Sparse Precipitation Downscaling and Multisensor Retrieval

Ardeshir Mohammad Ebtehaj [a,b] and Efi Foufoula-Georgiou [a]

a., Department of Civil Engineering, Saint Anthony Falls Laboratory and National Center for Earth-Surface Dynamics, University of Minnesota, USA
b., School of Mathematics, Minnesota Center for Industrial Mathematics, University of Minnesota, USA

UNIVERSITY OF MINNESOTA
Driven to Discover℠

Goal and Motivation

❑ Goals

- Develop state of the art algorithms for DOWNSCALING, optimal FUSION and RETRIEVAL of multi-sensor precipitation data via SPARSE REPRESENTATION and BAYESIAN statistical estimation techniques.

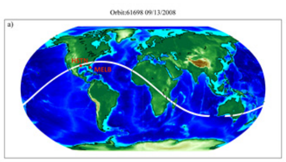
❑ Motivation:

- As multi-sensor precipitation data will be available routinely from multiple sensors (TRMM, GPM, NEXRAD, GAUGES, ...), the need for a combined high resolution precipitation product with less uncertainty becomes imperative for hydro-meteorological applications, while taking into account non-Gaussian statistics of precipitation due to presence of local extreme rain-cells.
- Operational algorithms for resolution-enhancement (downscaling) in real time need to be computationally efficient and robust
- Sparsity of precipitation images promises a new class of data-driven retrieval algorithms which can employ the coincidental DPR and radiometer (GMI) local information to retrieve more detailed precipitation information over the bands of the radiometer swath where DPR information is absent.

Collected Data Set

❑ Data Set:

- The data set used in this study is populated by near-surface reflectivity images from two hundred independent storms coincidently observed by TRMM and NEXRAD precipitation radars. The TRMM-2A25 and NEXRAD (level III) long range reflectivity products over two TRMM ground validation (GV) sites: Huston, Texas (HSTN) and, Melbourne, Florida (MELB), were collected on the basis of the TRMM overpass information provided by the TRMM-GV Office at the Goddard Space Flight Center, Maryland.



NEXRAD:

Horizontal Resolution: 1 km^2
Range: 460 km
Minimum reflectivity detection: 5 dBZ

TRMM:

Horizontal Resolution: 4.5 km^2
Minimum reflectivity detection: 17 dBZ

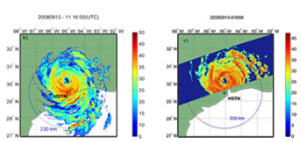


Fig.1 TRMM orbital track (a) and coincidental NEXRAD (b) and TRMM (c) observations of a hurricane over HSTN site at 09/13/2008

Statistics of Precipitation In Transform Domains

❑ Fourier and Wavelet Domains

- Energy is spread over a relatively wide range of frequencies in the Spectrum of Rainfall images. Conversely, the distribution of the wavelet coefficients exhibits extended tails significantly thicker than the Gaussian domain, implying that a large number of these coefficients are zero or very close to zero (SPARSITY)!

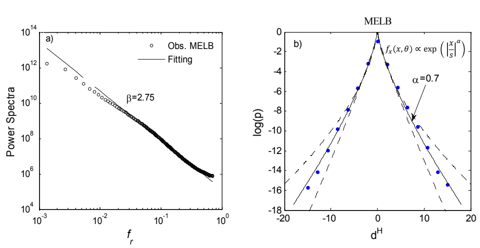


Fig.2 (a) ensemble average Fourier Spectrum of 100 independent storm images over TRMM-GV site in Melbourne Florida and, (b) the average histogram of the Horizontal wavelet coefficients of the same dataset.

Rainfall Sparsity and Applications in Multi-sensor Fusion

❑ Gaussian Scale Mixture (GSM): A Sparse Probability Model for Rainfall Fusion in the Wavelet Domain

- A large family of elliptically symmetric density functions with thicker tail than the Gaussian case (e.g., Laplace, Double exponential) can be generated as a mixture of Gaussian random variables [Andrew and Mallows ,1973; West, 1987; and Wainwright, 2001].

$$\boldsymbol{d} =^{d} \sqrt{z}\boldsymbol{u}$$

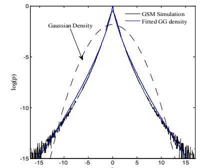


Fig.3 A sample simulation of the GSM using the log-multiplier, resembles the wavelet coefficients of the rainfall images

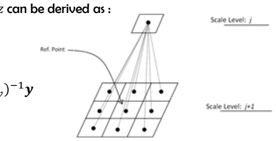
where, z is a positive independent scalar random variable, the so called mixing random variable or the multiplier, $\boldsymbol{u}$ is a zero-mean Gaussian vector with a given covariance matrix Σ_u and $\boldsymbol{d}$ is the family of Gaussian Scale Mixtures (GSM).

- Given a set of independent observations $\boldsymbol{y}$ of a GSM random vector $\boldsymbol{d} \in \mathbb{R}^n$ in Gaussian noise $\boldsymbol{y} = \boldsymbol{d} + v$, where $v \sim \mathcal{N}(0, \Sigma_v)$ and assuming a lognormal density $z \sim \mathcal{LN}(\mu_z, \sigma_z)$ as the a priori density function for the multiplier process, the maximum a posterior estimate of z can be derived as :

$$z_{MAP} = argmax_z\{\log p(z|\boldsymbol{y})\} = argmax_z\{\log p(\boldsymbol{y}|z) + logp(z)\}$$

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

$$\mathbb{E}\{\boldsymbol{d}|\boldsymbol{y}, z\} = z\Sigma_{\boldsymbol{u}}(z\Sigma_{\boldsymbol{u}} + \Sigma_v)^{-1}\boldsymbol{y}$$



where, v_n are the components of the vector $\mathrm{V} = Q^T S^{-1}\boldsymbol{y}$, S is the square root of the error covariance (i.e. $SS^T = \Sigma_v$) and (Λ, Q) contains the eigenvalues λ_n and eigenvectors of the square matrix $S^{-1}\Sigma_{\boldsymbol{u}}S^{-T}$.

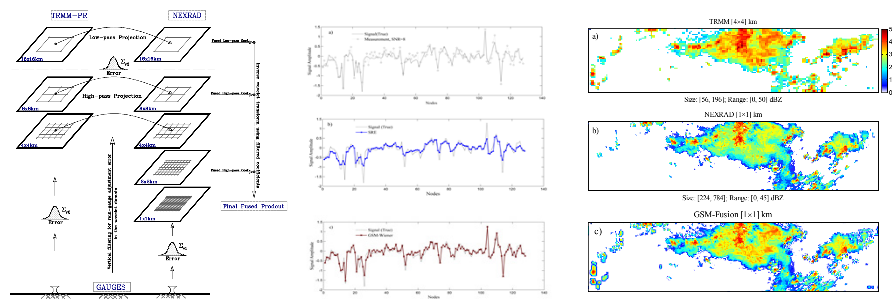
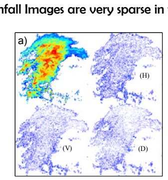


Fig.4: (Left panel) sketch of the fusion scheme, (middle panel) implementation on a 1D non-Gaussian multiscale process similar to that of observed in rainfall and (right panel), the results of fusion of the coincidental TRMM-PR with the ground-based NEXRAD base reflectivity for a storm over HSTN on 28 June 1998 at 18:13:00 UTC.

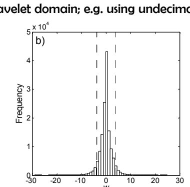
Rainfall Sparse Representation

❑ Sparsity and Recurrence of Precipitation Patches

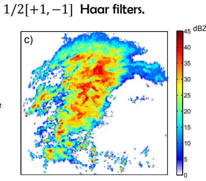
- Rainfall Images are very sparse in the wavelet domain; e.g. using undecimated 1D low-pass $1/2[+1,+1]$ and high-pass $1/2[+1,-1]$ Haar filters.



Dim: 2×[450,600]; Range: [0,45]



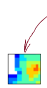
Only keeping top 20% of the wavelet coefficients more than 99.8% of the energy is preserved in the reconstructed rainfall image



Dim: [450,600]; Range: [0,45]

Fig.5: 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 (H), vertical (V) and diagonal (D) subbands, (b) histogram of the horizontal wavelet coefficients (W_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 (b) contains 80% of the wavelet coefficients that was set to zero for reconstruction of the reflectivity image, shown in (c).

- Local neighborhoods (patches) of precipitation images recur within different regions of the same storm or across different storm environments.



Precipitation patches are referred to as local neighborhoods of precipitation reflectivity values (e.g., 3x3 , 5x5 , etc.)

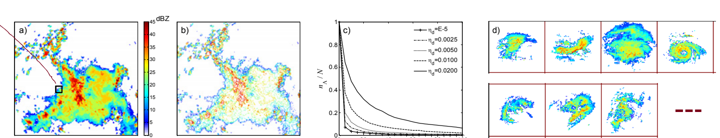


Fig.6: Recurrence of similar precipitation patches: (a) a sample precipitation reflectivity image over the TRMM (HSTN) GV-site on 1998/06/28 (18:13:00 UTC), (b) regions of high gradient (top 25\%) used for sampling of important patches, (c) a probability measure of finding similar patches in a test database and, (d) some sample reflectivity images of the test database.

Sparse Downscaling (SPaD) and Multi-sensor Retrieval

❑ Sparse Inverse Estimator

- Recently *Candes and Tao* [2006] and *Donoho* [2006] showed that recovery of a high-resolution signal from its downsampled version can go beyond Nyquist frequency, provided that the signal of interest is sparse in an appropriately chosen basis or say "dictionary".
- Estimation of a high-resolution signal/image $x = [x_1, x_2, \dots, x_m]^T \in \mathbb{R}^m$, from its low-resolution counterpart $y = \mathbb{R}^n$, where $n \leq m$, can be recast as an inverse problem, where $H \in \mathbb{R}^{n \times m}$ is a degradation operator

$$\boldsymbol{y} = H\boldsymbol{x} + \boldsymbol{e}. \qquad \text{ill-posed !}$$

- If x can be well approximated by its projection $x_s = \Phi c$ onto a redundant high-resolution dictionary $\Phi \in \mathbb{R}^{m \times M}$ $(M > m)$, where the number of non-zero elements of the coefficients are much smaller than the dimension of x (Sparsity Prior!), $\|c\|_0 \ll m$, then:

$$\boldsymbol{y} = H\Phi\boldsymbol{c} + \boldsymbol{e}'$$

- A priori sparsity implies that among many solutions of vector c, those with minimum number of non-zero elements are the solutions of interest, obeying the fidelity constraint,

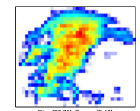
$$\hat{\boldsymbol{c}} = \arg\max_{\boldsymbol{c}} \|\boldsymbol{c}\|_0 \qquad \text{s.t.} \qquad \|\boldsymbol{y} - \Psi\boldsymbol{c}\|_2 \leq \epsilon' \qquad \text{NP-Hard !}$$

Where $\Psi = \mathrm{H}\Phi$ is the low-resolution dictionary and $\|\boldsymbol{e}'\|_2 = \epsilon'$.

- This can be solved by iterative numerical algorithms such as the *Orthogonal Matching Pursuit* (OMP) by *Mallat* [1993] or an l_1-relaxation strategy by *Chen and Donoho* (1995).

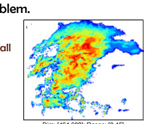
$$\hat{\boldsymbol{c}} = \arg\max_{\boldsymbol{c}} \|\boldsymbol{c}\|_1 \qquad \text{s.t.} \qquad \|\boldsymbol{y} - \Psi\boldsymbol{c}\|_2 \leq \epsilon'$$

- Forming the high (Φ) and low-resolution (Ψ) dictionaries from the patches of precipitation images in a training database of precipitation images, this inverse estimator can be applied to solve rainfall downscaling problem.



Dim: [29,38]; Range: [0,40]

Fig.7: Result of the SPaD while the rainfall image is included in the dictionaries. The MSE is less than 1% of the total energy.



Dim: [464,608]; Range: [0,45]

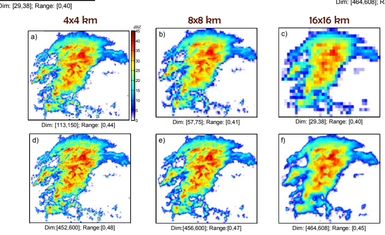


Fig.8: Results of the SPaD for downscaling of low-resolution rainfall data from 4,8,16 km in grid spacing down to 1x 1 km.

(a) The downscaled field is unique with reduced estimation error.

(b) The method is robust to measurement noise.

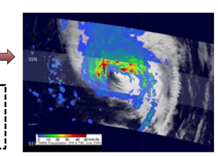
(c) The solution is smooth enough and free of the blockiness.

NEXT STATIONS !

A) Sparse Multisensor Precipitation Retrieval

B) Multiscale Sparse Data Assimilation

Dictionaries of coincidental Brightness Temperature and Reflectivity may permit recovery of rainfall reflectivity outside of the DPR coverage where only the GMI is available !



Summary of Findings

- A new method for multi-sensor precipitation data fusion is developed in the wavelet domain (Gaussian scale mixture on wavelet trees) which allows us to integrate different sources of multiscale precipitation data including : (1) Rain-gauge (2) Ground-based and (3) Satellite precipitation images, while taking into account the intrinsic statistical structure of precipitation images.
- A new downscaling framework is presented via recent development in sparse signal approximation community which shows superior performance compared to the conventional stochastic models. This estimation framework promises a new class of multi-sensor retrieval algorithms for multiscale data assimilation methods for estimation of temporal non-linear geophysical processes.

Acknowledgements:
This research was funded by NASA-GPM (Grant : NNX07AD33G) and a University of Minnesota Interdisciplinary Doctoral Fellowship (IDF).

References :

1- Ebtehaj, M., and E. Foufoula-Georgiou (2010), " Orographic Signature on Multiscale Statistics of Extreme Rainfall: A Storm Scale Study ", *J. Geophys. Res.*,
2- Ebtehaj, A.M., and E. Foufoula-Georgiou (2011), " Statistics of Precipitation Images and Cascade of Gaussian Scale Mixtures in the Wavelet Domain ", *J. Geophys. Res.*, 116, D14110, doi:10.1029/2010JD015177,2011
3- Ebtehaj, A.M., E. Foufoula-Georgiou (2011), " Adaptive non-Gaussian Fusion of Multi-sensor precipitation in the wavelet domain " *J. Geophys. Res.*, doi:10.1029/2011JD016219
4- Ebtehaj A.M., E. Foufoula Georgiou. G. Lerman (2011), Sparse Precipitation Downscaling , *J. Geophys. Res.*, Under Review