

**Hydro-meteorological Inverse Problems via Sparse
Regularization:
Advanced frameworks for rainfall spaceborne estimation**

**A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY**

Mohammad Ebtehaj

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREES OF
Doctor of Philosophy**

EFI FOUFOULA-GEORGIU

September, 2013

**© Mohammad Ebtahaj 2013
ALL RIGHTS RESERVED**

Acknowledgments

There are many people that have earned my gratitude for their contribution to my life and especially during my graduate school at the University of Minnesota. Foremost, I would like to express my sincere thanks to my adviser Professor Efi Foufoula-Georgiou, who unconditionally supported my academic endeavors and paved the way in all stages of this thesis. I am also grateful to Professor Gilad Lerman who showed me new research directions and encouraged me to blend my engineering skills with mathematical knowledge. Last but not least, my thanks also go to Professor Chris Paola and Professor Heinz Stefan who have always been inspiring teachers to me and kindly accepted to serve as committee members of this thesis.

Dedications

To my parents: Eshrat Madelat Saffari and Ali Akabar Ebtahaj who scarified their youth for their children and have been the origin of my academic aspirations.

To my lovely wife: Sara Khanzadi who has been a patient witness to all of the ups and downs of the past few years and supported me through every step of the way with honesty.

Abstract

The past decades have witnessed a remarkable emergence of new spaceborne and ground-based sources of multiscale remotely sensed geophysical data. Apart from applications related to the study of short-term climatic shifts, availability of these sources of information has improved dramatically our real-time hydro-meteorological forecast skills. Obtaining improved estimates of hydro-meteorological states from a single or multiple low-resolution observations and assimilating them into the background knowledge of a prognostic model have been a subject of growing research in the past decades. In this thesis, with particular emphasis on precipitation data, statistical structure of rainfall images have been thoroughly studied in transform domains (i.e., Fourier and Wavelet). It is mainly found that despite different underlying physical structure of storm events, there are general statistical signatures that can be robustly characterized and exploited as a prior knowledge for solving hydro-meteorological inverse problems such rainfall downscaling, data fusion, retrieval and data assimilation. In particular, it is observed that in the wavelet domain or derivative space, rainfall images are *sparse*. In other words, a large number of the rainfall expansion coefficients are very close to zero and only a small number of them are significantly non-zero, a manifestation of the non-Gaussian probabilistic structure of rainfall data. To explain this signature, relevant family of probability models including Generalized Gaussian Density (GGD) and a specific class of conditionally linear Gaussian Scale Mixtures (GSM) are studied. Capitalizing on this important but overlooked property of precipitation, new methodologies are proposed to optimally integrate and improve resolution of spaceborne and ground-based precipitation data. In particular, a unified framework is proposed that ties together the problems of downscaling, data fusion and data assimilation via a regularized variational approach, while taking into account the underlying sparsity in an appropriately chosen transform domain. This framework seeks solutions beyond the paradigm of the classic least squares by imposing a proper regularization. The results suggest that sparsity-promoting regularization can reduce uncertainty of estimation in hydro-meteorological inverse problems of downscaling, data fusion, and data assimilation. In continuation of the proposed methodologies, we also introduce a new data driven framework for multisensor spaceborne rainfall retrieval problem and present some preliminary and promising results.

Contents

| | |
|---|-------------|
| List of Tables | viii |
| List of Figures | ix |
| 1 An Overview | 1 |
| 1.1 Introduction and problem statement | 1 |
| 2 Rainfall statistics and Orographic signature | 6 |
| 2.1 Introduction | 6 |
| 2.2 Storm Environment | 8 |
| 2.3 Statistical Signature of Orography on Rainfall Cells | 10 |
| 2.4 Multiscale Analysis of Spatial Rainfall | 12 |
| 2.4.1 Scaling Manifestations in the Frequency and Real Space Domain | 12 |
| 3 Statistics of precipitation images | 20 |
| 3.1 Introduction | 20 |
| 3.2 Precipitation Data and Elementary Statistics | 21 |
| 3.3 Spectral Signature | 26 |
| 3.4 Statistics of Subband Components in the Wavelet Domain | 30 |
| 3.5 Wavelet Decomposition and Marginal Statistics | 31 |
| 3.6 Joint Statistics of the Wavelet Coefficients | 37 |
| 3.6.1 Scale-to-scale dependence | 37 |
| 3.6.2 Intra scale dependence | 39 |
| 3.7 GSM for the Wavelet Coefficients of Rainfall Images | 40 |
| 3.7.1 Cascade of Gaussian Scale Mixtures on Wavelets Trees | 41 |
| 3.7.2 GSM on Wavelet Trees vs. Multiplicative Random Cascades | 46 |

| | | |
|----------|---|-----------|
| 4 | Rainfall fusion via GSM filter | 49 |
| 4.1 | Linear Fusion of Multisensor Precipitation Data in the Spatial Domain . . . | 50 |
| 4.1.1 | Principle of Linear Least Squares Estimation | 50 |
| 4.1.2 | Scale Recursive Estimation of Multi-sensor Precipitation data | 52 |
| 4.2 | Precipitation Probability Model in the Wavelet Domain | 57 |
| 4.3 | GSM Optimal Estimation in the Wavelet Domain | 59 |
| 4.3.1 | Basics of the Framework | 59 |
| 4.3.2 | Global Versus Local Filtering | 62 |
| 4.3.3 | Synthetic 1-D Implementation | 63 |
| 4.4 | GSM Multi-sensor Fusion of Precipitation Data | 67 |
| 4.4.1 | Concepts | 67 |
| 4.4.2 | A Case Study on Precipitation Data | 70 |
| 5 | Sparse precipitation downscaling | 74 |
| 5.1 | Introduction | 74 |
| 5.2 | Sparse Inverse Estimator | 77 |
| 5.3 | On Sparsity of Rainfall Images | 79 |
| 5.4 | Recurrence of Rainfall Patches and Group Sparsity | 80 |
| 5.5 | Precipitation Sparse Downscaling | 82 |
| 5.5.1 | Conceptual Framework | 82 |
| 5.5.2 | SPaD via Dictionary Learning | 85 |
| 5.5.3 | Results | 86 |
| 5.5.3.1 | Inputs and Parameters | 86 |
| 5.5.3.2 | Downscaling Performance and Error Analysis | 88 |
| 6 | Regularized downscaling, data fusion and assimilation | 93 |
| 6.1 | Introduction | 93 |
| 6.2 | Discrete Inverse problems: Conceptual Framework | 95 |
| 6.3 | Regularized Downscaling | 96 |
| 6.3.1 | Problem Formulation | 96 |

| | | |
|----------|---|------------|
| 6.3.2 | Examples on Rainfall DS | 99 |
| 6.3.2.1 | Problem Formulation and Settings | 99 |
| 6.3.2.2 | Results of the rainfall downscaling (DS) | 103 |
| 6.4 | Regularized Data Fusion | 107 |
| 6.4.1 | Problem Formulation | 107 |
| 6.4.2 | Example on Rainfall DF | 109 |
| 6.5 | Regularized Variational Data Assimilation | 111 |
| 6.5.1 | Classic Variational Data Assimilation | 113 |
| 6.5.2 | Regularized Variational Data Assimilation | 115 |
| 6.5.2.1 | Background | 115 |
| 6.5.2.2 | A Generalized Framework to Regularize Variational Data Assimilation in Transform Domains | 116 |
| 6.5.2.3 | Solution Method via Quadratic Programing | 118 |
| 6.5.3 | Examples on Linear Advection-Diffusion Equation | 119 |
| 6.5.3.1 | Problem Statement | 119 |
| 6.5.3.2 | Assimilation Set Up | 121 |
| 6.5.3.3 | Results of Assimilation Experiments | 124 |
| 6.5.3.4 | Selection of the regularization parameters | 131 |
| 7 | Combined Radar Radiometer rainfall retrieval via Sparse Representation | 132 |
| 7.1 | Introduction | 132 |
| 7.2 | A Sketch of the Proposed algorithm | 134 |
| 7.3 | Data | 135 |
| 7.4 | Rain/No-rain Classification | 136 |
| 7.5 | Some Preliminary Results | 138 |
| 8 | Conclusion | 143 |
| | Bibliography | 145 |
| A | Appendix A. Details of Some Derivations in Chapter 4 | 162 |
| A.1 | Maximum a Posteriori Estimate of the Log-normal Multiplier | 162 |
| A.2 | Parameters of the MAP estimator | 164 |

| | | |
|----------|---|------------|
| B | Appendix B. Statistical Interpretations and Solution Methods | 165 |
| B.1 | Statistical Interpretations | 165 |
| B.1.1 | Regularized Variational Downscaling and Data Fusion Problem . . . | 165 |
| B.1.2 | Regularized Variational Data Assimilation | 166 |
| B.2 | Gradient Projection for Huber regularization | 168 |
| B.3 | Quadratic Programming form of the ℓ_1 -norm RVDA | 168 |
| B.4 | Upper Bound of the Regularization Parameter | 169 |
| B.5 | Gradient Projection Method for Sparsity-promoting Data Assimilation . . . | 170 |

List of Tables

| | | |
|-----|--|-----|
| 3.1 | Error statistics of TRMM-PR versus ground based NEXRAD. | 25 |
| 3.2 | Parameters of rainfall power spectral density models. | 30 |
| 3.3 | Scaling of first and second order moments of rainfall data. | 37 |
| 4.1 | Results of SRE rainfall fusion | 57 |
| 4.2 | Results of the GSM rainfall fusion. | 73 |
| 5.1 | Results of sparse precipitation downscaling. | 92 |
| 6.1 | Quantitative results of variational downscaling via Huber regularization. . . | 107 |
| 6.2 | quantitative results of the rainfall data fusion via Huber regularization. . . | 111 |
| 6.3 | Results of the ℓ_1 -norm RVDA for the white background error. | 127 |
| 6.4 | Results of the ℓ_1 -norm RVDA for the AR(2) background error. | 130 |

List of Figures

| | | |
|------|---|----|
| 2.1 | Digital elevation model and spatial mean of the rain cells. | 9 |
| 2.2 | Second order statistics of rain cells with respect to the underlying elevation. | 11 |
| 2.3 | Semi-log relationship of rainfall intensity with the underlying elevation. | 13 |
| 2.4 | Rainfall intensity and its log-transformation. | 14 |
| 2.5 | Rapidan storm 2D ensemble power spectral density. | 15 |
| 2.6 | Rapidan storm radially averaged power spectral density. | 16 |
| 2.7 | Spectral slope versus elevation in Rapidan storm. | 17 |
| 2.8 | Rapidan storm variogram. | 18 |
| 2.9 | Distribution of the gradient in Rapidan storm. | 19 |
| 3.1 | Frequency of the studied storms snapshots. | 23 |
| 3.2 | Sample representation of studied storm. | 24 |
| 3.3 | TRMM-PR versus ground based NEXRAD. | 25 |
| 3.4 | error of TRMM-PR versus ground based NEXRAD. | 26 |
| 3.5 | 1D and 2D spectral density of NEXRAD snapshots | 27 |
| 3.6 | 1D and 2D spectral density of TRMM-PR orbital observations | 29 |
| 3.7 | Sparsity of rainfall data in the wavelet domain. | 33 |
| 3.8 | Rainfall versus Brownian motion in the wavelet domain. | 34 |
| 3.9 | Rainfall versus Brownian motion in the wavelet domain. | 35 |
| 3.10 | Multi-fractal structure of rainfall in the wavelet domain. | 36 |
| 3.11 | Joint histogram of the rainfall wavelet coefficients. | 38 |
| 3.12 | Spatial correlation of the rainfall wavelet coefficients. | 40 |
| 3.13 | Graphical Tree-like models for rainfall modeling | 43 |

| | | |
|------|--|-----|
| 3.14 | Gaussian scale mixtures and rainfall wavelet modeling. | 44 |
| 3.15 | Gaussian scale mixtures parametrization for rainfall modeling. | 45 |
| 3.16 | Simulation results of the proposed Gaussian scale mixture probability model. | 47 |
| 4.1 | Graphical Tree model and notation | 52 |
| 4.2 | summary of rainfall statistics in Fourier and wavelet domain | 54 |
| 4.3 | SRE fusion of the TRMM-PR and ground-based NEXRAD observations. | 56 |
| 4.4 | Rainfall scale-to-scale and intra scale dependence | 60 |
| 4.5 | Rainfall intra scale correlation in the wavelet domain. | 60 |
| 4.6 | GSM simulation with log-multiplier. | 61 |
| 4.7 | Schematic of a local neighborhood. | 64 |
| 4.8 | A 1D GSM- filtering experiment. | 66 |
| 4.9 | Schematic of GSM-Fusion for rainfall merging. | 68 |
| 4.10 | Rain Gauge filtering using GSM. | 71 |
| 4.11 | GSM fusion fusion of TRMM-PR with ground-based NEXRAD. | 72 |
| 5.1 | Rainfall sparsity in the wavelet domain. | 80 |
| 5.2 | Rainfall patch base sparsity | 82 |
| 5.3 | Learned rainfall dictionaries. | 88 |
| 5.4 | Spad results for exact dictionaries. | 89 |
| 5.5 | Results of sparse precipitation downscaling. | 91 |
| 6.1 | Huber norm vs ℓ_2 and ℓ_1 -norm | 98 |
| 6.2 | Mathematical models for smoothing and downsampling of rainfall images. | 100 |
| 6.3 | Uniform (low-pass) smoothing and (high-pass) Laplacian Filter. | 101 |
| 6.4 | Sparsity in the Laplacian domain. | 102 |
| 6.5 | Variational downscaling using Huber norm regularization. | 104 |
| 6.6 | Zooming view of the variational downscaling results. | 105 |
| 6.7 | q-q plots of the downscaled fields vs high resolution observations. | 106 |
| 6.8 | Rainfall data fusion via Huber norm regularization. | 110 |
| 6.9 | q-q plots of the fused rainfall fields vs high resolution observations. | 111 |

| | | |
|------|---|-----|
| 6.10 | Initial conditions and their evolutions with the linear advection-diffusion. . . | 122 |
| 6.11 | Noisy and downsampled observations used in RD-VAR experiments. | 123 |
| 6.12 | Condition number of the background error correlation matrices. | 124 |
| 6.13 | Sample path of the AR(1) and AR(2), used for background error modeling. | 125 |
| 6.14 | Results of the ℓ_1 -norm RVDA for the white background error. | 126 |
| 6.15 | Results of the ℓ_1 -norm RVDA for the AR(1) background error. | 128 |
| 6.16 | Results of the ℓ_1 -norm RVDA for the AR(1) background error. | 129 |
| 6.17 | Mean squared error vs regularization parameter in RVDA experiments. . . . | 131 |
| | | |
| 7.1 | TRMM sensor package. | 133 |
| 7.2 | Schematic of sparse retrieval. | 134 |
| 7.3 | Multispectral Brightness temperature and surface rain t orbital track No. 7 on June, 28, 1998. | 136 |
| 7.4 | Matrix scatter plot of rainy versus none rainy pixels over ocean. | 137 |
| 7.5 | Matrix scatter plot of rainy versus none rainy pixels over land. | 139 |
| 7.6 | PR, 2A12 versus sparse retrieval at orbital track orbital track No. 7 on June, 28, 1998. | 140 |
| 7.7 | PR, 2A12 versus sparse retrieval for Hurricane Danielle. | 141 |
| 7.8 | Histogram of PR and 2A12 versus sparse retrieval. | 142 |

Chapter 1

An Overview

1.1 Introduction and problem statement

In parallel to the growing technologies for earth remote sensing observatory systems, for enhancing our environmental forecast skills, we have witnessed an increasing interest to improve the accuracy of observations and integrate them with predictive models. Remote sensing observations are typically noisy and coarse-scale representations of a true state variable of interest without sufficient details for fine-scale environmental modeling. In addition, environmental predictions are not perfect as prognostic models often suffer either from inadequate characterization of the underlying physics or inaccurate initialization. Given these limitations, several classes of estimation problems present themselves as continuous challenges for the atmospheric, hydrologic, and oceanic science communities. These include: (1) Downscaling which refers to the class of problems for enhancing the resolution of a measured or modeled state variable of interest by producing a fine-scale representation with reduced uncertainty; (2) Data fusion, to produce an improved estimate from a suite of noisy observations at different scales; (4) Data assimilation which deals with estimating initial conditions of a predictive model consistent with the available incomplete observations and the underlying dynamics; and (5) Data retrieval for recovery of state variable of interest from indirect and non-linearly related noisy observations. In this thesis, we revisit the problems of downscaling (DS), data fusion (DF), and data assimilation (DA) focusing on a common thread between them as variational inverse problems. New approaches are presented via applying proper regularization in transform domains and efficient solution methods are explored. In this thesis, we direct our concentration on hydro-meteorological problems with particular emphasis on land-surface hydrologic applications and in particular rainfall related inverse problems. In addition, a new framework for spaceborne rainfall data retrieval (DR) is briefly explained and some promising results are presented.

In land-surface hydrologic studies, downscaling (DS) of precipitation and soil moisture

observations has received considerable attention, using a relatively wide range of methodologies. DS methods in hydro-meteorology and climate studies generally fall into three main categories namely, dynamic downscaling, statistical downscaling, and variational downscaling. Dynamic downscaling often uses a regional physical model to reproduce fine-scale details of the state of interest consistent with the large-scale observations or outputs of a global circulation model (e.g., *Reichle et al.*, 2001a; *Castro et al.*, 2005; *Zupanski et al.*, 2010). Statistical downscaling methods encompass a large group of methods that typically use empirical multiscale statistical relationships, parameterized by observations or other environmental predictors, to reproduce realizations of fine-scale fields. Precipitation and soil moisture statistical downscaling has been mainly approached via spectral or (multi)fractal interpolation methods, capitalizing on the presence of a power law spectrum or statistical self-similarities in precipitation or soil moisture fields (*Lovejoy and Mandelbrot*, 1985; *Lovejoy and Schertzer*, 1990; *Gupta and Waymire*, 1993; *Kumar and Foufoula-Georgiou*, 1993a,b; *Perica and Foufoula-Georgiou*, 1996a,b; *Veneziano et al.*, 1996; *Wilby et al.*, 1998b,a; *Deidda*, 2000; *Kim and Barros*, 2002; *Rebora et al.*, 2005; *Badas et al.*, 2006; *Merlin et al.*, 2006, , among others). In variational approaches, a direct cost function is defined whose optimal point is the desired fine-scale field which can be obtained via using an optimization method. Recently along this direction, *Ebtehaj et al.* (2012) cast the rainfall DS problem as an inverse problem using sparse regularization to address the intrinsic rainfall singularities and non-Gaussian statistics.

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

multiple scales (e.g., *Gorenburg et al.*, 2001; *Tustison et al.*, 2003; *Bocchiola*, 2007; *Van de Vyver and Roulin*, 2009; *Wang et al.*, 2011), see also (*Kumar*, 1999) for soil moisture applications. Recently, using the Gaussian scale mixture probability model and an adaptive filtering approach *Ebtehaj and Foufoula-Georgiou* (2011b) proposed a fusion methodology in the wavelet domain to merge TRMM-PR and ground-based NEXRAD measurements, aiming to preserve the non-Gaussian structure and local extremes of precipitation fields.

Data assimilation (DA) has played an important role in improving the skills of environmental forecasts and has become by now a necessary step in operational predictive models (see *Daley*, 1993). DA amounts to integrating the underlying knowledge from the *observations* into the first guess or the *background* state, typically provided by a physical model from the previous forecast step. The goal is then to obtain an improved estimate of the current state of the system with reduced uncertainty, the so called *analysis*. The analysis is then used to forecast the state at the next time step and so on (see *Daley*, 1993; *Kalnay*, 2003, for a comprehensive review). One of the most common approaches to the data assimilation problem relies on variational techniques (e.g., *Sasaki*, 1958; *Lorenc*, 1986; *Talagrand and Courtier*, 1987; *Courtier and Talagrand*, 1990; *Parrish and Derber*, 1992; *Zupanski*, 1993; *Courtier et al.*, 1994; *Reichle et al.*, 2001b; *Margulis and Entekhabi*, 2003, among many others). In these methods, one explicitly defines a cost function, typically quadratic, whose unique minimizer is the analysis state. Very recently *Freitag et al.* (2012) and *Ebtehaj and Foufoula-Georgiou* (2013) proposed a regularized variational data assimilation scheme to improve assimilation results in the presence of sparsity in a pre-selected basis.

Data retrieval (DR) is also an essential component in hydro-meteorological estimation problems, especially for remote sensing observations. DR is typically referred to the class of nonlinear hydro-meteorological inverse problems. In this class of problems, we typically seek estimates of a particular state of interest from noisy and non-linearly related observations. Data retrieval is a necessary step for many operational data assimilation problems such as the spaceborne rainfall data assimilation. Spaceborne rainfall detection and estimation heavily rely on remotely sensed passive microwave signals which encode the rainfall radiant and scattering fluxes. These energy fluxes are not linearly dependent to the surface rainfall intensity and rainfall estimation from these observations requires to solve a highly nonlinear inverse problem. The existing non-linearly is typically governed by complex physically based cloud resolving models that can not be easily framed into the context of classic formulation of inverse problems. Therefore, typically we solve a rainfall retrieval problem to map radiant energy fluxes onto the rainfall space prior to any data assimilation experiment.

Inspired by statistical structure of rainfall data and ubiquity of non-Gaussian sparsity in many natural processes, the central goal of this thesis is to propose a unified sparsity promoting framework for the class of DS, DF, DA, and DR problems. The proposed

framework relies on the recent developments in mathematical formalisms of inverse problems (e.g., *Elad*, 2010). These recent developments have received a great deal of attention in the statistical regression and imaging sciences in the past decades (e.g., *Tibshirani*, 1996; *Chen et al.*, 1998; *Wainwright and Simoncelli*, 2000; *Chen et al.*, 2001), but are relatively new to the geophysical community. Presented methodologies are examined through downscaling and data fusion of rainfall remotely sensed observations, which has fundamental applications in flash flood predictions, especially in small watersheds (*Rebora et al.*, 2005; *Siccardi et al.*, 2005; *Rebora et al.*, 2006). We show that the presented methodologies allow us to improve the quality of rainfall estimation and reduce the uncertainty of estimation by recovering the small-scale high-intense rainfall extreme features, which have been lost due to the possible low-resolution constraints of the sensor. For the family of data assimilation problems, the promise of the presented framework, is demonstrated via simple example using the linear advection-diffusion equation, which has a key role in study of land-surface heat and mass fluxes (e.g., *Milly*, 1982; *Entekhabi et al.*, 1994; *Liang et al.*, 1999; *Caparrini et al.*, 2003; *Bateni and Entekhabi*, 2012). The results demonstrate that the accuracy of the analysis and forecast cycles in a data assimilation problem can be markedly improved, compared to the classic variational methods, especially when the initial state is sufficiently sparse in a properly chosen basis. For the data retrieval (DR) problem, we also introduce a novel supervised statistical learning algorithm that allows us to estimate surface rain rates from indirect spaceborne measurements of atmospheric brightness temperature. In this study we use Tropical Rainfall Measuring Mission (TRMM) satellite observations. In this satellite there are multiple sensors including an active precipitation radar (PR) and a passive radiometer (TMI) that provide a valuable library of coincidentally observed reflectivity and radiance of precipitable water in the vertical structure of atmosphere. The PR sensor is more accurate than the TMI but has a more confined areal coverage. Therefore, the motivation is to learn from these libraries of coincidental observations to enhance the accuracy of the low-resolution sensor (TMI) over the areas where the high-resolution sensor (PR) is blind. The results of the presented algorithm is extremely promising and comparable with the in-place physically based algorithms, developed in the past two decades. The results show a promising path towards further developments of the algorithm for operational purposes. This thesis is structured as follows:

Chapter 2 provides an extensive case study of high-resolution stage II Doppler radar data of the Rapidan storm, June 1995, over the Appalachian Mountains. This chapter provides preliminary insights into rainfall statistics in the Fourier and wavelet domain and their possible connections with the underlying orographic features. Chapter 3 is devoted to an extensive statistical analysis of high-resolution precipitation images obtained from the stage III ground-based radar observations over two ground validation sites (in Texas and Florida) of the Tropical Rainfall Measuring Mission (TRMM) satellite. In this chapter we study the rainfall statistics and their scaling behavior in the wavelet domain and introduce relevant

probability models with particular emphasize on the heavy tailed non-Gaussian and sparse structure of the rainfall fields. Chapter 4 explains a filtering methodology for fusion of multiscale multisensor precipitation data using the Gaussian Scale Mixtures (GSM) probability model. Examples are provided to merge coincidental TRMM observations with the ground-based NEXRAD radar data. Building upon the observed rainfall sparsity, Chapter 5 introduces a new rainfall downscaling methodology via ℓ_1 -norm regularization and dictionary learning. Inspired by the previous findings, new methodologies are presented in Chapter 6, that ties together the problems of variational downscaling (DS), data fusion (DF), and data assimilation (DA) via sparsity-promoting regularization. Chapter 7 is devoted to present the new data retrieval (DR) algorithm. Some promising results are provided in this chapter on retrieval of surface rain rates, given passive multispectral observations of the TMI sensor on board of the TRMM satellite.

Chapter 2

Orographic signature on multiscale statistics of extreme rainfall: A storm-scale study

Rainfall intensity and spatio-temporal patterns often show a strong dependence on the underlying terrain. The main objective of this work is to study the statistical signature imprinted by orography on the spatial structure of rainfall and its temporal evolution at multiple scales. The results of an extensive analysis of the high-resolution stage II Doppler radar data of the Rapidan storm, June 1995, over the Appalachian Mountains is reported in this study. The orographic signature on the elementary statistical structure of the precipitation fields is studied via a variable intensity thresholding scheme. This signature is further explored at multiple scales via analysis of the dependence of precipitation fields on the underlying terrain both in spatial and Fourier domains. This chapter provides insights on non-Gaussian structure of precipitation and serves as the main motivation to study the rainfall statistical structure in the wavelet domain which is the subject of the next chapter.

2.1 Introduction

There is evidence to suggest that the small]scale space]time structure of intense mountainous storms (subkilometer and subhour scales) has important hydrologic and geomorphic implications both in the short]term (e.g., triggering landslides and flash floods) and in the long]term evolution of landscapes (*Montgomery et al.*, 2001; *Roe*, 2005). Also the consequences of extreme mountainous storms are often disastrous for the local and regional economy. For instance, the storm of 27 June 1995 over Madison County in Virginia caused a peak rainfall accumulation exceeding 600 mm in 6 h (*Smith et al.*, 1996) which produced flash flooding and triggered massive shallow landsliding and destructive debris flows across

the region. Three fatalities and a total loss of property exceeding 200 million dollars were reported. During the event of 28 July 1997, the city of Fort Collins in Colorado experienced a catastrophic flash flood that caused five fatalities and more than 200 million dollars in damage. Maximum accumulated rainfall for that storm exceeded 254 mm within 6 h (*Lin et al.*, 2001).

In the long-term average, it is well known that the windward side of a mountainous range perpendicular to the moist wind flow receives more precipitation than the leeward side and leaves the dried air parcels to descend, resulting in a moisture-depleted region downslope called the rain shadow. This precipitation regime causes distinct climatic differences across the divide. This phenomenon is the first order and the most straightforward effect of the orographic signature on the spatial precipitation pattern. However, the full interaction of orography and rainfall spatiotemporal distribution is actually more complex and involves a variety of physical mechanisms. Forced lifting, differential advection of moist air aloft, convective currents initiated by mountain valley thermal gradients and cooling of warm moist air over snow covered mountains, individually or in combination, may enhance the intensity of rainfall production (*Barros and Kuligowski*, 1998). *Roe* (2005) categorized the mechanisms of orographic precipitation as: (1) stable upslope ascent, (2) partial blocking of the impinging air mass, (3) down valley flow induced by evaporative cooling, (4) leeside convergence, (5) convection owing to mechanical lifting above the level of free convection, and (6) seeder and feeder mechanism.

Orographic terrains fuel the storm by intensifying the destabilizing forces and accelerating the updraft movement of moist laden air masses. A significant number of the most heavy and disastrous precipitation events has been observed during the warm seasons over mesoscale mountainous ranges with several hundred kilometers in width such as the Rockies and the Appalachians in the United States and the Alps in Europe. Based on the analysis of several extreme storms in the United States (i.e., Black Hill flood, SD., 1972; Big Thompson Canyon, CO., 1976; Fort Collins, CO., 1997; Madison County, VA., 1995), it can be concluded that the occurrence of such extreme events emerges from the coincidence of some common synoptic and mesoscale hydro-meteorological mechanisms. Coexistence of a low level and efficient moist airflow, steep gradient and strong quasi-stationary convective systems are common features in summer time orographically accelerated extreme storms (*Lin et al.*, 2001; *Bousquet and Smull*, 2003). In such cases, most commonly, an area of intense pressure gradients is formed across the region under a quasi-stationary synoptic system which develops a low-level strongly moist upslope airflow. Orographic lifting provides the required energy to release the conditional instability and trigger a heavy rainfall. The quasi-stationary nature of the synoptic system develops a slow motion storm and magnifies more the hydromorphologic responses such as flash flooding and landsliding. Interestingly, studies of such storms in the United States show that a high value of the convective potential energy (CAPE) is not consistently observed in the storm environment (*Lin et al.*,

2001).

In parallel to studying the driving physical mechanisms of precipitating systems, in recent decades considerable efforts have been expended in understanding and quantifying the statistical structure of precipitation over different spatiotemporal scales (*Lovejoy and Mandelbrot*, 1985; *Lovejoy and Schertzer*, 1985, 1990; *Gupta and Waymire*, 1990, 1993; *Kumar and Foufoula-Georgiou*, 1993a,b; *Veneziano et al.*, 1996; *Deidda*, 2000; *Deidda et al.*, 2004, 2006; *Badas et al.*, 2006). Optimal parameterization and modeling of the multiscale behavior of rainfall fields often requires understanding of the conditional dependence of the statistical scaling parameters on some measures of the underlying physical processes. Along these lines, *Over and Gupta* (1994) investigated how the structure of rainfall at the mesoscale varies with time according to the large-scale dynamical forcing. *Perica and Foufoula-Georgiou* (1996a,b) established predictive relationships between statistical characteristics of mesoscale rainfall and thermodynamics of the storm environment. *Harris et al.* (1996); *Nykanen and Harris* (2003); *Badas et al.* (2006) investigated the spatial structure of orographic rainfall in the context of multifractal analysis. To advance the previous findings and to gain insight toward a practical conditional parameterization of statistical downscaling schemes over complex terrains, the main goal of this work is to present the results of an extensive study on quantifying the scaling structure of an orographic precipitation system and relating it to the underlying topography in Fourier and wavelet domains. The rainfall data of an extreme storm over Madison County, Virginia (latitude 37.9859N to 38.8763N, longitude 77.8341W to 78.9714W) on 27 June 1995 are used in this study. The precipitation fields at 1 km×km resolution (100×100 bins) are derived from the Sterling WSR-88D volume scan reflectivity data at every 6 min, using the Z-R conversion relationship $R = aZ^b$ with $a = 0.0425$ and $b = 0.71$ (*Smith et al.*, 1996). The underlying topographic elevations of the studied area at the same scale range between 66 and 1130 m above the mean sea level (see Figure 2.1a).

In section 2.2, a concise description of the storm physical environment and the orographic precipitation mechanism is presented. In section 2.3, we establish a pyramidal variable thresholding method for tracking the signature of orographic features especially on precipitation tail statistics. In section 2.4, the statistics and spatial scaling properties of the studied rainfall event are presented and discussed in real space, and in Fourier and wavelet domains.

2.2 Storm Environment

In this section, we briefly describe the mesoscale atmospheric conditions and the associated hydrologic and geomorphic responses of the Madison County storm of 27 June 1995. The reader is referred to *Pontrelli et al.* (1999) and *Smith et al.* (1996) for a comprehensive

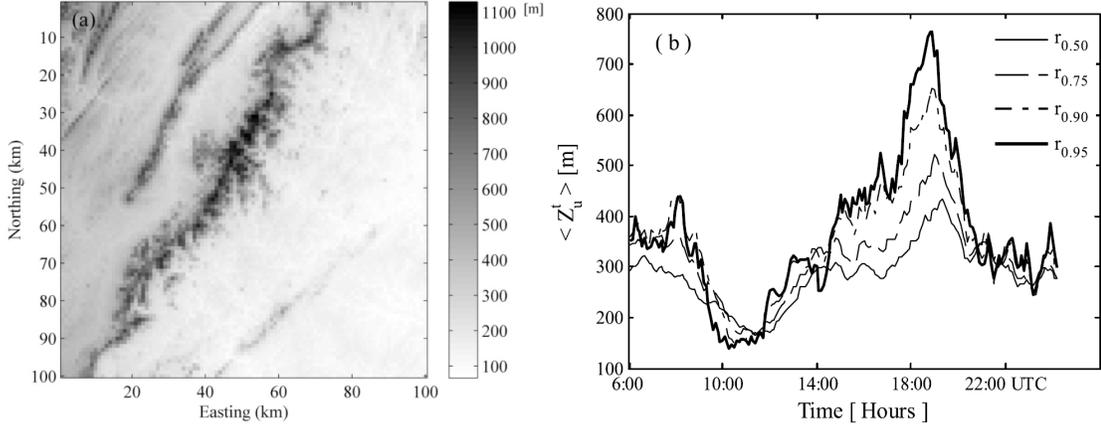


Figure 2.1: (a) Digital Elevation Model (DEM) of the studied area and (b) the spatial mean of the elevations underneath the rain cells $\langle Z_u^t \rangle$ defined at four different quantile levels $u = 0.50, 0.75, 0.90, 0.95$ (see text for definition).

study of the storm environment and the measured hydrologic response, respectively. The activity of this storm can be divided into three main phases: “prestorm phase” from 00:00 UTC to 05:00 UTC, “storm phase” from 05:00 to 20:00 UTC and the “post-storm phase” which extends from 20:00 to 24:00 UTC. It needs to be mentioned that the “storm phase” consists of two major convective activities across the region: the first and larger one in spatial extent took place over the Piedmont plains and moved toward Madison County from 0500 to 10:00 UTC; the second and larger one in intensity initiated near 12:00 UTC, moved very slowly over the Madison county and produced flash flooding and massive debris flows.

During the prestorm phase, a high-pressure system was positioned over New England pushing cold air southeastward into a warm low-pressure system centered off the coast of the Carolinas. Establishment of the synoptic high and low-pressure systems over the region initiated a pronounced easterly cool and moist maritime airflow, impinging the eastern flank of the Appalachians. The lifted index value for the surface layer was slightly negative and the CAPE was relatively small ($\sim 150 \text{ Jkg}^{-1}$) across the region. Atmospheric sounding at Sterling Radar Station (VA), indicated that the entire troposphere was near saturation. A high-pressure system built more southward, pressing the cold front into northern Virginia. Eventually, the easterly wind component and geographic features of the region in the Piedmont plains and Madison County triggered the first convective system around 0700 UTC. This front advanced southward and left the Madison County nearly rain-free and dissipated at 1100 UTC. This stage of the storm did not produce a major flash flooding but just increased the soil moisture content and left the region more vulnerable to excess precipitation. Around 1200 UTC, sounding information from the Sterling radar station signified that another strong convection was imminent. The level of free convection estimated at around 337m above the ground level, CAPE increased nearly to $600 \text{ [J kg}^{-1}]$, the lifted

index lowered slightly to -2, and the precipitable water was about 50 mm (*Pontrelli et al.*, 1999). Triggered by orography, the second strong convective system formed across the region during the late morning and early afternoon and prolonged to near 2000 UTC. This part of the storm was characterized by intense and slow moving rainy cells (~ 1.2 [ms⁻¹]) that produced a peak discharge of 3000 [m³s⁻¹] at 1845 UTC in the Rapidan watershed near Ruckersville station (*Smith et al.*, 1996) and mobilized hundreds of shallow landslides into massive debris flows (*Wieczorek et al.*, 2004). By 1800 UTC, the storm overtook the southern border of the Madison County and finally around 20:00 UTC, the areal extent of the precipitation decreased and the system dissipated significantly. Some scattered convective activities were observed, but by 2400 UTC these activities disappeared completely. The most striking feature of this storm was its interaction with the underlying orography which along with other meteorological factors made the storm intense and very stationary (*Smith et al.*, 1996).

2.3 Statistical Signature of Orography on Rainfall Cells

Several observations of spatial rainfall (e.g., *Le Cam*, 1961; *Gupta and Waymire*, 1990) support the fact that the spatial organization of rainfall possesses a hierarchical structure during the storm, small clusters of intense rainy cells are embedded inside the less intense and larger extent clusters of rain cells. In this section, we propose a simple methodology for extracting areas of intense rainfall activity and quantifying their underlying elevation for the purpose of characterizing dynamically the signature of topography on the statistics of rainfall intensity. At every snapshot of radar images, we define as *rain cells* all pixels of the rainfall fields over which the intensity exceeds a preassigned quantile of the overall storm rainfall intensity distribution. Delineating the geometry of the rain cells in time by this thresholding process, we dynamically analyzed the spatial statistics of the within-cells rainfall intensity with respect to the elevation of the topographic features underneath those cells.

Let $F(\cdot)$ denote the joint cumulative probability distribution of the nonzero rainfall intensities during the whole storm duration and $r_u = F^{-1}(u)$ be the quantile level associated with the probability of non non-exceedance u (e.g., $u = 0.5$ to 0.95). At every radar image at time t , we denote by $\mathbf{R}_t[k, l]$ the positive precipitation intensity in [mm/h] at location k, l and define as *rain cells* the set of values \mathbf{R}_u^t over which the intensity exceeds the quantile value r_u ,

$$\mathbf{R}_u^t = \{ \mathbf{R}^t[k, l] \mid \mathbf{R}^t[k, l] \geq r_u \}. \quad (2.1)$$

Likewise, let us denote the elevations underlying those rain cells as,

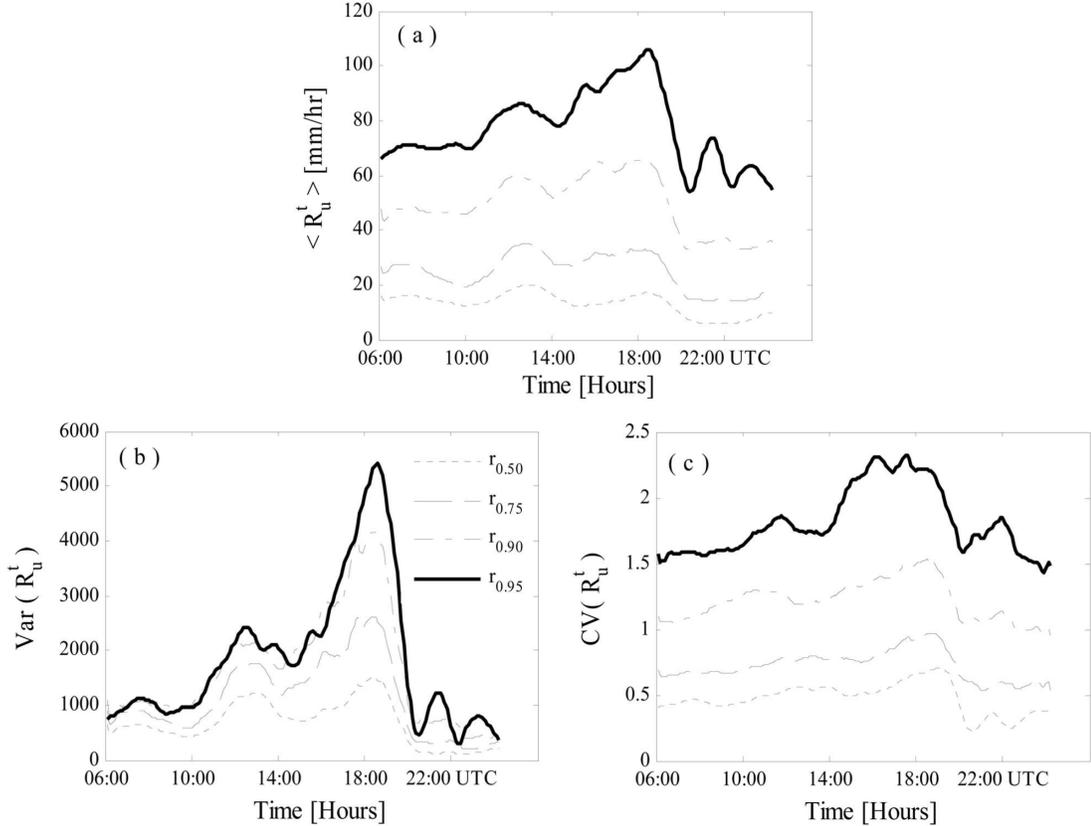


Figure 2.2: Time evolution of the (a) mean and (b) variance of rain cell intensities for different quantile levels. It is observed that while the storm moves toward higher topographic elevations (see Figure 2.1), both the mean and variance of the rain cells increase. It is also noted that the standard deviation cannot be linearly normalized by the mean as reflected in (c) the coefficient of variation, which does not remain invariant during the storm evolution.

$$\mathbf{Z}_u^t = \{ \mathbf{Z}^t [k, l] \mid \mathbf{R}^t [k, l] \geq r_u \}. \quad (2.2)$$

To track the dynamics of the rainfall field's elementary statistics with respect to the underlying topography, the spatial average of the underlying elevations of the rain cells, $\langle \mathbf{Z}_u^t \rangle$, is studied for different quantile levels. It can be observed (see Figure 2.1b) that during the first phase of the storm from 06:00 to 12:00 UTC, as the $\langle \mathbf{Z}_u^t \rangle$ elevation curves decrease in time and collapse into each other for different quantile levels u , the rain cells tend to be distributed over a relatively flat area. However, during the second phase of the storm from 12:00 to 20:00 UTC, the magnitude of $\langle \mathbf{Z}_u^t \rangle$ and its variability increases which means that the rain cells move upslope and get distributed over a region of steeper gradients.

To quantify the signature of the underlying elevation on the statistics of rainfall intensity in time, we also tracked the evolution of the spatial average $\langle \mathbf{R}_u^t \rangle$ and the spatial variance $\text{Var}(\mathbf{R}_u^t)$ of the rain cell intensity for different quantile levels, over the whole storm duration.

The results, reported in Figure 2.2, taken together with the results of Figure 2.2 show that there is a dependency among the tail statistics and underlying elevations of the rain cells. In other words, while the intense rain cells propagate over higher elevations during the storm evolution, the mean and the variance of the intensities of those cells grow in time (see Figures 2.2a and 2.2b), in a way that over high-elevation topography the variance grows even faster than the mean, especially for high quantiles (see Figure 2.2c, which depicts the coefficient of variation). Furthermore, analysis of the spatiotemporal average of rain cell intensities as a function of the underlying mean elevation for a range of quantiles defining the rain cells reveals an unexpected exponential relationship (see Figure 2.3) as,

$$\bar{\mathbf{R}}_u = z_0 \exp(k\bar{\mathbf{Z}}_u), \quad (2.3)$$

where k and z_0 are constants, and $\bar{\mathbf{R}}_u = \mathbb{E}[\langle \mathbf{R}_u^t \rangle]$ and $\bar{\mathbf{Z}}_u = \mathbb{E}[\langle \mathbf{Z}_u^t \rangle]$ represent the ensemble (over all radar images during the storm evolution) spatial average of the rain cell intensities and the underlying elevations for the quantile level u , respectively. This exponential increase depicts the fact that the space time dependence of extreme rain cluster intensity on topography, accelerates exponentially as one considers higher and higher order quantiles. This relationship, which has to be investigated further for other storm systems, can provide a useful predictive relationship for parameterization of the orographic signature on rainfall elementary statistics.

2.4 Multiscale Analysis of Spatial Rainfall

In previous section, rain cells were defined as rainy areas over which rain intensity exceeded a specified quantile threshold and it was found that the underlying topography gives rise to significant impacts on the first and second order statistics of the rainfall fields. By changing the quantile levels, cells of different size (scales) were investigated and it was shown that for the smaller scales (higher quantiles), the within cell rainfall intensities and their variability became larger as rain cells moved toward higher elevations. In this section, using a formal multi-resolution representation in real space, Fourier and wavelet domains, we pursue a different but complementary multiscaling analysis of the precipitation fields with respect to the underlying elevations.

2.4.1 Scaling Manifestations in the Frequency and Real Space Domain

The histogram of the spatial rainfall intensities is often very skewed with a broad dynamic range. To map the narrow range of many low-intensity values into a wider range of output levels, the logarithmic transformation is often an effective approach. To circumvent the

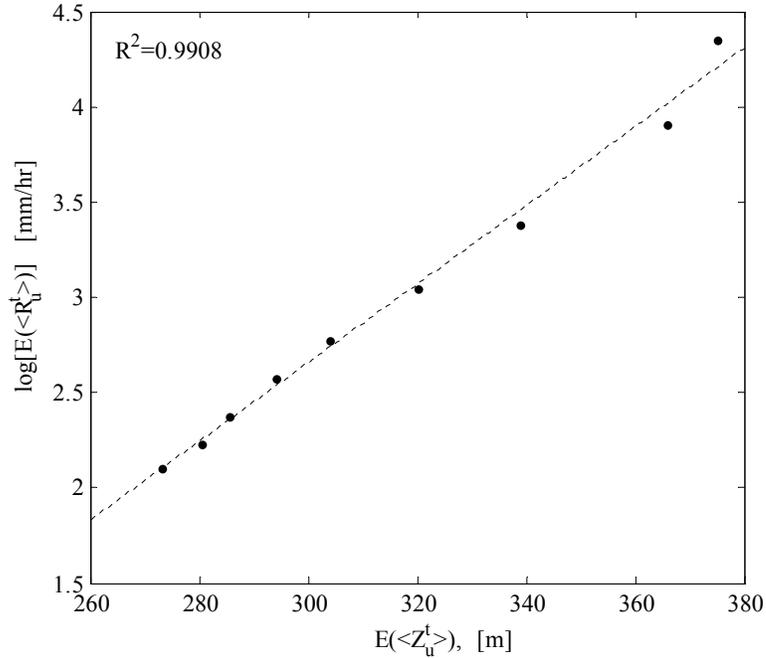


Figure 2.3: A semilog linear relationship indicates that the space time average of the rain cell intensity increases exponentially with respect to their underlying elevations. Every circle represents a quantile level r_u , where u varies from 0.2 to 0.95 in increments of 0.10.

problem of the zero intensity values in rainfall fields, a constant is typically added before the logs are taken, i.e., the transformation $\log(\mathbf{R}_t[k, l] + c)$ is applied. However, since the location of zero intensity values is known anyway, the following invertible transformation is a preferred alternative,

$$\psi^t[k, l] = \begin{cases} 0 & \text{if } \mathbf{R}^t[k, l] = 0 \\ \log(\mathbf{R}^t[k, l]) & \text{otherwise} \end{cases}, \quad (2.4)$$

where the $\log(\cdot)$ denotes natural logarithm. In particular, this transformation allows the entire real axis to be available for statistical computation, it makes the histogram of the field more Gaussian like (e.g., see Figure 2.4f) and as such it facilitates the detection and estimation of scale invariance (*Ruderman, 1994; Huang and Mumford, 1999*). It is worth noting that, as the log function is a monotonic transformation, the overall scaling law of the process would not be distorted in Fourier and real space under this transformation. In the sequel, unless otherwise noted, analyses are performed on the log-transformed rainfall fields according to (2.4).

Knowing that the inner product and norms in $L^2(R)$ are conserved by the Fourier transform (i.e., Parseval theorem), the Fourier expansion has been widely used for detecting the second order scaling laws of geophysical processes. Using the 2D Discrete Fourier Transformation,

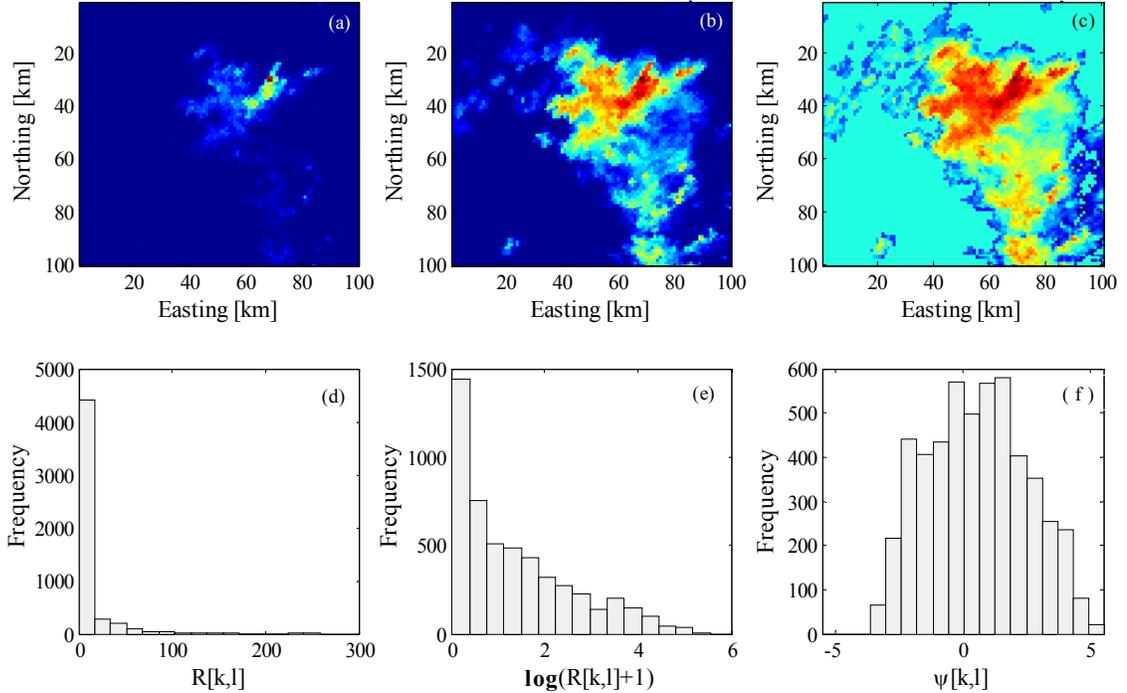


Figure 2.4: (a) Image of the Rapidan storm radar snapshot $\mathbf{R}[k, l]$ at 0800 UTC. (b) Image of the $\log(\mathbf{R}[k, l] + 1)$. (c) Image of the transformed rainfall fields according to equation (2.3). (d, e, and f) The corresponding histograms of the nonzero intensity values, respectively.

the power spectra (i.e., square of the absolute values of the Fourier coefficients) of the $\psi^t[k, l]$ have been computed for each of the radar images over the whole storm duration of 18h (180 radar snapshots at every 6 min). The output of the transformation is rearranged by moving the zero frequency components to the center of the spectrum for visualization purposes. To reduce the measurement noise effect and have a robust estimation in the spectral domain, an ensemble representation is obtained by averaging the spectra of rainfall images over the entire period of the storm. A contour plot of the ensemble power spectral density in the frequency domain is shown in Figure 2.5. It is observed that the storm spectral signature exhibits an anisotropic pattern and interestingly, this shape shows an elliptic regularity with its major axis perpendicular to the Appalachian ridge line (see the dashed line in Figure 2.5) for low frequencies and aligned to the mountain ridge line at higher frequency components. This directional behavior speaks for the organization and elongation of intense rain cells (small scale features) perpendicular to the mountain ridge, an isotropic energy distribution at intermediate scales and a coherency along the prevailing topography at much larger scales.

The 2D spectral density is mapped onto a one dimensional representation via averaging over all angles about the center of the spectrum. As shown in Figure 2.6, for the radial frequency interval $[0.03, 0.50]$ cycles/pixel corresponding to scales of 2 to 32 km in real space, a clear

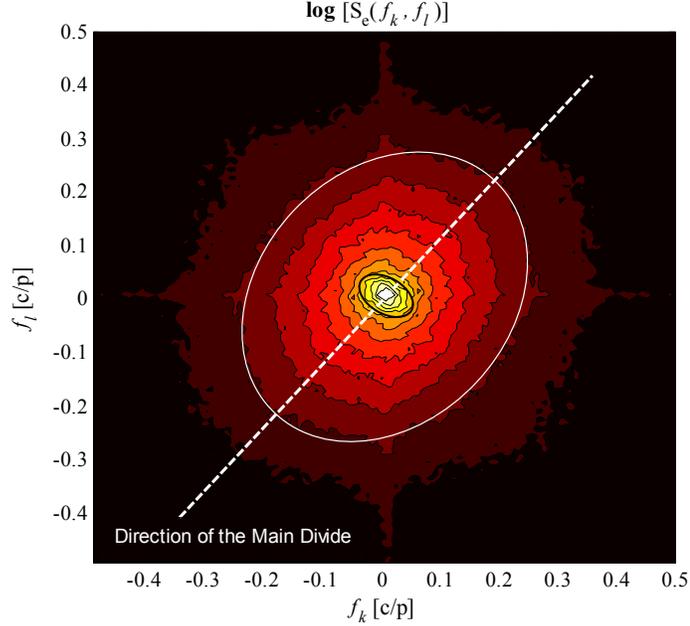


Figure 2.5: Two dimensional ensemble average spectral signature of the Rapidan storm. The power spectrum is centered at frequencies $f_k = f_l = 0$ and divided by 12 equally spaced contour lines. The dashed line displays the approximate direction of the underlying main divide.

log-log linearity of the ensemble spectrum of the rainfall field is observed. This confirms that the spatial distribution of the energy (i.e., variance) over radial frequencies $f_r = \sqrt{f_k^2 + f_l^2}$ admits a power law scaling relationship in the following form:

$$S_R(f_r) \propto \frac{1}{f_r^{\beta_e}}, \quad (2.5)$$

where the slope of the ensemble spectrum is $\beta_e \sim 2.5$.

It is interesting to ask how the spectral slope varies as the storm evolves over the mountain range. For the same frequency interval, the estimated exponents of the directionally averaged spectrum for each of the radar images are plotted versus time and $\langle \mathbf{Z}_{0.95}^t \rangle$ in Figures 2.7a and 2.7b, respectively. It is seen that the spectral slope ranges from 2.1 to 2.7 with a dependency on the underlying topography during the storm period from 0600 to 2000 UTC. In short, while the storm was developing over the low elevation terrain ($\langle \mathbf{Z}_{0.95}^t \rangle < 300\text{m}$), the spectral slope is larger indicating a stronger spatial self coherence (i.e., the energy content drops off faster from larger scale features down to cells of very small size). When the storm evolves over the mountain barrier ($\langle \mathbf{Z}_{0.95}^t \rangle > 400\text{m}$) the energy is distributed more uniformly across different ranges of scales, which implies that the spatial auto-correlation structure of the fields weakens when the orographic effect is more significant. In between (elevations of 300 to 400 m) the storm is in a transitional state with a widely varying

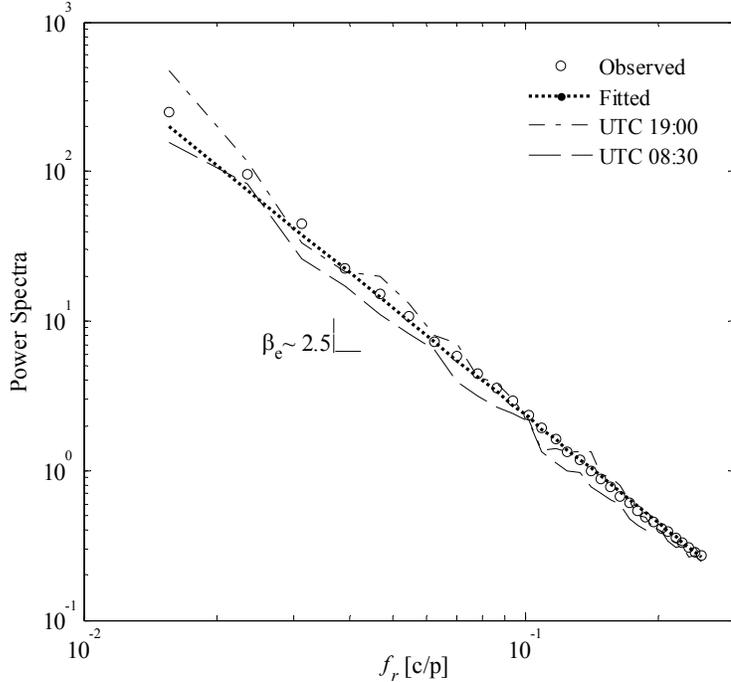


Figure 2.6: Open circles show the one dimensional representation of the rainfall field ensemble average Fourier power spectrum, and the dotted line represents a fitted least square linear model ($R^2 = 0.96$). The angular averaged power spectra of the rainfall fields at two different instants of time are also depicted (dashed lines). The log-log linearity of the power spectrum confirms the presence of second order scaling in the range of 2–32 km.

spectral slope (see Figure 2.1b).

This scaling behavior in the Fourier domain can be translated into real space which is more conducive to understanding the correlation structure associated with the observed power law spectral signature. Defining the 2D auto-correlation function of an image as the expected value of the component-wise product of the image with itself at different spatial translations $[d_k, d_l]$,

$$C [d_k, d_l] = \langle \psi [k, l], \psi [k + d_k, l + d_l] \rangle, \quad (2.6)$$

(Ruderman, 1994) analytically proved that for a field with power law spectrum of slope $\beta > 2$, the directional average of its autocorrelation function has the following form:

$$C (d_r) = c_1 - c_2 d_r^{\beta-2}, \quad (2.7)$$

where c_1 and c_2 are positive constants and $d_r = \sqrt{d_k^2 + d_l^2}$. By expanding equation (2.6), it can be shown that for a field with this correlation structure, the variogram

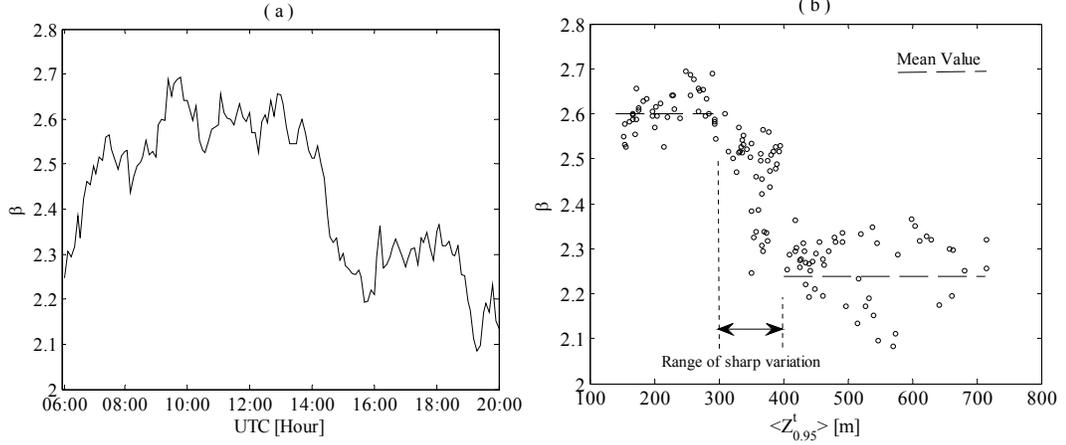


Figure 2.7: (a) Spectral slope of each of the radar images versus time during the storm period and (b) spectral slope versus the rain cell elevation $\langle Z_{0.95}^t \rangle$. Spectral slope decreases as the average underlying elevation of the rain cells increases. This dependence implies that the surface of the rainfall images becomes rougher in a geometrical sense with weaker correlation structure while it is intensified by orographic interaction.

$$\gamma(d_r) = \left\langle |\psi[k, l] - \psi[k + d_k, l + d_l]|^2 \right\rangle, \quad (2.8)$$

also exhibits a similar power law relationship.

The variograms of the rainfall fields are shown in Figure 2.8 in a log-log plot and the estimated slope in the scaling range of $2 \leq d_r \leq 32$ km confirms the spectral analysis in an average sense and shows that the variogram follows a power law scaling with an estimated slope of ~ 0.4 – 0.5 . By inspection, the sum of the auto-correlation function in equation (2.7) with $\beta > 2$ includes a divergent p series which in other words signifies a non-summable auto-correlation function (see, *Beran*, 1994) over radial translations and speaks for the evidence of long range dependence in spatial rainfall intensities.

Up to this point, the second order scale invariance properties of the rainfall field in the Fourier domain and real space have been studied and some connections of this scaling behavior with respect to the underlying elevations have been explored. In the domain of Gaussian self similar random fields this characterization can fully explain the scaling properties of the process. However, if the process is not Gaussian, higher order statistical moments over different scales (resolutions) need to be computed to test the presence of scaling. This can be accomplished by “coarse graining” the field at the original (highest) resolution to obtain fields at successively lower resolutions. Defining the original spatial rainfall as an image of size $N \times N$, a coarse grained field at scale λ can be obtained by passing a non-overlapping averaging filter of size $\lambda \times \lambda$ over the field,

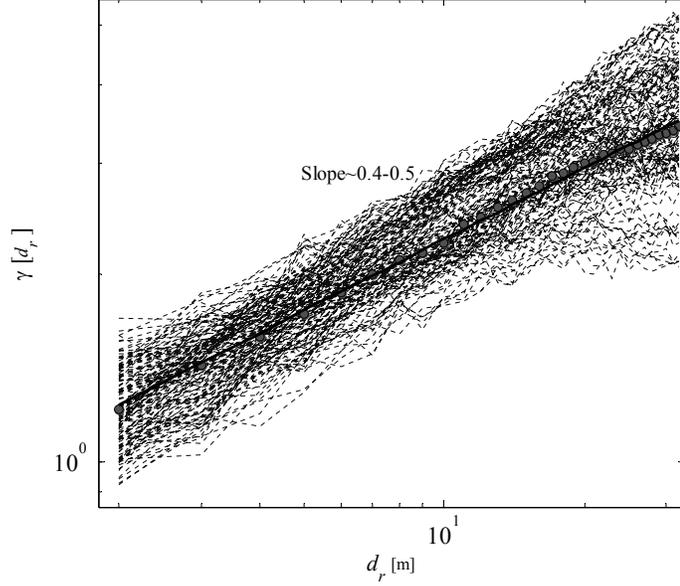


Figure 2.8: Variogram of all storm radar snapshots in a log-log plot. The observed slopes confirm the power law behavior of the spectra (see Figure 2.6) and long range dependence.

$$\psi_\lambda [k', l'] = \frac{1}{\lambda^2} \sum_{k, l}^\lambda \psi [k, l], \quad (2.9)$$

where the $\psi_\lambda [k', l']$ is a field of the same areal coverage as $\psi [k, l]$ but decimated by a factor λ in each dimension. As formation of local intense convective cells imposes a sharp spatial variation on precipitation intensity values, the magnitude of the gradient of a precipitation field implicitly represents the intensity of the rainfall convective activities. It is known that for a strictly self-similar Gaussian field the distributions of the local gradient $\nabla\psi_\lambda [k', l']$ standardized by their corresponding means are independent of λ and collapse into an identical shape of the Rayleigh density function (*Ruderman, 1994*).

For this purpose, we focused on the distributions of the magnitude of the local gradients of the coarse grained fields at different scales λ ,

$$|\nabla\psi_\lambda [k', l']| = \sqrt{\left(\frac{\partial\psi_\lambda [k', l']}{\partial k'}\right)^2 + \left(\frac{\partial\psi_\lambda [k', l']}{\partial l'}\right)^2}. \quad (2.10)$$

Distributions of the local gradients of the rainfall images are presented in log probability for $\lambda \in \{1, 2\}$ km in Figure 2.9. It is seen that the histograms do not collapse on each other, particularly over the tail, which implies that simple scaling may not hold true. The tail is also thicker than the Rayleigh distribution and in general the peaks of the histograms locate before the maximum of the Rayleigh density. The same trend is revealed for larger scales; however, naturally the histograms tend gradually to the shape of the Rayleigh distribution

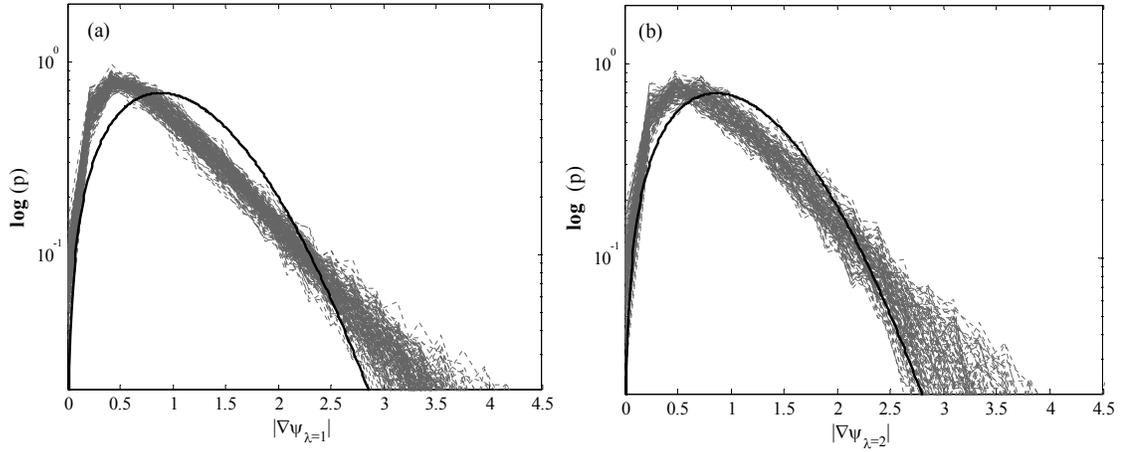


Figure 2.9: Dashed lines represent the distribution of the gradient of the normalized log intensity rainfall fields $\psi[k, l]$ for $\lambda \in \{1, 2\}$ km over the entire period of the storm. Solid line is a Rayleigh distribution fitted to the average histogram. The tail behavior indicates a deviation from Gaussianity.

as the scale increases. Interestingly, it is also observed that even by averaging the field over larger scales (16×16 km) the gradient fields cannot be fully explained in the Gaussian domain. This mainly arises because of the strong spatial correlation of the field, knowing that the central limit theorem holds true only for independent or weakly dependent random variables.

Chapter 3

Statistics of Precipitation Reflectivity Image

To estimate precipitation intensity in a Bayesian framework, given multiple sources of noisy measurements, a priori information about the multiscale statistics of precipitation is essential. In this chapter, statistics of remotely sensed precipitation reflectivity imageries are studied using two different data sets of randomly selected storms for which coincident ground based and spaceborne precipitation radar data were available. Two hundred reflectivity images of independent storm events were collected over two ground validation sites of the Tropical Rainfall Measurement Mission (TRMM) in the United States. Comparing ground-based and spaceborne images, second order statistics of the measurement error is characterized. The average spectral signature and second order scaling properties of those images are documented at different orientations in the Fourier domain. Decomposition of images using bandpass multiscale oriented filters reveals remarkable non-Gaussian marginal statistics and scale-to-scale dependence. Our results show that despite different physical storm structures, there are some inherent statistical properties which can be robustly parametrized and exploited as a priori information for parsimonious multiscale estimation of precipitation fields. A particular mixture of Gaussian random variables in the wavelet domain was found to be a suitable probability model that can reproduce the non-Gaussian marginal distribution as well as the scale-to-scale joint statistics of precipitation reflectivity data, important for properly capturing extremes and the coherent multiscale features of rainfall fields.

3.1 Introduction

In the past decades, a considerable research effort has been devoted to developing parsimonious stochastic models of space-time rainfall. The related theories of multiscale process

representation, e.g., in Fourier or wavelet domains, have proven to be useful for quantifying the rainfall variability at multiple scales. A large body of these developments has exploited the way that the second order statistics of the rainfall process vary across different scales (i.e. $1/f$ spectra). Beyond this, observing non-Gaussian characteristics of precipitation fields and scaling in higher order statistical moments, the theory of Multi-fractals and Multiplicative Random Cascades has extensively been used to capture these distinct properties of the rainfall fields (Lovejoy and Schertzer, 1990; Gupta and Waymire, 1990, 1993). Simultaneously, it has been shown that oriented subband encoding of precipitation fields using wavelets can lead to an efficient and rich multiscale representation of spatial rainfall (Kumar and Foufoula-Georgiou, 1993a,b). Subsequently, an appreciable amount of work has been devoted to extracting the dependency of the parameters of those stochastic models to the underlying physics of the storm (Over and Gupta, 1994; Perica and Foufoula-Georgiou, 1996a,b; Harris et al., 1996; Badas et al., 2006; Nykanen and Harris, 2003).

The purpose of this chapter is to: (1) demonstrate that precipitation reflectivity images exhibit some remarkably regular multiscale statistical characteristics, mainly related to non-Gaussian (heavy tail) marginals and scale-to-scale dependency, and (2) introduce a new modeling framework based on Gaussian Scale Mixtures (GSM) on wavelet trees which can be explored towards non-Gaussian, multiscale multi-sensor data fusion of precipitation fields. In section 3.2, we present basic statistics from a diverse array of precipitation reflectivity images collected coincidentally from ground-based NEXRAD and the spaceborne Precipitation Radar (PR) aboard the TRMM satellite for two TRMM Ground Validation (GV) sites in Texas and Florida. In section 3.3, an extensive analysis and comparison of these images in the Fourier domain is undertaken. In section 3.4, the statistics of these precipitation reflectivity images in the wavelet domain (using an advantageous Undecimated Orthogonal Discrete Wavelet transform) are presented and their marginal and joint statistics are quantified. A novel model based on the GSM on wavelet trees is introduced in section 3.5 and its potential for reproducing the observed heavy tails and covariance of the rainfall wavelet coefficients at multiple scales is demonstrated. The potential application of this model is also briefly discussed.

3.2 Precipitation Data and Elementary Statistics

A major portion of the available remotely sensed precipitation data is acquired via imaging in the microwave band of the electromagnetic spectrum. For active microwave sensors, such as ground or spaceborne radars, the precipitation fields are retrieved via physical or statistical relationship from the reflectivity images obtained as a result of the detected back-scattered energy of microwave signals emitted from the precipitation radar. On the other hand, for passive microwave sensors such as the TRMM Microwave Imager (TMI),

the precipitation fields are retrieved indirectly via conditional inversion of the observed *brightness temperature* (e.g., *Kummerow et al.*, 1996). In this study, we use coincidental reflectivity data of the spaceborne TRMM precipitation radar (PR) and the land-based NEXRAD radar to demonstrate that despite different physical structures of the studied storms, the near-surface images of precipitation reflectivity exhibit remarkably regular and stable statistical properties, which can be explicitly characterized within a novel formalism based on GSM in the wavelet domain.

Specifically, the data set used in this study is populated by near-surface reflectivity images from two hundred independent storms coincidentally observed by TRMM and NEXRAD precipitation radars. The TRMM-2A25 and NEXRAD (level III) long range reflectivity products over two TRMM-GV sites: Houston, Texas (HSTN) and, Melbourne, Florida (MELB), were collected on the basis of the TRMM overpass information provided by the GV Office at the Goddard Space Flight Center, Maryland. Using orthodromic distance, the NEXRAD product provides reflectivity at an horizontal resolution of about 1 km and up to the range of 460 kms with minimum reflectivity detection of 5 dBZ. The TRMM, 2A25 product provides an orbital track that spans a swath of 250 km at nadir with a resolution of about 4-4.5 km and minimum detection sensitivity of 17 dBZ. Note that as the quantitative comparison of the two sensors is of interest in this research, the NEXRAD near-surface long-range reflectivity product was selected to maximize the coincidental coverage between the two sensors. Obviously, rainfall rate estimation from this single level reflectivity product via a Z-R relationship needs to be limited to lower ranges (e.g. < 230 km) to minimize the range effect estimation errors.

The data set used in our study comprises reflectivity images of 95 and 105 storm events from both sensors over the HSTN and MELB sites from 1998 to 2010, respectively (see Figure 3.1). Concerning the sufficiency of the data for robust statistical inference, the images were carefully selected from storm events with adequate aerial coverage during the TRMM overpasses. The data set spans a wide range of storms with different physical structures and geometrical shapes ranging from highly localized convective storms, to frontal and synoptically induced hurricane systems (see Figure 3.2). It is emphasized that no attempt was made to convert these reflectivity images to precipitation intensity values, a task that would be a research topic by itself, given the diversity of storms and the ground-based radar range dependent estimation issues. In the rest of the chapter we refer to these reflectivity fields as *precipitation reflectivity images* or *precipitation images*.

Focusing on characterization of the error variance, these sensors were compared over the intensity range detectable by both. Accordingly, the mean reflectivity (in dBZ) of the TRMM images was compared with the mean of the corresponding NEXRAD images, conditional on reflectivity values exceeding 17 dBZ; see Figure 3.3 and 3.4. For this case, the standard bias was found to be -2% and -1.8% for the HSTN and MELB sites, respectively. This indicates

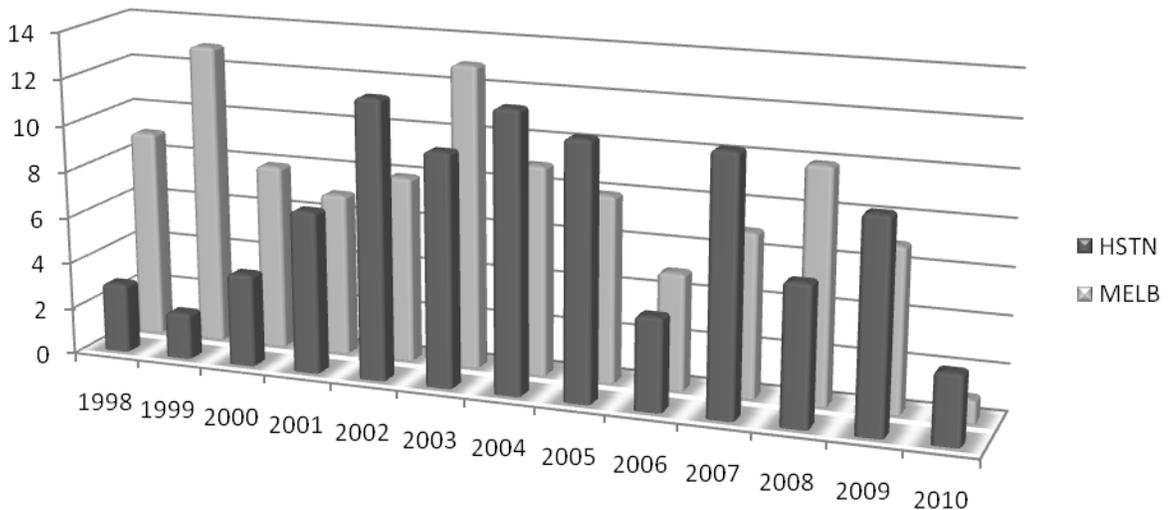


Figure 3.1: For statistical analysis, a total of 200 independent storms were selected from the TRMM-PR and the NEXRAD reflectivity dataset at the TRMM ground validation (GV) sites of Houston, TX (HSTN) and Melbourne, FL (MELB). The distribution of these events by year is shown above.

that the TRMM-PR overestimates the reflectivity intensity in the range that both of the sensors can detect reflected echoes. This bias is not unexpected and is mainly due to the inherent differences in the way that the two sensors interrogate the vertical profile of the atmosphere. The variance of error is estimated and reported in Table 3.1 based on two different definitions of signal-to-noise ratio metric. This characterization has an important implication in the context of linear multisensor fusion of precipitation products (e.g., *Chou et al.*, 1994; *Gorenburg et al.*, 2001; *Tustison et al.*, 2003). To this end, the bias was adjusted to zero via enforcing the regression line to pass through the origin and also the data pairs with normalized residual values (by the standard deviation) beyond the interval $[-2, 2]$ were excluded from the estimation process (see Figure 3.4). The latter treatment makes the estimation more robust to probable outliers.

The Kullback-Leibler (KL) divergence, also *known* as the relative entropy, was also studied to characterize the degree of proximity of the marginal densities of the observations, provided by the two sensors. The KL divergence is defined as:

$$\mathcal{KL}(p_1|p_0) = \sum_x \log \left(\frac{p_1}{p_0} \right) p_1 \quad (3.1)$$

where $p_j = p(x|\mathcal{H}_j)$ is the conditional marginal density of the precipitation reflectivity values under different measurement hypotheses with $j = 0, 1$ corresponding to TRMM and NEXRAD observations, respectively. The \mathcal{KL} divergence is a positive quantity which is equal to zero if and only if the compared densities are equal almost everywhere in their

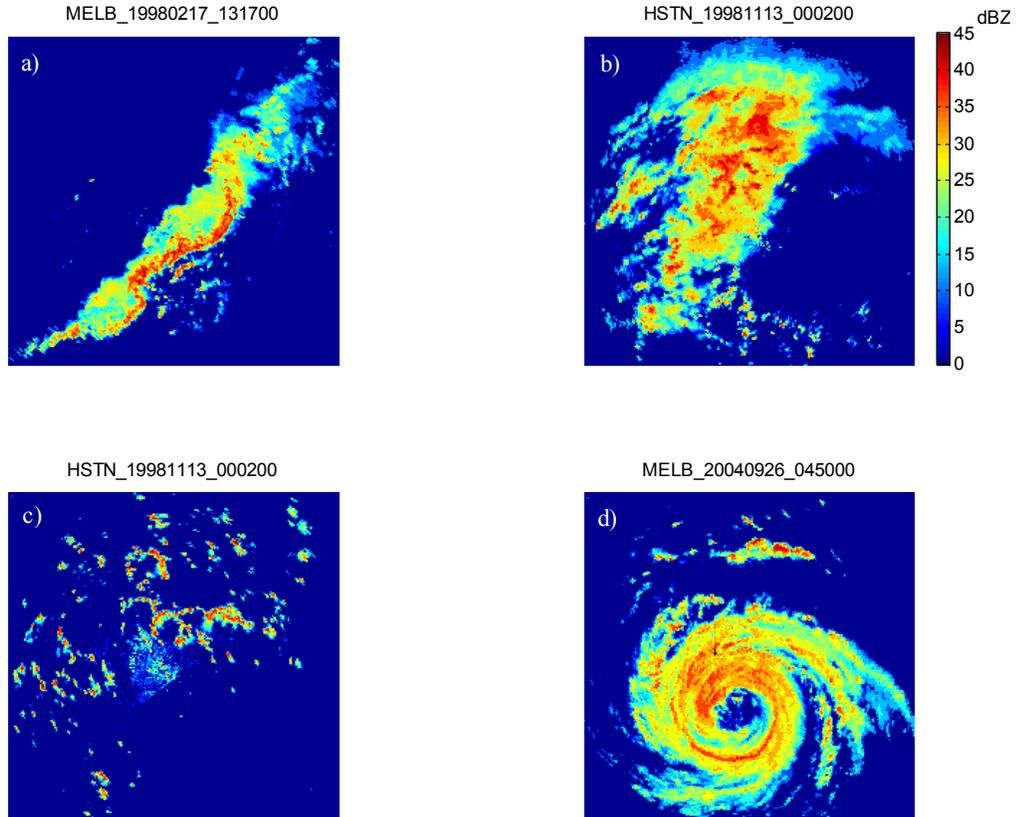


Figure 3.2: The collected datasets summarized in Figure 3.1 span a wide range of storms with different spatial structures and geometrical shapes. The NEXRAD reflectivity images for four selected storms are shown above; they are labeled according to the GV site, date and time in UTC.

domain. The \mathcal{KL} is not a conventional distance since it is not symmetric and does not satisfy the triangle inequality for three arbitrary densities. Yet, it has been shown to be a useful measure of density mismatch in statistical modeling (*Levy, 2008*). As can be seen from Table 3.1, this metric demonstrates a statistically significant deviation from zero for both GV sites implying a deviation of the marginal densities of reflectivity values provided by the two sensors. This particular observation along with the least squares analysis of the data set indicates that on the average the overall quality of the selected TRMM-PR overpass observations in the MELB site is superior to that of the HSTN site.

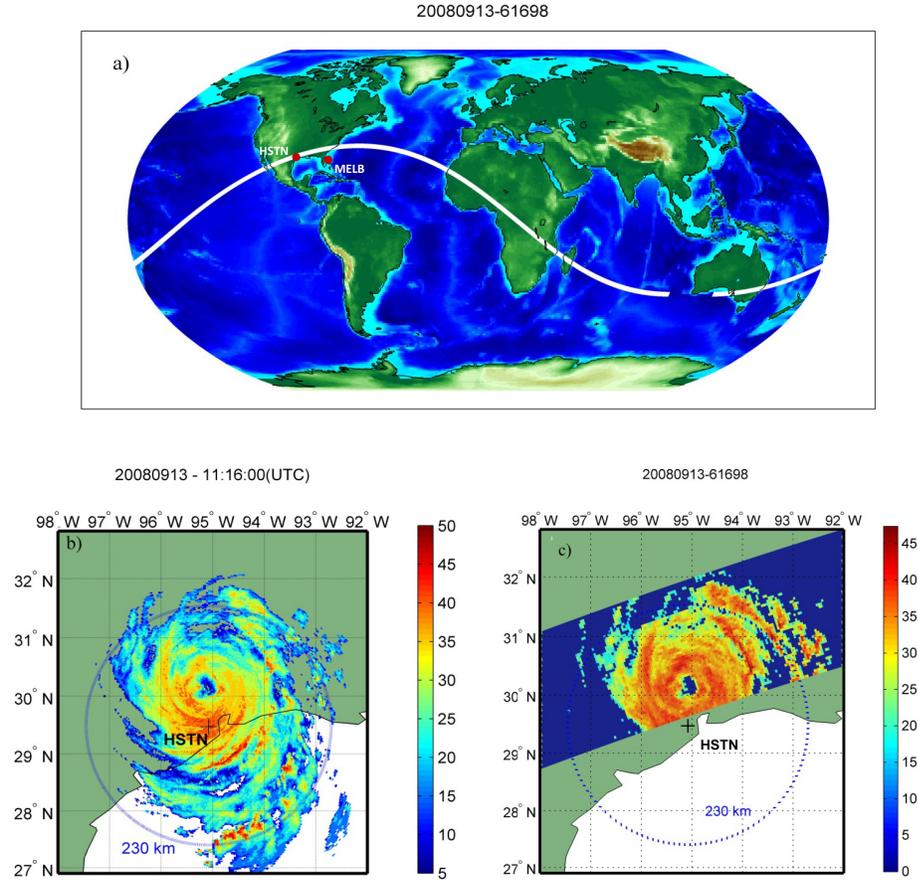


Figure 3.3: (a) The geographic locations of the study sites (MELB and HSTN) and the orbital track No.61698 of the TRMM satellite which captured a hurricane storm over Texas On 09/13/2008; (b,c) Reflectivity images of the storm captured by ground-based NEXRAD at 11:16:00 UTC and the coincidental TRMM-2A25 overpass, respectively.

| | | HSTN | | MELB | |
|------------------|--|--------------------------|-----------------|---------------------------|-----------------|
| | | NEXRAD | TRMM | NEXRAD | TRMM |
| SNR ₁ | | 11.9(10.4-12.9) | 13.0(11.6-13.6) | 12.4(11.2-13.6) | 13.6(13.0-14.4) |
| SNR ₂ | | 8.4(5.8-9.75) | 13.0(11.6-13.6) | 9.0(7.45-10.0) | 7.9(6.5-9.6) |
| \mathcal{KL} | | 1.0488 (0.7959 - 1.5681) | | 0.6749 (0.6108 - 0.7494) | |

Table 3.1: Standardized error variance in terms of two signal-to-noise ratio (SNR) metrics and Kullback-Leibler divergence of the marginal histograms of the TRMM and NEXRAD coincidental reflectivity observations. Values in parentheses indicate the 95% quantile range of estimation. The two different metrics of SNR are: $SNR_1 = 10 \log_{10}(\mu_s/\sigma_n)$ and $SNR_2 = 10 \log_{10}(\sigma_s/\sigma_n)$, where μ_s and σ_s are the mean and standard deviation of the signal and σ_n is the noise standard deviation.

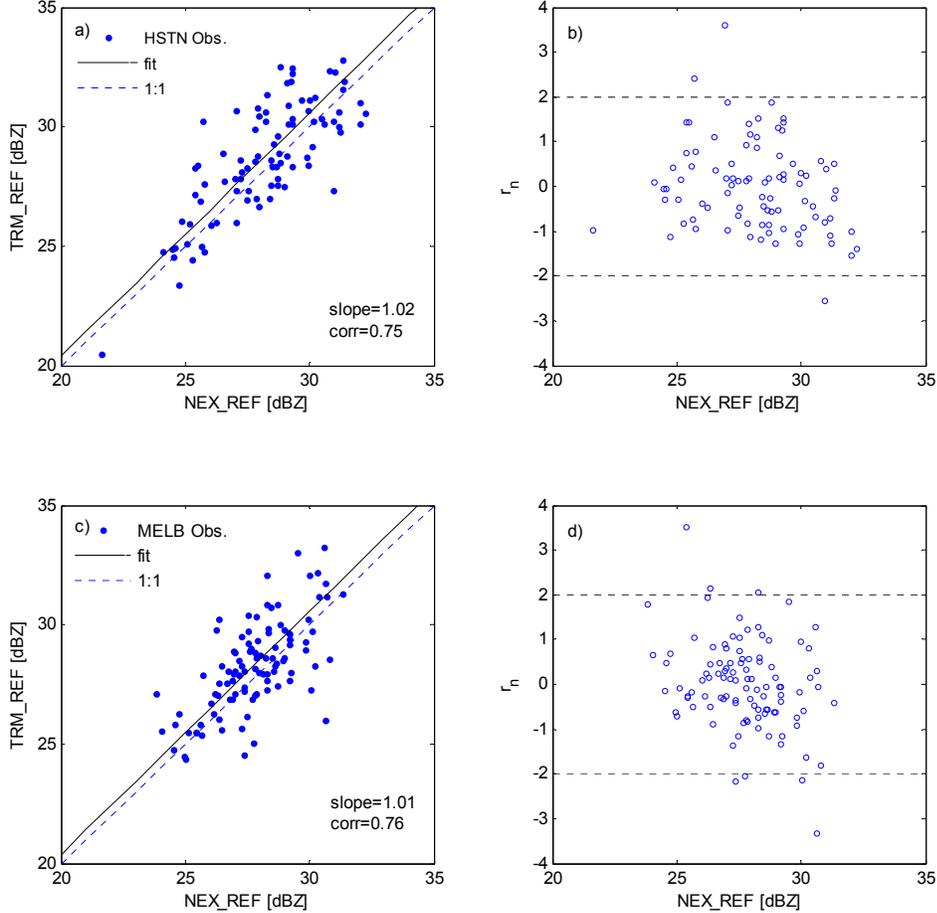


Figure 3.4: (a,c) TRMM versus NEXRAD reflectivity values in HSTN and MELB sites, respectively. The data pairs are the spatially averaged reflectivity values of coincidental pairs of images computed over the range of intensity values which is detectable by both sensors (≥ 17 dBZ). The solid line is the best least squares fitting and the broken line is the 1:1 line. The corresponding normalized regression residuals are shown in (b,d), with the $[-2, 2]$ lines marked to indicate the values that fall outside the ± 2 times standard deviation of residuals.

3.3 Spectral Signature

Several studies (e.g., *Lovejoy and Schertzer, 1990; Harris et al., 1996, 2001*, among others) have reported the presence of scale-invariance in the form of $f^{-\beta}$ average Fourier spectrum (i.e. $\mathbb{E}[|\mathcal{F}(f)|^2]$) in precipitation fields. The Fourier transformation, as an approximation to the Karhunen-Loève expansion, allows us to decouple the correlation structure of the rainfall fields into a set of almost uncorrelated Fourier coefficients with a nearly diagonal covariance matrix. Therefore, knowing that the inner product in $\mathbf{L}^2(\mathbb{R})$ is conserved under the Fourier transformation (i.e. Parseval's Theorem), the one dimensional representation of the average power spectrum $\mathbb{E}[|\mathcal{F}(f)|^2] = Af^{-\beta}$ is indeed diagonalization of the covariance in the frequency domain. Besides the information content of the spectral decay rate as

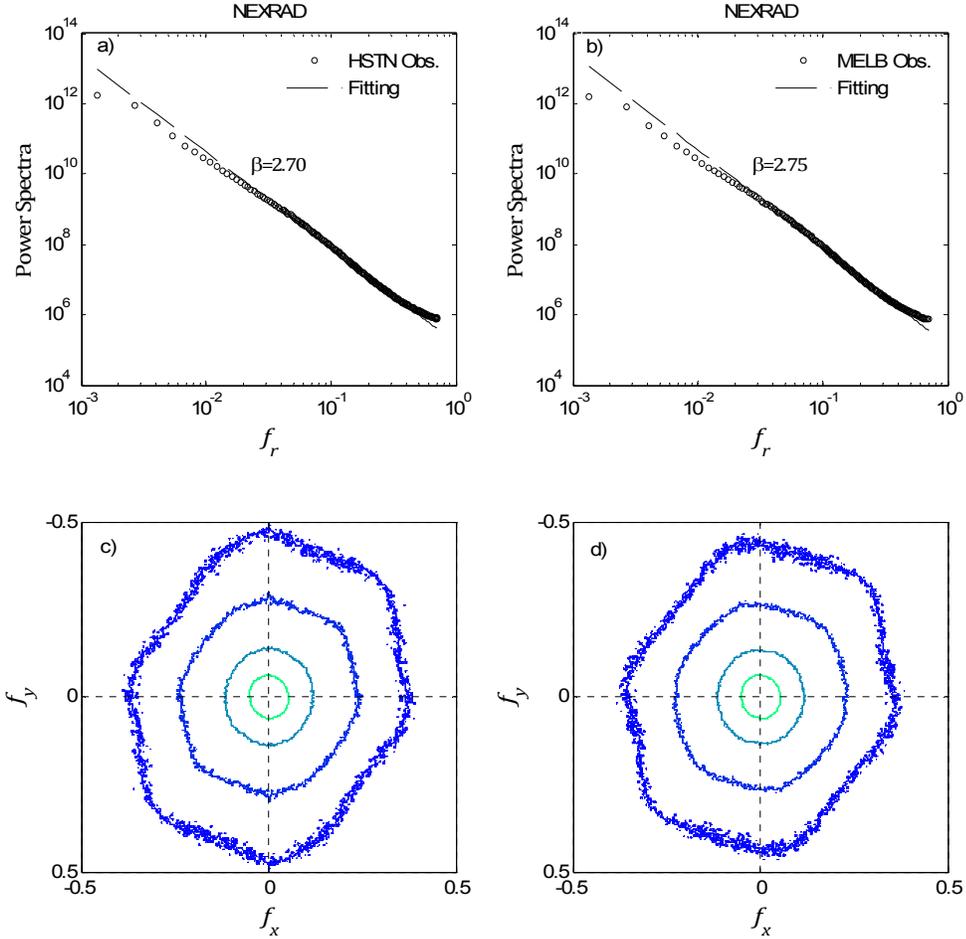


Figure 3.5: (a, b) Radially averaged spectra for the ensemble of NEXRAD reflectivity images at the Houston and Melbourne GV sites, respectively. (c,d) 2D ensemble spectra for the same datasets depicting directional anisotropy at small scales (large frequencies).

depicting the second order scaling law and degree of differentiability (smoothness) of a field, this diagonal representation of the covariance yields a computationally more efficient least square optimal filtering in the Fourier domain which also might be useful for filtering of high dimensional strongly correlated rainfall fields.

By construction, the Fourier spectrum of an image is insensitive to spatial translation, but it is not rotation invariant and can explain the anisotropy of a field. Accordingly, in addition to the energy distribution of the intensity values in the frequency domain, the 2D spectrum depicts the orientation of the edges and regions of sharp gradients in a 2D field. For instance, it has been reported that as horizontal and vertical edges are dominant in man-made scenes (e.g. cities), the spatial distribution of the spectrum of these images is more elongated along the vertical and horizontal orientations. In light of this, studying the spatial distribution, orientation and total energy of the precipitation reflectivity images in the spectral domain might be useful not only for exploring scale invariance and optimal

estimation but also for studying the regional organization of storm systems for retrieval applications.

To this end, we compute here a more general representation of the mean spectral signature of the precipitation images at different orientations θ ,

$$\mathbb{E} [|\mathcal{F}(f, \theta)|^2] = A(\theta)f^{-\beta(\theta)} \quad (3.2)$$

where $\mathcal{F}(\cdot)$ denotes the Fourier transformation in polar coordinates, $A(\theta)$ is a prefactor and $\beta(\theta)$ is the drop-off rate of the spectrum at angle θ . Using discrete Fourier transform, the square of the absolute values of the Fourier coefficients were calculated to obtain the 2D power spectrum for each individual image. This provides a set of Fourier power spectra which can be averaged over the entire dataset for each site (i.e. 95 images over HSTN and 105 images over MELB) to obtain the so called ensemble power spectrum, see Figure 3.5c,d. Using the least squares regression in a log-log scale, the power spectral model in the form of equation (3.2) can be fitted at different orientations to each individual or ensemble 2D spectra. The regressions were performed in the radial frequency interval of $[0.03, 0.50]$ cycle/pixel [c/p] corresponding to the pseudo spatial scale (i.e. Euclidean distance) of 2-32 km. Table 3.2 reports the results of directional estimation of power spectral slopes for the NEXRAD datasets. It is observed that the estimated spectral slopes vary between 2.35 to 2.75 for the HSTN site and between 2.45 to 2.85 for the MELB site. By averaging a 2D power spectrum over all angles, a one dimensional representation of the spectrum can also be obtained in which the parameters in equation (3.2) are independent of orientation. Figure 3.5a,b shows the radially averaged ensemble spectra for the NEXRAD datasets. The estimated drop-off rate of the radially averaged ensemble spectra in two sites is about 2.70-2.75, which implies that the precipitation images are globally much smoother than many other natural images (e.g., $\beta = 2$) (*Ruderman, 1994*). Interestingly, despite the different geographic locations of the two sites and different physical structures of the storms, the statistics of the spectral parameters vary within a very narrow range (Table 3.2). This observation implies that the spectral signature (i.e. spatial correlation structure) of the near-surface reflectivity images may not be a discriminatory measure of the physical structure of the storms. On the other hand, this universal behavior gives us a priori knowledge about the correlation structure of these type of rainfall images which can be useful for noise removal and optimal estimation of precipitation data in the Fourier domain.

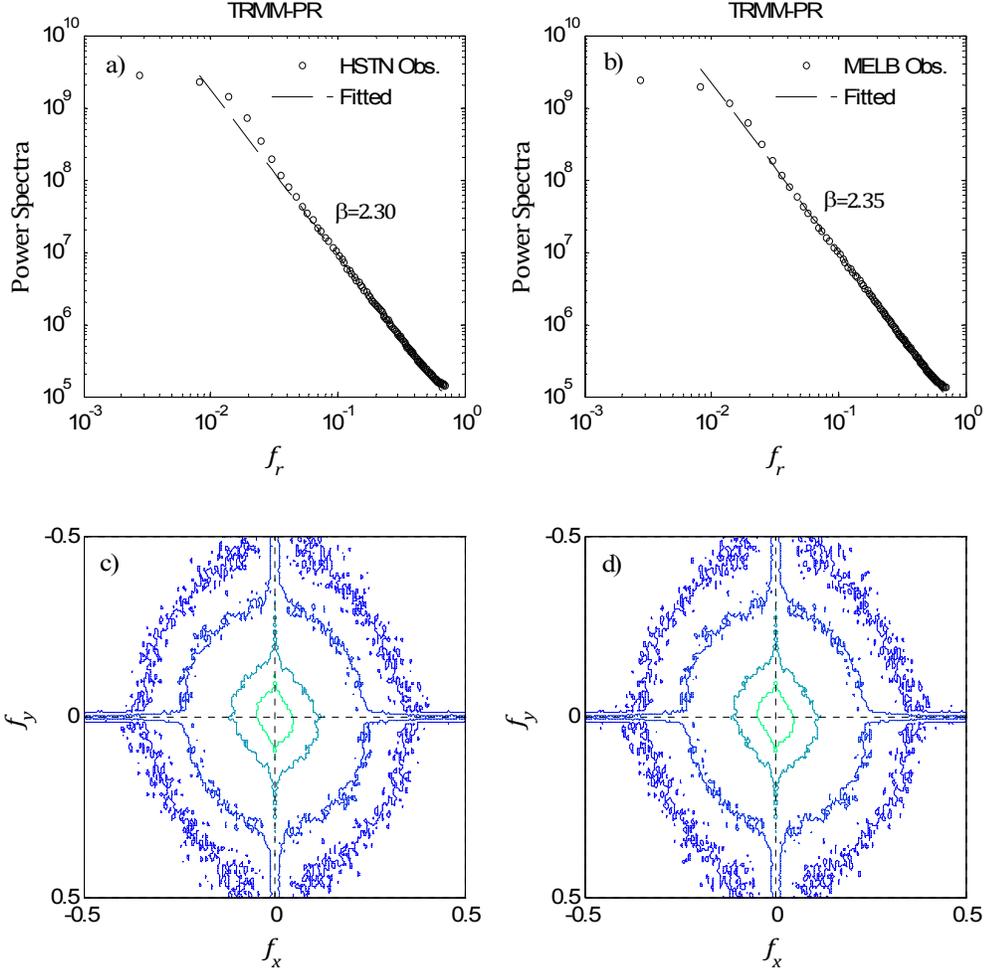


Figure 3.6: (a, b) Radially averaged spectra for the ensemble of TRMM-PR reflectivity images in the Houston and Melbourne GV sites, respectively. The smaller values of the spectral slopes compared to those of the corresponding NEXRAD reflectivity images of Figure 3.5 imply that the reflectivity images produced by the spaceborne TRMM-PR exhibit a weaker spatial correlation than those produced by the ground-based NEXRAD. (c, d) 2D ensemble spectra of the TRMM-PR data at HSTN and MELB sites, respectively, indicate similarity to the corresponding NEXRAD spectra but also show significant spectral leakage due to the artificial edge effects introduced by the swath boundaries (see text for more explanation).

As mentioned before, the shape of the spectrum can also speak for the regional organization of the rain-cells. Pronounced abrupt changes in the spatial domain intensity values (i.e. horizontal edges) cause spectral skewness (elongation) in the perpendicular direction (i.e. vertical direction) at the frequency domain (*Gonzalez and Woods, 2008*). In the collected storm images, for the low frequency components of less than $0.1 [c/p]$, the spectral signature shows a more dense and isotropic behavior, meaning that on average the large scale features of the storms do not have any particular spatial orientation. However, for high frequency components (i.e. small scale features) the ensemble power spectra are tilted and

more elongated towards the North-East (NE) or South-West (SW) directions (see Figure 3.5c,d). This similar asymmetric signature in both sites may mainly arise due to a regionally governing synoptic meteorological condition that gives rise to a directionally dominant formation of the rain patches with a length scale smaller than 10 km.

| HSTN | | MELB | | |
|----------------------|--------------------|-----------------|--------------------|-----------------|
| | $\log [A(\theta)]$ | $\beta(\theta)$ | $\log [A(\theta)]$ | $\beta(\theta)$ |
| $\bar{\theta}$ | 5.18(0.18) | 2.68 (0.12) | 5.11(0.19) | 2.75(0.10) |
| $\theta = 0.0^\circ$ | 4.86(0.22) | 2.66 (0.22) | 4.80(0.27) | 2.72(0.26) |
| $\theta = 90^\circ$ | 5.02(0.25) | 2.71 (0.27) | 4.92(0.25) | 2.80(0.21) |
| $\theta = 135^\circ$ | 4.50(0.19) | 2.75 (0.22) | 4.46(0.22) | 2.85(0.21) |
| $\theta = 45^\circ$ | 4.85(0.20) | 2.35 (0.20) | 4.75(0.19) | 2.45(0.19) |

Table 3.2: Estimated parameters of the spectral model in (3.2) for NEXRAD data at multiple directions ; the values in parentheses are the standard deviations. The parameters reported for $\bar{\theta}$ denote those obtained from the radially averaged ensemble spectra.

Due to the limited swath width and flight orientation, the TRMM-PR orbital observations often cannot capture the entire spatial extent of the storm events. TRMM-PR products often provide a cropped version of the whole storm with abrupt changes of intensity values on the swath boundaries. These artificial edges contaminate the spectral signature and give rise to some spectral leakages (see Figure 3.6), which do not allow us to properly study the directional organization of the rain-cells (i.e. edges) from this product. Although this boundary effect may be handled, for instance by padding the TRMM-PR images with mirror reflection of themselves across the boundaries, this obviously does not add any new information that can be exploited to study the spatial organization of the rain-cells. However, the decay rate of the radially averaged ensemble spectrum at similar frequency bands confirms that the reflectivity images of the TRMM sensor exhibit slightly weaker correlation structure (more irregular) compared to the NEXRAD products.

3.4 Statistics of Subband Components in the Wavelet Domain

Natural processes exhibit variability over a broad range of scales, often manifesting itself in isolated singularities in the form of edges or nested areas of intense activity. The decay of the Fourier spectrum captures the global distribution of variance over scales without providing information about the local distribution of the process variability at different scales. Using a set of multiscale bandpass filters at different orientations has been found to be extremely useful for extracting the information content of the local jump discontinuities and abrupt fluctuations of these fields (e.g., *Kumar and Foufoula-Georgiou, 1993a; Perica and Foufoula-Georgiou, 1996b; Huang and Mumford, 1999*). Spatial precipitation fields are highly clustered and exhibit strong correlation along with sparseness (zeroes) in

the real domain, mainly recognizable as the presence of oriented edges between rain and no-rain areas. Consequently, precipitation images often exhibit a stronger sparseness condition in the wavelet domain as the coherent cells and broad homogeneous areas of the precipitation images would map into near-zero wavelet high-pass coefficients. This often manifests itself in the marginal histogram of the wavelet subbands having a sharp peak at the center (i.e., around zero) and extended heavy tails which cannot be modeled in a Gaussian framework. As a simple treatment to overcome this leptokurtic behavior, *Perica and Fofoula-Georgiou (1996b)* proposed a Gaussian density model for the so-called “standardized rainfall fluctuations”, defined as the high-pass orthogonal wavelet coefficients divided by their corresponding low-pass coefficients. Although, that treatment can partially model the observed thick tail behavior, it cannot flexibly account for the cusp singularity or large mass of the wavelet coefficients around the center of the distribution.

In this study, we demonstrate that the Generalized Gaussian (GG) density which has been widely used for statistical modeling of the high-pass wavelet subbands of natural images (e.g., *Huang and Mumford, 1999*) can be employed to fully characterize the marginal statistics and heavy tail properties of the precipitation reflectivity images. As the rainfall imageries generally suffer from a considerable number of zero intensity values at the background (non-rainy areas within the field of view), a method is also presented to allow characterization of only the “relevant zeroes”, i.e., the zeroes that correspond to small isolated dry areas within the storm domain and to those corresponding to the storm edges but not the background zeroes (see, *Kumar and Fofoula-Georgiou, 1993b*). In addition, it is shown that despite the decorrelation capacity of the wavelet transformation (*Wornell and Oppenheim, 1992*), the wavelet coefficients of the rainfall images exhibit a weak correlation structure and a considerably regular higher order scale-to-scale dependence, which needs to be addressed for proper multi-resolution modeling of precipitation imageries.

3.5 Wavelet Decomposition and Marginal Statistics

The Orthonormal Discrete Wavelet Transformation (OWT) (*Mallat, 1989*) decomposes a 2D-signal $f(x, y)$ of size $K \times L$ into a pair of almost uncorrelated expansion coefficients $d_{m,k,l}^i$ (called wavelet coefficients) and orthonormal basis functions $\psi_{m,k,l}^i(x, y)$ in the form of

$$\psi_{m,k,l}^i(x, y) = 2^{m/2} \psi^i(2^m x - k, 2^m y - l) \quad (3.3)$$

$$f(x, y) = \frac{1}{\sqrt{KL}} \sum_{i=H,V,D} \sum_{m=-\infty}^{m=+\infty} \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} d_{m,k,l}^i \psi_{m,k,l}^i(x, y) \quad (3.4)$$

where $\psi_{m,k,l}^i(x, y)$ is the wavelet basis function at subband $i = \{H, V, D\}$ (i.e. Horizontal, Vertical and Diagonal directions in this study), scale m and translation indices (k, l) .

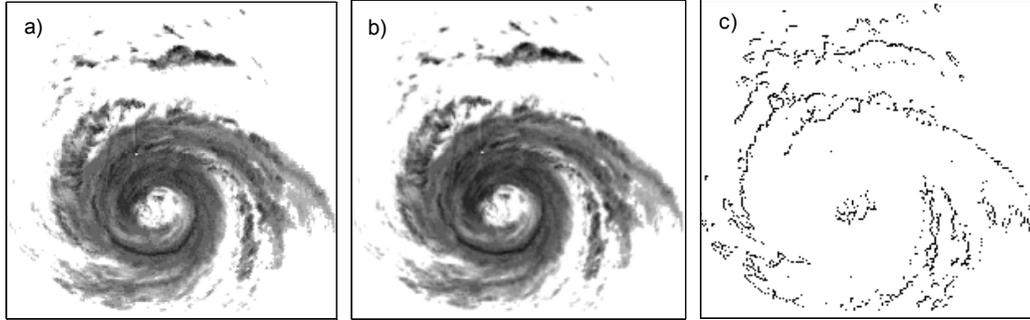
This representation uses the orthogonal wavelet bases functions $\psi_{m,k,l}^i(x, y)$ in a “critically-sampling rate” (meaning that the size of the input signal N is equal to the total size of the output subbands). Owing to the orthogonality of the bases and critical sampling rate, the inverse transformation allows a perfect reconstruction with a computational complexity of the order of $\mathcal{O}(N)$. However, this critical sampling rate makes the wavelet representation shift-variant and imposes significant aliasing in each individual subband. Although, the aliasing artifacts will cancel out in the reconstruction phase, this would be troublesome for processing and parametrization of each individual subband (*Nason and Silverman, 1995; Simoncelli and Freeman, 1995*).

In this study, a shift-invariant Undecimated Orthogonal Discrete Wavelet Transform (UOWT) (*Nason and Silverman, 1995*) is used for decomposition of the precipitation reflectivity images and statistical characterization of their wavelet coefficients. This decomposition produces nearly alias-free and overcomplete subband information. The latter property is another great advantage over the conventional OWT in which the size of the signal is downsampled by a factor of two at each level of decomposition, giving rise to inferential problems in subband parametrization of rainfall images with small wetted area. Obviously, the advantages of this overcomplete frame expansion, come at the expense of a higher computational complexity of the order of $\mathcal{O}(N \log N)$.

It is noted that in the wavelet domain, background zeroes will remain zeroes at consecutive scales in both high and low-pass subbands, while a range of zero intensity values within the storm domain (i.e. those zero intensity pixels which define the boundaries of the wetted areas of the storm from the background zeros) will become non-zeroes from fine-to-coarse scales. This observation provides an efficient means of eliminating the background zeroes while keeping the zeroes of interest (and their locations). In this study, to resolve this issue, the conditional marginal densities of high-pass subbands are estimated given that the low-pass coefficients at the same location and scale are positive (see Figure 3.7).

Figure 3.8b shows the marginal distribution of the wavelet coefficients (in log-probability scale) computed, as discussed above, from the NEXRAD reflectivity data of a June, 2007 storm over the Melbourne site (see Figure 3.8a). A highly leptokurtic behavior is observed which is in contrast, for example, with the Gaussian marginal distributions of the wavelet coefficients of a fractional Brownian surface, as shown in Figure 3.8c,d (note that a Gaussian density is an inverted parabola in a log-probability scale). It is important to note that both the 2D fractional Brownian surface (the slope of its power spectrum is $2H + 2$, where $0 < H < 1$ is the self-similarity index) and the precipitation image exhibit similar power law spectrum (i.e. $1/f$ law) but their marginal statistics are drastically different.

The Generalized Gaussian (GG) family, also known as the Generalized Laplace, has often been used to model the marginals of the wavelet coefficients in the context of natural images (*Huang and Mumford, 1999*). The early form of this class of density functions was



Size: [544, 624]

Figure 3.7: (a) NEXRAD reflectivity image of a storm over the HSTN site at 01/15/2007 at original resolution of 1 km; (b) first level low-pass subband image of the storm using an Undecimated Orthogonal Wavelet Transform (UOWT); (c) pixel-wise difference of images (b) and (a). At each level of the wavelet decomposition, as a result of the convolution of the field with the wavelet scaling function, a set of zero values next to the edges of the wetted areas becomes non-zero and subsequently the areas of non-zero pixels progressively grow from fine-to-coarse scales.

first presented by *Subbotin* (1923). The zero-mean parameterization of this family can be described by a shape $\alpha \in (0, \infty)$ and a width parameter $s \in (0, \infty)$ as

$$f_x(x, \theta) = \frac{\alpha}{2s\Gamma(1/\alpha)} \exp\left(-\left|\frac{x}{s}\right|^\alpha\right), \quad (3.5)$$

where, $\theta \in \{s, \alpha\}$ and $\Gamma(\cdot)$ denotes the standard gamma function $\Gamma(a) = \int_0^\infty e^{-t} t^{a-1} dt$, $a > 0$ (*Nadarajah*, 2005). This parameterization allows a concise characterization of a symmetric probability continuum spanning a wide range of distributions, including the Dirac delta function ($\alpha \rightarrow 0$) to the uniform density ($\alpha \rightarrow \infty$) in the limiting case. The tail probability of this family is summable and admits the classical central limit theorem. Therefore, it is only a suitable probability model for signals with finite energy in $\mathbf{L}^2(\mathbb{R})$ inner product space. It is worth noting that as this family is a subclass of the generalized gamma density function. *Martín and Pérez* (2009) proposed a gamma mixture representation which allows a pseudo random number generation scheme for this family. Letting \mathcal{G} be a gamma random variate $\mathcal{G} \sim \Gamma(\text{shape} = 1 + 1/\alpha, \text{scale} = 1)$, the GG random variables in equation (3.5) can be generated by $X \sim s\mathcal{G}^{1/\alpha}\mathcal{U}$, where \mathcal{U} is a uniform density function on $[-1, 1]$.

Given a set of sample wavelet coefficients $\mathbf{d}_j \in (d_{m,1}^i, d_{m,2}^i, \dots, d_{m,n}^i)^T$ of precipitation reflectivity images at subband i and scale m , where $j = 1, 2, \dots, n$ correspond to the spatial locations of these coefficients, the parameters of the fitted GG probability density can be estimated using the Method of Moments (MOM) or the Maximum Likelihood (ML) estimation method. The density of the GG distribution in the form of equation (3.5) can be fully characterized given the second and fourth order central moments of the sampled data:

$$\mathbb{E}[\mathbf{d}_j^2] = \frac{s^2\Gamma(3/\alpha)}{\Gamma(1/\alpha)}$$

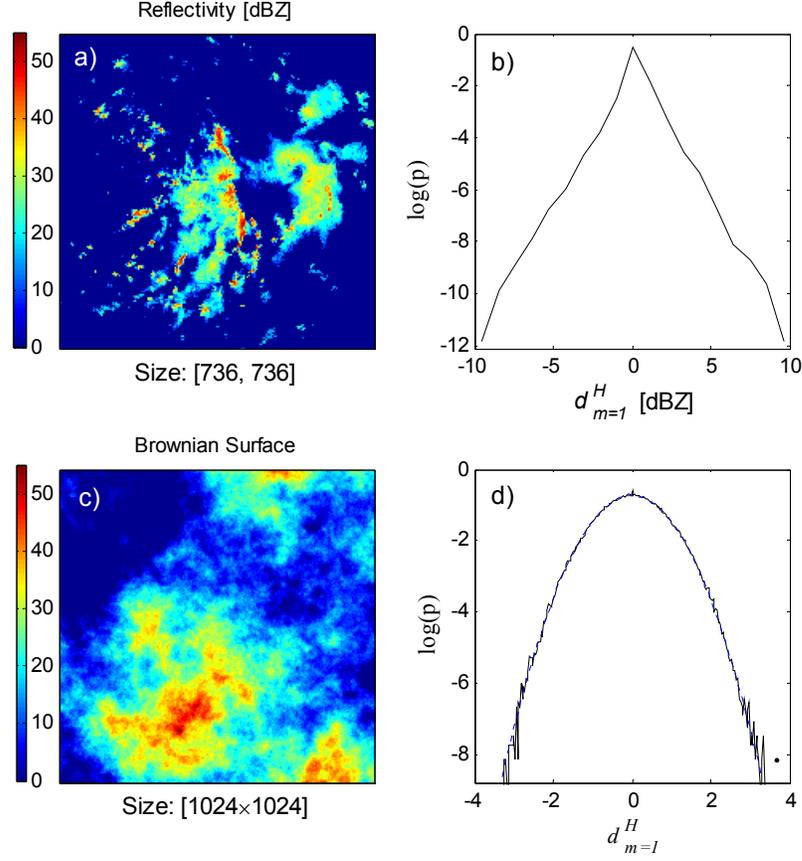


Figure 3.8: (a) NEXRAD reflectivity image of a storm over the MELB site on 07/05/2007 at 20:00:00 UTC, and (b) the associated histogram of the wavelet coefficients normalized by the standard deviation. Sharper peak and heavier tail than the Gaussian case is a typical statistical feature of the rainfall images in the wavelet domain. (c) Positive part of a 2D fractional Brownian surface with self-similarity index of 0.5 and (d) the associate Gaussian marginal histogram of the normalized horizontal wavelet high-pass coefficients.

$$\mathbb{E} [\mathbf{d}_j^4] = \frac{s^4 \Gamma(5/\alpha)}{\Gamma(1/\alpha)} \quad (3.6)$$

Accordingly, the shape parameter α can be estimated from the sample kurtosis of the wavelet coefficients as defined above, by numerically solving the following nonlinear equation,

$$\kappa [\mathbf{d}_j] = \frac{\Gamma(1/\alpha) \Gamma(5/\alpha)}{\Gamma^2(3/\alpha)} \quad (3.7)$$

and knowing the shape parameter, the width parameter can be estimated using equation (3.5). A closed form set of equations is also derivable to estimate the parameters in a ML sense. Specifically, maximizing the log-likelihood function

$$\theta_{ML} = \underset{\theta}{\operatorname{argmax}} \left\{ \log \left[\prod_{j=1}^n f(\mathbf{d}_j, \theta) \right] \right\} \quad (3.8)$$

yields the following set of nonlinear equations that can be solved numerically,

$$-\frac{n}{s} + \frac{\alpha}{s^{\alpha+1}} \sum_{j=1}^n |\mathbf{d}_j|^{\alpha} = 0$$

$$\frac{n}{\alpha} \left\{ \frac{1}{\alpha} \Psi\left(\frac{1}{\alpha}\right) + 1 \right\} - \sum_{j=1}^n \left| \frac{\mathbf{d}_j}{s} \right| \log \left| \frac{\mathbf{d}_j}{s} \right| = 0 \quad (3.9)$$

where, $\Psi(\cdot)$ the is the digamma function $\Psi(a) = \frac{d}{da} \log \Gamma(a)$.

Employing the ML estimation method, Figure 3.9 depicts the fitted Generalized Gaussian distribution to the average histogram of the first horizontal subband coefficients for all precipitation images in the Texas and Melbourne sites. It can be observed that the GG density can explain impressively well the heavy tail non-Gaussian features of the rainfall fields in the wavelet domain. As the GG density can be fully characterized knowing the second and fourth order statistical moments in equation (3.6), the evolution of the density at multiple scales can also be studied via characterization of the scaling properties of the second and fourth order moments.

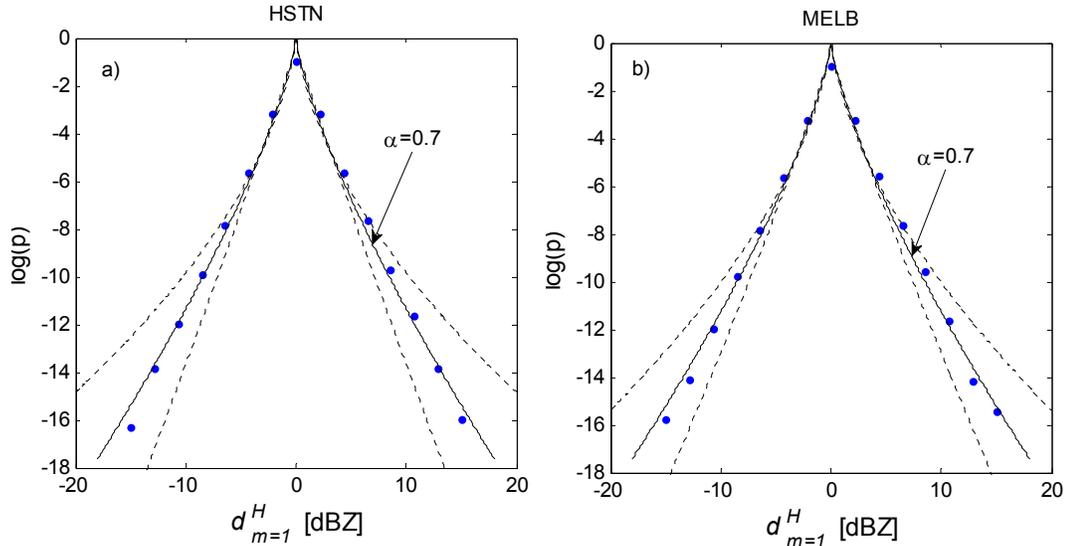


Figure 3.9: The empirical log-histogram of the horizontal subband coefficients normalized by the standard deviation, at one level of decomposition in HSTN (a) and MELB (b). The dots show the empirical histogram averaged over all storms in each site (see Figure 3.1) and the solid line is the Maximum Likelihood fitted GG distributions. The shape parameter for the average histogram is calculated consistently around 0.7 for both sites. The dashed lines present the 95% quantiles associated with the estimated parameters for each individual dataset.

As an orthogonal overcomplete wavelet representation is used in this study, the Parseval's theorem guarantees that the 2-norm is conserved in the transformed domain. Hence, as expected from spectral analysis, the scaling of the second order statistics in the wavelet

domain shall remain a power law. Observations (see Figure 3.10) also demonstrate that the fourth order moment of the wavelet coefficients obeys a power law scaling, which allows us to derive parametric expressions to describe the evolution of the GG density at multiple scales of interest. This can be further formalized in the framework of a stochastic multifractal representation (e.g., *Abry et al.*, 2004) in which, the q^{th} order moment of the wavelet coefficients in a particular subband can be explained as

$$\mathbb{E}[|\mathbf{d}_{m,j}|^q] = c_q 2^{m\tau_q} \quad (3.10)$$

where, c_q is a prefactor and τ_q characterizes the scaling law of the process in a finite range of scales. In the case that the scaling exponent τ_q can be uniquely expressed as $\tau_q = qH$, with the self-similarity index H independent of q , the process is called *mono-fractal*. In this case, the tail thickness of the marginal distribution of the process remains scale-invariant.

Decomposing all of the precipitation reflectivity images at four levels of decomposition ($m = 1$ to 4, spatial scales of 2 to 16 km), τ_q is estimated in a least squares sense for $q \in \{2, 4\}$ (Figure 3.10). The estimated values of τ_2 and τ_4 are summarized in Table 3.3 for all of the subband coefficients. It can be observed that the value of $\tau_4 - 2\tau_2$ is not equal to zero as one would obtain for a mono-fractal process. Instead, it is found that $\tau_4 - 2\tau_2 < 0$, implying that the kurtosis of the coefficients shrinks from fine-to-coarse scales. This can be seen from the kurtosis evolution at different scales given as:

$$\kappa(\mathbf{d}_{m,j}) = \frac{c_4}{(c_2)^2} 2^{m(\tau_4 - 2\tau_2)}. \quad (3.11)$$

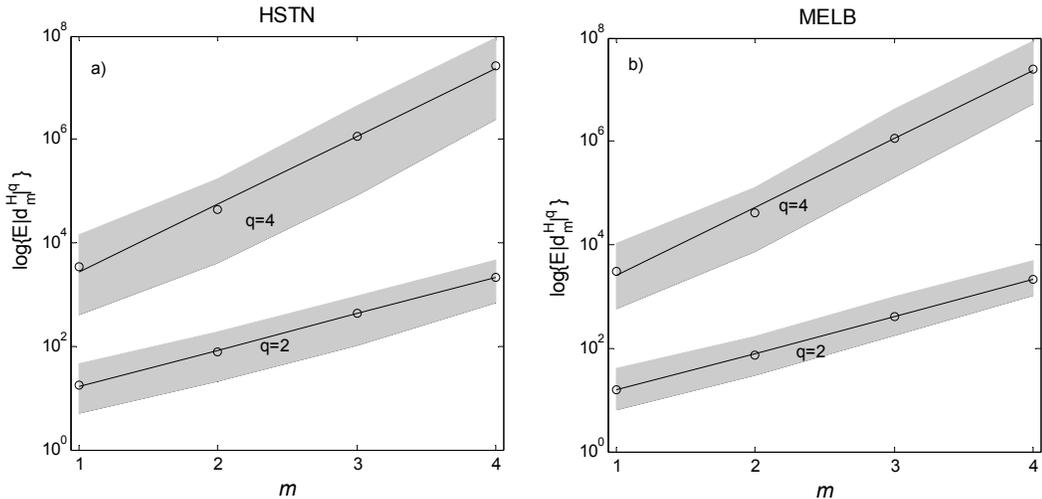


Figure 3.10: The average moment scaling law of the high-pass horizontal subband coefficients (wavelet coefficients) in the HSTN (a) and MELB (b) sites, respectively. $m = 1, \dots, 4$ denotes the spatial scales of 2 to 16 kms. This information can be exploited to characterize the evolution of the Generalized Gaussian (GG) density across a range of scales of interest. It appears that the scaling laws can be estimated consistently for both sites. The shaded areas denote the 95% quantile range of estimation. See Table 3.3 for estimated values of the scaling exponents.

This multiscale behavior of the kurtosis implies a multifractal scaling law of the wavelet coefficients. However, this nonlinear scaling (i.e. shrinkage of the tail) can be consistently studied via a linear (in log-log scale) characterization of the second and fourth order moments, individually. Accordingly, given any prior information about the scaling exponents and the wavelet coefficients at any particular scale (from which the GG parameters can be directly estimated), the evolution of the marginal density in terms of the parameters of the GG distribution can be fully explained at any scale of interest.

| HSTN | | | | MELB | | |
|----------|----------------|---------------|----------------|---------------|----------------|---------------|
| | H | V | D | H | V | D |
| τ_2 | 1.01(0.9-1.1) | 1.00(0.9-1.1) | 0.91 (0.8-1.0) | 1.03(0.9-1.1) | 1.02(0.90-1.1) | 0.93(0.8-1.0) |
| τ_4 | 1.89(1.7-2.10) | 1.86(1.6-2.0) | 1.69(1.5-1.9) | 1.91(1.7-2.1) | 1.88(1.7-2.0) | 1.75(1.6-1.9) |

Table 3.3: Estimated scaling exponents of the second and fourth order statistical moments for the NEXRAD data set at different orientations (horizontal H, vertical V, and diagonal D) in the range of scales of interest (2 to 16 km).

3.6 Joint Statistics of the Wavelet Coefficients

3.6.1 Scale-to-scale dependence

Similar to the Fourier expansion, it is theoretically proven by *Wornell and Oppenheim* (1992) that the discrete orthogonal wavelet transform for $1/f$ processes is an approximate Karhunen-Loève-like expansion that can decompose a correlated process into a set of uncorrelated expansion coefficients and orthogonal bases functions. For a 1D Gaussian scaling process, such as fractional Brownian motion with self-similarity index H , it has been theoretically shown (*Tewfik and Kim*, 1992) that the covariance of wavelet coefficients decays in the order of,

$$\mathbb{E} [\mathbf{d}_{m,k} \mathbf{d}_{r,l}^T] \sim \mathcal{O} \left(|2^m k - 2^r l|^{2(H-R)} \right) \quad (3.12)$$

where, m and r denote different scales, (k, l) are translation indices and R is the number of vanishing moments of the chosen wavelet (i.e. $\int x^p \psi_{m,k}(x) = 0, p = 0, 1, \dots, R - 1$). Obviously, according to equation (3.12), the decorrelation is not perfect for nearby coefficients and the decay rate also depends on the order of the vanishing moments of the selected wavelet; the larger the number of vanishing moments, the larger the decorrelation rate. On the contrary, due to the presence of multi-oriented edges and strong local dependencies of the intensity values in precipitation images, this whitening effect is more complicated especially in an overcomplete representation.

Although, the Haar wavelet has the least number of vanishing moments i.e. $R = 1$, it has some appealing features especially for analyzing the $1/f$ scaling and interpreting the rainfall

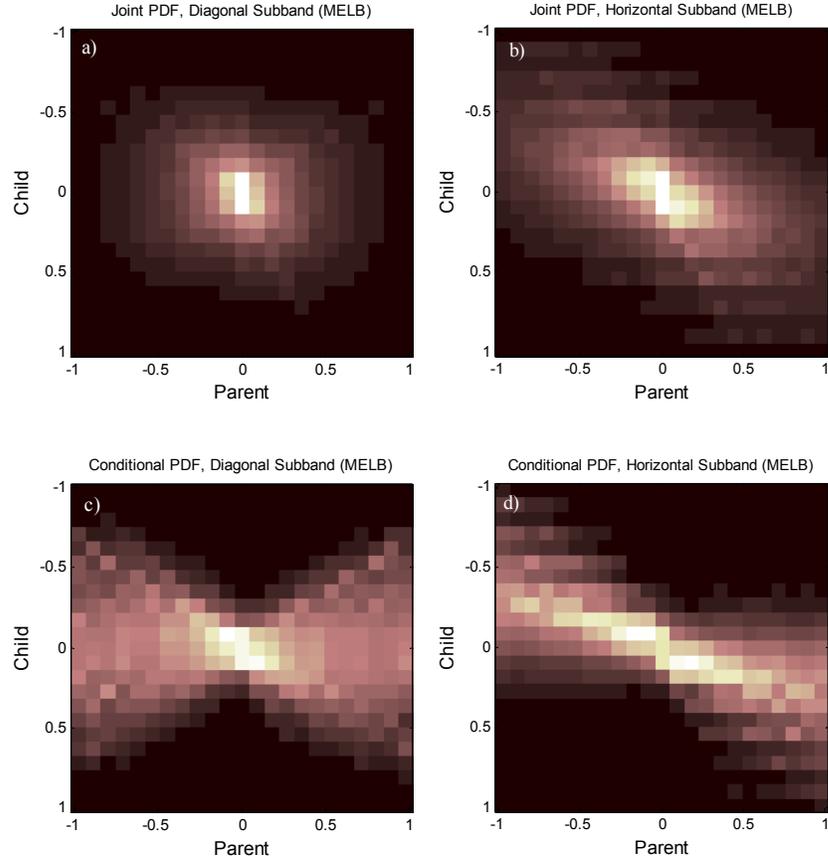


Figure 3.11: The average joint histogram of the diagonal (a) and vertical (b) high-pass wavelet coefficients at the MELB site are shown along with the corresponding conditional histograms in (c) and (d). The bow-tie shape of the conditional histograms manifests the scale-to-scale dependence and the tilted shape indicates the presence of a non-diagonal covariance structure.

wavelet coefficients as simple first-order increment of the field. For practical implementation on finite domain images, this wavelet has the shortest support among all of the wavelets and thus it does not need a periodic extension of the analyzed signal. Due to the presence of the pronounced edges in the rainfall images separating the rainy areas from the background zeros, this naturally implies that the distribution of the high-pass coefficients of the Haar wavelet with finite support has the least amount of cusp singularities at the center, which makes it more tractable for statistical parametrization. In addition, selection of the Haar wavelet is consistent with the assumption that the retrieved precipitation product is the arithmetic average representation of the highly irregular precipitation process at a particular scale (the low-pass filter corresponding to the Haar wavelet is like a box averaging filter. Obviously, using the Haar wavelet, there exist more pronounced intra-scale and scale-to-scale dependency, which needs to be characterized for proper stochastic modeling of the rainfall images.

The 2D-joint and conditional histograms of the Haar wavelet coefficients of the precipitation reflectivity images have been estimated to study the scale-to-scale dependence of the rainfall wavelet subbands at two adjacent scales. The relationships of the coarse and next finer scale coefficients are studied under the name of *parent* and *child* dependency. Figure 3.11 shows the average 2D-joint and conditional histograms of the wavelet coefficients for the MELB dataset. The conditional histogram is just a remapped version of the 2D-joint histogram in which, given the parent value, every n^{th} vertical bin is independently normalized into a probability scale such that $\sum_n p(child|parent) = 1$.

The shape of the joint histograms (Figure 3.11 top panels) clearly denotes that the conditional probability of the children given the parent is not uniform all over the domain and there exists higher order dependency that cannot be completely eliminated under the wavelet transformation. Indeed, the shape of the computed conditional histograms (Figure 3.11 bottom panels) denotes that the variance of the children depends on the parent magnitudes and larger parents give rise to children with larger variance. The tilted Bow-tie shape of the vertical subband, see Figure (3.11b) and (3.11d), also signifies the presence of off-diagonal non-zero elements on the covariance matrix of the parent and child coefficients. All of these confirm that the wavelet transformation cannot completely eliminate the scale-to-scale correlation and higher order dependence in the rainfall fluctuations. We remind the reader that the analysis in the reflectivity domain is related to analysis in the log-rainfall domain, and thus “rainfall fluctuations” here and in the sequel literally refer to fluctuations of the log-transformed rainfall fields. This kind of statistical dependencies are also observed for the HSTN-NEXRAD dataset (not shown here) for all nearby wavelet coefficients at all orientations.

3.6.2 Intra scale dependence

In addition to the fact that the wavelet transform does not completely decorrelate (globally) the precipitation images across scales, in this part we show that the wavelet coefficients of the precipitation images also exhibit a local intra-scale dependence structure. Panel (b) of Figure 3.12 shows the absolute value of the vertical high-pass subband of the storm image in panel (a). Although, it seems that the coefficients are not strongly correlated in a global sense, they are locally structured, especially near the major edges. The estimated local covariance matrices of a 5×5 neighborhood of the wavelet coefficients for all orientations is shown in panels (c), (d) and (e) of Figure 3.12. These covariance matrices are estimated using a bootstrap resampling scheme. Blocks of the neighborhood coefficients $\mathbf{d}_{m,k,l}^i$ are sampled with replacement, then for each j^{th} sampled block, the $\mathbf{d}_j \mathbf{d}_j^T$ is computed, where here \mathbf{d}_j is the column-wise vectorized version of $\mathbf{d}_{m,k,l}^i$ in a fixed order. For n -bootstrap samples, the covariance matrix $\Sigma_{\mathbf{d}}$ is estimated as follows:

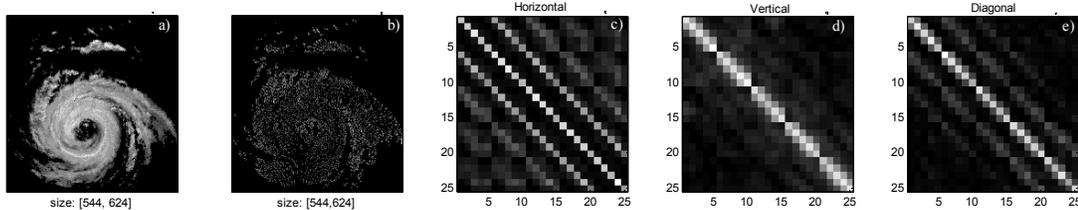


Figure 3.12: NEXRAD image of a storm over MELB on 09/26/2004 at 04:50:00 (UTC) (a) and the associated vertical subband image (b) are shown. Images of the covariance matrices of a 5×5 neighborhood for the horizontal (c), vertical (d) and diagonal (e) subband images signify the presence of an intra-scale dependence structure in the wavelet coefficients.

$$\Sigma_{\mathbf{d}} = \frac{1}{n} \sum_{j=1}^n \mathbf{d}_j \mathbf{d}_j^T \quad (3.13)$$

For sufficiently large n , this estimate guarantees convergence in probability to the true population value.

The off-diagonal non-zero elements of the estimated covariance matrices signify the imperfect local intra-scale whitening effect of the wavelet transformation (see Figure 3.12 panels (c) to (e)), which is more significant for the vertical and horizontal subbands.

All of these findings corresponding to the existence of a dependent structure among the wavelet coefficients challenge most of the available stochastic spatial rainfall models in which an uncorrelated reconstruction scheme is proposed to explain the high frequency features of the field. Note that the uncorrelated reconstruction schemes do not offer a means for controlling the tail evolution of the high frequency features and naturally, cannot accurately reproduce the small-scale extreme fluctuations in the precipitation images. For applications such as stochastic downscaling or multi-resolution fusion in the wavelet domain, the results reported herein imply that one has to consider a probability model for the wavelet coefficients, or say small scale features of precipitation images, which can reproduce a heavy tailed spatially correlated process with higher order scale-to-scale statistical dependence. In the next section, we propose a formalism within which the intra-scale correlation, scale-to-scale higher order dependence and heavy tail marginals can be simultaneously and parsimoniously reproduced in the wavelet domain.

3.7 GSM for the Wavelet Coefficients of Rainfall Images

In this section a stochastic model is introduced which allows us to capture the laid out features of the rainfall images in the wavelet domain. Basically, this model is capable of reproducing a class of heavy tail multiscale processes with a desired covariance structure together with a specific signature of higher order scale-to-scale dependence on the conditional

histogram. To this end, the basic idea is to exploit the construction proposed by *Wainwright et al.* (2001) in which the high-pass wavelet coefficients are modeled via a mixture of Gaussian random variables on a tree-like structure. Specifically, it will be shown that the wavelet coefficients can be decoupled into a mixture of two different Gaussian processes in which one controls the covariance and second order scaling, while the other one takes into account the tail and higher order scale-to-scale dependence.

3.7.1 Cascade of Gaussian Scale Mixtures on Wavelets Trees

Andrews and Mallows (1974) and *Wainwright et al.* (2001) showed that a set of heavy tailed symmetric density functions including the Laplace, t -distribution, logistic, standard power exponential and even stable distribution, can be generated as a mixture of Gaussian random variables, called Gaussian Scale Mixtures (GSM):

$$\mathbf{d} \stackrel{d}{=} \sqrt{z}\mathbf{u} \quad (3.14)$$

where $\stackrel{d}{=}$ stands for equality in distributions, z is a positive independent scalar random variable, the so called mixing random variable or the multiplier, \mathbf{u} is a zero-mean Gaussian vector with a given covariance matrix $\Sigma_{\mathbf{u}}$ and \mathbf{d} is the family of Gaussian Scale Mixtures (GSM). In particular, knowing that the z and \mathbf{u} are independent, the n -dimensional GSM has the following density function which can be specified with different choices of the random variable z :

$$f_D(\mathbf{d}) = \int_0^{\infty} f_{d|z}(\mathbf{d}|z) f_z(z) dz = \int_0^{\infty} \frac{1}{(2\pi)^{\frac{n}{2}} |z\Sigma_{\mathbf{u}}|^{1/2}} \exp\left(-\frac{\mathbf{d}^T (z\Sigma_{\mathbf{u}})^{-1} \mathbf{d}}{2}\right) f_z(z) dz \quad (3.15)$$

The discrete version of equation (3.15) resembles the statistical concept of estimating a symmetric distribution by summing zero mean Gaussian kernel densities whose covariances have been randomized by a positive random variable z . For instance, choosing z from the family of exponential density functions yields a representation of the family of Laplace distributions. Several classes of heavy tailed distributions can be produced in the context of the GSM; however, for the GG family with $0 < \alpha < 1$, a closed form expression for the density of the mixing random variable z does not exist.

By construction, one of the key properties of the GSM is that,

$$\Sigma_{\mathbf{d}} = \mathbb{E}[z] \Sigma_{\mathbf{u}} \quad (3.16)$$

which implies that the covariance structure of the GSM can be fully explained by the covariance matrix $\Sigma_{\mathbf{u}}$ and the mean of the multiplier process. Therefore, without loss of generality setting $\mathbb{E}[z] = 1$, the whole covariance structure of the GSM can be explained by the covariance of the \mathbf{u} . Accordingly, to generate a GSM with a desirable covariance,

similar to that of the rainfall wavelet coefficients, we need to adopt a mechanism which allows us to efficiently generate a (weakly) correlated Gaussian random field in a multiscale framework. For this purpose, the general class of multi-resolution linear Gauss-Markov processes defined on a regular tree-like structure (Figure 3.13) is of particular interest as follows:

$$\mathbf{x}(s) = A(s)\mathbf{x}(s\bar{\gamma}) + \mathbf{B}(s)\mathbf{w}(s) \quad (3.17)$$

where $\mathbf{x}(s)$ is the state of the process at node s and $s\bar{\gamma}$ is the parent node to which $\mathbf{x}(s)$ is connected at the next coarser scale, $\mathbf{A}(s)$ is the transition matrix, specified at each node of the tree and determining the coarse-to-fine scale dynamics of the process, $\mathbf{w}(s) \sim \mathcal{N}(0, \mathbf{I})$ and $\mathbf{B}(s)\mathbf{w}(s)$ is a Gaussian white noise with covariance $\mathbf{Q}(s) = \mathbb{E}[\mathbf{B}(s)\mathbf{B}(s)^T]$ (Chou *et al.*, 1994). The random vector $\mathbf{x}(s)$ at each node of the tree has a Gaussian distribution $\mathcal{N}(0, \Sigma_{\mathbf{x}(s)})$ with the following coarse-to-fine scale dynamics for the evolution of the covariance:

$$\Sigma_{\mathbf{x}(s)} = \mathbf{A}(s)\Sigma_{\mathbf{x}(s\bar{\gamma})}\mathbf{A}(s)^T + \mathbf{Q}(s) \quad (3.18)$$

known as the discrete Lyapunov equation. According to this recursion, the strength of the scale-to-scale dependence is determined by the value of the transition matrix $\mathbf{A}(s)$. When this prefactor tends to zero, the Markovian structure of the tree weakens and the process would be roughly uncorrelated from scale-to-scale with a nearly diagonal covariance $\Sigma_{\mathbf{x}(s)} \cong \mathbf{Q}(s)$. This construction provides a very flexible multiscale covariance structure, within which the special case of scale-to-scale stationarity can also be achieved by setting $\mathbf{A}(s) = \mathbf{A}$, $\mathbf{B}(s) = \mathbf{B}$, $\Sigma_{\mathbf{x}(s)} = \Sigma_{\mathbf{x}(s\bar{\gamma})}$ in equation (3.18) and adjusting \mathbf{A} and \mathbf{B} accordingly. However, due to the observed second-order scaling of the wavelet coefficients of the rainfall images, the non-stationary scale-to-scale construction of $\mathbf{u}(s)$ is of interest in this study. To this end, a stationary process $\mathbf{x}(s) \sim \mathcal{N}(0, \Sigma_{\mathbf{x}(0)})$ can be generated according to the dynamics in equation (3.17), where $\Sigma_{\mathbf{x}(0)}$ is the covariance at the root node, and then the scale-to-scale non-stationarity can be imposed by setting,

$$\mathbf{u}(s) = 2^{-j(s)\tau_2}\mathbf{x}(s) \quad (3.19)$$

where $j(s)$ represents the scale level on the wavelet tree from coarse-to-fine scales and τ_2 represents the geometric decay rate of the variance of the wavelet coefficients across dyadic scales (see equation (3.10) and Figure 3.10).

Recalling that $\Sigma_{\mathbf{d}(s)} = \Sigma_{\mathbf{u}(s)}$, this geometric decay rate indeed guarantees the scaling law of the variance of the wavelet coefficients,

$$\text{diag}(\Sigma_{\mathbf{d}(s)}) \cong 2^{-j(s)\tau_2}\text{diag}(\Sigma_{\mathbf{d}(0)}) \quad (3.20)$$

which leads to the presence of dyadic self-similarity and $1/f$ spectrum in the reconstructed field.

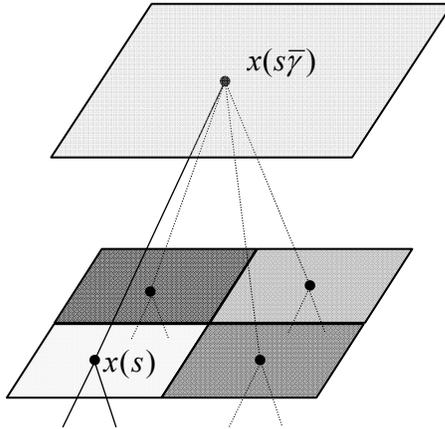


Figure 3.13: Schematic of a multi-resolution regular quad-tree, where each node on the cascade $\mathbf{x}(s)$ is connected to a unique parent $\mathbf{x}(s\bar{\gamma})$ node. On a quad-tree, $\mathbf{x}(s)$ indeed represents a 3-tuple including the scale level and the pair of spatial positions.

In addition to the scale-to-scale dependence, due to the tree-like Markovian construction, the nearby nodes at the same scale also exhibit a dependent structure as long as they share the same parent. Consequently, according to the proposed construction, this framework also allows us to capture not only the observed global scale-to-scale statistical structure, but also the local intra-scale correlation among the wavelet coefficients. However, in the simplest case, one may still decide to just assume completely uncorrelated wavelet coefficients (i.e. $A(s) = 0$) and pursue a white reconstruction phase accordingly.

Furthermore, simulating a GSM random variable with a desirable marginal density naturally requires a priori information and optimal estimation of the multiplier density function from the available data. Focusing on the marginal density of the GSM model, equation (3.14) can be written as

$$\log |\mathbf{d}(s)| = 1/2 \log [z(s)] + \log |\mathbf{u}(s)|, \quad (3.21)$$

Knowing that the convolution of two functions in the real space is equivalent to the product of their Fourier transforms in the frequency domain, the density of $\log |z(s)|$ can be computed non-parametrically, given a set of observations of the $\mathbf{d}(s)$ (see, *Portilla et al.*, 2001). The density of $\log |\mathbf{d}(s)|$ is indeed the convolution of the densities in the right hand side of equation (3.21) and hence, a rescaled version of the $\log [z(s)]$ distribution can be estimated by deconvolving the density of $\log |\mathbf{u}(s)|$ from the empirical histogram of the $\log |\mathbf{d}(s)|$. Note that in our case, $\mathbf{d}(s)$ will be the wavelet coefficients obtained from the wavelet transformation of the precipitation reflectivity images as discussed earlier. Figure 3.14(a), displays the results of the deconvolution problem for the MELB data set horizontal subbands at one level of decomposition. The log-histogram of the $\log [z(s)]$ is an inverted parabola and remarkably Gaussian which implies that the multiplier can be well explained

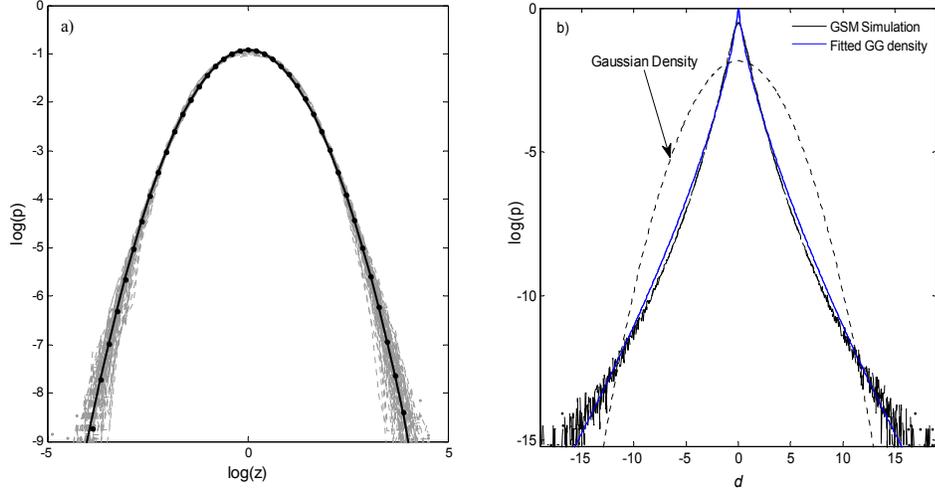


Figure 3.14: (a) Estimated densities (dashed lines) of the logarithm of the multiplier for the horizontal subband coefficients of the MELB data are well approximated by a Gaussian distribution (inverted parabola in log-probability). Note that the variance of the coefficients is normalized to one before performing the deconvolution. The solid circles show the $\mathcal{N}(0, 1)$ and the solid line is the average histogram of the data. (b) Simulated marginal of the GSM random variables using a log-normal multiplier is shown versus the fitted Generalized Gaussian (GG) density with $\alpha = 0.7$ as found from precipitation reflectivity data (see Figure 3.9). The Gaussian density is also shown for comparison.

by a scalar log-normal random variate $z(s) \sim \mathcal{LN}(\mu_z(s), \sigma_z(s))$, where $\mu_z(s)$ and $\sigma_z(s)$ are the mean and variance of the $\log[z(s)]$. As anticipated, numerical simulation of the GSM random variables in equation (3.14) using a log-normal multiplier shows that this mixture can reproduce reasonably well the GG density, see Figure 3.14(b), in close resemblance to the histogram of the rainfall wavelet coefficients (e.g., compare to Figure 3.9).

At a particular scale, knowing that $\mathbb{E}[\mathbf{d}(s)^4] = \mathbb{E}[z(s)^2] \mathbb{E}[\mathbf{u}(s)^4]$ and $\mathbb{E}[\mathbf{u}(s)^4] = 3 \{\mathbb{E}[\mathbf{u}(s)^2]\}^2$, a closed form expression for the kurtosis of the wavelet coefficients is derivable, $\kappa[\mathbf{d}(s)] = 3\mathbb{E}[z(s)^2]$. Assuming $\mathbb{E}[z(s)] = 1$ leads to $\mu_z(s) + \frac{\sigma_z(s)^2}{2} = 0$, which yields:

$$\kappa[\mathbf{d}(s)] = 3 \exp[\sigma_z(s)^2] \quad (3.22)$$

Equating (3.22) and (3.7), Figure 3.15 provides the relationship between the tail of the GG distribution and the variance of the log-multiplier. Assuming the GG density as a parametric model for the marginal histogram of the wavelet coefficients, this tells us that the tail of the GSM can be fully controlled by the variance of the $\log[z(s)]$. Accordingly, obtaining the sample kurtosis of the wavelet coefficients, the distribution of the multiplier can be fully characterized. It is worth noting that, according to equation (3.22) and the positivity of $\sigma_z(s)$, the proposed GSM construction does not allow a thinner tail than the Gaussian case and therefore this model is only suitable for generating GG marginals with $0 < \alpha \leq 2$. On the other hand, the scale-to-scale evolution of $\sigma_z(s)$ can be further

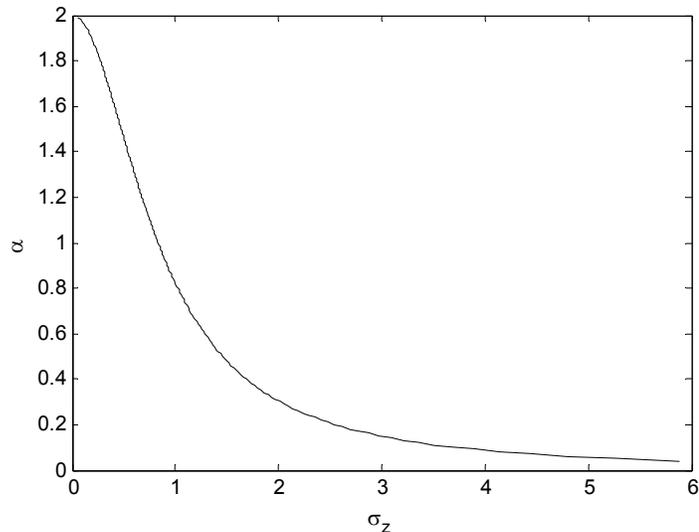


Figure 3.15: The relationship between the shape parameter α of the Generalized Gaussian (GG) density and the log-multiplier standard deviation σ_z .

expressed in terms of the multifractal properties of the rainfall fields manifested on the kurtosis statistic in equation (3.11).

$$\sigma_z^2(s) = \log\left(\frac{c_4}{3(c_2)^2}\right) + m(\tau_4 - 2\tau_2)\log(2) \quad (3.23)$$

This shows that the GSM log-multiplier allows us to collapse the scaling information content of the second and fourth order moments of the process into a single parameter of the log-multiplier, which eventually controls the thick tail properties of the marginal and higher order scale-to-scale dependence. In other words, in this construction $\mathbf{u}(s)$ captures second order statistics and the associated scaling of the wavelet coefficients, independent of $z(s)$ which addresses the heavy tail statistics and higher order dependency.

Figure 3.16 shows the results of a numerical experiment that demonstrates how the proposed construction can generate similar statistical signatures to those found for the wavelet coefficients of precipitation images. To this end, assuming the identity matrix as the normalized covariance of the wavelet coefficients at the root node, a cascade of stationary multiscale processes $\mathbf{x}(s)$ is generated by setting $\mathbf{A}(s) = \eta I$ and $\mathbf{B}(s) = \sqrt{1 - \eta^2} I$. Subsequently, the second order scaling law of the process is imposed on the cascade according to equation (3.19). The strength of the scale-to-scale correlation can be adjusted by η . For example, a nearly uncorrelated scale-to-scale construction can be achieved by sending η to zero and on the contrary, while η tends to unity the Markovian property is much stronger and the cascade produces a highly correlated field. Indeed, a larger value of the η increases the off-diagonal entries of the parent and child covariance matrix and gives rise to a tilted joint histogram.

Note that the heavy tail property of the marginal density and the higher order scale-to-scale dependence in terms of the observed bow-tie shape of the joint histogram, only depend on the variance of the log-multiplier process which characterizes the shape of the tail. Empirically, it seems that the type of dependence in the conditional histogram (shape of the bow-tie) is tightly related to the thickness of the tail, meaning that, for heavier tail (i.e. larger $\sigma_z(s)$) the high order parent-to-child dependency is more pronounced (Figure 3.16).

3.7.2 GSM on Wavelet Trees vs. Multiplicative Random Cascades

Multiplicative random cascades, in their canonical form, have been of central importance to stochastic simulation of geophysical processes and especially precipitation data (Gupta and Waymire, 1993). This class of stochastic models, with the coarse-to-fine scale recursion described below, allows us to generate multi-fractal measures with similar statistical properties as those typically observed in rainfall across a finite range of scales:

$$\mathbf{x}(s) = \mathbf{x}(s\bar{\gamma}) \zeta(s) \tag{3.24}$$

where, $\zeta(s)$ represents an independent identically distributed (*i.i.d.*) random multiplier at each node s of the tree with $\mathbb{E}[\zeta(s)] = 1$, also known as the cascade generator. The multiplicative structure of this model is a key factor which imposes the desired multifractal properties (e.g., Mandelbrot *et al.*, 1997) and the parent-to-child scale dynamics in the sense that the larger parents are more potent to generate larger children, a property that has been amply documented in the precipitation fields. However, this construction is nonlinear by its nature and hence the state estimation of equation (3.24) given a set of noisy observations, even under an affine observation equation:

$$\mathbf{y}(s) = \mathbf{x}(s) + \mathbf{w}(s) \tag{3.25}$$

where $w(s) \sim \mathcal{N}(0, \mathbf{R}(s))$, is not a trivial task. At first glance, it seems that by working in the log-space to linearize the model equation,

$$\log[\mathbf{x}(s)] = \log[\mathbf{x}(s\bar{\gamma})] + \log[\zeta(s)] \tag{3.26}$$

where, $\mathbb{E}\{\log[\zeta(s)]\} = 0$, the linear estimation theory of additive Markovian multiscale models (e.g., Gorenburg *et al.*, 2001; Tustison *et al.*, 2003) can be invoked, while preserving the multiscale properties of the multiplicative random cascade. We need to note however that, although the mean is conserved in this log-transformation, the higher order parent-to-child dependency is not preserved, given that $\zeta(s)$ is a sequence of *i.i.d.* random variates. For instance, it is easy to check that the conditional variance of $\mathbf{x}(s)$ given $\mathbf{x}(s\bar{\gamma})$ in equation (3.24) depends linearly on the magnitude of $\mathbf{x}(s\bar{\gamma})$ while in equation (3.26) this variance is only characterized by the noise term $\zeta(s)$. In other words, a proper additive construction

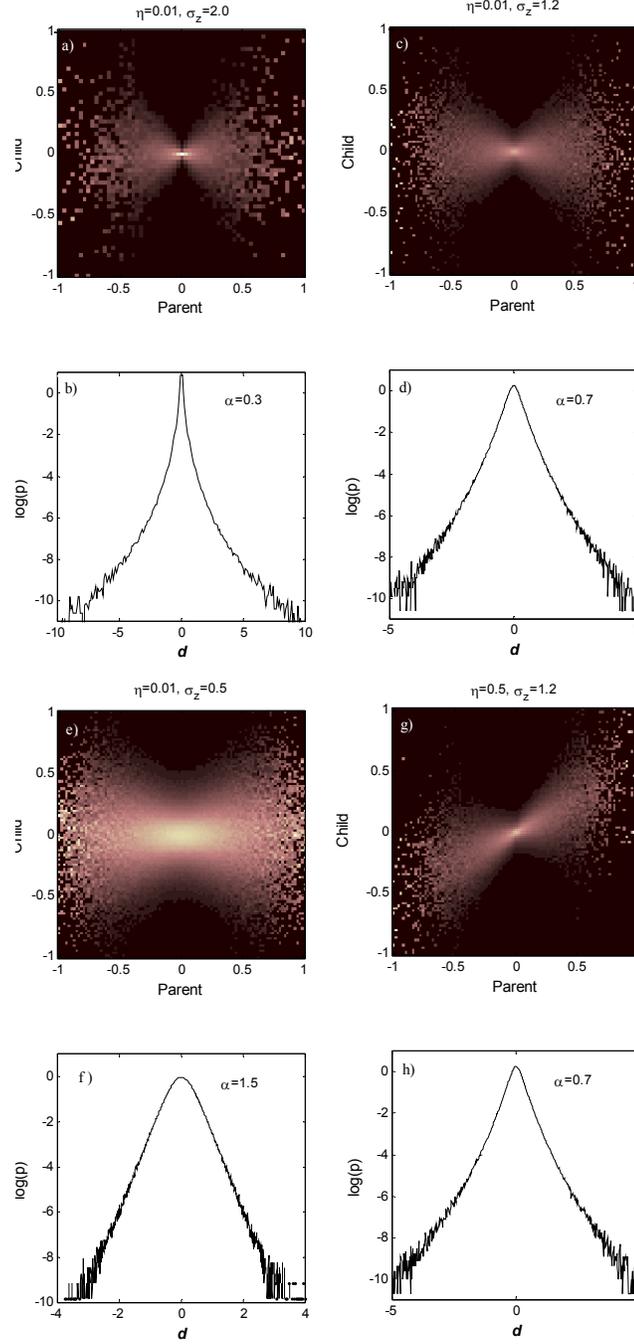


Figure 3.16: Marginal and joint statistics of simulated GSM cascades for various choices of parameters (η, σ_z) . A larger variance of the multiplier σ_z increases the thickness of the tail (smaller α) and the significance of the scale-to-scale dependence and η controls the directionality of the bow-tie shape of the joint histogram. (a,b) $\eta = 0.01, \sigma_z = 2.0(\alpha = 0.3)$; (c,d) $\eta = 0.01, \sigma_z = 1.2(\alpha = 0.7)$; (e,f) $\eta = 0.01, \sigma_z = 0.5(\alpha = 1.5)$; and (g,h) $\eta = 0.5, \sigma_z = 1.2(\alpha = 0.7)$.

requires the derivation of an appropriate noise term that can take into account this high order dependency (e.g. a correlated noise), which cannot be definitely explained by an *i.i.d.* $\zeta(s)$. A very important implication of this deduction is that, using multiplicative random cascades, implementation of the linear state estimation theory for precipitation data in log-transformed space cannot fully capture the distinct statistical signature of the rainfall process.

On the contrary, the GSM cascade on the wavelet tree has an additive construction, which allows a subtle and explicit characterization of the wavelet detail coefficients to properly account for the statistical structure of these fields. Indeed, given an estimate of $z(s)$, the density of $\mathbf{d}(s)$ is Gaussian and hence the conditional estimation of the wavelet coefficients becomes a linear problem. This is a great advantage of the GSM construction in the wavelet domain which eventually permits exploiting the well established linear estimation techniques.

Chapter 4

Adaptive Fusion of Multisensor Precipitation using Gaussian Scale Mixtures in the Wavelet Domain

In the previous chapter it is 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 chapter, 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 then 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 chapter is structured as follows:

In the context of precipitation multisensor fusion, the standard linear Gaussian estimation method is explained and implemented in Section 4.1. Practical aspects of implementation and shortcomings of this method are also discussed in this section. Section 4.2 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 4.3, basic theoretical and practical concepts of optimal estimation in the wavelet domain using the GSM probability model are explained. A synthetic 1D example is also presented to elaborate the main advantages of the proposed algorithm compared with the SRE method. Section 4.4 describes the implementation of the new model for precipitation estimation and data fusion by applying the algorithm to a real storm event coincidentally measured by ground-based NEXRAD and TRMM-PR sensors.

4.1 Linear Fusion of Multisensor Precipitation Data in the Spatial Domain

Since minimization of linear quadratic error linear functionals can be cast as a convex optimization 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 multi-sensor precipitation observations builds upon the basic theory of least squares estimation, a brief explanation of the relevant theoretical aspects is presented in this section.

4.1.1 Principle of Linear Least Squares Estimation

Consider a set of noisy measurements $\mathbf{y} \in \mathbb{R}^n$ of a parameter vector $\mathbf{x} \in \mathbb{R}^m$ with joint covariance matrix of the form $\Sigma = [\Sigma_{\mathbf{x}}, \Sigma_{\mathbf{xy}}; \Sigma_{\mathbf{yx}}, \Sigma_{\mathbf{y}}]$. 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,

$$\hat{\mathbf{x}} = m_{\mathbf{x}} + \Sigma_{\mathbf{xy}}\Sigma_{\mathbf{y}}^{-1}(\mathbf{y} - m_{\mathbf{y}}) \quad (4.1)$$

$$\Sigma_{\hat{\mathbf{x}}} = \Sigma_{\mathbf{x}} - \Sigma_{\mathbf{xy}}\Sigma_{\mathbf{y}}^{-1}\Sigma_{\mathbf{yx}} \quad (4.2)$$

where $m_{\mathbf{x}} = \mathbb{E}[\mathbf{x}]$ and $m_{\mathbf{y}} = \mathbb{E}[\mathbf{y}]$ (see, *Levy, 2008*).

Casting this problem in the context of a linear measurement equation of the form $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v}$ in the Gaussian noise environment $v \sim \mathcal{N}(0, \Sigma_{\mathbf{v}})$, where $C \in \mathbb{R}^{n \times m}$ is the measurement matrix, and knowing $\Sigma_{A\mathbf{x}, B\mathbf{y}} = A\Sigma_{\mathbf{xy}}B^T$ for any matrices A and B of relevant size, the above expression can be further expanded as follows.

$$\hat{\mathbf{x}} = m_{\mathbf{x}} + \Sigma_{\mathbf{x}}\mathbf{H}^T(\mathbf{H}\Sigma_{\mathbf{x}}\mathbf{H}^T + \Sigma_{\mathbf{v}})^{-1}(\mathbf{y} - \mathbf{H}m_{\mathbf{x}}) \quad (4.3)$$

$$\Sigma_{\hat{\mathbf{x}}} = \Sigma_{\mathbf{x}} - \Sigma_{\mathbf{x}}\mathbf{H}^T(\mathbf{H}\Sigma_{\mathbf{x}}\mathbf{H}^T + \Sigma_{\mathbf{v}})^{-1}\mathbf{H}\Sigma_{\mathbf{x}} \quad (4.4)$$

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 $\mathbb{E} \left[|\mathbf{x} - \hat{\mathbf{x}}|^2 \right] \leq \mathbb{E} \left[|\mathbf{x} - \text{span}\{\mathbf{y}\}|^2 \right]$. 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 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 $\mathbb{E} \left[|\mathbf{x} - \mathbb{E}[\mathbf{x}|\mathbf{y}]|^2 \right] \leq \mathbb{E} \left[|\mathbf{x} - f(\mathbf{y})|^2 \right]$, 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 requires 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.

A least squares estimation paradigm was introduced by *Kalman (1960)* for the estimation of discrete time linear Gauss-Markov stochastic processes; i.e., $\mathbf{x}_t = \mathbf{A}_t \mathbf{x}_{t-1} + \mathbf{w}_t$ where \mathbf{A}_t is the temporal transition matrix and $\mathbf{w}_t \sim \mathcal{N}(0, \Sigma_{\mathbf{w}_t})$ 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 $\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{v}_t$, where \mathbf{H}_t relates the system state to the measurements and $\mathbf{v}_t \sim \mathcal{N}(0, \Sigma_{\mathbf{v}})$. 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. In MAR representation, a multiresolution process is naturally defined on a tree-like graph structure \mathcal{T} (Figure 3.13), where each node $s \in \mathcal{T}$ on the tree is a 3-tuple which indicates the signal quantity $\mathbf{x}(s)$ in a specific translational offset and scale level:

$$\mathbf{x}(s) = \mathbf{A}(s) \mathbf{x}(s\bar{\gamma}) + \mathbf{w}(s) \quad (4.5)$$

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

$$\Sigma_{\mathbf{x}(s)} = \mathbf{A}(s) \Sigma_{\mathbf{x}(s\bar{\gamma})} \mathbf{A}^T(s) + \Sigma_{\mathbf{w}(s)} \quad (4.6)$$

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

$$\mathbf{y}(s) = \mathbf{H}(s) \mathbf{x}(s) + \mathbf{v}(s) \quad (4.7)$$

where $\mathbf{H}(s)$ specifies the linear relationship between $\mathbf{x}(s)$ and what is measured at each node and $\mathbf{v}(s)$ is a Gaussian white noise independent of $\mathbf{w}(s)$ with covariance $\Sigma_{\mathbf{v}(s)} = \mathbb{E}[\mathbf{v}(s)\mathbf{v}(s)^T]$. 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 in (4.5).

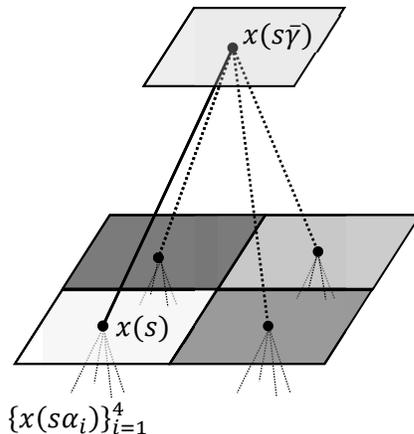


Figure 4.1: Regular quad-tree structure typically used to model 2D multiresolution Markov fields in which each parent node gives birth to four children nodes. In this representation, each node on the tree represents the process magnitude in a spatial region and specific location, while each level of the tree corresponds to a particular resolution (scale) of the process.

4.1.2 Scale Recursive Estimation of Multi-sensor Precipitation data

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., 2003; Bocchiola, 2007; Van de Vyver and Roulin, 2009*). Typically, a regular quad-tree \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 $j(s)$ gives rise to four children nodes, namely $\{s\alpha_i\}_{i=1}^4$, at scale $j(s)+1$ (Figure 4.1). Scale recursive estimation of multi-sensor 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?

Typically, in this setting the coarse resolution precipitation data (i.e., satellite observations or General Circulation Model predictions) are considered as model information $\mathbf{x}(s)$ at the root node of the quad-tree, 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 $\mathbf{y}(s)$.

In the context of rainfall data, it has long been argued (*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., $\mathbf{r}(s) = \mathbf{r}(s\bar{\gamma})\zeta(s)$, where $\zeta(s)$ represents a driving random

component known as the cascade generator with $\mathbb{E}[\zeta(s)] = 1$, in a microcanonical form. Working on high resolution rainfall data, *Menabde et al.* (1997) proposed a log-normal density for the $\zeta(s)$. To treat this nonlinear recursion and make it consistent with the settings in (4.5) and (4.7), typically the SRE method is performed in the log-transformed rainfall; i.e., $\log[\mathbf{r}(s)] := \mathbf{x}(s)$, or say a shifted version of the reflectivity fields

$$\log[\mathbf{r}(s)] = \log[\mathbf{r}(s\bar{\gamma})] + \log[\zeta(s)], \quad (4.8)$$

where, $\log[\zeta(s)]$ is a Gaussian white noise, equivalent to the term $w(s)$ in (4.5). In terms of the first order and marginal statistics, this transformation seems fine; however, the log-transformation can not 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., $\text{var}[\mathbf{r}(s)|\mathbf{r}(s\bar{\gamma})] = (\mathbf{r}(s\bar{\gamma}))^2 \text{var}[\zeta(s)]$, which is not the case in (4.8), 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 (Figure4.2b).

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 multi-sensor precipitation data. Here, an example is provided which uses the SRE framework to merge precipitation given from TRMM-PR satellite and ground-based NEXRAD coincidental precipitation reflectivity imageries. We assumed that the reflectivity images can be partially explained by the linear MAR model in (4.5). The tree is assumed stationary in the sense that $\mathbf{A}(s) = \mathbf{I}$ and as explained the “model” information $\mathbf{x}(s)$ is obtained from the coarse resolution TRMM near-surface reflectivity images at $\approx 4 \times 4$ km and the “observations” $\mathbf{y}(s)$ 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 $\mathbf{H}(s) = \mathbf{I}$. To address the self-similarity and commonly observed $1/f$ spectrum in the rainfall reflectivity images (Figure 4.2a), it is assumed that the variance of the driving noise term $\mathbf{w}(s)$, decays geometrically from coarse-to-fine scales by assigning $\Sigma_{\mathbf{w}(s)} \propto 2^{-H j(s)} \mathbf{I}$, 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 drop-off rate of the power spectrum of the synthesized fields 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

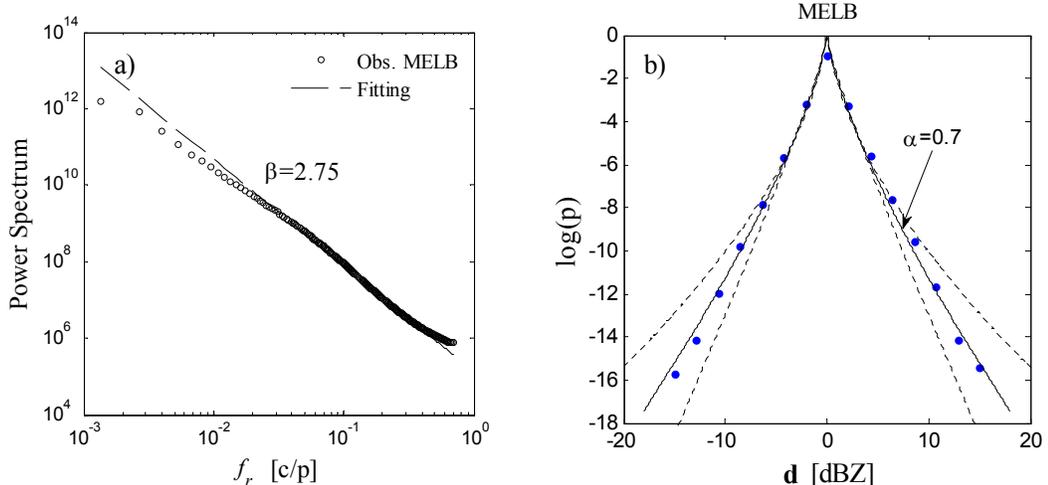


Figure 4.2: (a) Radially averaged ensemble spectrum of 105 NEXRAD precipitation single level near-surface reflectivity images at resolution 1×1 km over the Ground validation site of the TRMM satellite in Melbourne, Florida. Frequencies are in cycle per pixel [c/p] which are equivalent to the inverse of pseudo spatial scale [km^{-1}]. (b) the 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 $p(x) \propto \exp(-|x/s|^\alpha)$. The dashed lines are the 95% estimation quantiles.

down-sampling by a factor of two, 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 up-sampled by a factor of 2 and convolved with a nearest neighborhood filter 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 (*Gonzalez and Woods, 2008*). Recursive implementation of this encoding procedure yields characterization of the scale-to-scale detail information and estimation of the noise term $\mathbf{w}(s)$ in (4.5). 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 $\mathbf{x}(s)$ at higher resolution of interest on the tree.

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 in (4.5).

Using SRE, Figure 4.3 shows the fusion of high resolution (i.e., 1×1 km) NEXRAD observations of a storm event over the HSTN site on 1998/06/28 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 $\Sigma_{\mathbf{v}(s)}$ compared to the process covariance at each node will yield a fusion product which is more inclined towards 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 (2011a)* characterized the diagonal elements of this covariance for the dataset used in this study in terms of a signal-to-noise ratio measure in decibel (dB) scale; i.e., $\text{SNR} = 10 \log_{10} \left(\frac{\sigma^2(s)}{\sigma_{\mathbf{v}(s)}^2} \right)$, where $\sigma(s)$ is the standard deviation of the NEXRAD image and $\sigma_{\mathbf{v}(s)}$ 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 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.

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,

$$\text{Entropy} = - \sum_k p(\mathcal{Z}_k) \log_2 [p(\mathcal{Z}_k)] \quad (4.9)$$

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 per pixel (see, *Gonzalez and Woods, 2008*). In the studied storm images of Figure 4.3, using $\text{SNR} = 11.0$ 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

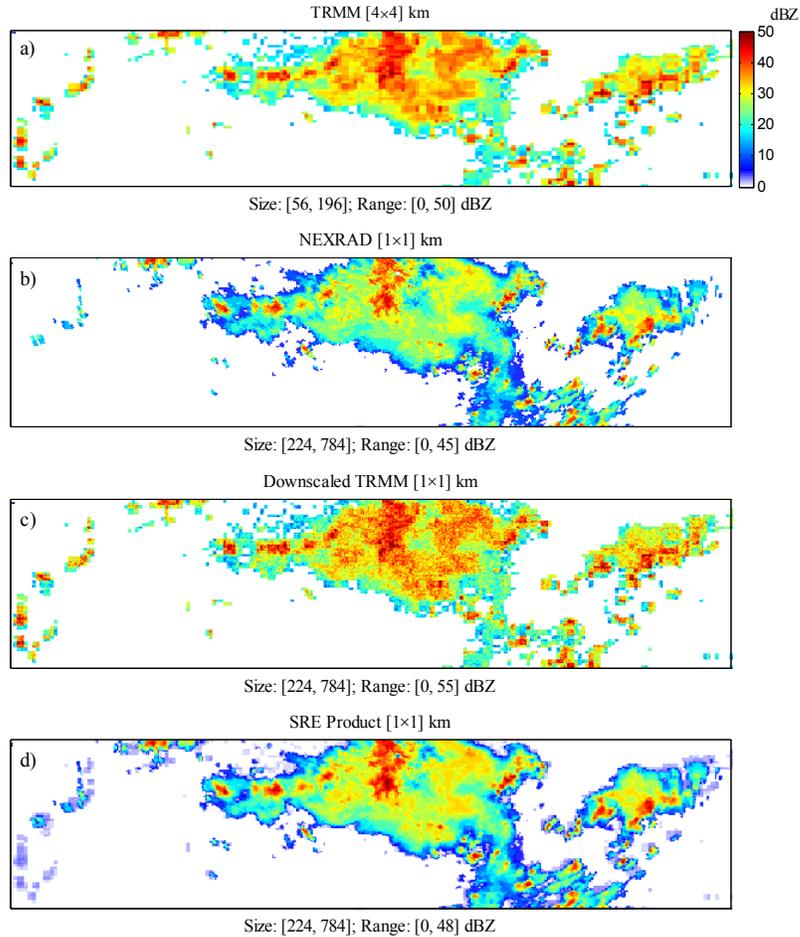


Figure 4.3: (a) TRMM-PR near surface single level reflectivity image over the HSTN site on 1998/06/28 at 18:13:00 UTC ; (b) the coincidental NEXRAD base-reflectivity image of the storm; (c) the downscaled TRMM observations according to the linear MAR model ($H = 0.6$) at resolution 1×1 km and (d) the fused product according to the SRE framework, assuming $\text{SNR} = 11.0$ dB, see text for discussion.

this entropy increase can be due to the growth of the wetted area as a natural result of the fusion process.

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 $\mathcal{Z} \in \mathbb{R}^{n \times m}$ at two scales of interest are computed and compared.

$$\|\mathcal{Z}\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^m |z_{i,j}|^2} = \sqrt{\text{tr}(\mathcal{Z}^T \mathcal{Z})} \quad (4.10)$$

Table 4.1 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.

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

Table 4.1: Frobenius norm of the TRMM-PR, NEXRAD observations and SRE-fused product, row-wised normalized by the maximum of each row. TRMM-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.

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 hydro-geomorphic 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 (*Wainwright et al.*, 2001). In the next sections, 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 1D example.

4.2 Precipitation Probability Model in the Wavelet Domain

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.

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; 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 ,

$$\mathbf{d} \stackrel{d}{=} \sqrt{z} \mathbf{u} \quad (4.11)$$

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, 2011a*), there is no closed form expression for the distribution of the multiplier. 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.

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 using Bayes’ theorem, the GSM multivariate density can be characterized as:

$$p_D(\mathbf{d}) = \int_0^\infty p(\mathbf{d}|z) p_Z(z) dz = \int_0^\infty \frac{\exp\left(-\frac{\mathbf{d}^T (z \Sigma_{\mathbf{u}})^{-1} \mathbf{d}}{2}\right)}{(2\pi)^{N/2} (\det |z \Sigma_{\mathbf{u}}|)^{1/2}} p(z) dz \quad (4.12)$$

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.

4.3 GSM Optimal Estimation in the Wavelet Domain

4.3.1 Basics of the Framework

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:

$$\mathbf{y} = \mathbf{d} + \mathbf{v} \quad (4.13)$$

where $\mathbf{v} \sim \mathcal{N}(0, \Sigma_{\mathbf{v}})$ and assuming $\mathbb{E}[z] = 1$, without loss of generality, equation (4.11) and (4.13) result in:

$$\begin{aligned} \Sigma_{\mathbf{y}|z} &= z\Sigma_{\mathbf{u}} + \Sigma_{\mathbf{v}} \\ \Sigma_{\mathbf{y}} &= \Sigma_{\mathbf{u}} + \Sigma_{\mathbf{v}} \end{aligned} \quad (4.14)$$

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

$$p(\mathbf{y}|z) = \frac{1}{(2\pi)^{n/2} (\det |z\Sigma_{\mathbf{u}} + \Sigma_{\mathbf{v}}|)^{1/2}} \exp\left(\frac{-\mathbf{y}^T (z\Sigma_{\mathbf{u}} + \Sigma_{\mathbf{v}})^{-1} \mathbf{y}}{2}\right) \quad (4.15)$$

With no a priori assumption on $p_Z(z)$ and perfect whitening effect of the wavelet transform, (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 (Figure 4.4a,b 4.5). 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.

Studying the heavy tail properties of these images in the wavelet domain, the log-normal density was found empirically as a suitable prior assumption about the distribution of the multiplier $p_Z(z)$ (Figure 4.6). 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:

$$\hat{z}_{MAP} = \arg \max_z \{\log p(z|\mathbf{y})\} = \arg \max_z \{\log p(\mathbf{y}|z) + \log p(z)\} \quad (4.16)$$

Assuming the multiplier has a log-normal 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} ,

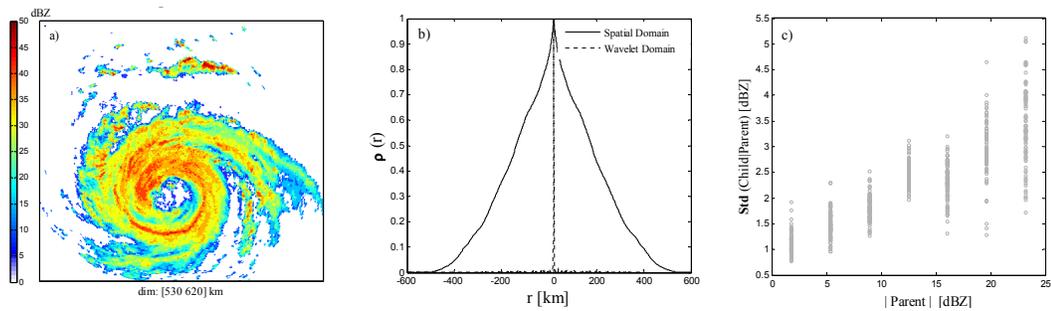


Figure 4.4: (a) A NEXRAD precipitation long-range reflectivity image at resolution 1×1 km over the MELB site on 09/26/2004 at 04:50:00 UTC ; (b) the central horizontal transect of the 2D correlation function in real (solid line) and wavelet domain (broken line). This shows that the correlation quickly decays in a short range of spatial lags in the wavelet domain as opposed to the spatial domain; and (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.

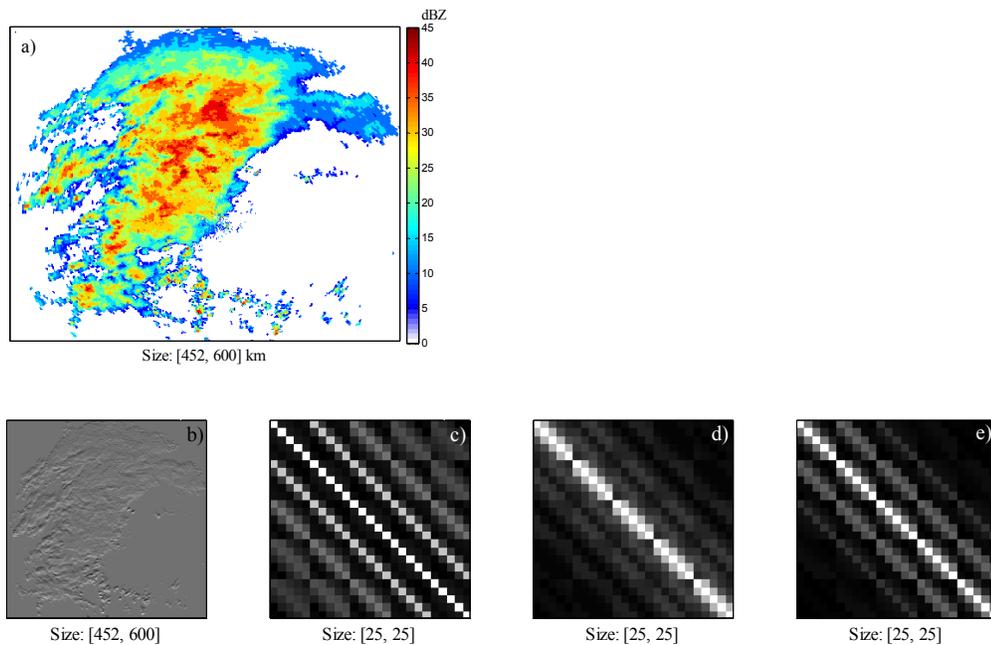


Figure 4.5: (a) a NEXRAD long-range reflectivity image (1×1 km) of a sample storm over MELB and (b) its wavelet horizontal subband image. The off-diagonal elements of the image representation of the covariance matrices of neighborhoods of size 5×5 for horizontal (c), vertical (d) and diagonal (e) subbands of the reflectivity image in (a), 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).

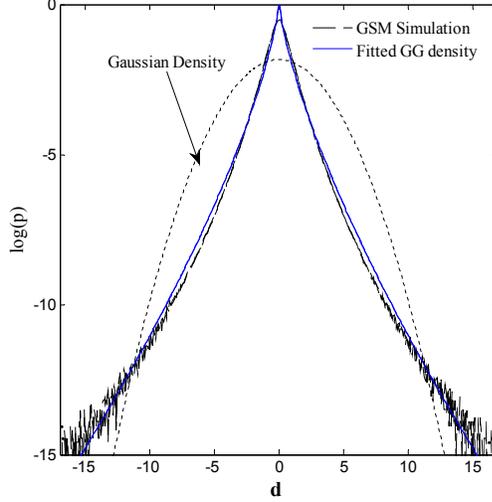


Figure 4.6: The broken line is the standardized marginal histogram (unit standard deviation) of a generated Gaussian Scale Mixture (GSM) random variable using a log-normal multiplier with $\sigma_z \approx 1.2$ in (3.12) and the solid line shows the fitted Generalized Gaussian (GG) density with $\alpha \approx 0.7$. The inverted parabola shows the Gaussian distribution in log-probability for comparison purposes.

$$f(\hat{z}_{MAP}) = \frac{\log \hat{z}_{MAP} + 3\sigma_z^2/2}{\hat{z}_{MAP}\sigma_z^2} + \frac{1}{2} \sum_{n=1}^N \frac{\hat{z}_{MAP} - \lambda_n^{-1}(v_n^2 - 1)}{(\hat{z}_{MAP} + \lambda_n^{-1})^2} = 0 \quad (4.17)$$

where v_n are the components of the vector $\mathcal{V} = \mathbf{Q}^T \mathbf{S}^{-1} \mathbf{y}$, \mathbf{S} is the square root of the error covariance matrix (i.e., $\mathbf{S}\mathbf{S}^T = \Sigma_{\mathbf{v}}$) and (Λ, \mathbf{Q}) contains the eigenvalues $\lambda_n \in \Lambda$ and eigenvectors of the positive semi-definite matrix $\mathbf{S}^{-1} \Sigma_{\mathbf{u}} \mathbf{S}^{-T}$ (Portilla *et al.*, 2001, 2003). Different numerical methods such as bisection, secant, false position or Brent's method can be used to find the root of equation (4.17), (see Appendix A). We used a multi-dimensional bisection method, 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.

Note that, solving equation (4.17) requires full characterization of the a priori log-normal 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 (4.17). Given the variance of the noise σ_v^2 and knowing that $\sigma_{\mathbf{u}}^2 = \sigma_{\mathbf{y}}^2 - \sigma_v^2$, it can be shown after some algebra that this parameter can be estimated as (see Appendix A):

$$\sigma_z^2 = \log(\mathbb{E}[z^2]) = \log\left(\frac{\mathbb{E}[\mathbf{y}^4]/3 - 2\sigma_{\mathbf{y}}^2\sigma_v^2 + \sigma_v^4}{(\sigma_{\mathbf{y}}^2 - \sigma_v^2)^2}\right) \quad (4.18)$$

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.

In the form of a general linear measurement equation $\mathbf{y} = \mathbf{H}\mathbf{d} + \mathbf{v}$, referring to the conditional Gaussian density of the GSM model and expressions in (4.3) and (4.4), the conditional expectation of the zero mean noisy wavelet coefficients and its covariance can be written as :

$$\hat{\mathbf{d}} = \mathbb{E}[\mathbf{d}|\mathbf{y}, z] = z\Sigma_{\mathbf{u}}\mathbf{H}^T [\mathbf{H}(z\Sigma_{\mathbf{u}})\mathbf{H}^T + \Sigma_{\mathbf{v}}]^{-1} \mathbf{y} \quad (4.19)$$

$$\Sigma_{\hat{\mathbf{d}}} = z\Sigma_{\mathbf{u}} - z^2\Sigma_{\mathbf{u}}\mathbf{H}^T [\mathbf{H}(z\Sigma_{\mathbf{u}})\mathbf{H}^T + \Sigma_{\mathbf{v}}]^{-1} \mathbf{H}\Sigma_{\mathbf{u}} \quad (4.20)$$

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

$$\mathbb{E}[\mathbf{d}|\mathbf{y}, \hat{z}_{MAP}] = \hat{z}_{MAP}\Sigma_{\mathbf{u}} (\hat{z}_{MAP}\Sigma_{\mathbf{u}} + \Sigma_{\mathbf{v}})^{-1} \mathbf{y} \quad (4.21)$$

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.

4.3.2 Global Versus Local Filtering

Implementation of the filter discussed above in the wavelet domain requires estimation of $\Sigma_{\mathbf{u}}$ for each subband, which can be obtained in (4.14) 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. 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.

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 (Figure 4.7). 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:

$$\Sigma_{\mathbf{u}} \approx \frac{\sum_{i=1}^{M \times L} (\mathbf{y}\mathbf{y}^T)_i}{M \times L} - \Sigma_{\mathbf{v}} \quad (4.22)$$

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 (4.21), this covariance only needs to be estimated once for each subband. On the other hand, the multiplier has to be estimated locally according to (4.17) for every neighborhood location, while it slides over the entire subband domain. The conditional expectation in (4.21) 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} .

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

4.3.3 Synthetic 1-D Implementation

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

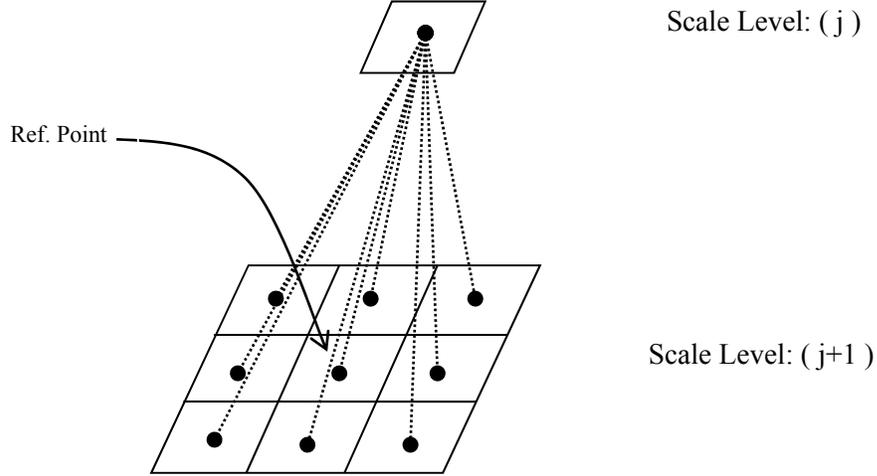


Figure 4.7: A general pyramidal neighborhood of size $K = 10$ for each individual wavelet subband where a cluster of 3×3 pixels (children) is connected to one pixel at the next coarser scale (parent). In this multiscale representation of a generalized neighborhood, scale-to-scale and short range intra-scale 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.

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, 2011a*), a one dimensional GSM process using a log-normal multiplier is simulated over a dyadic Markov tree. First, a multiscale non-stationary Gaussian process $\mathbf{u}(s)$ is produced on a dyadic tree according to the coarse-to-fine scale dynamics in (4.5). 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 element-wise by a sequence of random variables drawn from a log-normal density at different levels of the tree to produce a multiscale GSM process on a tree-like structure,

$$\begin{aligned} \mathbf{u}(s) &= \mathbf{A}(s)\mathbf{u}(s\bar{\gamma}) + \mathbf{B}(s)\mathbf{w}(s) \\ \mathbf{d}(s) &= \sqrt{z(s)}\mathbf{u}(s) \end{aligned} \tag{4.23}$$

where $\mathbf{w}(s) \sim \mathcal{N}(0, \mathbf{I})$ and $\mathbf{B}(s) = 2^{-Hj(s)/2}$. Setting $\sigma_v^2 = 0$ in (4.18), observe that the kurtosis $\kappa(\cdot)$ of a simulated log-normal GSM can be solely determined by the variance of the multiplier z ,

$$\kappa[\mathbf{d}(s)] = 3 \exp[\sigma_z^2(s)] \quad (4.24)$$

As the marginal distribution of the log-normal GSM resembles the family of Generalized Gaussian densities, 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 $\kappa(\cdot) = \Gamma(1/\alpha)\Gamma(5/\alpha)/\Gamma^2(3/\alpha)$.

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 1D 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 (Figure4.4) 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 4.8 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.

Analogous to the explained 1D case, it is expected that estimating multi-sensor 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 hydro-meteorological importance.

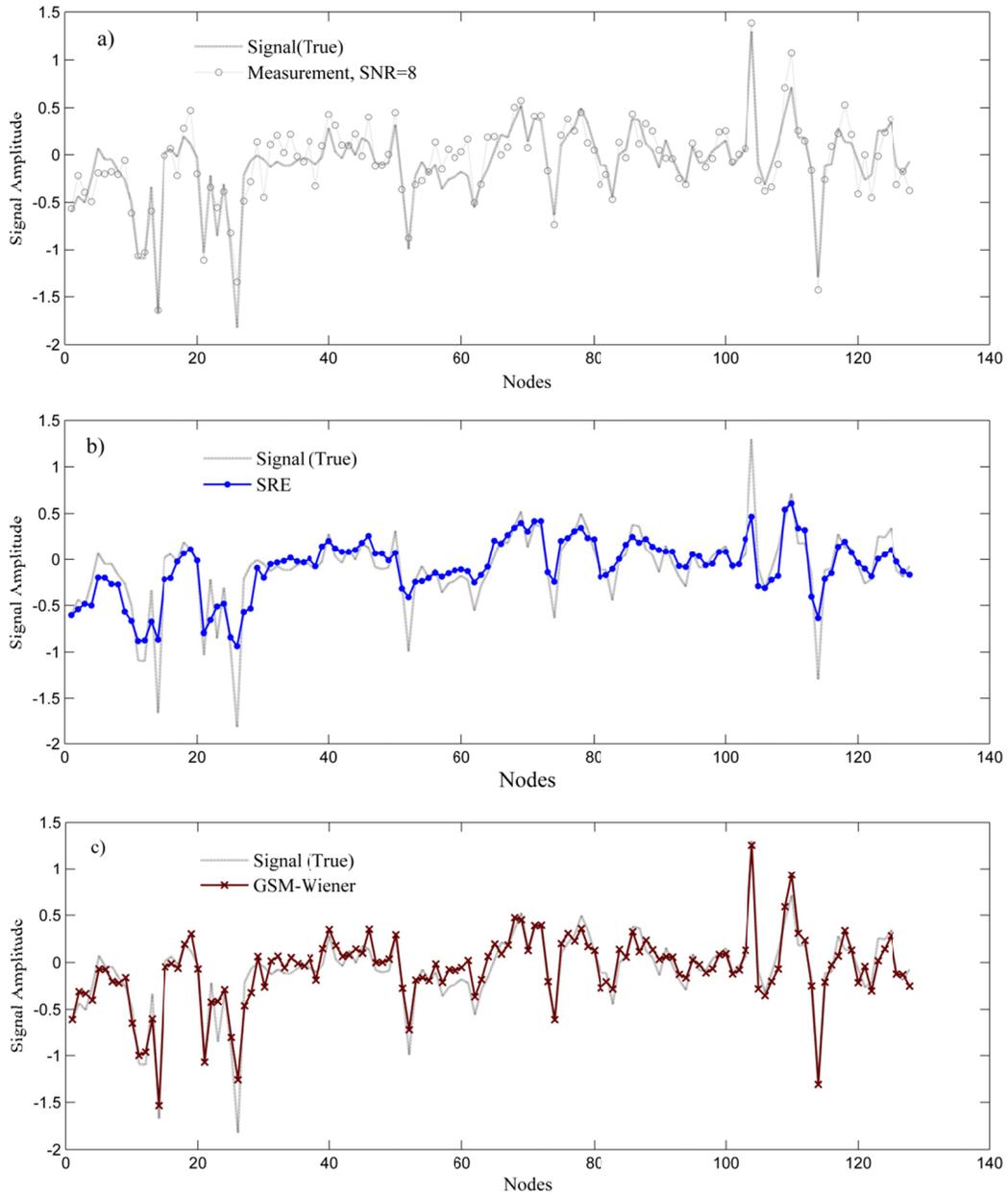


Figure 4.8: (a) A synthetic simulation of a 1D multiscale GSM using a log-normal multiplier and its noisy measurement. The sample path is generated with $\mathbf{A}(s) = 0.9$ and $H = 0.7$. The variance of the log-multiplier is set such that to reproduce a sample path with a marginal histogram similar 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 SRE (b) and GSM-Wiener (c) shows the superior performance of the GSM-filter especially over the singular points and extreme fluctuations.

4.4 GSM Multi-sensor Fusion of Precipitation Data

4.4.1 Concepts

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 (*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 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 spatio-temporal scales. For instance, it has been concluded that direct comparison of the real time rain-gauge data as a point-wise representation of the highly irregular rainfall process with a remotely sensed areally averaged rainfall products may not lead to a conclusive statistical characterization of the measurement error (see, *Ciach and Krajewski, 1999; Wang and Wolff, 2009*). Indeed, for a proper characterization of the measurement error norms, the spatio-temporal 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.

To address the scaling issues involved in precipitation multi-sensor 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 (NEXRAD) and low (TRMM-PR) resolution 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 (Figure 4.9).

Basically, by comparing different sources of rainfall measurements, three different error covariance matrices need to be characterized for proper implementation of the proposed

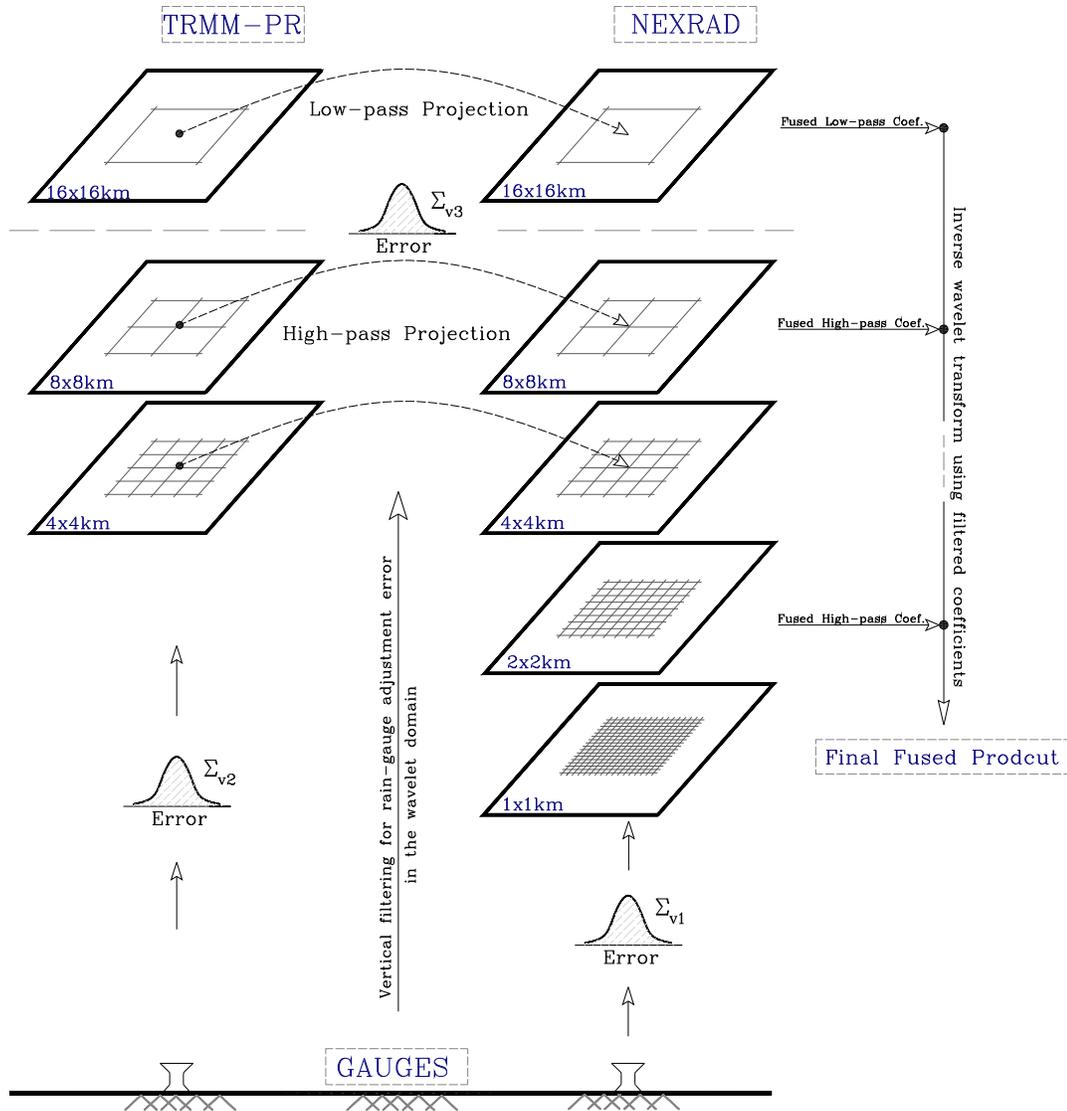


Figure 4.9: Schematic of the proposed GSM-fusion technique in the wavelet domain. In the “*Vertical Filtering*” stage, filtering for rain-gauge adjustment error is performed on each available unbiased source of remotely sensed precipitation reflectivity image; i.e., TRMM-PR and NEXRAD, independently. 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.

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.

The proposed GSM multi-sensor multiscale methodology consists of two major steps namely *Vertical Filtering* and *Lateral Projection* (Figure 4.9). 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 (4.21), the rain-gauge adjustment errors are first filtered out of the wavelet high-pass subbands of the TRMM and NEXRAD data from fine-to-coarse scales, 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 (Figure 4.9).

Theoretically speaking, using equation (4.21) in lateral projection phase, we can keep decomposing the coarse and low resolution 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 non-zero 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, the fused low-pass coefficients can be fused using a conventional least squares formalism as expressed in (4.3) and (4.4)

$$\mathbb{E} [\mathbf{x}_l | \mathbf{y}_l] \cong m_{\mathbf{x}_l} + \Sigma_{\mathbf{x}_l} (\Sigma_{\mathbf{x}_l} + \Sigma_{v3})^{-1} (\mathbf{y}_l - m_{\mathbf{y}_l}), \quad (4.25)$$

where, \mathbf{y}_l and $m_{\mathbf{y}_l}$ denote the NEXRAD low-pass coefficients and their mean in a local spatial neighborhood; $m_{\mathbf{x}_l}$ and $\Sigma_{\mathbf{x}_l}$ are thus the average and covariance of the TRMM low-pass coefficients in that neighborhood, respectively. Obviously, as there is not any 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 in (4.25).

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.

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 intense rain-cells.

It is also observed that over the decomposition levels 2 to 4 (i.e., 4 to 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.

4.4.2 A Case Study on Precipitation Data

The TRMM-PR and NEXRAD coincidental reflectivity image of a storm on 1998/06/28 over the HSTN site, used for the SRE implementation (in section 4.3.2), is also 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 $\text{SNR} = 13.0$ dB for construction of the diagonal rain-gauge adjustment error covariance $\Sigma_{\mathbf{v}_1}$, Figure 4.10 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 filtering would be done at minimum expense corresponding to the removal of the frequent local extremes of interest.

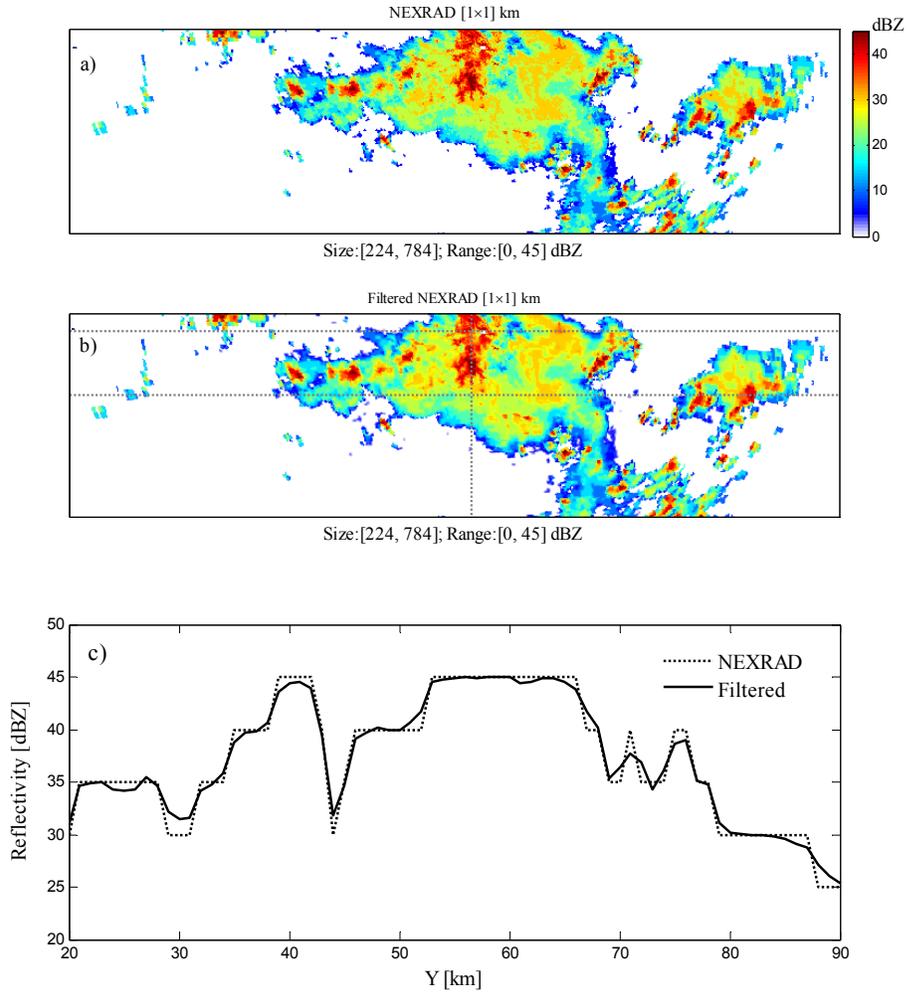


Figure 4.10: (a) Original NEXRAD reflectivity image over the HSTN site on 1998/06/28 at 18:13:00 UTC and (b) the rain-gauge filtered image assuming $\text{SNR} = 13.0$ dB. (c) A vertical transect of the image, bounded by two horizontal dashed lines in (b), shows a 1D representation of the rain-gauge filtering effects on the reflectivity image.

Figure 4.11 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 (4.25). Qualitative comparison of Figure 4.11a and b, 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 4.11c 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.

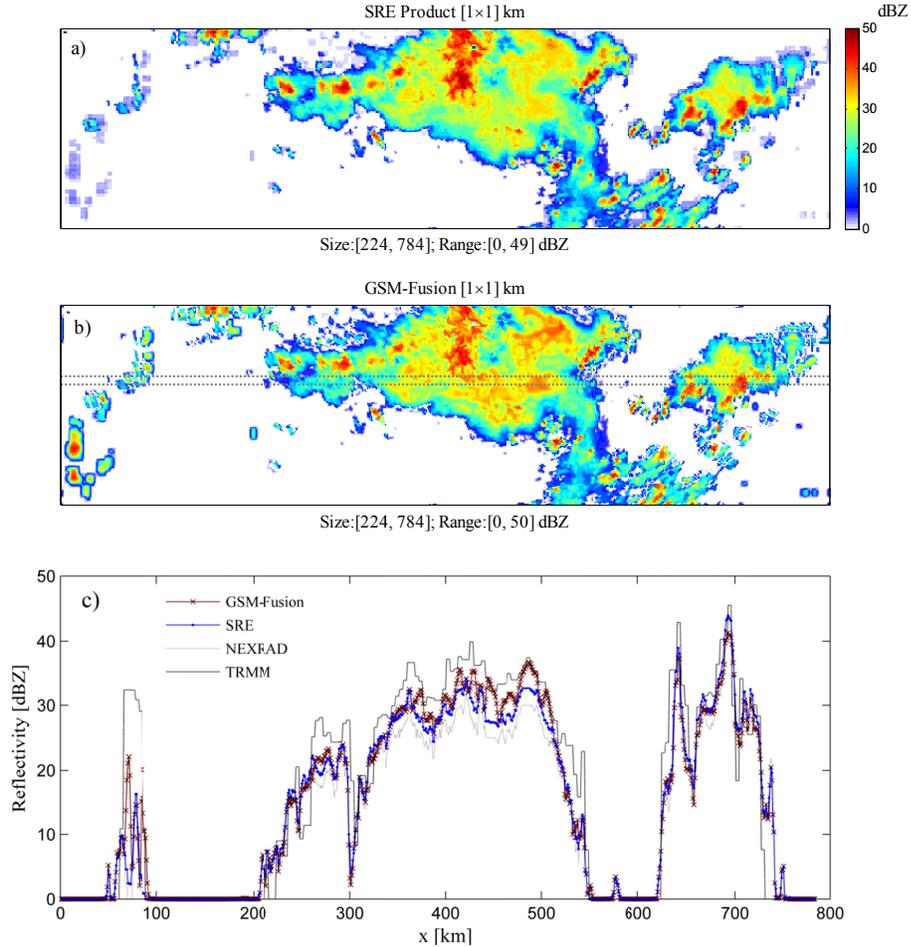


Figure 4.11: (a) SRE-Fusion of the NEXRAD and TRMM-PR reflectivity snapshots over the HSTN site on 1998/06/28 at 18:13:00 UTC, assuming $\text{SNR} = 11.0$ dB; (b) GSM-fusion of the same reflectivity snapshots, assuming $\text{SNR} = 24$ dB for rain-gauge errors (Σ_{v_i} , $i = 1, 2$) and $\text{SNR} = 16$ dB for Σ_{v_3} . Using the same order of error variance, the GSM-fusion extracts 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 (b), compares qualitatively the fusion performance of different fusion methods. For quantitative comparison please refer to Table 4.2.

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 4.2, shows the entropy, F -norm and statistical measures of the associated rain rates of the two fused images. To 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 the GSM product shows a larger mean and larger maximum values compared to the SRE method (see Figure 4.11c). 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 analyses of other storm cases and by using as “ground truth” highly accurate rain-gauge data.

| | 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 |
| Max rain rate [mm/h] | 60.06 | 27.86 | 55.59 | 48.45 |

Table 4.2: Statistical measures of the fusion products (1×1 km) by the GSM-Wiener and SRE methodologies are compared at the same order of measurement error between NEXRAD and TRMM-PR images. Entropy and Frobenius norms are row-wise normalized by the maximum values.

Chapter 5

Sparse precipitation downscaling

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

5.1 Introduction

The coarseness of the native spatio-temporal scales of some remotely sensed atmospheric state variables and also General Circulation Models (GCM) of the climate system is often restrictive for many terrestrial applications (*Wilby et al.*, 1998a,b). *Hostetler* (1994) reported that the largest prediction and parameter estimation error of both GCMs and hydrologic models occurs at the interfacial scales at which terrestrial and climate models need to be linked.

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

During the past decades, special attention has been devoted to developing different classes of stochastic models, $\mathcal{M}(\mathcal{Y}, \theta, \zeta)$, to reproduce consistently the missing high-frequency part of the rainfall fields, from a low-resolution input (\mathcal{Y}) using a set of parameters (θ) and a random generator (ζ) (*Lovejoy and Mandelbrot, 1985; Gupta and Waymire, 1993; Perica and Foufoula-Georgiou, 1996a; Menabde et al., 1997; Deidda et al., 2006; Deidda, 2000*). To parameterize the random generator, a large number of these models often relies on the observed statistical regularities of the precipitation fields such as power law Fourier spectrum and scaling of the higher-order statistical moments of rainfall fluctuations, often referred to as multifractality. Although, these stochastic generators can be designed to be statistically consistent with the observed small-scale rainfall variability, the spatial coherency of the generated high-resolution fields is often not satisfying. Blockiness and non-uniqueness of the downscaled fields are typical drawbacks of this sort of stochastic generators, enforcing their interpretation in an ensemble mode; see *Ferraris et al. (2003)* for a review of available statistical/mathematical downscaling models.

Apart from hydrometeorological applications, enhancing the resolution and quality of low-resolution imaging, often referred to as super-resolution in the image processing community, has been a subject of interest for many years. The state-of-the-art approaches in this area are currently relying on sparse approximation (*Mallat and Zhang, 1993; Chen et al., 1998*) and statistical learning (*Donoho and Stark, 1989; Tibshirani, 1996*), which interestingly have some roots in subsurface geophysics (*Claerbout and Muir, 1973*). The central results suggest that a high-resolution signal can be recovered from a low-resolution one via solving a

constrained optimization, provided that the signal of interest exhibits a sufficiently sparse representation (a large fraction of close to zero coefficients) in an appropriately chosen domain (e.g., wavelet) (e.g., *Yang et al.*, 2010; *Mallat and Yu*, 2010; *Elad*, 2010; *Zeyde et al.*, 2010).

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

Section 5.2 is devoted to explain notation and terminology. In Section 5.3, a sparse inverse estimator and its connection with other estimation paradigms is explained. Sparsity of the spatial rainfall images is elaborated in the wavelet domain in Section 5.4. We provide quantitative evidence in Section 5.5, that the organized small-scale features of a storm snapshot may recur within the larger storm domain or also across different storm environments. This indicates that the small-scale rainfall features which might have been lost in the low-resolution sensing and/or modeling, might be able to be reconstructed using a database of appropriately collected high and low-resolution precipitation images. In Section 5.6, the implementation details of the sparse inverse estimator for rainfall downscaling are explained. Section 5.7, points out some concluding remarks and future research directions.

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

5.2 Sparse Inverse Estimator

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

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{e}. \quad (5.1)$$

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

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

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

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

Substitution of \mathbf{x}_S into (5.1) results in

$$\mathbf{y} = \mathbf{H}\Phi \mathbf{c} + \mathbf{e}', \quad (5.2)$$

where $\mathbf{e}' = \mathbf{H}(\mathbf{x} - \mathbf{x}_S) + \mathbf{e}$. This implies that the low-resolution observation \mathbf{y} has similar representation in a transformed dictionary $\Psi = \mathbf{H}\Phi \in \mathbb{R}^{n \times M}$ which is naturally more redundant than Φ ($n \leq m$). Thus, by finding an appropriate sparse representation $\hat{\mathbf{c}}$ of \mathbf{y} in the so called *low-resolution dictionary* Ψ , the same coefficients $\hat{\mathbf{c}}$ can be used to reconstruct the desired high-resolution signal \mathbf{x} using the corresponding atoms of the *high-resolution dictionary* Φ that is, $\hat{\mathbf{x}} = \Phi \hat{\mathbf{c}}$. Given only equation (5.2), due to the redundancy of Ψ and possible observation noise, the representation is not obviously unique. The a priori assumption on sparsity of the signal, naturally implies that among many solutions, those with the minimum number of non-zero elements are the optimal ones. Following that the estimated representation coefficients shall obey the fidelity constraint imposed by (5.2), the optimal coefficients can be obtained by solving

$$\mathbf{c} = \arg \min_{\mathbf{c}} \|\mathbf{c}\|_0 \quad \text{subject to} \quad \|\mathbf{y} - \Psi \mathbf{c}\|_2 \leq \epsilon', \quad (5.3)$$

where $\epsilon' = \|\mathbf{e}'\|_2$. However, this gives rise to a Non-deterministic Polynomial-time hard

(NP-hard) problem for which a direct solution is still hopeless.

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

Chen et al. (1998) proposed a new strategy known as the Basis Pursuit (BP) which suggests that under some sparsity constraints on the solution, in lieu of $\|\mathbf{c}\|_0$ the closest convex penalization function $\|\mathbf{c}\|_1$ leads to the same sparsest solution. Using a Lagrangian penalization, the ℓ_1 -regularized dual version in (5.3) follows

$$\hat{\mathbf{c}} = \arg \min_{\mathbf{c}} \lambda \|\mathbf{c}\|_1 + \frac{1}{2} \|\mathbf{y} - \Psi \mathbf{c}\|_2^2, \quad (5.4)$$

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

It is worth noting that, in the statistical machine learning community, the problem of solving (5.4) is the Lagrangian form of the known Least Absolute Shrinkage and Selection Operator (LASSO) by (*Tibshirani, 1996*). From the Bayesian statistical standpoint, the minimization in (5.4) is also equivalent to the Maximum A Posteriori (MAP) estimation of the coefficients, $\hat{\mathbf{c}}_{MAP} = \arg \max_{\mathbf{c}} \{p(\mathbf{y}|\mathbf{c})p(\mathbf{c})\}$, given that the additive noise \mathbf{e} is Gaussian and the a priori distribution of the coefficients can be well explained by the family of Laplace densities, i.e. $p(c_i) \propto \exp(-|c_i|)$ (*Lewicki and Sejnowski, 2000*). In the context of the MAP estimator, the terms $\|\mathbf{c}\|_1$ and $\|\mathbf{y} - \Psi \mathbf{c}\|_2^2$ in (5.4) represent the contribution of the logarithm of the prior density $p(\mathbf{c})$ and the log-likelihood function $\log p(\mathbf{y}|\mathbf{c})$, respectively. Note that, the logarithm of the Laplace and Gaussian density is a monotonic and concave function, and by a sign change, the minimization in (5.4) and the MAP estimator are equivalent.

Effective implementation of the described sparse recovery method requires that the dictionary pairs for the high (Φ) and low-resolution (Ψ) atoms have been appropriately determined which will be briefly explained in the context of our study in the following sections.

5.3 On Sparsity of Rainfall Images

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

In this study we use the undecimated Haar wavelet transform by *Nason and Silverman (1995)*, with a single layer of decomposition, to demonstrate sparsity of the wavelet coefficients of precipitation images. The transform coefficients at different subbands (directions) are obtained by filtering the image with the one dimensional low-pass $[+1, +1]/2$ and high-pass $[+1, -1]/2$ Haar kernels and their transpose successively. It is demonstrated in Figure 5.1a that the majority of the wavelet coefficients in the Horizontal (H), Vertical (V) and Diagonal (D) subbands are close to zero. As is evident, the histogram of the wavelet coefficients (see, Figure 5.1b for the horizontal subband) also shows concentration of the coefficients around the origin. Keeping only the top largest 20% of the wavelet coefficients in absolute value and setting the rest to zero via a hard thresholding (see, *Donoho, 1995*); Figure 5.1c shows the reconstructed rainfall image. Using only about 20% of the entire wavelet coefficients, the reconstructed rainfall image contains 99.8% of the total energy of the original image (sum of pixel-wise squares) and is visually indistinguishable from the original one. This evidence confirms that the rainfall reflectivity images exhibit considerable sparseness in the wavelet domain which is a direct consequence of the piecewise smoothness and spatial coherency of the precipitation images.

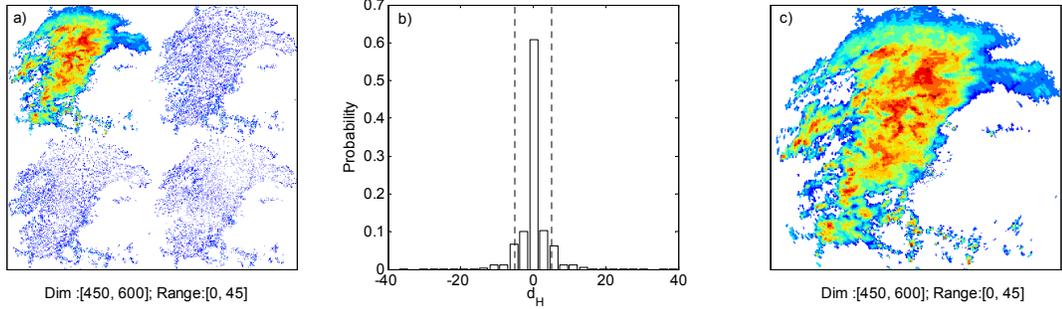


Figure 5.1: Evidence on sparsity of precipitation images in the wavelet domain: (a) A storm reflectivity snapshot at the TRMM GV-site in Houston, TX (HSTN) on 1998/11/13 (00:02:00 UTC) and the absolute values of the wavelet coefficients in the horizontal (H), vertical (V) and diagonal (D) subbands, (b) probability histogram of the horizontal wavelet coefficients (d_H) and (c) the reconstructed field using the top largest 20% of the wavelet coefficients in absolute values. The bounded area by the dashed lines in (b) contains 80% of the wavelet coefficients that was set to zero for reconstruction of the reflectivity image, shown in (c).

5.4 Recurrence of Rainfall Patches and Group Sparsity

The observed precipitation patterns are the result of a wide spectrum of highly non-linear and complex atmospheric processes operating over different range of scales, from micro-physical to regional and global. As a result, it is hard to imagine that we might be able to find very similar (in mathematical norms) precipitation patterns over relatively large storm-scales. However, finding similar features over sufficiently small sub-storm scales seems more feasible. The goal of this section is to provide evidence for the recurrence of small-scale “patches” of precipitation features (patches being small neighborhoods of the sampled rainfall fields) within different regions of the same storm or across different storm environments. The motivation is that, if for each patch of an arbitrary storm snapshot, we can find at least a few similar patches in a representative precipitation database, this will promise that the small-scale precipitation variability in that storm may be well approximated by an optimal and possibly linear combination of those similar patches.

To this end, a sample storm reflectivity image is selected (see Figure 5.2a) and N “sample patches” of size 5×5 , $\{\mathbf{y}_i \in \mathbb{R}^{25}\}_{i=1}^N$, are extracted over regions of significant magnitudes of rainfall gradients (see Figure 5.2b). The precipitation gradient was selected here only as a metric to guide the sampling of important patches as, naturally, localized regions of high rainfall gradients contain significant high-frequency information and define the within-storm spatial geometry and variability. Then, the high-pass fluctuations of all rainfall patches, the so called “rainfall features” $\{\mathbf{y}'_i\}_{i=1}^N$, are obtained by subtracting the local mean of each patch, that is $\mathbf{y}'_i = \mathbf{y}_i - \bar{\mathbf{y}}_i$. The same sampling procedure is performed to extract same size *test patches* $\{\mathbf{z}'_j\}_{j=1}^M$ of rainfall features, where naturally $M \gg N$, from a database

containing an independent set of about 100 storm NEXRAD (level III) reflectivity snapshots over the TRMM satellite Ground Validation (GV) site in Huston, TX (see Chapter 3 and for (Ebtehaj and Foufoula-Georgiou, 2011a) for a detailed description of the data set). For each i^{th} sample patch, we computed the Euclidean distance $\mathbf{d}_{ij} = \left\| \mathbf{y}'_i - \mathbf{z}'_j \right\|_2$ with all of the test patches and appropriately normalized it to the range of $[0, 1]$. Here, the similarity measure among the patches is defined as the normalized Euclidean distance below a certain threshold value η . Choosing the threshold values η , linearly spaced from 0.01 to 0.05 in five intervals, we determined the number of sample patches n_Λ in the sample rainfall image, that we could find at least an m_Λ number of similar test patches in our database. We then reported m_Λ versus the ratio n_Λ/N , for different choices of η in Figure 5.2c. As the ratio n_Λ/N is always positive and ranges between $[0, 1]$, it can be interpreted as a lower bound for the probability of finding at least an m_Λ number of similar patches within any larger database that contains our test database as a subset. Figure 5.2c suggests that the probability of finding a small number (small m_Λ) of very similar (small η) patches is high and drops rapidly as the number of similar patches is increased. This finding indicates that the rainfall fluctuations are not only pixel-wise sparse, but also may exhibit structured sparsity in a small neighborhood of rainfall intensities, which we refer to as *group sparsity*. As is evident, specializing the database for particular environmental conditions (e.g., orographic precipitation), may increase the chance of finding a very small number of similar features and thus the efficiency of group sparsity.

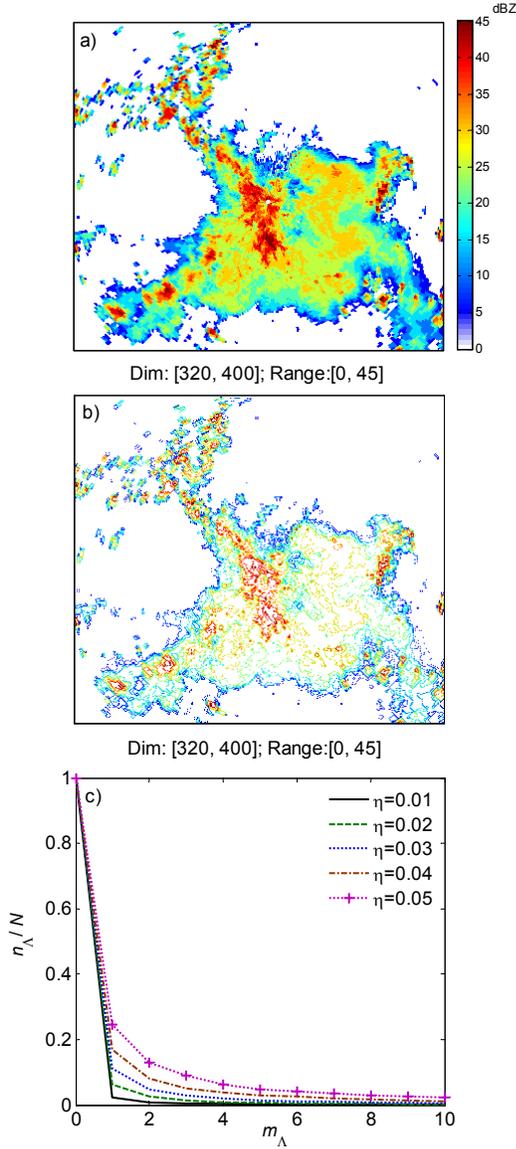


Figure 5.2: Recurrence of small-scale precipitation patches and group sparsity: (a) a sample precipitation reflectivity image over the TRMM (HSTN) GV-site on 1998/06/28 (18:13:00 UTC), (b) regions of high gradient (top 25%) used for sampling of important patches, and (c) a probability measure (n_Λ/N) of finding at least m_Λ number of similar patches (i.e., close in Euclidean distance determined by η) in the selected test database.

5.5 Precipitation Sparse Downscaling

5.5.1 Conceptual Framework

The proposed sparse restoration of a high-resolution rainfall image \mathcal{X} from a coarse-scale observation \mathcal{Y} , requires the availability of a representative set of N_T coincidental pairs of low and high-resolution rainfall images $\{(\mathcal{Z}_l, \mathcal{Z}_h)\}_{l,h=1}^{N_T}$, as a training set. Note that throughout

this study, we have adopted a simplified notation by which the resolution indices (l) and (h), are also served as counting indices for the number of used images. Ideally, the low-resolution part of the training set can be extracted from a spaceborne sensor or a GCM-output; while the high-resolution counterparts can be the result of a coincidental ground based high-resolution sensing device or a fine-scale predictive model (e.g., NEXRAD or a fine grid cloud resolving numerical model). Elaborating on the merits of the methodology, in this study we synthetically generated the low-resolution rainfall reflectivity images by smoothing and downsampling the available NEXRAD dataset. For instance, a low-resolution dataset $\{\mathcal{Z}_l\}_{l=1}^{N_T}$ at scale 8×8 km is obtained by first smoothing each high-resolution NEXRAD image (1×1 km) using an average spatial filter of size 8×8 and then downsampling by an integer factor of $\sqrt{s} = 8$, called the scaling ratio. Note that, given the pair of high and low-resolution training images, a set of “residual images” can be produced as

$$\mathcal{R} = \mathcal{Z}_h - \mathbf{Q}\mathcal{Z}_l, \quad (5.5)$$

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

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

$$\begin{aligned} f_1 = f_2^T &= [+1, -1] \\ f_3 = f_4^T &= [+1, -2, +1]. \end{aligned} \quad (5.6)$$

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

for brevity.

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

Deciding to choose the low-resolution rainfall patches of size $\sqrt{n} \times \sqrt{n}$ making thus $\mathbf{y} \in \mathbb{R}^n$, the sample and training feature patches can be extracted from the feature images $\{\mathcal{Y}', \{\mathcal{Z}'_l\}_{l=1}^{N_T}\}$ and stacked in \mathbf{y}' and $\mathbf{z}'_l \in \mathbb{R}^m$, where owing to the suggested filters in (5.6), here $m = 4n$. The column-wise concatenation of \mathbf{z}'_l in $\Psi \in \mathbb{R}^{m \times M}$ forms the desired empirical low-resolution rainfall dictionary. Now, in the feature space, the idea is to find a few columns of the low-resolution dictionary Ψ that their linear combinations can well approximate the low-resolution feature patch \mathbf{y}' of interest. This idea can be recast into the following sparse encoding problem, similar to equation (5.3), which indeed assumes $\mathbf{y}' = \Psi \mathbf{c} + \mathbf{e}$ with $\epsilon = \|\mathbf{e}\|_2$,

$$\hat{\mathbf{c}} = \arg \min_{\mathbf{c}} \|\mathbf{c}\|_0 \quad \text{subject to} \quad \|\mathbf{y}' - \Psi \mathbf{c}\|_2 \leq \epsilon. \quad (5.7)$$

As explained in Section 2, the estimated coefficients for each patch can now be used to combine the corresponding columns of a high-resolution dictionary to recover the relevant high-frequency features of interest for that patch. To this end, due to the size enlargement of the high-resolution field by a factor of s , the patches of size $\sqrt{sn} \times \sqrt{sn}$ need to be extracted from the residual images $\{\mathcal{R}_l\}_{l=1}^{N_T}$, obtained from the training dataset, and be concatenated column-wise, in a vector form $\mathbf{r} \in \mathbb{R}^{sn}$, to shape the desired empirical high-resolution dictionary, i.e., $\Phi \in \mathbb{R}^{sn \times M}$. Recall that the columns of these empirical dictionaries (Ψ, Φ) contain the patches of low-resolution rainfall features $\mathbf{z}'_l \in \mathbb{R}^m$ and the corresponding high-resolution rainfall residuals $\mathbf{r} \in \mathbb{R}^{sn}$, respectively. As a result, the obtained sparse representation in (5.7) is then being used to combine the columns of Φ and restore the high-resolution rainfall patches as $\hat{\mathbf{x}} = \mathbf{Q}\mathbf{y} + \hat{\mathbf{r}}$, where $\hat{\mathbf{r}} = \Phi \hat{\mathbf{c}}$. Applying this inverse estimator for all overlapping patches of the given low-resolution rainfall image, the entire high-resolution rainfall image can then be recovered. To reduce blocking artifacts and impose more coherency on the downscaled field, the estimated patches are averaged over their overlapping areas. The entire methodology, referred to as SPaD (**S**parse **P**recipitation **D**ownscaling), is also sketched algorithmically as follows:

Algorithm 5.1 : Sparse Precipitation Downscaling (SPaD)

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

Steps:

1. Apply the filters in (5.6) to \mathcal{Y} and $\{\mathcal{Z}_l\}_{l=1}^{N_T}$ to obtain the feature images $\{\mathcal{Y}', \{\mathcal{Z}'_l\}_{l=1}^{N_T}\}$ (i.e., 4 output images per 1 input image).
2. Extract all $\sqrt{n} \times \sqrt{n}$ overlapping patches from $\{\mathcal{Z}'_l\}_{l=1}^{N_T}$ (per the 4 filters) to shape the low-resolution dictionary $\Psi \in \mathbb{R}^{m \times M}$.
3. Compute $\{\mathcal{R}_l\}_{l=1}^{N_T}$ in (5.5) and extract all $\sqrt{sn} \times \sqrt{sn}$ patches to form the high-resolution rainfall dictionary $\Phi \in \mathbb{R}^{sn \times M}$.
4. **For** each feature patch $\mathbf{y}' \in \mathbb{R}^m$
 - Solve $\hat{\mathbf{c}} = \arg \min_{\mathbf{c}} \|\mathbf{c}\|_0$ s.t. $\|\mathbf{y}' - \Psi \mathbf{c}\|_2 \leq \epsilon$
 - Compute the residuals $\hat{\mathbf{r}} = \Phi \hat{\mathbf{c}}$
 - Compute the high-resolution rainfall patch $\hat{\mathbf{x}} = \Phi \mathbf{y}' + \hat{\mathbf{r}}$.

End

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

5.5.2 SPaD via Dictionary Learning

Collecting a large set of training patches of low-resolution precipitation features $\{\mathbf{z}'_l\}_{l=1}^M$ as the atoms of Ψ , often leads to a very large-scale sparse coding problem, which cannot be easily handled by the Basis Pursuit type of methods in a reasonable time and with sufficient accuracy. Often, when the number of dictionary atoms is much larger than the patch size, the maximum absolute value of the cross correlation between the atoms of Ψ , called the Mutual coherence, tends to unity and a stable sparse solution is hardly achievable via l_1 -norm convex relaxation of equation (5.7) in (5.4), (e.g., *Kim et al.*, 2007; *Elad*, 2010). As an alternative to a crude selection of the sample precipitation features as atoms of the empirical high and low-resolution dictionary, one can use a tunable and compact selection, in which the atoms are well adapted to the patches of rainfall images based on a learning procedure. The core idea is to optimally estimate a pair of reduced dimension low and high-resolution dictionary $(\Psi_r, \Phi_r) \in \mathbb{R}^{m \times K}$, where $K \ll M$, given a large number of low and high-resolution training patches. Typically, the learned reduced size low-resolution dictionary Ψ_r is obtained such that it can reproduce each training patch \mathbf{z}'_l by a sparse combination of its atoms in a linear setting $\mathbf{z}'_l = \Psi_r \mathbf{c}_l + \mathbf{e}$ with added noise \mathbf{e} , where

$\mathbf{c}_l \in \mathbb{R}^K$, $\|\mathbf{c}_l\|_0 \ll m$. Note that in this setting the goal is to have an optimal estimate of the Φ_r and the representation coefficients \mathbf{c}_l , simultaneously. It naturally follows to solve

$$\min_{\Psi_r, \mathbf{c}_l} \sum_{l=1}^M \|\mathbf{z}'_l - \Psi_r \mathbf{c}_l\|_2^2 \quad \text{subject to} \quad \|\mathbf{c}_l\|_0 \leq \text{const.} \quad \forall l, \quad (5.8)$$

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

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

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

Here, the dictionary Ψ_r is a reduced version of Ψ which can sparsely approximate the low-resolution features of all M training patches. By construction, using the same representation for high-resolution recovery, estimation of the reduced high-resolution dictionary Φ_r only requires the second step of a typical dictionary learning process (Zeyde et al., 2010) and follows:

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

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

5.5.3 Results

5.5.3.1 Inputs and Parameters

For construction of the training database, two hundred independent rainfall reflectivity images at resolution 1×1 km were selected over the two GV-sites of the TRMM satellite in Houston, Texas (HSTN) and Melbourne, Florida (MELB) from 1997 to 2010; see chapter 3. These high-resolution images are first smoothed and downsampled by scaling ratios $\sqrt{s} \in \{4, 8, 16\}$ to form the low-resolution components of the training set. Then, these low-resolution images are transformed into the feature space by applying the explained set

of filters in (5.6). Note that, as the NEXRAD data are in a resolution of $\sim 1 \times 1$ km, these scaling ratios also refer to the spatial grid spacing of the low-resolution rainfall data in kilometers.

Selection of the patch size is obviously an input parameter in the presented SPaD framework. Although the optimal patch size cannot be theoretically determined, for larger patch size the downscaling result is generally smoother. We empirically found that SPaD performs well for patch sizes in the range of 3×3 to 7×7 . Here we report the results using overlapping patches of size 3×3 . At least $M = 150, 100$ and 50 thousands of training high and low-resolution precipitation patch pairs were extracted to form the initial low and high-resolution empirical dictionaries (Ψ, Φ) for the selected scaling ratios (i.e., resolutions of 4,8 and 16 km), respectively. To obtain the residual images in (5.5), we adopted a bicubic interpolation operator. While extracting patches from each rainfall image, to screen the possible degenerate patches due to the zero rainfall background effect and piecewise constant property of rainfall images, those patch pairs ($\mathbf{z}'_l, \mathbf{r}$) were selected for which the variance of the residual patch \mathbf{r} exceeded a certain threshold (i.e., 50th percentile of the variance of all residual patches).

To implement the sparse estimator and solve problem (5.7), we examined the performance of both a greedy and also an ℓ_1 -regularization strategy, by employing the Orthogonal Matching Pursuit (OMP) and the ℓ_1 -regularized Least Squares (known as the ℓ_1 - ℓ_s) by *Kim et al.* (2007). The later method, is an interior point method which solves a quadratic programming reformulation of equation (5.7), using a preconditioned conjugate gradient method.

Employing the OMP in a greedy implementation, we set a fixed number of non-zero representation coefficients (i.e., $\|\mathbf{c}\|_0 = 3$) as the stopping criterion in (5.7) to find relevant atoms of the empirical dictionaries (Ψ, Φ). However, for the convex ℓ_1 -regularized strategy, we used a pair of learned low and high-resolution dictionaries (Ψ_r, Φ_r) of different sizes $K \in \{256, 512, 1024\}$ for the studied scaling ratios, respectively. These dictionaries were learned by solving problem (5.8) with a fixed number of non-zero elements in \mathbf{c}_l (i.e., $\|\mathbf{c}_l\|_0 = 3$) and employing the OMP as a sparse coding method. The chosen size of the dictionary is a design issue which requires to consider a trade-off between the accuracy of the recovery and the computational cost. It is empirically found that the selected size of the dictionaries leads to decent results with modest computational cost. Note that, at any scale of interest, the first guess of the reduced low-resolution dictionary for the explained iterative learning, is randomly selected from rainfall patches in the feature space (i.e., columns of Ψ). Figure 5.3 shows the initial and trained atoms of the learned high-resolution precipitation dictionary, designed for downscaling from scale 4-to-1 km in grid spacing. For selection of the regularization parameter λ , we simply chose $\lambda = 0.1 \|\Psi_r^T \mathbf{y}\|_\infty$, as recommended by *Kim et al.* (2007), which we also found empirically to perform well.

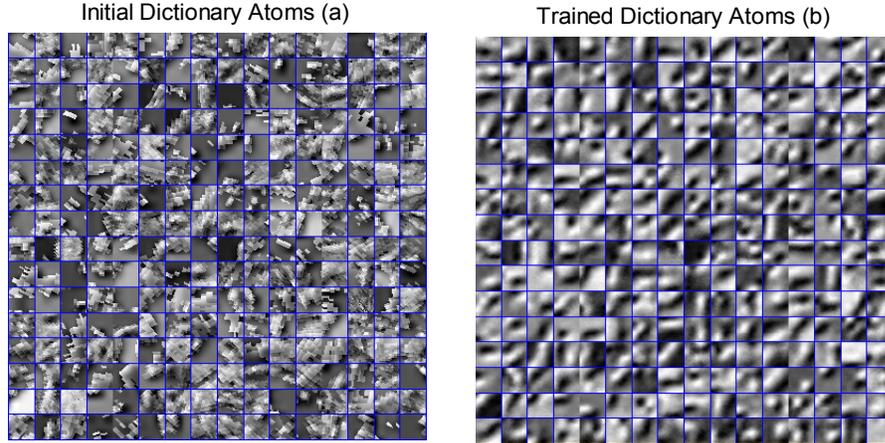


Figure 5.3: (a) Initial rainfall residual patches $\mathbf{r} \in \mathbb{R}^{sn}$ (i.e., patch size 3×3 and downscaling ratio $\sqrt{s} = 4$) at scale 1×1 km, randomly sampled from 150000 training rainfall patches to learn a high-resolution dictionary Φ_r of rainfall features with $K = 256$ atoms for downscaling from resolution 4-to-1 km in grid spacing, and (b) the learned precipitation dictionary after 50 iterations.

5.5.3.2 Downscaling Performance and Error Analysis

The base-reflectivity snapshots over the HSTN site for the storm of 1998/11/13 (00:02:00 UTC) and 1998/06/28 (18:13:00 UTC), shown in Figure 5.1 and 5.2, were selected to study the quantitative performance of the proposed downscaling scheme. In contrast to the stochastic type of downscaling models in which the high-frequency details are generated randomly to reproduce high-resolution realizations with consistent marginal statistics, the proposed method recovers the optimal high-resolution estimate via a supervised learning from our a priori knowledge, obtained by a set of historical observations. As a result, the final product is not blocky, compared to the stochastic models, and exhibits reduced estimation error. Here, by no means we intend to exhaustively compare the SPaD with stochastic downscaling models; this can be the topic of future studies. We note however that, due to the intrinsic randomness in the stochastic downscaling models, a comparative error analysis is not straightforward. Indeed in these type of models at each realization, the error of estimation (e.g., in a mean square error sense) increases with high probability as the added details are quite random and the downscaled field only preserves marginal statistics and not exact geometric and coherent structures of the storm image. At the same time, error estimation in an ensemble sense does not add much insight as the ensemble average tends to the original input image when the number of ensembles goes to infinity (e.g., consider the case of multiplicative random cascades in canonical form). Therefore, when using these type of stochastic models, the estimation error can not be readily improved even in an ensemble sense, while the comparison also remains ambiguous in the single realization sense.

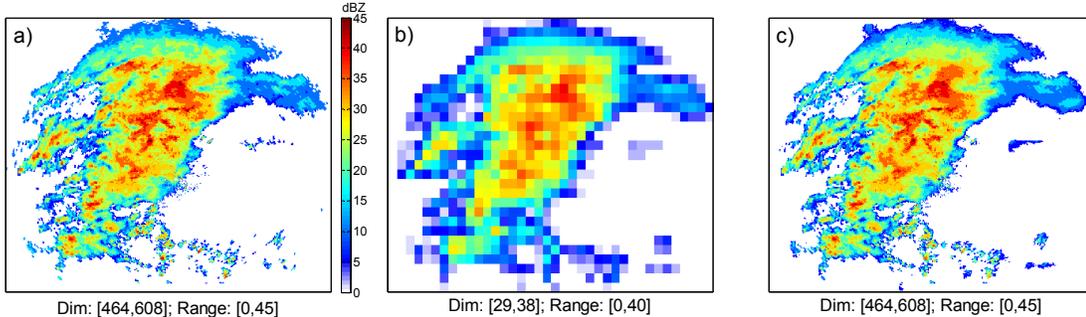


Figure 5.4: The SPaD performance in a case when the exact high and low-resolution patches are included in the training database. (a) The original high-resolution NEXRAD reflectivity image at resolution 1×1 km, (b) the coarse scale version of the image at scale 16×16 km generated by smoothing and downsampling, and (c) the SPaD results. The pixel-wise Mean Squares Error (MSE), by comparing the original high-resolution with the SPaD output, is 2.10 in $[\text{dBZ}]^2$ which is around 1% of the mean of the squares of rainfall reflectivity values.

First to show the effectiveness of the SPaD, we tried to downscale a 16×16 km coarse-scale version of the snapshot on 1998/11/13 while including the corresponding high and low-resolution patches of the rainfall image in the training dataset. The results in Figure 5.4 show how well this method can perform by finding the relevant patches, recover the high-frequency features and estimate the small-scale geometrical structures of the rainfall image of interest for a relatively coarse scale observation. This result is promising that it demonstrates the algorithmic success of our methodology, provided that a suitable prior knowledge is provided in the training database.

Figure 5.5 shows the performance of the SPaD in a case where the image of interest is not itself part of the dataset, but rather an independent training database is used. As is evident, the effectiveness of the algorithm degrades as the coarseness of the original image increases (e.g., 16×16 km) and the recovered geometrical structure of the small-scale features is prone to be distorted significantly. Although, the reference high-resolution rainfall field is of course not available in practical cases, it is instructive here to have a quantitative assessment of the downscaling success by comparing the SPaD results $\hat{\mathcal{X}}$ with the original high-resolution rainfall image \mathcal{X} . Defining the error matrix $\mathcal{E} = \mathcal{X} - \hat{\mathcal{X}}$, we have chosen some conventional fidelity metrics including; (a) Normalized Bias Ratio: $\text{NBR} = \frac{\overline{\mathcal{E}}}{\overline{\mathcal{X}}} \times 100$, where $\overline{(\cdot)}$ denotes the entry-wise expected value (mean); (b) Mean Square Error: $\text{MSE} = \overline{\mathcal{E}^2}$; (c) Mean Absolute Error: $\text{MAE} = \overline{|\mathcal{E}|}$, where $|\cdot|$ denotes the entry-wise absolute value, and (d) Peak Signal-to-Noise Ratio: $\text{PSNR} = 20 \log \left(\frac{L}{\sqrt{\text{MSE}}} \right)$ in decibel (dB), where L refers here to the dynamic range of the reference high-resolution image.

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

\mathcal{E} within the n^{th} neighborhood

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

Comparing to a reference image, many different realizations of precipitation fields may exhibit similar marginal statistics (e.g., MSE); however, their geometrical and coherent structures may be quite different. To address the SpaD performance for recovering the distinct structure and geometrical patterns of rain-cells in the studied precipitation images, a complementary metric, called the Structural SIMilarity (SSIM) by *Wang et al.* (2004), is also employed. On the contrary to the explained fidelity measures which only take into account the first and second order marginal statistics of the error, this metric considers the covariance structure as well, to overcome some common deficiencies of the MSE type of quality measures. Comparing images \mathcal{U} and \mathcal{V} , the SSIM is a local metric which is calculated for neighborhoods \mathbf{u} and \mathbf{v} (typically 8×8 pixels) of those images

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

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

Our numerical experiments show that the bias of the SPaD is within an acceptable range; i.e., the absolute value of the normalized bias ratio is less than 2% in the worst case and less than 0.4% in an average sense; see Table 5.1. By comparing the results of the SPaD with the observed low-resolution observation, the MSE metric shows significant improvement ranging from 40 to 50%. Computing the MMAE with a neighborhood of size 10×10 km, the recovery of the local maxima (intense rain cells) has been enhanced up to 15%. In terms of the PSNR the sparse downscaling shows an order of 2 to 3 dB of improvement and the SSIM index is also increased up to 10-15%.

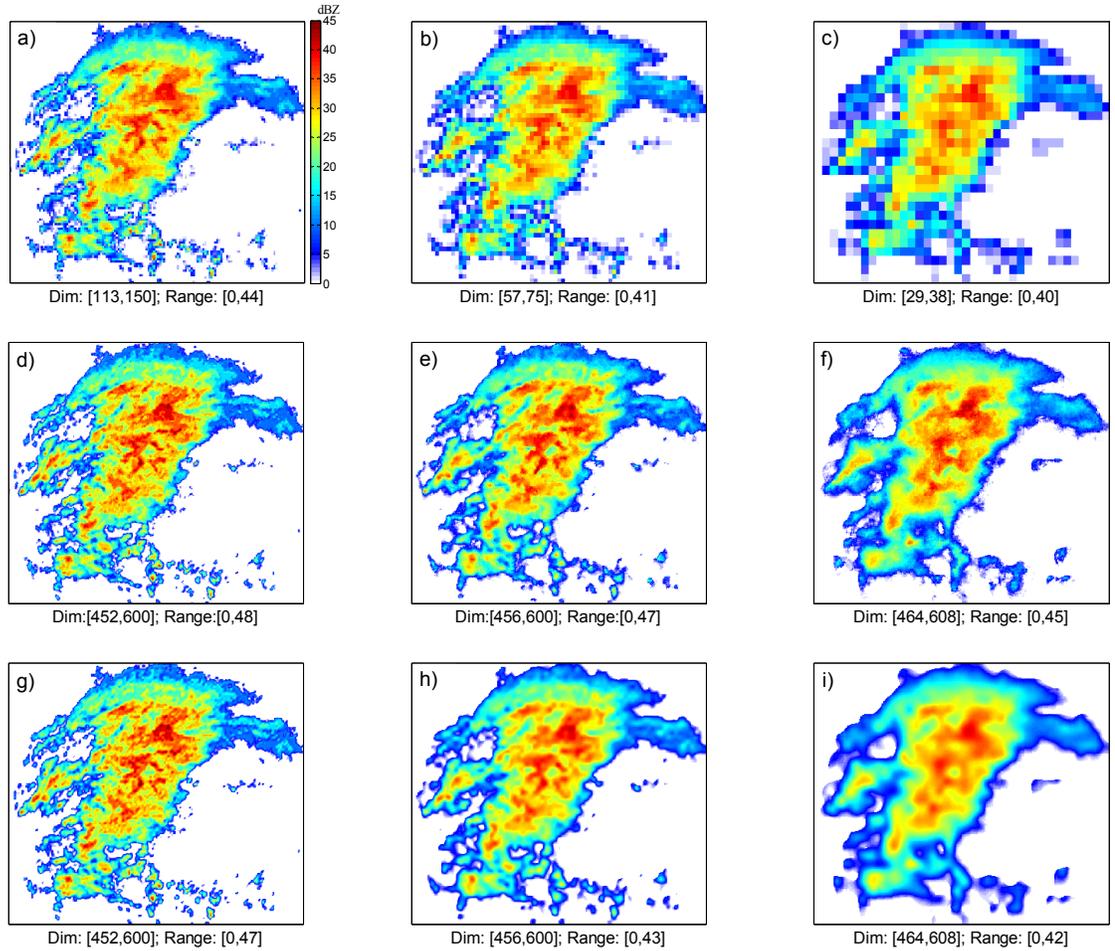


Figure 5.5: The qualitative performance of SPaD, given 200 regionally selected independent storm snapshots as the training database. (a) to (c) The low-resolution observations synthetically generated by smoothing and downsampling of the original 1×1 km fields to scales of 4,8 and 16 km in grid spacing. (d)-to-(f) Results of the downsampled rainfall images at resolution 1×1 km using the OMP method for selection of training rainfall patches as the “atoms” of the low and high-resolution dictionaries. (g)-to-(i) Results of the downsampled rainfall images at resolution 1×1 km employing the ℓ_1 -regularization strategy together with the learned dictionaries. As is evident, the SPaD results are smoother, when using the pair of the learned dictionaries of rainfall patches.

These preliminary results suggest that the accuracy of the downscaling may depend on the resolution of the original observation, complexity of the storm image and albeit the scaling ratio. However, on the basis of the chosen fidelity measures, the overall quality of the recovery is not significantly deteriorated for larger scaling ratios in our experiments. As expected, in general, smoother results were obtained using the learned dictionaries as the atoms are indeed weighted average of original rainfall patches. The SPaD results via dictionary learning led to better quality measures, typically for spatial scales less than < 4 km in grid spacing. For larger scales, the results of the SPaD with crude selection of rainfall patches as dictionary atoms and employing the OMP sparse coding, show better

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

Table 5.1: Error statistics obtained by comparing the low-resolution images with: (a) the high-resolution (true) ones, (b) Downscaling via SPaD with greedy selection of atoms by OMP algorithm, and (c) Downscaling via SPaD with ℓ_1 -regularization together with dictionary learning. For each metric (see text for definitions), the first and second rows show the obtained quantities for two reflectivity snapshots on 1998/11/13 (00:02:00 UTC) and 1998/06/28 (18:13:00 UTC), respectively. Note that, for each storm snapshot at a particular scale, the values of MSE, MAE and MMAE are row-wise normalized by their maximum values. The results are shown for downscaling from scales (4, 8, 16 km) down to 1 km in grid spacing.

[†]Note: NBR: Normalized Bias Ratio, MSE: Mean Squared Error, MAE: Maximum Absolute Error, MMAE: Mean Maximum Absolute Error, PSNR: Peak Signal to Noise Ratio, SSIM: Structural Similarity Index

error statistics. For instance, it is clear in Figure 5.5f and 5.5i that although the overall quality measures are not significantly different, due to the smoothness of the trained atoms, the dynamic range of the rainfall image cannot be well recovered in larger scaling ratios by the learned dictionaries.

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

Chapter 6

A Unified Variational Framework for Downscaling, Data Fusion and Assimilation

In previous chapters we have discussed some methodologies for rainfall downscaling and data fusion. Here, we introduce a unified variational framework that ties together the problems of downscaling, data fusion and data assimilation as ill-posed inverse problems. This framework seeks solutions beyond the classic least squares estimation paradigms by imposing a proper regularization, which is a constraint consistent with the degree of smoothness and probabilistic structure of the underlying state. We review relevant smoothing norm regularization methods in derivative space and extend classic formulations of the aforementioned problems with particular emphasis on land surface hydro-meteorological applications. The results of this chapter suggest that proper regularization can lead to more accurate results in hydro-meteorological downscaling, data fusion and data assimilation problems. In particular, using the Tikhonov and Huber regularization in the derivative space, the promise of the proposed framework is demonstrated in static downscaling and fusion of synthetic multi-sensor precipitation data, while a variational data assimilation experiment is presented using the linear advection-diffusion equation relevant to its fundamental applications in hydrologic study of land-surface heat and mass fluxes.

6.1 Introduction

The central goal of this chapter is to propose a unified framework for the class of downscaling (DS), data fusion (DF), and data assimilation (DA) problems by recasting them as discrete linear inverse problems using a relevant regularization in the derivative space, aiming to solve them more accurately compared to the classic weighted least-squares (WLS)

formulations. The proposed framework relies on the recent developments in mathematical formalisms of inverse problems which have received a great deal of attentions in the statistical regression and imaging sciences in the past decades (e.g., *Tibshirani*, 1996; *Chen et al.*, 2001), but are relatively new to the geophysical community. To the best of our knowledge, the only applications similar to the presented methodologies in this chapter, can be found in (*Ebtehaj et al.*, 2012) and (*Freitag et al.*, 2012) for rainfall downscaling and data assimilation of sharp fronts, respectively. To this end, some of the basic questions we are addressing in this chapter are: (1) How can we recast hydro-meteorological downscaling and data fusion problem as a variational inverse problem? (2) What are the proper regularization methods and how can we properly choose them? (3) How effective are those methods and what are the main advantages? and (4) Can we extend the concept of regularization to improve the results of the variational data assimilation in land surface hydro-meteorological problems?

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

$$\int_0^1 \mathcal{H}(r, t) x(t) dt = y(r), \quad 0 \leq r \leq 1, \quad (6.1)$$

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

might be more irregular and suffer from frequent and different sorts of discontinuities (e.g., rainfall). In fact, it can be shown that the proper choices of regularization not only yield unique and stable solutions but also reinforce the underlying regularity of the true state in the solution. It is important to note that, different regularity conditions are theoretically consistent with different statistical signatures in the true state, a fact that may guide proper design of the regularization, as explored in this study.

Presented methodologies for DS and DF problems are examined through synthetic downscaling and data fusion of rainfall remotely sensed observations, which has fundamental applications in flash flood predictions, especially in small watersheds (*Rebora et al.*, 2005; *Siccardi et al.*, 2005; *Rebora et al.*, 2006). We show that the presented methodologies allow us to improve the quality of rainfall estimation and reduce the uncertainty by recovering the small-scale high-intense rainfall extreme features, which have been lost due to the possible low-resolution constraints of the sensor. For the DA family of problems, the promise of the presented framework, is demonstrated via an elementary example using the heat equation, which has a key role in study of land-surface heat and mass fluxes (e.g., *Milly*, 1982; *Entekhabi et al.*, 1994; *Liang et al.*, 1999; *Caparrini et al.*, 2003; *Bateni and Entekhabi*, 2012). The results promise that the accuracy of the analysis and forecast cycles in a DA problem can be markedly improved, compared to the classic variational methods, especially when the initial state is discontinuous.

Section 6.2 provides conceptual insight into the discrete inverse problems. Section 6.3 describes the DS problem in detail, as a primitive building block for the other studied problems. Important classes of regularization methods are explained and their statistical interpretation is briefly discussed from the Bayesian point of view. Examples on rainfall downscaling are presented in this section by taking into account the specific regularity and statistical distribution of the rainfall fields in the derivative space. Section 6.4 is devoted to the regularized DF class of problems with examples and results on remotely sensed rainfall data. The regularized DA problem is discussed in Section 6.5.

6.2 Discrete Inverse problems: Conceptual Framework

In this section, we briefly explain the conceptual key elements of discrete linear inverse estimation relevant to the problems at hand and leave further details for the next sections. Analogous to equation (6.1), linear discrete inverse problems typically amount to estimating the true high-resolution m -element state vector $\mathbf{x} \in \mathbb{R}^m$ from the following observation model:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v}, \tag{6.2}$$

where $\mathbf{y} \in \mathbb{R}^n$ denotes the observations (e.g., output of a sensor), $\mathbf{H} \in \mathbb{R}^{n \times m}$ is an n -by- m observation operator that maps the state space onto the observation space, and $\mathbf{v} \sim \mathcal{N}(0, \mathbf{R})$

is the Gaussian error in \mathbb{R}^n . Note that, the observation operator, which is a discrete representation of the kernel in (6.1), and the noise covariance are supposed to be known or properly calibrated. Depending on the relative dimension of \mathbf{y} and \mathbf{x} , this linear system can be under-determined ($m \gg n$) or over-determined ($m \ll n$). In the under-determined case, there are infinite different \mathbf{x} 's that satisfy (6.2) while for the over-determined case a unique solution may not exist. As is evident, the DS problem belongs to the class of under-determined systems because the sensor output is a coarse-scale and noisy representation of the true state. However, the class of DF and DA problems fall into the category of over-determined systems, as the total size of the observations and background state exceeds the dimension of the true state.

In each of the above cases, we may naturally try to obtain a solution with minimum error variance by solving a linear (weighted) least squares (WLS) problem. However, for the under-determined case the solution still does not exist, while for the over-determined case it is commonly ill-conditioned and sensitive to the observation noise (see, Section 4). Therefore, the minimum variance WLS treatment can not properly make the above inverse problems well-posed. To obtain a unique and stable solution, the basic idea of regularization is to further constrain the solution. For instance, among many solutions that fit the observation model in (6.2), we can obtain the one with minimum energy, mean-squared curvature or total variation. The choice of this constrain or regularization, highly depends on a priori knowledge about the underlying regularity of \mathbf{x} . For sufficiently smooth \mathbf{x} we naturally may promote a solution with minimum mean-squared curvature to impose the desired smoothness on the solution. However, if the state is non-smooth and contains frequent jumps and discontinuities, a solution with minimum total variation might be a better choice. In subsequent sections, we explain these concepts in more details for the DS, DF, and DA problems with relevant examples.

6.3 Regularized Downscaling

6.3.1 Problem Formulation

To put the DS problem in a linear inverse estimation framework, we recognize that in the observation model of equation (6.2), the true high-resolution (HR) state $\mathbf{x} \in \mathbb{R}^m$ has a larger dimension than the low-resolution (LR) observation vector $\mathbf{y} \in \mathbb{R}^n$, that is $m \gg n$. Throughout this work, a notation is adopted in which the vector $\mathbf{x} \in \mathbb{R}^m$ may also represent, for example a 2D field $X \in \mathbb{R}^{\sqrt{m} \times \sqrt{m}}$ which is vectorized in a fixed order (e.g., lexicographical).

As explained in the previous section, the DS problem naturally amounts to obtaining an HR or fine-scale estimate $\hat{\mathbf{x}}$ of the true state, with minimum weighted least squares (WLS)

error, as follows:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\mathbf{R}^{-1}}^2 \right\}, \quad (6.3)$$

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

Using the theory of Lagrange multipliers the dual form of the constrained version of the optimization in (6.3) is

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\mathbf{R}^{-1}}^2 + \lambda \|\mathbf{L}\mathbf{x}\|_2^2 \right\}, \quad (6.4)$$

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

$$\hat{\mathbf{x}} = (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + 2\lambda \mathbf{L}^T \mathbf{L})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{y}, \quad (6.5)$$

provided that $\mathbf{L}^T \mathbf{L}$ is positive definite (*Tikhonov et al., 1977; Hansen, 1998; Golub et al., 1999; Hansen, 2010*). As is evident, the \mathbf{L} transformation also plays a key role in the solution of the regularized DS problem. For instance, choosing an identity matrix in (6.4) implies that we are looking for a solution with the smallest Euclidean norm (energy), while if \mathbf{L} represents a derivative operator, the above regularization term minimizes the energy in the derivative space, which naturally imposes extra smoothness on the solution.

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

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\mathbf{R}^{-1}}^2 + \lambda \|\mathbf{L}\mathbf{x}\|_1 \right\}, \quad (6.6)$$

where, the ℓ_1 -norm is $\|\mathbf{x}\|_1 = \sum_i |x_i|$. Choosing \mathbf{L} as a derivative operator in (6.6), in effect,

we minimize a measure of total variation of the state of interest. It is well understood that in this case, we typically better recover discontinuities and local jump singularities compared to the ℓ_2 -norm regularization in the derivative domain.

Note that, contrary to the Tikhonov regularization in (6.4), the ℓ_1 -norm regularization is a *non-smooth* convex optimization as the regularization term is non-differentiable and the conventional iterative gradient descent methods are no longer applicable in their standard forms.

One of the common approaches to treat the non-differentiability in (6.6) is to replace the ℓ_1 -norm with a smooth approximation, the so-called Huber norm, $\|\mathbf{x}\|_{\text{Hub}} = \sum_i \rho_T(x_i)$, where

$$\rho_T(x) = \begin{cases} x^2 & |x| \leq \tau \\ \tau(2|x| - \tau) & |x| > \tau \end{cases}, \quad (6.7)$$

and τ denotes a non-negative threshold (Figure 6.1). The Huber norm is a hybrid norm that behaves similar to the ℓ_1 -norm for values greater than the threshold τ while for smaller values it is identical to the ℓ_2 -norm. From the statistical regression point of view, the sensitivity of a norm as a penalty function to the outliers depends on the (relative) values of the norm for large residuals. If we restrict ourselves to convex norms, the least sensitive ones to the large residuals or say the outliers are those with linear behavior for large input arguments (i.e., ℓ_1 and Huber). Because of this property, these norms are often called *robust* norms, (*Huber*, 1964, 1981; *Boyd and Vandenberghe*, 2004). Throughout this chapter, for solving (6.6), we use the Huber relaxation due to its simplicity, efficiency and adaptivity to all of the concerning classes of DS, DF, and DA problems.

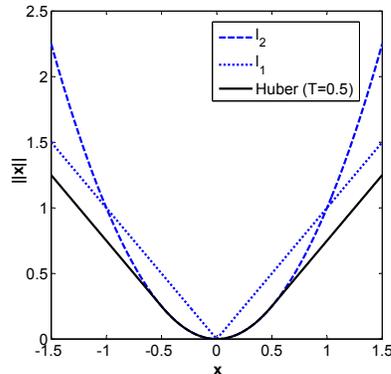


Figure 6.1: The Huber penalty is a smooth relaxation of the ℓ_1 -norm which acts quadratically for input values smaller than the threshold, while it behaves linearly for larger inputs. For heavy tailed inputs, linear penalization in the regularization term is advantageous compared to the quadratic penalization in which the overall cost function becomes dominated by a few large values in the tail of the distribution.

In general, the first term in (6.4 and 6.6) measures how well the solution approximates the

given (noisy) data, while the second term imposes a specific regularity on the solution. In effect, the regularizer plays a trade off role between making the fidelity to the observations sufficiently large, while not imposing too much regularity (the desired degree of smoothness) on the solution. The smaller the value of λ , the more weight is given to fitting the (noisy) observations which typically results in solutions that are less regular and prone to overfitting. On the other hand, the larger the value of λ , the more weight is given to the regularization term which may result a biased and overly regulated solution. Clearly, the goal is to find a balance between the two terms such that the solution is sufficiently close to the observations while obeying the underlying degree of regularity.

Important to note that, from the probabilistic point of view, the WLS problem (6.3) can be viewed as the maximum likelihood (ML) estimator of the HR field. On the other hand, the regularized problems (6.4) and (6.6) can be viewed as the Bayesian maximum a posteriori (MAP) estimator of the HR field. Indeed, the regularization terms refers to the *prior* knowledge about the probabilistic distribution of the state of interest . In other words, in (6.4) and (6.6), we implicitly assume that under the the chosen transformation \mathbf{L} , the state of interest can be well explained by the family of multivariate Gaussian $p(\mathbf{x}) \propto \exp\left(-\lambda \|\mathbf{L}\mathbf{x}\|_2^2\right)$ and Laplace $p(\mathbf{x}) \propto \exp(-\lambda \|\mathbf{L}\mathbf{x}\|_1)$ densities, respectively. Similarly, selecting the Huber norm can also be interpreted by assuming that $\log p(\mathbf{x}) \propto \sum_i \rho_T(x_i)$, which is equivalent to consider the *Gibbs* density function as the prior probability model (*Geman and Geman, 1984; Schultz and Stevenson, 1994*) (see Appendix B). The equivalence between the regularization, which imposes constraints on the regularity of the solution, and its Bayesian interpretation, which takes into account the prior probabilistic knowledge about the state of interest, is very insightful. This relationship establishes an important duality which can guide the selection of regularization method depending on the statistical properties of the state of interest in the real or derivative space.

6.3.2 Examples on Rainfall DS

6.3.2.1 Problem Formulation and Settings

As is evident, to downscale a remotely sensed hydro-meteorological state, using the explained discrete regularization methods, we need to have proper mathematical models for the downgrading operator and also a priori knowledge about the form of the regularization term. Clearly, in the presented framework, the downgrading operator needs to be a linear approximation of the sampling property of the sensor. If a sensor directly measures the state of interest while its maximum frequency channel is smaller than the maximum frequency content of the state (e.g., precipitation), the result of the sensing would be a smoothed and possibly down-sampled version of the true state. Thus, each element of the observed state in a grid-scale might be considered as an LR-representation of the true state, lacking the HR

possible choice is to assume that the sensor observes a coarse grained (i.e., non-overlapping box averaging) and noisy version of the true state. In other words, to produce a field at the grid-scale of $s_c \times s_c$ from a 1×1 , this assumption is equivalent to selecting a uniform smoothing kernel of size $s_c \times s_c$ followed by a down-sampling operation with ratio s_c (Figure 6.3a).

$$\text{a) } \frac{1}{s_c^2} \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}_{s_c \times s_c} \quad \text{b) } \nabla^2 = \frac{4}{(\kappa+1)} \begin{bmatrix} \frac{\kappa}{4} & \frac{1-\kappa}{4} & \frac{\kappa}{4} \\ \frac{1-\kappa}{4} & -1 & \frac{1-\kappa}{4} \\ \frac{\kappa}{4} & \frac{1-\kappa}{4} & \frac{\kappa}{4} \end{bmatrix}$$

Figure 6.3: (a) A uniform smoothing (low-pass) kernel of size $s_c \times s_c$. (b) The discrete (high-pass) generalized Laplacian filter of size 3×3 , where κ is a parameter ranging between 0 to 1. The Laplacian coefficients, obtained by filtering the 2D state with the Laplacian kernel, are approximate measures of the second order derivative. Throughout this chapter, we choose $\kappa = 0.5$ which corresponds to the standard second order differencing operation.

The error covariance matrix \mathbf{R} in observation model (6.2) plays a very important role on the results of the DS problem from both the mathematical and practical perspectives. Mathematically speaking, when the error is spatially white, the error covariance matrix is diagonal without any smoothing effect on the result (e.g., *Gaspari and Cohn*, 1999); however, the spatially correlated observation errors give rise to smoother results. Moreover, correlated errors with finite correlation length give rise to band error covariance matrices, which are prone to ill-conditioning. This ill-conditioning is typically more severe in case of ensemble error covariance estimation when the number of samples are typically much less than the observational dimension of the problem (e.g., *Ledoit and Wolf*, 2004). Practically speaking this error term speaks for the instrument error across the observation space, very relevant to rainfall remote sensing such as the ground-based NEXRAD sensing. Although, practical characterization of this error term is not in the scope of this study, for operation purposes, this term need to be properly estimated and calibrated based on factual evidence (e.g., *Hossain and Anagnostou*, 2005, 2006; *Maggioni et al.*, 2012; *AghaKouchak et al.*, 2012).

The choice of the regularization term also plays a very important role on the accuracy of the DS solution. Figure 6.4a demonstrates a NEXRAD reflectivity snapshot (1×1 km) over the Texas TRMM satellite ground validation site, while Figure 6.4b displays the standardized histogram of the discrete Laplacian coefficients (second order differences) and the fitted exponential of the form $p(x) \propto \exp(-\lambda|x|)$. It is seen that the analyzed rainfall image exhibits (nearly) sparse representation in the derivative space with a large mass around zero and heavier tail than the Gaussian. Throughout this chapter, we use the Laplacian for \mathbf{L} not only for its sparsifying effect on rainfall fields but also because of our experimental observations about its stabilizing role and computational adaptability for rainfall downscaling and data fusion. Although not shown here, the universality of this

structure can be observed in other rainfall reflectivity fields, denoting that the choice of the Laplace prior and ℓ_1 -norm regularization is preferred in the rainfall DS problems rather than the choice of the Tikhonov regularization; see, *Ebtehaj and Foufoula-Georgiou* (2011a) for a thorough survey of rainfall statistics in derivative space.

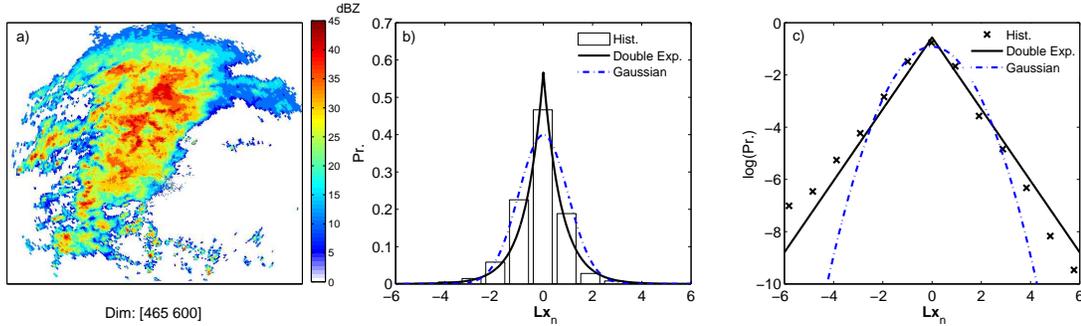


Figure 6.4: A rainfall reflectivity field and the distribution of its standardized Laplacian coefficients, $Lx_n = Lx/std(Lx)$, where $std(\cdot)$ is the standard deviation. (a) NEXRAD reflectivity snapshot at the TRMM GV-site in Houston, TX (HSTN) on 11/13/1998 (00:02:00 UTC) at scale 1×1 km. (b) The histogram of the standardized Laplacian coefficients, with $\kappa = 0.5$ (Figure 6.3b), and (c) their corresponding log-histogram. Note that, the zero coefficients over the non-rainy background have been excluded from the histogram analysis. The solid line in (b) is the least squares fitted exponential of the form $p(x) \propto \exp(-\lambda|x|)$ and the dash-dot line shows a standard normal distribution for comparison. The log-histogram in (c) contrasts the heavy tailed structure of the Laplacian coefficients versus the Gaussian distribution, clearer than the original histogram in (b).

This well behaved non-Gaussian structure in the derivative space mainly arises due to the presence of spatial coherent and correlated patterns in the rainfall fields which contain sharp transitions (large gradients) and isolated singularities (high-intensity rain-cells). In effect, over the large areas of uniform rainfall intensity, a measure of derivative translates rainfall intensities into a large number of (near) zero values; however, over the less frequent jumps and isolated high-intensity rain-cells, values of the derivative measure are markedly larger than zero and form the tails. Note that, this non-Gaussianity is due to the intrinsic spatial structure of rainfall reflectivity fields and can not be resolved by a logarithmic or power-law transformation (e.g., Z - R relationship). It is easy to see that after applying the Z - R relationship on the reflectivity fields, the shape of the rainfall histogram remains non-Gaussian and still can be explained by the Laplace density (not shown here). In practice, the histogram of the derivatives may exhibit a thicker tail than the Laplace density, requiring a heavier tail model such as the Generalized Gaussian Density (GGD) of the form $p(x) \propto \exp(-\lambda|x|^p)$, where $p < 1$ (see, *Ebtehaj and Foufoula-Georgiou*, 2011a). However, using such a prior model gives rise to a non-convex optimization problem in which convergence to the global minimum can not be easily guaranteed. Hence, the choice of the ℓ_1 -norm (the Laplace prior) for rainfall downscaling is indeed the closest convex relaxation that can partially fulfill the strict statistical interpretation of the rainfall field in

derivative domains.

Following our observations related to the distribution of the rainfall derivatives, using the Huber-norm regularization, we cast the rainfall DS as the following constrained variational problem:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\mathbf{R}^{-1}}^2 + \lambda \|\mathbf{L}\mathbf{x}\|_{\text{Hub}} \right\} \quad (6.8)$$

s.t. $\mathbf{x} \succeq 0$.

Obviously, the constraint is due to the non-negativity of the rainfall fields. In this study, we adopted the gradient projection (GP) method (*Bertsekas, 1999*, pp. 228), to solve the above variational problem (see Appendix B).

6.3.2.2 Results of the rainfall downscaling (DS)

The same rainfall snapshot shown in Figure 6.4 has been used to examine the performance of the proposed regularized DS methodologies. Throughout the chapter, to make the reported parameters independent of the intensity range, the rainfall reflectivity fields are first scaled into the range between 0 and 1; however, the downscaling results are presented in the true range.

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

Both the Huber and Tikhonov regularization methods were examined to downscale the observations from scales 4×4 and 8×8 km down to 1×1 km (Figure 6.5). A very small amount of white noise (i.e., standard deviation of 1e-3) was added to the LR observations (equation 6.2), giving rise to a diagonal error covariance matrix. In both of the regularization methods, for downscaling from 4-to-1 and 8-to-1 km in grid spacing, the regularization parameter λ was set to 5e-3 and 1e-2, respectively. These values are selected through trial and error; however, there are some formal methods for automatic estimation of this parameter, which are left for future work (e.g., *Hansen, 2010*, chap. 5). In our experiments, it turned out that small values of the Huber threshold τ , typically less than 10% of the field maximum range of variability, lead to a successful recovery of isolated singularities and local extreme rainfall cells (Figure 6.6 and 6.7).

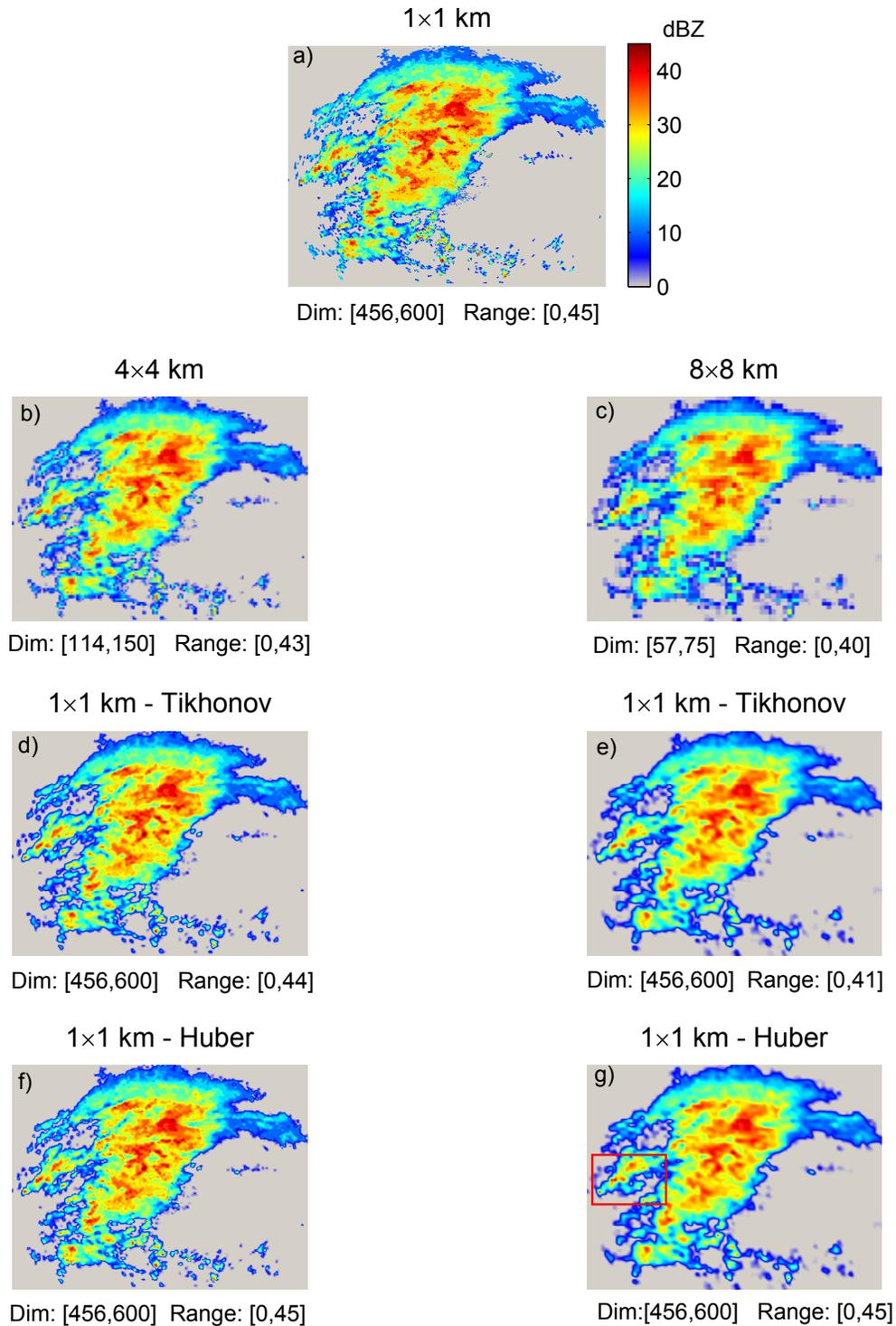


Figure 6.5: Sample results of the rainfall regularized Downscaling (DS). (a) True HR rainfall reflectivity: NEXRAD snapshot at the TRMM GV-site in Houston, TX (HSTN) on 11/13/1998 (00:02:00 UTC) at resolution 1×1 km. (b-c) The synthetically generated, 4×4 and 8×8 km, coarse-scale and noisy observations of the true rainfall reflectivity field. Left column: Tikhonov (d) and Huber (f) regularization results for downscaling from 4-to-1 km ($\tau = 0.02$). Right column: Tikhonov (e) and Huber (g) regularized DS for downscaling from 8-to-1 km ($\tau = 0.04$). Zooming views of the delineated box in (g) are shown in Figure 6.6.

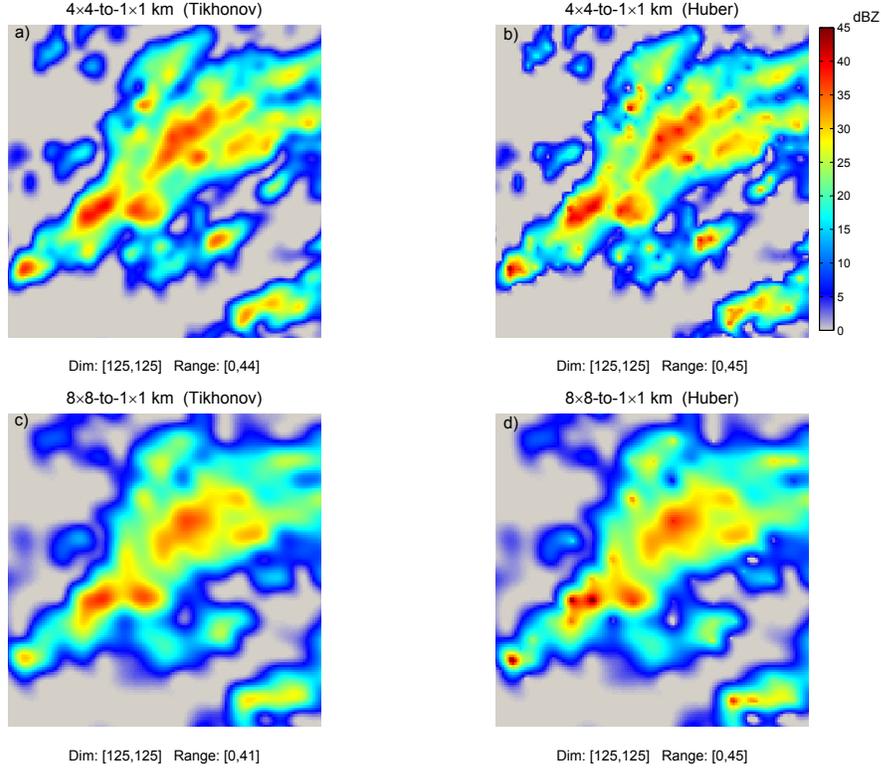


Figure 6.6: A zooming view for comparing qualitatively the Tikhonov (a, c) versus Huber (b, d) regularization, for the Downscaling (DS) example in Figure 6.5. The results indicate a marginally improved performance by the Huber regularization, especially for smaller scaling ratios. The Huber regularization yields sharper results and is more capable to recover high-intensity rainfall cells and the correct range of variability; see Table 6.1 for quantitative comparison using a suit of metrics and Figure (6.7).

In the studied snapshot, coarse graining of the rainfall reflectivity fields to the scales of 4×4 and 8×8 kilometers was equivalent to losing almost 20 and 30 percent of the rainfall energy in reflectivity domain in terms of the relative Root Mean Squared Error (RMSE), $RMSE = \|\mathbf{x} - \hat{\mathbf{x}}\|_2 / \|\mathbf{x}\|_2$ (see, Table 6.1). Note that, to compute the RMSE of the LR observations, the size of those fields was extended to the size of the true field using the nearest neighborhood interpolation, that is, each LR pixel was replaced with $s_c \times s_c$ pixels with the same intensity value. In addition to the relative RMSE measure, we also used three other metrics: (1) relative mean absolute error (MAE), $MAE = \|\mathbf{x} - \hat{\mathbf{x}}\|_1 / \|\mathbf{x}\|_1$; (2) A logarithmic measure often called the peak signal-to-noise ratio (PSNR), $PSNR = 20 \log_{10} (\max(\hat{\mathbf{x}}) / \text{std}(\mathbf{x} - \hat{\mathbf{x}}))$ where $\text{std}(\cdot)$ denotes the standard deviation and; (3) The structural similarity index (SSIM) by *Wang et al.* (2004). The PSNR in decibel (dB), represents a measure that not only contains RMSE information but also encodes the recovered range. The latter metric varies between -1 and 1 and the upper bound refers to the case where the estimated and reference (true) fields are perfectly matched. The SSIM metric is popular in the image processing community as it takes into account not

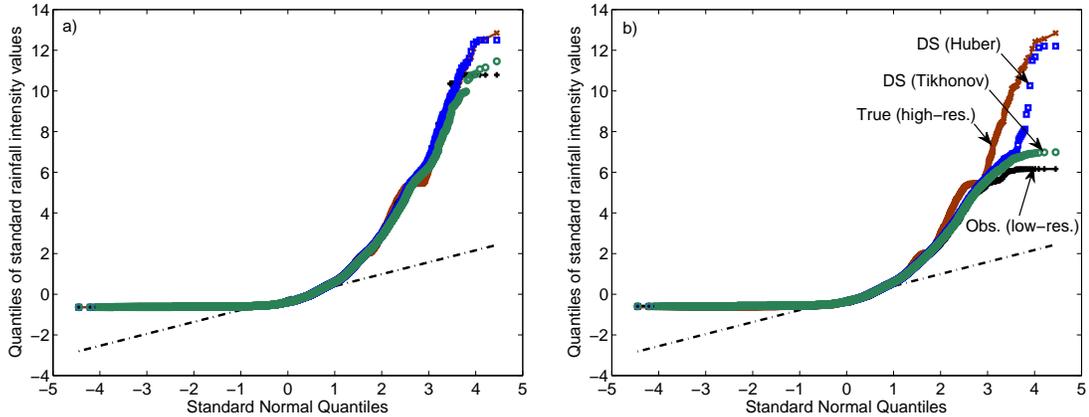


Figure 6.7: The quantiles of the standard Normal density versus the standardized distribution of the recovered rain rates [mm/hr], using $Z = 300R^{1.4}$ relationship, for the true HR field (red \times), the observed LR field (black $+$), the downsampled HR fields via the Tikhonov regularization (green \circ), and the Huber-norm regularization (blue \square), respectively. (a, b) The quantile-quantile plots for the HR fields obtained by downscaling from 4×4 to 1×1 km and 8×8 to 1×1 km, respectively. The rainfall quantile values are only for the positive rainy part of the fields and are standardized by subtracting the mean and dividing by the standard deviation. The qq-plots signify that the Huber regularization performs better than the Tikhonov, especially over the tails which represent the recovery of high-intensity and extreme rainfall values from LR observations.

only the marginal statistics such as the RMSE but also the correlation structure between the estimated and reference image. This metric seems very promising for analyzing the forecast mismatch with observations in hydro-meteorological studies, especially when the large-scale systematic errors (e.g., displacement error) might be more dominant than the random errors; see *Ebtehaj et al.* (2012) for applications of SSIM in rainfall downscaling.

On average, it is seen that almost 25% of the lost relative energy of the rainfall reflectivity fields can be restored via the regularized DS (Table 6.1). The ℓ_2 -norm regularization led to smoother results and as the scaling ratio grows, this regularization was almost incapable to recover the peaks and the correct variability range of the rainfall reflectivity field (Figure 6.6). Typically, as expected, the Huber-norm regularization results are slightly better than the Tikhonov ones, but not significantly. For large scaling ratios (i.e., $> 4 \times 4$ km) the results of those methods tended to coincide in terms of the selected lump quality metrics such the RMSE. However, using the Huber prior, the recovered range was markedly better than that by the Tikhonov regularization as reflected in the PSNR metric and recovered range. For example in downscaling from 8-to-1 km \times km via the Tikhonov regularization, the maximum recovered reflectivity values are around 41 [dBZ], while using the Huber norm regularization the maximum values are 45 [dBZ] (Figure 6.5). Employing the classic $Z - R$ relationship for the NEXRAD products (i.e. $Z = 300R^{1.4}$), one can easily check that the rain-rates associated with the above reflectivity values are approximately 15 and 28 [mm/hr], respectively. Therefore, although the lump quality metrics are comparable for

the two methods in the reflectivity domain, the main advantage of the Huber-norm over the ℓ_2 -norm is the recovery of local extreme rain rates (Figure 6.7). It is clear from quantile-quantile plots in Figure 6.7a,b that for smaller scaling ratio $s_c = 4$, the Huber regularization can very well follow the tail and body of the true rainfall distribution. However, the body of the recovered rainfall distribution falls below the true rainfall distribution for larger scaling ratio $s_c = 8$, indicating that in some high-intensity areas the method still underestimates the true field.

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

Table 6.1: Results showing the effectiveness of the proposed regularized DS by reducing the estimation error and increasing the accuracy of the rainfall fields. The first two columns refer to the values of the quality metrics obtained by comparing the constructed LR observations with true 1×1 km reflectivity field. The other columns show the obtained metrics by comparing the downscaled fields with the true rainfall field. The performance of the Huber prior is slightly better than the Tikhonov regularization, especially for the small scaling ratios (i.e., $\leq 4 \times 4$ km).[†] RMSE: relative root mean squared error; MAE: relative maximum absolute error; SSIM: structural similarity; and PSNR: peak signal to noise ratio (See section 3.2.2 for definitions).

6.4 Regularized Data Fusion

6.4.1 Problem Formulation

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

$$\mathbf{y}^i = \mathbf{H}^i \mathbf{x} + \mathbf{v}, \quad (6.9)$$

where $n_i \ll m$, $\mathbf{H}^i \in \mathbb{R}^{n_i \times m}$ and $\mathbf{v}^i \sim \mathcal{N}(0, \mathbf{R}^i)$ denotes uncorrelated Gaussian error in \mathbb{R}^{n_i} , $\mathbb{E}_{i \neq j} [\mathbf{v}^i (\mathbf{v}^j)^T] = 0$. Compared to the DS family of problems, a DF problem is more constrained in the sense that usually there are more equations than the number of unknowns, $\sum_i^N n_i \gg m$, giving rise to an over-determined linear system. As previously explained, naturally the linear weighted least squares (WLS) estimate of the true state,

given the series of N observations, amounts to solving the following optimization problem:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^N \left(\|\mathbf{y}^i - \mathbf{H}^i \mathbf{x}\|_{(\mathbf{R}^i)^{-1}}^2 \right) \right\}. \quad (6.10)$$

Notice that the solution of the above problem not only contains information about all of the available observations (Fusion) but also, with proper design of the observation operators, allows us to obtain an HR estimate of the state of interest (Downscaling). Clearly, the inverse of each covariance matrix in (6.10) plays the role of the relative contribution or weight of each observation \mathbf{y}^i in the cost function. In other words, if the elements of covariance matrix of a particular observation vector are large compared to those of the other observation vectors, naturally, the contribution of that observation to the obtained solution would be less significant.

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

$$\begin{aligned} \begin{bmatrix} \mathbf{y}^1 \\ \vdots \\ \mathbf{y}^N \end{bmatrix} &= \begin{bmatrix} \mathbf{H}^1 \\ \vdots \\ \mathbf{H}^N \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{v}^1 \\ \vdots \\ \mathbf{v}^N \end{bmatrix} \\ \Rightarrow \underline{\mathbf{y}} &= \underline{\mathbf{H}} \mathbf{x} + \underline{\mathbf{v}} \end{aligned} \quad (6.11)$$

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

$$\underline{\mathbf{R}} = \mathbb{E} [\underline{\mathbf{v}} \underline{\mathbf{v}}^T] = \begin{bmatrix} \mathbf{R}^1 & & 0 \\ & \ddots & \\ 0 & & \mathbf{R}^N \end{bmatrix}. \quad (6.12)$$

Therefore, the DF problem can be recast as the classic problem of estimating the true state from the augmented observation model of $\underline{\mathbf{y}} = \underline{\mathbf{H}} \mathbf{x} + \underline{\mathbf{v}}$. Hence, setting the gradient of the cost function in equation (6.10) to zero, yields the following linear system:

$$(\underline{\mathbf{H}}^T \underline{\mathbf{R}}^{-1} \underline{\mathbf{H}}) \hat{\mathbf{x}} = \underline{\mathbf{H}}^T \underline{\mathbf{R}}^{-1} \underline{\mathbf{y}}. \quad (6.13)$$

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

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\underline{\mathbf{y}} - \underline{\mathbf{H}}\mathbf{x}\|_{\underline{\mathbf{R}}^{-1}}^2 + \lambda \psi_{\mathbf{L}}(\mathbf{x}) \right\}, \quad (6.14)$$

s.t. $\mathbf{x} \succeq 0$

where the convex regularization function $\psi_{\mathbf{L}}(\mathbf{x})$ can take different penalty norms such as: the Tikhonov $\|\mathbf{L}\mathbf{x}\|_2^2$; the ℓ_1 -norm $\|\mathbf{L}\mathbf{x}\|_1$; or the Huber-norm $\|\mathbf{L}\mathbf{x}\|_{\text{Hub}}$. As is evident, similar to the DS problem, solution of (6.10) is equivalent to the frequentist ML estimator of the HR field while (6.14) is the Bayesian MAP estimator. For further explanations and statistical interpretations please see Appendix B.

6.4.2 Example on Rainfall DF

To quantitatively analyze the effectiveness of the proposed regularized DF methodology for rainfall data, we reconstructed two synthetic LR and noisy observations from the original HR NEXRAD reflectivity snapshot. To resemble different sensor constraints we chose different smoothing and down-sampling operations for each of the reconstructed field. The first observation field \mathbf{y}^1 was produced, at resolution 6×6 km, using a simple averaging filter of size 6×6 followed by a down-sampling ratio of $s_c = 6$. Analogously, the second field \mathbf{y}^2 was generated at scale 12×12 km using a Gaussian smoothing kernel of size the 12×12 with a standard deviation of 4. A white Gaussian noise, with standard deviation of 1e-2 and 2e-2 was also added, respectively, to resemble the measurement random error. Roughly speaking, this selection of the error magnitudes implies that the degree of confidence (relative weight) on the observation at 6×6 km is twice that of the one at 12×12 km. Here, we only restrict our consideration to the Huber norm regularization because of its consistency with the underlying rainfall statistics and its better performance in recovery of the rainfall heavy tailed structure (Figure 6.7). To solve the DF problem, we have used the same settings for the Gradient Projection (GP) method as explained in Appendix B.

The solution of the ill-conditioned WLS formulation or the ML estimator in (6.10) is blocky, out of range, and severely affected by the amplified inverted noise (Figure 6.8c). On the other hand, the regularized DF can properly restore a fine-scale and coherent estimate of the rainfall field. The results show that more than 30% of the uncaptured subgrid energy of the examined rainfall reflectivity field can be restored through solving the proposed methodology (Table 6.2).

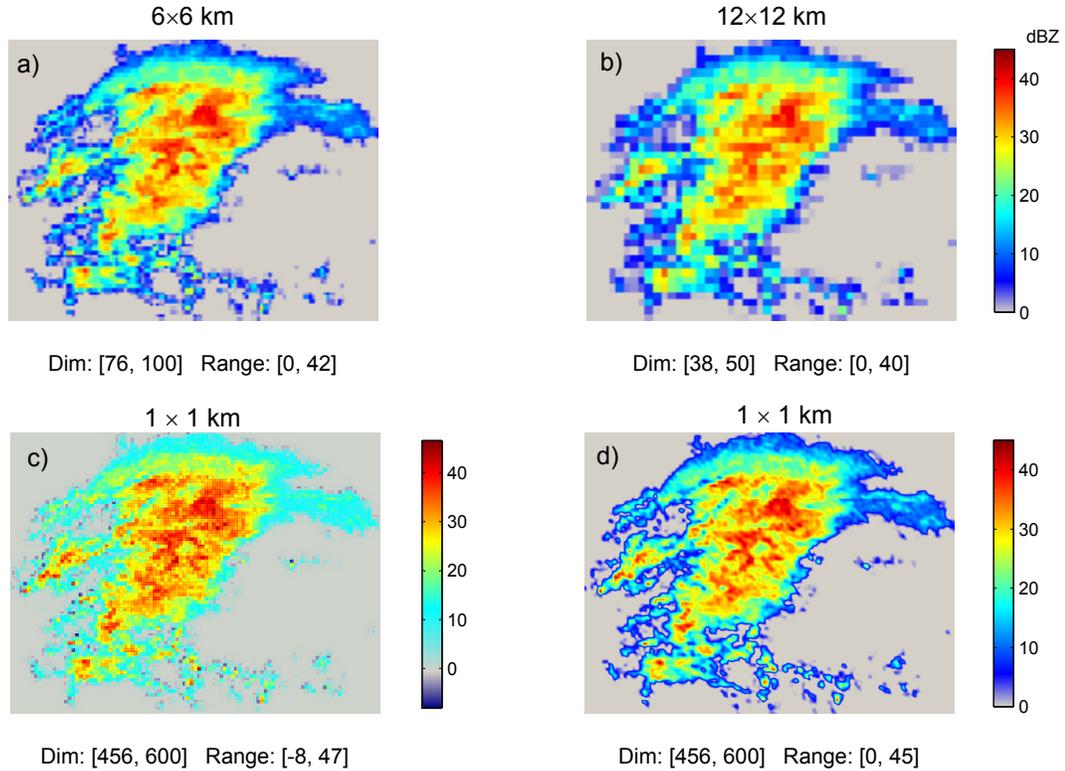


Figure 6.8: Data Fusion and Downscaling of multi-sensor remotely sensed rainfall reflectivity fields using the Huber regularization. (a-b) Reconstructed LR and noisy rainfall observations at scale 6 and 12 km in grid spacing. (c) The results of the WLS solution in (6.10), and (d) the solution of the regularized DF using the Huber norm with $\lambda = 1e-3$ and $\tau = 1e-2$.

As is evident, improvements of the selected fidelity measures in the DF problem are more pronounced than the results of the DS experiment. This naturally arises, because more observations are available in the DF problem than the DS one and thus the solution is better constrained. In terms of the selected lump metrics, analogous to the DS problem, we observed that the Huber-norm regularization is marginally better than the Tikhonov regularization, which is not reported here. However, as expected, in terms of recovery of the heavy tailed structure of the rainfall, it is verified that the Huber-norm regularization can capture the lost extreme values much better than the Tikhonov regularization (see Figure 6.9). It is clear from Figure (6.9) that the Huber-norm regularization very well captures the local extreme rainfall intensity values while the Tikhonov regularization falls short and can only partially recover those extreme intensity values.

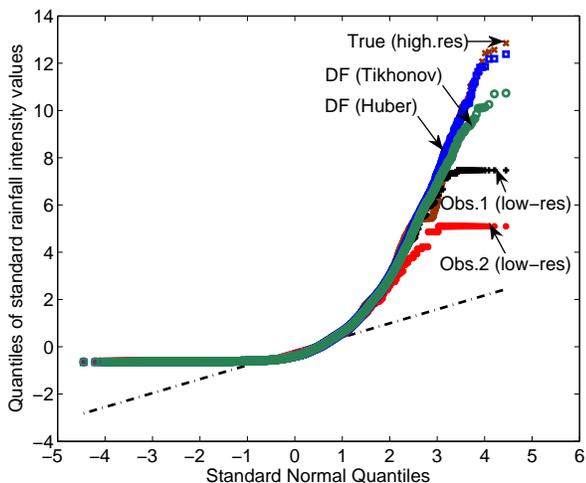


Figure 6.9: Quantiles of the standardized distribution of the recovered rain rates [mm/hr], using $Z = 300R^{1.4}$ relationship, versus standard normal quantiles. It is clear that the Huber-norm regularization results in better recovery of the rainfall extremes than the Tikhonov regularization. Evidently, because of extra information coming from multiple sensory data, the results of the extreme rain recovery in the DF experiment are improved compared to the DS results, see Figure (6.7).

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

Table 6.2: Values of the selected fidelity metrics in the rainfall DF experiment using the Huber regularization, see section 3.2.2 for the definitions. Here, the first two columns refer to comparison of the LR (6×6 and 12×12 km) observations with the true rainfall field and the last column presents the metrics obtained by comparing the DF results with the true field.

6.5 Regularized Variational Data Assimilation

Environmental prediction models are initial value problems and their forecast skills highly depend on the quality of their initialization. Data assimilation (DA) seeks the best estimate of the initial condition of a (numerical) model, given observations and physical constraints coming from the underlying dynamics (see, *Daley, 1993; Kalnay, 2003*). This important problem is typically addressed by two major classes of methodologies, namely sequential and variational methods (*Ide et al., 1997; Law and Stuart, 2012*). The sequential methods are typically built on the theory of mathematical filtering and recursive weighted least-squares (*Ghil et al., 1981; Ghil, 1989; Ghil and Malanotte-Rizzoli, 1991; Evensen, 1994a*;

Anderson, 2001; Moradkhani et al., 2005; Zhou et al., 2006; van Leeuwen, 2010, among others), while the variational methods are mainly rooted in the theories of constrained mathematical optimization and batch mode weighted least-squares (WLS) (e.g., Sasaki, 1970; Lorenc, 1986, 1988; Courtier and Talagrand, 1990; Zupanski, 1993, among others).

Although, recently the sequential methods have received a great deal of attention, the variational methods are still central to the operational weather forecasting systems. Classic formulation of the variational data assimilation (VDA) typically amounts to defining a (constrained) weighted least-squares penalty function whose optimal solution is the best estimate of the initial condition, the so-called *analysis* state. This penalty function typically encodes the weighted sum of the costs associated with the distance of the unknown true state to the available observations and previous model forecast, the so-called *background* state. Indeed, the penalty function enforces the solution to be close enough to both observations and background state in the weighted mean squared sense, while the weights are characterized by the observations and the background error covariance matrices. On the other hand, the constraints typically enforce the analysis to follow the underlying prognostic equations in a weak or strong sense (see, Sasaki, 1970; Daley, 1993, p.369). Typically, when we constrain the analysis only to the available observations and the background state at every instant of time, the variational data assimilation problem is called 3D-Var (e.g., Lorenc, 1986; Parrish and Derber, 1992; Lorenc et al., 2000; Kleist et al., 2009). On the other hand, when the analysis is also constrained to the underlying dynamics and available observations in a window of time, the problem is called 4D-Var (e.g., Zupanski, 1993; Rabier et al., 2000; Rawlins et al., 2007).

Inspired by the theories of smoothing spline and kriging interpolation in geostatistics, the first signs of using regularization in variational data assimilation trace back to the work by Wahba and Wendelberger (1980) and Lorenc (1986), where the motivation was to impose smoothness over the class of twice differentiable analysis states. More recently, Johnson et al. (2005b) argued that, in the classic VDA problem, the sum of the squared or ℓ_2 -norm of the weighted background error resembles the Tikhonov regularization (Tikhonov et al., 1977). Specifically, by the well-known connections between the Tikhonov regularization and spectral filtering via singular value decomposition (SVD) (e.g., see Hansen, 1998; Golub et al., 1999; Hansen et al., 2006), a new insight was provided into the interpretation and the stabilizing role of the background state on the solution of the classic VDA problem (see, Johnson et al., 2005a). Instead of using the ℓ_2 -norm of the background error, Freitag et al. (2010) and Budd et al. (2011) suggested to modify the classic VDA cost function using the sum of the absolute values or ℓ_1 -norm of the weighted background error. This assumption requires to statistically assume that the background error is heavy tailed and can be well approximated by the family of Laplace densities (e.g., Tibshirani, 1996; Lewicki and Sejnowski, 2000). For data assimilation of sharp atmospheric fronts, Freitag et al. (2012) kept the classic VDA cost function while further proposed to regularize the analysis

state by constraining the ℓ_1 -norm of its derivative coefficients.

In this section, motivated by our previous work in using ℓ_1 -norm regularization for down-scaling of remotely sensed rainfall observations (Ebtehaj *et al.*, 2012), we extend the previous studies (e.g., Freitag *et al.*, 2012) in regularized variational data assimilation (RVDA) by: (a) proposing a generalized regularization framework for assimilating low-resolution and noisy observations while the initial state of interest exhibits sparse representation in an appropriately chosen basis (i.e., wavelet, discrete cosine transform); (b) demonstrating the promise of the methodology in assimilation test problems using advection-diffusion dynamics with different error structure; and (c) proposing an efficient solution method for large-scale data assimilation problems. Here, we show that if sparsity in a pre-selected basis holds, this prior information can serve to improve the accuracy and stability of data assimilation problems. To this end, using prototype studies, different initial conditions are selected, which are sparse under the wavelet and spectral discrete cosine transformation (DCT). The promise of the ℓ_1 -norm RVDA is demonstrated via assimilating down-sampled and noisy observations in a 4D-Var setting by strongly constraining the solution to the governing advection-diffusion equation. In a broader context, we delineate the roadmap and explain how we may exploit sparsity, while the underlying dynamics and observation operator might be nonlinear. Particular attention is given to explain Monte Carlo driven approaches that can incorporate a sparse prior in the context of ensemble data assimilation.

6.5.1 Classic Variational Data Assimilation

At the time of model initialization t_0 , the goal of data assimilation can be stated as that of obtaining the *analysis* state as the best estimate of the true initial state, given noisy and low-resolution *observations* and the erroneous *background* state, while the analysis needs to be consistent with the underlying model dynamics. The background state in VDA is often considered to be the previous-time forecast provided by the prognostic model. By solving the VDA problem, the analysis is then being used as the initial condition of the underlying model to forecast the next time step and so on. In the following, we assume that the unknown true state of interest at the initial time t_0 is an m -element column vector in discrete space denoted by $\mathbf{x}_0 = [x_{0,1}, \dots, x_{0,m}]^T \in \mathbb{R}^m$, the noisy and low-resolution observations in the time interval $[t_0, \dots, t_k]$ are $\mathbf{y}_i \in \mathbb{R}^n$, $i = 1, \dots, k$, where $n \ll m$. Suppose that the observations are related to the true states by the following observation model

$$\mathbf{y}_i = \mathcal{H}(\mathbf{x}_i) + \mathbf{v}_i, \tag{6.15}$$

where $\mathcal{H} : \mathbb{R}^m \rightarrow \mathbb{R}^n$ denotes the nonlinear observation operator that maps the state space into the observation space, and $\mathbf{v}_i \sim \mathcal{N}(0, \mathbf{R}_i)$ is the Gaussian observation error with zero mean and covariance \mathbf{R}_i .

Taking into account the sequence of available observations, $\mathbf{y}_i \in \mathbb{R}^n$, $i = 0, \dots, k$, and denoting the background state and its error covariance by $\mathbf{x}_0^b \in \mathbb{R}^m$ and $\mathbf{B} \in \mathbb{R}^{m \times m}$, the 4D-Var problem amounts to obtaining the analysis at initial time as the minimizer of the following WLS cost function:

$$\mathcal{J}_{4D}(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_k) = \sum_{i=0}^k \left(\frac{1}{2} \|\mathbf{y}_i - \mathcal{H}(\mathbf{x}_i)\|_{\mathbf{R}_i^{-1}}^2 \right) + \frac{1}{2} \|\mathbf{x}_0^b - \mathbf{x}_0\|_{\mathbf{B}^{-1}}^2, \quad (6.16)$$

while the solution is constrained to the underlying model equation,

$$\mathbf{x}_i = \mathcal{M}_{0,i}(\mathbf{x}_0), \quad i = 0, \dots, k. \quad (6.17)$$

Here, $\|\mathbf{x}\|_{\mathbf{A}}^2 = \mathbf{x}^T \mathbf{A} \mathbf{x}$ denotes the *quadratic-norm*, while \mathbf{A} is a positive definite matrix and the function $\mathcal{M}_{0,i} : \mathbb{R}^m \rightarrow \mathbb{R}^m$ is a nonlinear model operator that evolves the initial state in time from t_0 to t_i .

Let us define $\mathbf{M}_{0,i}$ to be the Jacobian of $\mathcal{M}_{0,i}$ and restrict our consideration only to a linear observation operator, that is $\mathcal{H}(\mathbf{x}_i) = \mathbf{H}\mathbf{x}_i$, and thus the 4D-Var cost function reduces to

$$\mathcal{J}_{4D}(\mathbf{x}_0) = \sum_{i=0}^k \left(\frac{1}{2} \|\mathbf{y}_i - \mathbf{H}\mathbf{M}_{0,i}\mathbf{x}_0\|_{\mathbf{R}_i^{-1}}^2 \right) + \frac{1}{2} \|\mathbf{x}_0^b - \mathbf{x}_0\|_{\mathbf{B}^{-1}}^2. \quad (6.18)$$

By defining $\underline{\mathbf{y}} = [\mathbf{y}_0^T, \dots, \mathbf{y}_k^T]^T \in \mathbb{R}^N$, where $N = n(k+1)$, $\underline{\mathbf{H}} = [(\mathbf{H}\mathbf{M}_{0,0})^T, \dots, (\mathbf{H}\mathbf{M}_{0,k})^T]^T$, and

$$\underline{\mathbf{R}} = \begin{bmatrix} \mathbf{R}_0 & 0 & \cdots & 0 \\ 0 & \mathbf{R}_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{R}_k \end{bmatrix},$$

the 4D-Var problem (6.18) further reduces to minimization of the following cost function:

$$\mathcal{J}_{4D}(\mathbf{x}_0) = \frac{1}{2} \|\underline{\mathbf{y}} - \underline{\mathbf{H}}\mathbf{x}_0\|_{\underline{\mathbf{R}}^{-1}}^2 + \frac{1}{2} \|\mathbf{x}_0^b - \mathbf{x}_0\|_{\mathbf{B}^{-1}}^2. \quad (6.19)$$

Clearly, (6.19) is a smooth quadratic function of the initial state of interest \mathbf{x}_0 . Therefore, by setting the derivative to zero, it has the following analytic minimizer as the analysis state,

$$\mathbf{x}_0^a = (\underline{\mathbf{H}}^T \underline{\mathbf{R}}^{-1} \underline{\mathbf{H}} + \mathbf{B}^{-1})^{-1} (\underline{\mathbf{H}}^T \underline{\mathbf{R}}^{-1} \underline{\mathbf{y}} + \mathbf{B}^{-1} \mathbf{x}_0^b). \quad (6.20)$$

Throughout this study, we used Matlab built-in function `pcg.m`, described by *Bai et al.* (1987), for obtaining classic solutions of the 4D-Var in equation (6.20).

Accordingly, it is easy to see (e.g., *Daley*, 1993, p.39) that the analysis error covariance is the inverse of the Hessian of (6.19), as follows:

$$\mathbb{E} \left[(\mathbf{x}_0^a - \mathbf{x}_0) (\mathbf{x}_0^a - \mathbf{x}_0)^T \right] = (\underline{\mathbf{H}}^T \underline{\mathbf{R}}^{-1} \underline{\mathbf{H}} + \mathbf{B}^{-1})^{-1}. \quad (6.21)$$

It can be shown that the analysis in the above classic 4D-Var is the conditional expectation of the true state given observations and the background state. In other words, the analysis in the classic 4D-Var problem is the unbiased minimum mean squared error (MMSE) estimator of the true state (*Levy*, 2008, chap.4).

6.5.2 Regularized Variational Data Assimilation

6.5.2.1 Background

As is evident, when the Hessian (i.e., $\underline{\mathbf{H}}^T \underline{\mathbf{R}}^{-1} \underline{\mathbf{H}} + \mathbf{B}^{-1}$) in the classic VDA cost function in (6.19) is ill-conditioned, the VDA solution is likely to be unstable with large estimation uncertainty. To study the stabilizing role of the background error, motivated by the well-known relationship between the Tikhonov regularization and spectral filtering (e.g., *Golub et al.*, 1999), *Johnson et al.* (2005b,a) proposed to reformulate the classic VDA problem analogous to the standard form of the Tikhonov regularization (*Tikhonov et al.*, 1977). Accordingly, using a change of variable $\mathbf{z}_0 = \mathbf{C}_B^{-1/2} (\mathbf{x}_0 - \mathbf{x}_0^b)$, letting $\mathbf{B} = \sigma_b^2 \mathbf{C}_B$ and $\underline{\mathbf{R}} = \sigma_r^2 \underline{\mathbf{C}}_R$, where \mathbf{C}_B and $\underline{\mathbf{C}}_R$ are the correlation matrices, the classic variational cost function was proposed to be reformulated as follows:

$$\mathcal{J}_{4D}(\mathbf{z}_0) = \|\mathbf{f} - \mathbf{G}\mathbf{z}_0\|_2^2 + \mu \|\mathbf{z}_0\|_2^2. \quad (6.22)$$

where the ℓ_2 -norm is $\|\mathbf{x}\|_2 = (\sum_{i=1}^m x_i^2)^{1/2}$, $\mu = \sigma_r^2 / \sigma_b^2$, $\mathbf{G} = \underline{\mathbf{C}}_R^{-1/2} \underline{\mathbf{H}} \mathbf{C}_B^{1/2}$, and $\mathbf{f} = \underline{\mathbf{C}}_R^{-1/2} (\underline{\mathbf{y}} - \underline{\mathbf{H}} \mathbf{x}_0^b)$. Hence, by solving

$$\mathbf{z}_0^a = \underset{\mathbf{z}_0}{\operatorname{argmin}} \{ \mathcal{J}_{4D}(\mathbf{z}_0) \},$$

the analysis can be obtained as, $\mathbf{x}_0^a = \mathbf{x}_0^b + \mathbf{C}_B^{1/2} \mathbf{z}_0^a$. Having the above reformulated problem, (*Johnson et al.*, 2005a) provided new insights into the role of the background error covariance matrix on improving condition number and thus stability of the classic VDA problem.

To tackle data assimilation of sharp fronts, following the above reformulation, *Freitag et al.* (2012) suggested to add the smoothing ℓ_1 -norm regularization as follows:

$$\mathbf{z}_0^a = \underset{\mathbf{z}_0}{\operatorname{argmin}} \left\{ \mathcal{J}_{R4D}(\mathbf{z}_0) + \lambda \left\| \underline{\Phi} \left(\mathbf{C}_B^{1/2} \mathbf{z}_0 + \mathbf{x}_0^b \right) \right\|_1 \right\}, \quad (6.23)$$

where the ℓ_1 -norm is $\|\mathbf{x}\|_1 = \sum_{i=1}^m |x_i|$; the non-negative λ is called the regularization parameter; and Φ is proposed to be an approximate first-order derivative operator as follows:

$$\Phi = \begin{bmatrix} -1 & 1 & & 0 \\ & \ddots & \ddots & \\ 0 & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{(m-1) \times m}.$$

Notice that problem (6.23) is a non-smooth optimization as the derivative of the cost function does not exist at the origin. *Freitag et al.* (2012) recast this problem into a quadratic programming (QP) with both equality and inequality constraints where the dimension of the proposed QP is three times larger than that of the original problem. It is also worth noting that, the reformulations in (6.22) and (6.23) assume that the error covariance matrices are stationary (i.e., $\mathbf{B} = \sigma_b^2 \mathbf{C}_B$, $\mathbf{R} = \sigma_r^2 \mathbf{C}_R$) and the error variance is distributed uniformly across all of the problem dimensions. However, without loss of generality, a covariance matrix $\mathbf{B} \in \mathbb{R}^{m \times m}$ can be decomposed as $\mathbf{B} = \text{diag}(\mathbf{s}) \mathbf{C}_B \text{diag}(\mathbf{s})$, where $\mathbf{s} \in \mathbb{R}^m$ is the vector of standard deviations (*Barnard et al.*, 2000). Therefore, while one can have an advantage in stability of computation in (6.22) and (6.23), the stationarity assumptions and computations of the square roots of the error correlation matrices might be restrictive in practice.

In the subsequent sections, beyond ℓ_1 regularization of the first order derivative coefficients, we present a generalized framework to regularize the VDA problem in a properly chosen transform domain or basis (e.g., wavelet, Fourier, DCT). The presented formulation includes smoothing ℓ_1 and ℓ_2 -norm regularization as two especial cases and does not require any explicit assumption about the stationarity of the error covariance matrices. We recast the ℓ_1 -norm regularized variational data assimilation (RVDA) into a QP with lower dimension and simpler constraints compared to the presented formulation by *Freitag et al.* (2012). Furthermore, we introduce an efficient gradient-based optimization method, suitable for large scale data assimilation problems. Some results are presented via assimilating low-resolution and noisy observations into the linear advection-diffusion equation in a 4D-Var setting.

6.5.2.2 A Generalized Framework to Regularize Variational Data Assimilation in Transform Domains

In a more general setting, to regularize the solution of the classic VDA problem, one may constrain the magnitude of the analysis in the norm sense as follows:

$$\begin{aligned} \mathbf{x}_0^a &= \underset{\mathbf{x}_0}{\text{argmin}} \{ \mathcal{J}_{RAD}(\mathbf{x}_0) \} \\ \text{s.t. } & \|\Phi \mathbf{x}_0\|_p^p \leq \text{const.} \end{aligned} \quad (6.24)$$

where $\Phi \in \mathbb{R}^{m \times m}$ is any appropriately chosen linear transformation, and the ℓ_p -norm is $\|\mathbf{x}\|_p = (\sum |x_i|^p)^{1/p}$ with $p > 0$. By constraining the ℓ_p -norm of the analysis, we implicitly make the solution more stable. In other words, we bound the magnitude of the analysis state and reduce the instability of the solution due to the potential ill-conditioning of the classic cost function. Using the theory of Lagrange multipliers, the above constrained problem can be turned into the following unconstrained one:

$$\mathbf{x}_0^a = \underset{\mathbf{x}_0}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\underline{\mathbf{y}} - \underline{\mathbf{H}}\mathbf{x}_0\|_{\underline{\mathbf{R}}^{-1}}^2 + \frac{1}{2} \|\mathbf{x}_0^b - \mathbf{x}_0\|_{\mathbf{B}^{-1}}^2 + \lambda \|\Phi\mathbf{x}_0\|_p^p \right\}. \quad (6.25)$$

where the non-negative λ is the Lagrange multiplier or regularization parameter. As is evident, when λ tends to zero the regularized analysis tends to the classic analysis in (6.20), while larger values are expected to produce more stable solutions but with less fidelity to the observations and background state. Therefore, in problem (6.25), the regularization parameter λ plays an important trade-off role and ensures that the magnitude of the analysis is constrained in the norm sense while keeping it sufficiently close to observations and background state. Notice that although in special cases there are some heuristic approaches to find an optimal regularization parameter (e.g., *Hansen and O’Leary, 1993; Johnson et al., 2005b*), typically this parameter is selected empirically based on the problem at hand.

It is important to note that, from the probabilistic point of view, the regularized problem (6.25) can be viewed as the maximum a posteriori (MAP) Bayesian estimator. Indeed, the constraint of regularization refers to the *prior* knowledge about the probabilistic distribution of the state as $p(\mathbf{x}) \propto \exp\left(-\lambda \|\Phi\mathbf{x}\|_p^p\right)$. In other words, we implicitly assume that under the chosen transformation Φ the state of interest can be well explained by the family of multivariate Generalized Gaussian Density (e.g., *Nadarajah, 2005*) which includes the multivariate Gaussian ($p = 2$) and Laplace ($p = 1$) densities as special cases. As is evident, because the prior term is not Gaussian, the posterior density of the above estimator does not remain in the Gaussian domain and thus characterization of the a posteriori covariance is not straightforward in this case.

From an optimization view point, the above RVDA problem is convex with a unique global solution (analysis) when $p \geq 1$; otherwise, it may suffer from multiple local minima. For the special case of the Gaussian prior ($p = 2$) the problem is smooth and resembles the well-known smoothing norm Tikhonov regularization (*Tikhonov et al., 1977; Hansen, 2010*). However, for the case of the Laplace prior ($p = 1$) the problem is non-smooth, and it has received a great deal of attention in recent years for solving sparse ill-posed inverse problems (see, *Elad, 2010*, and references there in). It turns out that the ℓ_1 -norm regularization promotes sparsity in the solution. In other words, using this regularization, it is expected that the number of non-zero elements of $\Phi\mathbf{x}_0^a$ be significantly less than the observational dimension. Therefore, if we know a priori that a specific Φ projects a large number of elements of the state variable of interest onto (near) zero values, the ℓ_1 -norm is a proper

choice of the regularization term that can yield improved estimates of the analysis state (e.g., *Chen et al.*, 2001; *Candes and Tao*, 2006; *Elad*, 2010).

In the subsequent sections, we focus on the 4D-Var problem under the ℓ_1 -norm regularization as follows:

$$\mathbf{x}_0^a = \underset{\mathbf{x}_0}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\underline{\mathbf{y}} - \underline{\mathbf{H}}\mathbf{x}_0\|_{\underline{\mathbf{R}}^{-1}}^2 + \frac{1}{2} \|\mathbf{x}_0^b - \mathbf{x}_0\|_{\mathbf{B}^{-1}}^2 + \lambda \|\Phi\mathbf{x}_0\|_1 \right\}. \quad (6.26)$$

It is important to note that the presented formulation in (6.26) shares the same solution with the problem in (6.23) while in a more general setting, it can handle non-stationary error covariance matrices and does not require additional computational cost to obtain their square roots.

6.5.2.3 Solution Method via Quadratic Programming

Due to the separability of the ℓ_1 -norm, one of the most well-known methods, often called basis pursuit (see, *Chen et al.*, 1998; *Figueiredo et al.*, 2007), can be used to recast the ℓ_1 -norm RVDA problem in (6.26) to a constrained quadratic programming. Here, let us assume that $\mathbf{c}_0 = \Phi\mathbf{x}_0$, where \mathbf{x}_0 and \mathbf{c}_0 are in \mathbb{R}^m and split \mathbf{c}_0 into its positive $\mathbf{u}_0 = \max(\mathbf{c}_0, 0)$ and negative $\mathbf{v}_0 = \max(-\mathbf{c}_0, 0)$ components such that $\mathbf{c}_0 = \mathbf{u}_0 - \mathbf{v}_0$. Having this notation, we can express the ℓ_1 -norm via a linear inner product operation as $\|\mathbf{c}_0\|_1 = \mathbf{1}_{2m}^T \mathbf{w}_0$, where $\mathbf{1}_{2m} = [1, \dots, 1]^T \in \mathbb{R}^{2m}$ and $\mathbf{w}_0 = [\mathbf{u}_0^T, \mathbf{v}_0^T]^T$. Thus, problem (6.26) can be recast as a smooth constrained quadratic programming problem on non-negative orthant as follows:

$$\begin{aligned} \underset{\mathbf{w}_0}{\operatorname{minimize}} \quad & \left\{ \frac{1}{2} \mathbf{w}_0^T \begin{bmatrix} \mathbf{Q} & -\mathbf{Q} \\ -\mathbf{Q} & \mathbf{Q} \end{bmatrix} \mathbf{w}_0 + \left(\lambda \mathbf{1}_{2m} + \begin{bmatrix} \mathbf{b} \\ -\mathbf{b} \end{bmatrix} \right)^T \mathbf{w}_0 \right\} \\ \text{s.t.} \quad & \mathbf{w}_0 \succcurlyeq 0, \end{aligned} \quad (6.27)$$

where, $\mathbf{Q} = \Phi^{-T} (\underline{\mathbf{H}}^T \underline{\mathbf{R}}^{-1} \underline{\mathbf{H}} + \mathbf{B}^{-1}) \Phi^{-1}$, $\mathbf{b} = -\Phi^{-T} (\underline{\mathbf{H}}^T \underline{\mathbf{R}}^{-1} \underline{\mathbf{y}} + \mathbf{B}^{-1} \mathbf{x}_0^b)$, and $\mathbf{w}_0 \succcurlyeq 0$ denotes element-wise inequality.

Clearly, given the solution $\hat{\mathbf{w}}_0$ of (6.27), one can easily retrieve $\hat{\mathbf{c}}_0$ and thus the analysis state is $\mathbf{x}_0^a = \Phi \hat{\mathbf{c}}_0$.

The constraint of the QP problem (6.27) is simpler than the formulation suggested by (*Freitag et al.*, 2012) and allows us to use efficient and convergent gradient projection methods (e.g., *Bertsekas*, 1976; *Serafini et al.*, 2005; *Figueiredo et al.*, 2007), suitable for large-scale VDA problems. The dimension of the above problem seems twice that of the original problem; however, because of the existing symmetry in this formulation, the computational burden remains at the same order as the original classic problem (see, appendix B). Another important observation is that, choosing an orthogonal transformation (e.g.,

orthogonal wavelet, DCT, Fourier) for Φ is very advantageous computationally, as in this case $\Phi^{-1} = \Phi^T$.

Conceptually, adding relevant regularization terms, we enforce the analysis to follow a certain regularity and become more stable (*Hansen, 2010*). Here, by regularity, we refer to a certain degree of smoothness in the analysis state. For instance if we think of Φ as a first order derivative operator, using the smoothing ℓ_2 -norm regularization ($\lambda \|\Phi \mathbf{x}_0\|_2^2$), we enforce the energy of the solution's increments to be minimal, which naturally imposes more smoothness. Therefore, using the smoothing ℓ_2 -norm regularization in a derivative space, is naturally suitable for continuous and smooth physical states. On the other hand, for piece-wise smooth physical states with isolated singularities and jumps, it turns out that the use of the smoothing ℓ_1 -norm regularization ($\lambda \|\Phi \mathbf{x}_0\|_1$) in a derivative domain is very advantageous. Using this norm in derivative space, we implicitly constrain the total variation of the solution which prevents imposing extra smoothness on the solution. Proper selection of the smoothing norm and Φ may fall into the category of statistical model selection which is briefly explained in the following subsections.

As briefly explained previously, more stability of the solution comes from the fact that we constrain the magnitude of the solution by adding the regularization term and preventing the solution to blow up due to the ill-conditioning of the VDA problem (see, e.g., *Hansen, 1998; Johnson et al., 2005a*). In ill-conditioned classic VDA problems, it is easy to see that the inverse of the Hessian in (6.21) may contain very large elements which spoil the analysis. However, by regularization and making the problem well-posed, we shrink the size of the elements of the covariance matrix and reduce the estimation error. We need to emphasize that this improvement in the analysis error covariance, naturally comes at the cost of introducing a small bias in the regularized solution whose magnitude can be kept small by proper selection of the regularization parameter λ (see, e.g., *Neumaier, 1998*).

It is important to note that, for the smoothing ℓ_1 -norm regularization in (6.27), it is easy to show that the regularization parameter is bounded as $0 < \lambda < \|\mathbf{b}\|_\infty$, where the infinity-norm is $\|\mathbf{x}\|_\infty = \max(|x_1|, \dots, |x_m|)$. For those values of λ greater than the upper bound, clearly the analysis state in (6.27) is the zero vector with maximum sparsity (see, appendix B).

6.5.3 Examples on Linear Advection-Diffusion Equation

6.5.3.1 Problem Statement

The advection-diffusion equation is a parabolic partial differential equation with a drift and has fundamental applications in various areas of applied sciences and engineering. This equation is indeed a simplified version of the general Navier-Stocks equation for a

divergence free and incompressible Newtonian fluid where the pressure gradient is negligible. In a general form, this equation for a quantity of $\mathbf{x}(s, t)$ is

$$\begin{aligned} \frac{\partial \mathbf{x}(s, t)}{\partial t} + a(s, t) \nabla \mathbf{x}(s, t) &= \epsilon \nabla^2 \mathbf{x}(s, t), \\ \mathbf{x}(s, 0) &= \mathbf{x}_0(s), \end{aligned} \tag{6.28}$$

where $a(s, t)$ represents the velocity and $\epsilon \geq 0$ denotes the viscosity constant.

The linear ($a = \text{const.}$) and inviscid form ($\epsilon = 0$) of (6.28) has been the subject of modeling, numerical simulation, and data assimilation studies of advective atmospheric and oceanic flows and fluxes. For example, *Lin et al.* (1998) argued that the mechanism of rain-cell regeneration can be well explained by a pure advection mechanism, *Jochum and Murtugudde* (2006) found that Tropical Instability Waves (TIWs) need to be modeled by horizontal advection without involving any temperature mixing length. The nonlinear inviscid form (e.g., Burgers' equation) has been used in the shallow water equation and has been subject of oceanic and tidal data assimilation studies (e.g., *Bennett and McIntosh*, 1982; *Evensen*, 1994b). The linear and viscid form ($\epsilon > 0$) has fundamental applications in modeling of atmospheric and oceanic mixing (e.g., *Smith and Marshall*, 2009; *Lanser and Verwer*, 1999; *Jochum and Murtugudde*, 2006, chap. 6), land-surface moisture and heat transport (e.g., *Afshar and Marino*, 1978; *Hu and Islam*, 1995; *Peters-Lidard et al.*, 1997; *Liang et al.*, 1999), surface water quality modeling (e.g., *Chapra*, 2008, chap. 8), and subsurface mass and heat transfer studies (e.g., *Fetter*, 1994).

Here, we restrict our consideration only to the linear form and present a series of test problems to demonstrate the effectiveness of the ℓ_1 -norm RVDA in a 4D-Var setting. It is well understood that the general solution of the linear viscid form of (6.28) relies on the principle of superposition of linear advection and diffusion. In other words, the solution at time t is obtained via shifting the initial condition by at , followed by a convolution with the fundamental Gaussian kernel as follows:

$$\mathcal{D}(s, t) = (4\pi\epsilon t)^{-1/2} \exp\left(\frac{-|s|^2}{4\epsilon t}\right), \tag{6.29}$$

where the standard deviation is $\sqrt{2\epsilon t}$. As is evident, the linear shift of size at also amounts to obtaining the convolution of the initial condition with a Kronecker delta function as follows:

$$\mathcal{A}(s - at) = \begin{cases} 1 & s = at \\ 0 & \text{otherwise} \end{cases}. \tag{6.30}$$

6.5.3.2 Assimilation Set Up

Prognostic Equation and Observation Model It is well understood that (circular) convolution in discrete space can be constructed as a (circulant) Toeplitz matrix-vector product (e.g., *Chan and Jin, 2007*). Therefore, in the context of a discrete advection-diffusion model, the temporal diffusivity and spatial linear shift of the initial condition can be expressed in a matrix form by $\mathbf{D}_{0,i}$ and $\mathbf{A}_{0,i}$, respectively. In effect, $\mathbf{D}_{0,i}$ represents a Toeplitz matrix, for which its rows are filled with discrete samples of the Gaussian Kernel in (6.29), while the rows of $\mathbf{A}_{0,i}$ contain a properly positioned Kronecker delta function.

Thus, for our case, the underlying prognostic equation; i.e., $\mathbf{x}_i = \mathbf{M}_{0,i} \mathbf{x}_0$, may be expressed as follows:

$$\mathbf{x}_i = \mathbf{A}_{0,i} \mathbf{D}_{0,i} \mathbf{x}_0. \quad (6.31)$$

In this study, the low-resolution constraints of the sensing system are modeled using a linear smoothing filter followed by a down-sampling operation. Specifically, we consider the following time-invariant linear measurement operator

$$\mathbf{H} = \frac{1}{4} \begin{bmatrix} 1111 & 0000 & \cdots & 0000 \\ 0000 & 1111 & \cdots & 0000 \\ \vdots & \vdots & \vdots & \vdots \\ 0000 & 0000 & \cdots & 1111 \end{bmatrix} \in \mathbb{R}^{n \times m}, \quad (6.32)$$

which maps the higher-dimensional state to a lower-dimensional observation space. In effect, each observation point is then an average and noisy representation of the four adjacent points of the true state.

Initial States To demonstrate the effectiveness of the proposed ℓ_1 -norm regularization in (6.26), we consider four different initial conditions which exhibit sparse representation in the wavelet and DCT domains (Figure 6.10). In particular, we consider: (a) a flat top-hat, which is a composition of zero-order polynomials and can be sparsified theoretically using the first order Daubechies wavelet (DB01) or the Haar basis; (b) a quadratic top-hat which is a composition of zero and second order polynomials and theoretically can be well sparsified by wavelets with vanishing moments of order greater than three (*Mallat, 2009, pp.284*); (c) a window sinusoid; and (d) a squared exponential function which exhibits nearly sparse behavior in the DCT basis. In other words, in the high-frequencies due to the discontinuity in derivative decay sufficiently fast in the DCT domain. All of the initial states are assumed to be in \mathbb{R}^{1024} and are evolved in time with a viscosity coefficient $\epsilon = 4 \text{ [L}^2/\text{T]}$ and velocity $a = 1 \text{ [L/T]}$. The assimilation interval is assumed to be between 0 and $T = 500 \text{ [T]}$, where the observations are sparsely available over this interval at every 125 [T] time steps (Figure 6.10 and 6.11).

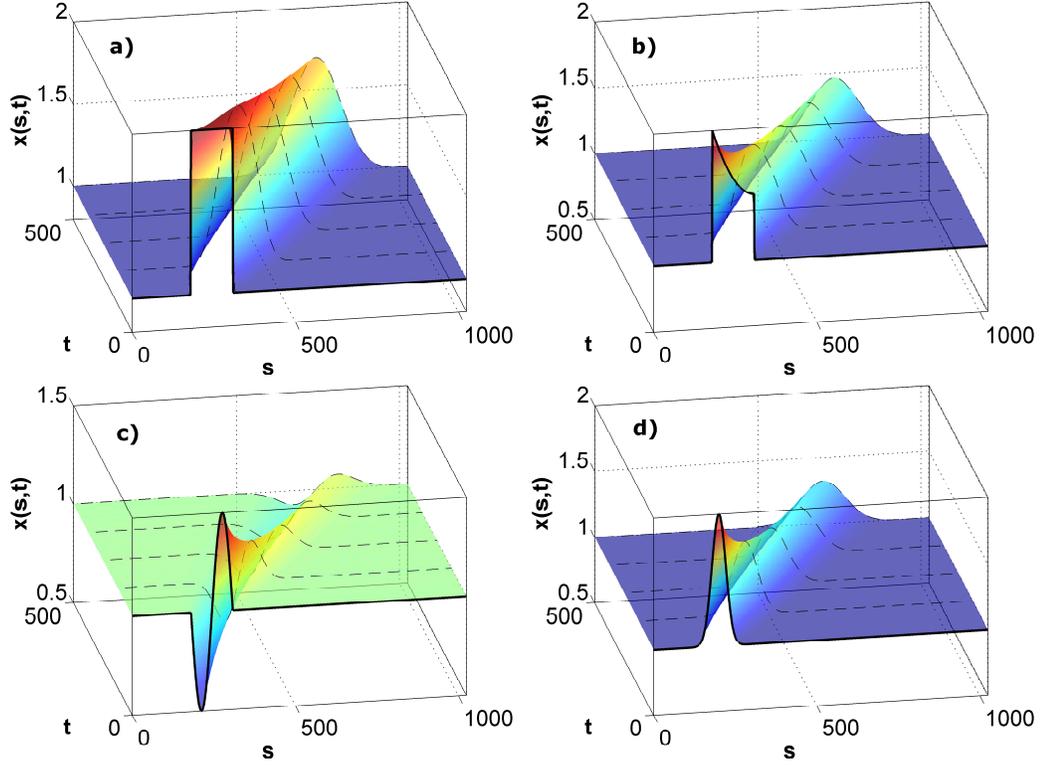


Figure 6.10: Initial conditions and their evolutions with the linear advection-diffusion equation: (a) flat top-hat (FTH), (b) quadratic top-hat (QTH), (c) window sinusoid (WS), and (d) squared-exponential (SE). The first two initial conditions (a, b) exhibit sparse representation in the wavelet domain while the next two (c, d) show nearly sparse representation in the discrete cosine domain (DCT). Initial conditions are evolved under the linear advection-diffusion equation (6.28) with $\epsilon = 4 [L^2/T]$ and $a = 1 [L/T]$. The broken lines show the time instants where the low-resolution and noisy observations are available in the assimilation interval.

Observation and Background Error The observations and background errors are important components of a data assimilation system that determine the quality and information content of the analysis. Clearly, the nature and behavior of the errors are problem-dependent and need to be carefully investigated in a case by case study. It needs to be stressed that from a probabilistic point of view, the presented formulation for the ℓ_1 -norm RVDA assumes that both of the error components are unimodal and can be well explained by the class of Gaussian covariance models. Here, for observation error, we only consider a stationary white Gaussian measurement error, $\mathbf{v} \sim \mathcal{N}(0, \mathbf{R})$, where $\mathbf{R} = \sigma_r^2 \mathbf{I}$ (Figure 6.11).

However, as discussed in (*Gaspari and Cohn, 1999*), the background error can often exhibit a correlation structure. In this study the first and second order auto-regressive (AR) Gaussian Markov processes, are considered for mathematical simulation of a possible spatial correlation in the background error; see *Gaspari and Cohn (1999)* for a detailed discussion about the error covariance models for data assimilation studies.

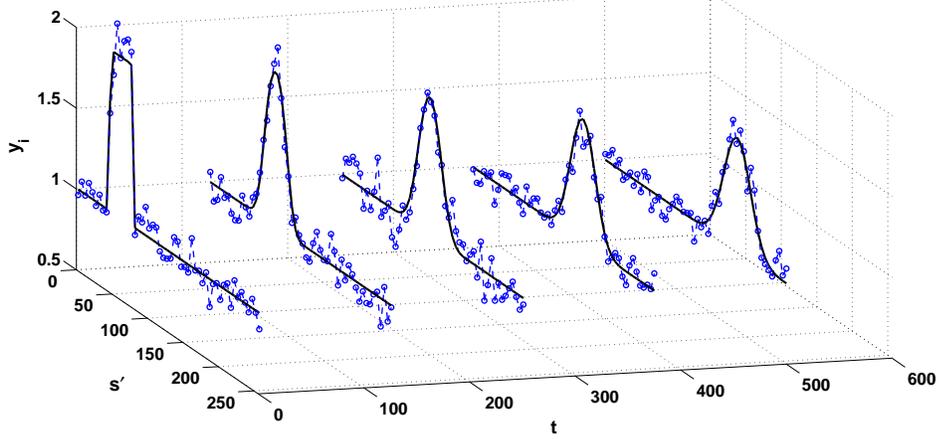


Figure 6.11: A sample representation of the available low-resolution (solid lines) and noisy observations (broken lines with circles) in every 125 [T] time steps in the assimilation window for the flat top-hat initial condition. Here, the observation error covariance is set to $\mathbf{R} = \sigma_r^2 \mathbf{I}$ with $\sigma_r = 0.08$ equivalent to $\text{SNR} = 20 \log(\sigma_{x_0}/\sigma_r) \approx 12$ dB.

The AR(1), also known as the Ornstein-Uhlenbeck process in infinite dimension, has an exponential covariance function $\rho(\tau) \propto e^{-\alpha|\tau|}$. In this covariance function, τ denotes the lag either in space or time, and the parameter α determines the decay rate of the correlation. The inverse of the correlation decay rate $l_c = 1/\alpha$ is often called the characteristic correlation length of the process. The covariance function of the AR(1) model has been studied very well in the context of stochastic process (e.g., *Durrett*, 1999) and estimation theory (e.g., *Levy*, 2008). For example, it is shown by *Levy* (2008, p. 298) that the eigenvalues are monotonically decreasing which may give rise to a very ill-conditioned covariance matrix in the discrete space, especially for small α or large correlation length. The covariance function of the AR(2) is more complicated than the AR(1); however, it has been shown that in special cases, its covariance function can be explained by $\rho(\tau) \propto e^{-\alpha|\tau|} (1 + \alpha|\tau|)$ (*Gaspari and Cohn*, 1999; *Stein*, 1999, p. 31). Note that, both of these covariance models are stationary and also isotropic as they are only a function of the magnitude of the correlation lag (*Rasmussen and Williams*, 2006, pp. 82). Consequently, the discrete background error covariance is a Hermitian Toeplitz matrix and can be decomposed into a scalar standard deviation and a correlation matrix as $\mathbf{B} = \sigma_b^2 \mathbf{C}_b$, where

$$\mathbf{C}_b = \begin{bmatrix} \rho(0) & \rho(1) & \cdots & \rho(m) \\ \rho(1) & \rho(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho(1) \\ \rho(m) & \cdots & \rho(1) & \rho(0) \end{bmatrix} \in \mathbb{R}^{m \times m}.$$

For the same values of α , it is clear that the AR(2) correlation function decays slower than that of the AR(1). Figure 6.12 shows empirical estimation of the condition number of the

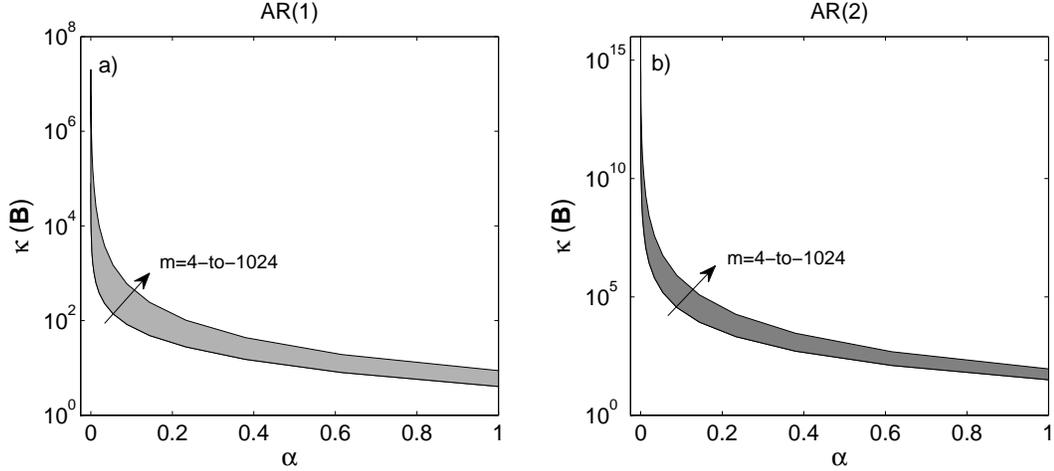


Figure 6.12: Empirical condition numbers of the background error covariance matrices as a function of parameter α and problem dimension (m) for the AR(1) in (a) and AR(2) in (b). The parameter α varies along the x-axis and m varies along the different curves of the condition numbers with values between 4 and 1024. We recall that $\kappa(\mathbf{B})$ is the ratio between the largest and smallest singular values of \mathbf{B} . In (a) the covariance matrix is $\mathbf{B}_{ij} = e^{-\alpha|i-j|}$ and in (b) $\mathbf{B}_{ij} = e^{-\alpha|i-j|} (1 + \alpha|i-j|)$, $1 \leq i, j \leq m$. It is seen that the condition numbers of the AR(2) model are significantly larger than those of the AR(1) model for the same values of the parameter α .

reconstructed correlation matrices at different dimensions ranging from $m = 4$ to 1024. As is evident, the error covariance of the AR(2) has a larger condition number than that of AR(1) for the same value of the parameter α . Clearly, as the background error plays a very important role on the overall condition number of the Hessian in the cost function in (6.19), an ill-conditioned background error covariance makes the solution more unstable with larger uncertainty around the obtained analysis.

Figure 6.13 shows a sample path of the chosen error models for the background error. Generally speaking, a correlated error contains large-scale (low-frequency) components that can corrupt the main spectral components of the true state at the same frequency range. Therefore, this type of error can superimpose with the large-scale characteristic features of the initial state and its removal is naturally more difficult than that of the white error via a data assimilation methodology.

6.5.3.3 Results of Assimilation Experiments

In this subsection, we present the results of the proposed regularized data assimilation as expressed in equation (6.26). We first present the results for the white background error and then discuss the correlated error scenarios. As previously explained, the first two initial conditions exhibit sharp transitions and are naturally sparse in the wavelet domain. For those initial states (Figure 6.10a, b) we have used classic orthogonal wavelet

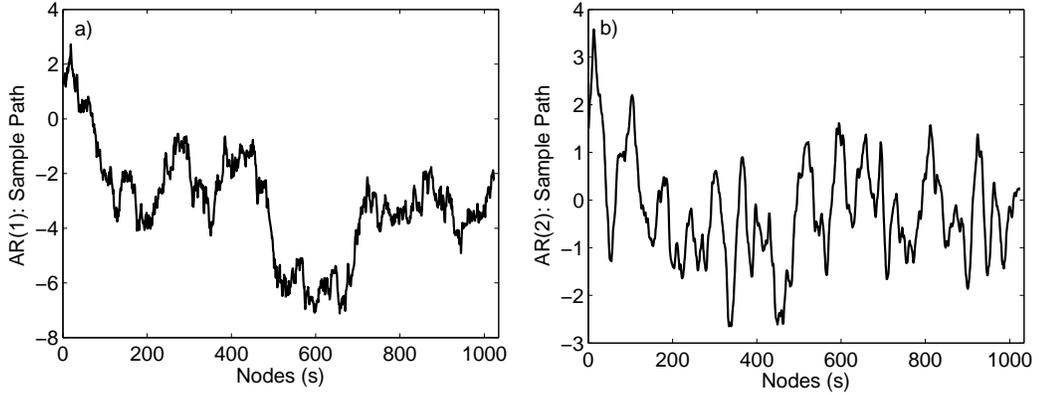


Figure 6.13: Sample paths of the used correlated background error: (a) the sample path for the AR(1) covariance matrix with $\alpha^{-1} = 150$, and (b) the sample path for the AR(2) covariance matrix with $\alpha^{-1} = 25$. The paths are generated by multiplying a standard white Gaussian noise $\mathbf{e} \sim \mathcal{N}(0, \mathbf{I})$ from the left by the lower triangular matrix \mathbf{L} , obtained by Cholesky factorization of the background error covariance matrix, that is $\mathbf{B} = \mathbf{L}\mathbf{L}^T$. It is seen that for small α , the sample paths exhibit large scale oscillatory behavior that can potentially corrupt low-frequency components of the underlying state.

transformation by *Mallat* (1989). Indeed, the columns of $\Phi \in \mathbb{R}^{1024 \times 1024}$ in this case contain the chosen wavelet basis that allow us to decompose the initial state of interest into its wavelet representation coefficients, as $\mathbf{c} = \Phi \mathbf{x}$ (forward wavelet transform). On the other hand, due to the orthogonality of the chosen wavelet $\Phi \Phi^T = \mathbf{I}$, rows of Φ^T contain the wavelet basis that allows us to reconstruct the initial state from its wavelet representation coefficients, that is $\mathbf{x} = \Phi^T \mathbf{c}$ (inverse wavelet transform). We used a full level of decomposition without any truncation of wavelet decomposition levels to produce a fully sparse representation of the initial state. For example, in our case where $\mathbf{x} \in \mathbb{R}^{1024}$, we have used ten levels of decomposition.

For the last two initial states (Figure 6.10c, d) we used DCT transformation (e.g., *Rao and Yip*, 1990) which expresses the state of interest by a linear combination of the oscillatory cosine functions at different frequencies. It is well understood that this basis has a very strong compaction capacity to capture the energy content of sufficiently smooth states and sparsely represent them via a few elementary cosine waveforms. Note that, this transformation is also orthogonal ($\Phi \Phi^T = \mathbf{I}$) and contrary to the Fourier transformation, the expansion coefficients are real.

White Background Error For the white background and observation error covariance matrices ($\mathbf{B} = \sigma_b^2 \mathbf{I}$, $\mathbf{R} = \sigma_r^2 \mathbf{I}$), we considered $\sigma_b = 0.10$ (SNR $\cong 10.5$ dB) and $\sigma_r = 0.08$ (SNR $\cong 12$ dB), respectively. Some results are shown in Figure 6.14 for the selected initial conditions. It is clear that the ℓ_1 -norm regularized solution markedly outperforms the classic 4D-Var solutions in terms of the selected metrics. Indeed, in the regularized analysis the

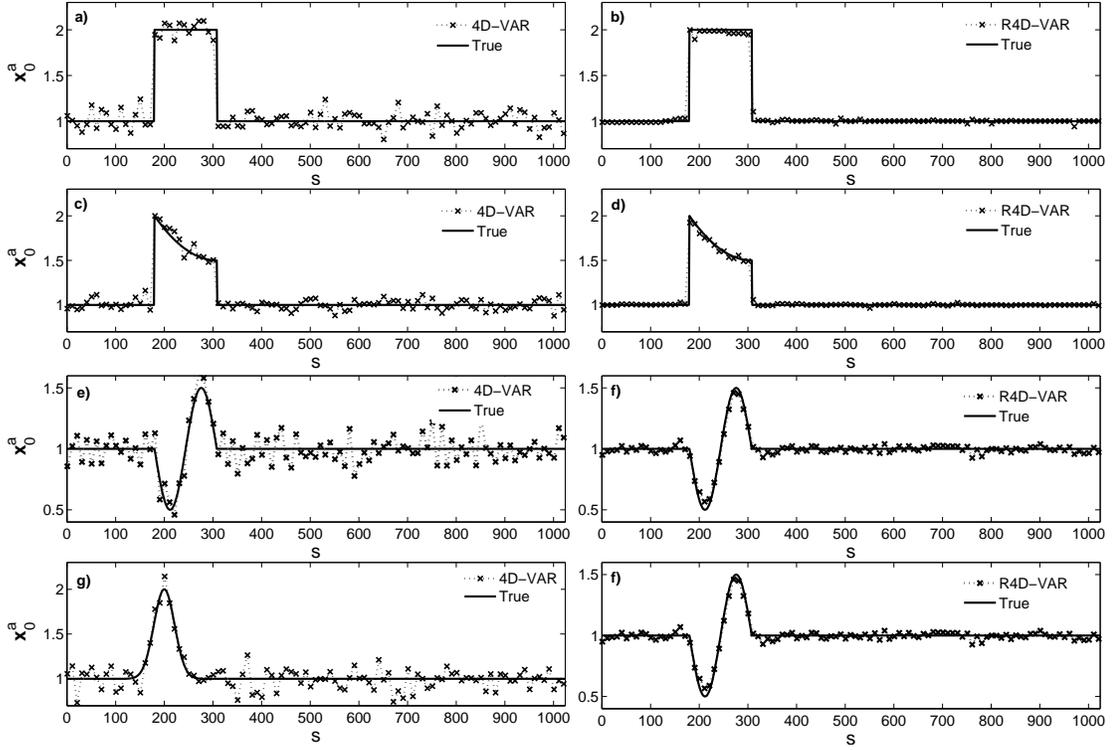


Figure 6.14: The results of the classic 4D-Var (left panel) versus the results of ℓ_1 -norm R4D-Var (right panel) for the tested initial conditions in a white Gaussian error environment. The solid lines are the true initial conditions and the crosses represent the recovered initial states or the analysis. In general, the results of the classic 4D-Var suffer from overfitting while the background and observation errors are suppressed and the sharp transitions and peaks are effectively recovered in the regularized analysis.

error is sufficiently suppressed and filtered, while characteristic features of the initial state are well-preserved. On the other hand, classic solutions typically over-fitted and followed the background state rather than extracting the true state. As a result, we can argue that for the white error covariance the classic 4D-Var has a very weak filtering effect which is an essential component of an ideal data assimilation scheme. This over-fitting may be due to the redundant (over-determined) formulation of the classic 4D-Var; see (*Hawkins, 2004*) for a general explanation on overfitting problems in statistical estimators and also see *Daley (1993, p.41)*.

The average of the results for 30 independent runs is reported in Table 6.3. Three different lump quality metrics are examined as follows:

$$\begin{aligned}
 \text{MSE}_r &= \|\mathbf{x}_0^t - \mathbf{x}_0^a\|_2 / \|\mathbf{x}_0^t\|_2 \\
 \text{MAE}_r &= \|\mathbf{x}_0^t - \mathbf{x}_0^a\|_1 / \|\mathbf{x}_0^t\|_1 \\
 \text{BIAS}_r &= \left| \overline{(\mathbf{x}_0^t - \mathbf{x}_0^a)} \right| / |\overline{\mathbf{x}_0^t}|
 \end{aligned} \tag{6.33}$$

Table 6.3: Expected values of the MSE_r , MAE_r , and BIAS_r , defined in (6.33), for 30 independent runs. The background and observation errors are white ($\mathbf{B} = \sigma_b^2 \mathbf{I}$, $\mathbf{R} = \sigma_r^2 \mathbf{I}$), where $\sigma_b = 0.10$ (SNR $\cong 10.5$ dB) and $\sigma_r = 0.08$ (SNR $\cong 12$ dB). The initial conditions are: flat top-hat (FTH), quadratic top-hat (QTH), window sinusoid (WS), and squared-exponential (SE). The results are reported for both the classic 4D-Var and the regularized 4D-Var (R4D-Var).

| White Background Error | | | | | | |
|------------------------|----------------|--------|----------------|--------|-----------------|-----------|
| | MSE_r | | MAE_r | | BIAS_r | |
| | R4D-Var | 4D-Var | R4D-Var | 4D-Var | R4D-Var | 4D-Var |
| FTH | 0.0188 | 0.0690 | 0.0099 | 0.0589 | 0.0016 | 0.0004 |
| QTH | 0.0152 | 0.0515 | 0.0083 | 0.0414 | 0.0030 | 0.0016 |
| WS | 0.0296 | 0.0959 | 0.0229 | 0.0771 | 0.0038 | 0.0022 |
| SE | 0.0316 | 0.0899 | 0.0235 | 0.0728 | 0.0018 | 4.26e - 5 |

namely, relative mean squared error (MSE_r), relative mean absolute error (MAE_r), and relative Bias (BIAS_r). In (6.33) \mathbf{x}_0^t denotes the true initial condition, \mathbf{x}_0^a is the analysis, and overbar denotes the expected value. It is seen that based on the selected lump quality metrics, the ℓ_1 -norm R4D-Var significantly outperforms the classic 4D-Var. In general, the MAE_r metric is improved more than the MSE_r metric in the presented experiments. The best improvement is obtained for the flat top-hat initial condition (FTH), where the sparsity is very strong compared to the other initial conditions. In other words, the ℓ_1 -norm R4D-Var is more effective for stronger sparsity of the initial state. The MSE_r metric is improved almost three orders of magnitude, while the MAE_r improvement reaches up to six orders of magnitude in the FTH initial condition. We need to note that although the trigonometric functions can be sparsely represented in the DCT domain, here we used a window sinusoid, which suffers from discontinuities over the edges and can not be perfectly sparsified in the DCT domain. However, we see that even in a weaker sparsity, the results of the ℓ_1 -norm R4D-Var are still much better than the classic solution.

Correlated background error In this part, the background error $\mathbf{B} = \sigma_b^2 \mathbf{C}_b$ is considered to be correlated. As previously discussed, typically longer correlation length creates ill-conditioning in the background error covariance matrix and makes the problem more unstable. On the other hand, the correlated background error covariance imposes smoothness on the analysis (see, *Gaspari and Cohn, 1999*), improves filtering effects, and makes the classic solution to be less prone to overfitting. In this subsection, we examine the effect of correlation length on the solution of data assimilation and compare the results of the sparsity promoting R4D-Var with the classic 4D-Var. Here, we do not apply any preconditioning as the goal is to emphasize on the stabilizing role of the ℓ_1 -norm regularization in the presented formulation. In addition, for brevity, the results are only reported for the top-hat and window sinusoid initial condition, which are solved in the wavelet and DCT domains, respectively.

a) Results for the AR(1) background error As is evident, in this case, the background state is defined by adding AR(1) correlated error to the true state (6.15a,d) which is known to us for these experimental studies. Figure 6.15 demonstrates that in the case of correlated error the classic 4D-Var is less prone to overfitting compared to the case of the uncorrelated error in Figure 6.14. Typically in the flat top-hat initial condition (FTH) with sharp transitions, the classic solution fails to capture those sharp jumps and becomes spoiled around those discontinuities (Figure 6.15b). For the trigonometric initial condition (WS), the classic solution is typically overly smooth and can not capture the peaks (Figure 6.15e). These deficiencies in classic solutions typically become more pronounced for larger correlation lengths and thus more ill-conditioned problems. On the other hand, the ℓ_1 -norm R4D-Var markedly outperforms the classic method by improving the recovery of the sharp transitions in FTH and peaks in WS (Figure 6.15).

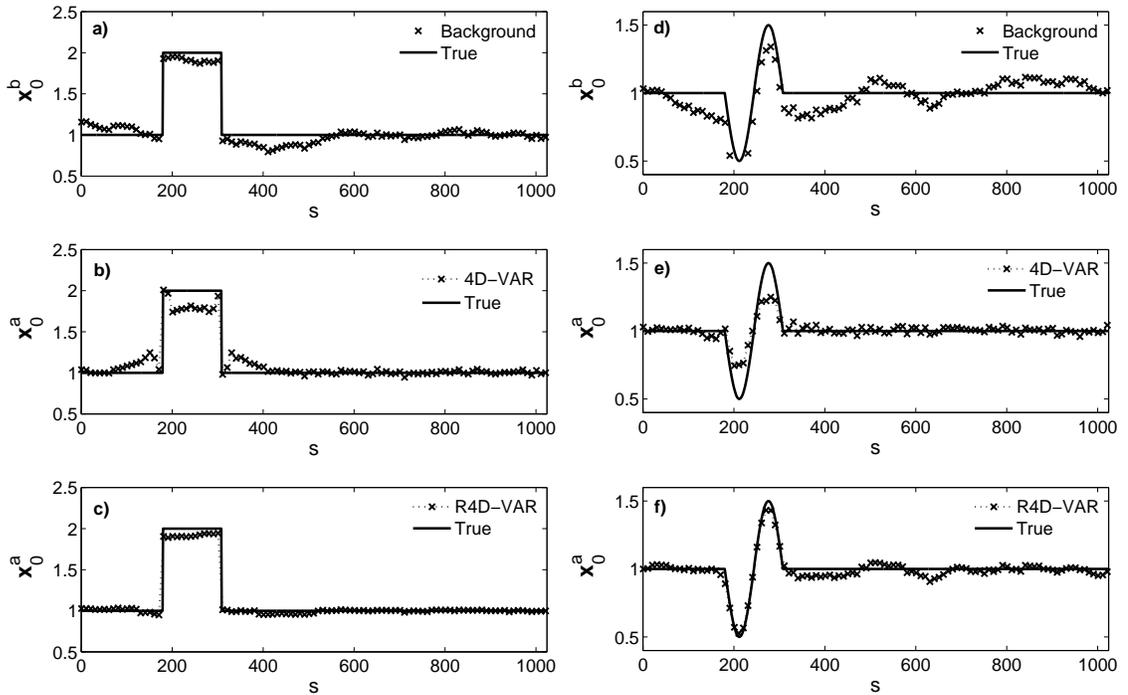


Figure 6.15: Comparison of the results of the classic 4D-Var (b, e) and ℓ_1 -norm R4D-Var (c, f) for the top-hat (left panel) and window sinusoid (right panel) initial conditions. The background states in (a) and (d) are defined by adding correlated errors using an AR(1) covariance model of $\rho(\tau) \propto e^{-\alpha|\tau|}$, where $\alpha = 1/250$. The results show that the ℓ_1 -norm R4D-Var improves recovery of sharp jumps and peaks and results in a more stable solution compared to the classic 4D-Var; see Figure 6.16 for quantitative results.

We examined a wide range of applicable correlation lengths, $\alpha^{-1} \in \{1, 10, 25, 50, 250, 1000\}$, which correspond to decades of variations ranging from 10^1 to 10^6 in the condition number $\kappa(\mathbf{B})$ of the background error covariance matrices (see Figure 6.12a). The assimilation results using different correlation lengths are demonstrated in Figure 6.16. To have a robust

conclusion about comparison of the proposed R4D-Var with the classic 4D-Var, the plots in this figure demonstrate the expected values of the quality metrics for 30 independent runs.

It can be seen that for small error correlation lengths ($\alpha^{-1} \lesssim 25$), the improvement of the R4D-Var is very significant while in the medium range ($25 \lesssim \alpha^{-1} \lesssim 50$) the classic solution becomes more competitive and closer to the regularized analysis. As previously mentioned, this improvement in the classic solutions is mainly due to the smoothing effect of the background covariance matrix. However, for larger correlation lengths ($\alpha^{-1} \gtrsim 50$), the differences of the two methods are more drastic as the classic solutions become more unstable and fail to capture the underlying structure of the initial state of interest. In general, we see that the MSE_r and MAE_r metrics are improved for all examined background error correlation lengths. As expected, the regularized solutions are slightly biased compared to classic solutions; however, the magnitude of the bias is not significant compared to the mean value of the initial state (see Figure 6.16). Figure 6.16 also shows a very important outcome of regularization which implies that the R4D-Var is almost insensitive to the studied range of correlation length and thus condition number of the problem. This confirms the stabilizing role of regularization and needs to be further studied for large scale and operational data assimilation problems. Another important observation is that, for extremely correlated background error, the classic R4D-Var may produce analysis with larger bias than the proposed R4D-Var (Figure 6.16c).

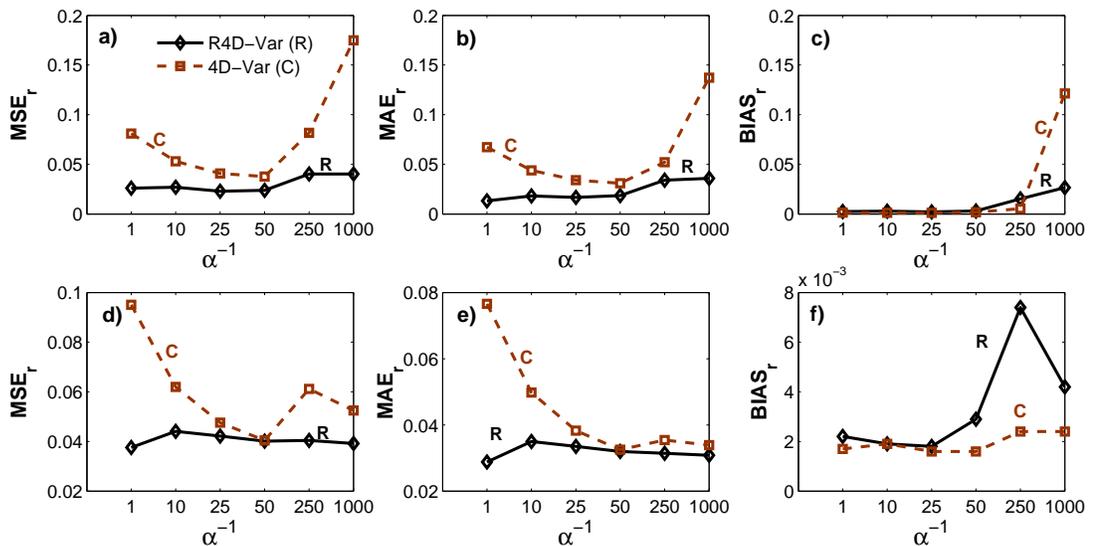


Figure 6.16: Comparison of the results of the proposed ℓ_1 -norm R4D-Var (solid lines) and the classic 4D-Var (broken lines) under the AR(1) background error for different correlation characteristic length scales (α^{-1}). Top panel: (a-c) the chosen quality metrics for the top-hat initial condition (FTH); Bottom panel: (d-f) the metrics for the window sinusoid initial condition (WS). These results, averaged over 30 independent runs, demonstrate significant improvements in recovering the analysis state by the proposed ℓ_1 -norm R4D-Var compared to the classic 4D-Var.

This unexpected result might be due to the presence of spurious bias in the background state coming from a strongly correlated error. In other words, a strongly correlated error may shift the mean value of the background state significantly and create a large bias in the solution of the classic 4D-Var. In this case, the improved performance of the R4D-Var may be due to its stronger stability and filtering properties.

b) Results for the AR(2) background error The AR(2) model is suitable for errors with higher order Markovian structure compared to the AR(1) model. As is seen in Figure (6.13), the condition number of the AR(2) covariance matrix is much larger than the AR(1) for the same values of the parameter α in the studied covariance models. Here, we limited our experiments to fewer characteristic correlation lengths of $\alpha^{-1} = \{1, 5, 25, 50\}$. We constrained our considerations to $\alpha^{-1} \lesssim 50$, because for larger values (slower correlation decay rates) the condition number of \mathbf{B} exceeds 10^8 and almost both methods failed to obtain the analysis without any preconditioning effort.

Table 6.4: Expected values of the MSE_r , MAE_r , and BIAS_r defined in (6.33), for 30 independent runs. The background and observation errors are modeled by the first order auto-regressive ($\mathbf{B} = \sigma_b^2 \mathbf{C}_B$) and white ($\mathbf{R} = \sigma_r^2 \mathbf{I}$) Gaussian processes, where $\sigma_b = 0.10$ (SNR $\cong 10.5$ dB) and $\sigma_r = 0.08$ (SNR $\cong 12$ dB). The parameter α denotes the correlation decay rate in the AR(2) covariance function $\rho(\tau) \propto e^{-\alpha|\tau|} (1 + \alpha|\tau|)$. The studied initial conditions are: flat top-hat (FTH), and window sinusoid (WS) and the results are reported for both the classic 4D-Var and the regularized 4D-Var (R4D-Var). The dash lines in the table denote that the classic method failed to return a solution without any pre-conditioning.

| AR(2) – Background Error | | | | | | | |
|--------------------------|---------------|----------------|--------|----------------|--------|-----------------|--------|
| | α^{-1} | MSE_r | | MAE_r | | BIAS_r | |
| | | R4D-Var | 4D-Var | R4D-Var | 4D-Var | R4D-Var | 4D-Var |
| FTH | 1 | 0.0254 | 0.0754 | 0.0162 | 0.0629 | 0.0023 | 0.0016 |
| | 5 | 0.0328 | 0.0643 | 0.0212 | 0.0534 | 0.0043 | 0.0018 |
| | 25 | 0.0722 | - | 0.0608 | - | 0.0187 | - |
| | 50 | 0.0742 | - | 0.0582 | - | 0.0268 | - |
| WS | 1 | 0.0363 | 0.0887 | 0.0272 | 0.0715 | 0.0029 | 0.0012 |
| | 5 | 0.0708 | 0.0906 | 0.0571 | 0.0529 | 0.0106 | 0.0017 |
| | 25 | 0.0877 | - | 0.0710 | - | 0.0243 | - |
| | 50 | 0.0898 | - | 0.0747 | - | 0.0361 | - |

In our case study, for $\alpha^{-1} \lesssim 25$, where $\kappa(\mathbf{B}) \lesssim 10^6$, the proposed R4D-Var outperforms the 4D-Var similar to what has been explained for the AR(1) error in the previous subsection. However, we found that for $25 \lesssim \alpha^{-1} \lesssim 50$, where $10^6 \lesssim \kappa(\mathbf{B}) \lesssim 10^8$, without proper preconditioning, the used conjugate gradient algorithm fails to obtain the analysis state in the 4D-Var (Table 6.4). On the other hand, due to the role of the proposed regularization, the R4D-Var remains sufficiently stable; however, its effectiveness deteriorated compared to the cases where the condition numbers were lower. This observation verifies the known

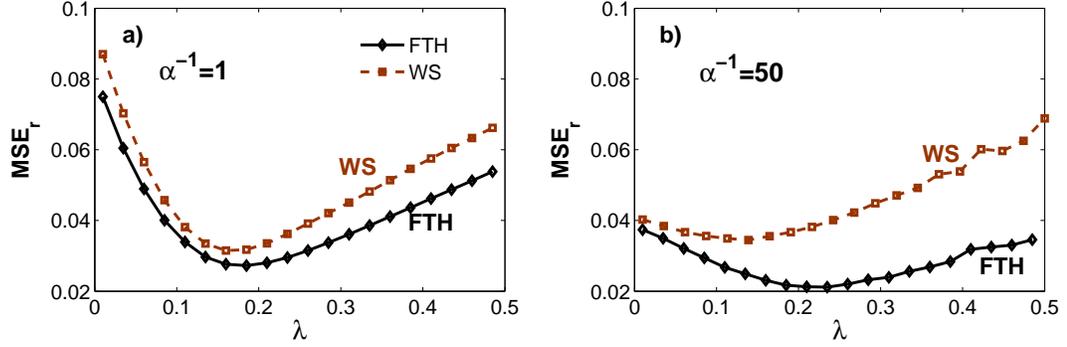


Figure 6.17: The relative mean squared error versus the regularization parameter obtained for the AR(1) background error for different characteristic correlation length (a) $\alpha^{-1} = 1$, and (b) $\alpha^{-1} = 50$. FTH and WS denote the flat top-hat and window sinusoid initial conditions, respectively.

role of the proposed regularization for improving the condition number of the variational data assimilation problem.

6.5.3.4 Selection of the regularization parameters

As previously explained, the regularization parameter λ plays a very important role in making the analysis sufficiently faithful to the observations and background state, while preserving the underlying regularity of the analysis. To the best of our knowledge, no general methodology exists which will produce an exact and closed form solution for the selection of this parameter, especially for the proposed ℓ_1 -norm regularization (see, Hansen, 2010, chap.5). Here, we chose the regularization parameter λ by trial and error based on a minimum mean squared error criterion (Figure 6.17). As a rule of thumb, we found that in general $\lambda \lesssim 0.05 \|\mathbf{b}\|_\infty$ yields reasonable results. We also realized that under similar error signal-to-noise ratio, the selection of λ depends on some important factors such as, the pre-selected basis, the degree of ill-conditioning of the problem, and more importantly the ratio between the dominant frequency components of the state and the error.

Chapter 7

Combined Radar Radiometer rainfall retrieval via Sparse Representation

7.1 Introduction

In the past decades, a large number of precipitation events have been coincidentally detected and estimated by multiple sensors with different accuracies and error structures. For instance, the TRMM satellite has provided a valuable library of multisensor precipitation data. Aboard of the TRMM satellite, there are two main sensory devices, namely TRMM microwave imager (TMI) and precipitation radar (PR), that are designed to detect and measure atmospheric precipitable water content and in particular near surface precipitation rates.

The TMI instrument is a nine-channel passive microwave radiometer, which measures atmospheric electromagnetic radiant fluxes at five central frequencies of 10.65, 19.35, 21.30, 37.0, 85.5 GHz. Radiance information is provided in horizontal and vertical polarizations for all channels except 21.3 GHz, for which radiant flux is only available for vertical polarization. After post-boost period (August 24, 2001), the TMI sensor spans a swath width of around 878 km at each orbital revolution of the TRMM satellite around the Earth. Spatial resolution of the TMI sensor approximately ranges from 5 to 65 km for the frequency channels of 85.5 to 10.65 GHz, respectively.

TRMM-PR is the first spaceborne single polarization radar which works at frequency 13.796 GHz and provides three-dimensional maps of the storm structure. The PR has a vertical resolution of 250 meters (at nadir) for the first four kilometer of troposphere and provides vertical profiles of the rain and snow from the surface up to the height of about 20 kilometers. This sensor spans a swath width of around 250 km with spatial resolution of 4 to 5 kilometer (Figure 7.1).

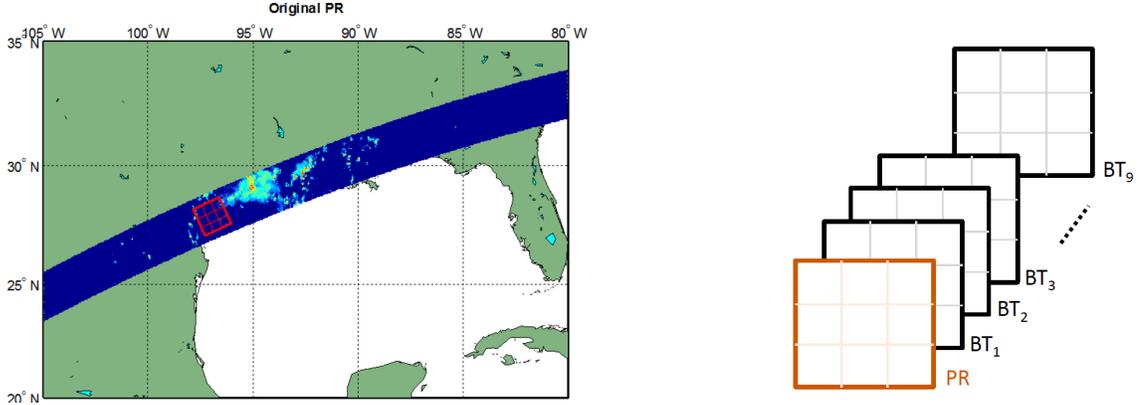


Figure 7.2: Schematic of a patch base rainfall retrieval algorithm. A patch is a local neighborhood of nearby pixels shown schematically in left panel. Right panel shows a three-dimensional patch of the PR rainfall rain rates and TMI brightness temperature (BT) values. As is evident, Euclidean dimension of the BT patches are nine times larger than that of the PR patches.

7.2 A Sketch of the Proposed algorithm

Similar to the Sparse Precipitation Downscaling (SPaD) framework by *Ebtehaj et al.* (2012), explained in Chapter 5, we suggest posing the rainfall retrieval problem as a non-linear inverse problem as follows:

$$\mathbf{Y}_T = \mathcal{H}(\mathbf{X}_R) + \mathbf{v}, \quad (7.1)$$

where the goal is to estimate the rainfall field (\mathbf{X}_R) from noisy and multispectral images of atmospheric brightness temperature (\mathbf{Y}_T). Here, \mathbf{v} denotes an additive Gaussian noise, $\mathcal{H}(\cdot) : \mathbf{X}_R \rightarrow \mathbf{Y}_T$ is the observation operator that maps multispectral measurements of brightness temperature (BT) onto the space of rainfall intensity values. Notice that in expression (7.1), we aim to recover a high-resolution rainfall field from multiple low-resolution BT fields. Therefore, provided that the effective dimension of \mathbf{Y}_T is greater than the the dimension of \mathbf{X}_R , we are dealing with an overdetermined inverse problem.

This inverse problem is naturally ill-posed due to the existing non-linearity of the measurement operator and observation noise. To implicitly characterize the nonlinear observation operator $\mathcal{H}(\cdot)$, the proposed retrieval method relies on statistical learning from two empirical dictionaries containing coincidental patches (e.g., a local neighborhood of 3×3 pixels, Figure 7.2) of TMI multispectral channels and the PR rain rate estimates. In particular, the TMI (\mathbf{D}_T) and PR (\mathbf{D}_R) dictionaries are very fat matrices (with more columns than rows), which contain a large set of multispectral BT patches and their corresponding rain rates as their columns, respectively. As is evident, these dictionaries share the same column dimension while the row dimension of the \mathbf{D}_R is nine times larger than that of the \mathbf{D}_T .

Now assume that, given knowledge of the coincidental dictionaries, a patch of multispectral BT values ($\mathbf{y}_T \in \mathbb{R}^n$) is given over the outer swath, where the PR has no coverage and we

are interested to estimate the corresponding PR rain rates ($\mathbf{x}_R \in \mathbb{R}^m$). For each BT patch, the idea is to develop a mechanism for finding a few similar columns of the BT-dictionary (columns of $\mathbf{D}_T \in \mathbb{R}^{n \times M}$) that their weighted linear combination can well approximate the BT patch of interest ($\mathbf{y}_T \in \mathbb{R}^n$). In a norm notation, this idea amounts to solving the following constrained optimization,

$$\underset{\mathbf{c}}{\text{minimize}} \|\mathbf{y}_T - \mathbf{D}_T \mathbf{c}\|_2^2 \quad \text{subject to} \quad \|\mathbf{c}\|_1 \leq \text{const.} \quad (7.2)$$

The elements of optimization vector $\mathbf{c} \in \mathbb{R}^M$ (representation coefficients) indeed contain the weights that linearly combine few columns of the BT-dictionary \mathbf{D}_T . Notice that the cost function in equation (7.2) penalizes the mean squared error while the constraint promotes sparsity in the representation coefficients. In other words, $\hat{\mathbf{c}}$ as the minimizer of problem (7.2) can well approximate \mathbf{y}_T while it has very few non-zero elements compared to the dimension of \mathbf{y}_T . Obtaining $\hat{\mathbf{c}}$, we may use the same coefficients to combine the corresponding columns of the coincidental PR-dictionary ($\mathbf{D}_R \in \mathbb{R}^{m \times M}$) to reproduce an estimate of the related rain rates, that is $\hat{\mathbf{x}}_R = \mathbf{D}_R \hat{\mathbf{c}}$. We can repeat this process for all of patches (overlapping or non-overlapping) of the BT-field fields and estimate the corresponding values of rainfall intensity. Hereafter, this algorithm is called the Sparse Precipitation Retrieval (SPaR).

7.3 Data

For preliminary examination of the proposed algorithm, so far we have collected 1B11 and 2A25 TRMM orbital products over the entire year of 2002. The 1B11 product contains multispectral images of BT values (*Kummerow et al.*, 1996; *Hong et al.*, 1999) registered onto two different grid points for high-frequency (85 GHz) and low-frequency channels (< 85 GHz). The 2A25 products provides a large set of information about vertical structure of atmospheric moisture and precipitable water content including the final surface rain rates obtained from the latest available radar profiling algorithm (e.g., *Iguchi et al.*, 2000). Information about land, ocean, and coastal surface positions is obtained from orbital product 1C21, which uses Digital Terrain Elevation Dataset (DTED) provided by NASA-JPL. For verification of the proposed algorithm, we compare the preliminary results with the TRMM 2A12 product, which uses the most updated Goddard profiling algorithm by *Kummerow et al.* (2010).

As previously explained, the TMI and PR final orbital products are provided at different spatial resolutions and over two independent and irregularly spaced grid points in geographical coordinate system (i.e., latitude, longitude). In this study, to obtain the required dictionaries, we mapped both of the products onto the PR grid points, which are more regular than those of the TMI grids. Figure 7.3 shows brightness temperature images of

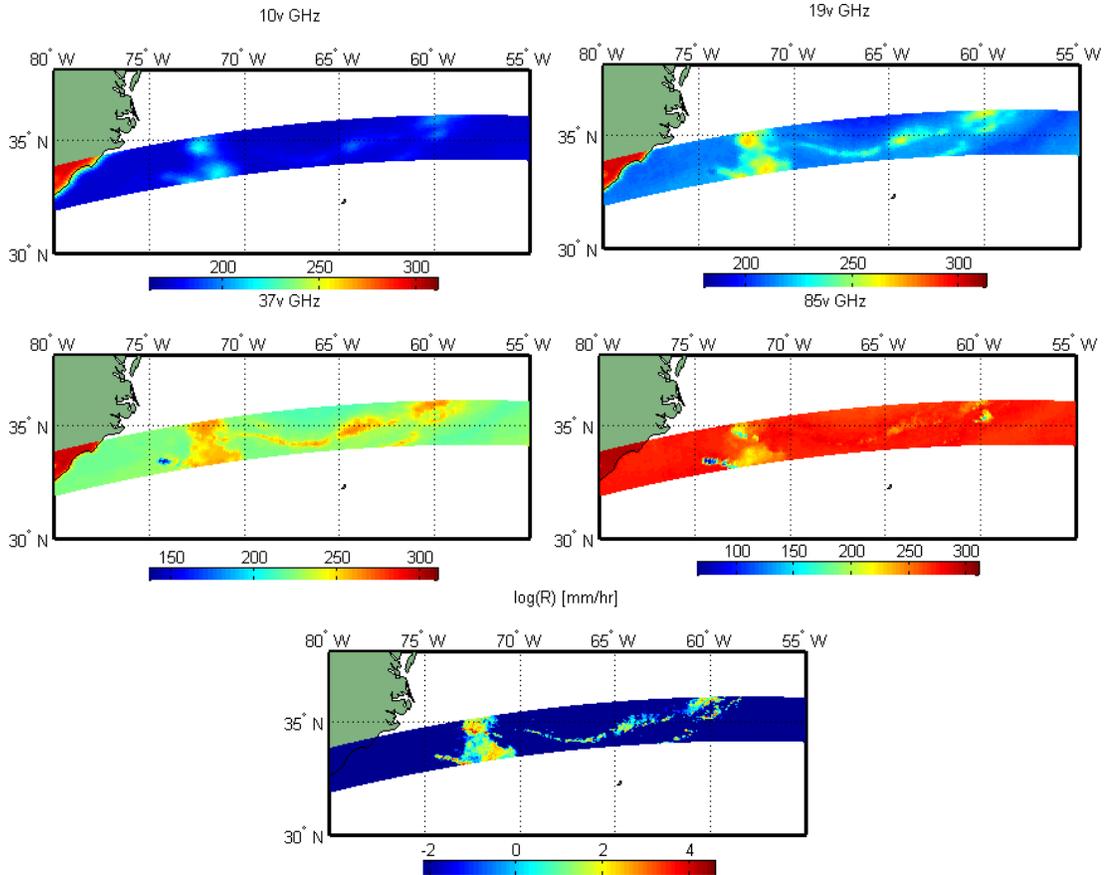


Figure 7.3: 1B11 Brightness Temperature (BT) and 2A25 surface rain rates for a particular rainfall event over ocean at orbital track No. 7 on June, 28, 1998. Top panel and middle panel (from left to right): vertical polarization of the BT images at central frequencies 10, 19, 37, and 85 GHz. Bottom panel: surface rain rates obtained from 2A25 algorithm. Notice that the presented BT images are mapped onto the PR grids over the inner swath.

the 1B11 product, mapped onto the PR grid points, and 2A25 surface rain rates at orbital track No. 7 on June, 28, 1998.

7.4 Rain/No-rain Classification

In this section, we show grouped scatter plots to gain insight about the nature of rain/no-rain classification problems given multispectral BT channels. Figure 7.4 shows the matrix-scatter plot of the TMI channels over *ocean* for fifty thousand randomly selected rainy (purple) and non-rainy (green) pairs of BT values. The data are collected from one hundred randomly selected orbital observations in the year of 2002¹.

¹Throughout this report, I only use the data collected in the year of 2002 for simplicity in coding and testing of the proposed algorithm. Eventually, the the final results will be documents for a more comprehensive and larger data set.

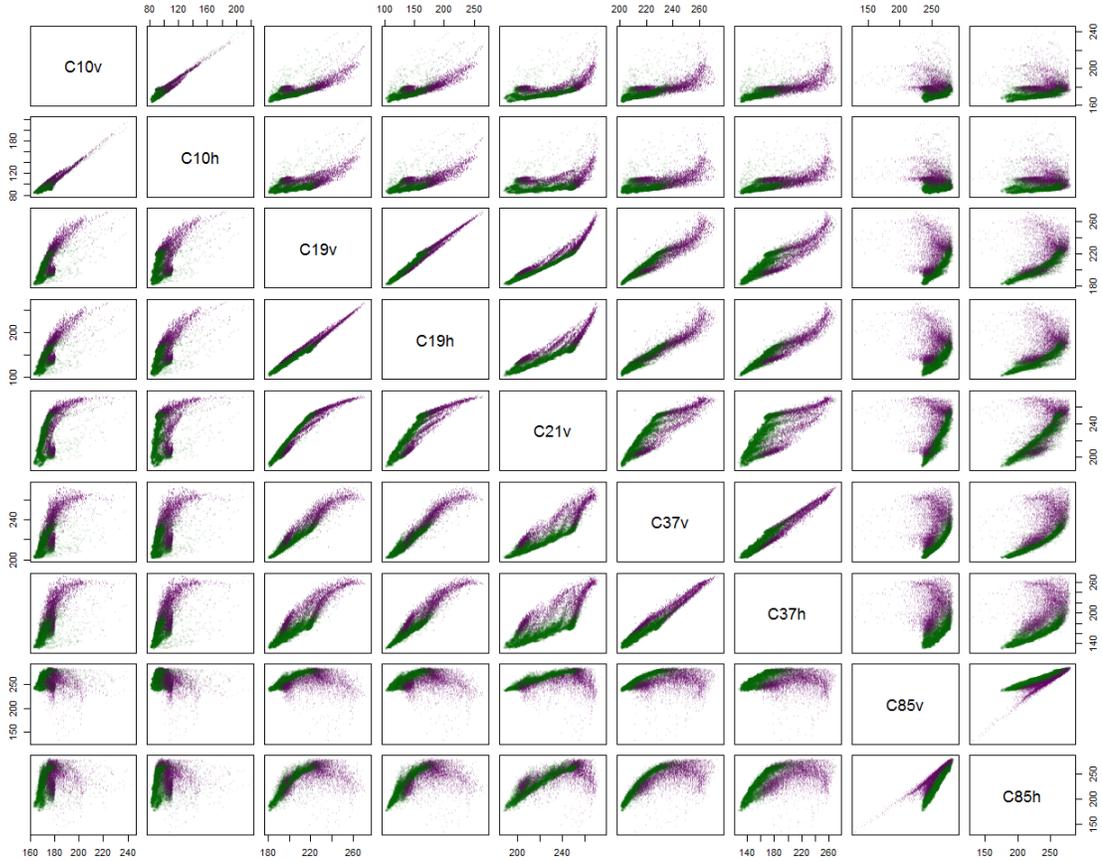


Figure 7.4: Matrix-scatter plot of rainy (purple) and non-rainy (green) pixels for fifty thousand pairs of the TMI multispectral BT values over ocean. The corresponding radar observations are used to discriminate between rainy and non-rainy pixels. The rainy and non-rainy profiles are relatively well separated, especially in some particular channels. Separability of rainy and non-rainy BT-profiles over ocean is more pronounced than those collected over land (see, 7.5).

For rainfall detection-estimation purposes, Figure 7.4 mainly reveals the following important points:

1. Data exhibit some redundancy. In other words, most of the horizontal and vertical polarizations are linearly related (except 85 GHz). Therefore, dimensionality reduction may improve computational efficiency and reduce observational noise.
2. Classification of rainy and non-rainy areas based on the multispectral BT values seems beyond a linear discriminant analysis and decision surfaces are typically nonlinear. It is interesting to note that for some channel pairs (e.g., C10 and C37 GHz) the rainy and non-rainy values almost leave on linear subspaces. However, there are also other channel pairs for which the rainy and non-rainy values leave on nonlinear manifolds (e.g., C85 and C19 GHz).

3. Non-rainy pixels leave on more coherent and lower dimensional manifolds than the rainy pixels. This observation tells us that potentially the singular values of the non-rainy BT values decay faster than those of the rainy ones. In other words, the non-rainy profiles can be better explained by fewer principle components than principle components needed for the rainy ones. This observation motivates us to further study the behavior of the BT values after applying the Karhunen-Loéve transformation (KLT) or proper orthogonal decomposition (POD).

Figure 7.5 demonstrates similar matrix-scatter plot for the rainy (purple) and non-rainy (green) pairs of the BT values over *land*. Compared to the previous plots (Figure 7.4), it can be seen that in most cases, the non-rainy areas are more scattered and effectively leave in higher dimensions. This signature is consistent with the fact that the background temperature over land shows more variability than the background temperature over ocean. As is evident, the rainy and non-rainy pairs are more blended than those of the ocean data. Surprisingly, channel 21.3 GHz (C27v) shows collinearity with its nearby vertically polarized channels (i.e., C19v, C37v). In channel 85 GHz, separability of rainy and non-rainy pairs is not as clear as it is over the ocean. Unlike to the ocean data, the horizontal and vertical polarizations in this channel show collinearity and seem to leave in a linear subspace with less pronounced separability.

The most important observations are: (a) The rainy pairs over land are more *clustered* rather than leaving on a coherent subspace or manifold, (b) The rainy and non-rainy pairs show more overlapping regions over land than over ocean.

Therefore, it seems clear that the rainfall discriminate analysis is more difficult over land than over ocean and the family of linear Bayesian classifiers may not be the best choice for rain/no-rain classification problem. Instead, we need to focus on *local* and adaptive detection-estimation frameworks (e.g., SPaR), which are capable to effectively approximate nonlinear and discontinuous decision boundaries.

7.5 Some Preliminary Results

I have randomly collected about three million 1-by-1 patches ² of the BT arrays and their coincidental PR measurements from the previously explained data set and stored them as columns of the BT (\mathbf{D}_T) and PR (\mathbf{D}_R) dictionaries. The SPaR algorithm can be easily used for classification/detection purposes as well. At this stage of the algorithmic implementation, which is very naive, the probability of detecting rainy pixels from multispectral BT

²At this stage the algorithm is coded only for pixel-wise estimation (for simplicity) and patch-wise estimation scheme will be developed in near future. Improved results are expected for the patch-wise estimation scheme, as the inverse problem in (7.2) is further constrained by additional spatial information.

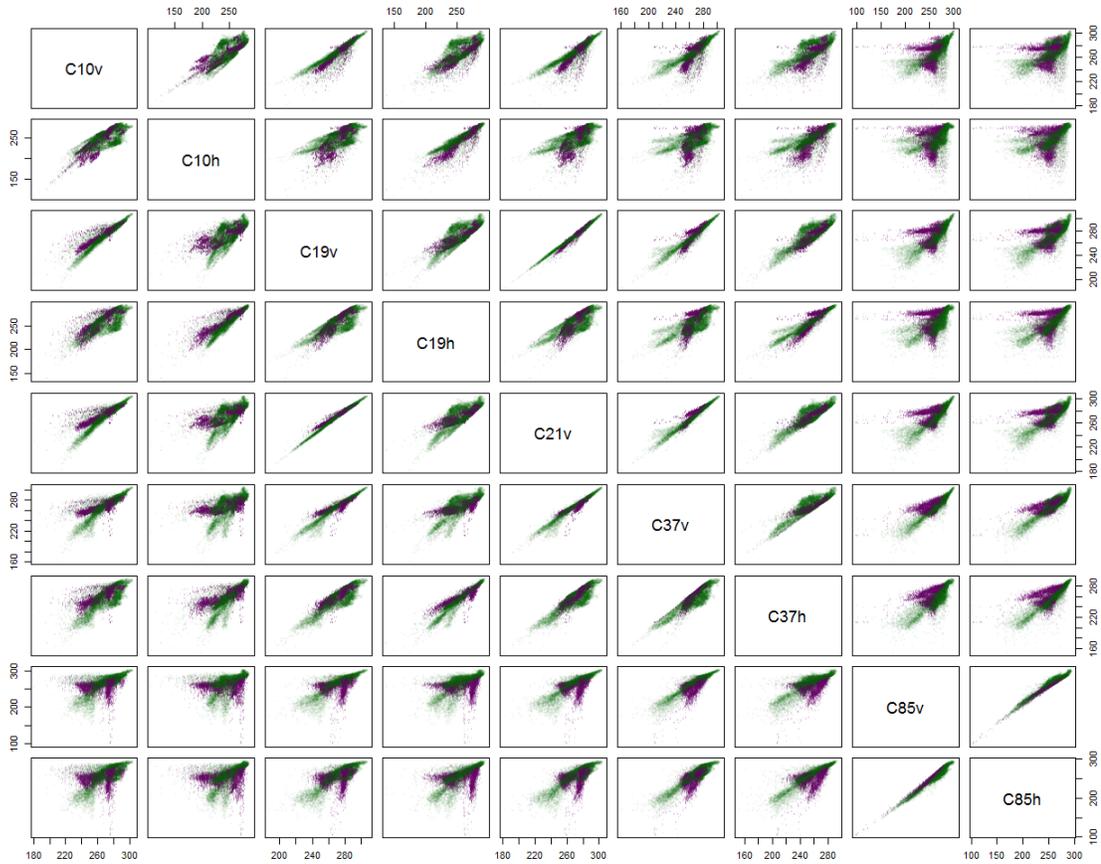


Figure 7.5: Matrix-scatter plot of rainy (purple) and non-rainy (green) pairs of the TMI channels over land. This plot encodes fifty thousand profiles of the BT values, obtained from one hundred randomly selected orbital information in the year of 2002. The corresponding radar observations are used to discriminate between rainy and non-rainy pairs. This plot shows that the structure of the BT data over land is more complex and irregular than the structure of the BT data over ocean. The rainy pairs are separately clustered and exhibit large overlapping regions with the non-rainy pairs.

measurements exceeds 93%. In other words, the probability of misclassification or false detection is less than 7%. This probability is obtained from more than a million independently selected validation cases. In other words, we randomly sampled more than one million of multispectral BT values from the data set (independent of the collected dictionaries) and correctly classified more than 93% of them as rainy and non-rainy profiles.

Figure 7.6 shows a preliminary result of a retrieval (estimation) test example for that particular storm shown in Figure 7.3. This sample result (bottom panel in Figure 7.6) is very comparable with 2A12 product, which is based on the currently operational Goddard profiling algorithm (Gprof-2010). Gprof algorithm uses a Bayesian approach to match the observed brightness temperatures to hydrometeor profiles derived from cloud resolving models (CRMs).

Comparing the SPaR (bottom panel in Figure 7.6) and 2A12 product (middle panel in

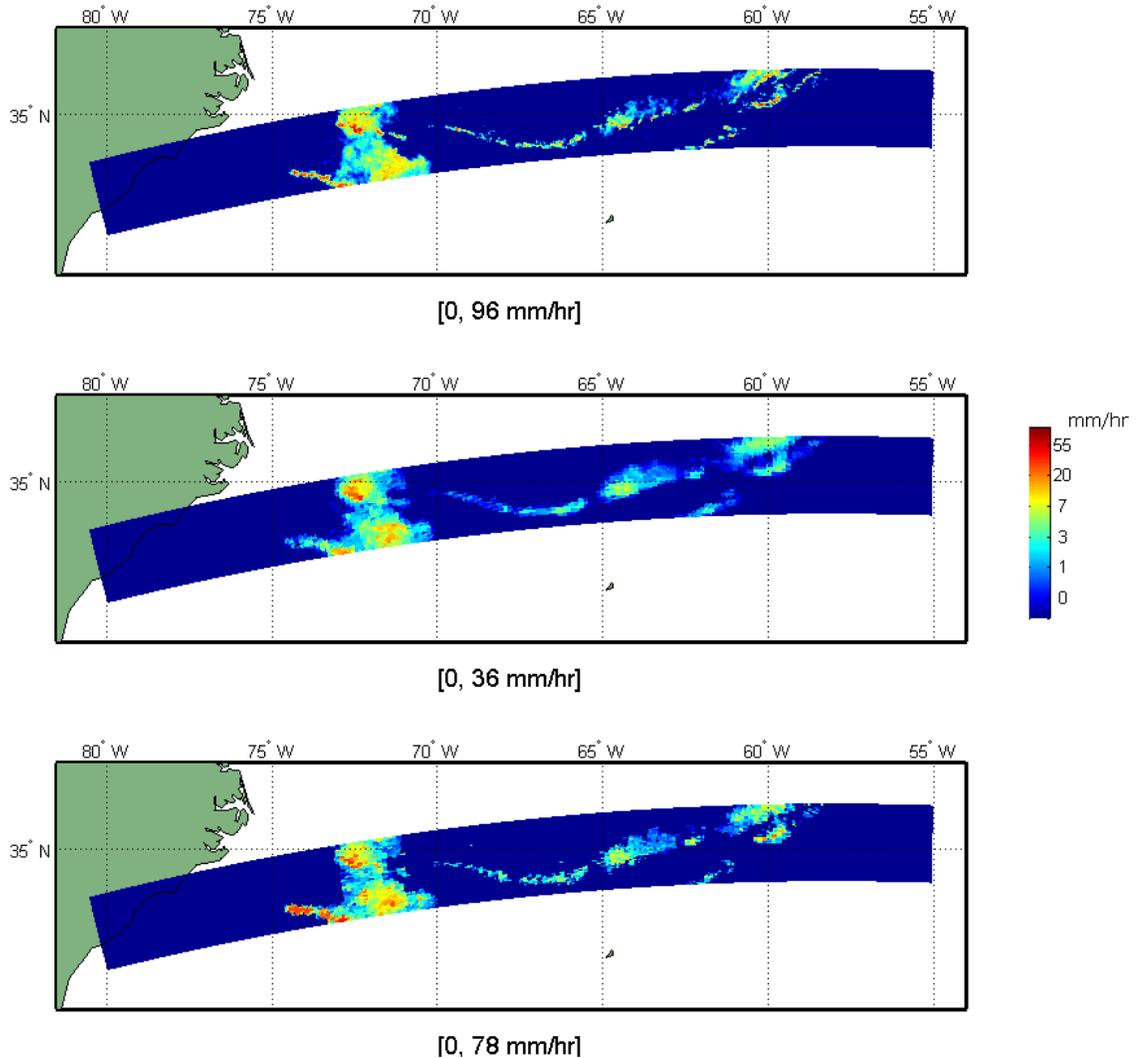


Figure 7.6: Comparing the retrieved rain rates for a sample event over ocean at orbital track No. 7 on June, 28, 1998. (Top panel) PR-2A25 surface rain rates, (middle panel) TMI-2A12, and (bottom panel) SPaR rainfall retrieved rain rates in [mm/hr].

Figure 7.6) with the PR-2A25 (top panel in Figure 7.6), we see that the SPaR rainfall recovery is relatively sharper and captures more small scale structures of the storm than that of the 2A12 product. The maximum rain rate in PR-2A25 is 96 [mm/hr], while the maximum retrieved values in 2A12 and SPaR is about 36 [mm/hr] and 78 [mm/hr], respectively. Although, the SPaR still underestimates the retrieved extreme rain rates; however, the results are markedly improved compared to the 2A12 products.

Figure 7.7 shows another rainfall retrieval experiment for Hurricane Danielle 08/29/2010 (UTC 09:48:00). Hurricane Danielle, the first major hurricane of the 2010 Atlantic season, intensified into a Category 4 storm back on the 27th of August with sustained winds estimated at 115 knots (~ 132 mph) by the National Hurricane Center as it was moving northwest towards Bermuda. We see that both of the 2A12 and SPaR algorithm under-

estimate the extreme precipitation rates; however, the SPaR performs slightly better and improves the recover range by almost 50%. The mean squared error is also improved by more than 15%, considering the PR as the ground truth observation. The SPaR retrieval is again sharper and contains more structural details of the hurricane. Typically, 2A12 algorithm retrieves larger areas of low rain rates surrounding the high-intense rain cells (Figure 7.7). These low rain rates are typically way below the accuracy of the PR sensor, which is greater than 0.7 [mm/hr]. The SPaR is naturally closer to the PR-2A25, however, it is feasible to use CRM's BT profiles in the SPaR dictionaries and improve its capacity to recover low-rain rates, if it is needed.

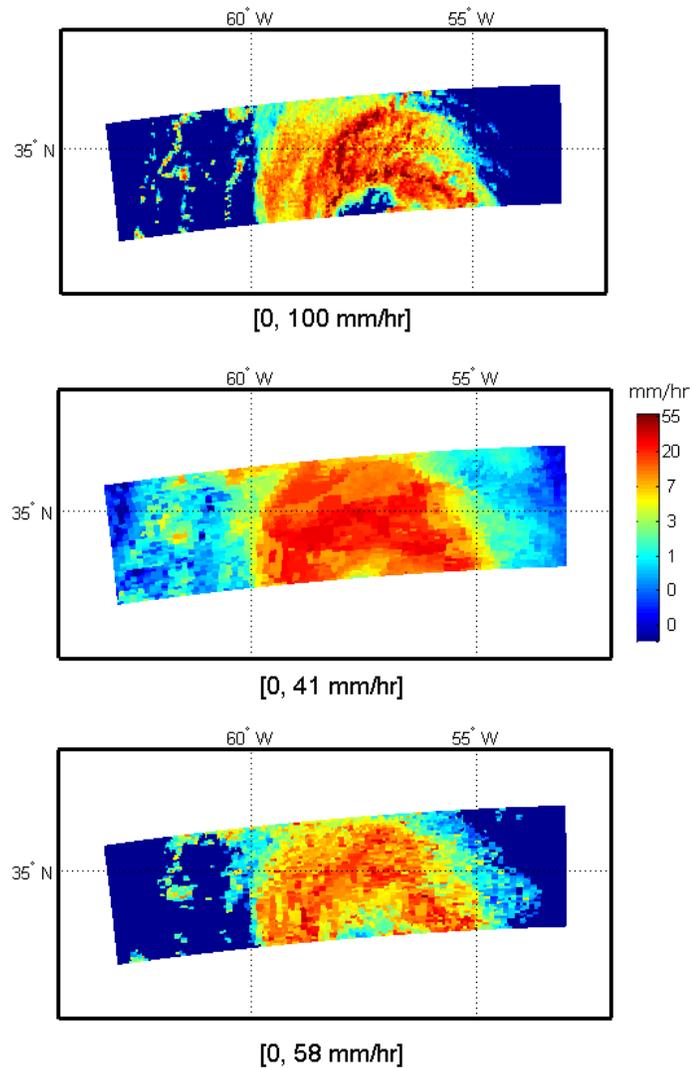


Figure 7.7: Comparing the retrieved rain rates for Hurricane Danielle 08/29/2010 (UTC 09:48:00). (Top panel) PR-2A25 surface rain rates in [mm/hr], (middle panel) TMI-2A12 and (bottom panel) SPaR rainfall retrieved rain rates.

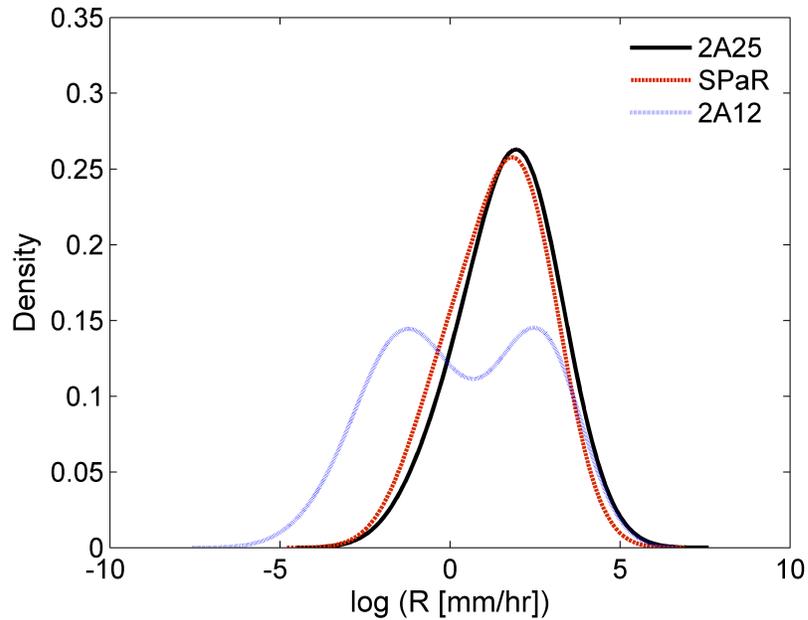


Figure 7.8: Kernel density estimates of the retrieved rain rates for the Hurricane Danielle over the demonstrated snapshot in Figure 7.7. The estimated probability density functions of the 2A12 (dotted blue) and SPaR (dashed red) are demonstrated versus the 2A25 (solid black), which is considered to be the best and ground truth estimate of the rain rates. The SPaR and 2A25 show a good agreement, while the tail of the 2A25 is naturally thicker than the SPaR. On the other hand, 2A12 is bimodal and is not very well consistent with the density of the 2A25 product.

Figure 7.8 shows a kernel density estimation (*Bowman and Azzalini., 1997*) of the probability density function of the retrieved rain rates over the demonstrated snapshot of the Hurricane Danielle in Figure 7.7. It is clear that in the sense of probability distribution the SPaR method very well matches the PR-2A25 while the 2A12 is relatively bimodal and is not very close to the PR observations. This bimodality mainly arises due to the retrieval of low rain rates in the 2A12, which does not exist in 2A25.

Chapter 8

Conclusion

In this thesis, we extensively studied statistical structure of precipitation (reflectivity) images in spatial and transform domains. We confirmed previous findings on power law behavior of the rainfall fields in the Fourier domain and provided new insights about the dependence of the rainfall spectral signature with the underlying topographic features (*Ebtehaj and Foufoula-Georgiou, 2010*). Statistical structure of rainfall fields was thoroughly investigated in wavelet domain for a relatively large set of the NEXRAD high-resolution rainfall images, collected in two ground validation sites of the TRMM satellite (*Ebtehaj and Foufoula-Georgiou, 2011a*). We observed that the distribution of rainfall wavelet coefficients exhibit non-Gaussian heavy tailed behavior and can be well explained by the family of Generalized Gaussian Density (GGD). The GGD is a flexible three parameters density which contains the Gaussian and Laplace distributions as its special cases. This non-Gaussian and well parameterizable structure of rainfall fields prompted us to exploit non-linear filtering methodologies for multiscale fusion of multisensor precipitation observations. To this end, inspired by the developments in the theories of Gaussian scale mixtures (GSM), we exploited a conditionally linear probability model to develop non-Gaussian rainfall fusion methodologies that allows us to preserve heavy tailed structure of rainfall fields (*Ebtehaj and Foufoula-Georgiou, 2011b*). Explaining the rainfall wavelet coefficients (smoothed derivatives) by the GGD density, directed us to explore the sparsity of rainfall fields in the wavelet domain. We found that a large number of the rainfall wavelet coefficients are very close to zero (forming a large mass at the center of the density), while a few of them are significantly non-zero (forming the heavy tails of the density). Taking advantage of the rainfall sparsity, we proposed a new variational data driven approach to the problem of rainfall downscaling (*Ebtehaj et al., 2012*), using ℓ_1 -norm regularization and dictionary learning. We extended this idea and suggested a unified sparsity-promoting variational framework that ties together the hydro-meteorological inverse problems of downscaling (DS), data fusion (DF), and data assimilation (DA) (*Ebtehaj and Foufoula-Georgiou, 2013*). This idea was extended and examined via simple but important variational data assimilation prob-

lems of land surface heat and mass fluxes using linear advection-diffusion equation (*Ebtehaj et al.*, 2013). Inspired by the rainfall sparsity, a new supervised statistical learning approach was also proposed for spaceborne rainfall retrieval problem. The preliminary results of the implemented algorithm were compared with the currently operation algorithm called Gprof-2010 by *Kummerow et al.* (2010). The comparison showed that the results are much sharper than those of the Gprof-2010 algorithm and capitalized on improved recovery of extreme rainfall intensity values form spaceborne passive microwave observations.

In general, in this thesis we emphasized on the importance of proper sparse regularization in hydro-meteorological inverse problems and its connection with the statistical properties of the underlying state of interest. We showed that a sparse regularization in an appropriately chosen domain not only makes hydro-meteorological inverse problems sufficiently well-posed but leads to a more accurate solution with improved uncertainty of estimate, compared to the classic least squares solutions.

Although the focus of this thesis was on rainfall data, the presented frameworks can be potentially applied to other hydro-meteorological inverse problems, such as soil moisture downscaling, fusion, and data assimilation. Future research can be devoted to explore the ℓ_1 -norm regularization for solving non-linear data assimilation problems. In addition, exploring dictionary learning approaches (see, *Ebtehaj et al.*, 2012) for rainfall retrieval problem is a promising research direction.

Bibliography

- Abry, P., S. Jaffard, and B. Lashermes (2004), Revisiting scaling, multifractal, and multiplicative cascades with the wavelet leader lens, in *Proc. SPIE*, vol. 5607, pp. 103–117.
- Adler, R., G. Huffman, A. Chang, R. Ferraro, P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, et al. (2003), The version 2 global precipitation climatology project GPCP monthly precipitation analysis (1979-present), *J. Hydrometeor.*, *4*(6), 1147–1167.
- Afshar, A., and M. A. Marino (1978), Model for simulating soil-water content considering evapotranspiration, *Journal of Hydrology*, *37*(3–4), 309 – 322, doi:10.1016/0022-1694(78)90022-7.
- AghaKouchak, A., A. Mehran, H. Norouzi, and A. Behrangi (2012), Systematic and random error components in satellite precipitation data sets, *Geophys. Res. Lett.*, *39*(9), doi: 10.1029/2012GL051592.
- Aharon, M., M. Elad, and A. Bruckstein (2006), K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation, *IEEE Trans. Signal. Proces.*, *54*(11), 4311–4322, doi:10.1109/TSP.2006.881199.
- Anderson, J. L. (2001), An Ensemble Adjustment Kalman Filter for Data Assimilation, *Mon. Wea. Rev.*, *129*(12), 2884–2903, doi:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.
- Andrews, D. F., and C. L. Mallows (1974), Scale Mixtures of Normal Distributions, *J. R. Stat. Soc.*, *36*(1), pp. 99–102.
- Badas, M. G., R. Deidda, and E. Piga (2006), Modulation of homogeneous space-time rainfall cascades to account for orographic influences, *Nat. Hazard. Earth. Sys.*, *6*(3), 427–437, doi:10.5194/nhess-6-427-2006.
- Bai, Z., J. Demmel, J. Dongarra, A. Ruhe, and H. Van Der Vorst (1987), *Templates for the solution of algebraic eigenvalue problems: a practical guide*, vol. 11, SIAM, Philadelphia.

- Barnard, J., R. McCulloch, and X. Meng (2000), Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage, *Stat. Sinica*, 10(4), 1281–1312.
- Barros, A., and R. Kuligowski (1998), Orographic effects during a severe wintertime rainstorm in the Appalachian Mountains, *Mon. Wea. Rev.*, 126(10), 2648–2672.
- Bateni, S. M., and D. Entekhabi (2012), Surface heat flux estimation with the ensemble Kalman smoother: Joint estimation of state and parameters, *Water Resour. Res.*, 48(8), doi:10.1029/2011WR011542.
- Bennett, A. F., and P. C. McIntosh (1982), Open Ocean Modeling as an Inverse Problem: Tidal Theory, *J. Phys. Oceanogr.*, 12(10), 1004–1018, doi:10.1175/1520-0485(1982)012<1004:OOMAAI>2.0.CO;2.
- Beran, J. (1994), *Statistics for long-memory processes*, vol. 61, 315 pp., Chapman and Hall, New York, monogr. Stat. Appl. Prob.
- Bertsekas, D. (1976), On the Goldstein-Levitin-Polyak gradient projection method, *IEEE Trans. Automat. Contr.*, 21(2), 174 – 184, doi:10.1109/TAC.1976.1101194.
- Bertsekas, D. P. (1999), *Nonlinear Programming*, 2nd ed., 794 pp., Athena Scientific, Belmont, MA.
- 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.
- Bousquet, O., and B. F. Smull (2003), Observations and impacts of upstream blocking during a widespread orographic precipitation event, *Quarterly Journal of the Royal Meteorological Society*, 129(588), 391–409, doi:10.1256/qj.02.49.
- Bouttier, F., and P. Courtier (2002), Data assimilation concepts and methods, *Meteorological training course lecture series. ECMWF*, p. 59.
- Bowman, A. W., and A. Azzalini. (1997), *Applied Smoothing Techniques for Data Analysis*, New York: Oxford University Press.
- Boyd, S., and L. Vandenberghe (2004), *Convex optimization*, 716 pp., Cambridge University Press, New York.
- Budd, C., M. Freitag, and N. Nichols (2011), Regularization techniques for ill-posed inverse problems in data assimilation, *Computers & Fluids*, 46(1), 168–173, doi:10.1016/j.compfluid.2010.10.002.

- Burt, P., and E. Adelson (1983), The Laplacian pyramid as a compact image code, *IEEE Trans. Commun.*, *31*(4), 532–540.
- Candes, E., and T. Tao (2006), Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?, *IEEE Trans. Inform. Theory.*, *52*(12), 5406–5425, doi: 10.1109/TIT.2006.885507.
- Caparrini, F., F. Castelli, and D. Entekhabi (2003), Mapping of land-atmosphere heat fluxes and surface parameters with remote sensing data, *Bound.-Layer Meteor.*, *107*(3), 605–633.
- Castro, C. L., S. Pielke, Roger A., and G. Leoncini (2005), Dynamical downscaling: Assessment of value retained and added using the Regional Atmospheric Modeling System RAMS, *J. Geophys. Res.*, *110*(D5), D05,108, doi:doi:10.1029/2004JD004721.
- Chan, R. H.-F., and X.-Q. Jin (2007), *An Introduction to Iterative Toeplitz Solvers*, SIAM, Philadelphia.
- Chapra, S. C. (2008), *Surface Water Quality Modeling*, Waveland Press, Inc. IL, USA.
- Chen, S., D. Donoho, and M. Saunders (2001), Atomic Decomposition by Basis Pursuit, *SIAM rev.*, *43*(1), 129–159.
- Chen, S. S., D. L. Donoho, and M. A. Saunders (1998), Atomic decomposition by basis pursuit, *SIAM J. Sci. Comput.*, *20*, 33–61.
- Chou, K., A. Willsky, and A. Benveniste (1994), Multiscale recursive estimation, data fusion, and regularization, *IEEE Trans. Automat. Contr.*, *39*(3), 464–478, doi:10.1109/9.280746.
- Ciach, G. J., and W. F. Krajewski (1999), On the estimation of radar rainfall error variance, *Adv. Water Resour.*, *22*(6), 585–595.
- Claerbout, J., and F. Muir (1973), Robust modeling with erratic data, *Geophysics*, *38*, 826, doi:DOI:10.1190/1.1440378.
- Coifman, R., and D. Donoho (1995), Translation invariant de-noising, *Lecture Notes in Statist.*, *103*, 125–150.
- Cooley, J. W., and J. W. Tukey (1965), An algorithm for the machine calculation of complex Fourier series, *Math. Comp.*, *19*(90), 297–301.
- Courtier, P., and O. Talagrand (1990), Variational assimilation of meteorological observations with the direct and adjoint shallow-water equations, *Tellus A*, *42*(5), 531–549.

- Courtier, P., J.-N. Thépaut, and A. Hollingsworth (1994), A strategy for operational implementation of 4D-VAR, using an incremental approach, *Quart. J. Roy. Meteor. Soc.*, *120*(519), 1367–1387, doi:10.1002/qj.49712051912.
- Daley, R. (1993), *Atmospheric data analysis*, 472 pp., Cambridge University Press.
- Deidda, R. (2000), Rainfall downscaling in a space-time multifractal framework, *Water Resour. Res.*, *36*(7), 1779–1794.
- Deidda, R., M. G. Badas, and E. Piga (2004), Space-time scaling in high-intensity Tropical Ocean Global Atmosphere Coupled Ocean-Atmosphere Response Experiment (TOGA-COARE) storms, *Water Resources Research*, *40*(2), n/a–n/a, doi:10.1029/2003WR002574.
- Deidda, R., M. G. Badas, and E. Piga (2006), Space-time multifractality of remotely sensed rainfall fields, *Journal of Hydrology*, *322*(1–4), 2 – 13.
- Donoho, D. (1995), De-noising by soft-thresholding, *IEEE Trans. Inform. Theory.*, *41*(3), 613–627, doi:10.1109/18.382009.
- Donoho, D. L., and P. B. Stark (1989), Uncertainty principles and signal recovery, *SIAM J. Appl. Math.*, *49*, 906–931.
- Durrett, R. (1999), *Essentials of Stochastic Processes*, Springer-Verlag, N.Y.
- Ebtehaj, A., M. Zupanski, G. Lerman, and E. Foufoula-Georgiou (2013), Variational Data Assimilation via Sparse Regularization, *Tellus*, under Review.
- Ebtehaj, A. M., and E. Foufoula-Georgiou (2010), Orographic signature on multiscale statistics of extreme rainfall: A storm-scale study, *J. Geophys. Res.*, *115*, D23112, doi:10.1029/2010JD014093.
- Ebtehaj, A. M., and E. Foufoula-Georgiou (2011a), Statistics of precipitation reflectivity images and cascade of Gaussian-scale mixtures in the wavelet domain: A formalism for reproducing extremes and coherent multiscale structures, *J. Geophys. Res.*, *116*, D14110, doi:10.1029/2010JD015177.
- Ebtehaj, A. M., and E. Foufoula-Georgiou (2011b), Adaptive fusion of multisensor precipitation using Gaussian-scale mixtures in the wavelet domain, *J. Geophys. Res.*, *116*, D22110, doi:10.1029/2011JD016219.
- Ebtehaj, A. M., and E. Foufoula-Georgiou (2013), On Variational Downscaling, Fusion and Assimilation of Hydro-meteorological States: A Unified Framework via Regularization, *Water Resour. Res.*, doi:10.1002/wrcr.20424, in press.

- Ebtehaj, A. M., E. Foufoula-Georgiou, and G. Lerman (2012), Sparse regularization for precipitation downscaling, *J. Geophys. Res.*, *116*, D22110, doi:10.1029/2011JD017057.
- Elad, M. (2010), *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, 376 pp., Springer Verlag.
- Elad, M., and A. Feuer (1997), Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images, *IEEE Trans. Image. Process.*, *6*(12), 1646–1658, doi:10.1109/83.650118.
- Engan, K., S. O. Aase, and J. H. Husoy (1999), Method of optimal directions for frame design, *paper presented at ICASSP 99, IEEE*, phoenix, Ariz.
- Engan, K. S., K., and J. H. Husoy (2007), A family of iterative LS-based dictionary learning algorithms, ILS-DLA, for sparse signal representation, *Digital Signal Process.*, *17*, 32–49.
- Entekhabi, D., H. Nakamura, and E. Njoku (1994), Solving the inverse problem for soil moisture and temperature profiles by sequential assimilation of multifrequency remotely sensed observations, *IEEE Trans. Geosci. Remote Sens.*, *32*(2), 438–448, doi:10.1109/36.295058.
- Evensen, G. (1994a), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, *99*(C5), 10,143–10,162.
- Evensen, G. (1994b), Inverse methods and data assimilation in nonlinear ocean models, *Physica D: Nonlinear Phenomena*, *77*(1–3), 108 – 129, doi:10.1016/0167-2789(94)90130-9, <ce:title>Special Issue Originating from the 13th Annual International Conference of the Center for Nonlinear Studies Los Alamos, NM, USA, 17–21 May 1993 </ce:title>.
- Ferraris, L., S. Gabellani, N. Rebora, and A. Provenzale (2003), A comparison of stochastic models for spatial rainfall downscaling, *Water Resour. Res.*, *39*(12), doi:10.1029/2003WR002504.
- Fetter, C. (1994), *Applied Hydrogeology*, Prentice Hall, New Jersey, USA, Fourth Edition.
- Figueiredo, M., R. Nowak, and S. Wright (2007), Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems, *IEEE J. Sel. Topics Signal Process.*, *1*(4), 586–597, doi:10.1109/JSTSP.2007.910281.
- Freitag, M. A., N. K. Nichols, and C. J. Budd (2010), L1-regularisation for ill-posed problems in variational data assimilation, *PAMM*, *10*(1), 665–668, doi:10.1002/pamm.201010324.

- Freitag, M. A., N. K. Nichols, and C. J. Budd (2012), Resolution of sharp fronts in the presence of model error in variational data assimilation, *Quart. J. Roy. Meteor. Soc.*, doi:10.1002/qj.2002.
- Fristedt, B., N. Jain, and N. Krylov (2007), *Filtering and prediction: a primer*, vol. 38, American Mathematical Society (SIAM).
- Gaspari, G., and S. E. Cohn (1999), Construction of correlation functions in two and three dimensions, *Quart. J. Roy. Meteor. Soc.*, 125(554), 723–757, doi:10.1002/qj.49712555417.
- Geman, S., and D. Geman (1984), Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intell.*, (6), 721–741.
- Ghil, M. (1989), Meteorological data assimilation for oceanographers. Part I: Description and theoretical framework, *Dyn. Atmos. Oceans*, 13(3), 171–218.
- Ghil, M., and P. Malanotte-Rizzoli (1991), Data Assimilation in Meteorology and Oceanography, pp. 141 – 266, Elsevier, doi:10.1016/S0065-2687(08)60442-2.
- Ghil, M., S. Cohn, J. Tavantzis, K. Bube, and E. Isaacson (1981), Applications of Estimation Theory to Numerical Weather Prediction, in *Dynamic Meteorology: Data Assimilation Methods, Applied Mathematical Sciences*, vol. 36, edited by L. Bengtsson, M. Ghil, and E. Källén, pp. 139–224, Springer New York, doi:10.1007/978-1-4612-5970-1_5.
- Golub, G., P. Hansen, and D. O’Leary (1999), Tikhonov regularization and total least squares, *SIAM J. Matrix Anal. Appl.*, 21(1), 185–194.
- Gonzalez, R., and R. Woods (2008), *Digital Image Processing*, Pearson/Prentice Hall.
- Gorenburg, I. P., D. McLaughlin, and D. Entekhabi (2001), Scale-recursive assimilation of precipitation data, *Adv. Water Resour.*, 24(9–10), 941–953, doi:10.1016/S0309-1708(01)00033-1.
- Gupta, V., and E. Waymire (1993), A statistical analysis of mesoscale rainfall as a random cascade, *J. Appl. Meteorol.*, 32, 251–251.
- Gupta, V. K., and E. Waymire (1990), Multiscaling properties of spatial rainfall and river flow distributions, *J. Geophys. Res.*, 95(D3), 1999–2009, doi:10.1029/JD095iD03p01999.
- Hansen, P. (1998), *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*, vol. 4, Society for Industrial Mathematics (SIAM), Philadelphia.
- Hansen, P. (2010), *Discrete inverse problems: insight and algorithms*, vol. 7, Society for Industrial & Applied Mathematics (SIAM), Philadelphia, PA, USA.

- Hansen, P., and D. O’Leary (1993), The use of the L-curve in the regularization of discrete ill-posed problems, *SIAM J Sci Comput*, *14*(6), 1487–1503.
- Hansen, P., J. Nagy, and D. Óleary (2006), *Deblurring images: matrices, spectra, and filtering*, vol. 3, Society for Industrial & Applied Mathematics (SIAM), Philadelphia, PA, USA.
- Harris, D., M. Menabde, A. Seed, and G. Austin (1996), Multifractal characterization of rain fields with a strong orographic influence, *J. Geophys. Res.*, *101*(D21), 26,405–26,414, doi:10.1029/96JD01656.
- Harris, D., E. Foufoula-Georgiou, K. K. Droegemeier, and J. J. Levit (2001), Multiscale statistical properties of a high-resolution precipitation forecast, *J. Hydrometeor.*, *2*(4), 406–418.
- Hawkins, D. M. (2004), The problem of overfitting, *J. Chem. Inf. Comput. Sci.*, *44*(1), 1–12.
- Hong, Y., C. D. Kummerow, and W. S. Olson (1999), Separation of Convective and Stratiform Precipitation Using Microwave Brightness Temperature, *J. Appl. Meteor.*, *38*(8), 1195–1213, doi:10.1175/1520-0450(1999)038<1195:SOCASP>2.0.CO;2.
- Hong, Y., K. Hsu, S. Sorooshian, and X. Gao (2004), Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system, *J. Appl. Meteorol.*, *43*(12), 1834–1853.
- Hossain, F., and E. N. Anagnostou (2005), Numerical investigation of the impact of uncertainties in satellite rainfall estimation and land surface model parameters on simulation of soil moisture, *Adv. Water Resour.*, *28*(12), 1336–1350.
- Hossain, F., and E. N. Anagnostou (2006), A two-dimensional satellite rainfall error model, *IEEE Trans. Geosci. Remote Sens.*, *44*(6), 1511–1522.
- Hostetler, S. (1994), The (continuing) problem of discordant scales, *Clim. Change*, *27*(4), 345–350.
- Hu, Z., and S. Islam (1995), Prediction of Ground Surface Temperature and Soil Moisture Content by the Force-Restore Method, *Water Resources Research*, *31*(10), 2531–2539, doi:10.1029/95WR01650.
- Huang, J., and D. Mumford (1999), Statistics of natural images and models, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 637–663, doi:10.1109/CVPR.1999.786990.

- Huber, P. (1964), Robust estimation of a location parameter, *Ann. of Math. Stat.*, 35(1), 73–101.
- Huber, P. (1981), *Robust statistics*, vol. 1, John Wiley & Sons, Inc., New York.
- Huffman, G., R. Adler, B. Rudolf, U. Schneider, and P. Keehn (1995), Global precipitation estimates based on a technique for combining satellite-based estimates, rain gauge analysis, and NWP model precipitation information, *J. Climate*, 8(5), 1284–1295.
- Huffman, G., R. Adler, M. Morrissey, D. Bolvin, S. Curtis, R. Joyce, B. McGavock, and J. Susskind (2001), Global precipitation at one-degree daily resolution from multisatellite observations, *J. Hydrometeor.*, 2(1), 36–50.
- Huffman, G., D. Bolvin, E. Nelkin, D. Wolff, R. Adler, G. Gu, Y. Hong, K. Bowman, and E. Stocker (2007), The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, *J. Hydrometeor.*, 8(1), 38–55.
- Ide, K., P. Courtier, M. Gill, and A. Lorenc (1997), Unified notation for data assimilation: Operational, sequential, and variational, *J. Meteor. Soc. Japan*, 75, 181–189.
- Iguchi, T., T. Kozu, R. Meneghini, J. Awaka, and K. i. Okamoto (2000), Rain-Profiling Algorithm for the TRMM Precipitation Radar, *J. Appl. Meteor.*, 39(12), 2038–2052, doi:10.1175/1520-0450(2001)040<2038:RPAFTT>2.0.CO;2.
- Jochum, M., and R. Murtugudde (2006), Temperature advection by tropical instability waves, *J. Phys. Oceanogr.*, 36(4), 592–605.
- Johnson, C., B. J. Hoskins, and N. K. Nichols (2005a), A singular vector perspective of 4D-Var: Filtering and interpolation, *Quart. J. Roy. Meteor. Soc.*, 131(605), 1–19, doi: 10.1256/qj.03.231.
- Johnson, C., N. K. Nichols, and B. J. Hoskins (2005b), Very large inverse problems in atmosphere and ocean modelling, *Int. J. Numer. Meth. Fl.*, 47(8-9), 759–771, doi:10.1002/fld.869.
- Kalman, R. E. (1960), A new approach to linear filtering and prediction problems, *J Basic eng-T ASME*, 82(1), 35–45.
- Kalnay, E. (2003), *Atmospheric modeling, data assimilation, and predictability*, 341 pp., Cambridge University Press, New York.
- Kim, G., and A. Barros (2002), Downscaling of remotely sensed soil moisture with a modified fractal interpolation method using contraction mapping and ancillary data, *Remote Sens. Environ.*, 83(3), 400–413.

- Kim, S.-J., K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky (2007), An Interior-Point Method for Large-Scale ℓ_1 -Regularized Least Squares, *IEEE J. Sel. Topics Signal Process.*, *1*(4), 606–617, doi:10.1109/JSTSP.2007.910971.
- Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord (2009), Introduction of the GSI into the NCEP Global Data Assimilation System, *Wea. Forecasting*, *24*(6), 1691–1705, doi:10.1175/2009WAF2222201.1.
- Kumar, P. (1999), A multiple scale state-space model for characterizing subgrid scale variability of near-surface soil moisture, *IEEE Trans. Geosci. Remote.*, *37*(1), 182–197, doi:10.1109/36.739153.
- Kumar, P., and E. Foufoula-Georgiou (1993a), A multicomponent decomposition of spatial rainfall fields. 2. Self-similarity in fluctuations, *Water Resour. Res.*, *29*(8), 2533–2544.
- Kumar, P., and E. Foufoula-Georgiou (1993b), A multicomponent decomposition of spatial rainfall fields. 1. Segregation of large- and small-scale features using wavelet transforms, *Water Resour. Res.*, *29*(8), 2515–2532.
- Kummerow, C., W. Olson, and L. Giglio (1996), A simplified scheme for obtaining precipitation and vertical hydrometeor profiles from passive microwave sensors, *IEEE Trans. Geosci. Remote.*, *34*(5), 1213–1232.
- Kummerow, C., W. Barnes, T. Kozu, J. Shiue, and J. Simpson (1998), The Tropical Rainfall Measuring Mission (TRMM) Sensor Package, *J. Atmos. Oceanic Technol.*, *15*(3), 809–817, doi:10.1175/1520-0426(1998)015<0809:TTRMMT>2.0.CO;2.
- Kummerow, C. D., S. Ringerud, J. Crook, D. Randel, and W. Berg (2010), An Observationally Generated A Priori Database for Microwave Rainfall Retrievals, *J. Atmos. Oceanic Technol.*, *28*(2), 113–130, doi:10.1175/2010JTECHA1468.1.
- Lanser, D., and J. Verwer (1999), Analysis of operator splitting for advection–diffusion–reaction problems from air pollution modelling, *Journal of Computational and Applied Mathematics*, *111*(1–2), 201–216, doi:10.1016/S0377-0427(99)00143-0.
- Law, K. J. H., and A. M. Stuart (2012), Evaluating Data Assimilation Algorithms, *Mon. Wea. Rev.*, *140*(11), 3757–3782, doi:10.1175/MWR-D-11-00257.1.
- Le Cam, L. (1961), A stochastic description of precipitation, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 3, pp. 165–186.
- Ledoit, O., and M. Wolf (2004), A well-conditioned estimator for large-dimensional covariance matrices, *J. Multivariate Anal.*, *88*(2), 365–411.

- Levy, B. C. (2008), *Principles of Signal Detection and Parameter Estimation*, 1 ed., 639 pp., Springer Publishing Company, New York, USA, doi:10.1007/978-0-387-76544-0.
- Lewicki, M., and T. Sejnowski (2000), Learning overcomplete representations, *Neural Comput.*, *12*(2), 337–365.
- Liang, X., E. F. Wood, and D. P. Lettenmaier (1999), Modeling ground heat flux in land surface parameterization schemes, *J. Geophys. Res.*, *104*(D8), 9581–9600.
- Lin, Y.-L., R. L. Deal, and M. S. Kulie (1998), Mechanisms of cell regeneration, development, and propagation within a two-dimensional multicell storm, *J. Atmos. Sci.*, *55*(10), 1867–1886.
- Lin, Y.-L., S. Chiao, T.-A. Wang, M. L. Kaplan, and R. P. Weglarz (2001), Some Common Ingredients for Heavy Orographic Rainfall, *Wea. Forecasting*, *16*(6), 633–660, doi:10.1175/1520-0434(2001)016<0633:SCIFHO>2.0.CO;2.
- Lorenc, A. (1988), Optimal nonlinear objective analysis, *Quart. J. Roy. Meteor. Soc.*, *114*(479), 205–240.
- Lorenc, A. C. (1986), Analysis methods for numerical weather prediction, *Quart. J. Roy. Meteor. Soc.*, *112*(474), 1177–1194, doi:10.1002/qj.49711247414.
- Lorenc, A. C., S. P. Ballard, R. S. Bell, N. B. Ingleby, P. L. F. Andrews, D. M. Barker, J. R. Bray, A. M. Clayton, T. Dalby, D. Li, T. J. Payne, and F. W. Saunders (2000), The Met. Office global three-dimensional variational data assimilation scheme, *Quart. J. Roy. Meteor. Soc.*, *126*(570), 2991–3012, doi:10.1002/qj.49712657002.
- Lovejoy, S., and B. Mandelbrot (1985), Fractal properties of rain, and a fractal model, *Tellus A*, *37*(3), 209–232.
- Lovejoy, S., and D. Schertzer (1985), Generalized Scale Invariance in the Atmosphere and Fractal Models of Rain, *Water Resour. Res.*, *21*(8), 1233–1250, doi:10.1029/WR021i008p01233.
- Lovejoy, S., and D. Schertzer (1990), Multifractals, Universality Classes and Satellite and Radar, *J. Geophys. Res.*, *95*(D3), 2021–2034.
- Maggioni, V., R. H. Reichle, and E. N. Anagnostou (2012), The Impact of Rainfall Error Characterization on the Estimation of Soil Moisture Fields in a Land Data Assimilation System, *J. Hydrometeorol.*, *13*(3), 1107–1118, doi:10.1175/JHM-D-11-0115.1.
- Mallat, S. (1989), A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Trans. Pattern Anal. Mach. Intell.*, *11*(7), 674–693, doi:10.1109/34.192463.

- Mallat, S. (2009), *A wavelet tour of signal processing: the sparse way*, 3rd ed., 805 pp., Elsevier /Academic Press.
- Mallat, S., and G. Yu (2010), Super-Resolution With Sparse Mixing Estimators, *IEEE Trans. Image. Process.*, *19*(11), 2889–2900, doi:10.1109/TIP.2010.2049927.
- Mallat, S., and Z. Zhang (1993), Matching pursuits with time-frequency dictionaries, *IEEE Trans. Signal. Proces.*, *41*(12), 3397–3415, doi:10.1109/78.258082.
- Mandelbrot, B. B., F. Adlai, and L. Calvet (1997), A multifractal model of asset return, *Cowles Found. Discuss. Pap.*, *1164*, 31, new Haven, Ct.
- Margulis, S. A., and D. Entekhabi (2003), Variational Assimilation of Radiometric Surface Temperature and Reference-Level Micrometeorology into a Model of the Atmospheric Boundary Layer and Land Surface, *Mon. Wea. Rev.*, *131*(7), 1272–1288, doi:10.1175/1520-0493(2003)131<1272:VAORST>2.0.CO;2.
- Martín, J., and C. Pérez (2009), Bayesian analysis of a generalized lognormal distribution, *Comput. Stat. Data Anal.*, *53*(4), 1377–1387, doi:http://dx.doi.org/10.1016/j.csda.2008.11.029.
- Masunaga, H., and C. Kummerow (2005), Combined radar and radiometer analysis of precipitation profiles for a parametric retrieval algorithm, *J. Atmos. Oceanic Technol.*, *22*(7), 909–929.
- Menabde, M., D. Harris, A. Seed, G. Austin, and D. Stow (1997), Multiscaling properties of rainfall and bounded random cascades, *Water Resour. Res.*, *33*, 12.
- Merlin, O., A. Chehbouni, Y. Kerr, and D. Goodrich (2006), A downscaling method for distributing surface soil moisture within a microwave pixel: Application to the Monsoon’90 data, *Remote Sens. Environ.*, *101*(3), 379–389.
- Milly, P. C. D. (1982), Moisture and heat transport in hysteretic, inhomogeneous porous media: A matric head-based formulation and a numerical model, *Water Resour. Res.*, *18*(3), 489–498.
- Montgomery, D. R., G. Balco, and S. D. Willett (2001), Climate, tectonics and the morphology of the Andes, *Geology*, doi:10.1130/0091-7613(2001)029<0579:ctatmo>2.0.co;2.
- Moradkhani, H., K.-L. Hsu, H. Gupta, and S. Sorooshian (2005), Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, *Water Resour. Res.*, *41*(5), W05,012–.
- 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, *Lecture Notes in Statist.*, 103, 281–299.
- Neumaier, A. (1998), Solving Ill-Conditioned and Singular Linear Systems: A Tutorial on Regularization, *SIAM Rev.*, 40(3), 636–666, doi:10.1137/S0036144597321909.
- Nichols, N. K. (2010), *Mathematical Concepts of Data Assimilation*, chap. Part 1, pp. 13–39, Data assimilation: Making sense of observations, Springer-Verlag Berlin Heidelberg.
- Nykanen, D., and D. Harris (2003), Orographic influences on the multiscale statistical properties of precipitation, *J. Geophys. Res.*, 108(D8), doi:10.1029/2001JD001518.
- Nykanen, D. K., E. Foufoula-Georgiou, and W. M. Lapenta (2001), Impact of small-scale rainfall variability on larger-scale spatial organization of land-atmosphere fluxes, *J. Hydrometeor.*, 2(2), 105–121.
- Over, T. M., and V. K. Gupta (1994), Statistical Analysis of Mesoscale Rainfall: Dependence of a Random Cascade Generator on Large-Scale Forcing, *J. Appl. Meteor.*, 33(12), 1526–1542, doi:10.1175/1520-0450(1994)033<1526:SAOMRD>2.0.CO;2.
- Parrish, D. F., and J. C. Derber (1992), The National Meteorological Center’s Spectral Statistical-Interpolation Analysis System, *Mon. Wea. Rev.*, 120(8), 1747–1763, doi:10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2.
- Perica, S., and E. Foufoula-Georgiou (1996a), Linkage of scaling and thermodynamic parameters of rainfall: Results from midlatitude mesoscale convective systems, *J. Geophys. Res.*, 101(D3), 7431–7448, doi:10.1029/95JD02372.
- Perica, S., and E. Foufoula-Georgiou (1996b), Model for multiscale disaggregation of spatial rainfall based on coupling meteorological and scaling, *J. Geophys. Res.*, 101(D21), 26–347.
- Peters-Lidard, C. D., M. S. Zion, and E. F. Wood (1997), A soil-vegetation-atmosphere transfer scheme for modeling spatially variable water and energy balance processes, *J. Geophys. Res.*, 102(D4), 4303–4324, doi:10.1029/96JD02948.
- Pontrelli, M. D., G. Bryan, and J. M. Fritsch (1999), The Madison County, Virginia, Flash Flood of 27 June 1995, *Wea. Forecasting*, 14(3), 384–404, doi:10.1175/1520-0434(1999)014<0384:TMCVFF>2.0.CO;2.
- Portilla, J., V. Strela, M. Wainwright, and E. Simoncelli (2001), Adaptive Wiener denoising using a Gaussian scale mixture model in the wavelet domain, in *Proc 8th IEEE Int’l Conf on Image Proc.*, vol. 2, pp. 37–40 vol.2, doi:10.1109/ICIP.2001.958418.

- Portilla, J., V. Strela, M. Wainwright, and E. Simoncelli (2003), Image denoising using scale mixtures of Gaussians in the wavelet domain, *IEEE Trans. Image. Process.*, *12*(11), 1338–1351, doi:10.1109/TIP.2003.818640.
- Rabier, F., H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons (2000), The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics, *Quart. J. Roy. Meteor. Soc.*, *126*(564), 1143–1170.
- Rao, K., and P. Yip (1990), *Discrete Cosine Transform: Algorithms, Advantages, Applications*, Boston: Academic Press.
- Rasmussen, C., and C. Williams (2006), *Gaussian processes for machine learning*, vol. 1, MIT press Cambridge, MA.
- Rauch, H., F. Tung, and C. Strieble (1965), Maximum likelihood estimates of linear dynamic systems, *AIAA J.*, *3*(8), 1445–1450.
- Rawlins, F., S. P. Ballard, K. J. Bovis, A. M. Clayton, D. Li, G. W. Inverarity, A. C. Lorenc, and T. J. Payne (2007), The Met Office global four-dimensional variational data assimilation scheme, *Quart. J. Roy. Meteor. Soc.*, *133*(623), 347–362, doi:10.1002/qj.32.
- Rebora, N., L. Ferraris, J. Von Hardenberg, A. Provenzale, et al. (2005), Stochastic downscaling of LAM predictions: an example in the Mediterranean area, *Advances in Geosciences*, *2*, 181–185.
- Rebora, N., L. Ferraris, J. Von Hardenberg, A. Provenzale, et al. (2006), Rainfall downscaling and flood forecasting: a case study in the Mediterranean area, *Nat. Hazard. Earth. Sys.*, *6*(4), 611–619.
- Reichle, R., D. Entekhabi, and D. McLaughlin (2001a), Downscaling of radio brightness measurements for soil moisture estimation: A four-dimensional variational data assimilation approach, *Water Resour. Res.*, *37*(9), 2353–2364.
- Reichle, R., D. McLaughlin, and D. Entekhabi (2001b), Variational data assimilation of microwave radio brightness observations for land surface hydrology applications, *IEEE Trans. Geosci. Remote.*, *39*(8), 1708–1718, doi:10.1109/36.942549.
- Roe, G. H. (2005), Orographic Precipitation, *Annu. Rev. Earth Planet. Sci.*, *33*, 645–671, doi:10.1146/annurev.earth.33.092203.122541.
- Ruderman, D. L. (1994), Statistics of natural images, *Network Comput. Neural Syst.*, *5*, 517–548, doi:10.1088/0954-898X/5/4/006.
- Sasaki, Y. (1958), An objective analysis based on variational method, *J. Meteor. Soc. Japan*, *36*, 77–88.

- Sasaki, Y. (1970), Some basic formalisms in numerical variational analysis, *Mon. Weather Rev.*, *98*(12), 875–883.
- Schultz, R., and R. Stevenson (1994), A Bayesian approach to image expansion for improved definition, *IEEE Trans. Image. Process.*, *3*(3), 233–242.
- Schuermans, J., and M. Bierkens (2007), Effect of spatial distribution of daily rainfall on interior catchment response of a distributed hydrological model, *Hydrol. Earth Syst. Sci.*, *11*, 677–693.
- Serafini, T., G. Zanghirati, and L. Zanni (2005), Gradient projection methods for quadratic programs and applications in training support vector machines, *Optim. Methods Softw.*, *20*(2-3), 353–378, doi:10.1080/10556780512331318182.
- Siccardi, F., G. Boni, L. Ferraris, and R. Rudari (2005), A hydrometeorological approach for probabilistic flood forecast, *J. Geophys. Res.*, *110*(D5), D05,101.
- Simoncelli, E., and W. Freeman (1995), The steerable pyramid: a flexible architecture for multi-scale derivative computation, in *International Conference on Image Processing, 1995. Proceedings.*, vol. 3, pp. 444–447 vol.3, doi:10.1109/ICIP.1995.537667.
- Smith, J. A., M. L. Baeck, M. Steiner, and A. J. Miller (1996), Catastrophic rainfall from an upslope thunderstorm in the central Appalachians: The Rapidan Storm of June 27, 1995, *Water Resources Research*, *32*(10), 3099–3113, doi:10.1029/96WR02107.
- Smith, K. S., and J. Marshall (2009), Evidence for Enhanced Eddy Mixing at Middepth in the Southern Ocean, *J. Phys. Oceanogr.*, *39*(1), 50–69, doi:10.1175/2008JPO3880.1.
- Smith, M. B., V. I. Koren, Z. Zhang, S. M. Reed, J. J. Pan, and F. Moreda (2004), Runoff response to spatial variability in precipitation: an analysis of observed data, *J. Hydrol.*, *298*(1-4), 267–286.
- Sorooshian, S., K. Hsu, G. Xiaogang, H. Gupta, B. Imam, and D. Braithwaite (2000), Evaluation of PERSIANN system satellite-based estimates of tropical rainfall, *Bull. Amer. Meteor. Soc.*, *81*(9), 2035–2046.
- Stein, M. L. (1999), *Interpolation of Spatial Data*, Springer-Verlag, New York.
- Strela, V., J. Portilla, and E. Simoncelli (2000), Image denoising using a local Gaussian scale mixture model in the wavelet domain, in *paper presented at SPIE Conference on Wavelet Applications in Signal and Image Processing, Int. Soc. for Opt. Eng., San Diego, Calif.*
- Subbotin, M. (1923), On the law of frequency errors, *Math. Sb*, *31*, 296–301.

- Talagrand, O., and P. Courtier (1987), Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory, *Quart. J. Roy. Meteor. Soc.*, *113*(478), 1311–1328.
- Tewfik, A., and M. Kim (1992), Correlation structure of the discrete wavelet coefficients of fractional Brownian motion, *IEEE Trans. Inform. Theory.*, *38*(2), 904–909, doi:10.1109/18.119750.
- Tibshirani, R. (1996), Regression Shrinkage and Selection via the Lasso, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, *58*(1), 267–288.
- Tikhonov, A., V. Arsenin, and F. John (1977), *Solutions of ill-posed problems*, Winston & Sons. Washington, DC.
- Tustison, B., E. Foufoula-Georgiou, and D. Harris (2003), Scale-recursive estimation for multisensor Quantitative Precipitation Forecast verification: A preliminary assessment, *J. Geophys. Res.*, *107*, 8377.
- 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*(D8), D08,104.
- van Leeuwen, P. J. (2010), Nonlinear data assimilation in geosciences: an extremely efficient particle filter, *Q.J.R. Meteorol. Soc.*, *136*(653), 1991–1999.
- Veneziano, D., R. L. Bras, and J. D. Niemann (1996), Nonlinearity and self-similarity of rainfall in time and a stochastic model, *J. Geophys. Res.*, *101*(D21), 26,371–26,392.
- Vivoni, E., D. Entekhabi, R. Bras, and V. Ivanov (2007), Controls on runoff generation and scale-dependence in a distributed hydrologic model, *Controls on runoff generation and scale-dependence in a distributed hydrologic model*, *11*(5), 1683–1701.
- Wahba, G., and J. Wendelberger (1980), Some New Mathematical Methods for Variational Objective Analysis Using Splines and Cross Validation, *Mon. Wea. Rev.*, *108*(8), 1122–1143, doi:10.1175/1520-0493(1980)108<1122:SNMMFV>2.0.CO;2.
- Wainwright, M. J., and E. P. Simoncelli (2000), Scale mixtures of Gaussians and the statistics of natural images, edited, *Advances in Neural Information Processing Systems*, *12*, 855–861, edited by S. A. Solla, T. K. Leen, and K.-R. Muller, , MIT Press, Cambridge, Mass.
- 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. Harmon. Anal.*, *11*(1), 89 – 123, doi:10.1006/acha.2000.0350.

- Wang, J., and D. B. Wolff (2009), Evaluation of TRMM ground-validation radar-rain errors using rain gauge measurement, *J. Appl. Meteorol. Climatol.*, *49*(2), 310–324, doi:10.1175/2009JAMC2264.
- Wang, S., X. Liang, and Z. Nan (2011), How much improvement can precipitation data fusion achieve with a multiscale Kalman Smoother-based framework?, *Water Resour. Res.*, *47*(1), W00H12,, doi:doi:10.1029/2010WR009953.
- Wang, Z., A. Bovik, H. Sheikh, and E. Simoncelli (2004), Image quality assessment: From error visibility to structural similarity, *IEEE Trans. Image. Process.*, *13*(4), 600–612.
- Wieczorek, G. F., G. S. Mossa, and B. A. Morgan (2004), Regional debris-flow distribution and preliminary risk assessment from severe-storm events in the Appalachian Blue Ridge Province, *Tech. rep.*, Springer, New York.
- Wilby, R., H. Hassan, and K. Hanaki (1998a), Statistical downscaling of hydrometeorological variables using general circulation model output, *J. Hydrol.*, *205*(1), 1–19.
- Wilby, R., T. Wigley, D. Conway, P. Jones, B. J. M. Hewitson, and D. S. Wilks (1998b), Statistical downscaling of general circulation model output: A comparison of methods, *Water Resour. Res.*, *34*(11), 2995–3008, doi:doi:10.1029/98WR02577.
- Woods, R., and M. Sivapalan (1999), A synthesis of space-time variability in storm response: Rainfall, runoff generation, and routing, *Water Resour. Res.*, *35*(8), 2469–2485, doi: 10.1029/1999WR900014.
- Wornell, G. W., and A. Oppenheim (1992), Estimation of fractal signals from noisy measurements using wavelets, *IEEE Trans. Signal. Proces.*, *40*(3), 611–623, doi:10.1109/78.120804.
- Yang, J., J. Wright, T. Huang, and Y. Ma (2010), Image super-resolution via sparse representation, *IEEE Trans. Image. Process.*, *19*(11), 2861–2873.
- Younger, P. M., J. E. Freer, and K. J. Beven (2009), Detecting the effects of spatial variability of rainfall on hydrological modeling within an uncertainty analysis framework, *Hydrol. Process.*, *23*(14), 1988–2003.
- Zeyde, R., M. Elad, and M. Protter (2010), On single image scale-up using sparse representations, *paper presented at 7th International Conference on Curves and Surfaces, SMAI-AFA, Avignon, France.*
- Zhou, Y., D. McLaughlin, and D. Entekhabi (2006), Assessing the performance of the ensemble Kalman filter for land surface data assimilation, *Mon. Weather Rev.*, *134*(8), 2128–2142.

Zupanski, D., S. Q. Zhang, M. Zupanski, A. Y. Hou, and S. H. Cheung (2010), A Prototype WRF-Based Ensemble Data Assimilation System for Dynamically Downscaling Satellite Precipitation Observations, *J. Hydrometeor.*, *12*(1), 118–134, doi:10.1175/2010JHM1271.1.

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

Appendix A

Appendix A. Details of Some Derivations in Chapter 4

A.1 Maximum a Posteriori Estimate of the Log-normal Multiplier

Recall that the Bayesian Maximum a Posteriori (MAP) estimator of z is defined as:

$$\hat{z}_{MAP} = \arg \max_z \{ \log [p_{Z|\mathbf{Y}}(z|\mathbf{y})] \} = \arg \max_z \{ \log [p_{\mathbf{Y}|Z}(\mathbf{y}|z)] + \log [p_Z(z)] \} \quad (\text{A.1})$$

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 $p_{\mathbf{Y}|Z}(\mathbf{y}|z)$ is Gaussian with covariance $\Sigma_{\mathbf{y}|z} = z\Sigma_{\mathbf{u}} + \Sigma_v$, first let us focus on the term $\log [p_{\mathbf{Y}|Z}(\mathbf{y}|z)]$ which leads to the derivation of the ML estimator; and then by incorporating the a priori term, the MAP estimator in (4.17) will be derived.

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

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

Therefore, the log-likelihood function can be written as:

$$\log [p_{\mathbf{Y}|Z}(\mathbf{y}|z)] = -\frac{1}{2} \mathbf{y}^T (\Sigma_{\mathbf{y}|z})^{-1} \mathbf{y} - \frac{1}{2} \log [\det |\Sigma_{\mathbf{y}|z}|] + C \quad (\text{A.3})$$

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

$$\begin{aligned}
\Sigma_{\mathbf{y}|z} &= z\Sigma_{\mathbf{u}} + \Sigma_v \\
&= z\Sigma_{\mathbf{u}} + SS^T \\
&= S [S^{-1} (z\Sigma_{\mathbf{u}}) S^{-T} + I] S^T
\end{aligned} \tag{A.4}$$

where S is the square root of $\Sigma_v = SS^T$ which can be computed using Cholesky or eigenvalue decomposition. Noting that $S^{-1}\Sigma_{\mathbf{u}}S^{-T}$ is also a positive semi-definite matrix, which can be diagonalized by an eigenvalue decomposition (i.e., spectral factorization) as $S^{-1}\Sigma_{\mathbf{u}}S^{-T} = Q\Lambda Q^T$, where $\{Q, \Lambda\}$ are matrices containing orthogonal eigenvectors $QQ^T = I$ and positive eigenvalues $\lambda_n \in \Lambda$, respectively. Therefore, diagonalization in Eq.(A.4) can be written as:

$$\Sigma_{\mathbf{y}|z} = SQ(z\Lambda + I)Q^T S^T \tag{A.5}$$

Using this diagonalized version of the covariance matrix, equation (A.3) can be further expanded as follows:

$$\begin{aligned}
\log [p_{\mathbf{Y}|Z}(\mathbf{y}|z)] &= -\frac{1}{2}\mathbf{y}^T (SQ(z\Lambda + I)Q^T S^T)^{-1} \mathbf{y} - \frac{1}{2} \log [\det |SQ(z\Lambda + I)Q^T S^T|] + C \\
&= -\frac{1}{2} (Q^T S^{-1}\mathbf{y})^T (z\Lambda + I)^{-1} (Q^T S^{-1}\mathbf{y}) - \frac{1}{2} \log [\det |SQ(z\Lambda + I)Q^T S^T|] + C \\
&= -\frac{1}{2}\mathcal{V}^T (z\Lambda + I)^{-1}\mathcal{V} - \frac{1}{2} \log [\det |z\Lambda + I|] + C'
\end{aligned} \tag{A.6}$$

where the vector $\mathcal{V} = Q^T S^{-1}\mathbf{y}$. 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 (A.6) with respect to z , we have:

$$\begin{aligned}
\frac{\partial \log [p_{\mathbf{Y}|Z}(\mathbf{y}|z)]}{\partial z} &= \frac{1}{2}\mathcal{V}^T \left(\Lambda (z\Lambda + I)^{-2} \right) \mathcal{V} - \frac{1}{2} \sum_{n=1}^N \frac{\lambda_n}{z\lambda_n + 1} \\
&= \frac{1}{2} \sum_{n=1}^N \frac{\lambda_n v_n^2}{(z\lambda_n + 1)^2} - \frac{1}{2} \sum_{n=1}^N \frac{\lambda_n}{z\lambda_n + 1} \\
&= \frac{1}{2} \sum_{n=1}^N \frac{\lambda_n^{-1}(v_n^2 - 1) - z}{(z + \lambda_n^{-1})^2}
\end{aligned} \tag{A.7}$$

Note that assuming a non-informative density for the multiplier, setting expression in (A.7) equal to zero, the root gives the maximum likelihood estimator of z .

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

$$\log [p_Z(z)] = -\frac{(\log z - \mu_z)^2}{2\sigma_z^2} - \log(z) + C \quad (\text{A.8})$$

$$\frac{\partial \log [p_Z(z)]}{\partial z} = \frac{-\log z + \mu_z - \sigma_z^2}{z\sigma_z^2} \quad (\text{A.9})$$

Then combining (A.7) and (A.9), leads to the derivation in (4.17) :

$$\frac{\log \hat{z}_{MAP} + \frac{3}{2}\sigma_z^2}{\hat{z}_{MAP}\sigma_z^2} + \frac{1}{2} \sum_{n=1}^N \frac{\hat{z}_{MAP} - \lambda_n^{-1}(v_n^2 - 1)}{(\hat{z}_{MAP} + \lambda_n^{-1})^2} = 0 \quad (\text{A.10})$$

From practical point of view, the $\log \hat{z}_{MAP}$ term in (A.10) 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 in (A.10) can take opposite signs on the boundaries.

A.2 Parameters of the MAP estimator

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

$$\mathbb{E}[\mathbf{y}^4] = \mathbb{E}[z^2] \mathbb{E}[\mathbf{u}^4] + 6\mathbb{E}[z] \mathbb{E}[\mathbf{u}^2] \mathbb{E}[v^2] + \mathbb{E}[v^4] \quad (\text{A.11})$$

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

$$\mathbb{E}[z^2] = \frac{\mathbb{E}[\mathbf{y}^4]/3 - 2\sigma_u^2\sigma_v^2 - \sigma_v^4}{\sigma_u^4} \quad (\text{A.12})$$

As $\sigma_u^2 = \sigma_y^2 - \sigma_v^2$, expression (A.12) can be updated to:

$$\mathbb{E}[z^2] = \frac{\mathbb{E}[\mathbf{y}^4]/3 - 2\sigma_y^2\sigma_v^2 + \sigma_v^4}{(\sigma_y^2 - \sigma_v^2)^2} \quad (\text{A.13})$$

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 :

$$\mathbb{E}[z^2] = \exp(\sigma_z^2) \quad (\text{A.14})$$

which assuming $\sigma_v^2 = 0$ leads to equation (4.18).

Appendix B

Appendix B. Statistical Interpretations and Solution Methods

B.1 Statistical Interpretations

B.1.1 Regularized Variational Downscaling and Data Fusion Problem

From the *frequentist* statistical point of view, it is easy to show that the WLS solution of (6.3) is equivalent to the maximum likelihood estimator (ML)

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

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

$$\begin{aligned} \hat{\mathbf{x}}_{ML} &= \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \frac{1}{2}(\mathbf{y} - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}) \right\} \\ &= \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\mathbf{R}^{-1}}^2 \right\}, \end{aligned}$$

which is identical to the WLS solution of problem (6.3).

It is important to note that in the ML estimator, \mathbf{x} is considered to be a deterministic variable (fixed) while \mathbf{y} has a random nature.

On the other hand, in the *Bayesian* perspective, the regularized solution of equation (6.4) is equivalent to the maximum a posteriori (MAP) estimator

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

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

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

The first term, $-\log p(\mathbf{y}|\mathbf{x})$, is just the negative log-likelihood as appeared in the ML estimator and the second term is called the *prior* which accounts for the a priori knowledge about the density of the state vector \mathbf{x} . Accordingly, the proposed Tikhonov regularization in (6.4) is equivalent to the MAP estimator assuming that the state, or the linearly transformed state $\mathbf{L}\mathbf{x}$, can be explained by a multivariate Gaussian of the following form:

$$\log p(\mathbf{x}) \propto \mathbf{x}^T \mathbf{Q} \mathbf{x},$$

where the covariance is $\mathbf{Q} = \mathbf{L}^T \mathbf{L}$ (Tikhonov et al., 1977; Elad and Feuer, 1997; Levy, 2008).

Clearly, the choice of the ℓ_1 -norm in equation (6.6), implies that $\log p(\mathbf{x}) \propto \|\mathbf{L}\mathbf{x}\|_1$ or say the transformed state can be well explained by a multivariate *Laplace* density with heavier tail than the Gaussian case (e.g., Tibshirani, 1996; Lewicki and Sejnowski, 2000). As previously explained, the Huber-norm regularization can also be interpreted as the *Gibbs* prior probability model $\log p(\mathbf{x}) \propto \sum_i \rho_T(x_i)$ for the state of interest (Geman and Geman, 1984; Schultz and Stevenson, 1994).

Obviously, based on the selected type of regularization, statistical interpretation of the DF regularized class of problems is also similar to what was explained for the DS problem. In other words, given the augmented observation model in (6.25), it is easy to see that the solution of (6.10) is the ML estimator while (6.14) can be interpreted as the MAP estimator with a prior density depending on the form of the regularization term.

B.1.2 Regularized Variational Data Assimilation

Statistical interpretation of the classic variational DA problems is a bit tricky compared to the DS and DF class of problems, mainly because of the involvement of the background information in the cost function. Lorenc (1986) derived the 3D-VAR cost function using Bayes theorem and called it the ML estimator (see, e.g., Lorenc, 1988; Bouttier and Courtier, 2002). More recently, it has been argued that the 4D-VAR, and thus as a special case the 3D-VAR cost function, can be interpreted via the Bayesian MAP estimator (Johnson et al., 2005b; Freitag et al., 2010; Nichols, 2010). For notational convenience, here we

only explain the statistical interpretation of the 3D-VAR and its regularized version which can be easily generalized for the case of the 4D-VAR problem.

As discussed earlier, the ML estimator is basically a frequentist view to estimate the most likely value of an unknown deterministic variable from (indirect) observations with random nature. The ML estimator intuitively requires to find the state that maximizes the likelihood function as $\hat{\mathbf{x}}_{ML} = \underset{\mathbf{x}}{\operatorname{argmax}} p(\mathbf{y}|\mathbf{x})$. Let us assume that, at the initial time step t_0 , the background \mathbf{x}_0^b is just a (random) realization of the true deterministic initial state \mathbf{x}_0 . In other words, we consider $\mathbf{x}_0^b = \mathbf{x}_0 + \mathbf{w}$, where the error \mathbf{w} can be well explained by a zero mean Gaussian density $\mathcal{N}(0, \mathbf{B})$, uncorrelated with the observation error, $\mathbb{E}[\mathbf{w}\mathbf{v}^T] = 0$. Here, the background state is treated similar to an observation with random nature. Thus, let us recast the problem of obtaining the analysis as a classic linear inverse problem by augmenting the available information in the form of $\underline{\mathbf{y}} = \underline{\mathbf{H}}\mathbf{x}_0 + \underline{\mathbf{v}}$, where $\underline{\mathbf{y}} = [(\mathbf{x}_0^b)^T, \mathbf{y}_0^T]^T$, $\underline{\mathbf{H}} = [\mathbf{I}, \mathbf{H}^T]^T$, and $\underline{\mathbf{v}} \sim \mathcal{N}(0, \underline{\mathbf{R}})$ with the following block diagonal covariance matrix

$$\underline{\mathbf{R}} = \begin{bmatrix} \mathbf{B} & 0 \\ 0 & \mathbf{R} \end{bmatrix}.$$

Notice that $\underline{\mathbf{R}}$ is block diagonal because the background and observation errors are uncorrelated. Following the augmented representation and applying $-\log(\cdot)$, we have $-\log p(\underline{\mathbf{y}}|\mathbf{x}_0) \propto 1/2(\underline{\mathbf{y}} - \underline{\mathbf{H}}\mathbf{x}_0)^T \underline{\mathbf{R}}^{-1}(\underline{\mathbf{y}} - \underline{\mathbf{H}}\mathbf{x}_0)$ and thus it is easy to see that the ML estimator in terms of the augmented observations, $\mathbf{x}_0^a = \underset{\mathbf{x}_0}{\operatorname{argmax}} p(\underline{\mathbf{y}}|\mathbf{x}_0)$, is equivalent to minimizing the 3D-VAR cost function in (6.29). Therefore, following this statistical interpretation, the classic 3D-VAR, can be derived via the frequentist ML estimator.

On the other hand, from the Bayesian perspective, the state of interest and the available observations are considered to be random and the MAP estimator is the optimal point which maximizes the posterior density as $\hat{\mathbf{x}}_{MAP} = \underset{\mathbf{x}}{\operatorname{argmax}} p(\mathbf{x}|\mathbf{y})$. Let us assume a priori that the (random) state of interest has a Gaussian density with the mean \mathbf{x}_b and covariance \mathbf{B} , that is $p(\mathbf{x}_0) \sim \mathcal{N}(\mathbf{x}_0^b, \mathbf{B})$. More formally, this assumption implies that the deterministic background is the central (mean) forecast and is related to the random true state via $\mathbf{x}_0 = \mathbf{x}_0^b + \mathbf{w}$, where $\mathbf{w} \sim \mathcal{N}(0, \mathbf{B})$. Therefore, using Bayes theorem it immediately follows that the 3D-VAR is the MAP estimator, $\mathbf{x}_0^a = \underset{\mathbf{x}_0}{\operatorname{argmax}} p(\mathbf{x}_0|\mathbf{y})$, assumed Gaussian prior for the true state of interest.

In conclusion, the regularized 3D-VAR in (6.31) might be interpreted as the MAP estimator, $\mathbf{x}_0^a = \underset{\mathbf{x}_0}{\operatorname{argmax}} p(\mathbf{x}_0|\underline{\mathbf{y}})$, with the prior density, $p(\mathbf{x}_0) \propto \lambda \psi_{\mathbf{L}}(\mathbf{x}_0)$, when we follow the frequentist approach to interpret the classic 3D-VAR as the ML estimator. On the other hand, taking the MAP interpretation for the classic 3D-VAR, the regularized version might be understood as the MAP estimator which also accounts for an extra and independent prior on the distribution of the state under \mathbf{L} transformation.

B.2 Gradient Projection for Huber regularization

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

$$\mathcal{J}(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\mathbf{R}^{-1}}^2 + \lambda \|\mathbf{L}\mathbf{x}\|_{\text{Hub}} \quad (\text{B.1})$$

$$\nabla \mathcal{J}(\mathbf{x}) = \mathbf{H}^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x}) + \lambda \mathbf{L}^T \rho'_T(\mathbf{L}\mathbf{x}), \quad (\text{B.2})$$

where

$$\rho'_T(x) = \begin{cases} 2x & |x| \leq \tau \\ 2\tau \text{sign}(x), & |x| > \tau. \end{cases} \quad (\text{B.3})$$

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

$$\begin{aligned} \hat{\mathbf{x}} &= \text{argmin} \{ \mathcal{J}(\mathbf{x}) \} \\ \text{s.t. } \mathbf{x} &\succeq 0. \end{aligned} \quad (\text{B.4})$$

We have used one of the primitive gradient projection (GP) methods to solve the above constrained DS problem (see, *Bertsekas*, 1999, pp. 228), for which more detailed explanations can be found in Section 9.4 of this appendix.

B.3 Quadratic Programming form of the ℓ_1 -norm RVDA

To obtain the quadratic programming (QP) form presented in (6.27), we follow the general strategy proposed in the seminal work by *Chen et al.* (2001). To this end, let us expand the ℓ_1 -norm regularized variational data assimilation (ℓ_1 -RVDA) problem in (6.26) as follows:

$$\underset{\mathbf{x}_0}{\text{minimize}} \left\{ \frac{1}{2} \mathbf{x}_0^T (\mathbf{B}^{-1} + \underline{\mathbf{H}}^T \underline{\mathbf{R}}^{-1} \underline{\mathbf{H}}) \mathbf{x}_0 - \left(\mathbf{B}^{-1} \mathbf{x}_0^b + \underline{\mathbf{H}}^T \underline{\mathbf{R}}^{-1} \underline{\mathbf{y}} \right)^T \mathbf{x}_0 + \lambda \|\Phi \mathbf{x}_0\|_1 \right\}. \quad (\text{B.5})$$

Assuming $\mathbf{c}_0 = \Phi \mathbf{x}_0 \in \mathbb{R}^m$, then the above problem can be rewritten as,

$$\underset{\mathbf{z}_0}{\text{minimize}} \left\{ \frac{1}{2} \mathbf{c}_0^T \mathbf{Q} \mathbf{c}_0 + \mathbf{b}^T \mathbf{c}_0 + \lambda \|\mathbf{c}_0\|_1 \right\}, \quad (\text{B.6})$$

where, $\mathbf{Q} = \Phi^{-T} (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}) \Phi^{-1}$ and $\mathbf{b} = -\Phi^{-T} (\mathbf{B}^{-1} \mathbf{x}_0^b + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{y})$. Having $\mathbf{c}_0 = \mathbf{u}_0 - \mathbf{v}_0$, where $\mathbf{u}_0 = \max(\mathbf{c}_0, 0) \in \mathbb{R}^m$ and $\mathbf{v}_0 = \max(-\mathbf{c}_0, 0) \in \mathbb{R}^m$ encode the positive and negative components of \mathbf{c}_0 , problem (B.6) can be represented as follows:

$$\begin{aligned} \underset{\mathbf{x}_0}{\text{minimize}} \quad & \left\{ \frac{1}{2} (\mathbf{u}_0 - \mathbf{v}_0)^T \mathbf{Q} (\mathbf{u}_0 - \mathbf{v}_0) + \mathbf{b}^T (\mathbf{u}_0 - \mathbf{v}_0) + \lambda \mathbf{1}_m^T (\mathbf{u}_0 + \mathbf{v}_0) \right\} \\ \text{s.t.} \quad & \mathbf{u}_0 \succcurlyeq 0, \mathbf{v}_0 \succcurlyeq 0 \end{aligned} \quad (\text{B.7})$$

Stacking \mathbf{u}_0 and \mathbf{v}_0 in $\mathbf{w}_0 = [\mathbf{u}_0^T, \mathbf{v}_0^T]^T$, the more standard QP formulation of the problem is immediately followed as:

$$\begin{aligned} \underset{\mathbf{w}_0}{\text{minimize}} \quad & \left\{ \frac{1}{2} \mathbf{w}_0^T \begin{bmatrix} \mathbf{Q} & -\mathbf{Q} \\ -\mathbf{Q} & \mathbf{Q} \end{bmatrix} \mathbf{w}_0 + \left(\lambda \mathbf{1}_{2m} + \begin{bmatrix} \mathbf{b} \\ -\mathbf{b} \end{bmatrix} \right)^T \mathbf{w}_0 \right\} \\ \text{s.t.} \quad & \mathbf{w}_0 \succcurlyeq 0. \end{aligned} \quad (\text{B.8})$$

Obtaining $\hat{\mathbf{w}}_0 = [\hat{\mathbf{u}}_0^T, \hat{\mathbf{v}}_0^T]^T \in \mathbb{R}^{2m}$ as the solution of (B.8), one can easily recover $\hat{\mathbf{c}}_0 = \hat{\mathbf{u}}_0 - \hat{\mathbf{v}}_0$ and thus the initial state of interest $\hat{\mathbf{x}}_0 = \Phi^{-1} \hat{\mathbf{c}}_0$.

The dimension of the QP representation (B.8) is twice that of the original ℓ_1 -RVDA problem (B.5). However, using iterative first order gradient based methods, which are often the only practical option for large-scale data assimilation problems, it is easy to show that the effect of this dimensionality enlargement is minor on the overall cost of the problem. Because, one can easily see that obtaining the gradient of the cost function in (B.8) only requires to compute

$$\begin{bmatrix} \mathbf{Q} & -\mathbf{Q} \\ -\mathbf{Q} & \mathbf{Q} \end{bmatrix} \mathbf{w}_0 = \begin{bmatrix} \mathbf{Q} (\mathbf{u}_0 - \mathbf{v}_0) \\ -\mathbf{Q} (\mathbf{u}_0 - \mathbf{v}_0) \end{bmatrix},$$

which mainly requires matrix-vector multiplication in \mathbb{R}^m (see; e.g., *Figueiredo et al.*, 2007).

B.4 Upper Bound of the Regularization Parameter

Here to derive the upper bound for the regularization parameter in the ℓ_1 -RVDA problem, we follow a similar approach as suggested for example by *Kim et al.* (2007). Let us refer back to the problem (B.6) which is convex but not differentiable at the origin. Obviously, \mathbf{c}_0^a is a minimizer if and only if the cost function $\mathcal{J}_{RAD}(\mathbf{c}_0)$ in (B.6) is sub-differentiable at \mathbf{c}_0^a and thus

$$0 \in \partial \mathcal{J}_{RAD}(\mathbf{c}_0^a),$$

where, $\partial\mathcal{J}_{RAD}(\mathbf{c}_0^a)$ denotes the sub-differential set at the solution point or analysis coefficients in the selected basis. Given that

$$\partial\mathcal{J}_{RAD}(\mathbf{c}_0^a) = \mathbf{Q}\mathbf{c}_0^a + \mathbf{b} + \lambda\partial(\|\mathbf{c}_0^a\|_1),$$

we have

$$-\mathbf{Q}\mathbf{c}_0^a - \mathbf{b} \in \lambda\partial(\|\mathbf{c}_0^a\|_1).$$

and thus for $\mathbf{c}_0^a = \mathbf{0}_m$, $\mathbf{0}_m = [0, \dots, 0]^T \in \mathbb{R}^m$, one can obtain the following vector inequality

$$-\lambda\mathbf{1}_m \preceq -\mathbf{b} \preceq \lambda\mathbf{1}_m,$$

which implies that $\|\mathbf{b}\|_\infty \leq \lambda$. Therefore λ must be less than $\|\mathbf{b}\|_\infty$ to obtain nonzero analysis coefficients in problem (B.6) and thus (B.5).

B.5 Gradient Projection Method for Sparsity-promoting Data Assimilation

Gradient projection (GP) method is an efficient and convergent optimization method to solve convex optimization problems over convex sets (see, *Bertsekas, 1999*, pp. 228). This method is of particular interest, especially, when the constraints form a convex set \mathcal{C} with simple projection operator. The cost function $\mathcal{J}_{RAD}(\mathbf{w}_0)$ in (6.27) is a quadratic function that need to be minimized on non-negative orthant $\mathcal{C} = \{\mathbf{w}_0 \mid w_{0,i} \geq 0 \forall i = 1, \dots, 2m\}$ as follows:

$$\begin{aligned} \hat{\mathbf{w}}_0 &= \operatorname{argmin} \{\mathcal{J}_{RAD}(\mathbf{w}_0)\} \\ &\text{s.t. } \mathbf{w}_0 \succeq 0. \end{aligned} \tag{B.9}$$

For this particular problem, the GP method amounts obtaining the following fixed point:

$$\mathbf{w}_0^* = [\mathbf{w}_0^* - \beta\nabla\mathcal{J}_{RAD}(\mathbf{w}_0^*)]^+, \tag{B.10}$$

where β is a stepsize along the descent direction and for every element of \mathbf{w}_0

$$[w_0]^+ = \begin{cases} 0 & \text{if } w_0 \leq 0 \\ w_0 & \text{otherwise,} \end{cases} \tag{B.11}$$

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

$$\mathbf{w}_0^{k+1} = [\mathbf{w}_0^k - \beta^k\nabla\mathcal{J}_{RAD}(\mathbf{w}_0^k)]^+. \tag{B.12}$$

Thus, if the descent at step k is feasible, that is $\mathbf{w}_0^k - \beta^k \nabla \mathcal{J}_{RAD}(\mathbf{w}_0^k) \succeq 0$, the GP iteration becomes an ordinary unconstrained steepest descent method, otherwise the result is mapped back onto the feasible set by the projection operator in (B.11). In effect, the GP method finds iteratively the closest feasible point in the constraint set to the solution of the original unconstrained minimization.

In our study, the stepsize β^k was selected using the *Armijo rule*, or the so-called *backtracking line search*, that is a convergent and very effective stepsize rule. This stepsize rule depends on two constants $0 < \xi < 0.5$, $0 < \varsigma < 1$ and assumed to be $\beta^k = \varsigma^{m_k}$, where m_k is the smallest non-negative integer for which

$$\mathcal{J}_{RAD} \left(\mathbf{w}_0^k - \beta^k \nabla \mathcal{J}_{RAD}(\mathbf{w}_0^k) \right) \leq \mathcal{J}_{RAD}(\mathbf{w}_0^k) - \xi \beta^k \nabla \mathcal{J}_{RAD}(\mathbf{w}_0^k)^T \nabla \mathcal{J}_{RAD}(\mathbf{w}_0^k). \quad (\text{B.13})$$

In our experiments the backtracking parameters are set to $\xi = 0.2$ and $\varsigma = 0.5$ (see, *Boyd and Vandenberghe*, 2004, pp.464 for more explanation). In our coding, the iterations terminate if $\frac{\|\mathbf{w}_0^k - \mathbf{w}_0^{k-1}\|_2}{\|\mathbf{w}_0^{k-1}\|_2} \leq 10^{-5}$ or the number of iterations exceeds 100.