

Adaptive fusion of multisensor precipitation using Gaussian-scale mixtures in the wavelet domain

Ardeshir Mohammad Ebtehaj^{1,2} and Efi Foufoula-Georgiou¹

Received 7 May 2011; revised 20 August 2011; accepted 22 August 2011; published 19 November 2011.

[1] The past decades have witnessed a remarkable emergence of new sources of multiscale multisensor precipitation data, including global spaceborne active and passive sensors, regional ground-based weather surveillance radars, and local rain gauges. Optimal integration of these multisensor data promises a posteriori estimates of precipitation fluxes with increased accuracy and resolution to be used in hydrologic applications. In this context, a new framework is proposed for multiscale multisensor precipitation data fusion which capitalizes on two main observations: (1) non-Gaussian statistics of precipitation images, which are concisely parameterized in the wavelet domain via a class of Gaussian-scale mixtures, and (2) the conditionally Gaussian and weakly correlated local representation of remotely sensed precipitation data in the wavelet domain, which allows for exploiting the efficient linear estimation methodologies while capturing the non-Gaussian data structure of rainfall. The proposed methodology is demonstrated using a data set of coincidental observations of precipitation reflectivity images by the spaceborne precipitation radar aboard the Tropical Rainfall Measurement Mission satellite and by ground-based weather surveillance Doppler radars.

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

1. Introduction

[2] Availability of multisensor observations and the desire for accurate high-resolution forecasts have always been a strong motivation to explore algorithms for optimal integration of multisensor, multiscale geophysical data into a more complete set of information with reduced uncertainty. In the last decade, a large set of multiscale precipitation data has been collected using locally distributed traditional networks of rain gauges and regional ground-based weather surveillance radars, e.g., Next Generation Radar (NEXRAD) and global spaceborne active and passive satellite sensors, e.g., Tropical Rainfall Measuring Mission (TRMM). The acquired remotely sensed precipitation data might be mathematically understood as the convolution of the highly irregular reflectivity field of the atmospheric hydrometeors with a low-pass operator, naturally corrupted with noise due to intrinsic measurement fluctuations or to sensor noise. Several studies have been conducted to characterize the measurement error of the remotely sensed precipitation data [e.g., Ciach and Krajewski, 1999; Wang and Wolff, 2009], whereas much less attention has been devoted to developing consistent and robust algorithms that can be used to filter out

these errors and optimally merge (fuse) the multisensor data for obtaining a posteriori estimates of the precipitation fields.

[3] The standard linear Gaussian filtering methods on Markov treelike structures, the so-called scale-recursive estimation (SRE) technique, has been commonly proposed to assimilate remotely sensed rainfall observations at different scales into a stochastic (e.g., multiplicative random cascade) model of rainfall fields [Gorenburg *et al.*, 2001; Tustison *et al.*, 2002; Gupta *et al.*, 2006; Bocchiola and Rosso, 2006; Bocchiola, 2007; Van de Vyver and Roulin, 2009]. The main advantage of this method is its efficiency to provide a recursive least squares solution for high-dimensional multiscale Gauss-Markov estimation problems. Consequently, by construction, fusion of multisensor rainfall data using SRE is based on the assumption that the multiscale statistical structure of the precipitation fields (or log transformed fields) is linear and can be explained in the Gaussian domain.

[4] Early observations signified that spatial rainfall exhibits a clustered behavior [e.g., LeCam, 1961], meaning that areas of high-intensity precipitation, referred to as “rainfall cells,” tend to occur in clusters within regions of lower rain rate. This had been the earliest motivation for modeling approaches that sought to represent the observed geometry and statistical structure of precipitation by means of clustered point processes [e.g., see Gupta and Waymire, 1993]. These isolated high-intensity clusters of rainfall cells manifest themselves in the tail statistics of the rainfall histogram which are remarkably thicker than the domain of Gaussian distributions, including the lognormal density [Ebtehaj and

¹Saint Anthony Falls Laboratory, Department of Civil Engineering and National Center for Earth-Surface Dynamics, University of Minnesota, Twin Cities, Minneapolis, Minnesota, USA.

²School of Mathematics, University of Minnesota, Twin Cities, Minneapolis, Minnesota, USA.

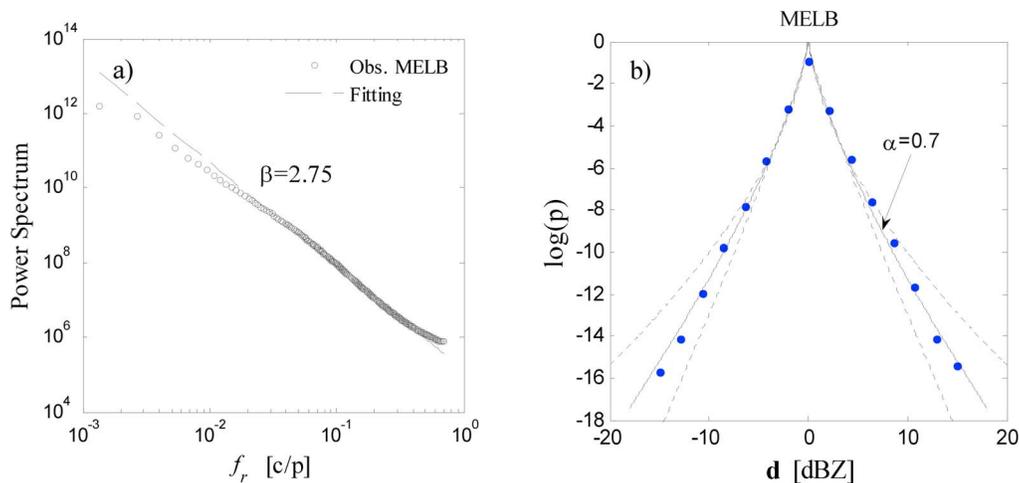


Figure 1. (a) Radially averaged ensemble spectrum of 105 NEXRAD precipitation single-level near-surface reflectivity images at resolution 1×1 km over the ground validation site of the TRMM satellite in Melbourne, Florida. Frequencies are in cycle per pixel (c/p), which are equivalent to the inverse of pseudo spatial scale km^{-1} . (b) Associated average histogram of the horizontal subband coefficients \mathbf{d} . The solid circles show the mean empirical histogram, and the solid line is the fitted generalized Gaussian distribution $p(x) \propto \exp(-|x|^\alpha)$. The dashed lines are the 95% estimation quantiles.

Foufoula-Georgiou, 2010, 2011]. These observations indicate that by using the SRE method for rainfall multiscale estimation problems, important high-order statistical characteristics of rainfall cannot be thoroughly captured and the result of the rainfall data fusion may be an overly smooth representation without adequate detailed features of the storm rainfall cells and extreme intensity values. This obviously calls for developing consistent and well-structured filtering and fusion methodologies that can efficiently address the distinct non-Gaussian estimation issues of high-dimensional precipitation fields and also permit incorporation of different sources of the multiscale measurement errors in the fusion process.

[5] Recent studies [*Ebtehaj and Foufoula-Georgiou*, 2011] demonstrated that a particular mixture of Gaussian random variables can well capture the observed heavy tail properties of the precipitation data in the wavelet domain. In this paper, we will explain how this probability model in the wavelet domain can be exploited for optimal multiscale fusion of multisensor precipitation data. Using the developed methodology, we present a case study in which, first, rain gauge corrected products are derived via filtering the measurement error from the coincidental observations of the TRMM-PR satellite and ground-based NEXRAD reflectivity data and, second, the rain gauge corrected products are merged in a multiscale framework. The advantages of the introduced method over the conventional linear Gaussian estimation technique are also discussed. Accordingly, this paper is structured as follows:

[6] Section 2 is devoted to briefly discussing the statistics and non-Gaussian structure of rainfall reflectivity images in the wavelet and Fourier domains. In the context of precipitation multisensor fusion, the standard linear Gaussian estimation method is explained and implemented in section 3. Practical aspects of implementation and shortcomings of this

method are also discussed in section 3. Section 4 explains the new proposed probability model, namely the Gaussian-scale mixture (GSM), for precipitation reflectivity images in the wavelet domain which can be used for consistent and robust multiscale multisensor data fusion. In section 5, basic theoretical and practical concepts of optimal estimation in the wavelet domain using the GSM probability model are explained. A synthetic one-dimensional example is also presented to elaborate the main advantages of the proposed methodology compared with the SRE method. Section 6 describes the implementation of the new model for precipitation estimation and data fusion by applying it to a real storm event coincidentally measured by ground-based NEXRAD and TRMM-PR sensors. Section 7 presents a summary of the study and points out some directions for future research.

2. Non-Gaussian Statistics of Precipitation Images

[7] *Ebtehaj and Foufoula-Georgiou* [2011] studied the statistical structure of a data set populated by near-surface precipitation images in decibels of reflectivity (dBZ), from two hundred independent storms coincidentally observed by NEXRAD and TRMM precipitation radars over two TRMM ground validation (GV) sites in Houston, Texas (HSTN) and Melbourne, Florida (MELB). It was demonstrated that the Fourier and wavelet decompositions of these fields permit concise parameterization across a range of spatial scales. Specifically, it was revealed that besides the power law decay of the Fourier spectra (see Figure 1a), the distribution of the wavelet coefficients (smoothed increments) shows a symmetric cusp singularity around the center with extended heavy tails significantly thicker than the Gaussian case (see Figure 1b). Although the conversion of rainfall reflectivity

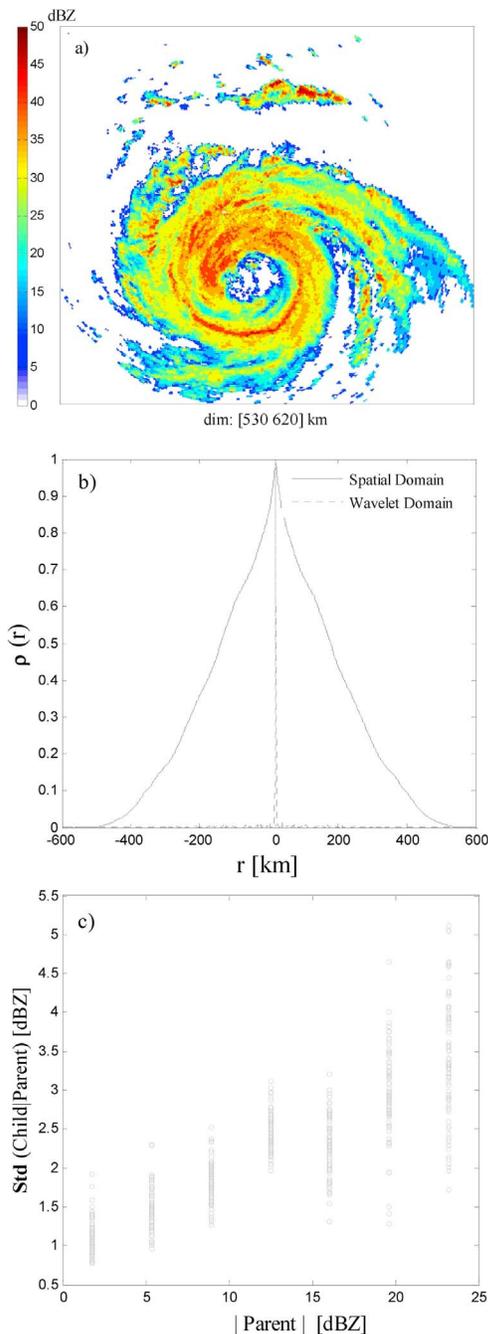


Figure 2. (a) A NEXRAD precipitation long-range reflectivity image at resolution 1×1 km over the MELB site on 26 September 2004 at 04:50:00 UTC. (b) The central horizontal transect of the two-dimensional correlation function in real (solid line) and wavelet (broken line) domain. This shows that the correlation quickly decays in a short range of spatial lags in the wavelet domain as opposed to the spatial domain. (c) Standard deviation of the wavelet coefficients of the reflectivity fields (children) conditioned on the magnitude of the coefficients at the next coarser scale (parents), obtained from 100 independent storm snapshots over the HSTN-GV site, demonstrating the scale-to-scale dependence of the reflectivity fluctuations. Note that here wavelet coefficients refer to the horizontal subbands at one level of decomposition.

in dBZ to precipitation intensity is not a trivial task [see, e.g., *Krajewski and Smith, 2002*], a commonly used Z - R relationship would imply that the wavelet coefficients of the log transformed rainfall intensity would maintain the singular structure shown in Figure 1b, up to a scaling factor. Using an orthogonal overcomplete stationary wavelet transform [Nason and Silverman, 1995; Coifman and Donoho, 1995], it has been found [Ebtehaj and Foufoula-Georgiou, 2011] that this sort of non-Gaussian heavy tail property of the rainfall fields in the wavelet domain can be well explained by a family of elliptically symmetric distribution functions, namely the generalized Gaussian (GG) density $p(x) \propto \exp(-|x|_s^\alpha)$, where α and s are positive real numbers which control the tail and width of the distribution, respectively (see Figure 1b). Another important observation made by Ebtehaj and Foufoula-Georgiou [2011] was that the wavelet transformation cannot completely whiten the strongly correlated spatial rainfall reflectivity images. It was revealed that despite the Karhunen-Loève-like decorrelation property (see Figures 2a and 2b) of the wavelet transform, Wornell [1990], the wavelet coefficients of precipitation reflectivity images at nearby subbands (same scale or across scales) exhibit considerable short-range dependence. Studying the relationship of the wavelet coefficients at two consecutive scales under the name of parent and child, Figure 2c demonstrates that the conditional variance of the children highly depends on the parent magnitude. Indeed, larger parents give rise to children with larger magnitude and spread. However, this scale-to-scale dependence is not the entire story as the empirical observations also signify that the rainfall fluctuations are also highly structured at each subband with noticeable intrascale short-range correlation. Specifically, decomposing a precipitation reflectivity image, Figure 3 shows the image representation of the covariance matrices of a local neighborhood of size 5×5 for the three different subbands at the same scale. This clearly shows that the off-diagonal elements of the covariance matrices are significant and should not be neglected for accurate synthetic reconstruction and/or estimation purposes in the wavelet domain. Note that, to make the statistical characterization and the results of the proposed fusion methodologies independent of the storm specific Z - R relationship, all of the analyses in this work are performed on precipitation reflectivity images in dBZ. Therefore, up to a scaling factor, the results can be generalized to the log rainfall intensity accepting a linear relationship between rainfall $R[\text{mm}/\text{hr}]$ and decibels of reflectivity Z ; i.e., $Z = 10 \log(Z/Z_0)$, via the common $Z = aR^b$ relationship, where Z refers to the reflectivity and $Z_0 = 1 \mu\text{m}^3$ is the reference level.

3. Linear Fusion of Multisensor Precipitation Data in the Spatial Domain

[8] Since minimization of quadratic error functionals is a convex problem with a unique solution, least squares estimation has played a central role in a large family of statistical inference and estimation problems. As in this study optimal estimation or fusion of multisensor precipitation observations builds upon the basic theory of least squares estimation, a brief explanation of the relevant theoretical aspects is presented in this section.

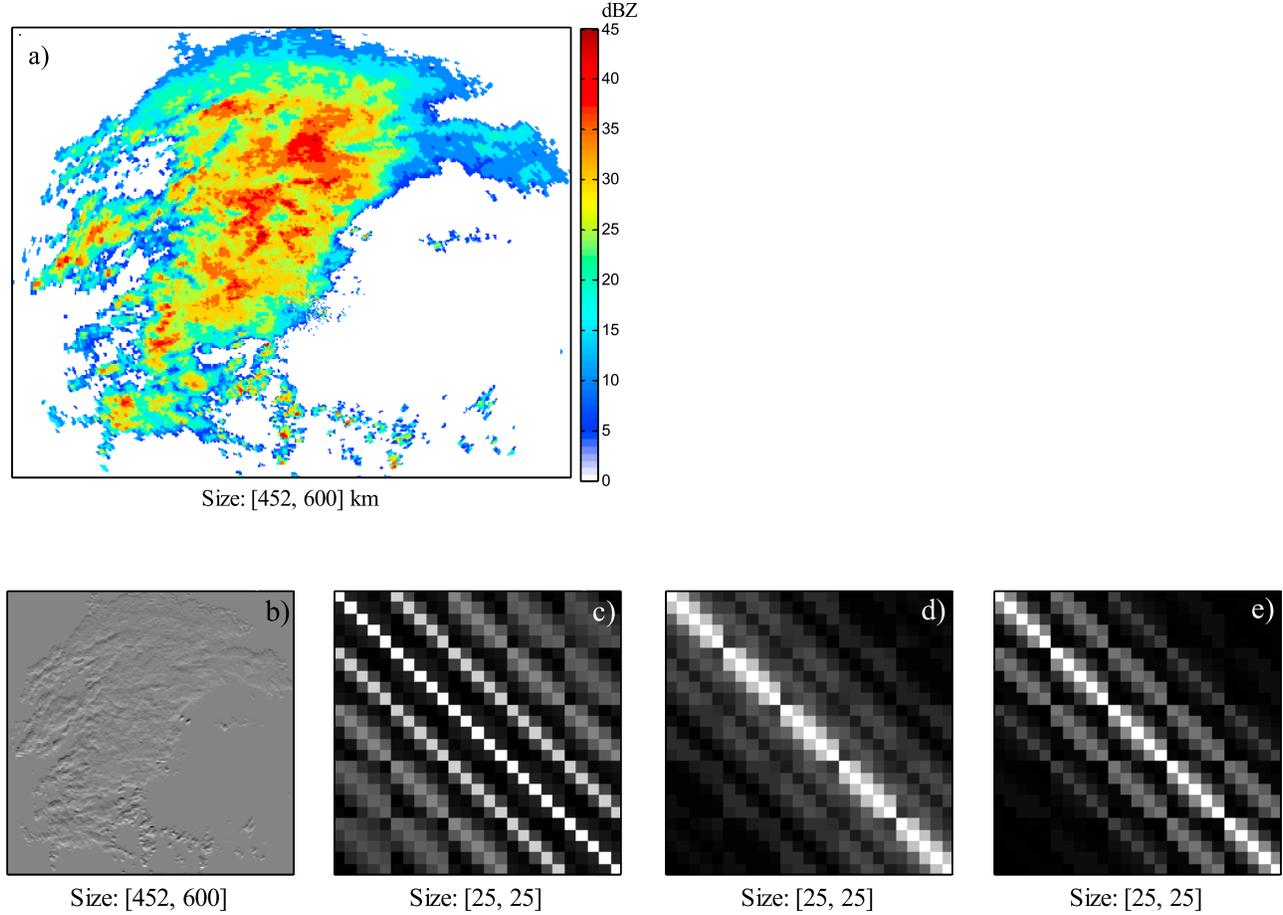


Figure 3. (a) A NEXRAD long-range reflectivity image (1×1 km) of a sample storm over MELB and (b) its wavelet horizontal subband image. The off-diagonal elements of the image representation of the covariance matrices of neighborhoods of size 5×5 for (c) horizontal, (d) vertical, and (e) diagonal subbands of the reflectivity image in Figure 3a show that the wavelet coefficients are weakly correlated in space at different subbands and the correlation almost vanishes on the boundaries. This shows that the covariance structure of the rainfall data in the wavelet domain can be reasonably well captured “locally” in a finite spatial range (in this resolution <5 km).

3.1. Principles of Least Squares Estimation

[9] Consider a set of noisy measurements $\mathbf{y} \in \mathbb{R}^N$ of a parameter vector $\mathbf{x} \in \mathbb{R}^K$ with joint covariance matrix of the form $\Sigma = [\Sigma_x, \Sigma_{xy}; \Sigma_{yx}, \Sigma_y]$. Without having any a priori assumption about the density of observations and parameter vectors, the Bayesian least squares estimator of \mathbf{x} and the associated covariance (Σ) of estimation can be derived as,

$$\begin{aligned} \hat{\mathbf{x}} &= m_x + \Sigma_{xy} \Sigma_y^{-1} (\mathbf{y} - m_y) \\ \Sigma_{\hat{\mathbf{x}}} &= \Sigma_x - \Sigma_{xy} \Sigma_y^{-1} \Sigma_{yx}, \end{aligned}$$

where $m_x = \mathbb{E}[\mathbf{x}]$ and $m_y = \mathbb{E}[\mathbf{y}]$ [see, e.g., *Levy, 2008*].

[10] Casting this problem in the context of a linear measurement equation of the form $\mathbf{y} = C\mathbf{x} + \mathbf{v}$ in the Gaussian noise environment $\mathbf{v} \sim \mathcal{N}(0, \Sigma_v)$, where $C \in \mathbb{R}^{N \times K}$ is the measurement matrix, and knowing $\Sigma_{A\mathbf{x}, B\mathbf{y}} = A\Sigma_{\mathbf{x}\mathbf{y}}B^T$ for any

matrices A and B of relevant size, the above expression can be further expanded as follows.

$$\hat{\mathbf{x}} = m_x + \Sigma_x C^T (C\Sigma_x C^T + \Sigma_v)^{-1} (\mathbf{y} - Cm_x) \quad (1)$$

$$\Sigma_{\hat{\mathbf{x}}} = \Sigma_x - \Sigma_x C^T (C\Sigma_x C^T + \Sigma_v)^{-1} C\Sigma_x. \quad (2)$$

[11] The least squares estimation $\hat{\mathbf{x}}$ of \mathbf{x} given \mathbf{y} , is indeed the projection of \mathbf{x} onto the linear subspace spanned by \mathbf{y} or say $\text{span}\{\mathbf{y}\}$, which is optimal in the sense that $\mathbb{E}[|\mathbf{x} - \hat{\mathbf{x}}|^2] \leq \mathbb{E}[|\mathbf{x} - \text{span}\{\mathbf{y}\}|^2]$. In the case that \mathbf{x} and \mathbf{y} are in the Gaussian domain (linear filtering), the least squares estimator is fully optimal in the sense that it coincides with the conditional expectation $\hat{\mathbf{x}} = \mathbb{E}[\mathbf{x}|\mathbf{y}]$. However, in the case of non-Gaussian distributions (nonlinear filtering), the conditional expectation is a nonlinear function of the measurements and

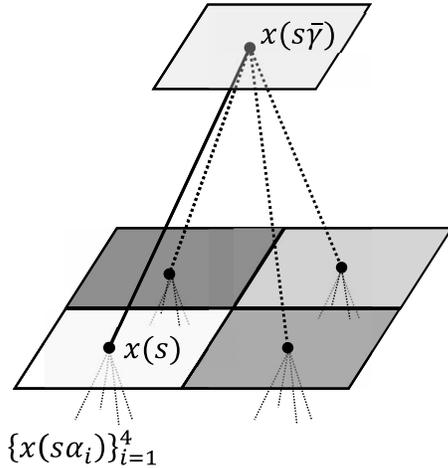


Figure 4. Regular quadtree structure typically used to model two-dimensional multiresolution Markov fields in which each parent node gives birth to four children nodes. In this representation, each node on the tree represents the process magnitude in a spatial region and specific location, while each level of the tree corresponds to a particular resolution (scale) of the process.

the least squares estimator is just a suboptimal linear estimation of the conditional expectation. Note that in general, the conditional expectation is optimal in the sense that $\mathbb{E} [|\mathbf{x} - \mathbb{E} [\mathbf{x}|\mathbf{y}]|^2] \leq \mathbb{E} [|\mathbf{x} - f(\mathbf{y})|^2]$, where $f(\mathbf{y})$ denotes any nonlinear function of the observations [see *Fristedt et al.*, 2007; *Levy*, 2008]. In practice, for high-dimensional problems obtaining this estimator may require inversion of the measurement covariance matrix, which might be computationally cumbersome especially in temporal systems while online measurements become available sequentially and cumulative in time.

[12] A least squares estimation paradigm was introduced by *Kalman* [1960] for the estimation of discrete time linear Gauss-Markov stochastic processes, i.e., $\mathbf{x}_t = A_{t-1}\mathbf{x}_{t-1} + w_{t-1}$, where A_t is the temporal transition matrix and $w_t \sim \mathcal{N}(0, \Sigma_w)$ is a white Gaussian noise vector, known as the model error. Minimizing the trace of the covariance matrix of the estimates, this formalism allows us to sequentially obtain the conditional expectation of the system state variables $\hat{\mathbf{x}}_t = \mathbb{E} [\mathbf{x}_t | \mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_t]$ in time, given the noisy observations in the framework of an affine measurement equation $\mathbf{y}_t = C_t \mathbf{x}_t + v_t$, where C_t relates the system state to the measurements and $v_t \sim \mathcal{N}(0, \Sigma_v)$. Obviously, for such a Gaussian dynamic system, Kalman filter (KF) is an optimal estimator as the conditional expectation, and the associated covariance can fully explain the entire probabilistic structure of the system. Replacing the notion of time with scale, the original idea of the linear estimation of temporal Gauss-Markov systems was further expanded by *Chou et al.* [1994] to the optimal estimation of multiresolution auto-regressive (MAR) Gaussian processes [e.g., *Luetgen et al.*, 1993; *Daniel and Willsky*, 1999; *Willsky*, 2002]. In MAR representation, a multiresolution process is naturally defined on a treelike graph structure \mathcal{T} (see Figure 4), where each node $s \in \mathcal{T}$ on

the tree is a 3-tuple which indicates the signal quantity $x(s)$ in a specific translational offset and scale level:

$$x(s) = A(s)x(s\bar{\gamma}) + w(s). \quad (3)$$

In this coarse-to-fine multiresolution dynamics, $s\bar{\gamma}$ denotes the parent node of s , $A(s)$ is the transition matrix and $w(s) \sim \mathcal{N}(0, \Sigma_{w(s)})$ is a Gaussian white noise. The system states $x(s)$ at each node s are distributed as $\mathcal{N}(0, \Sigma_{x(s)})$, where the covariance $\Sigma_{x(s)} = \mathbb{E} [x(s)x(s)^T]$ evolves according to the discrete Lyapunov equation across scales:

$$\Sigma_{x(s)} = A(s)\Sigma_{x(s\bar{\gamma})}A^T(s) + \Sigma_{w(s)}.$$

In this setting, the measurement equation at different scales is given by

$$y(s) = C(s)x(s) + v(s), \quad (4)$$

where $C(s)$ specifies the linear relationship between $x(s)$ and what is measured at each node and $v(s)$ is a Gaussian white noise independent of $w(s)$ with covariance $\Sigma_{v(s)} = \mathbb{E} [v(s)v(s)^T]$. Analogous to the two-pass smoother algorithm for temporal system dynamics by *Rauch et al.* [1965], known as the RTS-Smoother due to the author names, the presented two-pass scale-recursive estimation (SRE) algorithm by *Chou et al.* [1994] permits fusing (assimilating) such measurements into the estimation process of the MAR dynamics of equation (3).

3.2. Scale-Recursive Estimation of Multisensor Precipitation Data

[13] In the past decade, several research groups exploited the SRE framework for fusing coincidental multiscale soil moisture and precipitation data [e.g., *Kumar*, 1999; *Gorenburg et al.*, 2001; *Tustison et al.*, 2002; *Gupta et al.*, 2006; *Bocchiola and Rosso*, 2006; *Bocchiola*, 2007; *Van de Vyver and Roulin*, 2009]. Typically, a regular quadtree \mathcal{T}_q structure is being used to model coarse-to-fine scale dynamics of precipitation fields in the SRE framework, where each node s on the tree at scale-level $j(s)$ gives rise to four children nodes, namely $\{s\alpha_i\}_{i=1}^4$, at scale-level $j(s) + 1$ (see Figure 4). Scale-recursive estimation of multisensor precipitation data has some subtle technical issues. For example, it needs to be assumed that precipitation data can be modeled from coarse-to-fine scales by a Gaussian MAR dynamics. Furthermore, rainfall is a positive process and the background of precipitation fields is often filled with too many zero intensity values. How should these zeros be handled in the context of a stochastic MAR model while preserving the positivity condition and proper correlation structure of the precipitation data?

3.2.1. Implementation on Rainfall Fusion

[14] Typically, in this setting the coarse resolution precipitation data (i.e., satellite observations or general circulation model predictions) are considered as model information $x(s)$ at the root node of the quadtree, where a MAR model is being used to produce the rainfall information at higher resolutions of interest. Consequently, the available observations at finer resolution, such as the data from the NEXRAD

weather surveillance radars, are generally considered as measurements $y(s)$.

[15] In the context of rainfall data, it has long been argued [e.g., *Lovejoy and Schertzer*, 1990; *Gupta and Waymire*, 1993] that these fields can be explained using a nonlinear multiplicative scale-to-scale stochastic structure; e.g., $r(s) = r(s\bar{\gamma})\zeta(s)$, where $\zeta(s)$ represents a driving random component known as the cascade generator with $\mathbb{E}[\zeta(s)] = 1$, in a canonical form. Working on high-resolution rainfall data, *Menabde et al.* [1997] proposed a lognormal density for the $\zeta(s)$. To treat this nonlinear recursion and make it consistent with the settings in equations (3) and (4), typically the SRE method is performed in the log transformed rainfall; i.e., $\log[r(s)] = x(s)$, or say a shifted version of the reflectivity fields [e.g., *Gorenburg et al.*, 2001; *Tustison et al.*, 2002],

$$\log[r(s)] = \log[r(s\bar{\gamma})] + \log[\zeta(s)], \quad (5)$$

where $\log[\zeta(s)]$ is a Gaussian white noise, equivalent to the term $w(s)$ in equation (3). In terms of the first-order and marginal statistics, this transformation seems fine; however, the log transformation cannot completely transform a rainfall field into a Gaussian process and change the multiplicative scale-to-scale correlation into an additive structure. For instance, it is easy to check that the conditional variance in a multiplicative recursion depends on the magnitude of the process at the next coarser scale; i.e., $\text{var}[r(s)|r(s\bar{\gamma})] = (r(s\bar{\gamma}))^2 \text{var}[\zeta(s)]$, which is not the case in equation (5), as long as the term $\log[\zeta(s)]$ remains a “white” type of Gaussian noise at different scales. In effect, some important higher-order scale-to-scale statistical structures are ignored in linear estimation of rainfall data in the log transformed domain. Moreover, we have also shown that the marginal histogram of the rainfall reflectivity data (logarithm of rainfall through Z - R relationship) is far from being in the Gaussian domain of attraction (see Figure 1b) [*Ebtehaj and Foufoula-Georgiou*, 2010, 2011].

[16] Considering all of these major MAR model incompatibilities with the observed statistical structure of rainfall, the standard Gaussian multiscale filtering technique still provides a very efficient global least square estimator of the multiscale multisensor precipitation data. Here an example is provided which uses the SRE framework to merge precipitation given from TRMM-PR and ground-based NEXRAD coincidental precipitation reflectivity imageries. We assumed that the reflectivity images can be partially explained by the linear MAR model in equation (3). The tree is assumed stationary in the sense that $A(s) = I$ and as explained the “model” information $x(s)$ is obtained from the coarse resolution TRMM near-surface reflectivity images at $\approx 4 \times 4$ km and the “observations” $y(s)$ are set to the NEXRAD high-resolution reflectivity imageries at 1×1 km. We assumed that both sensors provide unbiased precipitation estimates in a global sense and hence we set $C(s) = I$. To address the self-similarity and commonly observed $1/f$ spectrum in the rainfall reflectivity images (see, e.g., Figure 1a), it is assumed that the variance of the driving noise term $w(s)$ decays geometrically from coarse-to-fine scales by assigning $\Sigma_{w(s)} \propto 2^{-Hj(s)}I$, where the scalar parameter $H > 0$ refers to the self-similarity index, and $j(s)$ represents coarse-to-fine scale levels at node s . This parameter controls the dropoff

rate of the power spectrum of the synthesized fields [*Daniel and Willisky*, 1999] and can be estimated from the available high-resolution NEXRAD data. To this end, we simply employed the concept of image pyramid encoding [*Burt and Adelson*, 1983]. The original NEXRAD image can be coarsened by smoothing and downsampling by a factor of 2, using an average filter of size 2×2 , to produce an “approximate” representation of the field at the next coarser scale. The approximate coarse scale image is then upsampled by a factor of 2 and convolved with a nearest neighborhood interpolator to produce the so-called “prediction field,” which will have the same dimensions as the original NEXRAD image. The difference between the prediction field and the original one indeed gives us the “detail information” which is needed to reconstruct perfectly the high-resolution original image given the low-resolution (approximate) version at the next coarser scale [see *Gonzalez and Woods*, 2008]. Recursive implementation of this encoding procedure yields characterization of the scale-to-scale detail information and characterization of the noise term $w(s)$ in equation (3). Indeed, we used the high-resolution NEXRAD precipitation data to estimate the required self-similarity exponent of the MAR model to provide TRMM rainfall information $x(s)$ at higher resolution of interest on the tree.

[17] The background effect in rainfall fields is significant, meaning that a major portion of the image is typically filled with zero intensity values. Therefore, crudely putting a coarse resolution rainfall image (e.g., TRMM observations) in the framework of the MAR dynamics will result in adding noise to the background of the image which is in contradiction with the positivity assumption of the rainfall fields and will lead to an invalid representation. To resolve this problem, we just prune the tree from coarse-to-fine scales, meaning that the zero intensity parent nodes in the TRMM data are assumed impotent in giving rise to nonzero children. By this assumption, background pixels (zeros) of the coarse resolution TRMM image remain zero in the finer scale images produced by the MAR model of equation (3).

3.2.2. Results of SRE Fusion

[18] Using SRE, Figure 5 shows the fusion of high-resolution (i.e., 1×1 km) NEXRAD observations of a storm event over the HSTN site on 28 June 1998 at 18:13:00 UTC with a coincidental TRMM-PR observation at resolution $\approx 4 \times 4$ km. The covariance of the measurement noise has a significant implication on the outcome of the fusion process. Choosing a large $\Sigma_{v(s)}$ compared to the process covariance at each node will yield a fusion product which is more inclined toward the model information (i.e., TRMM data) and vice versa. Obviously, estimation of the properties of this covariance requires some statistical inference from a large set of coincidental observations. *Ebtehaj and Foufoula-Georgiou* [2011] characterized the diagonal elements of this covariance for the data set used in this study in terms of a signal-to-noise ratio (SNR) measure in decibel (dB) scale; i.e., $\text{SNR} = 10 \log_{10} \left(\frac{\sigma^2(s)}{\sigma_{v(s)}^2} \right)$, where $\sigma(s)$ is the standard deviation of the NEXRAD image and $\sigma_{v(s)}$ represents the standard deviation of the error between the mean of the coincidental images of the TRMM-PR and NEXRAD observations. This error covariance can be characterized as a function of scale; however, in this study this measure is just estimated in an average sense at the TRMM resolution and

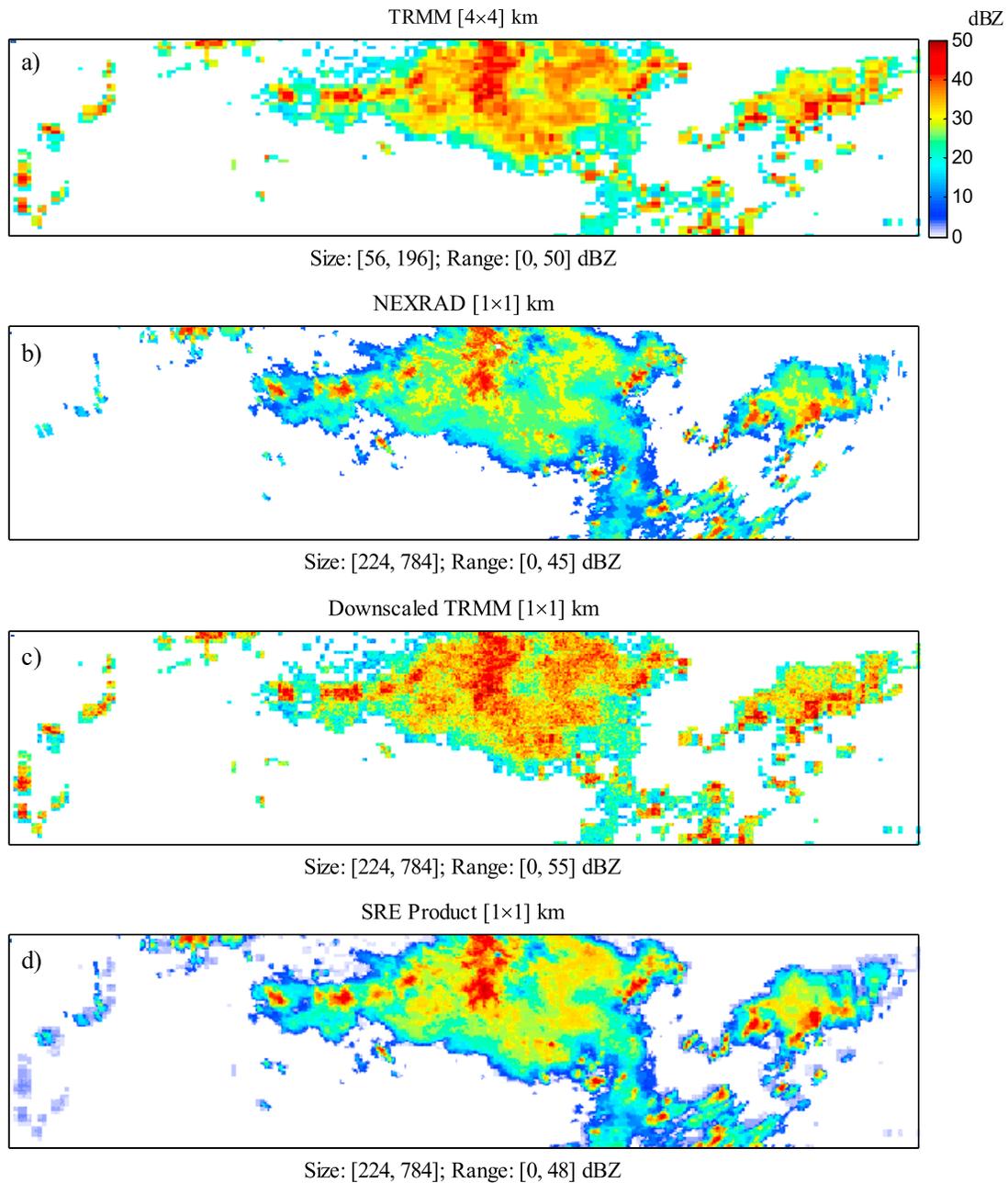


Figure 5. (a) TRMM-PR near-surface single-level reflectivity image over the HSTN site on 28 June 1998 at 18:13:00 UTC. (b) Coincidental NEXRAD base reflectivity image of the storm. (c) Downscaled TRMM observations according to the linear MAR model ($H = 0.6$) at resolution 1×1 km. (d) Fused product according to the SRE framework, assuming $\text{SNR} = 11.0$ dB (see text for discussion).

generalized to all scales. Our studies suggest that the SNR in the defined context ranges from 10.0 to 20.0 dB, meaning that the noise power is less than 10% of the rainfall energy.

[19] As a true representation of the precipitation fields is not generally available and all of the rainfall products are just estimates of the process, a quantitative evaluation of the performance of the fusion process in terms of error norms is not very straightforward. Nevertheless, it is generally expected that the information content of a field is augmented as a result of a fusion process. To address this fact, the

entropy or the expected value of the information content of the rainfall images can be estimated and compared,

$$\text{Entropy} = - \sum_k p(\mathcal{Z}_k) \log_2 [p(\mathcal{Z}_k)]$$

where $p(\cdot)$ is the probability measure of the rainfall reflectivity image intensity values \mathcal{Z}_k falling within the k^{th} bin of the image histogram. Using the $\log_2(\cdot)$ function, the entropy measures the average information content in terms of bits

Table 1. Frobenius Norm of the TRMM-PR and NEXRAD Observations and SRE-Fused Product, Rowwise Normalized by the Maximum of Each Row^a

Scale	Normalized Frobenius Norm		
	TRMM-PR	NEXRAD	SRE Fusion
4 × 4 km	1	0.76	0.80
1 × 1 km	1	0.78	0.84

^aTRMM-PR shows a larger F -norm in this case, and the fused product falls within the lower and upper bounds. In the SRE fusion method, the fine-resolution image of the TRMM radar is obtained using a synthetic realization of the MAR model, and the low-resolution NEXRAD field is obtained by successive dyadic coarsening as explained in the text.

per pixel [see *Gonzalez and Woods, 2008*]. In the studied storm images of Figure 5, using SNR = 11.0 dB, as the consequence of fusion, the average information content of the final high-resolution fused product was increased approximately by 33% compared to the original NEXRAD image. Apart from different probable sources of false detection (e.g., ground clutter) which need to be treated separately, due to the inherent differences in the way that the two sensors interrogate the vertical profile of the atmosphere, in fusion of precipitation snapshots there might be some spots that a sensor detects as rainy areas where the other sensor is blind. Therefore, naturally the wetted area (positive part) of the fused products is greater or equal compared to the individual original measurements. Consequently, a major part of this entropy increase can be due to the growth of the wetted area as a natural result of the fusion process.

[20] We can also compare matrix norms of the processed (fused) and unprocessed images (original) to quantify how this fusion process may affect the overall second-order marginal statistics of the fields. As a result of the fusion process, we do expect that the final processed image possesses a 2-norm measure which falls within the range of the 2-norm of the original input images; i.e., TRMM-PR and NEXRAD data. To this end, the Frobenius norm (F -norm) of the processed and unprocessed reflectivity images $\mathcal{Z} \in \mathbb{R}^{M \times N}$ at two scales of interest are computed and compared:

$$\|\mathcal{Z}\|_F = \sqrt{\sum_{i=1}^M \sum_{j=1}^N |z_{i,j}|^2} = \sqrt{\text{tr}(\mathcal{Z}^T \mathcal{Z})}.$$

Table 1 shows the normalized calculated F -norms for the originally observed TRMM-PR, NEXRAD images and their posterior estimates as a result of the fusion process. It turns out that the TRMM observation in this case has larger F -norm (energy) than the NEXRAD image at the same scale and the energy of the posteriori estimate lies within the range of the computed F -norms of the original observations.

[21] As we explained before, isolated singularities and jump discontinuities are typical features in the rainfall images which manifest themselves as a tail in the probability distributions of the wavelet coefficients, considerably thicker than the Gaussian case. These local extreme values often contain a large portion of the signal energy and may play a very important role in the hydrogeomorphic consequences

and risk assessment of extreme storm events. In standard linear Gaussian filtering schemes such as SRE, this important property is not explicitly addressed and often the results of these methods will lead to an overly smooth representation of the process [e.g., *Wainwright et al., 2001*]. In section 5, a new nonlinear adaptive estimation method is presented which exploits the regular statistical signature of the rainfall images in the wavelet domain to properly capture the non-Gaussian and singular structure of precipitation data. We further explain quantitatively the major advantages of the newly proposed method as compared to the standard linear Gaussian estimation in an illustrative synthetic one-dimensional example.

4. Precipitation Probability Model in the Wavelet Domain

[22] As described before, rainfall reflectivity images exhibit a highly non-Gaussian structure. This structure has a remarkably distinct and regular signature in the wavelet domain. We explain in this section how this regularity permits an adaptive exploitation of the common linear filtering methods for proper multiscale rainfall estimation. Note that in this study we have used a redundant orthogonal wavelet representation, the so-called “stationary wavelet transform” [*Nason and Silverman, 1995; Coifman and Donoho, 1995*], for multiscale subband encoding of the reflectivity images. This class of wavelet representation is shift invariant and produces an overcomplete set of nearly alias-free subband information which leads to a superior filtering performance and more accurate subband parameterization.

[23] It has been theoretically shown that a variety of well-known classes of elliptically symmetric thick tail probability distributions such as the t -distribution, symmetric gamma, double exponential and α -stable can be reproduced using a scale mixture of Gaussian random variables [*Andrews and Mallows, 1974; West, 1987; Wainwright and Simoncelli, 2000; Wainwright et al., 2001*]. Formally, a Gaussian-scale mixture (GSM) random vector \mathbf{d} is defined as the product of a zero mean Gaussian random vector $\mathbf{u} \sim \mathcal{N}(0, \Sigma_{\mathbf{u}})$, and a positive scalar multiplier random variable z ,

$$\mathbf{d} \stackrel{d}{=} \sqrt{z} \mathbf{u}, \quad (6)$$

where $\stackrel{d}{=}$ stands for equality in distributions. For instance, choosing z as an exponential random variable, the GSM will produce a random vector from the family of Laplace densities. However, in the case of the generalized Gaussian density with $0 < \alpha < 1$, the range typically observed for rainfall reflectivity images [*Ebtehaj and Foufoula-Georgiou, 2011*], there is no closed form expression for the distribution of the multiplier [*Wainwright et al., 2001*]. By construction, $\Sigma_{\mathbf{d}} = \mathbb{E}[z] \Sigma_{\mathbf{u}}$, hence without loss of generality assuming $\mathbb{E}[z] = 1$, the entire covariance structure of the GSM can be controlled by the covariance of the \mathbf{u} . Obviously, higher-order moments of the multiplier z will determine the shape or heavy tail properties of the GSM random vector.

[24] A key feature of the GSM is that the conditional density $p(\mathbf{d}|z)$ is a zero mean Gaussian process with covariance $z \Sigma_{\mathbf{u}}$. Integrating the joint density with respect to z and

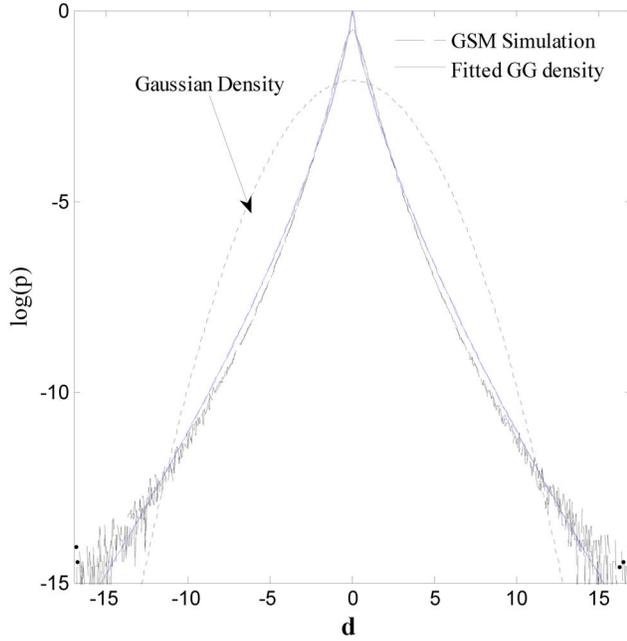


Figure 6. The broken line is the standardized marginal histogram (unit standard deviation) of a generated Gaussian-scale mixture (GSM) random variable using a lognormal multiplier with $\sigma_z \approx 1.2$ (see equation (12)), and the solid line shows the fitted generalized Gaussian (GG) density with $\alpha \approx 0.7$. The inverted parabola shows the Gaussian distribution in log probability for comparison purposes.

using Bayes' theorem, the GSM multivariate density can be characterized as [Wainwright *et al.*, 2001]:

$$p_D(\mathbf{d}) = \int_0^\infty p(\mathbf{d}|z)p_Z(z)dz = \int_0^\infty \frac{\exp\left(-\frac{\mathbf{d}^T(z\Sigma_{\mathbf{u}})^{-1}\mathbf{d}}{2}\right)}{(2\pi)^{N/2}(\det|z\Sigma_{\mathbf{u}}|)^{1/2}}p(z)dz.$$

[25] A finite dimensional version of this representation is reminiscent of the Gaussian kernel density estimation paradigm in the statistical literature. Indeed in discrete space, the probability mass function of the GSM is a convex weighted sum of different rescaled versions of some zero mean Gaussian kernels, where given a set of observations, the weights and bandwidths (i.e., $p_Z(z)$ and z) of the kernels need to be estimated in an optimal sense.

5. GSM Optimal Estimation in the Wavelet Domain

5.1. Basics of the Framework

[26] Given a set of independent observations $\mathbf{y} \in \mathbb{R}^N$ of a GSM random vector $\mathbf{d} \in \mathbb{R}^N$ ($\mathbf{d} = \sqrt{z}\mathbf{u}$) in a Gaussian noise:

$$\mathbf{y} = \mathbf{d} + \mathbf{v}, \quad (7)$$

where $\mathbf{v} \sim \mathcal{N}(0, \Sigma_{\mathbf{v}})$ and assuming $\mathbb{E}[z] = 1$, without loss of generality, equations (6) and (7) result in:

$$\begin{aligned} \Sigma_{\mathbf{y}|z} &= z\Sigma_{\mathbf{u}} + \Sigma_{\mathbf{v}} \\ \Sigma_{\mathbf{y}} &= \Sigma_{\mathbf{u}} + \Sigma_{\mathbf{v}}. \end{aligned} \quad (8)$$

In this case, the likelihood function of the multivariate GSM density can be expressed as follows:

$$p(\mathbf{y}|z) = \frac{1}{(2\pi)^{N/2}(\det|z\Sigma_{\mathbf{u}} + \Sigma_{\mathbf{v}}|)^{1/2}} \exp\left(\frac{-\mathbf{y}^T(z\Sigma_{\mathbf{u}} + \Sigma_{\mathbf{v}})^{-1}\mathbf{y}}{2}\right).$$

[27] With no a priori assumption on $p_Z(z)$ and perfect whitening effect of the wavelet transform, *Strela* [2000] and *Strela et al.* [2000] derived the maximum likelihood (ML) estimator for the multiplier z . However, as explained previously, it has been found that the wavelet decomposition cannot completely decorrelate the rainfall reflectivity images and the wavelet coefficients are highly structured with a short range of spatial dependence (see Figures 2a, 2b, and 3). This implies that the assumption about the diagonality of the covariance matrix of the wavelet coefficients might not be a good assumption for modeling of spatial rainfall.

[28] Studying the heavy tail properties of these images in the wavelet domain, the lognormal density was found empirically as a suitable prior assumption about the distribution of the multiplier $p_Z(z)$ [Ebtahaj and Foufoula-Georgiou, 2011] (see Figure 6). Accordingly, given a set of independent observations $\mathbf{y} \in \mathbb{R}^N$ and the a priori information about the density of z , the maximum a posteriori estimate (MAP) of the multiplier can be computed in a Bayesian setting:

$$\hat{z}_{MAP} = \arg \max_z \{ \log p(z|\mathbf{y}) \} = \arg \max_z \{ \log p(\mathbf{y}|z) + \log p(z) \}.$$

[29] Assuming the multiplier has a lognormal density $z \sim \mathcal{LN}(\mu_z, \sigma_z)$ with $\mathbb{E}[z] = 1$, where μ_z and σ_z are the mean and variance of the $\log(z)$, in a more general case where $\Sigma_{\mathbf{u}}$ and $\Sigma_{\mathbf{v}}$ are not diagonal, solving the above optimization problem leads to the following nonlinear expression which needs to be solved numerically for \hat{z}_{MAP} [Portilla *et al.*, 2001, 2003],

$$f(\hat{z}_{MAP}) = \frac{\log \hat{z}_{MAP} + 3\sigma_z^2/2}{\hat{z}_{MAP}\sigma_z^2} + \frac{1}{2} \sum_{n=1}^N \frac{\hat{z}_{MAP} - \lambda_n^{-1}(v_n^2 - 1)}{(\hat{z}_{MAP} + \lambda_n^{-1})^2} = 0, \quad (9)$$

where v_n are the components of the vector $\mathcal{V} = Q^T S^{-1} \mathbf{y}$; S is the square root of the error covariance matrix (i.e., $SS^T = \Sigma_{\mathbf{v}}$) and (Λ, Q) contains the eigenvalues $\lambda_n \in \Lambda$ and eigenvectors of the positive semidefinite matrix $S^{-1} \Sigma_{\mathbf{u}} S^{-T}$. Different numerical methods such as bisection, secant, false position or Brent's method can be used to find the root of equation (9) (see Appendix A). We used a multidimensional bisection method [see also Portilla *et al.*, 2001], one of the simplest and most primitive ones which always guarantees the convergence to a root as long as the search begins in an interval where the function $f(\cdot)$ takes opposite sign on the boundaries.

[30] Note that solving equation (9) requires full characterization of the a priori lognormal density $p_Z(z)$ and its parameters (μ_z, σ_z) . The assumption corresponding to the unity of the expected value of the multiplier reduces the number of unknown parameters to one, as $\mu_z + \frac{\sigma_z^2}{2} = 0$, and therefore it suffices only to estimate the variance σ_z^2 as it is the only parameter appearing in equation (9). Given the variance of the noise $\sigma_{\mathbf{v}}^2$ and knowing that $\sigma_{\mathbf{u}}^2 = \sigma_{\mathbf{y}}^2 - \sigma_{\mathbf{v}}^2$, it

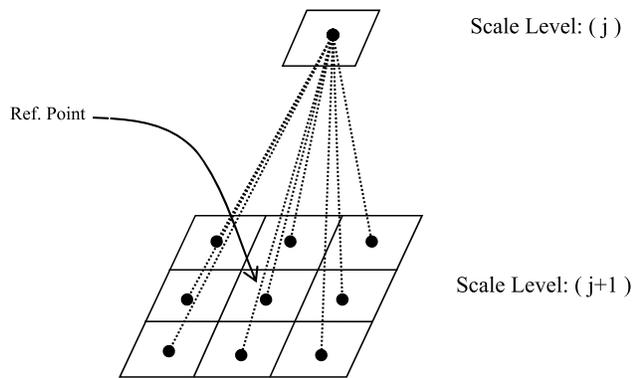


Figure 7. A general pyramidal neighborhood of size $N = 10$ for each individual wavelet subband where a cluster of 3×3 pixels (children) is connected to one pixel at the next coarser scale (parent). In this multiscale representation of a generalized neighborhood, scale-to-scale and short-range intrascale dependence of the wavelet coefficients can be explicitly captured in the local covariance matrix. The parent node information is only taken into account where coarser-scale subband data are available in the wavelet domain. In other words, for coarsest-scale subband information, structure of the generalized neighborhood reduces to a simple neighborhood of 3×3 pixels.

can be shown after some algebra that this parameter can be estimated as (see Appendix A):

$$\sigma_z^2 = \log(\mathbb{E}[z^2]) = \log\left(\frac{\mathbb{E}[y^4]/3 - 2\sigma_y^2\sigma_v^2 + \sigma_v^4}{(\sigma_y^2 - \sigma_v^2)^2}\right). \quad (10)$$

[31] Assuming that the wavelet coefficients of rainfall images $\mathbf{d} \in \mathbb{R}^N$ can be explained using a GSM model with a log prior density for the multiplier, filtering and optimal fusion of different sources of noisy measurements can be performed efficiently in the wavelet domain while characteristic heavy tail distributions and local singularities can be well captured.

[32] In the form of a general linear measurement equation $\mathbf{y} = \mathbf{C}\mathbf{d} + \mathbf{v}$, referring to the conditional Gaussian density of the GSM model and expressions in equations (1) and (2), the conditional expectation of the zero mean noisy wavelet coefficients and its covariance can be written as:

$$\hat{\mathbf{d}} = \mathbb{E}[\mathbf{d}|\mathbf{y}, z] = z\Sigma_{\mathbf{u}}\mathbf{C}^T[C(z\Sigma_{\mathbf{u}})\mathbf{C}^T + \Sigma_{\mathbf{v}}]^{-1}\mathbf{y}$$

$$\Sigma_{\hat{\mathbf{d}}} = z\Sigma_{\mathbf{u}} - z^2\Sigma_{\mathbf{u}}\mathbf{C}^T[C(z\Sigma_{\mathbf{u}})\mathbf{C}^T + \Sigma_{\mathbf{v}}]^{-1}\mathbf{C}\Sigma_{\mathbf{u}}.$$

[33] Assuming an unbiased system of measurement equations with identity measurement matrix $\mathbf{C} = \mathbf{I}$, given the MAP estimator of the multiplier, this conditional expectation can be simplified into:

$$\mathbb{E}[\mathbf{d}|\mathbf{y}, \hat{z}_{MAP}] = \hat{z}_{MAP}\Sigma_{\mathbf{u}}(\hat{z}_{MAP}\Sigma_{\mathbf{u}} + \Sigma_{\mathbf{v}})^{-1}\mathbf{y}. \quad (11)$$

[34] When applied to a local neighborhood of coefficients, this expression is indeed an adaptive local Wiener filter which in an individual subband (i.e., a particular scale of

interest) can be exploited to recover the “contaminated” wavelet coefficients in a Gaussian noise environment. Note that in this setting as the entire structure of the covariance matrix is incorporated, a correlated noise also can be used and there is no restriction on choosing only a white type of Gaussian noise (diagonal $\Sigma_{\mathbf{v}}$). Obviously, obtaining the filtered wavelet coefficients for each subband, a denoised version of the process of interest (e.g., rainfall reflectivity) can be recovered using the inverse wavelet transform, which is optimal both in the least squares and maximum likelihood sense in the wavelet domain.

5.2. Global Versus Local Filtering

[35] Implementation of the filter discussed above in the wavelet domain requires estimation of $\Sigma_{\mathbf{u}}$ for each subband, which can be obtained from equation (8) given the measurement error covariance $\Sigma_{\mathbf{v}}$. As the wavelet decomposition approximately whitens the precipitation fields, assuming a finite correlation length for the wavelet coefficients, this covariance can be estimated via characterization of the dependence of a local “neighborhood” of the wavelet coefficients. In high-dimensional problems, this local characterization not only makes the estimation process computationally more tractable but also leads to a superior estimation, in the sense that the local singular structures of interest (precipitation local extremes) can be better recovered from noisy observations [Strela, 2000; Portilla et al., 2001]. In effect, modulating the measurement covariance by the estimated multiplier, the significance of filtering is adaptively adjusted according to the local singular features of the field. For large multiplier values over the singular points of the process, the filtering is less significant and the filter accepts the observations close to the true values; however, when the multiplier modulates the $\Sigma_{\mathbf{u}}$ in the same order of magnitude as that of the noise covariance, the filter smooths out the observations by suppressing the noise.

[36] In general, a local neighborhood may include clusters of nearby wavelet coefficients from different subbands at multiple scales around a reference point. In this study, we use a pyramidal neighborhood of the wavelet coefficients which includes two clusters of the coefficients, each in an individual subband at two successive scales (see Figure 7). In this construction, the elements of a neighborhood of size $\sqrt{N} \times \sqrt{N}$ of the wavelet coefficients have to be stacked according to a fixed order into a vector form $\mathbf{y} \in \mathbb{R}^N$. Sliding the neighborhood over the entire subband of size $\mathbb{R}^{M \times L}$ in an overlapping manner, the sample covariance matrix $\Sigma_{\mathbf{u}} \in \mathbb{R}^{N \times N}$ can be estimated for each individual subband of large dimension as:

$$\Sigma_{\mathbf{u}} \approx \frac{\sum_{i=1}^{M \times L} (\mathbf{y}\mathbf{y}^T)_i}{M \times L} - \Sigma_{\mathbf{v}}.$$

[37] To resolve the block filtering boundary issues, each subband has been padded symmetrically with “mirror reflection” of itself around the boundaries. For implementation of the GSM-Wiener filter in equation (11), this covariance ($\Sigma_{\mathbf{u}}$) only needs to be estimated once for each subband. On the other hand, the multiplier has to be estimated locally according to equation (9) for every neighborhood location, while it slides over the entire subband domain. The

conditional expectation in equation (11) gives an estimate of the entire neighborhood elements, where only the central value needs to be kept as the posterior estimate for each point. This posteriori estimate of the central value is indeed a weighted average of all surrounding elements in the neighborhood while the weights are adaptively modulated by the estimated multiplier \hat{z}_{MAP} .

[38] Note that by construction the estimated $\Sigma_{\mathbf{u}}$ is always symmetric; however, it may not be positive semidefinite for high levels of noise. To ensure the positive semidefiniteness of $\Sigma_{\mathbf{u}}$, we first factorized the matrix using eigenvalue decomposition $\Sigma_{\mathbf{u}} = VDV^T$ and then only nonnegative eigenvalues $\{d_i\}_{i=1}^n$ were picked to reconstruct a positive semidefinite version of the estimated covariance matrix. Of course, if the leading eigenvalue becomes negative the subband information cannot be recovered at the assumed power of noise.

5.3. Synthetic One-Dimensional Implementation

[39] As previously demonstrated, one of the main advantages of the local GSM-Wiener filter is its adaptability to the local structure of the signal compared to its global standard Gaussian counterpart (i.e., SRE), leading thus to a superior estimation of non-Gaussian heavy tail precipitation fields with frequent isolated intense rainfall clusters. This potential achievement and verification may not be very clear while filtering out the measurement noise and fusing precipitation images especially when the true intensity values of the processes are not available. In this section, a synthetic study is conducted to show how this filter provides a superior framework to recover the true process from noisy observations of a one-dimensional multiscale process with non-Gaussian heavy tail marginals. For this purpose, analogous to the observed heavy tail multiscale structure of the rainfall fluctuations [Ebtehaj and Foufoula-Georgiou, 2011], a one-dimensional GSM process using a lognormal multiplier is simulated over a dyadic Markov tree. First, a multiscale nonstationary Gaussian process $u(s)$ is produced on a dyadic tree according to the coarse-to-fine scale dynamics of equation (3). The variance of the driving noise is tuned with a relevant geometrical decay rate from coarse-to-fine scales to reproduce an asymptotically dyadic self-similar process with $1/f$ spectrum. This process is multiplied elementwise by a sequence of random variables drawn from a lognormal density at different levels of the tree to produce a multiscale GSM process on a treelike structure,

$$\mathbf{u}(s) = A(s)\mathbf{u}(s\bar{\gamma}) + B(s)w(s)$$

$$\mathbf{d}(s) = \sqrt{z(s)}\mathbf{u}(s),$$

where $w(s) \sim \mathcal{N}(0, 1)$ and $B(s) = \mathcal{Z}^{-Hj(s)/2}$. Setting $\sigma_v^2 = 0$ in equation (10), observe that the kurtosis $\kappa(\cdot)$ of a simulated lognormal GSM can be solely determined by the variance of the multiplier z ,

$$\kappa[\mathbf{d}(s)] = 3 \exp[\sigma_z^2(s)]. \quad (12)$$

[40] As the marginal distribution of the lognormal GSM resembles the family of generalized Gaussian densities [Wainwright et al., 2001], this also implies that the shape of the equivalent GG density is only characterized by the

variance of the multiplier, knowing that the tail parameter (α) of a GG density can be uniquely estimated from the sample kurtosis $\kappa(\cdot) = \Gamma(\frac{1}{\alpha})\Gamma(\frac{3}{\alpha})/\Gamma^2(\frac{2}{\alpha})$ [see, e.g., Nadarajah, 2005].

[41] In this study, the sample path of the generated GSM signal is considered as the true values which have to be recovered, given a set of noisy observations $\mathbf{y}(s) = \mathbf{d}(s) + \mathbf{v}(s)$, where $\mathbf{v}(s) \sim \mathcal{N}(0, \Sigma_{\mathbf{v}(s)})$. For this particular case of one-dimensional simulation, the signal-to-noise ratio, was set on the order of 8 dB to generate the noisy measurements. To exploit the multiscale structure of the process, the generalized neighborhood includes a single reference point of the process and only a single parent node in the next coarser scale; i.e., $\Sigma_{\mathbf{u}}$ is a 2×2 matrix. This allows us to incorporate a local scale-to-scale correlation (see, e.g., Figure 2) and capture the parent and child dynamics, for improving the signal recovery. A realization of this synthetic simulation and the recovered signal, using the standard linear multiscale estimation (SRE) and the local GSM-Wiener filter is presented in Figure 8 for the seventh scale level on a dyadic tree with 2^7 leaf nodes. Qualitatively, the GSM-Wiener is outperforming the SRE method especially over the recovery of the peaks and singularities. Note that for both cases, the estimation process suppresses the noise over the regions where the signal is of low amplitude; however, the GSM-Wiener filter shows a better performance over the singular points. This can also be quantitatively evaluated in terms of some vector norms of error; i.e., $\|e\|_p$. A normalized measure $(\|e_S\|_p - \|e_G\|_p)/\|e_S\|_p$ is defined, where $\|e_S\|_p$ and $\|e_G\|_p$ are the p -norms of the error for the recovered signal using SRE and GSM-Wiener filters, respectively. For instance, in this particular case, assuming $p = 2$, the 2-norm (energy) of error is improved about 20 percent while this gain rose to about 45 percent for the infinity norm or the maximum absolute value of the error vector. This significant improvement implies that GSM-Wiener filtering can outperform standard Gaussian methods on the recovery of the commonly observed types of singularities in the precipitation fields, while also keeping the other common norms of the error even lower than the standard linear estimation algorithms.

[42] Analogous to the explained one-dimensional case, it is expected that estimating multisensor precipitation data using standard linear Gaussian filtering methods such as the SRE, may result in not properly capturing important singular features of the fields which can potentially be of great hydrometeorological importance.

6. GSM Multisensor Fusion of Precipitation Data

6.1. Conceptual Development

[43] In this section we describe how the explained GSM-Wiener filter can be employed for optimal estimation and fusion of the precipitation reflectivity images, given different sources of noisy observations. It has long been recognized that all of the active and passive precipitation sensors have their own specific measurement error structure [see, e.g., Wang and Wolff, 2009]. As explained previously, in recent decades significant effort has been devoted to error characterization of the remotely sensed precipitation products. Typically, using an appropriate Z - R relationship, this involves statistical comparison of the remotely sensed data with a

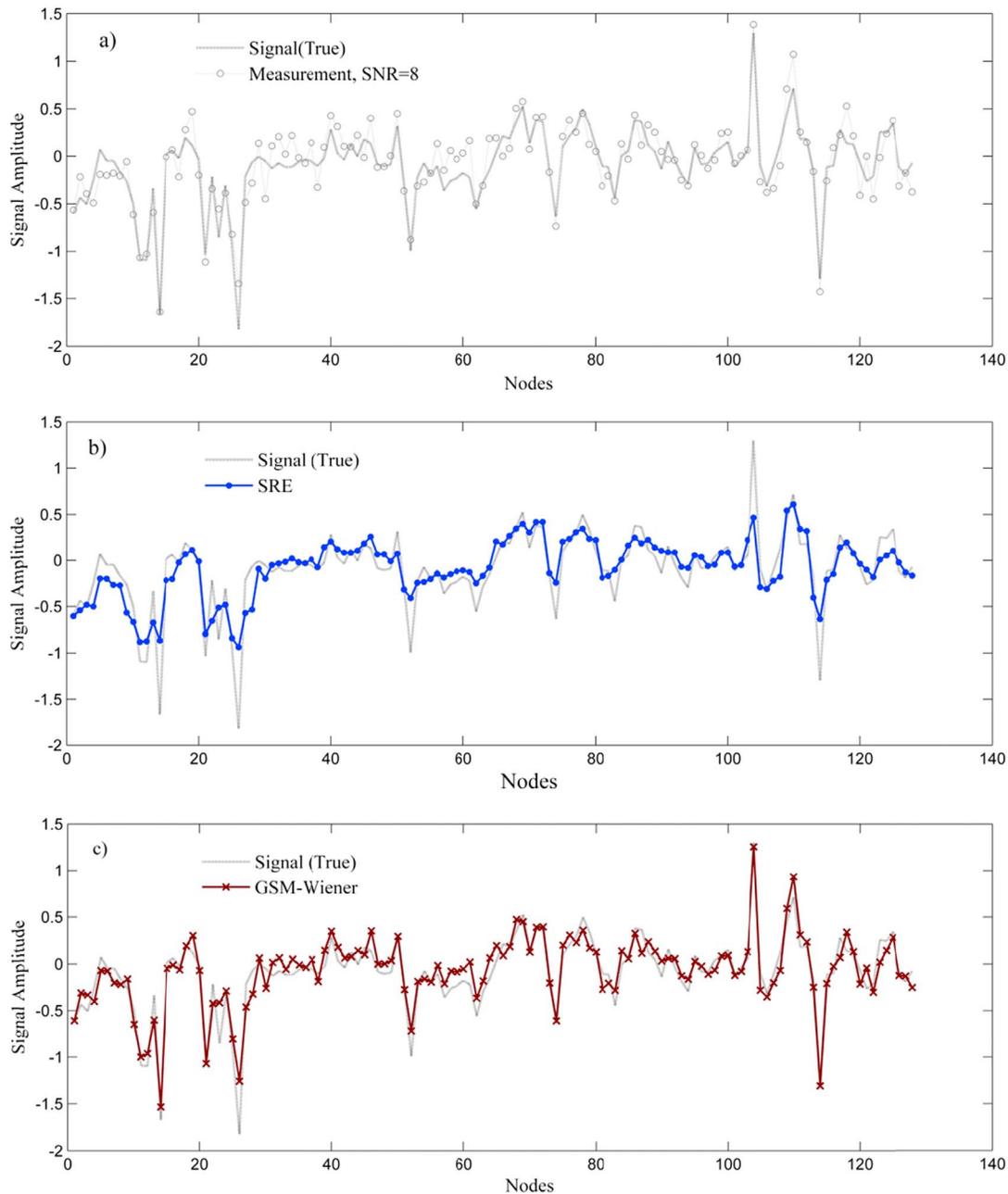


Figure 8. (a) A synthetic signal of a one-dimensional multiscale GSM (using a lognormal multiplier) and its noisy measurement. The sample path is generated with $A(s) = 0.9$ and $H = 0.7$. The variance of the log multiplier is set to reproduce a sample path with a marginal histogram similar to a generalized Gaussian distribution with $\alpha = 0.7$ ($\sigma_z \approx 1.20$) found in the rainfall NEXRAD reflectivity images. The signal recovery using the (b) SRE and (c) GSM-Wiener shows the superior performance of the GSM filter, especially over the singular points and extreme fluctuations.

reliable set of rain gauge observations either in the reflectivity or rainfall intensity domain. This comparison is not very straightforward as these sources of information often give precipitation estimates at different spatiotemporal scales. For instance, it has been concluded that direct comparison of the real time rain gauge data as a pointwise representation of the highly irregular rainfall process with a remotely sensed areally averaged rainfall products may not lead to a conclu-

sive statistical characterization of the measurement error [e.g., Ciach and Krajewski, 1999; Wang and Wolff, 2009]. Indeed, for a proper characterization of the measurement error norms, the spatiotemporal scaling gaps between these sources of information need to be properly addressed. Although this characterization is not explicitly addressed in this study, it is emphasized that the proposed fusion methodology requires this information as an input and the overall quality of the

fusion process highly depends on this error characterization. Note that the proposed fusion methodology is just a filtering method to estimate the conditional expectation of the unbiased noisy precipitation data. Therefore, it is also assumed that there is no systematic bias in the observation instruments and any sort of bias adjustment has to be performed prior to applying the presented fusion algorithm.

[44] To address the scaling issues involved in precipitation multisensor estimation, the new proposed fusion methodology possesses a multiscale filtering structure. Practical implementation of this methodology does not necessarily require a stochastic or physically based precipitation model to fill the scaling gaps between the available high-resolution (NEXRAD) and low-resolution (TRMM-PR) precipitation products. Indeed, as the estimation process is performed in a pyramidal data structure in the wavelet domain, in the presented case study, the high-resolution (i.e., <4 km) information of the final fused product would be solely based on the rain gauge corrected NEXRAD wavelet coefficients. However, in the scales where the TRMM data are also available; i.e., ≥ 4 km, the fusion process exploits all sources of information (see Figure 9).

[45] Basically, by comparing different sources of rainfall measurements, three different error covariance matrices need to be characterized for proper implementation of the proposed fusion technique: (1) Σ_{v1} , NEXRAD versus rain gauges; (2) Σ_{v2} , TRMM-PR versus rain gauges; and (3) Σ_{v3} , NEXRAD versus TRMM-PR. Hereafter, the error covariances Σ_{v1} and Σ_{v2} are called “rain gauge adjustment error.” Note that although the results of the fusion process will be more complete using all of the error covariance information, lack of knowledge about any of these error matrices is not prohibitive for practical implementation of the introduced methodology. This becomes more clear as we proceed to elaborate the method in detail.

[46] The proposed GSM multisensor multiscale methodology consists of two major steps namely “*Vertical Filtering*” and “*Lateral Projection*” (see Figure 9). In the vertical filtering phase, incorporation of the measurement errors is considered as a filtering problem of the sort $\mathbf{y} = \mathbf{x} + \mathbf{v}$, where \mathbf{y} denotes the remotely sensed precipitation observation and \mathbf{x} is the true precipitation process which is corrupted by a Gaussian noise \mathbf{v} . Substituting Σ_{v1} and Σ_{v2} as error covariance matrices in equation (11), the rain gauge adjustment errors are first filtered out from the wavelet high-pass subbands of the TRMM and NEXRAD data, independently. Afterwards, in the scales where both the TRMM and NEXRAD data become available (i.e., ≥ 4 km), given the characterized error covariance Σ_{v3} , the rain gauge corrected subband information of the TRMM images can be laterally projected at the same scales onto the subspace spanned by the rain gauge corrected subbands of the NEXRAD measurements. For the scales, where the TRMM-PR data are not available (i.e., <4 km) the high-pass subband information of the fused product will be solely obtained from the rain gauge corrected NEXRAD wavelet coefficients. At the last step, the error corrected wavelet coefficients at all scales are used to reconstruct the final fused product using the inverse wavelet transform (see Figure 9).

[47] Theoretically speaking, using equation (11) in lateral projection phase, we can keep decomposing the observations (i.e., TRMM and NEXRAD) into multiple levels until we end

up with a single valued low-pass subband and performing the GSM fusion on the high-pass coefficients over all the scales where the data from both sensors are available. This procedure might be computationally expensive and it seems reasonable to perform finite levels of the wavelet decomposition for fusion and denoising purposes, knowing that high-frequency noisy features of a signal are typically captured at the first levels of wavelet high-pass coefficients. Consequently, at a certain scale level, we eventually need to project the nonzero mean low-pass coefficients of the low-resolution products (TRMM) onto the similar subspace (same scale) spanned by the high-resolution data (NEXRAD). As the non-Gaussian features of the signals are typically captured in high-pass subbands in the wavelet domain [see *Ebtehaj and Foufoula-Georgiou*, 2011], the low-pass coefficients can be fused using a conventional least squares formalism as expressed in equations (1) and (2),

$$\mathbb{E}[\mathbf{x}_l|\mathbf{y}_l] \cong \mathbf{m}_{\mathbf{x}_l} + \Sigma_{\mathbf{x}_l}(\Sigma_{\mathbf{x}_l} + \Sigma_{\mathbf{v}_3})^{-1}(\mathbf{y}_l - \mathbf{m}_{\mathbf{y}_l}), \quad (13)$$

where $\mathbf{y}_l, \mathbf{m}_{\mathbf{y}_l}$ denote the NEXRAD low-pass coefficients and their mean in a local spatial neighborhood; and $\mathbf{m}_{\mathbf{x}_l}, \Sigma_{\mathbf{x}_l}$ are the average and covariance of the TRMM low-pass coefficients in that neighborhood, respectively. Obviously, as there is no lower-scale subband information available while fusing low-pass coefficients, the neighborhoods in this case just include a cluster of coefficients in a single subband and there is no information of parent nodes encoded in the covariance matrices of equation (13).

[48] Besides the input error covariances, a set of two parameters need to be determined in the presented fusion methodology, including: (1) the structure and size of the generalized neighborhood and (2) the levels of the wavelet decomposition. In this work, we did not perform a quantitative assessment of different choices of the parameters on the fusion results and simply chose values we found empirically to perform well.

[49] As explained previously, the correlation of the rainfall wavelet coefficients almost vanishes over a neighborhood of size 3 to 5 pixels (km) for the first level of subband coefficients. In general, it is found that increasing the size of the neighborhood (i.e., enlarging the estimated correlation domain) gives rise to a smoother and more blurred fusion product. On the other hand, smaller spatial neighborhoods generally generate a product which contains sharper and more detailed structure of the high-intensity rain cells.

[50] It is also observed that over the decomposition levels 2–4 (i.e., 4–16 km) the noise (observational error) can be well captured in the wavelet domain and the results of the fusion are satisfactory. Of course, for higher levels of decomposition the low-pass fusion takes place at larger scales which means that more detailed features of the fused product will be obtained from the higher-resolution data (e.g., NEXRAD) and incorporation of small-scale information of the low-resolution sensor (e.g., TRMM-PR) would be less significant.

6.2. A Case Study on Precipitation Data

[51] The TRMM-PR and NEXRAD coincidental reflectivity image of a storm on 28 June 1998 over the HSTN site, used for the SRE implementation (see section 3.2), is also

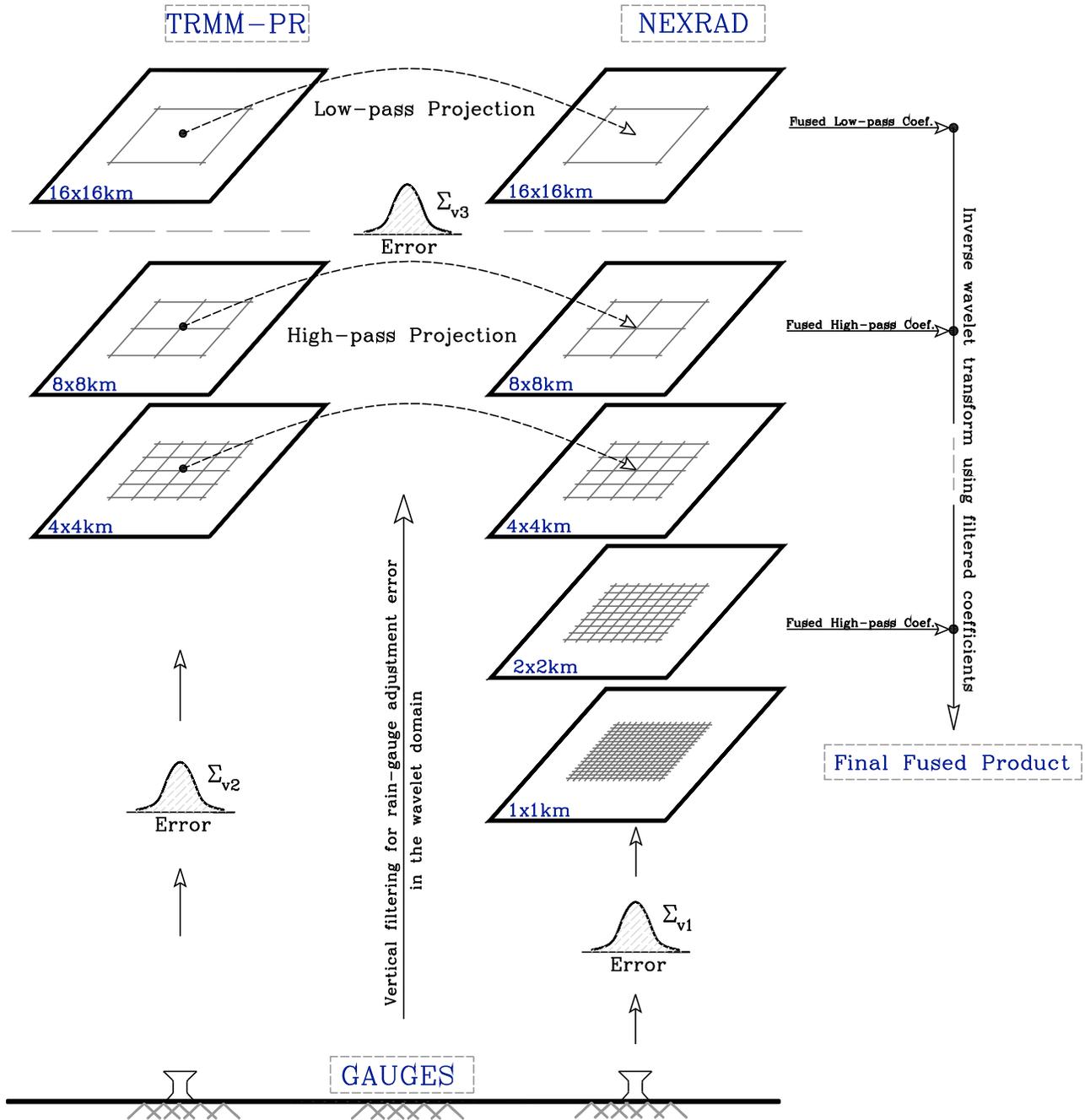


Figure 9. Schematic of the proposed GSM fusion technique in the wavelet domain. In the “vertical filtering” stage, filtering for rain gauge adjustment error is performed on each available unbiased source of remotely sensed precipitation reflectivity image, i.e., TRMM-PR and NEXRAD, independently. Then, the rain gauge corrected data can be fused in the wavelet domain in the “lateral projection” phase as explained in the text. Small-scale (<4 km) subband information of the final fused product is extracted from the rain gauge corrected NEXRAD data, while the larger-scale subbands contain information from all sources of the available rainfall data.

selected in this part for comparison purposes. In this study we have not explicitly characterized the error structure and only implicitly used a fraction of the rainfall power to produce observational error covariance. Assuming SNR = 13.0 dB for construction of the diagonal rain gauge adjustment error covariance Σ_{v1} , Figure 10 shows how incorporation of this error may be reflected in a rainfall reflectivity image. A

cluster of wavelet coefficients including a neighborhood of size 3×3 pixels and a single parent node right above the central point at the next coarser scale is considered for this filtering. Obviously, as we defined the rain gauge data fusion in the context of a denoising problem, the result would be a field smoother than the original NEXRAD image. However, using the presented adaptive method this

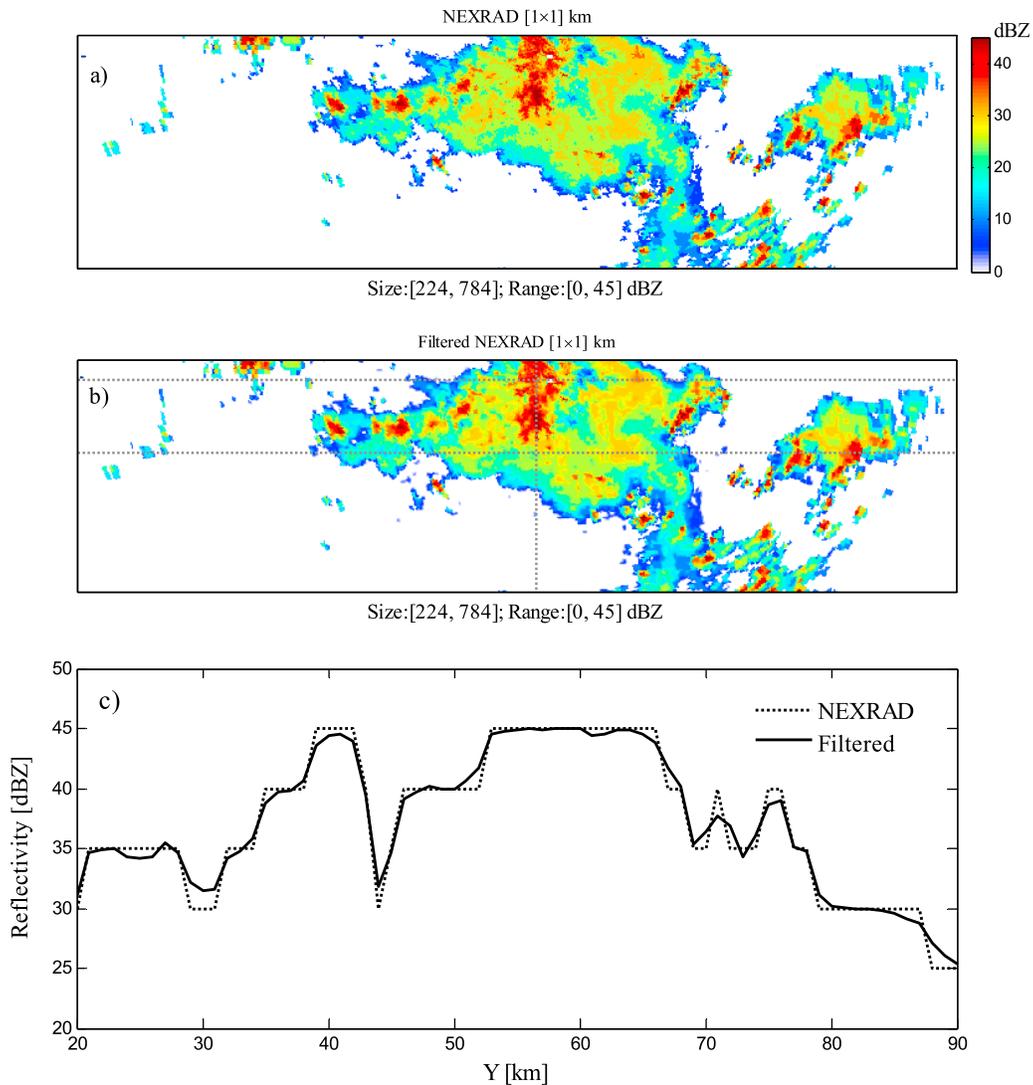


Figure 10. (a) Original NEXRAD reflectivity image over the HSTN site on 28 June 1998 at 18:13:00 UTC and (b) the rain gauge corrected image assuming $\text{SNR} = 13.0$ dB. (c) A vertical transect of the image, bounded by two horizontal dashed lines in Figure 10b, shows a one-dimensional representation of the rain gauge filtering effects on the reflectivity image.

filtering would be done at minimum expense corresponding to the removal of the frequent local extremes of interest.

[52] Figure 11 compares qualitatively the performance of the GSM-Wiener fusion (using a neighborhood of size 3×3 plus a single parent) and the standard SRE method with the same order of total error covariance. First, assuming $\text{SNR} = 24.0$ dB, filtering of the rain gauge adjustment error is performed independently on the TRMM and NEXRAD reflectivity images. Then, at scale 4×4 km with $\text{SNR} = 16.0$ dB, the rain gauge corrected TRMM data are projected onto the subspace of the low-pass coefficients of the rain gauge corrected NEXRAD data, using equation (13). Qualitative comparison of Figures 11a and 11b shows that the GSM fusion algorithm recovers more small-scale features and produces a fused rainfall reflectivity image with a more detailed structure. A one-dimensional transect in Figure 11c better demonstrates the performance of the proposed GSM fusion compared to the standard SRE method. It

seems that due to the local structure of the GSM fusion, this method also takes into account more information from the TRMM observations while preserving the storm structure and the local high-intensity values of the rain cells. For instance, in the left-hand side of the fused images it is clear that the SRE method significantly filtered out the TRMM observations when the NEXRAD data were absent; however, the GSM fusion method did incorporate the TRMM information in that part of the image.

[53] In the SRE fusion example we did not consider the rain gauge adjustment error. Therefore, to quantitatively compare the GSM and SRE fusion methodologies, it is more realistic to neglect the incorporation of rain gauge adjustment error in the GSM fusion and compare the fused products by considering only the same measurement error between NEXRAD and TRMM-PR; i.e., $\text{SNR} = 11.0$ dB. Table 2 shows the entropy, F -norm and statistical measures of the associated rain rates of the two fused images. To

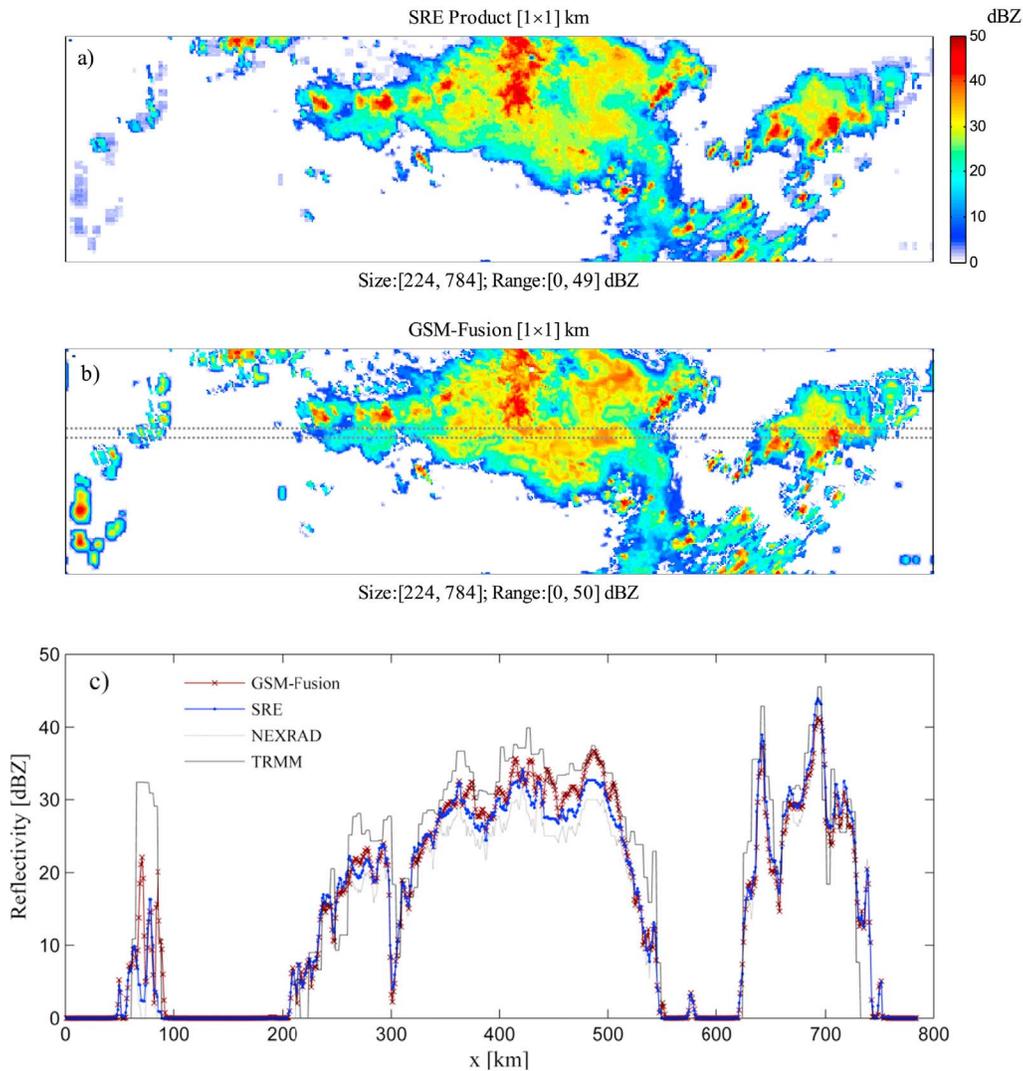


Figure 11. (a) SRE fusion of the NEXRAD and TRMM-PR reflectivity snapshots over the HSTN site on 28 June 1998 at 18:13:00 UTC, assuming SNR = 11.0 dB. (b) GSM fusion of the same reflectivity snapshots, assuming SNR = 24 dB for rain gauge errors ($\Sigma_{v,i}, i = 1, 2$) and SNR = 16 dB for $\Sigma_{v,3}$. Using the same order of error variance, the GSM fusion extracts a more detailed structure of the storm and incorporates more information from the TRMM data. (c) One-dimensional representation of the observed and estimated rainfall reflectivity averaged over the displayed band, delineated by the dashed lines in Figure 11b, compares qualitatively the performance of the employed fusion methods. For quantitative comparison, please refer to Table 2.

compute the rain rate, the commonly used $Z-R$ relationship of $Z = 300R^{1.4}$ has been used for all reflectivity images. In this case, the maximum entropy is obtained for the SRE product while it shows a smaller F -norm value compared to the GSM-fused product. This is good news in the sense that while the GSM product has larger energy and spatial variability, it exhibits a smaller uncertainty and more regularity in terms of the entropy measure. Recall that the entropy may be interpreted as the spontaneous behavior of the system state and hence a larger entropy indicates a less predictable behavior of the system. The rain intensities also denote that, apparently due to the local and adaptive structure of the GSM fusion, the TRMM-PR observations have been incorporated more significantly in the final fused product, as

Table 2. Statistical Measures of the Fusion Products (1×1 km) by the GSM-Wiener and SRE Methodologies Compared at the Same Order of Measurement Error^a

	TRMM-PR	NEXRAD	GSM Fusion	SRE Fusion
Normalized entropy	0.74	0.75	0.88	1
Normalized F norm	1	0.78	0.89	0.84
Mean rain rate (mm/h)	1.85	0.78	1.16	1.02
Maximum rain rate (mm/h)	60.06	27.86	55.59	48.45

^aEntropy and Frobenius norms are normalized by their respective maximum values.

the GSM product shows a larger mean and larger maximum values compared to the SRE method (see Figure 11c). Knowing that the wetted area in the studied storm snapshot is about $120 \times 10^3 \text{ km}^2$, the difference between the estimated rain budget from the original and fused products seems significant which denotes the importance of the precipitation fusion. The precipitation products that we have used in this study might not be the best from an operational standpoint and to substantiate more the practical benefits of the proposed precipitation fusion methodology, these findings need to be further investigated and solidified by more detailed analysis of other storm cases for which “ground truth” highly accurate rain gauge data are available.

7. Conclusion

[54] A new method is presented which allows us to integrate multiscale multiplatform precipitation measurements to provide a posteriori estimates of spatial rainfall. This method exploits the multiscale representation of precipitation images in the wavelet domain and specifically a particular class of Gaussian-scale mixture (GSM) distributions as the probability model of the wavelet high-pass coefficients of rainfall data. This algorithm is structured in a way that it can address the non-Gaussian statistics and capture the extreme intensity values of the precipitation fields in a Gaussian noise environment with a superior performance and many other advantages compared to the commonly used standard linear Gaussian fusion techniques. Exploiting the decorrelating effects of the wavelet decomposition, the posteriori estimates rely on the local multiscale spatial covariance structure of the rainfall fields in the wavelet domain. Therefore, as the wavelet decomposition converts a strongly correlated field into a set of weakly correlated subbands, in lieu of estimating the entire covariance of the original field in the spatial domain, a local representation of the covariance is estimated in each subband and used for optimal estimation and fusion. This makes the problem computationally more efficient while allowing to capture the non-Gaussian marginal statistics of these fields. Depending on the typical rainfall domain size for a single NEXRAD station and without any special optimization in coding style, running the MATLAB code of the algorithm on an Intel(R)-i7CPU with 2.80 GHz clock rate, takes in the order of less than 5 min.

[55] Using the developed fusion methodology, we can obtain a posteriori estimates of rainfall images as long as ground-based and spaceborne precipitation data are coincidentally available. Therefore, according to the revisiting time of the TRMM satellite or other sources of spaceborne precipitation data, we could update the satellite reflectivity images all over the contiguous United States using the national NEXRAD mosaic images. The fused rainfall product in this sense can be used in data assimilation systems to improve forecasts at the local and regional scales. Applicability of the developed methodology for obtaining posteriori estimates of spatial rainfall fields, given TRMM Microwave Imager (TMI) rainfall information, seems feasible and might be of great interest for future development of this work. It is worth nothing that, this fusion methodology can also be applied to rainfall data at different time scales (e.g., daily or monthly data), as long as the error covariance is properly

determined for that specific time scale. However, for larger-scale cumulative precipitation data the non-Gaussian signature is naturally less significant.

[56] Due to the local structure of the proposed estimation or fusion methodology, the range dependence effects of the sensors can also be easily incorporated in the error covariance matrices. This might be of great interest for estimation and fusion of the orbital satellite products produced by cross track instruments, where the measurement error shows a systematic dependence on the distance of the detected precipitation with respect to the centerline of the swath (e.g., Advanced Microwave Sounding Unit). Furthermore, in a broader context, applicability of the GSM probability model for data assimilation of non-Gaussian large-scale geophysical processes (e.g., soil moisture, atmospheric state variables) could also be of great interest for future research. Obviously, developing efficient algorithms in this context can significantly improve the shortcomings of the conventional Gaussian-based assimilation methods with respect to the typically observed extremes and singularities of interest in natural processes.

Appendix A: Details of Some Derivations

A1. Maximum a Posteriori Estimate of the Lognormal Multiplier

[57] Recall that the Bayesian maximum a posteriori (MAP) estimator of z is defined as:

$$\hat{z}_{MAP} = \arg \max_z \{ \log [p_{Z|Y}(z|\mathbf{y})] \} = \arg \max_z \{ \log [p_{Y|Z}(\mathbf{y}|z)] + \log [p_Z(z)] \} \quad (\text{A1})$$

where this estimator is equivalent to the maximum likelihood (ML) estimator when there is no informative assumption (i.e., uniform density) with respect to the a priori term $\log [p_Z(z)]$. Knowing that $p_{Y|Z}(\mathbf{y}|z)$ is Gaussian with covariance $\Sigma_{\mathbf{y}|z} = z\Sigma_{\mathbf{u}} + \Sigma_{\mathbf{v}}$, first let us focus on the term $\log [p_{Y|Z}(\mathbf{y}|z)]$ which leads to the derivation of the ML estimator; and then by incorporating the a priori term, the MAP estimator in equation (9) will be derived.

[58] For a noisy GSM observation vector $\mathbf{y} \in \mathbb{R}^N$, we have:

$$p_{Y|Z}(\mathbf{y}|z) = \frac{1}{(2\pi)^{N/2} (\det |\Sigma_{\mathbf{y}|z}|)^{1/2}} \exp \left(\frac{-\mathbf{y}^T (\Sigma_{\mathbf{y}|z})^{-1} \mathbf{y}}{2} \right). \quad (\text{A2})$$

[59] Therefore, the log likelihood function can be written as:

$$\log [p_{Y|Z}(\mathbf{y}|z)] = -\frac{1}{2} \mathbf{y}^T (\Sigma_{\mathbf{y}|z})^{-1} \mathbf{y} - \frac{1}{2} \log [\det |\Sigma_{\mathbf{y}|z}|] + C, \quad (\text{A3})$$

where C is a constant independent of z . The covariance is a positive semidefinite matrix, hence $\Sigma_{\mathbf{y}|z}$ can be diagonalized as follows:

$$\begin{aligned} \Sigma_{\mathbf{y}|z} &= z\Sigma_{\mathbf{u}} + \Sigma_{\mathbf{v}} \\ &= z\Sigma_{\mathbf{u}} + SS^T \\ &= S[S^{-1}(z\Sigma_{\mathbf{u}})S^{-T} + I]S^T, \end{aligned} \quad (\text{A4})$$

where S is the square root of $\Sigma_v = SS^T$ which can be computed using Cholesky or eigenvalue decomposition. Note that $S^{-1}\Sigma_u S^{-T}$ is also a symmetric positive semidefinite matrix, which can be diagonalized by an eigenvalue decomposition (i.e., spectral factorization) as $S^{-1}\Sigma_u S^{-T} = Q\Lambda Q^T$, where $\{Q, \Lambda\}$ are matrices containing orthogonal eigenvectors $QQ^T = I$ and positive eigenvalues $\lambda_n \in \Lambda$, respectively. Therefore, diagonalization in equation (A4) can be written as:

$$\Sigma_{y|z} = SQ(z\Lambda + I)Q^T S^T. \quad (\text{A5})$$

[60] Using this diagonalized version of the covariance matrix, equation (A3) can be further expanded as follows:

$$\begin{aligned} \log[p_{Y|Z}(y|z)] &= -\frac{1}{2}y^T(SQ(z\Lambda + I)Q^T S^T)^{-1}y \\ &\quad -\frac{1}{2}\log[\det|SQ(z\Lambda + I)Q^T S^T|] + C \\ &= -\frac{1}{2}(Q^T S^{-1}y)^T(z\Lambda + I)^{-1}(Q^T S^{-1}y) \\ &\quad -\frac{1}{2}\log[\det|SQ(z\Lambda + I)Q^T S^T|] + C \\ &= -\frac{1}{2}\mathcal{V}^T(z\Lambda + I)^{-1}\mathcal{V} - \frac{1}{2}\log[\det|z\Lambda + I|] + C', \end{aligned} \quad (\text{A6})$$

where the vector $\mathcal{V} = Q^T S^{-1}y$. Note that $z\Lambda + I$ is a diagonal matrix whose determinant is equal to the multiplication of its diagonal elements $\{z\lambda_n + 1\}_{n=1}^N$. Therefore, taking derivative of equation (A6) with respect to z , we have:

$$\begin{aligned} \frac{\partial \log[p_{Y|Z}(y|z)]}{\partial z} &= \frac{1}{2}\mathcal{V}^T(\Lambda(z\Lambda + I)^{-2})\mathcal{V} - \frac{1}{2}\sum_{n=1}^N \frac{\lambda_n}{z\lambda_n + 1} \\ &= \frac{1}{2}\sum_{n=1}^N \frac{\lambda_n v_n^2}{(z\lambda_n + 1)^2} - \frac{1}{2}\sum_{n=1}^N \frac{\lambda_n}{z\lambda_n + 1} \\ &= \frac{1}{2}\sum_{n=1}^N \frac{\lambda_n^{-1}(v_n^2 - 1) - z}{(z + \lambda_n^{-1})^2}. \end{aligned} \quad (\text{A7})$$

[61] Note that assuming a noninformative density for the multiplier, setting equation (A7) equal to zero, the root gives the maximum likelihood estimator of z .

[62] Assuming a lognormal density $p_Z(z; \mu_z, \sigma_z) = \frac{1}{z\sqrt{2\pi\sigma_z}} \exp(-\frac{(\log z - \mu_z)^2}{2\sigma_z^2})$, the derivative of the log likelihood is then given by:

$$\log[p_Z(z)] = -\frac{(\log z - \mu_z)^2}{2\sigma_z^2} - \log(z) + C \quad (\text{A8})$$

$$\frac{\partial \log[p_Z(z)]}{\partial z} = \frac{-\log z + \mu_z - \sigma_z^2}{z\sigma_z^2}. \quad (\text{A9})$$

[63] Then combining equations (A7) and (A9), leads to the derivation of equation (9):

$$\frac{\log \hat{z}_{MAP} + \frac{3}{2}\sigma_z^2}{\hat{z}_{MAP}\sigma_z^2} + \frac{1}{2}\sum_{n=1}^N \frac{\hat{z}_{MAP} - \lambda_n^{-1}(v_n^2 - 1)}{(\hat{z}_{MAP} + \lambda_n^{-1})^2} = 0. \quad (\text{A10})$$

[64] From a practical point of view, the $\log \hat{z}_{MAP}$ term in equation (A10) can dominate the magnitude of the other terms and ensures the convergence of the bisection method as there exists a range of initial values for \hat{z}_{MAP} , that the left-hand side of equation (A10) can take opposite signs on the boundaries.

A2. Parameters of the MAP Estimator

[65] Knowing that $y = \sqrt{z}u + v$ and assuming $\mathbb{E}[z] = 1$, we have:

$$\mathbb{E}[y^4] = \mathbb{E}[z^2]\mathbb{E}[u^4] + 6\mathbb{E}[u^2]\mathbb{E}[v^2] + \mathbb{E}[v^4]. \quad (\text{A11})$$

[66] Given that for the zero mean Gaussian distribution $\mathbb{E}[u^4] = 3\sigma_u^4$ and $\mathbb{E}[v^4] = 3\sigma_v^4$ leads to:

$$\mathbb{E}[z^2] = \frac{\mathbb{E}[y^4]/3 - 2\sigma_u^2\sigma_v^2 - \sigma_v^4}{\sigma_u^4}. \quad (\text{A12})$$

[67] As $\sigma_u^2 = \sigma_y^2 - \sigma_v^2$, equation (A12) can be updated to:

$$\mathbb{E}[z^2] = \frac{\mathbb{E}[y^4]/3 - 2\sigma_y^2\sigma_v^2 + \sigma_v^4}{(\sigma_y^2 - \sigma_v^2)^2}. \quad (\text{A13})$$

[68] As we assumed z is $\mathcal{LN}(\mu_z, \sigma_z)$ with $\mathbb{E}[z] = 1$, we have $\mu_z + \frac{1}{2}\sigma_z^2 = 1$, hence:

$$\mathbb{E}[z^2] = \exp(\sigma_z^2), \quad (\text{A14})$$

which, assuming $\sigma_v^2 = 0$ in equation (A13), leads to equation (12).

[69] **Acknowledgments.** This work was supported by NASA-GPM award NNX07AD33G and also Mathematics in Geosciences NSF grants EAR-0824084 and EAR-0835789. Partial support by the National Center for Earth-Surface Dynamics, an NSF Science and Technology Center (EAR-0120914), is also gratefully acknowledged. In addition, computations of this research were partly supported by the University of Minnesota Supercomputing Institute for Advanced Computational Research. We also appreciate the comments of three anonymous referees which substantially improved the presentation of our work.

References

- Andrews, D. F., and C. L. Mallows (1974), Scale mixtures of normal distributions, *J. R. Stat. Soc., Ser. B.*, 36, 99–102.
- 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.
- Bocchiola, D., and R. Rosso (2006), The use of scale recursive estimation for short term quantitative precipitation forecast, *Phys. Chem. Earth, Parts A, B, C*, 31(18), 1228–1239, doi:10.1016/j.pce.2006.03.019.
- Burt, P., and E. Adelson (1983), The Laplacian pyramid as a compact image code, *IEEE Trans. Commun.*, 31(4), 532–540.
- Ciach, J. G., and W. F. Krajewski (1999), On the estimation of radar rainfall error variance, *Adv. Water Resour.*, 22, 585–595.
- Chou, K. C., A. S. Willsky, and A. Benveniste (1994), Multiscale recursive estimation, data fusion, and regularization, *IEEE Trans. Autom. Control*, 39, 464–478.
- Coifman, R. R., and D. L. Donoho, (1995), Translation-invariant de-noising, in *Wavelets and Statistics, Lecture Notes Stat.*, vol. 103, pp. 125–150, Springer, New York.
- Daniel, M., and A. S. Willsky (1999), The modeling and estimation of statistically self-similar processes in a multiresolution framework, *IEEE Trans. Inf. Theory*, 45, 955–970.
- Ebtehaj, M., and E. Foufoula-Georgiou (2010), Orographic signature on multiscale statistics of extreme rainfall: A storm-scale study, *J. Geophys. Res.*, 115, D23112, doi:10.1029/2010JD14093.

- Ebtehaj, M., and E. Foufoula-Georgiou (2011), 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.
- Fristedt, B., N. Jain, and N. Krylov (2007), *Filtering and Prediction: A Primer*, *Student Math. Libr.*, vol. 8, 252 pp., Am. Math. Soc., Providence, R. I.
- Gonzalez, R. C., and R. E. Woods (2008), *Digital Image Processing*, 3rd ed., 954 pp., Prentice Hall, Upper Saddle River, N. J.
- 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. K., and E. C. Waymire (1993), A statistical analysis of mesoscale rainfall as a random cascade, *J. Appl. Meteorol.*, *32*, 251–267.
- Gupta, R., V. Venugopal, and E. Foufoula-Georgiou (2006), A methodology for merging multisensor precipitation estimates based on expectation-maximization and scale-recursive estimation, *J. Geophys. Res.*, *111*, D02102, doi:10.1029/2004JD005568.
- Kalman, R. (1960), A new approach to linear filtering and prediction problems, *J. Basic Eng.*, *82*, 35–45.
- Krajewski, W. F., and J. A. Smith (2002), Radar hydrology: Rainfall estimation, *Adv. Water Resour.*, *25*(8–12), 1387–1394, doi:10.1016/S0309-1708(02)00062-3.
- 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*, 182–197.
- LeCam, L. (1961), A stochastic description of precipitation, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 4, *Biology and Problems of Health*, vol. 4, edited by J. Neyman, pp. 165–186, Univ. of Calif. Press, Berkeley.
- Levy, B. C. (2008), *Principles of Signal Detection and Parameter Estimation*, 639 pp., Springer, New York.
- Lovejoy, S., and D. Schertzer (1990), Multifractals, universality classes and satellite and radar measurements of clouds and rain fields, *J. Geophys. Res.*, *95*(D3), 2021–2034.
- Luetgten, M. R., W. C. Karl, A. S. Willsky, and R. R. Tenney (1993), Multiscale representations of Markov random fields, *IEEE Trans. Signal Process.*, *41*, 3377–3396.
- Menabde, M., D. Harris, A. Seed, G. Austin, and D. Stow (1997), Multiscale properties of rainfall and bounded random cascades, *Water Resour. Res.*, *33*(12), 2823–2830, doi:10.1029/97WR02006.
- Nadarajah, S. (2005), A generalized normal distribution, *J. Appl. Stat.*, *32*(7), 685–694.
- Nason, G. P., and B. W. Silverman (1995), The stationary wavelet transform and some statistical applications, in *Wavelets and Statistics, Lecture Notes Stat.*, vol. 103, pp. 281–299, Springer, New York.
- Portilla, J., V. Strela, M. Wainwright, and E. Simoncelli (2001), Adaptive Wiener denoising using a Gaussian scale mixture model in the wavelet domain, in *Proceedings of 2001 International Conference on Image Processing*, pp. 37–40, IEEE Press, New York.
- Portilla, J., V. Strela, M. J. Wainwright, and E. P. Simoncelli (2003), Image denoising using scale mixtures of Gaussians in the wavelet domain, *IEEE Trans. Image Process.*, *12*(11), 1338–1351.
- Rauch, H., F. Tung, and C. Strieble (1965), Maximum likelihood estimates of linear dynamic systems, *AIAA J.*, *3*(8), 1445–1450.
- Strela, V. (2000), Denoising via block Wiener filtering in wavelet domain, in *Proceedings of the Third European Congress of Mathematics*, vol. 2, Birkhäuser, Basel, Switzerland.
- Strela, V., J. Portilla, and E. Simoncelli (2000), Image denoising using a local Gaussian scale mixture model in the wavelet domain, paper presented at SPIE Conference on Wavelet Applications in Signal and Image Processing, Int. Soc. for Opt. Eng., San Diego, Calif.
- Tustison, B., E. Foufoula-Georgiou, and D. Harris (2002), 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.
- Wainwright, M. J., and E. P. Simoncelli (2000), Scale mixtures of Gaussians and the statistics of natural images, *Adv. Neural Inf. Process. Syst.*, *12*, 855–861.
- Wainwright, M. J., E. P. Simoncelli, and A. S. Willsky (2001), Random cascades on wavelet trees and their use in analyzing and modeling natural images, *Appl. Comput. Harmonic Anal.*, *11*, 89–123.
- Wang, J., and D. B. Wolff (2009), Evaluation of TRMM ground-validation radar-rain errors using rain gauge measurements, *J. Appl. Meteorol. Climatol.*, *49*(2), 310–324, doi:10.1175/2009JAMC2264.
- West, M. (1987), On scale mixtures of normal distributions, *Biometrika*, *74*(3), 646–648.
- Willsky, A. S. (2002), Multiresolution Markov models for signal and image processing, *Proc. IEEE*, *90*, 1396–1458.
- Wornell, G. W. (1990), A Karhunen-Loève-like expansion for $1/f$ processes via wavelets, *IEEE Trans. Inf. Theory*, *36*, 859–861.

A. M. Ebtehaj and E. Foufoula-Georgiou, Saint Anthony Falls Laboratory, Department of Civil Engineering, University of Minnesota, Twin Cities, 2 Third Ave. SE, Minneapolis, MN 55414-2196, USA. (ebteh001@umn.edu)