993] have been extensively used to heuristically tackle problems of this sort. In this numerical method, the solution is obtained via iterative selection of the relevant dictionary atoms. Initializing the estimation residual by the observed signal, at each iteration, first the support of the representation coefficients $\mathbf{c}$ is updated by selecting an atom which has the maximum inner product with the estimation residual and then, given the support set, the values of the representation coefficients are being updated through an ordinary least squares. The iterations continue until a certain number of atoms is selected or the magnitude of the estimation residual falls below ϵ', when the observation error is accurately known.

[13] *Chen et al.* [1999] proposed a new strategy known as the Basis Pursuit (BP) which suggests that under some sparsity constraints on the solution, in lieu of $\|\mathbf{c}\|_0$ the closest convex penalization function $\|\mathbf{c}\|_1$ leads to the same sparsest solution [see also *Chen et al.*, 1999; *Elad*, 2010]. Using a Lagrangian penalization, the l_1-regularized version of equation (3) follows

$$\hat{\mathbf{c}} = \arg\min_{c} \lambda\|\mathbf{c}\|_1 + \frac{1}{2}\|\mathbf{y} - \mathbf{\Psi c}\|_2^2, \tag{4}$$

which can be recast into a linear programming problem and solved at large dimensions with modest computational cost. Note that in this Lagrangian form, we need to choose the non-negative regularization parameter $\lambda > 0$. The magnitude of λ reflects somehow the power of the noise in the observation and according to a rule of thumb, we can choose λ as approximately the ratio of the noise standard deviation to the standard deviation of the expected non-zero elements of the solution [see *Elad*, 2010]. A small λ reduces the role of the l_1 penalty term and hence leads to a more dense solution, while a larger one further sparsifies the solution up to a theoretical upper bound $\|\mathbf{\Psi}^T\mathbf{y}\|_\infty$, beyond which all elements of the solution tend to zero.

[14] It is worth noting that, in the statistical machine learning community, the problem of solving equation (4) is the Lagrangian form of the Least Absolute Shrinkage and Selection Operator (LASSO) by *Tibshirani* [1996]. From the Bayesian statistical standpoint, the minimization in equation (4) is also equivalent to the Maximum A Posteriori (MAP) estimation of the coefficients, $\hat{\mathbf{c}}_{MAP} = \arg\max_{\mathbf{c}}\{p(\mathbf{y}|\mathbf{c})p(\mathbf{c})\}$, given that the additive noise $\mathbf{e}$ is Gaussian and the a priori distribution of the coefficients can be well explained by the family of Laplace densities, i.e. $p(c_i) \propto \exp(-|c_i|)$ [see *Lewicki and Sejnowski*, 2000]. In the context of the MAP estimator, the terms $\|\mathbf{c}\|_1$ and $\|\mathbf{y} - \mathbf{\Psi c}\|_2^2$ in equation (4) represent the contribution of the logarithm of the prior density $p(\mathbf{c})$ and the log-likelihood function $\log p(\mathbf{y}|\mathbf{c})$, respectively. Note that, the logarithm of the Laplace and Gaussian density is a monotonic and concave function, and by a sign change, the minimization in equation (4) and the MAP estimator are equivalent.

[15] Effective implementation of the described sparse recovery method requires that the dictionary pairs for the high ($\mathbf{\Phi}$) and low-resolution ($\mathbf{\Psi}$) atoms have been appropriately determined which will be briefly explained in the context of our study in the following sections.

4. On Sparsity of Rainfall Images

[16] Study of precipitation signals and images in transform domains (e.g., Fourier or wavelet) has received considerable attention and revealed interesting multiscale features of precipitation structures. Numerous studies on the Fourier spectrum of rainfall images have provided evidence of power law scaling and spreading of the energy over a relatively wide range of frequencies. Conversely, it is found that the probability distribution of the wavelet coefficients of precipitation reflectivity images exhibits a large mass at zero and extended tails significantly thicker than those of the Gaussian distribution, implying that a large number of these coefficients are very close to zero and only a small portion of them contains significant amount of the rainfall energy [e.g., *Perica and Foufoula-Georgiou*, 1996; *Ebtehaj and Foufoula-Georgiou*, 2011a, 2011b].

[17] In this study we use the undecimated Haar wavelet transform by *Nason and Silverman* [1995], with a single layer of decomposition, to demonstrate sparsity of the wavelet coefficients of precipitation images. The transform coefficients at different subbands (directions) are obtained by filtering the image with the one dimensional low-pass [+1, +1]/2 and high-pass [+1, −1]/2 Haar kernels and their transpose successively. It is demonstrated in Figure 1a that the majority of the wavelet coefficients in the Horizontal, Vertical and Diagonal subbands are close to zero. As is evident, the histogram of the wavelet coefficients (see Figure 1b for the horizontal subband) also shows concentration of the coefficients around the origin. Keeping only the top largest 20% of the wavelet coefficients in absolute value and setting the rest to zero via a hard thresholding [see *Donoho*, 1995], Figure 1c shows the reconstructed rainfall image. Using only about 20% of the entire wavelet coefficients, the reconstructed rainfall image contains 99.8% of the total energy of the original image (sum of pixel-wise squares) and is visually indistinguishable from the original one. This evidence confirms that the rainfall reflectivity images exhibit considerable sparseness in the wavelet domain which is a direct consequence of the piecewise smoothness and spatial coherency of the precipitation images.

5. Recurrence of Rainfall Patches and Group Sparsity

[18] The observed precipitation patterns are the result of a wide spectrum of highly non-linear and complex atmospheric processes operating over different range of scales, from microphysical to regional and global. As a result, it is hard to imagine that we might be able to find very similar (in mathematical norms) precipitation patterns over relatively large storm-scales. However, finding similar features over sufficiently small sub-storm scales seems more feasible. The goal of this section is to provide evidence for the recurrence of

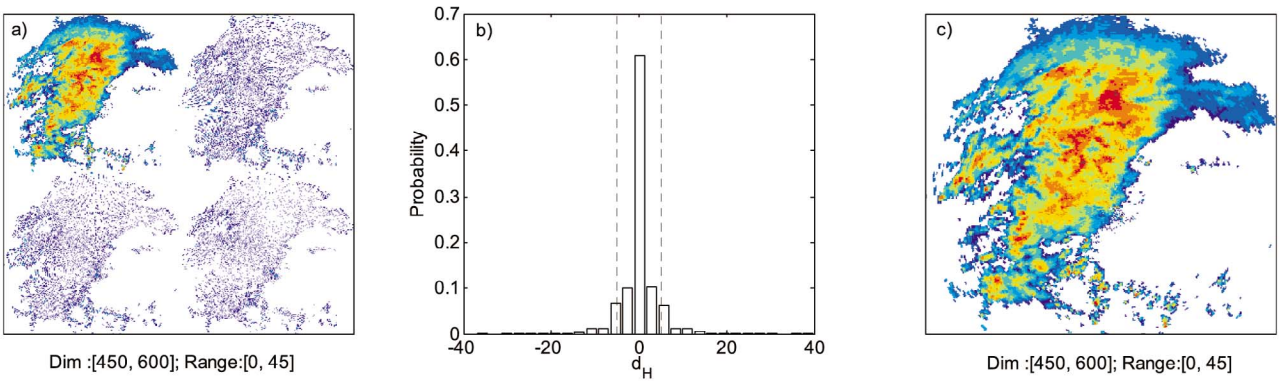

Figure 1. Evidence on sparsity of precipitation images in the wavelet domain: (a) A storm reflectivity snapshot at the TRMM GV-site in Houston, TX (HSTN) on 1998/11/13 (00:02:00 UTC) and the absolute values of the wavelet coefficients in the horizontal (top right), vertical (bottom left) and diagonal (bottom right) subbands, (b) probability histogram of the horizontal wavelet coefficients (d_H) and (c) the reconstructed field using the top largest 20% of the wavelet coefficients in absolute values. The bounded area by the dashed lines in Figure 1b contains 80% of the wavelet coefficients that was set to zero for reconstruction of the reflectivity image, shown in Figure 1c.

small-scale "patches" of precipitation features (patches being small neighborhoods of the sampled rainfall fields) within different regions of the same storm or across different storm environments. The motivation is that, if for each patch of an arbitrary storm snapshot, we can find at least a few similar patches in a representative precipitation database, this will promise that the small-scale precipitation variability in that storm may be well approximated by an optimal and possibly linear combination of those similar patches.

[19] To this end, a sample storm reflectivity image is selected (see Figure 2a) and N "sample patches" of size 5×5, $\{\mathbf{y}_i \in \mathbb{R}^{25}\}_{i=1}^{N}$, are extracted over regions of significant magnitudes of rainfall gradients (see Figure 2b). The precipitation gradient was selected here only as a metric to guide the sampling of important patches as, naturally, localized regions of high rainfall gradients contain significant high-frequency information and define the within-storm spatial geometry and variability. Then, the high-pass fluctuations of all rainfall patches, the so-called "rainfall features" $\{\mathbf{y}_i'\}_{i=1}^{N}$, are obtained by subtracting the local mean of each patch, that is $\mathbf{y}_i' = \mathbf{y}_i - \overline{\mathbf{y}}_i$. The same sampling procedure is performed to extract same size *test patches* $\{\mathbf{z}_j'\}_{j=1}^{M}$ of rainfall features, where naturally $M \gg N$, from a database containing an independent set of about 100 storm NEXRAD (level III) reflectivity snapshots over the TRMM satellite Ground Validation (GV) site in Huston, TX (see *Ebtehaj and Foufoula-Georgiou* [2011a] for a detailed description of the data set). For each ith sample patch, we computed the Euclidean distance $\mathbf{d}_{ij} = \|\mathbf{y}_i' - \mathbf{z}_j'\|_2$ with all of the test patches and appropriately normalized it to the range of [0, 1]. Here, the similarity measure among the patches is defined as the normalized Euclidean distance below a certain threshold value η. Choosing the threshold values η, linearly spaced from 0.01 to 0.05 in five intervals, we determined the number of sample patches n_Λ in the sample rainfall image, for which we could find at least an m_Λ number of similar test patches in our test database. We then reported m_Λ versus the ratio n_Λ/N, for different choices of η in Figure 2c. As the ratio n_Λ/N is always positive and ranges between [0, 1], it can be interpreted as a lower bound for the probability of finding at least an m_Λ number of similar patches within any larger database that contains our test database as a subset. Figure 2c suggests that the probability of finding a small number (small m_Λ) of very similar (small η) patches is high and drops rapidly as the number of similar patches is increased. This finding indicates that the rainfall fluctuations are not only pixel-wise sparse, but also may exhibit structured sparsity in a small neighborhood of rainfall intensities, which we refer to as *group sparsity*. As is evident, specializing the database for particular environmental conditions (e.g., orographic precipitation), may increase the chance of finding similar features and improve the group sparsity.

6. Precipitation Sparse Downscaling

6.1. Conceptual Framework

[20] The proposed sparse restoration of a high-resolution rainfall image $\mathcal{X}$ from a coarse-scale observation $\mathcal{Y}$, requires the availability of a representative set of N_T coincidental pairs of low and high-resolution rainfall images $\{(\mathcal{Z}_l, \mathcal{Z}_h)\}_{l,h=1}^{N_T}$, as a training set. Note that throughout this study, we have adopted a simplified notation by which the resolution indices (l) and (h), also serve as counting indices for the number of the studied images. Ideally, the low-resolution part of the training set can be extracted from a space-borne sensor or a GCM-output; while the high-resolution counterparts can be the result of a coincidental ground based high-resolution sensing device or a fine-scale predictive model (e.g., NEXRAD or a fine grid cloud resolving numerical model). Elaborating on the merits of the methodology, in this study we synthetically generated the low-resolution rainfall reflectivity images by smoothing and downsampling the available NEXRAD data set. For instance, a low-resolution data set $\{\mathcal{Z}_l\}_{l=1}^{N_T}$ at scale 8×8 km is obtained by first smoothing each high-resolution NEXRAD image (1×1 km) using an average spatial filter of size 8×8 and then downsampling by an integer factor of $\sqrt{s} = 8$,

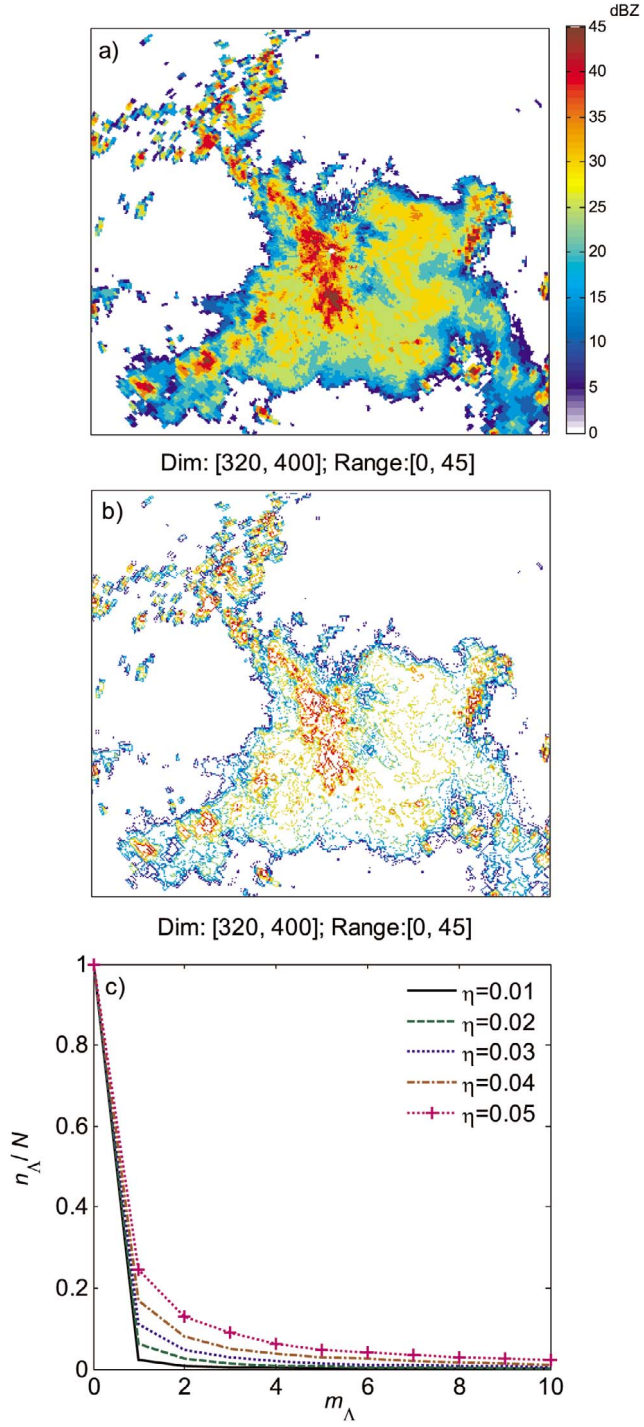

Figure 2. Recurrence of small-scale precipitation patches and group sparsity: (a) a sample precipitation reflectivity image over the TRMM (HSTN) GV-site on 1998/06/28 (18:13:00 UTC), (b) regions of high gradient (top 25%) used for sampling of important patches, and (c) a probability measure (n_Λ/N) of finding at least m_Λ number of similar patches (i.e., close in Euclidean distance determined by η) in the selected test database.

called the scaling ratio. Note that, given the pair of high and low-resolution training images, a set of "residual images" can be produced as

$$\mathcal{R} = \mathcal{Z}_h - \mathbf{Q}\mathcal{Z}_l, \tag{5}$$

with each image containing the high-frequency features lost by the low-resolution sensing. Here, $\mathbf{Q}$ denotes an interpolation operator (e.g., nearest neighborhood, bilinear, bicubic, Kriging) which is used to compute a low-pass estimate of the high-resolution precipitation image of interest.

[21] The core part of the proposed downscaling methodology is the inverse estimator explained in section 3 and relies on the group sparsity of precipitation patches. As explained earlier, the group sparsity is revealed for the high-pass components (features) of rainfall patches. To this end, the high-pass components of the low-resolution observation and of the training database need to be extracted, i.e., $\{\mathcal{Y}', \{\mathcal{Z}_l'\}_{l=1}^{N_T}\}$. Typically redundant transformations are preferred due to their robustness to noise, more sparsifying effect and greater flexibility to match the data structure. Here, similar to the previously explained undecimated Haar wavelet transform, vertical and horizontal features of precipitation images have been extracted using the following kernels,

$$\begin{aligned} f_1 &= f_2^T = [+1, -1] \\ f_3 &= f_4^T = [+1, -2, +1]. \end{aligned} \tag{6}$$

Concerning the precipitation images, the first two kernels (first order derivative) mainly characterize directional edges of the storm from zero intensity background and boundaries of high-intensity rain-cells while the others (second order derivative) encode high curvature regions within the body of the rain-cells, which are prone to be lost in a low-resolution sensing or large-scale modeling. These filters naturally give rise to four output high-pass images of the same size as the input image, leading to a 4:1 redundancy factor; i.e., $\left\{\mathcal{Y}_i', \left\{\mathcal{Z}_{l,i}'\right\}_{l=1}^{N_T}\right\}_{i=1}^{4}$, where hereafter the counting index i is dropped in our notation for brevity.

[22] In the proposed precipitation downscaling methodology, due to the evidence of group sparsity, the sparse inverse estimator is applied locally on rainfall patches rather than globally on the whole image. This approach not only makes the problem computationally more efficient but also permits an overlapping estimation scheme to avoid undesirable blocking artifacts in the final downscaled product.

[23] Deciding to choose the low-resolution rainfall patches of size $\sqrt{n} \times \sqrt{n}$ making thus $\mathbf{y} \in \mathbb{R}^n$, the sample and training feature patches can be extracted from the feature images $\{\mathcal{Y}', \{\mathcal{Z}_l'\}_{l=1}^{N_T}\}$ and stacked in $\mathbf{y}'$ and $\mathbf{z}_l' \in \mathbb{R}^m$, where owing to the suggested filters in equation (6), here $m = 4n$. The column-wise concatenation of $\mathbf{z}_l'$ in $\mathbf{\Psi} \in \mathbb{R}^{m \times M}$ forms the desired empirical low-resolution rainfall dictionary. Now, in the feature space, the idea is to find a few columns of the low-resolution dictionary $\mathbf{\Psi}$ that their linear combinations can well approximate the low-resolution feature patch $\mathbf{y}'$ of interest. This idea can be recast into the following sparse

encoding problem, similar to equation (3), which indeed assumes $\mathbf{y}' = \mathbf{\Psi c} + \mathbf{e}$ with $\epsilon = \|\mathbf{e}\|_2$,

$$\hat{\mathbf{c}} = \arg\min_c \|\mathbf{c}\|_0 \quad \text{subject to} \quad \|\mathbf{y}' - \mathbf{\Psi c}\|_2 \le \epsilon. \tag{7}$$

As explained in section 2, the estimated coefficients for each patch can now be used to combine the corresponding columns of a high-resolution dictionary to recover the relevant high-frequency features of interest for that patch. To this end, due to the size enlargement of the high-resolution field by a factor of s, the patches of size $\sqrt{sn} \times \sqrt{sn}$ need to be extracted from the residual images $\{\mathcal{R}_l\}_{l=1}^{N_T}$, obtained from the training data set, and be concatenated column-wise, in a vector form $\mathbf{r} \in \mathbb{R}^{sn}$, to shape the desired empirical high-resolution dictionary, i.e., $\mathbf{\Phi} \in \mathbb{R}^{sn \times M}$. Recall that the columns of these empirical dictionaries ($\mathbf{\Psi}$, $\mathbf{\Phi}$) contain the patches of low-resolution rainfall features $\mathbf{z}'_l \in \mathbb{R}^m$ and the corresponding high-resolution rainfall residuals $\mathbf{r} \in \mathbb{R}^{sn}$, respectively. As a result, the obtained sparse representation in equation (7) is then being used to combine the columns of $\mathbf{\Phi}$ and restore the high-resolution rainfall patches as $\hat{\mathbf{x}} = \mathbf{Qy} + \hat{\mathbf{r}}$, where $\hat{\mathbf{r}} = \Phi\hat{\mathbf{c}}$. Applying this inverse estimator for all overlapping patches of the given low-resolution rainfall image, the entire high-resolution rainfall image can then be recovered. To reduce blocking artifacts and impose more coherency on the downscaled field, the estimated patches are averaged over their overlapping areas. The entire methodology, referred to as SPaD (Sparse Precipitation Downscaling), is also sketched algorithmically as follows:

Algorithm 1: Sparse Precipitation Downloading (SPaD)

Inputs: Low-resolution rainfall $\mathcal{Y}$; patch size n; the scaling ratio s; a training set $\{(\mathcal{Z}_l, \mathcal{Z}_h)\}_{l,h=1}^{N_T}$.
Steps:

1. Apply the filters in equation (6) to $\mathcal{Y}$ and $\{\mathcal{Z}_l\}_{l=1}^{N_T}$ to obtain the feature images $\{\mathcal{Y}', \{\mathcal{Z}'_l\}_{l=1}^{N_T}\}$ (i.e., 4 output images per 1 input image).

2. Extract all $\sqrt{n} \times \sqrt{n}$ overlapping patches from $\{\mathcal{Z}'_l\}_{l=1}^{N_T}$ (per the 4 filters) to shape the low-resolution dictionary $\mathbf{\Psi} \in \mathbb{R}^{m \times M}$.

3. Compute $\{\mathcal{R}_l\}_{l=1}^{N_T}$ in equation (5) and extract all $\sqrt{sn} \times \sqrt{sn}$ patches to form the high-resolution rainfall dictionary $\mathbf{\Phi} \in \mathbb{R}^{sn \times M}$.

4. **For** each feature patch $\mathbf{y}' \in \mathbb{R}^m$

   - Solve $\hat{\mathbf{c}} = \arg\min_{\mathbf{c}} \|\mathbf{c}\|_0$ s.t. $\|\mathbf{y}' - \mathbf{\Psi c}\|_2 \le \epsilon$
   - Compute the residuals $\hat{\mathbf{r}} = \mathbf{\Phi}\hat{\mathbf{c}}$
   - Compute the high-resolution rainfall patch $\hat{\mathbf{x}} = \mathbf{Qy} + \hat{\mathbf{r}}$.

   End

Output: Recover the high-resolution rainfall image $\hat{\mathcal{X}}$ by putting together all patches $\hat{\mathbf{x}}$, while averaging on the overlapping areas.

6.2. SPaD via Dictionary Learning

[24] Collecting a large set of training patches of low-resolution precipitation features $\{\mathbf{z}'_l\}_{l=1}^{M}$ as the atoms of $\mathbf{\Psi}$, often leads to a very large-scale sparse coding problem, which cannot be easily handled by the Basis Pursuit type of methods in a reasonable time and with sufficient accuracy. Often, when the number of dictionary atoms is much larger than the patch size, the maximum absolute value of the cross correlation between the atoms of $\mathbf{\Psi}$, called the Mutual coherence, tends to unity and a stable sparse solution is hardly achievable via l_1-norm convex relaxation of equation (7) in the form of equation (4), [see e.g., *Kim et al.*, 2007; *Elad*, 2010]. As an alternative to a crude selection of the sample precipitation features as atoms of the empirical high and low-resolution dictionary, one can use a tunable and compact selection, in which the atoms are well adapted to the patches of rainfall images based on a learning procedure. The core idea is to optimally estimate a pair of reduced dimension low and high-resolution dictionary $(\mathbf{\Psi}_r, \mathbf{\Phi}_r) \in \mathbb{R}^{m \times K}$, where $K \ll M$, given a large number of low and high-resolution training patches. Typically, the learned reduced size low-resolution dictionary $\mathbf{\Psi}_r$ is obtained such that it can reproduce each training patch $\mathbf{z}'_l$ by a sparse combination of its atoms in a linear setting $\mathbf{z}'_l = \mathbf{\Psi}_r\mathbf{c}_l + \mathbf{e}$ with added noise $\mathbf{e}$, where $\mathbf{c}_l \in \mathbb{R}^K, \|\mathbf{c}_l\|_0 \ll m$. Note that in this setting the goal is to have an optimal estimate of the $\mathbf{\Phi}_r$ and the representation coefficients $\mathbf{c}_l$, simultaneously. It naturally follows to solve

$$\min_{\mathbf{\Psi}_r, \mathbf{c}_l} \sum_{l=1}^{M} \|\mathbf{z}'_l - \mathbf{\Psi}_r\mathbf{c}_l\|_2^2 \qquad \text{subject to} \quad \|\mathbf{c}_l\|_0 \le \text{const.} \ \forall l, \tag{8}$$

which is not well-posed and convex on both $\mathbf{\Psi}_r$ and $\mathbf{c}_l$ [*Engan et al.*, 1999; *Aharon et al.*, 2006; *Engan et al.*, 2007]. Using the Method of Optimal Direction (MOD), this problem can be tackled by following a nested iterative optimization strategy [*Engan et al.*, 1999, 2007]. This method, starts with a first guess of the dictionary $\mathbf{\Psi}_r^{(0)}$ and proceeds to estimate the representation $\mathbf{c}_l^{(0)}$ for $\forall l$, using a sparse encoding strategy (e.g., OMP). Then, at the ith iteration, given the obtained representation coefficients, the dictionary is being updated through an ordinary least squares

$$\mathbf{\Psi}_r^{(i+1)} = \mathbf{\Psi}\left(\mathbf{C}^{(i)}\right)^+,$$

where $\mathbf{C}^{(i)} \in \mathbb{R}^{K \times M}$ contains a column-wise concatenation of all estimated coefficients $\left\{\mathbf{c}_l^{(i)}\right\}_{l=1}^{M}$ at step ith; $\left(\mathbf{C}^{(i)}\right)^+ = \left(\mathbf{C}^{(i)}\right)^T\left[\mathbf{C}^{(i)}\left(\mathbf{C}^{(i)}\right)^T\right]^{-1}$ is the right Moore-Penrose pseudoinverse of $\mathbf{C}^{(i)}$ and $\mathbf{\Psi}$ is the original empirical low-resolution dictionary. This iterative procedure continues until a stopping criterion is met (e.g., number of iterations or an error criterion).

[25] Here, the dictionary $\mathbf{\Psi}_r$ is a reduced version of $\mathbf{\Psi}$ which can sparsely approximate the low-resolution features of all M training patches. By construction, using the same representation for high-resolution recovery, estimation of the reduced high-resolution dictionary $\mathbf{\Phi}_r$ only requires the

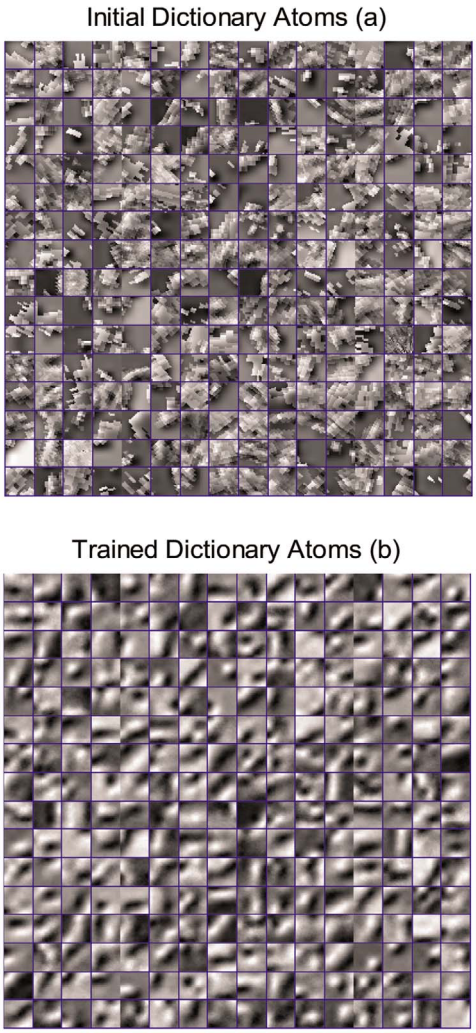

Figure 3. (a) Initial rainfall residual patches $\mathbf{r} \in \mathbb{R}^{sn}$ (i.e., patch size 3×3 with downscaling ratio $\sqrt{s} = 4$) at scale 1×1 km, randomly sampled from 150000 training rainfall patches to learn a high-resolution dictionary $\mathbf{\Phi}_r$ of rainfall features with $K = 256$ atoms for downscaling from resolution 4-to-1 km in grid spacing, and (b) the learned precipitation dictionary after 50 iterations.

second step of a typical dictionary learning process [see *Zeyde et al.*, 2010] and follows:

$$\mathbf{\Phi}_r = \mathbf{\Phi}\mathbf{C}^{+}.$$

In fact, given the final matrix of representation coefficients $\mathbf{C}$, obtained through the learning of the reduced low-resolution $\mathbf{\Psi}_r$ dictionary in the previous step, it only requires to use an ordinary least squares to obtain a learned high-resolution dictionary $\mathbf{\Phi}_r$.

6.3. Results

6.3.1. Inputs and Parameters

[26] For construction of the training database, two hundred independent rainfall reflectivity images at resolution 1×1 km were selected over the two GV-sites of the TRMM satellite in Houston, Texas (HSTN) and Melbourne, Florida (MELB) from 1997 to 2010 [see *Ebtehaj and Foufoula-Georgiou*, 2011a]. These high-resolution images are first smoothed and downsampled by scaling ratios $\sqrt{s} \in \{4, 8, 16\}$ to form the low-resolution components of the training set. Then, these low-resolution images are transformed into the feature space by applying the explained set of filters in equation (6). Note that, as the NEXRAD data are in a resolution of $\sim 1 \times 1$ km, these scaling ratios also refer to the spatial grid spacing of the low-resolution rainfall data in kilometers.

[27] Selection of the patch size is obviously an input parameter in the presented SPaD framework. Although the optimal patch size cannot be theoretically determined, for larger patch size the downscaling result is generally smoother. We empirically found that SPaD performs well for patch sizes in the range of 3×3 to 7×7. Here we report the results using overlapping patches of size 3×3. At least $M = 150$, 100 and 50 thousands of training high and low-resolution precipitation patch pairs were extracted to form the initial low and high-resolution empirical dictionaries ($\mathbf{\Psi}$, $\mathbf{\Phi}$) for the selected scaling ratios (i.e., resolutions of 4,8 and 16 km), respectively. To obtain the residual images in equation (5), we adopted a bicubic interpolation operator. While extracting patches from each rainfall image, to screen the possible degenerate patches due to the zero rainfall background effect and piecewise constant regions of rainfall images, those patch pairs $(\mathbf{z}'_l, \mathbf{r})$ were selected for which the variance of the residual patch $\mathbf{r}$ exceeded a certain threshold (i.e., 50th percentile of the variance of all residual patches).

[28] To implement the sparse estimator and solve equation (7), we examined the performance of both a greedy and also an l_1-regularization strategy, by employing the Orthogonal Matching Pursuit (OMP) and the l_1-regularized Least Squares (known as the $l_1 - l_s$) by *Kim et al.* [2007]. The later method, is an interior point method which solves a quadratic programming reformulation of equation (7) [see *Chen et al.*, 1999], using a preconditioned conjugate gradient method.

[29] Employing the OMP in a greedy implementation, we set a fixed number of non-zero representation coefficients (i.e., $\|\mathbf{c}\|_0 = 3$) as the stopping criterion in equation (7) to find relevant atoms of the empirical dictionaries ($\mathbf{\Psi}$, $\mathbf{\Phi}$). However, for the convex l_1-regularized strategy, we used a pair of learned low and high-resolution dictionaries ($\mathbf{\Psi}_r$, $\mathbf{\Phi}_r$) of different sizes $K \in \{256, 512, 1024\}$ for the studied scaling ratios, respectively. These dictionaries were learned by solving equation (8) with a fixed number of non-zero elements in $\mathbf{c}_l$ (i.e., $\|\mathbf{c}_l\|_0 = 3$) and employing the OMP as a sparse coding method. The chosen size of the dictionary is a design issue which requires to consider a trade-off between the accuracy of the recovery and the computational cost. It is empirically found that the selected size of the dictionaries leads to decent results with modest computational cost. Note that, at any scale of interest, the first guess of the reduced low-resolution dictionary for the explained iterative learning, is randomly selected from rainfall patches in the feature space (i.e., columns of $\mathbf{\Psi}$). Figure 3 shows the initial and trained atoms of the learned high-resolution precipitation dictionary, designed for downscaling from scale 4-to-1 km in grid spacing. For selection of the regularization parameter λ, we simply chose $\lambda = 0.1\|\mathbf{\Psi}_r^T \mathbf{y}\|_\infty$, as recommended by *Kim et al.* [2007], which we also found empirically to perform well.

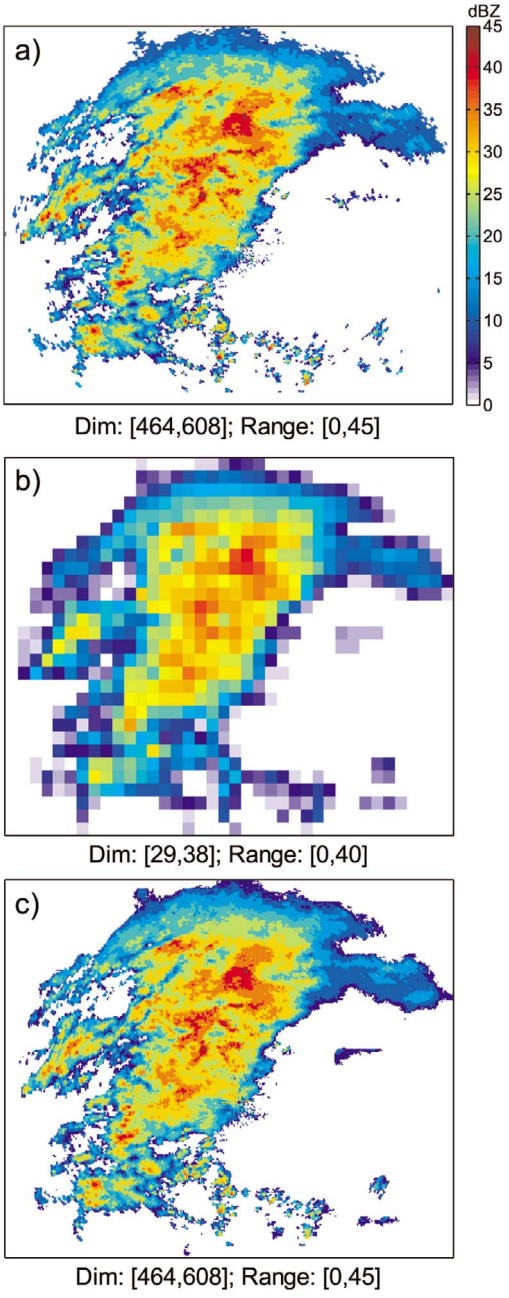

Figure 4. The SpaD performance in a case when the exact high and low-resolution patches are included in the training database. (a) The original high-resolution NEXRAD reflectivity image at resolution 1×1 km, (b) the coarse scale version of the image at scale 16×16 km generated by smoothing and downsampling, and (c) the SPaD results. The pixel-wise Mean Squares Error (MSE), by comparing the original high-resolution with the SPaD output, is 2.10 in $[\text{dBZ}]^2$ which is around 1% of the mean of the squares of rainfall reflectivity values.

6.3.2. Downscaling Performance and Error Analysis

[30] The base-reflectivity snapshots over the HSTN site for the storm of 1998/11/13 (00:02:00 UTC) and 1998/06/28 (18:13:00 UTC), shown in Figures 1 and 2, were selected to study the quantitative performance of the proposed downscaling scheme. In contrast to the stochastic type of downscaling models in which the high-frequency details are generated randomly to reproduce high-resolution realizations with consistent marginal statistics, the proposed method recovers the optimal high-resolution estimate via a supervised learning from our a priori knowledge, obtained by a set of historical training observations. As a result, the final product is not blocky, compared to the stochastic models, and exhibits reduced estimation error. Here, by no means we intend to exhaustively compare the SPaD with stochastic downscaling models; this can be the topic of future studies. We note however that, due to the intrinsic randomness in the stochastic downscaling models, a comparative error analysis is not straightforward. Indeed in these type of models at each realization, the error of estimation (e.g., in a mean square error sense) increases with high probability as the added details are quite random and the downscaled field only preserves marginal statistics and not exact geometric and coherent structures of the storm image. At the same time, error estimation in an ensemble sense does not add much insight as the ensemble average tends to the original input image when the number of ensembles goes to infinity (e.g., consider the case of multiplicative random cascades in canonical form). Therefore, when using these type of stochastic models, the estimation error can not be readily improved even in an ensemble sense, while the comparison also remains ambiguous in the single realization sense.

[31] First to show the effectiveness of the SPaD, we tried to downscale a 16×16 km coarse-scale version of the snapshot on 1998/11/13 while including the corresponding high and low-resolution patches of the rainfall image in the training data set. The results in Figure 4 show how well this method can perform by finding the relevant patches, recover the high-frequency features and estimate the small-scale geometrical structures of the rainfall image of interest for a relatively coarse scale observation. This result is promising that it demonstrates the algorithmic success of our methodology, provided that a suitable prior knowledge is included in the training database.

[32] Figure 5 shows the performance of the SPaD in a case where the image of interest is not itself part of the data set, but rather an independent training database is used. As is evident, the effectiveness of the algorithm degrades as the coarseness of the original image increases (e.g., 16×16 km) and the recovered geometrical structure of the small-scale features is prone to be distorted significantly. Although, the reference high-resolution rainfall field is of course not available in practical cases, it is instructive here to have a quantitative assessment of the downscaling success by comparing the SPaD results $\hat{\mathcal{X}}$ with the original high-resolution rainfall image $\mathcal{X}$. Defining the error matrix $\mathcal{E} = \mathcal{X} - \hat{\mathcal{X}}$, we have chosen some conventional fidelity metrics including: (1) Normalized Bias Ratio: NBR $= \left(\overline{\mathcal{E}}/\overline{\mathcal{X}}\right) \times 100$, where $\overline{(\cdot)}$ denotes the entry-wise expected value (mean); (2) Mean Square Error: MSE $= \overline{\mathcal{E}^2}$; (3) Mean Absolute Error: MAE $=$

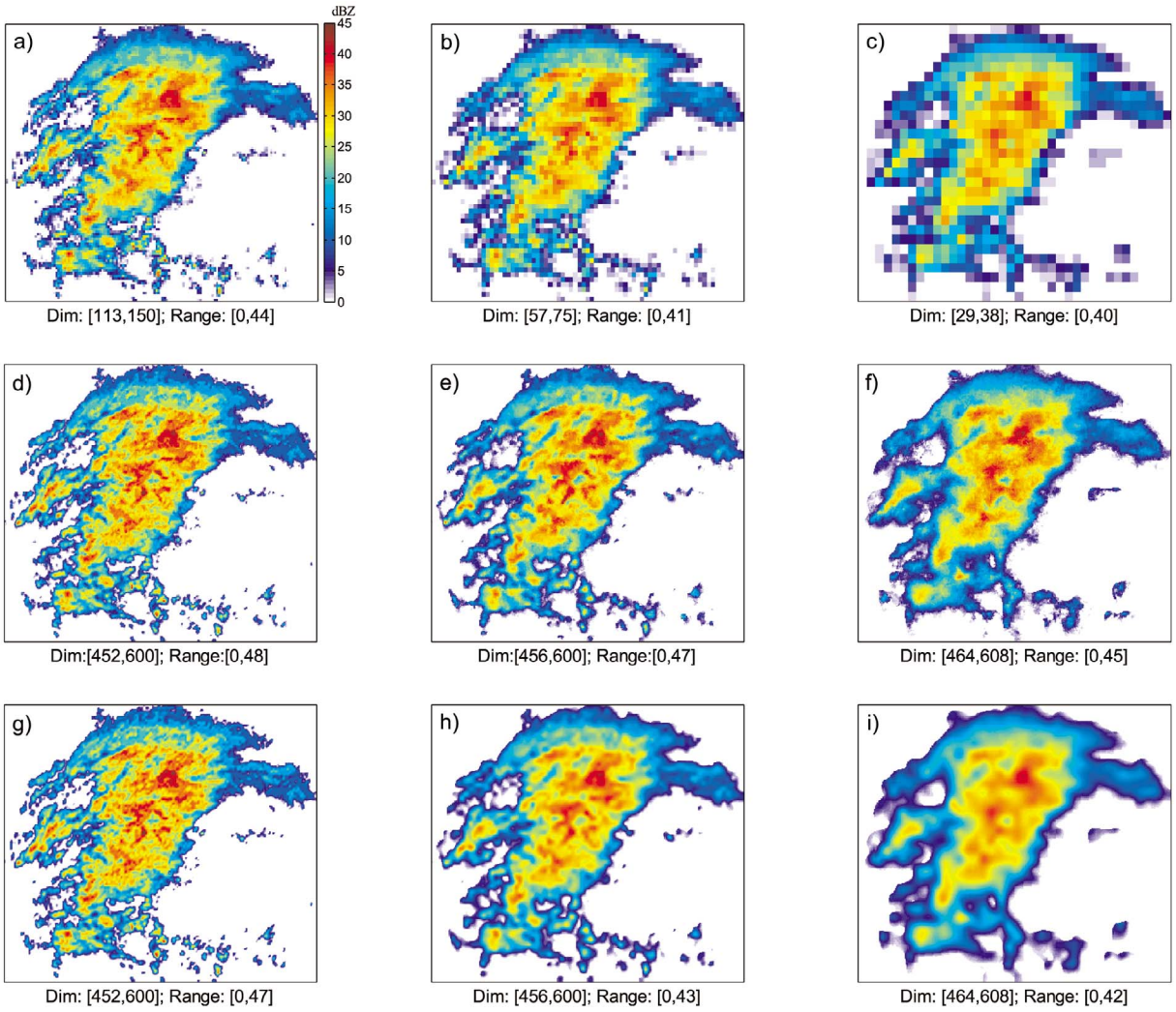

Figure 5. The qualitative performance of SPaD, given 200 regionally selected independent storm snapshots as the training database. (a, b, and c) The low-resolution observations synthetically generated by smoothing and downsampling of the original 1 × 1 km fields to scales of 4,8 and 16 km in grid spacing. (d, e, and f) Results of the downscaled rainfall images at resolution 1 × 1 km using the OMP method for selection of training rainfall patches as the "atoms" of the low and high-resolution dictionaries. (g, h, and i) Results of the downscaled rainfall images at resolution 1 × 1 km employing the l_1-regularization strategy together with the learned dictionaries. As is evident, the SPaD results are smoother, when using the pair of the learned dictionaries of rainfall patches.

$\overline{|\mathcal{E}|}$, where $|\cdot|$ denotes the entry-wise absolute value, and (4) Peak Signal-to-Noise Ratio: PSNR $= 20\log\left(\frac{L}{\sqrt{MSE}}\right)$ in decibel (dB), where L refers here to the dynamic range of the reference high-resolution image.

[33] To evaluate the SPaD performance in recovering the isolated singularities of the rainfall (i.e., local intense rain-cells), we defined a new quality metric which computes the Mean of the Maximum of the Absolute Error (MMAE) for all N overlapping neighborhoods of rainfall intensities (e.g., area of 10 × 10 km), where $\mathbf{e}_{ij}^n$ are the entries of the error matrix $\mathcal{E}$ within the nth neighborhood

$$\mathrm{MMAE} = \frac{1}{N}\sum_{n=1}^{N}\left(\max_{i,j}\left|\mathbf{e}_{ij}^n\right|\right).$$

Comparing to a reference image, many different realizations of precipitation fields may exhibit similar marginal statistics (e.g., MSE); however, their geometrical and coherent structures may be quite different. To address the SpaD performance for recovering the distinct structure and geometrical patterns of rain-cells in the studied precipitation images, a complementary metric, called the Structural SIMilarity (SSIM) by *Wang et al.* [2004], is also employed. Contrary to the previously explained fidelity measures which only take into account the first and second order marginal statistics of the error, this metric considers the covariance structure as well, to overcome some common deficiencies of the MSE type of quality measures. Comparing images $\mathcal{U}$ and $\mathcal{V}$, the SSIM is a local

Table 1. Error Statistics Obtained by Comparing the Low-Resolution Images With the High-Resolution (True) Ones (a), Downscaling via SPaD With Greedy Selection of Atoms by OMP Algorithm (b), and Downscaling via SPaD With l_1-Regularization Together With Dictionary Learning (c)[a]

	Metric Values								
	4 km × 4 km Scale			8 km × 8 km Scale			16 km × 16 km Scale		
Metrics[b]	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
NBR (%)	−0.00	0.43	0.41	−0.00	0.21	0.17	+0.00	−0.11	0.05
	−0.22	0.06	0.21	−0.67	−0.7	−0.59	−1.37	−1.99	−1.58
MSE	1.00	0.6	0.56	1.00	0.52	0.55	1.00	0.54	0.58
	1.00	0.58	0.55	1.00	0.58	0.62	1.00	0.59	0.66
MAE	1.00	0.82	0.80	1.00	0.72	0.74	1.00	0.70	0.75
	1.00	0.80	0.76	1.00	0.76	0.78	1.00	0.73	0.79
MMAE	1.00	0.81	0.77	1.00	0.83	0.81	1.00	0.84	0.84
	1.00	0.83	0.79	1.00	0.84	0.79	1.00	0.83	0.83
PSNR	24.00	26.25	26.50	20.50	23.35	23.15	18.30	21.10	20.70
	21.50	23.90	24.15	18.20	20.50	20.25	15.90	18.15	17.80
SSIM	0.73	0.80	0.82	0.58	0.68	0.69	0.52	0.60	0.60
	0.70	0.80	0.80	0.55	0.65	0.65	0.45	0.54	0.52

[a]For each metric, the first and second rows show the obtained quantities for two reflectivity snapshots on 1998/11/13 (00:02:00 UTC) and 1998/06/28 (18:13:00 UTC), respectively. Note that, for each storm snapshot at a particular scale, the values of MSE, MAE and MMAE are row-wise normalized by their maximum values. The results are shown for downscaling from scales (4, 8, 16 km) down to 1 km in grid spacing.

[b]NBR, Normalized Bias Ratio; MSE, Mean Squared Error; MAE, Maximum Absolute Error; MMAE, Mean Maximum Absolute Error; PSNR, Peak Signal-to-noise ratio; SSIM, Structural Similarity Index.

metric which is calculated for neighborhoods **u** and **v** (typically 8 × 8 pixels) of those images

$$\mathrm{SSIM}(\mathbf{x},\mathbf{y}) = \frac{(2\bar{\mathbf{u}}\bar{\mathbf{v}} + c_1)\left(2\overline{(\mathbf{u}-\bar{\mathbf{u}})(\mathbf{v}-\bar{\mathbf{v}})} + c_2\right)}{(\bar{\mathbf{u}}^2 + \bar{\mathbf{v}}^2 + c_1)\left(\overline{(\mathbf{u}-\bar{\mathbf{u}})^2} + \overline{(\mathbf{v}-\bar{\mathbf{v}})^2} + c_2\right)},$$

where c_1 and c_2 are two constants to stabilize the computation. This index is typically computed for all or a subgroup of patch pairs and the mean value of the index is taken as the overall quality measure, [see *Wang et al.*, 2004]. As is evident, this measure is symmetric and ranges between [−1, +1], while the upper bound is only reachable in the case of two identical images. Here, we do not intend to get into the detailed implications of using the SSIM index; however, this measure seems very promising for studying the forecast mismatch and verification in hydrometeorological studies (see also the forecast verification discussions by *Venugopal et al.* [2005]).

[34] Our numerical experiments show that the bias of the SPaD is within an acceptable range; i.e., the absolute value of the normalized bias ratio is less than 2% in the worst case and less than 0.4% in an average sense; see Table 1. By comparing the results of the SPaD with the observed low-resolution observations, the MSE metric shows significant improvement ranging from 40 to 50%. Computing the MMAE with a neighborhood of size 10 × 10 km, the recovery of the local maxima (intense rain cells) has been enhanced up to 15%. In terms of the PSNR the sparse downscaling shows an order of 2 to 3 dB of improvement and the SSIM index is also increased up to 10–15%.

[35] These preliminary results suggest that the accuracy of the downscaling may depend on the resolution of the original observation, complexity of the storm image and albeit the scaling ratio. However, on the basis of the chosen fidelity measures, the overall quality of the recovery is not significantly deteriorated for larger scaling ratios in our experiments. As expected, in general, smoother results were obtained using the learned dictionaries as the atoms are indeed weighted average of original rainfall patches. The SPaD results via dictionary learning led to better quality measures, typically for spatial scales less than <4 km in grid spacing. For larger scales, the results of the SPaD with crude selection of rainfall patches as dictionary atoms and employing the OMP sparse coding, show better error statistics. For instance, it is clear from Figures 5f and 5i that although the overall quality measures are not significantly different, due to the smoothness of the trained atoms, the dynamic range of the rainfall image cannot be well recovered in larger scaling ratios by the learned dictionaries.

[36] In this study no attempt has been made for computational parallelization and/or coding optimization. All of the computations have been run on a Windows operating system using an Intel(R)-i7 Central Processing Unit with 2.80 GHz clock rate. On the basis of the explained parameters, the code generally ran for less than 5 min to produce downscaled precipitation images of size ∼500 × 500 pixels, with scaling ratio $\sqrt{s} = 16$ (i.e. to downscale from 16 to 1 km). Typically, the lowest computational speed was for the scenarios of using the l_1-regularization scheme together with the learned dictionaries.

7. Concluding Remarks

[37] A new data-driven downscaling framework has been proposed, which exploits the recent developments in sparse signal estimation. It is shown that, in the range of the studied scales, the proposed model can recover the high frequency details of a remotely sensed low-resolution precipitation image from a representative set of training pairs of high and low-resolution rainfall images, via solving a constrained optimization problem. The proposed method results in unique and free of blockiness high-resolution rainfall image with reduced estimation error. The presented results seem encouraging toward larger-scale studies to explore the generality, performance and hydrometeorological practical implications.

[38] The choice of the embedded interpolation operator **Q** (see equation (5)) is a key factor in the performance of the proposed framework and employing the recently developed adaptive directional interpolation methods [e.g., *Mallat and Yu*, 2010] maybe worth exploring in further developments of the SPaD. Another issue worth exploring in the future is the customization of the high and low-resolution dictionaries to the storm environment (e.g., orographic rainfall), which might lead to more accurate downscaling products. The presented methodology is also flexible enough for potential development of a new class of combined radar-radiometer retrieval algorithms for spaceborne estimation of precipitation. Typically, spaceborne estimation of rainfall relies on the Precipitation Radar (PR), e.g., aboard the TRMM satellite, and the passive radiometer with different resolutions and spatial coverage. By design, the PR swath covers a more constrained (narrower) width than the radiometer but with more meteorological details including information on the 3D atmospheric structure. By collecting the coincidental PR and the radiometer brightness temperature frequency channels as the "atoms" of the explained joint dictionaries, a modified version of the presented framework seems feasible for retrieving more detailed precipitation information over the areas of the radiometer swath where PR information is absent.

[39] Note that the scales of applicability and effectiveness of the proposed methodology can not be precisely determined theoretically and need to be tested for different rainfall products and forecasting systems empirically. In addition, multiscale propagation of the estimation error in a closed form expression or numerically has not been covered in this study. However, a numerical estimation of the error seems feasible via bootstrap resampling of the dictionary atoms over a range of the regularization parameter λ [see, e.g., *Tibshirani*, 1996], which can be another potential subject for future research.

[40] **Acknowledgments.** This work has been mainly supported by NASA-GPM award NNX07AD33G and also a University of Minnesota Interdisciplinary Doctoral Fellowship (IDF). Partial support by the National Center for Earth-surface Dynamics, an NSF Science and Technology Center (EAR-0120914), is also gratefully acknowledged. Thanks also go to Arthur Hou and Sara Zhang at Goddard Space Flight Center for insightful discussions and for their interest in this work.

References

Aharon, M., M. Elad, and A. Bruckstein (2006), K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation, *IEEE Trans. Signal Process.*, *54*(11), 4311–4322.

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

Chen, S. S.,D. L. Donoho, and M. A. Saunders (1999), Atomic decomposition by basis pursuit, *SIAM J. Sci. Comput.*, *20*, 33–61.

Claerbout, J. F., and F. Muir (1973), Robust modeling with erratic data, *Geophysics*, *38*, 826–844.

Deidda, R. (2000), Rainfall downscaling in a space-time multifractal framework, *Water Resour. Res.*, *36*(7), 1779–1784, doi:10.1029/2000WR900038

Donoho, D. L. (1995), De-noising by soft-thresholding, *IEEE Trans. Inf. Theory*, *41*(3), 613–627.

Donoho, D. L., and P. B. Stark (1989), Uncertainty principles and signal recovery, *SIAM J. Appl. Math.*, *49*, 906–931.

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

Elad, M. (2010), *Sparse and Redundant Representations*, Springer, New York.

Engan, K., S. O. Aase, and J. H. Husøy (1999), Method of optimal directions for frame design, paper presented at ICASSP 99, IEEE, Phoenix, Ariz.

Engan, K., K. Skretting, and J. H. Husøy (2007), A family of iterative LS-based dictionary learning algorithms, ILS-DLA, for sparse signal representation, *Digital Signal Process.*, *17*, 32–49.

Ferraris, L., S. Gabellani, N. Rebora, and A. Provenzale (2003), A comparison of stochastic models for spatial rainfall downscaling, *Water Resour. Res.*, *39*(12), 1368, doi:10.1029/2003WR002504.

Gupta, V. K., and E. C. Waymire (1993), A statistical analysis of mesoscale rainfall as a random cascade, *J. Appl. Meteorol.*, *32*, 251–267.

Hostetler, S. (1994), Hydrologic and atmospheric models: The (continuing) problem of discordant scales, *Clim. Change*, *27*(4), 345–350.

Kim, S. J., K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky (2007), An interior-point method for large-scale l1-regularized least squares, *Select. Top. IEEE Trans. Signal Process. 1*(4), 606–617.

Lewicki, M. S., and T. J. Sejnowski (2000), Learning overcomplete representations, *Neur. Comput.*, *12*, 337–365

Lovejoy, S. (1981), Analysis of rain areas in terms of fractals, paper presented at 20th Conference on Radar Meteorology, Am. Meteorol. Soc., Boston, Mass.

Lovejoy, S., and B. Mandelbrot (1985), Fractal properties of rain and a fractal model, *Tellus, Ser. A*, *37*, 209–232, doi:10.1111/j.1600-0870.1985.tb00423.x.

Mallat, S., and G. Yu (2010), Super-resolution with sparse mixing estimators, *IEEE Trans. Image Process.*, *19*(11), 2889–2900.

Mallat, S., and Z. Zhang (1993), Matching pursuits with time-frequency dictionaries, *IEEE Trans. Signal Process.*, *41*(12), 3397–3415.

Menabde, M., D. Harris, A. Seed, G. Austin, and D. Stow (1997), Multiscaling properties of rainfall and bounded random cascades, *Water Resour. Res.*, *33*(12), 2823–2830, doi:10.1029/97WR02006.

Nason, G. P., and B. W. Silverman (1995), The stationary wavelet transform and some statistical applications, *Lect. Notes Stat.*, *103*, 281–299.

Nykanen, D., E. Foufoula-Georgiou, and W. Lapenta (2001), Impact of small-scale rainfall variability on larger-scale spatial organization of land-atmosphere fluxes, *J. Hydrometeorol.*, *2*(2), 105–121.

Perica, S., and E. Foufoula-Georgiou (1996), Model for multiscale disaggregation of spatial rainfall based on coupling meteorological and scaling descriptions, *J. Geophys. Res.*, *101*(D21), 26,347–26,361, doi:10.1029/96JD01870.

Rebora, N., L. Ferraris, J. von Hardenberg, and A. Provenzale (2006a), RainFARM: Rainfall downscaling by a filtered autoregressive model, *J. Hydrometeorol.*, *7*(4), 724–738.

Rebora, N., L. Ferraris, J. von Hardenberg, and A. Provenzale (2006b), Rainfall downscaling and flood forecasting: A case study in the Mediterranean area, *Nat. Hazards Earth Syst. Sci.*, *6*, 611–619, doi:10.5194/nhess-6-611-2006,

Schuurmans, J., and M. Bierkens (2007), Effect of spatial distribution of daily rainfall on interior catchment response of a distributed hydrological model, *Hydrol. Earth Syst. Sci.*, *11*, 677–693.

Smith, M. B., V. I. Koren, Z. Zhang, S. M. Reed, J. J. Pan, and F. Moreda (2004), Runoff response to spatial variability in precipitation: an analysis of observed data, *J. Hydrol.*, *298*(1-4), 267–286.

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

Venugopal, V., S. Basu, and E. Foufoula-Georgiou (2005), A new metric for comparing precipitation patterns with application to ensemble forecasts, *J. Geophys. Res.*, *110*, D08111, doi:10.1029/2004JD005395.

Vivoni, E., D. Entekhabi, R. Bras, and V. Ivanov (2007), Controls on runoff generation and scale-dependence in a distributed hydrologic model, *Hydrol. Earth Syst. Sci.*, *11*(5), 1683–1701.

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

Wilby, R. L., and T. Wigley (1997), Downscaling general circulation model output: A review of methods and limitations, *Prog. Phys. Geogr.*, *21*(4), 530–548.

Woods, R., and M. Sivapalan (1999), A synthesis of space-time variability in storm response: Rainfall, runoff generation, and routing, *Water Resour. Res.*, *35*(8), 2469–2485, doi:10.1029/1999WR900014.

Yang, J., J. Wright, T. S. Huang, and Y. Ma (2010), Image super-resolution via sparse representation, *IEEE Trans. Image Process. 19*(11), 2861–2873.

Younger, P. M., J. E. Freer, and K. J. Beven (2009), Detecting the effects of spatial variability of rainfall on hydrological modeling within an uncertainty analysis framework, *Hydrol. Process.*, *23*(14), 1988–2003.

Zeyde, R., M. Elad, and M. Protter (2010), On single image scale-up using sparse-representations, paper presented at 7th International Conference on Curves and Surfaces, SMAI-AFA, Avignon, France.

A. M. Ebtehaj and E. Foufoula-Georgiou, Department of Civil Engineering, Saint Anthony Falls Laboratory, University of Minnesota, 2 3rd Ave. SE, Minneapolis, MN 55414, USA. (ebteh001@umn.edu; efi@umn.edu)

G. Lerman, School of Mathematics, University of Minnesota, 127 Vincent Hall, 206 Church St. SE, Minneapolis, MN 55455, USA. (lerman@umn.edu)