June 1995

LIDS-TH-2302

Research Supported By:

Air Force Office of Scientific Research
AFOSR F49620-95-1-0083

Army Research Office
ARO DAAL03-92-G-0115

Army Research Projects Agency
F49620-93-1-0604

Office of Naval Research
ONR N0014-91-J-1004

Draper Laboratories
DL-W-467133

National Science Foundation
NSF DMS-9316624

NSERC '67 Fellowship (from the Natural Sciences & Engineering Research Council of Canada)

Application of Multiscale Estimation to Large Scale Multidimensional Imaging and Remote Sensing Problems

Paul Werner Fieguth

June 1995

LIDS-TH-2302

Sponsor Acknowledgments

AFOSR F49620-95-1-0083

ARO DAAL03-92-G-0115

ARPA F49620-93-1-0604

ONR N0014-91-J-1004

DL-W-467133

NSF DMS-9316624

NSERC '67 Fellowship (from the Natural Sciences & Engineering Research Council of Canada)

Application of Multiscale Estimation to Large Scale Multidimensional Imaging and Remote Sensing Problems

Paul Werner Fieguth

This report is based on the unaltered thesis of Paul Werner Fieguth submitted to the Department of Electrical Engineering and Computer Science in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Massachusetts Institute of Technology in June 1995.

This research was conducted at the M.I.T. Laboratory for Information and Decision Systems with research support gratefully acknowledged by the above mentioned sponsor(s).

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02139, USA

Application of Multiscale Estimation to Large Scale Multidimensional Imaging and Remote Sensing Problems

by

Paul Werner Fieguth

B.Sc., University of Waterloo (1991)
S.M., Massachusetts Institute of Technology (1993)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1995

Author .....
Department of Electrical Engineering and Computer Science
May 23, 1995

Certified by .....
Alan S. Willsky
Professor of Electrical Engineering
Thesis Supervisor

Accepted by .....
Frederick R. Morgenthaler
Graduate Officer, Department of EECS

Application of Multiscale Estimation to Large Scale Multidimensional Imaging and Remote Sensing Problems

by

Paul Werner Fieguth

Submitted to the Department of Electrical Engineering and Computer Science
on May 23, 1995, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering

Abstract

A recently developed multiresolution estimation framework offers the possibility of highly efficient statistical analysis, interpolation, and smoothing of extremely large data sets in a multiscale fashion. This framework enjoys a number of advantages not shared by other statistically-based methods, particularly in terms of the ability to evaluate estimates and error variances in a computationally efficient manner, however there remain several barriers which constrain the widespread use of this framework:

- The multiscale framework has been characterized as well-suited for large-scale estimation problems such as in remote sensing, however no such scientific endeavors have been undertaken which might motivate the use of the framework among scientists.
- Past research efforts have developed a rich class of multiscale models; however given the selection of a particular multiscale model structure or class, the identification of unknown parameters within the model remains unclear.
- The estimates produced by the estimator typically possess artifacts introduced by the multiscale structure; these artifacts are distracting to the human eye, limiting the use of the framework in certain image processing applications.

This thesis directly addresses each of the above limitations:

- Two problems of current scientific interest are addressed: the estimation of the ocean surface height from satellite data, and the estimation of the earth's gravitational equipotential. Both lines of research produce results of potential interest to the scientific community.
- We demonstrate a technique for estimating multiscale parameters in simple models by developing an estimator for the fractal dimension of fractional Brownian motion processes. Furthermore, for a $1/f$-like class of multiscale models a Cramer-Rao bound can be determined for the maximum-likelihood estimation of model parameters.

- Significant improvements in estimate smoothness are achieved using a novel overlapping multiscale framework capable of reducing artifacts below the level of detectability with a modest computational burden. The performance of the overlapping framework is demonstrated in the context of the surface reconstruction problem of computer vision.

Thesis Supervisor: Alan S. Willsky
Title: Professor of Electrical Engineering

Acknowledgements

There are a great number of people who have contributed significantly to my Ph.D. years at M.I.T.

Most significant, of course, is the help and encouragement which I received over the past few years from Alan Willsky. I thank him for his enthusiasm and for his careful supervision of my research, from which I benefited tremendously over the course of my Ph.D., and also for his attitudes and insights regarding the supervisor–student mentoring process, which may very well be some of the most valuable lessons that I take with me from MIT. I also appreciate the considerable assistance from Carl Wunsch, particularly his help on the remote sensing aspects of this thesis, and the research advice and assistance in technical writing from Clem.

I have also greatly appreciated the technical and social interaction with past and current SSG members: the opportunity for technical collaboration with Bill, the valuable (but largely ignored) love-life advice from Seema and Rachel, the regular checkups on my happiness by Hamid, the mathematical distractions of Ilya, the "real-world" education provided by association with Mickey (in particular the wonders of Channel 35 and the Netherlands) and, of course, regular doses of abuse from Mike.

Surely the greatest time-sink and impediment to graduation during my years at MIT has been the social-life of Ashdown House (an oxymoron to some, perhaps ...). Four people deserve significant mention: Scott (for at least ten thousand conversations on the same topic, may hell soon freeze over), Bruce (for two winters of suffering, but also for two years of experiences to be recounted for a lifetime), Alain (for inspiring conversations on the fine art of vegetarianism), and John (fellow invader from the north, for three years of warm friendship and a remarkable tolerance for childish pranks).

Finally my parents deserve thanks for their remarkable tolerance in a son who, after 27 years, has yet to contribute meaningfully to society. Thankyou.

My desire to obtain a Ph.D. degree has a rather curious and poorly motivated origin.

Since the age of about five years I have been conspicuously aware of a red book[25], sitting on the shelf at home. It was looked at occasionally, always with awe, and with a sense that I should never be capable of producing such a work myself, but furthermore with the expectation that should I ever complete such a work, then surely my goals in life would have been fulfilled.

After twenty years this quest is finally over; the same five year old smiles and anticipates the arrival of a new red book[28] to join the first. Although this work surely fails to meet the standards set by its motivator, I hope that it may be judged acceptable.

Contents

List of Figures

List of Tables

Nomenclature

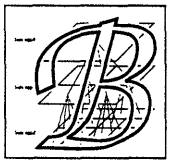
Symbol	Definition	Page Refs.
$\hat{\cdot}$	represents an estimate of $\cdot$	41 94 109
$\tilde{\cdot}$	represents an error of $\cdot$	42 51
$\|\| \cdot \|\|$	counts the number of elements in set $\cdot$	53 102 105
$\cdot_l$	indicates an overlapped quantity	154 154
0	root node of multiscale tree	49 252
f	frequency	32 59
h	combined geoid / ocean surface height	66
k	multiscale state vector dimension	58 155 182
m	tree scale index and function	31 49 60
o_m	degree of overlap between siblings on scale m	172 281
p	surface gradient along x direction	192 195
q	surface gradient along y direction	192 195
q	the order of a multiscale tree	49 171 251
r	lacunarity measure - scale to scale size ratio	205
s	multiscale node index	49 252
v	measurement noise	40 46 49
w	process noise	40 49
w_m	size of region aggregated by nodes on scale m	172 205 281
x	spatial axis	192 203
x	state vector	40 49 252

Symbol	Definition	Page Refs.
$\hat{x}$	state estimate	41 253
$\tilde{x}$	state estimation error	41 51 255
y	spatial axis	192 203
y	measurement vector	40 49 252
z	generic surface height	45 192
A	process dynamics matrix	41 49 252
B	process noise weighting matrix	49 252
C	measurement matrix	40 49 252
D	discrete fractional Gaussian noise process	95 98
E	expectation operator	40 57 60
F	fractional Brownian motion process	60 95 98
F	multiscale upwards model	253 259
G_x	state projection matrix into overlapping domain	154 161 174
G_y	measurement projection matrix into overlapping domain	154 165 174
H_x	estimate projection matrix from overlapping domain	154 162 174
H	fBm dimension parameter	60 95 100
I	identity matrix	41 46
J	degree and order of spherical harmonic series	120 122
L	generic linear operator	45 197
L	least squares estimation operator	42 163 170
M	total number of multiscale tree scales	49 252 282
N	geoid height	66 120 123
N	number of finest-scale tree nodes	51 155
P	probability function	55 104
P	covariance	40 40
$\tilde{P}$	estimation error covariance	42 46 51
P_o	prior covariance at multiscale root node	49 75
R	measurement error covariance	40 49 162
T	surface topography of the earth	117 119

Symbol	Definition	Page Refs.
W	number of multiscale trees to be averaged	149
Y	aggregated vector of measurements	45 53 193
$\mathcal{L}$	likelihood	53 94 257
$\mathcal{O}$	order operator: $g \in \mathcal{O}(f) \rightarrow g(x) \leq \alpha f(x) \ \forall \, x$	51 149 156
$\mathcal{O}$	overlapping tree parameterization vector	206 283
$\mathcal{M}$	set of tree nodes possessing measurements	49 94
$\mathcal{N}$	Gaussian random vector operator	46 94 258
$\mathcal{T}$	set of tree nodes	48 94
α	multiscale "child" operator	49 97 173
$\bar{\delta}$	multiscale "sibling" operator	251
$\bar{\gamma}$	multiscale "parent" operator	49 97 251
γ	scale to scale process noise ratio	101
μ	spectral power law	60 74 110
ϕ	spatially varying geoid error function	77 124 135
ψ	fractional overlap	173
ρ	correlation coefficient	108 111
σ	standard deviation	60 95 100
ζ	ocean surface elevation relative to the geoid	66 122 124
Λ	generic covariance matrix	53 57 102
Σ	actual realized covariance	154 159

Chapter 1

Introduction



roadly speaking, this thesis addresses the development of computationally efficient algorithms for solving statistical estimation problems; that is, estimation problems characterized by an explicit statistical prior model and by an explicit statistical measurement model. A wide variety of algorithms have been developed addressing such estimation problems, emphasizing varying degrees of statistical structure or computational efficiency. The classical tool to apply in such problems is the Kalman filter[4]; other examples (from the remote sensing community) include objective analysis[20] and kriging[91].

For the estimation problems of interest in this thesis, specifically *large* estimation problems in which estimation error statistics are required, the straightforward application of any of the above estimation methods fails to be practical for computational reasons. Indeed, the straightforward estimation approach implied by any of these methods, applied to a random field representing a square image, N pixels on a side, involves the inversion of a matrix having N^4 elements, requiring $\mathcal{O}(N^6)$ computations! Clearly if the estimation of images having $N > 256$ is to be commonplace a new approach is required.

The key to the approach to be used in this thesis is that we begin by focusing explicitly on scale. In particular, rather than starting with the statistical description of the phenomenon to be estimated at a single, fine scale of resolution – the typical Kalman filter implementation – we describe its statistical structure at a hierarchy

of scales. Such a multiscale approach is not new; signal processing has seen intense interest in multiscale / multiresolution methods, applied to a broad range of signal analysis, compression, and estimation problems. Much of this interest has been fueled by the following three motivations:

- Many natural and human systems exhibit features or behavior across a broad range of space or time scales (for example, the earth's oceans[27], critical phenomena in physics[112], or the distribution of galaxies in the universe).

- Many surface-reconstruction[44] and Markov random field problems[34, 35] use local models to characterize systems which possess long-range correlations. Iterative approaches to solving such problems on a single fine scale invariably lead to a kind of critical slowing down: the computational effort *per pixel* grows[1] with the size N of the random field. Multiresolution approaches[102], still using local models but now on a pyramidal hierarchy of scales, have been conspicuously effective in providing efficient solutions.

- The development of wavelet theory[68] has provided a powerful new rigorous framework for multiscale analysis.

Motivated by one or more of the above aspects, quite a variety of multiscale approaches have been developed. Among these multiscale approaches are methods such as multigrid[39, 74, 102] and hierarchical basis functions[100]. While such approaches have proven well-suited and computationally efficient for solving certain relaxation problems (in particular, elliptic PDE problems) these approaches have not, however, been particularly successful in attacking statistical estimation problems. Broadly speaking, there are two aspects of statistical estimation problems which are not adequately dealt with by most multiscale algorithms:

1. Most statistical estimation problems are characterized by an explicit prior statistical model, parameterized in terms of a number of random variables with

[1]The computational effort per pixel for solving Laplace's equation using Gauss-Jacobi/Seidel iterations grows as N^d for a field in d dimensions having N pixels on a side[38].

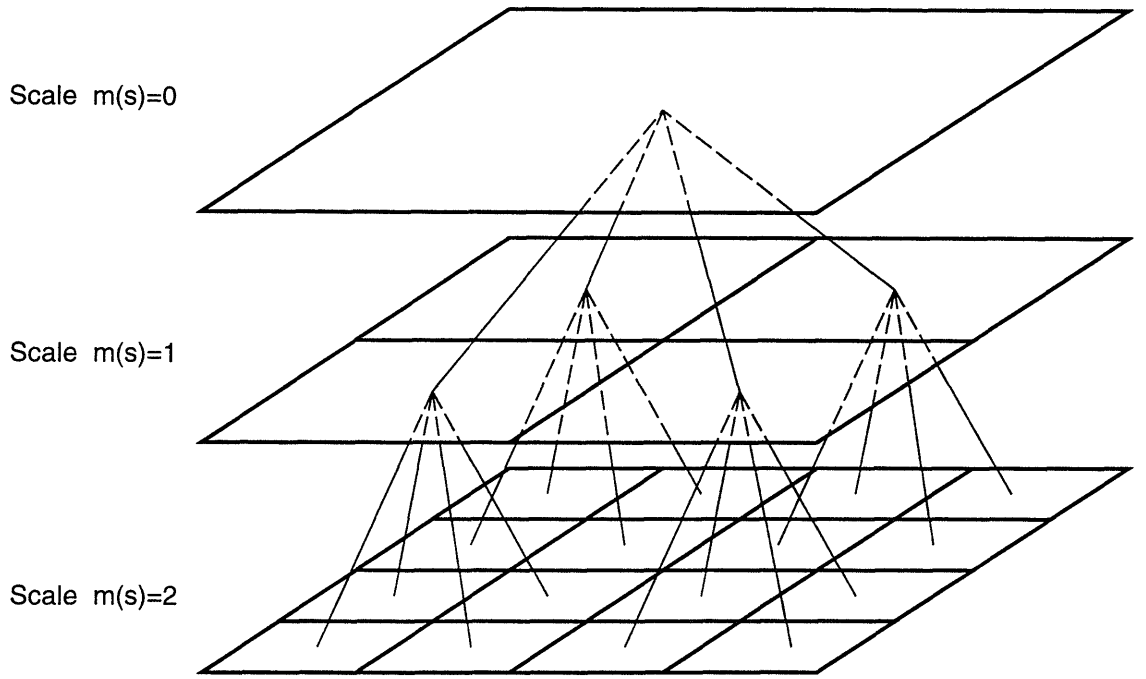


Figure 1-1: An example multiscale tree. The coarsest scale is shown at the top of the figure, with finer scales below. An explicit statistical relationship is specified between those tree nodes connected by thin lines. This figure shows only a relatively simple example; *any* tree (i.e., any acyclic graph) may actually be used.

specified probability distributions. The behavior of most multiscale algorithms, on the other hand, is parameterized by a set of variables which *implicitly* specify a prior model, but where the relation between these variables and the associated prior model is unclear and difficult to discern.

2. We are interested in those problems where estimation error statistics are required (e.g., in oceanographic remote sensing). The computation of these statistics for large problems is notoriously difficult.

   FFT methods are available which can accelerate such computations, but only under relatively strict conditions: the random process resides on a rectangular grid, is stationary, and is densely sampled with a stationary measurement noise covariance. The problems investigated in this thesis will not satisfy all of these conditions.

A recently-introduced class of multiscale stochastic models[14, 62] has begun to address both of the above two problems by developing a framework in which the models possess explicit statistical priors, and for which error statistics are not only computable, but indeed efficiently computable even for very large estimation problems. These multiscale stochastic models live on trees, such as the one shown in Figure 1-1. It should be emphasized that a great variety of tree structures are possible; the structure shown in the figure is just one possibility, although a rather common and convenient one for representing two-dimensional processes. This framework possesses further advantages: the production of estimates and estimation error variances on a hierarchy of scales (e.g., on each scale of the multiscale tree) facilitating resolution/accuracy tradeoffs leading to the direct extraction of estimates of coarser scale features, and the fusion of data of differing resolution with no change in algorithmic structure. It is this multiscale framework which forms the basis for this thesis.

This introduction should serve only to introduce the notion of multiscale estimation; a more thorough description of the motivation for multiscale approaches and a greater explanation of the multiscale estimation procedure may be found in Chapter 2. The next section will outline the contributions of this thesis, followed by a description of the thesis organization.

1.1 Thesis Contributions

This thesis continues the earlier research efforts of Chou[14] and Luettgen[62]. Specifically, to put the work of this thesis into context, at the point in time when this thesis research was begun the following research goals had been accomplished:

- The establishment and definition of the multiscale estimation framework.
- The development of the multiscale estimation algorithm and the multiscale likelihood calculation algorithm.
- The development of two classes of multiscale prior models: $1/f$ priors and Markov random field priors.

- The application of the multiscale framework to problems such as optical-flow estimation, texture synthesis, and texture discrimination.

The further contributions of this thesis, listed below, build upon these accomplishments.

1.1.1 Multiscale Model Identification

Given some random process with unknown or very complicated statistics, there are two basic steps in developing a model for the process:

1. the selection of the parameterized model (e.g., first-order Gauss-Markov, second-order Gauss-Markov, $1/f$ etc.),

2. the quantitative determination of the unknown parameters within the selected model.

The selection of a parameterized model is usually accomplished based on a physical or intuitive understanding of the random process in question; the determination of parameters forms the subject of system identification. Although these steps are still in their relative infancy with respect to multiscale models, we can report some progress.

Certain previous multiscale applications[63] determined multiscale model parameters by trial and error; with the development of the multiscale likelihood calculation algorithm[65] the maximum-likelihood estimation of such parameters is possible, at least in principle. A contribution of this thesis is the derivation of a Cramer-Rao bound for the maximum-likelihood estimation of the parameters for a particular class of multiscale models with $1/f$-like properties. The bound is compared with Monte-Carlo simulations and with tests on real remote-sensing data.

A further contribution is made in the estimation of the fractal dimension of fractional Brownian motion[69] processes. We present the development of a new $1/f$-like multiscale model which gives unbiased estimates of the fractal dimension, and compare its performance with other proposed estimators[49, 114].

1.1.2 Remote Sensing Applications

There are several aspects to estimation problems in remote sensing which make them well-suited to multiresolution methods:

1. The problems are typically very large and multidimensional, requiring the estimation of millions (or more) of random variables.

2. Since remote sensing measurements are typically made by satellite or by ship, the measurements are spatially irregularly sampled, precluding the use of FFT methods.

3. Many natural systems, the object of study in remote sensing, are characterized by $1/f$-like behavior, which is readily modeled using multiscale techniques.

4. Estimation error statistics are required in many applications to make proper use of the computed estimates; such error statistics are frequently very difficult or impossible to compute.

This thesis makes contributions to two remote sensing problems, oceanography and geodesy;[2] both of these problems are subject to the four characteristics just listed.

The oceanographic remote sensing problem involves the estimation of the height of the ocean surface, given measurements taken from a satellite in orbit. Empirical studies of the satellite data show the ocean surface to possess a $1/f$ like behavior, from which an appropriate multiscale model is developed. The contribution of this thesis to this oceanographic remote sensing problem is a demonstration to the remote sensing community of the applicability and efficiency of novel multiscale estimation techniques.

The geodetic remote sensing problem involves the estimation of the height of the geoid,[3] again using satellite data. A *joint* model, simultaneously estimating the ocean height and the geoid, is developed The contribution of this thesis to geodesy is more

[2]The study and determination of planetary gravitational fields.

[3]The geoid represents a surface of constant gravitational equipotential on the earth. In the absence of ocean currents and winds, the surface of the ocean would conform to the shape of the geoid.

scientific in nature: the models which we develop are strongly motivated by existing models in the geodetic literature, and the goal of the chapter is an improved set of geoid estimates – estimates of immediate use to researchers studying ocean altimetric data.

1.1.3 Multiscale Estimation and Smoothness

In the opinion of the author, one of the most significant stumbling blocks in the application of the multiresolution framework has been the presence of blocky artifacts in the resulting estimates. In many cases the quality or quantity of the measurements may not justify the production of smoother estimates[63], however there are at least two cases in which the production of smoother estimates may be required:

1. In certain applications (e.g., computer vision) the resulting estimates will be displayed and require smoothness for aesthetic reasons.

2. In other cases (e.g., oceanographic remote sensing) one may wish to compute gradients of the estimated field, in which case blocky discontinuities are unacceptable.

A significant contribution of this thesis is the development of a novel multiscale structure known as an overlapping tree, which is able produce smooth estimates, even with sparse measurements. This approach is not a specific multiscale model, rather it is a general technique which may be applied to a broad variety of multiscale models. We derive theoretical conditions for the applicability of the overlapping tree technique, and demonstrate its use in texture estimation.

1.1.4 Multiscale Implementation of Variational Priors

The final contribution of this thesis is to the problem of surface reconstruction in computer vision. Typically, surface reconstruction problems are formulated in terms of a variational cost function; the desired surface is the one which minimizes the cost function. The most common solution to this problem involves discretizing the partial

differential equation (PDE) resulting from the variational formulation, and solving the PDE by relaxation methods.

In this thesis we propose a novel alternative: rather than solving the variational equation directly we determine a multiscale model which has a similar behavior. Coupling this multiscale model with the overlapping technique just described results in an estimated surface which is computed quickly, for which estimation error statistics are available, and which is aesthetically pleasing.

1.2 Thesis Organization

Chapter 2 presents background material relevant to multiscale estimation. The chapter begins by presenting a broad overview of statistical modeling and optimal estimation, highlighting the computational difficulties that may be encountered in a straightforward implementation of optimal estimation techniques, and motivating alternative implementations such as the multiscale approach of this thesis. The discussion assumes relatively little in terms of a statistical background and was written to familiarize members of the remote sensing community with matters of optimal estimation. A second section parallels the development of the first, but explores an alternative manner of posing estimation problems using variational methods.

The next two sections discuss our multiscale estimation framework more specifically. We first introduce the multiscale tree and the basic nomenclature needed to talk about stochastic processes on such a tree. Next, we define the basic multiscale dynamic and measurement equations which characterize the class of multiscale models in our framework. Two subsections follow, containing high-level descriptions of the two most significant multiscale algorithms to be used in this thesis: the multiscale estimation algorithm and the multiscale likelihood calculation algorithm.

Next we present an overview of multiscale model development. This development falls into two broad classes: (i) relatively high order models motivated by Markov random field processes, and (ii) relatively low order models motivated by $1/f$ and simple variational prior models.

Finally we overview the specific computational implementation of the multiscale framework that was used for all of the examples in this thesis. Although such software considerations are not among the research subjects of this thesis, this particular implementation has been found to be conspicuously versatile and efficient.

Chapter 3 presents the first of the two remote sensing applications considered in this thesis: oceanography. We begin with a brief definition of terms and a discussion of the ocean elevation estimation problem and the reasons why this problem is of interest, both to researchers in signal processing and in remote sensing. We develop a $1/f$-like model for the ocean and show the results of applying this model to remote sensing measurements. Finally we demonstrate the versatility of our multiscale approach by developing a heterogeneous multiscale model to account for the effect of ocean currents or for the effect of space-varying geoid-errors.

Chapter 4 begins by presenting the problem of estimating the fractal dimension of a fractal Brownian motion process, beginning with the development of an appropriate model, followed by a demonstration of estimation results based on synthesized data sets. The second half of the chapter discusses the problem of system identification for multiscale models. We specify the $1/f$-like multiscale model class of interest possessing two free parameters, and determine a Cramer-Rao bound for the maximum likelihood estimation of these parameters. We present two examples comparing the Cramer-Rao bound to the actual performance of the estimator:

(i) based on synthesized measurements, and

(ii) based on the oceanographic measurements used in Chapter 3.

Chapter 5 presents the second remote sensing application: in this chapter we explore the joint estimation of the geoid and the ocean surface. The chapter begins by reviewing the various methods and measurements which have been used in the past for estimating the geoid. The next two sections seek to characterize the spatial variation of the geoid error by estimating high frequency terms of the geoid from topographical data, and by computing the distribution of statistically significant oceanographic residuals. Finally a joint model is posited and the estimated geoid is presented.

Chapter 6 presents the novel overlapping-tree technique. Although the overlapping-tree concept is not intrinsically complicated, a rigorous explanation of the concept is relatively difficult, and the reader is advised to read carefully and to study the (superficially simple) examples in some detail. The chapter begins with a discussion of smoothness, and presents several interpretations as to why the multiscale estimation approach may lead to blocky artifacts.

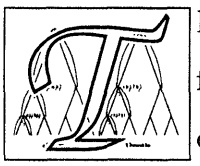
After some preliminaries, the discussion of the overlapping-tree technique begins with a simple example, which is meant to make some of the abstract overlapping-tree notions more concrete; the example is broken into two parts: multiscale modeling and multiscale estimation. Next we present a rigorous derivation of the conditions placed on the overlapping tree such that the resulting estimates equal the optimal least-squares estimates of interest. Next we discuss the manner in which a particular overlapping tree structure may be selected to be consistent with the conditions just derived. The chapter ends with three applications of the overlapping-tree technique to the estimation of a texture characterized by a Markov random field model.

Chapter 7 presents the work on multiscale approaches to surface reconstruction. The chapter begins by deriving the most common variational costs used for surface reconstruction based on the "membrane" and "thin-plate" models of classical physics[18]. The gradient integrability problem[31, 43] is discussed, and a variational equation which requires the simultaneous estimation of a surface and its gradients is presented. The next section of the chapter discusses the multiscale analog to each of the various components of the variational equation, which ultimately leads to a complete multiscale model appropriate for surface reconstruction. A wide variety of experimental results are shown based on problems with sparse or dense measurements and discontinuous surfaces; the results are compared with other relaxation methods (Gauss-Seidel, conjugate-gradient, and multigrid).

Chapter 8 summarizes the results of this thesis, presents the major contributions, and details a number of avenues for further research.

Chapter 2

Background



he purpose of this chapter is to introduce and to motivate the multiscale framework which is used throughout this thesis. Sections 2.1 and 2.2 discuss two alternative methods of formulating estimation problems, and discuss the computational difficulties which motivate an exploration of alternative approaches, such as our multiscale one. Section 2.3 presents the multiscale framework which we will use throughout this thesis. A subsection is dedicated to each of the two principle algorithms which have been developed for this framework: a multiscale estimation algorithm, and a multiscale whitening / likelihood calculation algorithm. Section 2.4 discusses the basic philosophy behind the development of multiscale models to solve estimation problems, and then presents simple multiscale analogs to three statistical prior models: Markov random field priors, $1/f$-like priors, and quadratic variational priors. Finally Section 2.5 outlines the manner in which the multiscale algorithms were implemented on a computer.

2.1 Statistical Models and Optimal Estimation

Consider the basic problem of estimating a collection of random variables, represented abstractly by the vector $\boldsymbol{x}$, based on a set of noise-corrupted measurements,

represented by $\boldsymbol{y}$:

$$\boldsymbol{y} = C\boldsymbol{x} + \boldsymbol{v} \qquad E[\boldsymbol{v}] = 0 \qquad E\left[\boldsymbol{v}\boldsymbol{x}^T\right] = 0 \qquad E\left[\boldsymbol{v}\boldsymbol{v}^T\right] = R \tag{2.1}$$

where $\boldsymbol{v}$ represents the measurement noise or error. In general, for the problems to be considered in this thesis the components of $\boldsymbol{v}$ are assumed to be uncorrelated but possibly with non-constant variances – i.e., R is diagonal but not a multiple of the identity. The matrix C describes the nature of the measurement process. Frequently the components of $\boldsymbol{x}$ represent a dense grid of unknowns, and C is a "selection matrix" indicating which of the components of $\boldsymbol{x}$ are measured and which x_i corresponds to each y_j. This is a convenient, but not necessary, arrangement; more general definitions of $\boldsymbol{x}$ and C are possible.

We can view our estimation problem as estimating the deviations of $\boldsymbol{x}$ from its mean, thus for simplicity, we assume that $\boldsymbol{x}$ is zero-mean and has prior covariance

$$E\left[\boldsymbol{x}\boldsymbol{x}^T\right] = P_x \tag{2.2}$$

For problems of substantial size, the explicit specification of the correlation structure of $\boldsymbol{x}$ through the full covariance matrix P_x is neither feasible nor useful unless P_x is extremely sparse with known structure – e.g., if P_x is banded, implying only local correlation among the components of $\boldsymbol{x}$. However, such sparse or banded structures are not particularly appropriate or useful for problems of interest here, as we are interested in representing phenomena possessing correlations at many (and not just local) scales. Furthermore, as we will see, banded or sparse covariance structures do not necessarily lead to simple algorithms for statistical data analysis.

Consequently we are led instead to construct an implicit model of the statistical structure of $\boldsymbol{x}$ of the form

$$M\boldsymbol{x} = \boldsymbol{w} \tag{2.3}$$

$$P_x^{-1} = M^T P_w^{-1} M \tag{2.4}$$

where P_w is the covariance of $\boldsymbol{w}$. There are several reasons why representations as in (2.3),(2.4) can be attractive. One is that processes with complex correlation structures can be represented in a very compact manner. For example consider the linear state space model

$$x(t+1) = Ax(t) + w(t) \qquad E\left[x(0)w^T(t)\right] = 0 \tag{2.5}$$

If we construct the vectors

$$\boldsymbol{x}^T = \left[x^T(0)\ x^T(1)\ x^T(2)\ \ldots\right] \qquad \boldsymbol{w}^T = \left[x^T(0)\ w^T(0)\ w^T(1)\ \ldots\right] \tag{2.6}$$

then we obtain a representation as in (2.3) with P_w block diagonal and M lower bidiagonal:

$$M = \begin{bmatrix} I & 0 & 0 & 0 & \ldots \\ -A & I & 0 & 0 & \ldots \\ 0 & -A & I & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{2.7}$$

As we now show, it is the *inverse* of P_x, which according to (2.4) involves only M and P_w, that is critical in constructing solutions to optimal estimation problems.

Specifically, the problem of interest here is the computation of the minimum variance linear estimate of $\boldsymbol{x}$ based on $\boldsymbol{y}$, as well as a statistical characterization of the error $\tilde{\boldsymbol{x}} = \boldsymbol{x} - \hat{\boldsymbol{x}}$. There are numerous ways in which to represent the solution to this problem, but the one that is most convenient for our discussion is that given by the normal equations for this least squares problem:

$$(P_x^{-1} + C^T R^{-1} C)\hat{\boldsymbol{x}} = C^T R^{-1} \boldsymbol{y} \tag{2.8}$$

This problem formulation and the normal equation solution are well known in many disciplines, however approximations or suboptimal solutions have generally been required in order to use (2.8) to estimate $\boldsymbol{x}$. Consider the formal explicit solution, $\hat{\boldsymbol{x}}$,

to (2.8) and the resulting error covariance $\tilde{P}_x$:

$$\hat{\boldsymbol{x}} = L\boldsymbol{y} = \tilde{P}_x C^T R^{-1} \boldsymbol{y} \tag{2.9}$$

$$\tilde{P}_x^{-1} = P_x^{-1} + C^T R^{-1} C \tag{2.10}$$

Note that if P_x has a sparse or banded structure, indicative of local correlations, this structure is *not* generally preserved either in the estimation gain matrix L (2.9) or in the estimator error covariance $\tilde{P}_x$. Thus simple, local, smoothing algorithms (e.g., local least squares, local interpolation) while efficient computationally, generally represent a suboptimal approximation to (2.9) even in situations in which they appear to be best matched, i.e., when the field to be interpolated has local correlations. Moreover, a very important point is that the statistical structure of the resulting estimation error field, $\tilde{P}_x$, is *not* local, despite locality in P_x. Furthermore the calculation of $\tilde{P}_x$ is generally prohibitively complex (since, in particular, the inversion of the prior covariance P_x is extremely demanding). Thus the use of simple local algorithms generally involves a compromise in statistical consistency, in the explicit and faithful use of prior statistical models and information, in the calculation of accurate error statistics, and in the ability to account for correlations at many scales.

The situation looks much different, however, if we examine the normal equations (2.8) directly. If we begin with an implicit model for $\boldsymbol{x}$ as in (2.3) – or equivalently with a decomposition of P_x^{-1} as in (2.4) with M and P_w^{-1} having sparse or local structure – then from (2.10) we see that this structure is maintained in $\tilde{P}_x^{-1}$ and in the normal equations. In particular, when the measurements are point measurements of components of $\boldsymbol{x}$ with uncorrelated errors – so that C is a selection matrix and R diagonal – then $C^T R^{-1} C$ is also diagonal, so that $\tilde{P}_x^{-1} = P_x^{-1} + C^T R^{-1} C$ maintains the same structure as P_x^{-1}.

The significance of these observations is considerable. For example, for the time-recursive state space model (2.5) with local measurements, i.e.,

$$y(t) = C_t x(t) + v(t) \tag{2.11}$$

we have from (2.4) and (2.7) that P_x^{-1} is block tridiagonal, a structure that is shared by $\tilde{P}_x^{-1}$. As a consequence, the normal equations can be solved in an extremely efficient fashion, namely Gaussian elimination – also known as the Kalman filter[4] – followed by back-substitution – known as the Rauch-Tung-Striebel (RTS) smoothing algorithm[89]. Furthermore in the process of performing these calculations we directly compute the diagonal elements of $\tilde{P}_x$ – i.e., the estimation error covariance matrices for $x(t)$ for each value of t. Moreover, perhaps less widely known, these calculations also yield a model for $\tilde{\boldsymbol{x}}$ without any additional work. In particular since $\tilde{P}_x^{-1}$ has the same structure as P_x^{-1}, we might hope to model $\tilde{\boldsymbol{x}}$ as

$$\tilde{M}\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{w}} \tag{2.12}$$

where $\tilde{\boldsymbol{w}}$ is block diagonal and $\tilde{M}$ has the same structure as M in (2.7) – i.e., so that $\tilde{\boldsymbol{x}}$ has a time-recursive model as in (2.5). Such a model does in fact exist, and its parameters are directly and very simply computable from the original model (2.5) parameters and from the error covariances computed by the Kalman filter and RTS smoother.

Furthermore, since we have a model (2.12) for the estimation errors in this time-recursive statistical estimation problem, we can use the measurement residuals

$$\tilde{\boldsymbol{y}} = \boldsymbol{y} - C\hat{\boldsymbol{x}} = C\tilde{\boldsymbol{x}} + \boldsymbol{v} \tag{2.13}$$

to detect statistically significant deviations from the assumed statistics. In addition, the recursive Kalman filter algorithm allows whitening of the data $\boldsymbol{y}$ and thus the efficient computation of likelihood functions, leading to statistically optimal methods for estimating parameters of the model (e.g., parameters embedded in M, P_w, C, and R).

The critical question, then, is whether we can find analogous classes of models for phenomena that vary in space rather than time, i.e., models that have a similar set of properties and that also allow us to capture rich classes of spatial phenomena including those with multiple correlation scales. One class of such models that has been

widely proposed used is the class of Markov random fields (MRF's). As discussed in [59], such fields have models as in (2.3) in which M is an elliptic (symmetric, positive definite) partial difference operator and where $P_w = M$. In this case $P_x^{-1} = M$, emphasizing the correspondence between models and inverse covariances. Furthermore such models can capture multiple correlation scales. Moreover $\tilde{M} = \tilde{P}_x^{-1}$ in (2.10) is also an operator of the same structure as M so that subsequent data assimilation stages, in which the error statistics at one stage form the prior model for the next, face structurally identical estimation problems. The normal equations in this case correspond to an elliptic partial differential equation and the error covariance to the inverse of an elliptic operator. Consequently the required computations for estimation, error covariance calculation, anomaly detections, and likelihood evaluation are not simple and can in fact be prohibitively complex except in the case of stationary models and uniform data (so that Fourier techniques can be applied).

Section 2.4 will present an alternative to MRF's for the modeling of random fields that overcomes these difficulties through the use of scale-recursive models, permitting the realization of the full set of advantages found for the time-recursive state model (2.5).

2.2 Variational Problems and Optimal Estimation

Variational methods[18, 111] offer an alternative means for the formulation and solving of estimation problems. Variational formulations lead directly to Euler-Lagrange[18] partial differential equations (PDEs) to be solved in order to obtain the desired reconstructions. Except in those specific cases where the surface model and the measurement statistics are homogeneous, permitting FFT techniques to be applied, the solution of these equations can be a significant computational task, especially for problems of large size. Furthermore, the calculation of reconstruction error covariances[51, 98] for such approaches are, for all practical purposes, completely in-

feasible, as their computation corresponds in essence to the calculation of the full inverse of the partial differential operator arising from the variational problem. As a result, we are motivated to explore the relationship between variational problems and estimation problems with the hope of developing an alternative solution technique.

Variational problems have a dual interpretation as statistical estimation problems[63, 99]. Specifically, a variational problem with quadratic costs (i.e., a least squares problem) may be interpreted as a Gaussian statistical model:

- A variational quadratic penalty term on the deviation between the estimated surface and its measurements corresponds to an estimation problem with a measurement model in additive white Gaussian noise.
- Quadratic penalty terms on various linear functionals of the process have the statistical interpretation as a prior Gaussian model on the unknown process.

We will make the above dual interpretations more concrete in the context of the following linear-functional quadratic-cost variational problem:

$$\min_z \quad \left\{ (Y - \mathcal{C}(z))^T \mathcal{R}^{-1} (Y - \mathcal{C}(z)) + \int \int_{R^2} \mathcal{L}(z)^T \mathcal{L}(z) dx dy \right\} \tag{2.14}$$

where z is the function to be estimated on R^2, $\mathcal{L}$ is a column vector of linear functionals of z, and where Y is a set of measurements modeled by the column of linear functionals $\mathcal{C}$. Although such a variational expression is elegantly represented in continuous space, the goal of implementing a practical estimator on a computer motivates the shift to discrete space.

Let $Z^T = [\ldots, z(x_i, y_j), \ldots]^T$ represent a vector of samples of $z(x, y)$; then (2.14) may be discretized as

$$\min_Z \left\{ (Y - CZ)^T \mathcal{R}^{-1} (Y - CZ) + \{LZ\}^T \{LZ\} \right\} \tag{2.15}$$

where L is a matrix representing a discrete approximation of the linear functionals in $\mathcal{L}$ over the discrete grid, and where C is a matrix which describes the measurements of Z.

Each discrete optimization problem of the form of (2.15) possesses an estimation counterpart. Specifically, the optimization of (2.15) corresponds exactly to the problem of estimating Z given the measurement model[1]

$$Y = CZ + v \qquad v \sim \mathcal{N}(0, \mathcal{R}) \tag{2.16}$$

and a prior model

$$LZ = w \qquad w \sim \mathcal{N}(0, I) \tag{2.17}$$

The solution to this estimation problem is given by the Euler-Lagrange equation:

$$\left(L^T L + C^T \mathcal{R}^{-1} C\right) \hat{Z} = C^T \mathcal{R}^{-1} Y \tag{2.18}$$

and for which the estimation error covariance is

$$\tilde{P} = \left(L^T L + C^T \mathcal{R}^{-1} C\right)^{-1} \tag{2.19}$$

The computation of (2.18) corresponds to the solution of a PDE, a computationally difficult task. However it is (2.19) that is orders of magnitude more complex, as it corresponds to the *complete inversion* of a higher-order PDE operator.

At this point in our discussion of variational problems we have reached the same impasse as in our discussion in the previous section on MRFs: the computational difficulty in solving (2.18) and (2.19). What we propose to do is to replace the prior (2.17) by a similar *multiscale* model, such that computing (2.18) and any element of (2.19) is easy; Section 2.4 will present a multiscale framework in which it will be possible to construct appropriate variational-like models.

[1]The notation $\mathcal{N}(\mu, R)$ represents a Gaussian random vector with mean μ and variance R.

2.3 Measurements and Prior Models

Given a Bayesian estimation problem, there is a dual interpretation which we shall find to be useful: the dual interpretation between statistical prior models and measurement models.

The following discrete estimation problem is motivated by the discussion on variational problems in Section 2.2. Consider a measurement model

$$Y = CZ + v \qquad v \sim \mathcal{N}(0, R) \tag{2.20}$$

and a prior model

$$\begin{bmatrix} L_1 \\ L_2 \end{bmatrix} Z = w \qquad w \sim \mathcal{N}(0, I) \tag{2.21}$$

The essential observation here is that portions of the "prior" model may be interpreted as "measurements"; specifically, the above estimation problem is the same as the following problem, now having a modified measurement model which incorporates one part of the prior information

$$\begin{bmatrix} Y \\ 0 \end{bmatrix} = \begin{bmatrix} C \\ L_1 \end{bmatrix} Z + \bar{v} \qquad \bar{v} \sim \mathcal{N}\left(0, \begin{bmatrix} R & 0 \\ 0 & I \end{bmatrix}\right) \tag{2.22}$$

and a corresponding prior model

$$L_2 Z = \bar{w} \qquad \bar{w} \sim \mathcal{N}(0, I) \tag{2.23}$$

The solution to both of these estimation problems is the same, given by the usual Euler-Lagrange equations:

$$\left(\begin{bmatrix} L_1 \\ L_2 \end{bmatrix}^T \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} + C^T R^{-1} C \right) \hat{Z} = C^T R^{-1} Y \tag{2.24}$$

and for which the estimation error covariance is

$$\tilde{P} = \left(\begin{bmatrix} L_1 \\ L_2 \end{bmatrix}^T \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} + C^T R^{-1} C \right)^{-1} \tag{2.25}$$

This relatively simple dual interpretation may be helpful in the development of efficient estimators under circumstances in which a process Z has a statistical prior of the form (2.21), such that we have a highly efficient estimator capable of capturing prior model L_2, but where the prior statistics of L_1 are not simple to capture efficiently. We shall find it useful to reinterpret L_1 as part of the measurement model. An example of such a reinterpretation will be given in Chapter 7.

2.4 Multiscale Processing

The multiscale models of interest in this thesis and originally introduced in [14, 62] are scale-recursive models defined on index sets that are organized as multilevel trees (a simple example of such a tree for a 2-D random field was illustrated in Figure 1-1). Each level of the tree corresponds to a different scale of resolution in the representation of the random field, with coarser scales toward the top of the tree, and where the components of $\boldsymbol{x}$ correspond to variables defined at the various nodes of the tree. This modeling framework is more flexible than the figure might suggest however, because it is applicable to higher dimensional trees or to asymmetric and unusually shaped trees. This flexibility can be used to match the particular multiscale structure of the phenomenon being modeled or to capture local differences in scale structure (e.g., if the field has finer scale details in particular regions). For the purposes of the discussion in this section the quadtree structure of Figure 1-1 will suffice.

Let $\mathcal{T}$ represent the set of nodes on a multiscale tree; let $s \in \mathcal{T}$ index the nodes of the tree. All of the tree structures to be used in this thesis can be described in terms of two parameters:

q represents the *order* of the tree; that is, q equals the number of descendants of

each node (except for those on the finest scale).[2]

M represents the number of scales (or levels) in the tree.

Each node $s \in \mathcal{T}$ is associated with scale $m(s)$, where $0 \leq m(s) < M$; larger values of $m(s)$ refer to finer levels of the tree. The root node of the tree is denoted by 0, $m(0) = 0$; the root is the unique node of the tree possessing no parent.

Two operators are used to traverse the tree:

$s\bar{\gamma}$ represents the parent node of $s \neq 0$;

$s\alpha_i$ represents the ith child node of s; $1 \leq i \leq q$.

The specific model class of interest here is inspired by the successes of the time-recursive model (2.5). In particular, components of $\boldsymbol{x}$ at these nodes are related by a coarse-to-fine recursion:

$$x(s) = A(s)x(s\bar{\gamma}) + B(s)w(s) \qquad \forall s \in \mathcal{T}, s \neq 0 \tag{2.26}$$

where $w(s)$ is a white noise process with identity covariance. The initial condition of the process at the root node is given by

$$E[x(0)] = 0 \qquad E[x(0)x^T(0)] = P_o \tag{2.27}$$

Moreover, the general measurement model associated with this framework also allows measurements at multiple scales:

$$y(s) = C(s)x(s) + v(s) \qquad \forall s \in \mathcal{M} \subseteq \mathcal{T} \tag{2.28}$$

where $v(s)$ is white, with covariance $R(s)$. $\mathcal{M}$ is an arbitrary subset of $\mathcal{T}$ and contains those nodes at which measurements are present. With only the occasional exception, the applications considered in this thesis will have all of the measurements at the

[2]In principle, different nodes on the tree could have different numbers of descendants; i.e., the order $q(s)$ varies with s. Such generality is permitted by our multiscale framework, but will not be needed at any point in this thesis.

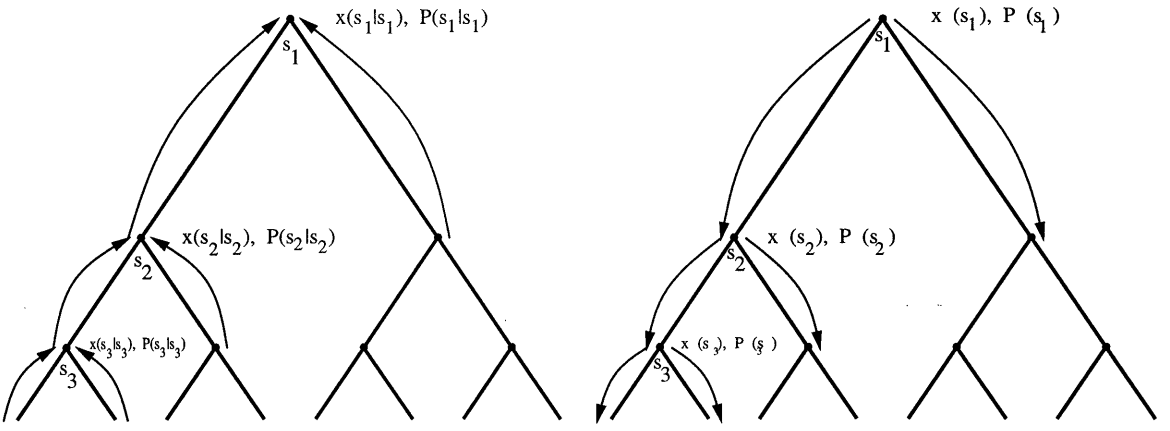


Figure 2-1: The processing of nodes in the multiscale framework proceeds in two passes. First (left figure) information is propagated up the tree; at each tree node, the conditional estimate is formed based on all measurements on that node and its descendants. Next (right figure) information is propagated back down the tree; at each tree node the estimate is formed based on all measurements on the tree.

finest scale – i.e., at a sparse and irregular subset of nodes at the lowest level on the tree – and we will focus principally on the estimates at this finest scale as well. However, the statistical algorithm for the model (2.26),(2.28) can handle data at multiple resolutions and produces estimates (and error statistics) at all scales.

Optimal estimation, error model characterization, data whitening and likelihood calculation have extremely efficient realizations for this class of multiscale models. Broadly speaking, these efficiencies are a result of the structure of the tree and the model (2.26),(2.28) which leads to a divide-and-conquer structure for statistical analysis: conditioned on any node on the tree, each of the subtrees connected to this node are conditionally independent (for example, conditioned on the top node in Figure 1-1, each of the four distinct subtrees below this node are conditionally independent). The following two subsections discuss the multiscale estimation and likelihood calculation algorithms.

2.4.1 Multiscale Estimation

The previous paragraph possessed the key to efficient estimation on multiscale trees: conditioned on any node on the tree, each of the subtrees connected to this node

are conditionally independent. Thus for any node s the processing of the data in the subtree beneath it can be decomposed into independent processing of the data in each of the descendent subtrees. As illustrated in Figure 2-1, optimal estimation of x (i.e., the collection of all $x(s)$'s) based on y (all $y(s)$'s) can be implemented as two sweeps on the tree. The fine-to-coarse sweep generalizes the Kalman filter and results in the calculation at each node s of the best linear estimate of $x(s)$ based on all of the data in the subtree below s. The usual predict and update steps of the Kalman filter are used without modification; the adaptation of the Kalman filter to the tree stems from the addition of a merge step, which combines the predicted values from several children at a common parent. Next the coarse-to-fine sweep generalizes the Rauch-Tung-Striebel algorithm and produces the best estimate and error variances at every node based on all of the data.

The resulting algorithm, the equations of which are summarized in Appendix B (see [14, 62] for greater details) involves only local calculations following the structure of the tree. Thus calculations for each node are performed once on each of the upward and downward sweeps. Furthermore, if N denotes the number of nodes at the finest scale of the tree, i.e., the number of pixels at the finest scale of resolution, then the total number of nodes on the tree is $\frac{4}{3}N$. Thus the total complexity of the algorithm is proportional to N, resulting in constant complexity per grid point independent of the size of the grid.[3]

The multiscale estimation equations yield a model for the error $\tilde{x}(s) = x(s) - \hat{x}(s)$ which has a multiscale form[64], so that subsequent data assimilation stages can be carried out in exactly the same fashion. Specifically,

$$\tilde{x}(s) \;=\; \tilde{P}(s \mid s)P_s^{-1}A(s)P_{s\bar{\gamma}}\tilde{P}^{-1}(s\bar{\gamma} \mid s)\tilde{x}(s\bar{\gamma}) + \check{w}(s) \tag{2.29}$$

where P_s is the prior covariance at node s and $\tilde{P}(s \mid t)$ represents the estimation error covariance of $x(s)$ given all observations in the subtree below node t.

[3]The computational complexity of the tree is $\mathcal{O}(Nk^3)$, where k represents the state vector length of $x(s)$. Constant complexity per grid point is achieved under the assumption (or assertion) that k be a fixed value, independent of N.

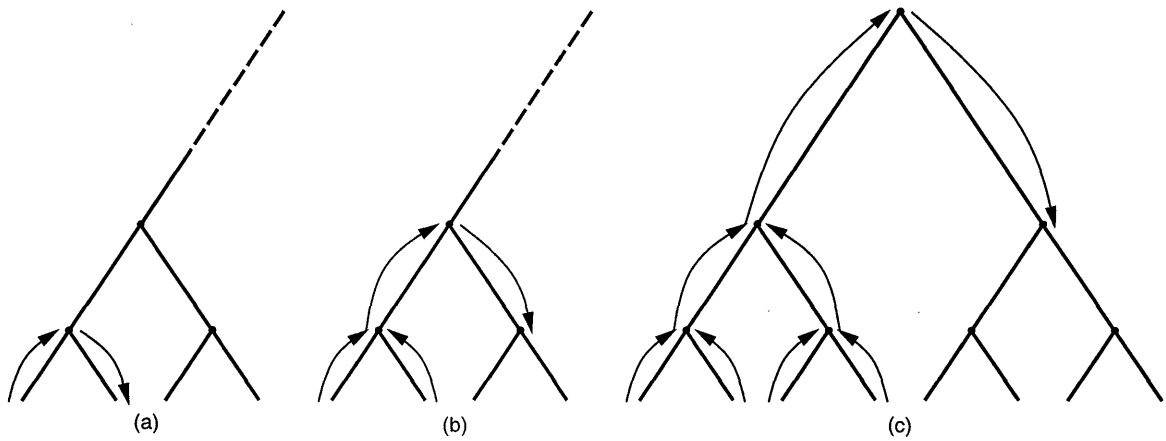


Figure 2-2: A rough sense of the information flow for the multiscale whitening algorithm. The whitening proceeds in a depth-first approach, first conditionally whitening each of the children of a parent node before whitening the parent. The notions suggested by this figure are made more precise in [62, 65] and in Appendix C.

The estimation error covariances $\tilde{P}(s \mid s)$ are computed for each node s on the tree; that is, the *block-diagonal* components of the full estimation error covariance matrix are computed. For problems of the size considered in this thesis (e.g., $N \approx 10^5$), the full error covariance matrix is too large (e.g., $\approx 10^{10}$ elements) to be calculated or even to be stored by any practical means. It may, however, be useful to calculate a specific subset of the off-diagonal elements.

The model in (2.29) permits the calculation of the estimation error covariance between any two arbitrary nodes in the tree (i.e., the calculation of arbitrary off-diagonal elements in the full error covariance matrix); an example of the application of this model will be demonstrated in the oceanographic context in Chapter 3.

2.4.2 Multiscale Likelihood Calculation

Another algorithm, closely related to the optimal estimation algorithm of the previous subsection, allows us to whiten the data and compute likelihoods in an equally efficient fashion[62, 65]. That is, given a set of multiscale model parameters $A(s)$, $B(s)$, $C(s)$, $R(s)$, P_o (and possibly other parameters specifying the structure of the tree such, as the order q) and a stacked vector Y of measurements, we can calculate the likelihood

$$\mathcal{L}\left[A(),B(),C(),R(),Y\right] = -\frac{1}{2}\log|\Lambda_y| - \frac{1}{2}Y^T\Lambda_Y^{-1}Y - \frac{\|Y\|}{2}\log 2\pi \quad (2.30)$$

where $\|Y\|$ counts the number of elements in Y, Λ_Y represents the covariance of random vector Y, and where we assume that the noise terms w,v of (2.26),(2.28) are Gaussian,

The direct evaluation of expressions such as (2.30) is typically difficult, even in the straightforward time-recursive case (2.5). Instead, in the time-recursive case such a likelihood calculation is made simpler by first computing the whitened residuals process; such a residuals process is computed by the Kalman filter. Fortuitously, a multiscale whitened residual process may be computed using the multiscale analog of the Kalman filter. Essentially, information is passed about the tree in a "pre-order"[4] traversal; a suggestive sketch of the information flow is shown in Figure 2-2.

The equations of the resulting algorithm are summarized in Appendix C (see [62, 65] for greater details). The computational effort of the whitening algorithm is similar to its estimation counterpart from the previous subsection; specifically, the algorithm involves only local calculations following the structure of the tree, thus two sets of calculations are performed at each node, leading to a total complexity of the algorithm which is a constant multiple of the number of grid points, independent of the size of the grid.

In principle, once we have a multiscale whitening algorithm (and hence a likelihood calculation algorithm) the estimation of multiscale model parameters is possible. Specifically, suppose that we have a family of multiscale models parameterized by a vector H; then an estimate of H may be determined by maximizing the likelihood function

$$\hat{H} = \arg_H \max \mathcal{L}\left[A(s,H),B(s,H),C(s,H),R(s,H),Y\right] \quad (2.31)$$

[4]There are three basic forms of tree traversal: "pre-order" (or depth-first), "in-order", and "post-order"; the prefix describes the time of processing the parent node relative to its children.

Chapter 4 will explore the identification of multiscale model parameters using approaches like that of (2.31).

2.5 Multiscale Modeling

Given the multiscale framework and the estimation and likelihood algorithms outlined in the previous section, one remaining challenge lies in the selection or determination of an appropriate multiscale model. In general, there are two basic approaches to solving computationally difficult estimation problems with some given prior statistics:

1. Replace the prior model with another model (e.g., a multiscale one) which is *similar* to the original, but whose optimal solution can be found efficiently.

2. Develop a suboptimal algorithm and apply it directly to the original problem.

Applications of the multiscale framework[63, 65] have tended to follow the former philosophy: we do not view our multiscale framework as an *approximation* to a given estimation problem, rather the multiscale approach allows us to solve the problem optimally under a multiscale prior that is similar; the development of multiscale models in this thesis will follow the same spirit.

In addition to the computational efficiencies admitted by the multiscale framework and its related algorithms of the last two subsections, multiscale models can also be used to capture the statistical structure of rich classes of phenomena. The following subsections will document several statistical models and their multiscale counterparts. These next subsections are more than just examples: we will have occasion to build upon these multiscale counterparts throughout this thesis.

2.5.1 Markov Random Field-Like Models

This subsection will describe exact and approximate multiscale counterparts to Markov random field models. Although we will never implement or make use of such multiscale models directly, the intuition behind the development of multiscale MRF counterparts will be useful in the investigation of the smoothness of multiscale

estimates in Chapter 6. A thorough development of the material in this section may be found in [61, 62].

Let $\Omega \subseteq \mathcal{R}^2$ be a closed set, and let Γ be the boundary set of Ω; Γ separates $\mathcal{R}^2$ into $\Omega \cap \Gamma^c$, Γ, and $\Omega^c \cap \Gamma^c$. A continuous-space, two-dimensional process $z(t), t \in \mathcal{R}^2$, is said to be a Markov random field if the process inside of Ω, $\{z(t) \mid t \in \Omega \cap \Gamma^c\}$, is independent of the process outside of Ω, $\{z(t) \mid t \in \Omega^c \cap \Gamma^c\}$, given the process on the boundary set Γ, $\{z(t) \mid t \in \Gamma\}$.

On a discrete lattice (instead of on a continuous plane) the definition of a MRF becomes more subtle[7, 22, 35]. Essentially, $z(t), t \in \mathcal{Z}^2$ is a MRF if there exists a neighborhood set D_t, such that

$$P_{z(t)|z(\tau)}(Z_t \mid \{Z_\tau \mid \tau \in D_t\}) = P_{z(t)|z(\tau)}(Z_t \mid \{Z_\tau \mid \tau \in \{\mathcal{Z}^2 \cap \{t\}^c\}) \tag{2.32}$$

For simplicity in this discussion, we restrict our attention to processes $z(t)$ that are Markov random fields under first order neighborhoods:

$$D_t = \{(t, t+1), (t+1, t), (t, t-1), (t-1, t)\} \tag{2.33}$$

Consider the lattice shown in Figure 2-3. Let $z(t)$ be a process defined on this lattice, such that $z(t)$ is a MRF under the first order neighborhood (2.33). Then conditioned on the process values

$$\{z(t) \mid t \in \text{Shaded region of Figure 2-3}\} \tag{2.34}$$

the four processes (one per quadrant) $\{z(t) \mid t \in Qi\}$ are all mutually independent.

Recall that a property, similar to this last statement, is asserted on the multiscale tree: conditioned on a parent node, the q children descendent from this parent are independent. Consider modeling the process $z(t)$ of Figure 2-3 on a quad-tree: let $x(0)$, the state at the root node of the tree, consist of the shaded elements of $z(t)$ in Figure 2-3; furthermore identify each of the quadrants Qi with the children of the root node; now proceed recursively down the tree. Clearly such an approach leads to

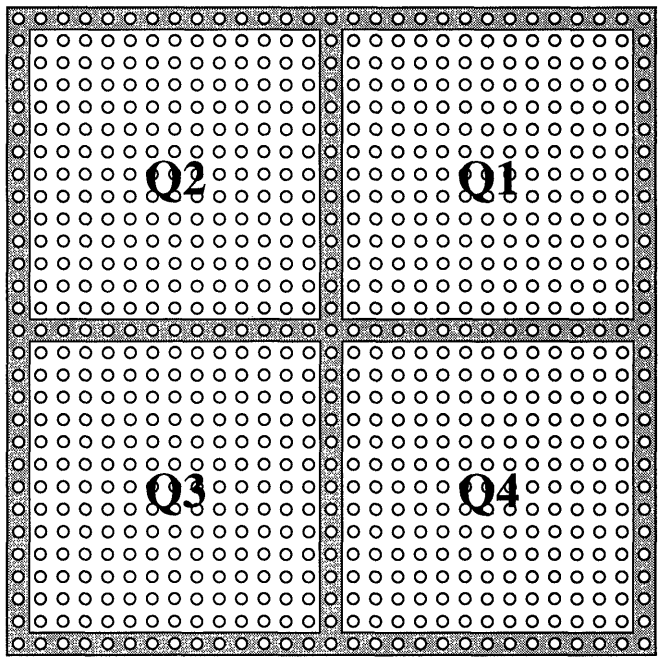


Figure 2-3: The figure shows a finite, two dimensional lattice. Consider a process $z(t)$ to be defined on the lattice sites, such that $z(t)$ is a MRF under the first order neighborhood (2.33). Then conditioned on the lightly-shaded states, the processes $z(t)$ in each of the four quadrants $Q1, \ldots, Q4$, bounded by the shaded region, are independent.

a multiscale model having large state vectors: for a first-order MRF representing N^2 pixels, the root node of the multiscale counterpart would contain $(6N - 9)$ elements. The development of multiscale counterparts to Nth order Markov random field processes may be accomplished by setting the state of each parent node to capture the boundary elements between and around its child nodes, but where the boundary is taken to be N pixels thick.

In principle, multiscale counterparts with smaller state dimensions may be formed by maintaining a low order approximation to the MRF elements; for example, the shaded elements in Figure 2-3 might be represented by a set of Fourier or wavelet coefficients. The arbitrary selection of a Fourier or wavelet basis is somewhat *ad-hoc*; the next subsection will describe a more systematic approach to determining lower-order representations.

2.5.2 Canonical Correlation-Based Models

The previous section outlined an approach for constructing multiscale analogs to Markov random field models; however there are two limitations to that approach:

- The previous subsection does not yield any systematic approaches for determining *good* reduced-state-dimension multiscale counterparts to MRF models.
- The previous subsection does not suggest a means by which to develop multiscale counterparts to non-Markov processes.

The method of canonical correlations[1, 3] has allowed the above issues to be addressed.

Let $x = [x_1 x_2]$ be a Gaussian random vector with a known correlation structure

$$E\left[xx^T\right] = \Lambda_x = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{12}^T & \Lambda_{22} \end{bmatrix} \tag{2.35}$$

We would like to determine two orthogonal matrices T_1, T_2 such that $\chi_1 = T_1 x_1$ and $\chi_2 = T_2 x_2$ are in canonical form:

$$E\left[\chi_1^T \chi_1\right] = I \qquad E\left[\chi_2^T \chi_2\right] = I \tag{2.36}$$

$$E\left[\chi_1^T \chi_2\right] = D = \text{diag}\left(\sigma_1, \ldots, \sigma_p, 0, \ldots, 0\right) \tag{2.37}$$

where $\sigma_1 \geq \ldots \geq \sigma_p > 0$. The appropriate orthogonal matrices are readily found using the singular value decomposition. Let

$$A = \Lambda_{11}^{-\frac{1}{2}} \Lambda_{12} \Lambda_{22}^{-\frac{1}{2}} \tag{2.38}$$

Compute the SVD

$$A = USV^T \qquad U^T U = I, V^T V = I \tag{2.39}$$

then

$$T_1 = U^T \Lambda_{11}^{-\frac{1}{2}}, \qquad T_2 = V^T \Lambda_{22}^{-\frac{1}{2}}, \qquad D = S \tag{2.40}$$

Essentially, the singular value decomposition of Λ_{12} quantifies all interdependencies between x_1 and x_2; therefore by conditioning on the first p elements of χ_1 and χ_2, the vectors x_1 and x_2 become independent. In other words, the linear functionals

$$\tau_1 = [I_p \;\; 0]\, T_1 \qquad \tau_2 = [I_p \;\; 0]\, T_2 \tag{2.41}$$

describe the information needed to decorrelate x_1 and x_2. Furthermore, the ordering of the singular values in the diagonal matrix D (2.37) suggests a natural *reduced* set of linear functionals:

$$\tau_{1,k} = [I_k \;\; 0]\, T_1 \qquad \tau_{2,k} = [I_k \;\; 0]\, T_2 \qquad k \leq p \tag{2.42}$$

The ability to maintain a set of values to conditionally decorrelate multiple random vectors can play a central role in the development of multiscale models. Two modifications to the canonical correlations approach described in this subsection are required in order to develop multiscale counterparts to arbitrary Gaussian processes:

1. The canonical correlations procedure essentially just captures the correlation between two random vectors. On the tree, such correlations must be captured on a hierarchy of scales.

2. The canonical correlations procedure conditionally decorrelates *two* vectors; on multiscale trees with order $q > 2$, the procedure must be generalized to q random vectors.

These generalizations have been researched[48] and have led to an effective multiscale modeling technique; an example of the use of such multiscale models will be presented in Chapter 6.

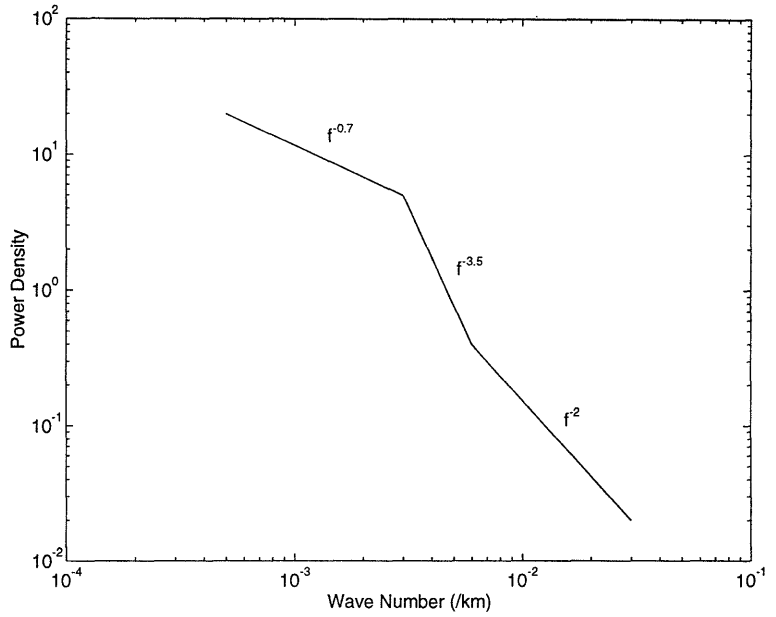


Figure 2-4: An example of a natural phenomenon characterized by a $1/f^{\mu}$-like behavior. This figure shows a rough characterization of the global power spectral density of the ocean surface (from [33]).

2.5.3 $1/f$-like Models

Our multiscale framework is directly suited to capturing phenomena that display a multitude of correlation scales. Of particular interest is the class of so-called $1/f$ models[113], i.e., processes that display $1/f^{\mu}$-like spectra over a significant range of frequencies. Many natural and human phenomena have been found to possess $1/f$-like spectral properties, which has led to considerable study of $1/f$ processes. For example, Figure 2-4 (from [33]) shows a typical power spectrum for the height of the ocean surface, modeled as a $1/f^{\mu}$-process with different values of μ over different wavenumber intervals.

One class of such processes that is frequently used because of its analytical convenience and tractability is the class of fractional Brownian motion (fBm) processes, introduced by Mandelbrot and Van Ness[69]. Let $F[k]$ be a sampled fBm process; i.e.,

$$F[k] = F(k\Delta t) \qquad k \in Z \tag{2.43}$$

for which the associated nonstationary covariance is

$$E\left[F[k], F[m]\right] = \frac{\sigma^2}{2}(\Delta t)^{2H}\left(|k|^{2H} + |m|^{2H} - |k-m|^{2H}\right) \tag{2.44}$$

where σ and H are scalar parameters which completely characterize the process: σ controls the overall power of the process, H determines the fractal dimension $D = 2 - H$ of the process. The modeling of fBm on multiscale trees, and the estimation of H given samples of a fBm process will be studied in Chapter 4.

The class of fBm processes, although interesting, represents only a single possibility among a broad array of $1/f$-like processes; furthermore the *exact* representation of fBm processes on the multiscale tree is not particularly convenient. More generally, phenomena with $1/f^{\mu}$-like spectra display so-called self-similar scaling properties in that the variability of such a phenomenon scales geometrically with the spatial resolution at which the variations are measured. Such scaling rules are captured very simply in our multiscale model through the imposition of a scaling relationship in the gain $B(s)$ in (2.26). Recall that $m(s)$ denotes the scale of a node s; thus the choice

$$A(s) = 1 \qquad B(s) = B_o 2^{(1-\mu)m(s)/2} \tag{2.45}$$

displays the same scaling behavior as that implied by a $1/f^{\mu}$ spectrum[115]. Changes in scaling laws, corresponding for example to the changes in spectral slope in Figure 2-4, can be captured simply by changing the value of μ over different ranges of scale. Local changes in scaling structure can also be easily accommodated by local modifications of $B(s)$.

There are several motivations for the use of the multiscale model of (2.45):

- The simplicity of the model makes it easy to implement and computationally very efficient.
- Since the model is described in terms of only two parameters the model identification process is relatively straightforward.
- The statistics of many natural processes are poorly understood, and in fact rep-

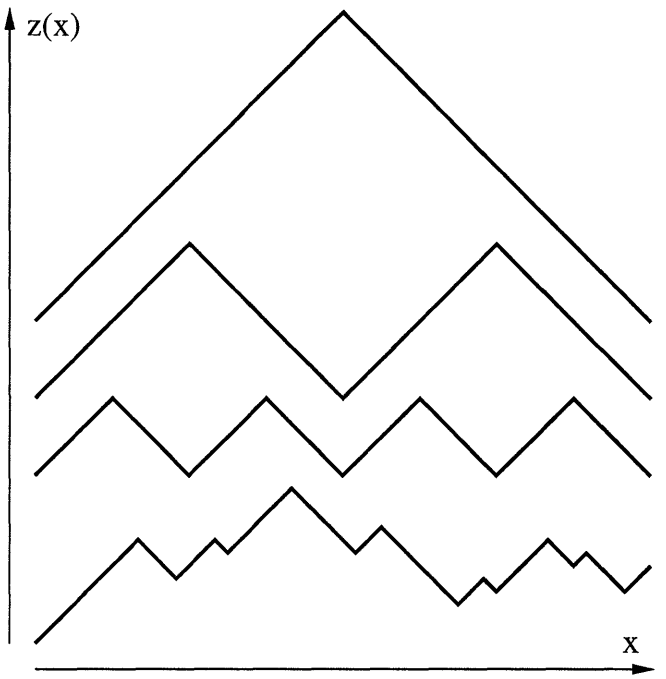


Figure 2-5: A collection of four curves, each equally penalized by the cost function $\int (dz/dx)^2 d\tau$

resent an active area of research. For such processes, detailed and sophisticated multiscale models based on poorly-known statistics are inappropriate; instead, a simple model (such as (2.45)) which captures the known basic scale to scale behavior of the process of interest may yield equally good results.

We will be using (2.45) as the basis for each of our remote sensing applications in Chapters 3, 4, and 5.

2.5.4 Variational-Like Models

The discussion in Section 2.2 motivates the development of multiscale prior models *similar* to the prior implied by simple variational constraints. This problem has been investigated in the past[14, 63, 99] for the case of a single quadratic penalty on the unknown function. The basic idea for this case is most easily visualized in 1-D; i.e., for variational costs of the form

$$\int \left(\frac{dz}{dx}\right)^2 d\tau \tag{2.46}$$

Under the variational cost function of (2.46), each of the four profiles for $z(x)$ depicted in Figure 2-5 incurs the same penalty. Indeed, in 1-D the penalty term (2.46) is equivalent to a Brownian motion prior model[63]; i.e., a process with a $1/f^2$ spectrum. Based on the discussion of $1/f$-like multiscale models in the previous subsection, we can posit the following model as a prior similar to that of (2.46):

$$z(s) = 1 \cdot z(s\bar{\gamma}) + B_o 2^{-m(s)/2} w(s) \tag{2.47}$$

As discussed in[63], an analogous interpretation also holds for 2-D variational penalties on derivatives of $z(x, y)$, leading to quadtree models of the same form as (2.47). We will be using and building upon this result extensively in Chapter 7.

2.6 Multiscale Framework Implementation

The computer implementation of the multiscale framework has proven to be a challenging and interesting task. Although computer implementations of such algorithms are not the focus of this thesis, a brief discussion is warranted.

The framework does *not* pose inherent difficulties to implementation; for example, a basic implementation of the multiscale estimator (see Appendix B) on a quadtree sufficient for preliminary oceanographic tests was accomplished in about two days. Any difficulties associated with implementing the framework stem from the remarkable variety of possible multiscale tree configurations and tree models. Indeed, we encountered the following dilemma:

- An implementation which is targeted to a particular tree structure and multiscale model can reap considerable computational efficiencies by taking advantage of the known structure, but a new or heavily modified computer program is likely to be needed for each different tree structure or model.
- An implementation which is completely general and tolerant of *any* tree structure whatsoever would be slow and awkward to use (even a simple tree would require a detailed and exact description, since nothing would be implicit or

assumed).

We have developed the following creative compromise:

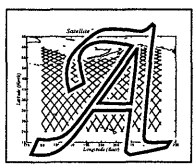
- Develop a core "engine", which implements the multiscale estimator, likelihood calculator, cross-covariance calculator, and sample path generator, without making *any* assumptions whatsoever regarding the tree structure or multiscale model.

- Create a set of interfaces which operate between this core engine and the user-supplied application program. Each interface asserts a different set of assumptions regarding the multiscale tree and model; these assumptions are asserted at compile-time, allowing the general engine code to be modified, simplifying itself based on the nature of the approximations asserted, and allowing it to achieve the associated computational efficiency.

The above approach has enabled a *single* piece of code to perform *all* of the multiscale calculations in this thesis in a computationally efficient manner. Further details and an overview of the organizational structure of the software may be found in Appendix A.

Chapter 3

Ocean Surface Estimation



significant estimation problem in oceanography concerns the interpolation of large oceanographic data sets; this chapter will describe the application of our multiscale framework to such a problem. This application, although significant in and of itself, is meant to illustrate the potential for the utility of the multiscale approach in broader contexts; e.g., the geodesy application in Chapter 5 is based heavily upon the results and insights acquired in the investigation of this chapter.

Section 3.1 presents an introduction and the necessary background for the altimetric interpolation problem. Section 3.2 discusses the development of an appropriate multiscale model, followed by experimental tests in Section 3.3. Background information on optimal estimation, which accompanies this chapter in its journal form[27] for the benefit of the remote sensing community, may be found in Chapter 2.

3.1 Introduction

The problem of estimating the shape of the ocean surface from satellite altimetry measurements is of considerable current interest both because of its importance in global ocean modeling and climate studies and because of the relatively recent launch of the joint American/French TOPEX/POSEIDON altimeter[32, 73, 95], a satellite-based platform capable of measuring ocean height to an unprecedented accuracy of

approximately 5 cm. The availability of data of this quality and coverage makes it possible to address a variety of scientific questions ranging from producing regularly gridded maps of ocean height (to be used, for example, in global ocean modeling studies) to the estimation of the spatial spectrum of ocean height variations. Achieving objectives such as these, however, presents daunting challenges to the data analyst, in particular in terms of the enormous size of the problems to be solved. The method of this chapter permits the production of statistically optimal results, with computational loads that are extremely modest.

The various quantities involved in the satellite altimetry problem are sketched in Figure 3-1. The TOPEX/POSEIDON altimeters are mounted on a satellite orbiting at an altitude of approximately 1500km. The altimeters use microwave ranging techniques to determine the precise distance, D, between the satellite and the surface of the ocean. GPS navigation and laser tracking from ground stations[78] determine the position, O, of the satellite in three-dimensional space with respect to an idealized, ellipsoidal, earth. Other details aside, the difference $h = O - D$ measures the height of the surface of the ocean with respect to the idealized ellipsoid.

In principle, the height h includes effects such as oceanic tides[90], solid body tides, and atmospheric pressure influences. Although the study and quantification of these effects are worthwhile efforts, we are not interested in estimating such quantities in this thesis, and we assume tidal and atmospheric perturbations to have been subtracted from h. With this said, $h = N + \zeta$ is made up of two principle quantities:

N represents the geoid[42, 50] , i.e., the gravitational equipotential surface at sea level. The geoid is very nearly an ellipsoid; deviations from the ellipsoidal shape are due to spatial fluctuations of the density of the earth's mantle and crust, and due to the earth's topography (e.g., mountains and valleys).

ζ represents the height of the ocean surface relative to the geoid. In the absence of ocean currents, $\zeta = 0$. Conversely, the spatial gradients of ζ allow elements of the ocean circulation pattern to be deduced[116]; this observation forms one of the main motivations for the estimation of ζ.

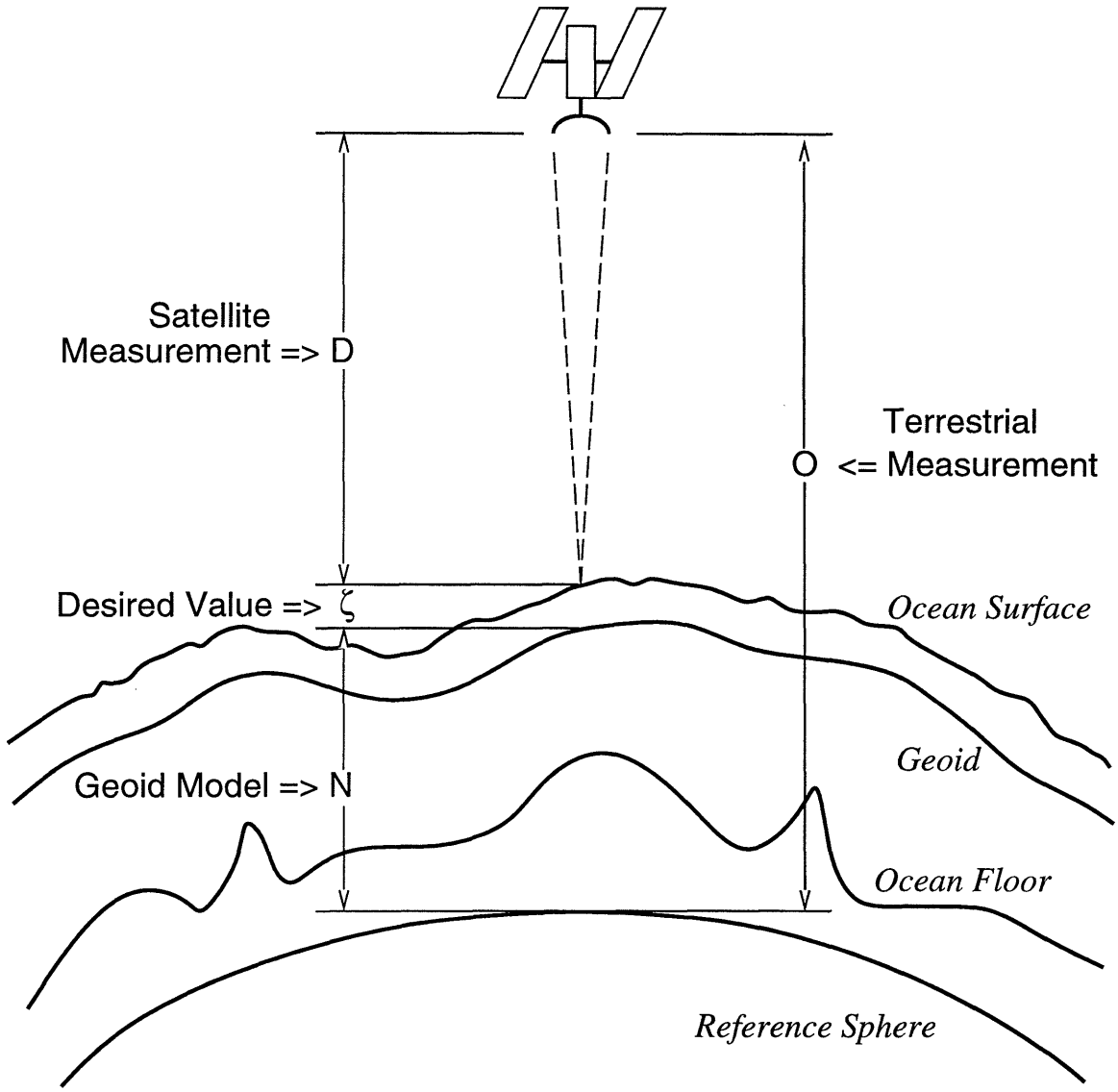


Figure 3-1: A general overview of the nomenclature associated with the TOPEX / POSEIDON measurements.

The importance of deducing ζ from satellite data cannot be overemphasized. It is impossible to match the breadth, uniformity, and frequency of ocean sampling obtained by satellites by any other means; ship-borne experiments, while capable of sampling the interior of the ocean, are slow to perform and prodigious in cost. The importance of determining the circulation pattern itself stems from the fact that the kinetic energy of the ocean circulation vastly exceeds that of the atmosphere, thus to properly model and understand the earth's climate and global warming necessitates an improved knowledge of oceanographic currents.

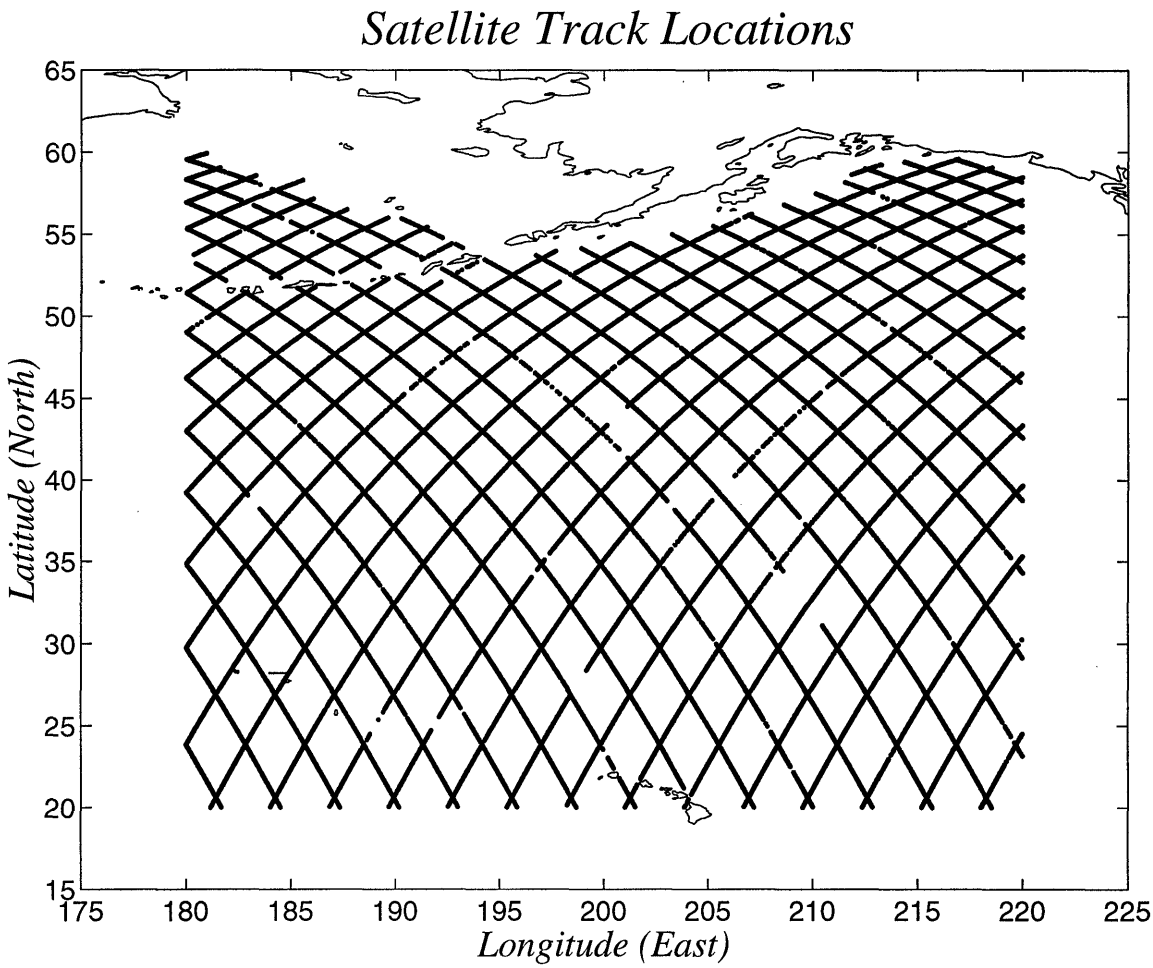


Figure 3-2: Set of TOPEX/POSEIDON measurement tracks in north Pacific. The region shown in this figure will be the focus of the estimation efforts in this chapter.

Estimates have been developed for N based on satellite tracking data, previous altimeter missions, and from direct measurements of gravity. The estimates of N used in this chapter are the OSU91A[87] estimates developed at the Ohio State University. The estimates are computed from a spherical harmonic basis to order and degree 360 (i.e., a resolution of approximately one degree).

Figure 3-2 depicts a region of the northeastern Pacific from Hawaii to Alaska. Overlaid on this region is the distribution of TOPEX/POSEIDON satellite measurements[78] over a typical ten day cycle. Successive measurements along a track are separated by approximately 7km or 0.06 degrees; the spacing between adjacent tracks is approximately 270km. Gridded images of the ocean are required

at fine scales, both in order to observe features of interest, and to produce numerical values compatible with fine scale ocean models. Even for the comparatively modest portion of the ocean shown in Figure 3-2, we must estimate ocean surface heights at more than 100,000 grid points based on roughly 20,000 altimetric measurements. For a full ocean basin, or for the entire global surface, the problem is of truly formidable proportions.

The size of data analysis problems such as this is not the only significant challenge. In many cases, including the one of interest here, significant spatial nonstationarities are present for several possible reasons:

1. The sampling pattern of the data is frequently nonuniform and irregular, including occasional periods of data dropout as shown in Figure 3-2.

2. The sensed phenomenon is itself nonstationary, exhibiting differing spatial scales and magnitudes of variability in different regions. Ocean surface statistics, for example, differ between regions containing vigorous currents such as the Kuroshio or the Gulf stream and those regions which are comparatively quiet such as the northeast Pacific.

3. The *quality* of measurements may also be nonstationary. Recall that the TOPEX / POSEIDON altimeter provides direct measurements, D, of the distance from the satellite to the ocean surface, whereas ζ is the fundamental quantity of interest as the ocean current field may be inferred from the derivative of this relative surface. The geoid estimates are subtracted from h to yield measurements of ζ; thus the complex and nonstationary error structure of the geoid estimates[87] translates directly into nonstationary errors in the altimetry measurements.

Such nonstationarities or irregularities in the data pattern present a major challenge[33, 117], as there is no regular structure that can be used to advantage. In particular Fourier methods, with their noteworthy efficiencies, cannot be applied directly or without significant approximations and idealizations.

Furthermore, in addition to the estimation of quantities such as ocean surface height, there are compelling reasons for desiring a characterization of the errors in these estimates. In particular, to assess the value of a set of estimates we must have a measure of their accuracy, requiring at the very least the calculation of error variances. Moreover, there are strong motivations for the characterization of the spatial correlation structure in the estimation errors. For example, the assimilation of ocean surface estimates into global circulation models[33], which effects a blending of the surface measurements and the underlying science, in principle requires the full specification of the error correlations so that accurate model/data combinations can be effected.

In addition, error covariance calculations are useful for a variety of other scientific reasons. For example, geoid estimates have errors due to unresolved, spatially localized perturbations such as sea mounts or trenches. Such errors can manifest themselves as outliers in the data, or more precisely in the residuals (data minus estimates); the availability of error statistics permits the identification of statistically significant outliers and the estimation of localized geoid corrections implied by these residuals.

Finally an important characteristic of many remote sensing problems, including the one examined here, is that the phenomenon under study exhibits behavior across a broad range of scales. For example, global ocean models predict behavior at (and interactions among) a vast range of spatial scales. Indeed, models for ocean height spectra[33] are typically described in terms of inverse power-law relationships. Such a spectral description corresponds directly to a scaling relationship between the expected amplitude and spatial scale of ocean features - i.e., it corresponds to a fractal model. Statistical modeling of the ocean surface and the processing of ocean height data must account for this multiscale structure.

A number of smoothing and data assimilation algorithms (e.g., objective analysis[20], kriging[91]) have been developed, each of which has emphasized varying degrees of statistical structure or computational efficiency. The combination of the issues we have mentioned – problem size, nonstationarity, statistical characterization

of errors, and accounting for correlation structures over a range of scales – has generally required that compromises be made in the statistical consistency and optimality of the results. The multiscale method to be illustrated in this chapter avoids the need to make such compromises.

The multiscale algorithm has a total computational complexity per grid point independent of the size of the grid, can accommodate nonstationarities in the model of the phenomenon or the data, and allows the complete characterization of error statistics; the results presented in Section 3.3 will highlight these capabilities. Our approach also produces estimates at a hierarchy of scales, facilitating resolution/accuracy tradeoffs and the direct extraction of estimates of coarser scale features. Finally, all of our satellite data is taken at the same level of resolution, however in principle we could incorporate data of differing resolution and coverage with no change in algorithmic structure.

3.2 Multiscale Model Selection

The experimental results of this chapter are based upon data taken over a single TOPEX/POSEIDON repeat cycle (about 10 days)[52]. The altimetric measurements of h are processed by subtracting the geoidal reference field[87, 88]; furthermore the usual corrections are applied to the data: ionospheric[75], tidal[90], orbital[78], and atmospheric pressure loading.

Recall the multiscale dynamic equations (2.26),(2.28) discussed in Section 2.2. The first tasks in determining a specific multiscale model are the selection of the multiscale tree structure, the nature of the model (in particular, the dimension of $x(s)$) and the specific model parameters (e.g., the $A(s)$, $B(s)$, $C(s)$ values of (2.26),(2.28)). Since the finest scale of a quad-tree (e.g., as in Figure 3-3) is a 2-D process, we are motivated to select such a quad-tree as the basis for our multiscale model. We let $x(s)$ be a scalar representing the ocean height at the particular scale and position corresponding to node s. The determination of the scalars $A(s), B(s)$ is then made by choosing these parameters to match certain characteristics of the TOPEX/POSEIDON data.

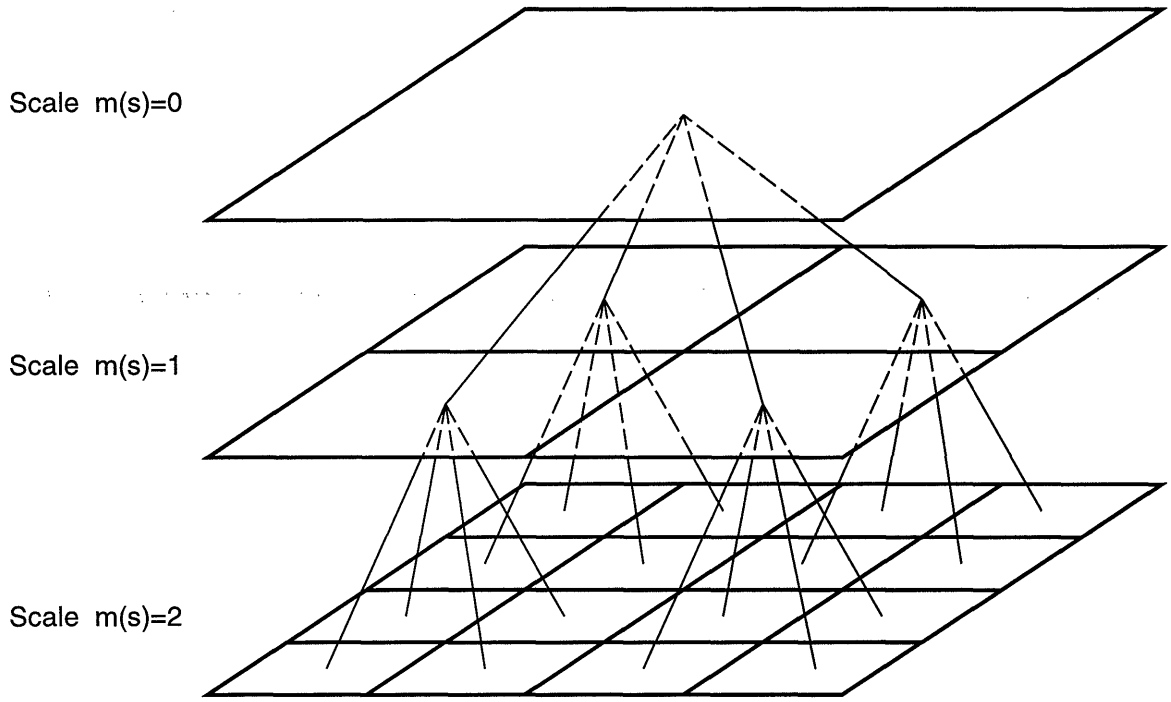


Figure 3-3: The multiscale tree structure to be used for the ocean elevation estimation problem.

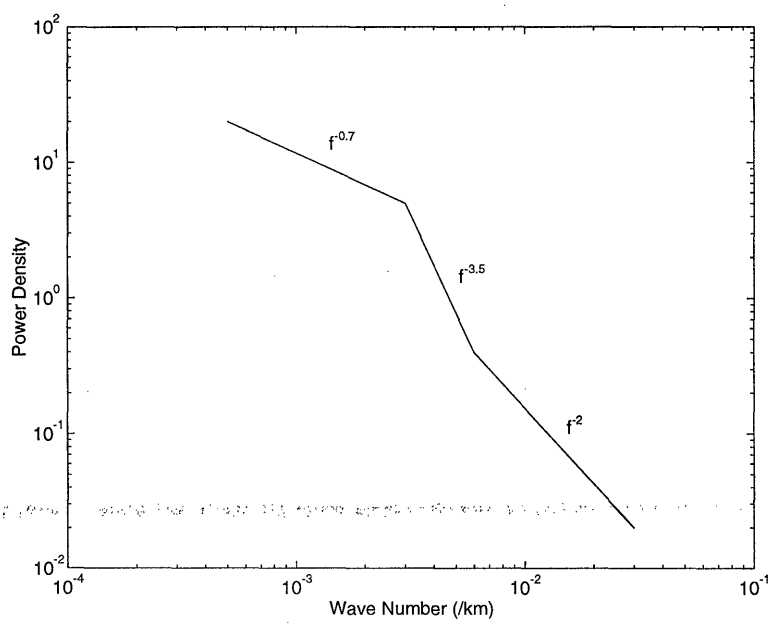


Figure 3-4: A rough characterization of the global power spectral density (from [33]); the characterization is $1/f$-like, which motivates the selection of a $1/f$-like multiscale model.

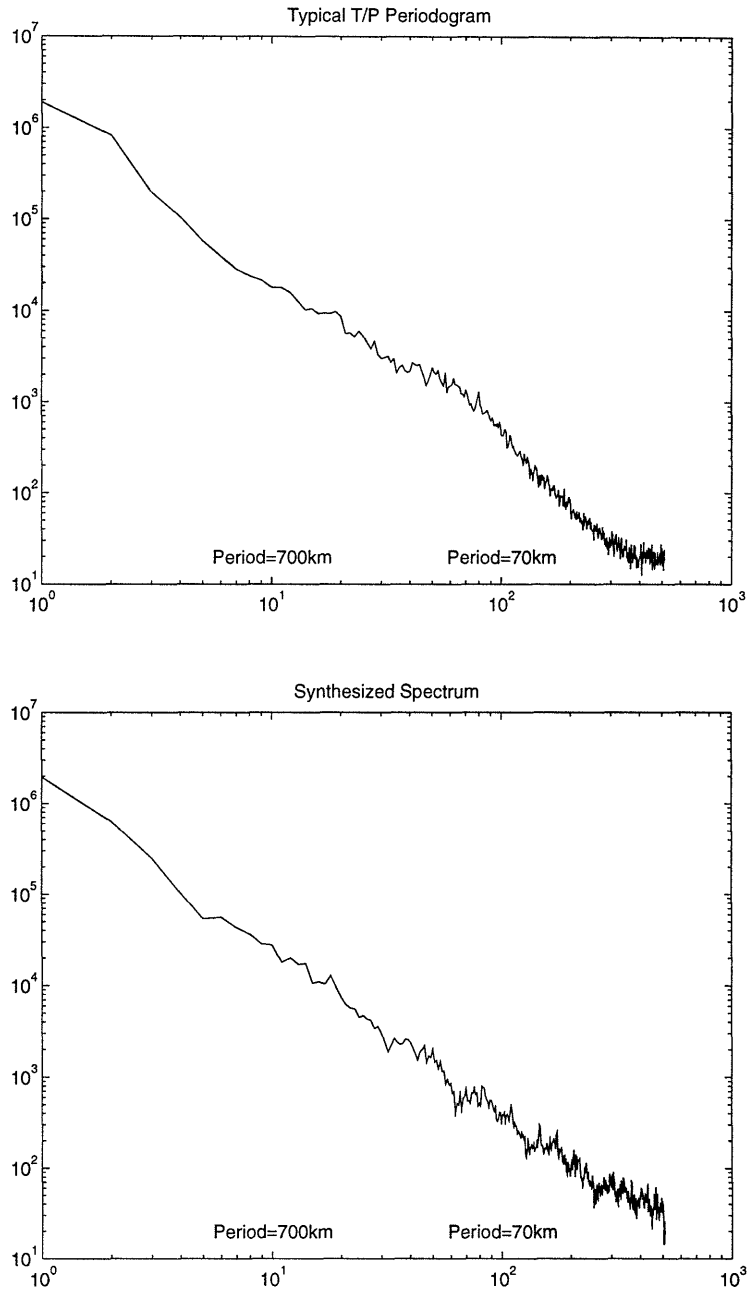


Figure 3-5: Top: Empirical power spectral density based on TOPEX/POSEIDON data.
Bottom: Power spectral density from multiscale model simulations.

Figure 3-4[33] shows a rough characterization of the global power spectrum of the ocean surface elevation ζ from the oceanographic literature. The spectrum is characterized as piecewise straight in the log-log domain of the figure, motivating the selection of a $1/f$-like multiscale model to represent the ocean surface elevation. Rather than relying upon such a global spectrum, we can determine the empirical power spectrum for the ocean surface within our region of interest directly from TOPEX/POSEIDON data: the top spectrum of Figure 3-5 shows such a periodogram. As in Figure 3-4, the spectrum falls as a relatively straight line in the log-log domain. From the discussion in Section 2.5.3, we know that the following model possesses $1/f^{\mu}$ characteristics:

$$A(s) = 1 \qquad B(s) = B_o 2^{(1-\mu)m(s)/2} \tag{3.1}$$

It is a simple matter to choose B_o, μ such that the power spectrum associated with the sample paths of the multiscale model (3.1) is similar to the empirical spectrum at the top of Figure 3-5. Specifically, the choice $B_o = 35cm, \mu = 2$ leads to the following multiscale model (where $x(s)$ is measured in cm):

$$x(s) = x(s\bar{\gamma}) + 35 \cdot 2^{-m(s)/2} w(s) \tag{3.2}$$

That is, the aggregate surface height of the ocean at some position and scale equals the aggregate height of its parent node, i.e., at the same spatial position but at a coarser scale, plus a perturbation offset whose variance decreases geometrically with scale. The model (3.2) corresponds to the power spectrum shown in the bottom half of Figure 3-5, which is very similar to the original empirical spectrum at the top of the figure. There is a second manner in which to determine the multiscale model parameters using multiscale likelihood techniques. This approach will be discussed in Chapter 4. We will continue to use (3.2) as the multiscale model for the computations described in the next section.

Next, the prior variance P_s of $x(s)$ at each node of the tree can be determined

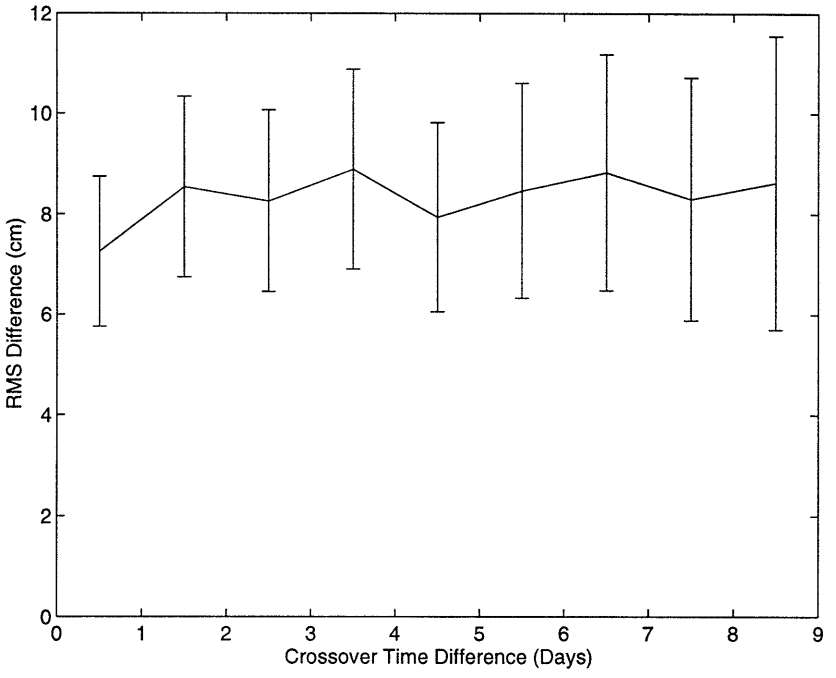


Figure 3-6: RMS statistics of altimetric offsets at orbit crossover points (i.e., at those points in physical space where two satellite orbital paths intersect) as a function of the time difference between the two paths. The figure supports the assumption that the ocean surface is relatively constant over periods of time up to ten days.

from a recursion obtainable directly from (2.26):

$$P_s = E\left[x(s)x^T(s)\right] = A(s)P_{s\bar{\gamma}}A^T(s) + B(s)B^T(s) \tag{3.3}$$

The recursion is initialized with the prior variance P_o of $x(0)$ at the root node of the tree. Roughly speaking, P_o can be thought of as specifying the prior level of uncertainly in the aggregate mean height of the ocean. Here, in order to avoid biasing our estimate of overall ocean height, we have set P_o to be very large ($\approx 10^5$).

The measurement model is straightforward, since our observations are direct measurements of the $\{x(s)\}$ on the finest scale of the tree, i.e., C is a selection matrix. Every node s on the finest scale obeys $m(s) = M - 1$, thus

$$C(s) = \begin{cases} 0 & m(s) < M-1 \text{ or } x(s) \text{ does not correspond to an observation.} \\ 1 & m(s) = M-1 \text{ and } x(s) \text{ corresponds to an observation point.} \end{cases} \tag{3.4}$$

The final parameter that needs to be specified is the measurement noise variance $R(s)$.

In particular, in this study we will consider three sources of error in the measurement data:

1. The error in estimating the distance from the satellite to the ocean surface; this error is assumed to be 5cm white Gaussian noise[32].

2. The error in assuming that the ocean is a static surface. Specifically, the surface model (3.2) is static and does not account for any time evolution of the surface, even though the measurements $\{y(s)\}$ are taken over time.

3. The error in the geoid model, which manifests itself as an error in the geoid-corrected TOPEX/POSEIDON data.

Item (2.) in the above list may be addressed by an examination of the satellite crossover statistics; a crossover is a point of intersection of two satellite orbits. Let the set of crossover points be given by

$$\{(ya_i, yd_i, ta_i, td_i, p_i)\} \tag{3.5}$$

where ya_i, yd_i represent the measured ocean elevations at the same physical point p_i in space, but on ascending and descending satellite orbits, at times ta_i, td_i respectively. The difference $ya_i - yd_i$ is independent of any error in the geoid which is a function of space only; the difference is primarily due to the ocean elevation change $\zeta(p_i, ta_i) - \zeta(p_i, td_i)$ over the time interval $\Delta t_i = ta_i - td_i$. Figure 3-6 plots the empirical variance of $ya - yd$ as a function of the time difference $\mid \Delta t \mid$. Increases in this variance with $\mid \Delta t \mid$ quantify the degree to which the ocean measurements do not correspond to a static surface. From Figure 3-6 the variances are seen to be relatively independent of $\mid \Delta t \mid$; that is, the assumption that the ocean surface is static over periods up to ten days appears to be well justified for our region of interest in the north Pacific.

Item (3.) in the above list concerns errors in the geoid estimates. The highest quality geoid models currently available[79, 87] are quite effective at capturing large scale and moderate scale geoid fluctuations, but are less accurate in regions of sharp local changes. Such a result is not surprising: the OSU91A geoid model which we

use is constructed as a spherical harmonic expansion (truncated to order and degree 360); the truncated expansion can exhibit Gibbs-like phenomena near abrupt changes. Furthermore, navigation errors in the satellite lead to errors in registering satellite measurements with points on the earth, and thus in areas of steep geoid gradient, such registration errors lead to greater uncertainty in the geoid reference field than in other regions in which the geoid is slowly varying. As a result, altimetric measurements in the vicinity of steep geoid slopes are determined relative to a poor geoid reference and therefore represent a less accurate assessment of the ocean surface height. Consequently we have used the following measurement variance model:

$$R(s) = (5\text{cm})^2 + \phi(\text{Geoid Slope}) \tag{3.6}$$

where $\phi()$ is an increasing function (detailed in Section 3.4).

Finally, it is important to make a comment about one of the consequences of using a simple scalar version of our multiscale model. In particular, the spatial position of the multiscale tree on the ocean is somewhat arbitrary; that is, there is no particularly natural orientation for the multiscale tree. Consequently we will want to make sure that the estimates produced by our algorithm are insensitive to the precise positioning of the tree. However, consider a node s at a relatively coarse scale on the tree. Since the state at each node is a scalar, the correlation between the four children of node s, each of which represent the height of the ocean over a large area, is captured by only one degree of freedom. In particular, the finer scale decompositions of each of these four descendants proceed completely independently, and as a consequence artifacts may appear along coarse tree boundaries due to inadequate correlation. Techniques which attenuate such artifacts will be studied in some detail in Chapter 6; in this chapter we will use a relatively simple method that is adequate for our purposes. Specifically, we compute ocean surface estimates for each of ten tree positions (each shifted with respect to the others) and average the results. It should be made clear that this is not at all like spatial low-pass filtering or interpolation, as strong nonstationarities, such as in the quality of the data as measured by $R(s)$, are

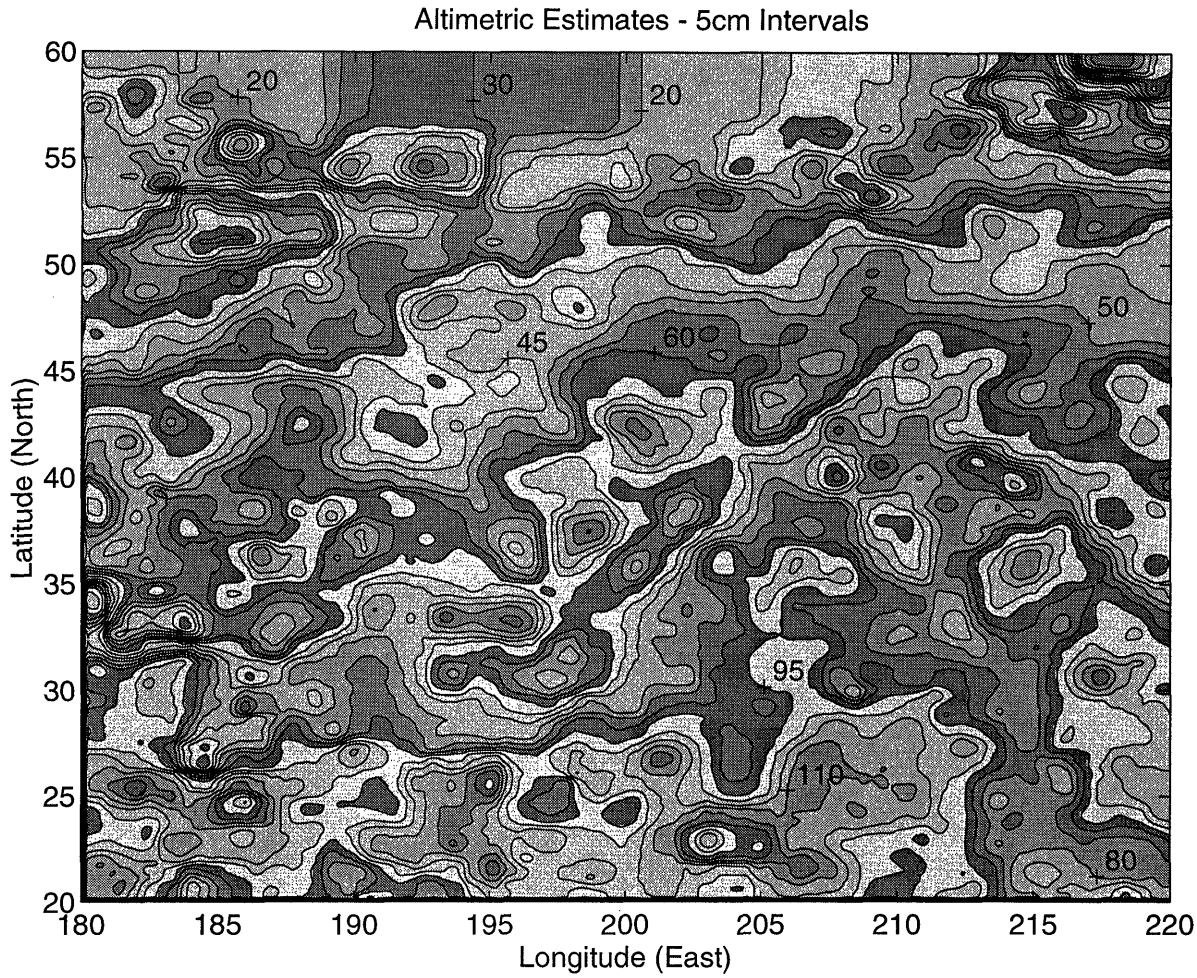


Figure 3-7: Estimates of the mean ocean elevation based on a single ten day set of data.

maintained.

3.3 Estimation Results

Given a collection of observations and the multiscale model as defined in the previous section, the multiscale estimation algorithm (detailed in Appendix B) permits rapid computation of multiscale estimates, estimation error variances, and measurement residuals.

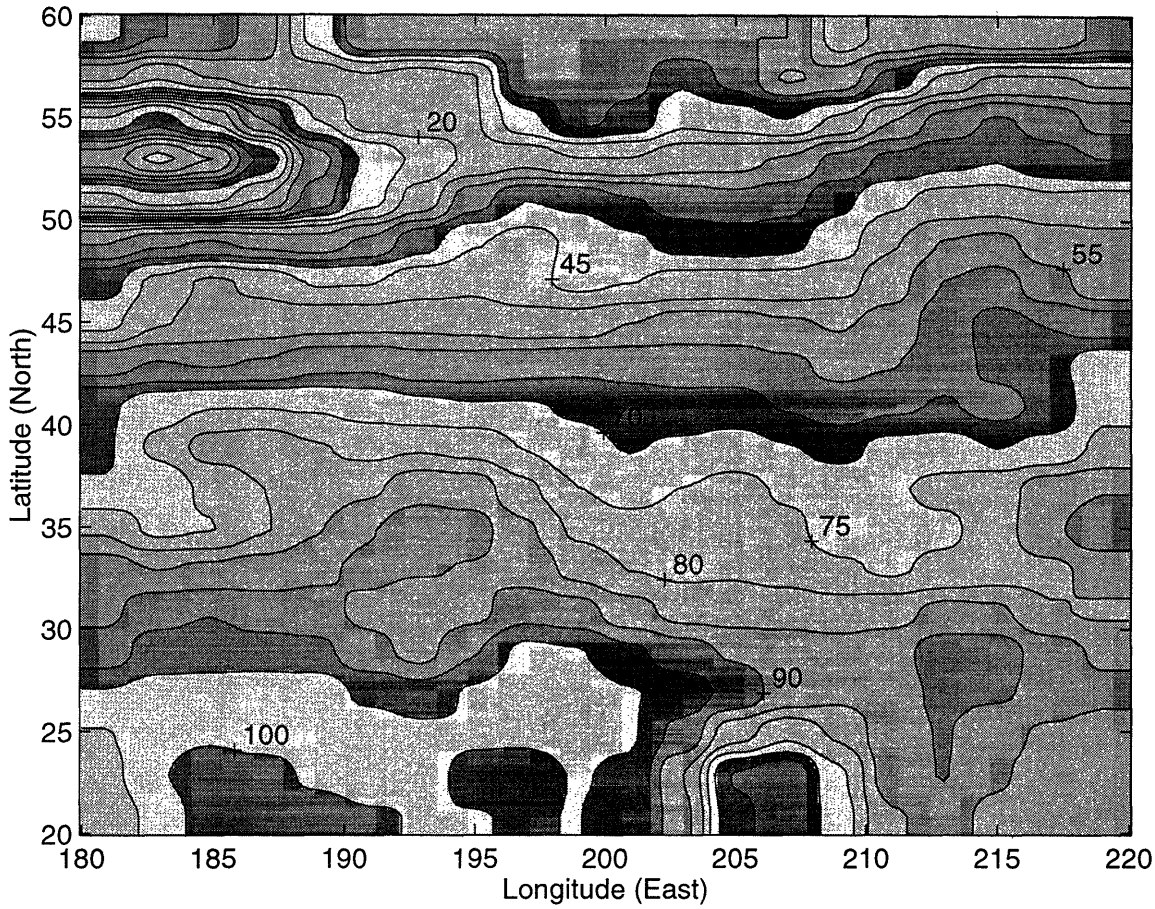


Figure 3-8: Typical example of objective mapping using standard oceanographic techniques, based on the same data set as in Figure 3-7.

3.3.1 Multiscale Estimates

A sample map of ocean surface estimates, taken from the finest scale of the tree, is shown in Figure 3-7. This map is based upon a single repeat cycle, or ten days, of data (about 20,000 data points). The 250,000 estimates and associated estimation covariance information were computed in less than one minute on a Sun Sparc-10 (the map is based on ten trees of estimates, each tree requiring 5 seconds of computation time). Although Figure 3-7 shows estimates on one scale only, the one minute of computer time produces estimates and error variances on all scales of the tree.

The ocean height variations shown in the figure are consistent with the known large-scale oceanographic behavior of the region (that is, a predominant gradient in the north-south direction with surface height offset on the order of one meter[52]).

Moreover, the estimates such as those shown in the figure offer far higher resolution than has heretofore been available (e.g., [12]). It is this very leap in resolution that makes the quantitative assessment of our results difficult - we have come across no other altimetric maps of sufficient resolution to compare with our plots. For example, Figure 3-8 [52] shows an ocean altimetric map for the same region of the ocean and the same period of time as we have considered. The figure, typical of the methods used by oceanographers, is based upon gridding followed by spatial filtering. Clearly a thorough validation of the enhanced resolution results provided by our method will require alternate methods such as integration with global circulation models, a problem that remains for the future. Nevertheless, the ability to produce such estimates efficiently is itself of significance.

3.3.2 Multiscale Error Variances

Estimation error variances corresponding to Figure 3-7 are shown in Figure 3-9. These values are based on the same ten day set of measurements as for the estimates just discussed; the distribution of measurement dropouts along the satellite tracks in this data set can be inferred from Figure 3-2. As before, the results are computed as the average over ten multiscale trees, still within the same one minute of computer time in which the estimates were computed.

Because of the spatially varying uncertainty in our measurements due to geoid model error, the occurrence of data dropouts, and the irregular pattern of data collection, we would expect that the uncertainty pattern in the optimal estimate of our ocean height map would be highly variable and would, to some extent, reflect these features. In particular, observe that the regions of lowest uncertainty (the lightly shaded regions in the figure) correspond to the points at which we have satellite measurements; a careful inspection of the figure will also reveal occasional darker breaks along these lines, corresponding to data dropouts. In addition, because of the spatially-varying noise model, the measurements near the Aleutian and Hawaiian chains (which induce a significant geoid gradient) are modeled as being noisier, resulting in elevated covariance values. The large region of uncertainty at the top of

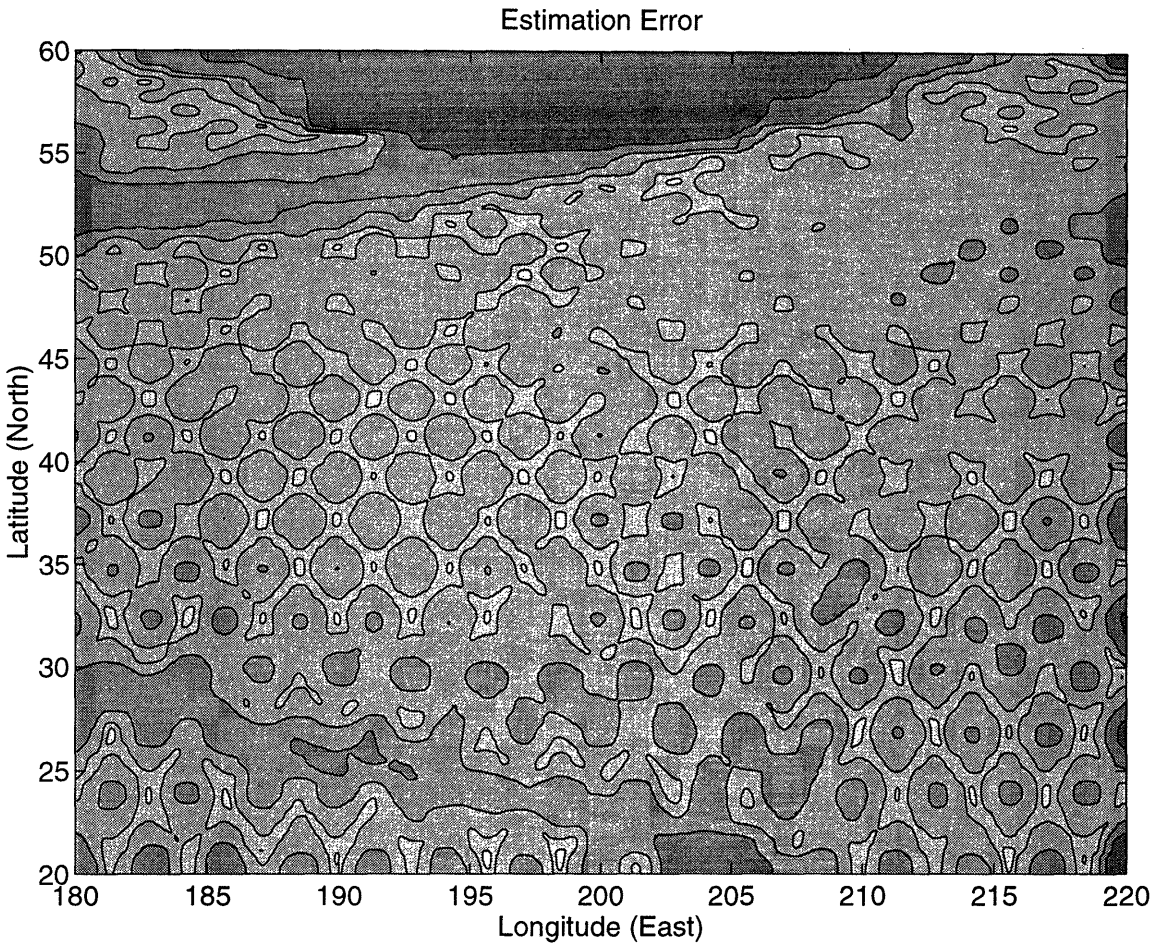


Figure 3-9: Estimation error variances based on one repeat cycle of data; darker regions represent greater uncertainty.

the figure is due to the Alaskan land mass.

Specific off-diagonal terms in the error covariance matrix may also be computed using (2.29) with equal computational ease (as compared to other approaches which would require the impractical calculation of the full error covariance matrix, containing $\approx 10^{10}$ elements). For example, by computing error covariances between a large ensemble of tree nodes (here 50,000 pairs of nodes, randomly positioned in longitude) one can determine averaged correlation coefficients of the estimation error as a function of longitudinal separation, as shown in Figure 3-10.

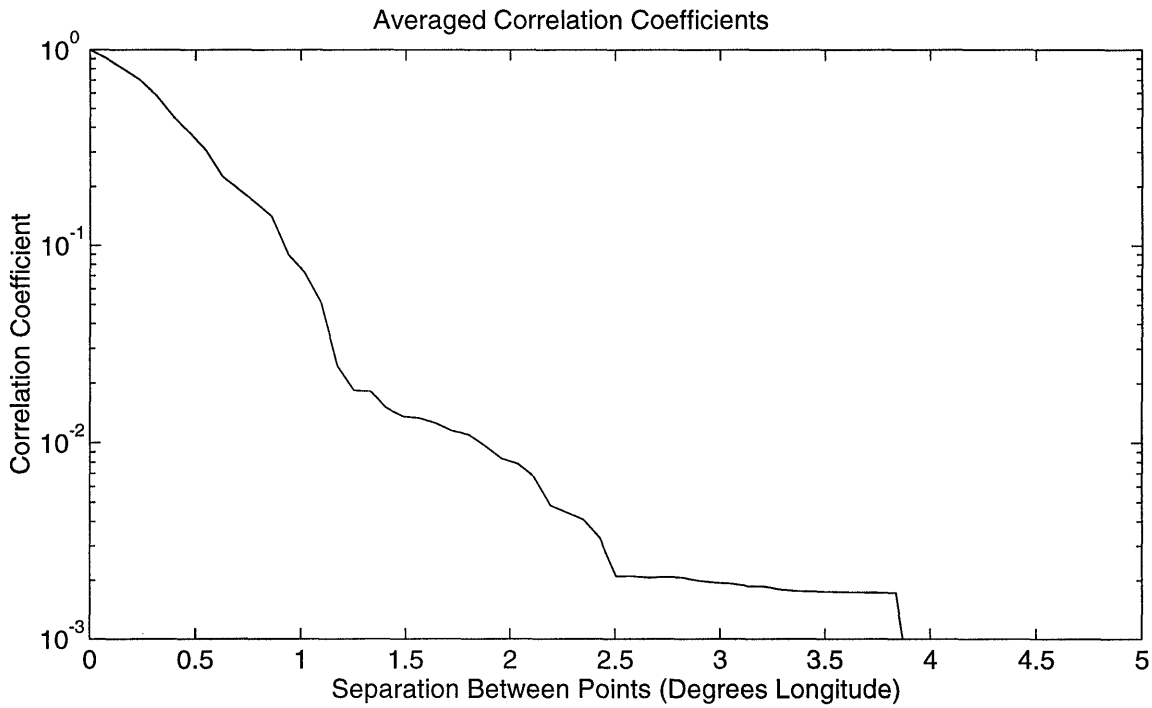


Figure 3-10: Off-diagonal elements of the estimation error covariance matrix can be computed. This figure shows a Monte-Carlo determination of the correlation coefficient of the estimation error of two points as a function their longitudinal separation.

3.3.3 Oceanographic Anomaly Estimates

The shape of the ocean surface, as was estimated in Figure 3-7, is characterized by a mean circulation shape upon which seasonal cycles and other variations (e.g., ocean eddies, which do not appear predictably as a function of season) are superimposed. Both the mean ocean shape and shorter period variations are of interest to oceanographers. However the shape of the ocean surface at any given point in time is dominated by the mean circulation shape; as a result oceanographers compute *anomalies*, which are the differences between the mean ocean surface and surface estimates based upon a short period of data.

Figure 3-11 shows a set of four such anomalies, computed as the difference between the estimated average ocean elevation (estimated using the model (3.2), but with a full year of satellite data) and ocean elevation estimates based upon relatively short time intervals. The four anomaly plots of Figure 3-11 are spaced 2.5 days apart in

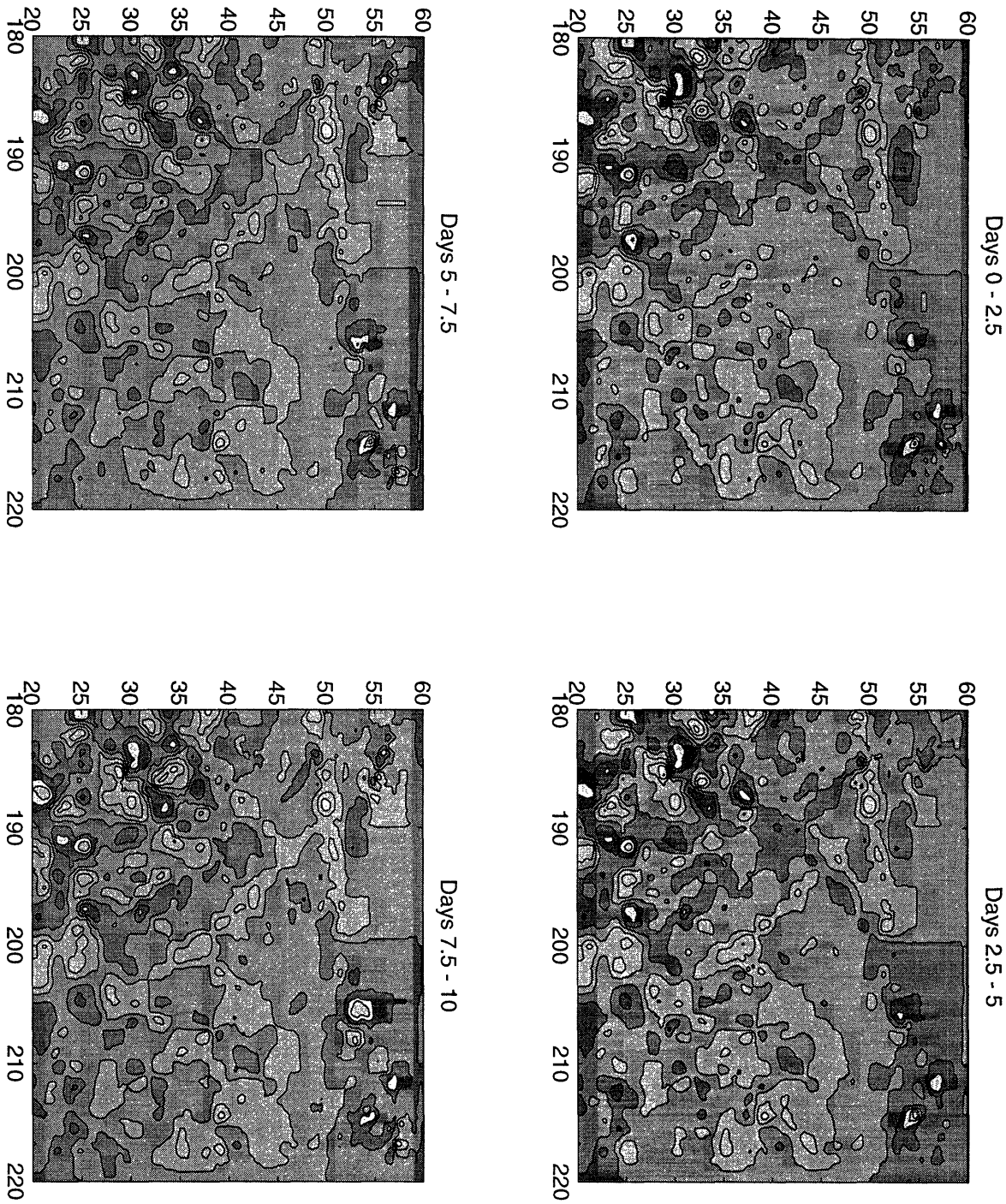


Figure 3-11: Collection of four anomaly plots: anomalies are calculated as the difference between ocean surface estimates based on a short window of data and the mean ocean surface elevation. Each of the four plots is based on a window of data about eight days in length; the window is shifted by 2.5 days from one plots to the next.

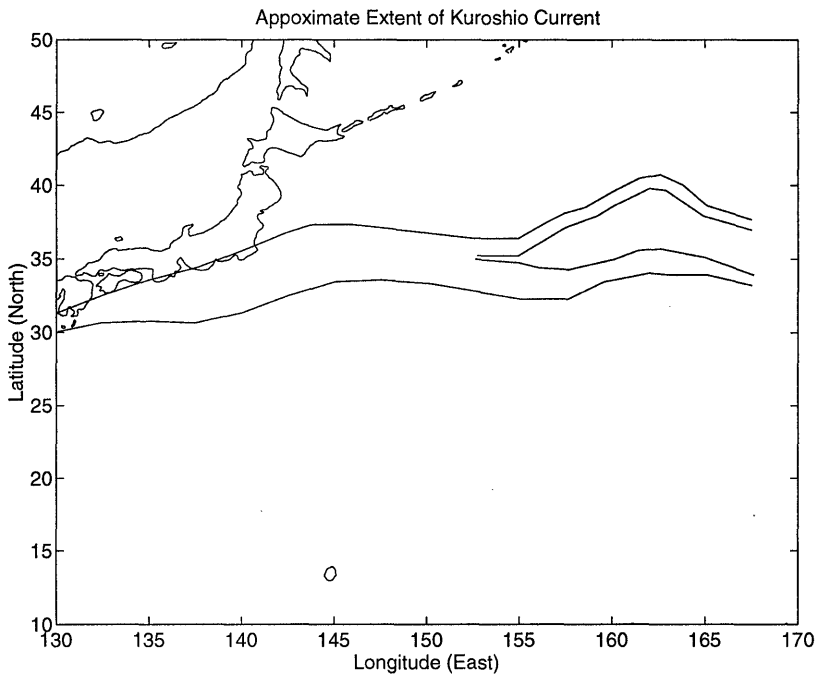


Figure 3-12: Approximate extent of the Kuroshio current off the coast of Japan.

time, each based on about eight days of data. The interpretation of such figures requires a good deal of knowledge and intuition regarding ocean dynamics. We can observe the evolution of a variety of features from one frame to the next, however a physical understanding or justification of these evolutions is well beyond the scope of this thesis.

3.3.4 Model Heterogeneities

One of the drawbacks with certain accelerated methods, such as those based on FFTs, is the need for stationarity or uniformity of the phenomenon being modeled. In contrast, our multiscale framework allows us to incorporate nonstationarities without sacrificing computational efficiency.

Consider, for example, the Kuroshio current in the northwest Pacific off the coast of Japan; the approximate extent of the current is illustrated in Figure 3-12. Due to the strength of this current, the gradient of the ocean surface in the neighborhood of the Kuroshio is approximately four times larger[97] than in relatively quiescent regions (the Pacific northeast, for example). To compensate for this effect, one can

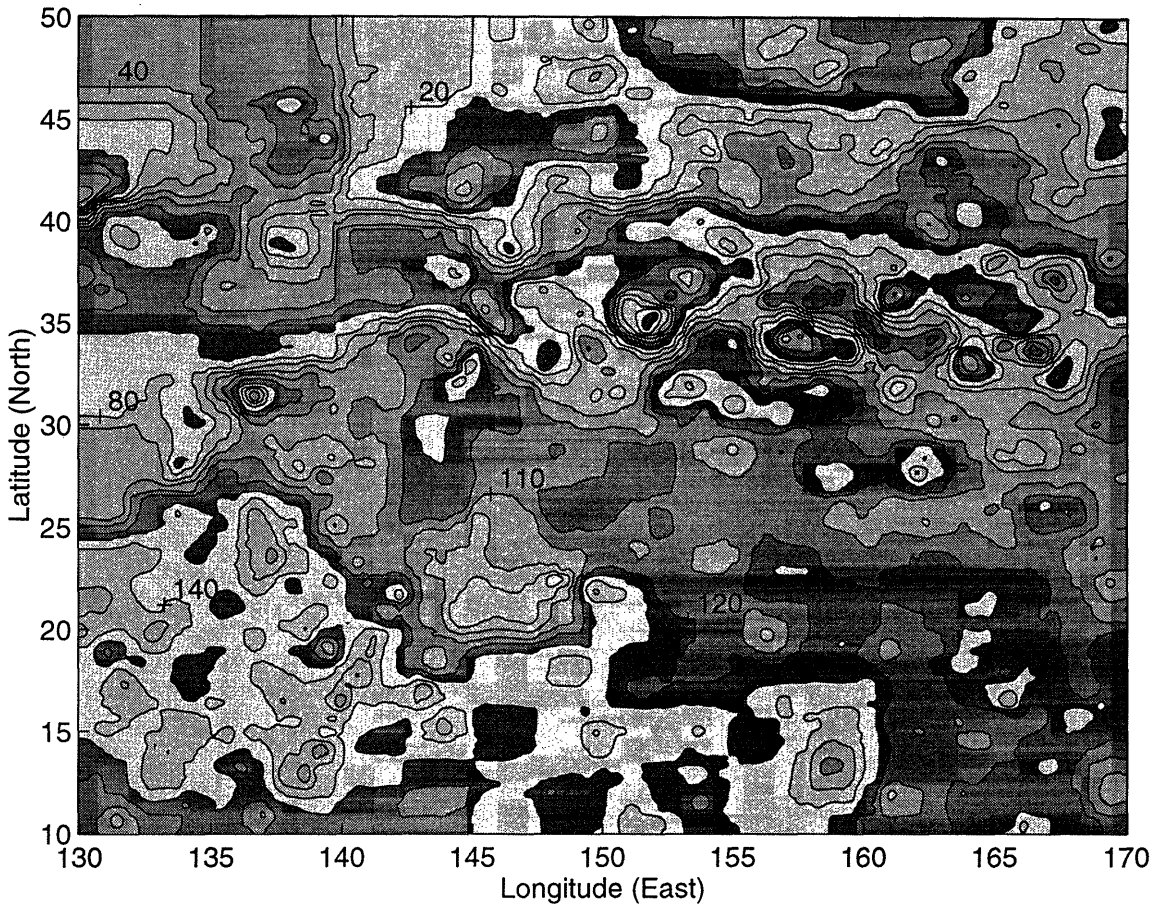


Figure 3-13: Estimates of ocean elevation (in cm) in the northwest Pacific using a nonstationary model which accounts for the increased surface gradients in the vicinity of the Kuroshio.

modify (3.1) by increasing those process noise values on those multiscale tree nodes which overlap part of the Kuroshio. Such a process noise is highly nonstationary, and by (3.3) implies a nonstationary prior covariance model. Since such adjustments to the process noise remain compatible with the multiscale framework of (2.26), (2.28), not only does our approach remain efficient in the face of such heterogeneities, but in fact the increase in computational burden over the homogeneous case is essentially nil.

Figures 3-13, 3-14 show estimates and error variances respectively for the northwest region of the Pacific, using a heterogeneous process noise model as detailed above. The distribution of the error variances shows the combined effects of irregular spatial sampling by the satellite, loss of satellite measurements over land (Japan),

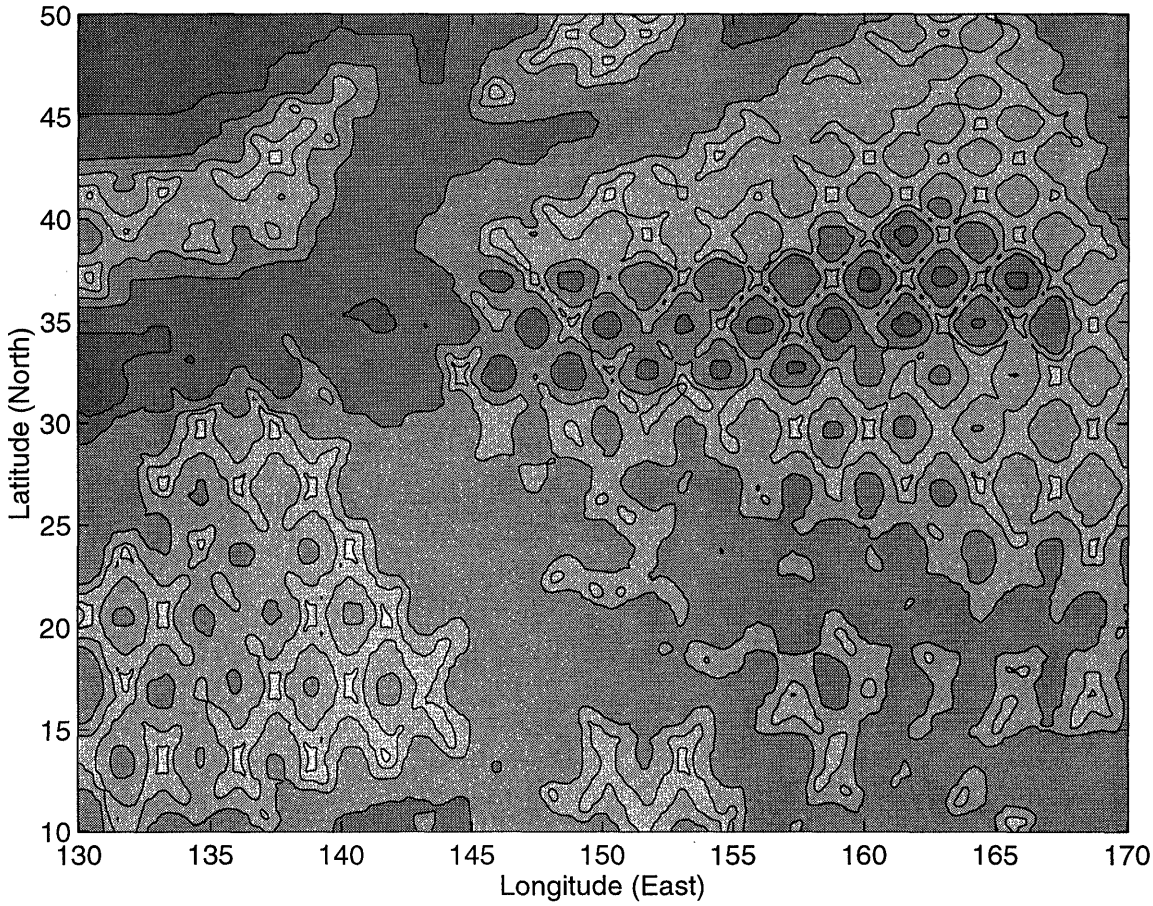
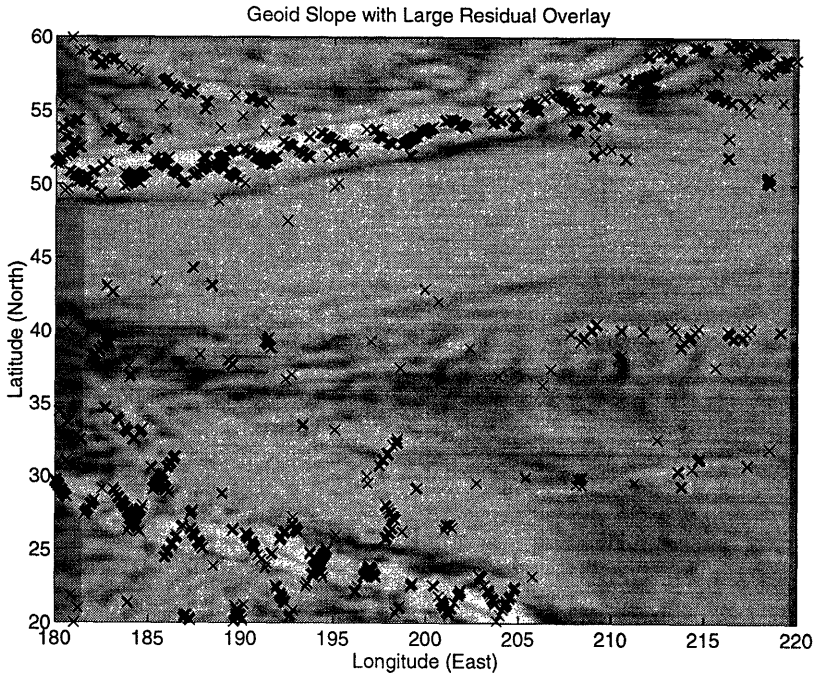


Figure 3-14: Estimation error variances corresponding to the elevation estimates of Figure 3-13. Darker regions represent greater uncertainty.

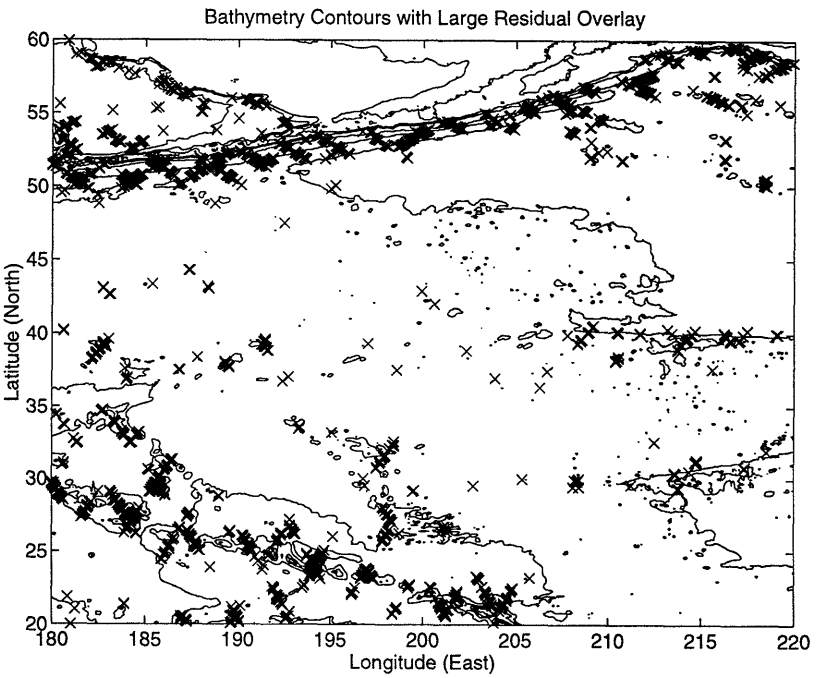
increased prior uncertainty over the Kuroshio, and nonstationary geoid-model error. For purposes of comparison, Figure 3-15 shows the differences in the altimetry estimates produced by multiscale models with and without Kuroshio compensation.

3.4 Calculation of Measurement Residuals

The examination of measurement residuals, the differences between measurements and the corresponding estimates, serves to test the validity of our multiscale models. In particular, by normalizing these residuals with respect to their expected standard deviations one can isolate statistically significant outliers. Such an approach may be used to argue the inclusion of the geoid slope dependent term in the measurement error (3.6). Figure 3-16 shows the distribution of statistically large residuals, calculated



(a)



(b)

Figure 3-16: Overlay of geoid gradient map (in (a)) and of ocean bathymetry contours (in (b)) with the distribution of locations of large residuals; regions of lighter shading represent steeper geoid gradient. The striking correlation exhibited in these figures motivates the correction of Figure 3-17.

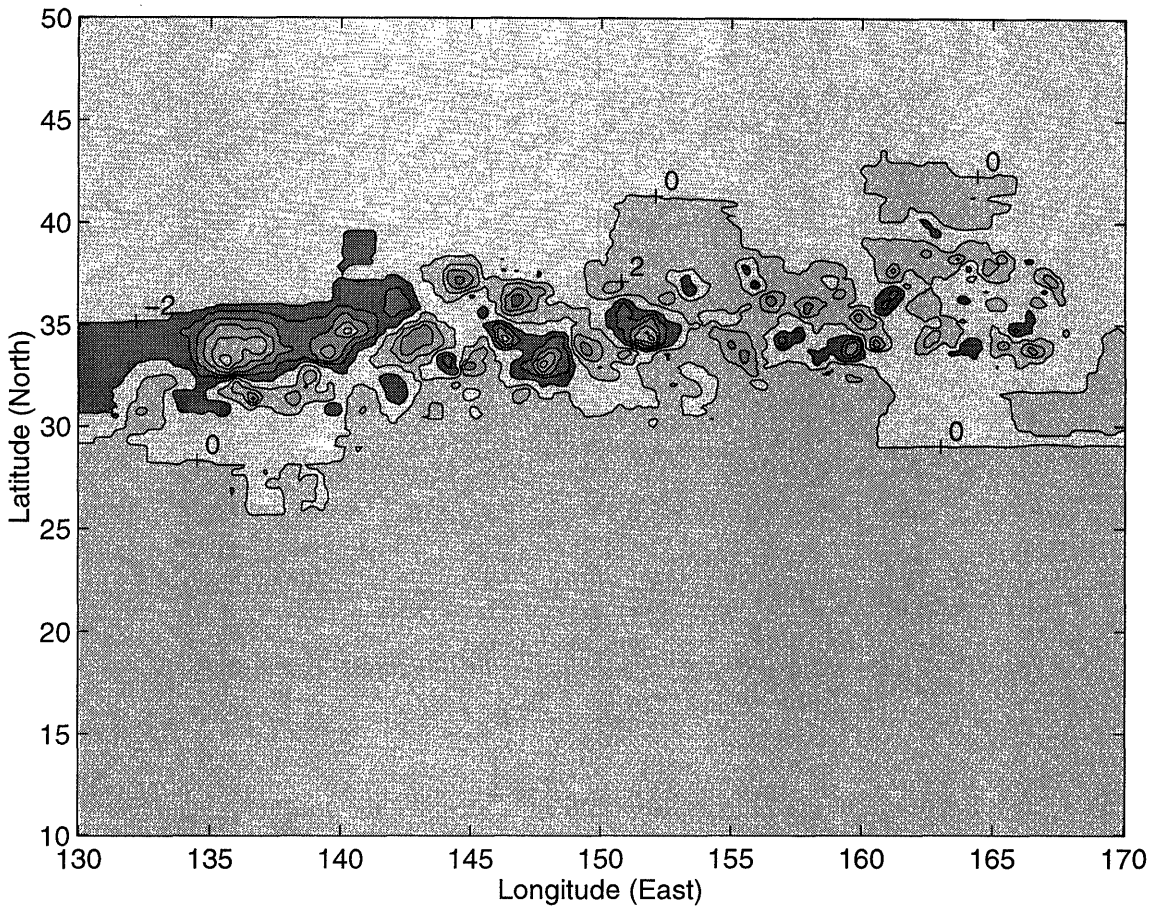


Figure 3-15: Differences (in cm) in the estimates produced by a homogeneous multi-scale model and a model accounting for the presence of the Kuroshio.

using a simple measurement noise model

$$R(s) = (5\text{cm})^2 \tag{3.7}$$

that is, a noise model which does not take any geoid model errors into account. Figure 3-16 also plots the geoid gradient; the correlation between significant residuals and steep geoid slope is unambiguous, and argues in favor of a geoid slope-corrected measurement noise model. As an additional comparison, the same locations of large residuals are shown superimposed on a plot of ocean bathymetry contours (the shape of the ocean bottom) in the bottom half of the figure. To the extent that bathymetry features are responsible for locally steep slopes in the geoid, the residual-bathymetry correlation does not come as a surprise. Such residual-geoid correlation immediately

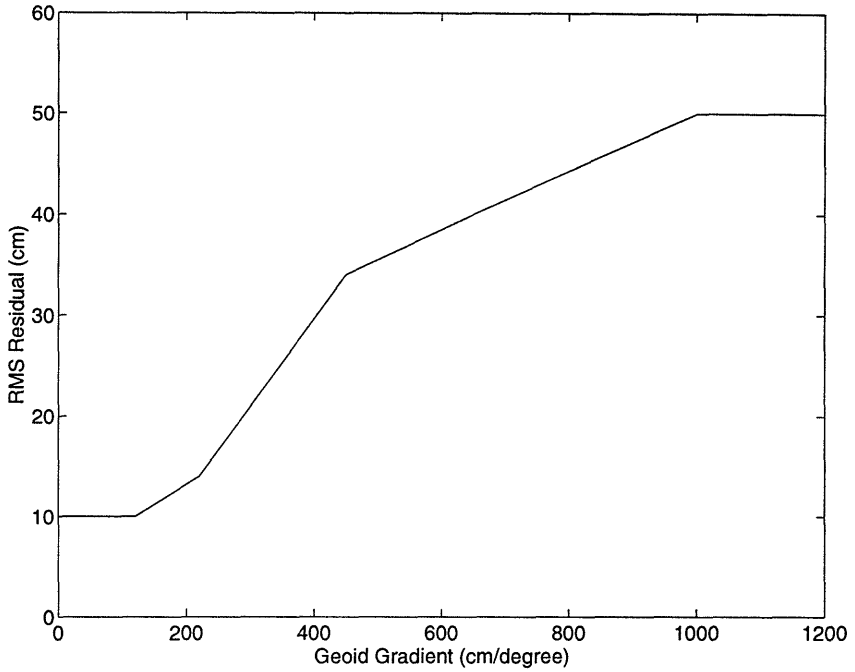


Figure 3-17: This figure shows a sketch of the dependence of root-mean-square value of measurement residuals as a function of the geoid gradient. This dependence is used as a basis for taking geoid errors into account.

motivates the development of models to estimate the geoid; the development of such a model is the subject of Chapter 5.

Figure 3-17 plots root mean square estimation residual magnitudes as a function of geoid slope. This figure leads to the form, shown in Figure 3-18, for the geoid-slope dependent term in the measurement noise model (function $\phi()$ of (3.6)) used for the other results in this section. Such a heterogeneous set of measurement noise variances may be used with no appreciable increase in computational burden (just as before, with the heterogeneous process noise model for the Kuroshio).

3.5 Conclusions

This chapter has demonstrated the application of the efficient multiscale estimation framework to the problem of ocean altimetry estimation based on irregularly sampled satellite measurements. A number of significant difficulties which have led to significant suboptimalities and approximations in many other estimation algorithms are

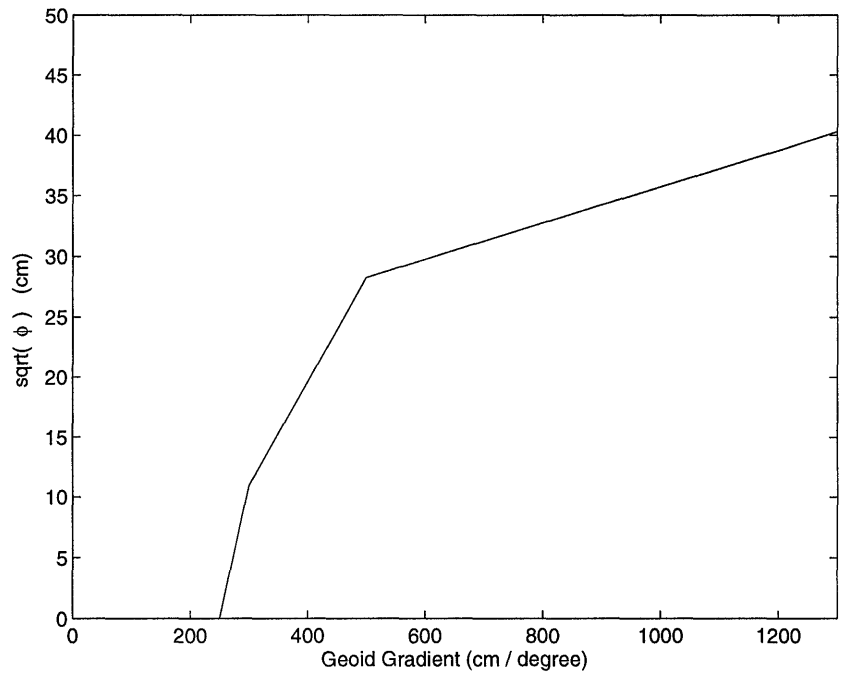


Figure 3-18: This figure plots the actual function $\phi()$ (see (3.6)) which was used throughout this chapter; the form of the function is based on Figure 3-17.

resolved by our approach: our multiscale framework possesses the efficiency to deal with truly enormous, possibly nonstationary, problems, computing both estimates and error variances with relative computational ease. Furthermore the concept of scale is made explicit, permitting the explicit characterization of phenomena possessing interactions across a number of scales.

Although throughout this chapter the ocean altimetry application has been used as a vehicle for demonstrating the use of the multiscale framework in such a modeling context, the success of the application motivates many further possible applications as well as extensions within the current context. With respect to the latter, we can point to several problems of considerable interest, including the following:

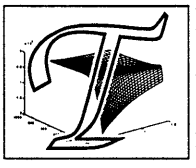
- The precise shape of power spectrum of the ocean remains a matter of current scientific interest. Multiscale likelihood methods[62] provide an efficient and statistically rigorous machinery for examining problems of identifying the statistical structure of random fields. The identification of parameters in $1/f$-like multiscale models such as (3.2) will be explored in Chapter 4.

- The distribution of measurement residuals (Figure 3-16) demonstrates clearly the presence of geoid error as well as suggesting a way in which to correct for it and thus provide local corrections to our estimate of the geoid. In particular, it is possible that *joint* estimation of the geoid and ocean height may simultaneously improve estimates of both of these quantities. This extension forms the subject of Chapter 5.

- There are a number of extensions of our multiscale modeling framework in the development of higher-order methods for estimating both surface height and surface gradients (a problem of independent interest in surface reconstruction problems in computer vision); such multiscale models will be developed in Chapter 7.

Chapter 4

Multiscale Parameter Identification



his chapter considers problems of multiscale system identification. After the introduction, Section 4.2 demonstrates the application of the multiscale framework to the problem of estimating the fractal dimension of a random process, and Section 4.3 deals with the development of a bound on the uncertainty of estimated model parameters for a particular class of $1/f$-like multiscale models.

4.1 Introduction

There are two basic steps in developing a multiscale model for an unknown random process:

1. the selection of a parameterized multiscale model (e.g., $1/f$-like models (Section 2.5.3), multiscale Gauss-Markov counterparts (Section 2.5.1), etc.),

2. the quantitative determination of the unknown parameters within the selected model.

The selection of a parameterized multiscale model may be accomplished based on a physical or intuitive understanding of the random process in question, or it may be

accomplished using automated model generation techniques[47] if certain covariance properties of the process are known. In any event, given a parameterized model the determination of its parameters forms the subject of system identification[60] and will be explored in this chapter.

Consider a parameterized multiscale model class, e.g.,

$$x(s) = A(s,H)x(s\bar{\gamma}) + B(s,H)w(s) \qquad w(s) \sim \mathcal{N}(0,I) \qquad \forall s \in \mathcal{T}, s \neq s_o \quad (4.1)$$

$$y(s) = C(s,H)x(s) + v(s) \qquad v(s) \sim \mathcal{N}(0,R(s,H)) \qquad \forall s \in \mathcal{M} \subseteq \mathcal{T} \quad (4.2)$$

which is similar to the basic multiscale equations (2.26),(2.28), except that (4.1),(4.2) are parameterized by a vector H, and the noise terms w, v are assumed to be Gaussian (to allow likelihood statistics to be calculated). In principle, given a set of observations $\{y(s), s \in \mathcal{M}\}$, an estimator for the vector H may be written abstractly as

$$\hat{H} = \arg_H \max \; \mathcal{L}\Big[A(s,H), B(s,H), C(s,H), R(s,H), y(s)\Big] \quad (4.3)$$

The remainder of this chapter will explore the application of the abstract parameter identification scheme of (4.3) in two contexts:

1. In Section 4.2 we will demonstrate the estimation of the fractal dimension of fractional Brownian motion[69] processes. We will present the development of a new $1/f$-like multiscale model which gives unbiased estimates of the fractal dimension, and will compare its performance with other proposed estimators[49, 114].

2. In Section 4.3 we will postulate a class of $1/f$-like multiscale models having two free parameters. This model class is of significant interest, since it is used in the remote sensing applications of Chapters 3 and 5. We will derive a Cramer-Rao bound for the maximum-likelihood estimation of the parameters for this class of multiscale models, and compare the bound with Monte-Carlo simulations and with tests on remote-sensing data.

4.2 Fractal Dimension Identification

Fractional Brownian motion (fBm) processes[69] were introduced briefly in Section 2.5.3. The second order (nonstationary) statistics of such processes are given by

$$E\left[F[k], F[m]\right] = \frac{\sigma^2}{2}(\Delta t)^{2H}\left(|k|^{2H} + |m|^{2H} - |k-m|^{2H}\right) \tag{4.4}$$

where σ and H are scalar parameters which completely characterize the process. The quantity H, which we wish to estimate, determines the fractal dimension $(2-H)$ of the process; the power σ is assumed to be known.

The exact maximum likelihood (ML) calculation for the fractal dimension of fBm is computationally difficult (see [105]); fractal estimators have been developed based on optimal algorithms applied to fBm-like $1/f$ models[114], and based on approximate or suboptimal algorithms developed directly from the fBm model[49, 105, 29]. Our philosophy in multiscale statistical modeling falls into the former class: the statistical self-similarity of fBm processes motivates us to develop a multiscale fBm counterpart, parameterized by H which will allow us to use our statistically-optimal multiscale likelihood calculation algorithm to estimate H in a manner similar to (4.3).

The approach in [114] is based on a $1/f$ process constructed using a wavelet basis in which the wavelet coefficients are independent, with variances that vary geometrically with scale with an exponent equal to H. Geometrically varying variances are a poor approximation to the fBm statistics for low H and lead to a biased estimator. The method in [49] determines the exact statistics of the Haar wavelet coefficients of the discrete fractional Gaussian noise (DFGN) process $D[k] = F[k+1] - F[k]$ and then develops an estimator by assuming, with some approximation, that the coefficients are uncorrelated; the use of the DFGN process D limits the estimator to those cases in which the fBm process is densely and uniformly sampled.

The goal of the research of this section, on the other hand, is the development of a fast estimator for H that functions under a broader variety of measurement circumstances, for example in the presence of gaps in the measured sequence, mea-

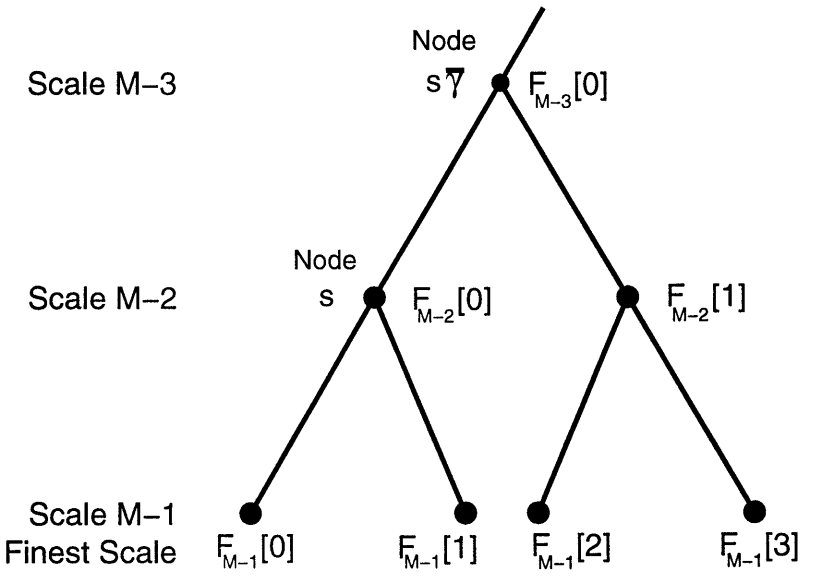


Figure 4-1: Dyadic tree structure used for the fractal Brownian motion estimator.

surement noise having a time-varying variance and higher dimensional processes (e.g., 2-D random fields). The basis for accomplishing this is the utilization of our multiscale framework, in particular the multiscale likelihood calculation algorithm[65]. The next section develops the multiscale estimator, followed by a description of estimation results.

4.2.1 Fractal Model Development

The statistical self-similarity of fBm makes the application of wavelets to the fBm process a logical choice. Kaplan and Kuo[49] apply the Haar wavelet to the incremental process $D[k]$, and Wornell and Oppenheim[114] apply higher order Daubechies wavelets to $F[k]$. We propose to use the multiscale framework to develop a Haar wavelet multiscale stochastic model which applies directly to $F[k]$. This choice of wavelet is motivated by the particularly simple realization of the Haar wavelet within our framework by using a dyadic tree structure (see Figure 4-1):

$$\text{Coarse Scales}: \quad \begin{cases} x(s) = \begin{bmatrix} 1 & +1 \\ 0 & 0 \end{bmatrix} x(s\bar{\gamma}) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} B(s,H)w(s) & s = s\bar{\gamma}\alpha_1 \\ x(s) = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} x(s\bar{\gamma}) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} B(s,H)w(s) & s = s\bar{\gamma}\alpha_2 \end{cases} \tag{4.5}$$

$$\text{Finest Scale}: \quad \begin{cases} x(s) = \begin{bmatrix} 1 & +1 \end{bmatrix} x(s\bar{\gamma}) + 0 \cdot w(s) & s = s\bar{\gamma}\alpha_1 \\ x(s) = \begin{bmatrix} 1 & -1 \end{bmatrix} x(s\bar{\gamma}) + 0 \cdot w(s) & s = s\bar{\gamma}\alpha_2 \\ y(s) = x(s) + v(s) \end{cases} \tag{4.6}$$

where $s = s\bar{\gamma}\alpha_1$ implies that s is the left descendant of its parent, similarly $s = s\bar{\gamma}\alpha_2$ the right. At coarse scales $x(s)$ consists of two scalars: a coarse approximation to the $1/f$ process, and a detail coefficient. The detail coefficient equals the difference in the coarse $1/f$-like representation between node s and its two children, where the sign of this difference depends on the parity of the child (i.e., left vs. right). At the finest scale $x(s)$ is a single scalar, representing a sample of a $1/f$-like process, and measurements of the actual fBm sequence appear as observations $y(s)$ at the finest scale. Note that the $1/f$ process described by (4.5),(4.6) does not yield a finest scale process that is exactly an fBm process (and thus, as with the technique in [114], our model does not exactly match the statistics of the process to be estimated). However by an appropriate choice of the remaining model parameters we can produce a process with a similar type of $1/f$ behavior.

The elements which remain to be determined in the above multiscale model are the $B(s, H)$: the standard deviation of the detail wavelet coefficients between node s and its children. Expressions for the statistics of the wavelet decomposition of fBm have been determined by others [29, 104], however the self-statistics for the special case of the Haar wavelet are easily computed as follows:

- Let $F_{M-1}[k] = F[k]$ which is the fBm process of interest.
- Define $F_{M-1-m}[k]$ as the process obtained by coarsening $F[k]$ m times, such that

$$F_m[k] = \left(F_{m+1}[2k] + F_{m+1}[2k+1]\right)/2 \tag{4.7}$$

  a relation which follows from the multiscale model of (4.5).

- From (4.7) it follows that

$$F_{M-m}[k] = \sum_{i=0}^{2^{m-1}-1} \frac{F\left[2^{m-1}k+i\right]}{2^{m-1}} \tag{4.8}$$

$$F_{M-m}[k+1] - F_{M-m}[k] = \sum_{i=0}^{2(2^{m-1}-1)} D\left[2^{m-1}k+i\right] \frac{2^{m-1} - |2^{m-1}-i-1|}{2^{m-1}} \tag{4.9}$$

$$\equiv \sum_{i=0}^{2(2^{m-1}-1)} D\left[2^{m-1}k+i\right] c_i \tag{4.10}$$

- From the stationarity of the increments process $D[k]$, and from the symmetry

$$F_m[k] - F_{m+1}[2k] = -\left\{F_m[k] - F_{m+1}[2k+1]\right\} \tag{4.11}$$

(4.10) reduces to the desired variance expression

$$E\left[\left(F_{M-m-1}[k] - F_{M-m}[2k]\right)^2\right] = \frac{1}{4} \sum_{i=-2(2^{m-1}-1)}^{2(2^{m-1}-1)} \lambda_{\mathrm{D}}[i] \sum_{j=-\min(0,i)}^{2(2^{m-1}-1)+\min(0,-i)} c_i c_{i+j} = B^2(s,H) \tag{4.12}$$

where s is any node on scale $(M-m-1)$ of the tree, and where λ_{D} is the covariance function of $D[k]$:

$$\lambda_{\mathrm{D}}[i] = \frac{\sigma^2}{2}\left[|i+1|^{2H} + |i-1|^{2H} - 2|i|^{2H}\right] \tag{4.13}$$

By way of comparison, it has been shown[114] that $1/f$ processes, of which fBm is a subset, may be approximated by wavelet synthesis in which the wavelet coefficient variances are an exponential function of scale:

$$B^2(s,H) = \beta 2^{-2Hm(s)} \qquad \text{i.e., } \log_2 \frac{B(s\bar{\gamma},H)}{B(s,H)} = H \tag{4.14}$$

Table 4.1 shows the scale to scale variance ratios as predicted by (4.12). The deviation from the approximate scaling law of (4.14) is most pronounced at low H; it is this

Variance Ratio:	$H = 0.25$	$H = 0.50$	$H = 0.75$	$H = 0.9$
$\log_2(B_0/B_1)$	0.250			
$\log_2(B_1/B_2)$	0.249	⋮		
$\log_2(B_2/B_3)$	0.247	0.500	⋮	
$\log_2(B_3/B_4)$	0.242	0.499	0.750	⋮
$\log_2(B_4/B_5)$	0.228	0.496	0.749	0.900
$\log_2(B_5/B_6)$	0.188	0.484	0.745	0.898
$\log_2(B_6/B_7)$	0.091	0.437	0.727	0.892
$\log_2(B_7/B_8)$	-0.084	0.292	0.650	0.861

Table 4.1: Haar wavelet coefficient standard-deviation ratios as a function of H and scale: B_m^2 represents the wavelet coefficient variance at scale m, where the coarsest scale is $m = 0$. The deviation of the variance progression from an exponential law is most pronounced at fine scales and for low values of H.

Variance Rule:		$H = 0.25$	$H = 0.50$	$H = 0.75$	$H = 0.90$
Variances assumed exponential with scale	$\hat{H}$:	0.05	0.40	0.70	0.91
Variances based on exact result	$\hat{H}$:	0.24	0.51	0.75	0.92

Table 4.2: The estimator in the top row is based on the premise that wavelet coefficient variances are exponentially distributed, leading to a biased estimate $\hat{H}$.

deviation which leads to a bias for those estimators based on (4.14), as shown in Table 4.2.

Our proposed estimator for H is based on the multiscale model of (4.5),(4.6), where the process noise terms $B(s, H)$ are given by (4.12). The estimator is implemented as outlined in (4.3), in which the likelihood maximization is performed using standard nonlinear techniques (e.g., the section search method of MATLAB).

4.2.2 Fractal Identification Results

Sixty four fBm sample paths, each having a length of 2048 samples, were generated using the Cholesky decomposition method of [66], precisely the same approach as in

Estimator		$H = 0.25$	$H = 0.50$	$H = 0.75$	$H = 0.90$
W.O.	$\hat{H}$	0.082	0.398	0.683	0.846
[114]	$\sigma_{\hat{H}}$	0.022	0.021	0.026	0.021
	$(H - \hat{H})_{\text{RMS}}$	0.169	0.109	0.072	0.058
K.K.	$\hat{H}$	0.252	0.499	0.748	0.899
[49]	$\sigma_{\hat{H}}$	0.017	0.017	0.017	0.017
	$(H - \hat{H})_{\text{RMS}}$	0.017	0.017	0.017	0.017
Multiscale	$\hat{H}$	0.249	0.503	0.768	0.919
Haar	$\sigma_{\hat{H}}$	0.011	0.019	0.050	0.109
	$(H - \hat{H})_{\text{RMS}}$	0.011	0.019	0.054	0.110

Table 4.3: Estimation results for three estimators, based on 64 fBm sample paths, each of length 2048 samples, with no additive noise. The experimental results for the first two estimators are from [49].

Kaplan and Kuo[49] whose experimental results form the basis of comparison with ours. The process power $\sigma^2 = 1$, and σ is assumed to be known.

The performance of three fBm estimators is compared in Table 4.3. An evident bias is present for low H in the estimator of Wornell and Oppenheim[114] as was argued earlier based on Table 4.1. Also recall that the multiscale model of (4.5),(4.6) assumed the wavelet detail coefficients to be uncorrelated; this assumption becomes progressively poorer as H increases[29], leading to an increase in the error variance for our estimator at large H. Nevertheless, our method still performs reasonably well over quite a wide range of values of H. Moreover, using the techniques developed in [61], we can construct higher-order multiscale models which account for most of the residual correlation in the wavelet coefficients. The same likelihood procedure applied to these higher-order models would then yield even better results, closely approaching the exact ML estimator based on the exact fBm statistics. However since fBm itself is an idealization, the benefit in practice of such higher-order models over that based on the low-order model (4.5),(4.6) depends upon the application.

Finally, as we have said, our approach applies equally well in a variety of other settings. For example, Table 4.4 illustrates the performance of our estimator under non-uniform sampling (by randomly discarding 10% of the measurements), and under

Circumstance		$H = 0.25$	$H = 0.50$	$H = 0.75$	$H = 0.90$
Irregular	$\hat{H}$	0.246	0.507	0.781	0.937
Sampling	$\sigma_{\hat{H}}$	0.033	0.044	0.076	0.124
	$(H - \hat{H})_{\text{RMS}}$	0.033	0.045	0.082	0.128
Nonstationary	$\hat{H}$	0.268	0.511	0.769	0.918
Measurement	$\sigma_{\hat{H}}$	0.011	0.019	0.051	0.109
Noise Variance	$(H - \hat{H})_{\text{RMS}}$	0.021	0.022	0.054	0.110

Table 4.4: Performance of the fBm estimator for two examples: irregular sampling (removal, at random, of 10% of the measurements), and nonstationary measurement noise variance (noise standard deviation $\sigma[k] = \frac{1}{2}\exp\{-[(k-1024)/500]^2\}$). In both cases the results are based on 64 fBm sample paths, each of length 2048 samples.

a varying measurement noise variance. Both of these special cases are accomplished with essentially no change in the algorithm (as opposed to the estimators of [114, 49]). In addition, by using a quadtree rather than a dyadic tree we can also apply these techniques in 2-D. An example of estimating non-isotropic fractal parameters for a 2-D random field based on irregular, nonstationary data will be presented in Section 4.3.4.

4.3 Scalar Model Identification

The previous section considered the estimation of the fractal dimension, motivated by a fairly specific class of processes (i.e., fBm). In this section we will consider a more generic $1/f$ model, motivated by the discussion in Section 2.5.3 and by the use of this model in remote sensing applications (Chapter 3, Chapter 5). The canonical $1/f$-like multiscale model is as follows:

$$x(s) = x(s\bar{\gamma}) + B\gamma^{m(s)}w(s) \qquad \begin{array}{c} x(0) \sim \mathcal{N}(0, P_o) \\ w(s) \sim \mathcal{N}(0,1) \end{array} \tag{4.15}$$

$$y(s) = x(s) + v(s) \qquad v(s) \sim \mathcal{N}(0, \sigma^2) \qquad m(s) = M - 1 \tag{4.16}$$

i.e., measurements are made at the finest level of the tree only.

Figure 4-2 shows a contour plot of the model likelihood, as a function of B and

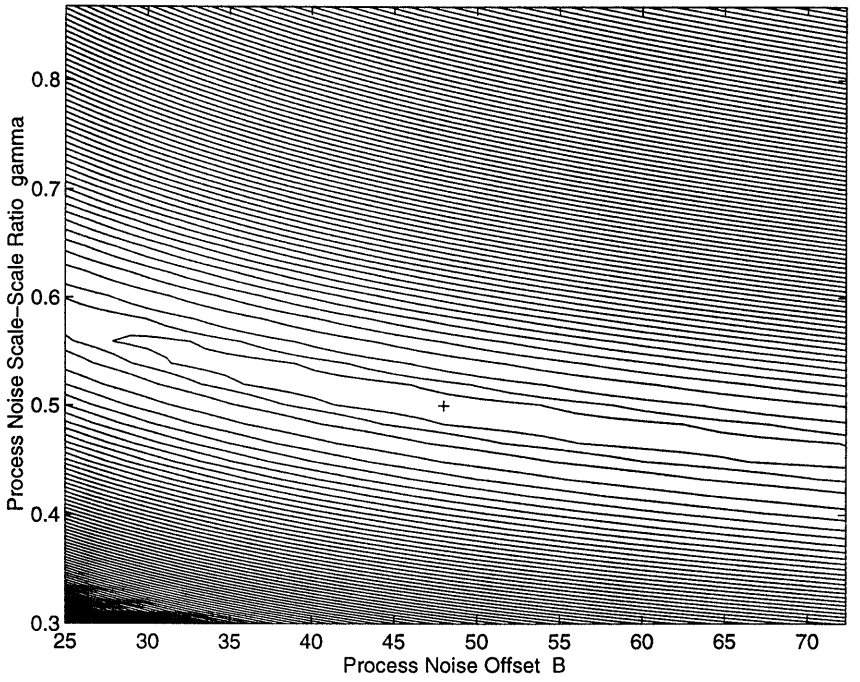


Figure 4-2: This figure presents a contour representation of the likelihood surface of the canonical model (4.15) based on synthesized data measurements on the finest scale. The likelihood surface is maximized at $B = 48, \gamma = 0.5$, marked by a '+', however the nearest likelihood contour, which is lower than the maximum by only 0.2%, extends for a very broad range of (B, γ) values.

γ, given a set of measurement data synthesized using the multiscale model of (4.15). This figure possesses one curious feature: the portion of the likelihood surface within 0.2% of the peak (at $B = 48, \gamma = 0.5$) is a ridge which extends over a considerable range of B and γ. It is figures such as this which motivated the exploration of Cramer-Rao bounds for the estimation of B and γ to acquire a deeper understanding of the associated likelihood space.

4.3.1 Cramer-Rao Bound

Let $\boldsymbol{y}$ represent the collection of observations at the finest scale of the tree and Λ_y the corresponding covariance matrix. Then the log-likelihood is given by

$$\ln\left\{p_{\boldsymbol{y}}(Y)\right\} = -\tfrac{1}{2}\ln\left\{(2\pi)^{\|\boldsymbol{y}\|}|\Lambda_y|\right\} + Y^T\Lambda_y^{-1}Y \tag{4.17}$$

where $\|\boldsymbol{y}\|$ represents the number of elements in vector $\boldsymbol{y}$. Furthermore, if we let χ represent the collection of finest scale states $x(s)$ at which we have measurements, then

$$\Lambda_y = \Lambda_\chi + \Lambda_v \tag{4.18}$$

where Λ_v is diagonal, but where Λ_χ possesses a great deal of structure (from the multiscale structure of the tree, that is, from (4.15)). Although analytical expressions for the determinant $|\Lambda_y|$ are readily derived[65], closed-form expressions for $|\Lambda_y|$ are very complicated and contribute little insight into the nature of (4.17).

Suppose that, in addition to the noisy measurements available on a subset of the finest scale, we assume the availability of *noiseless* measurements on *all* coarser-scale nodes of the tree. Although this may appear to be a poor assumption, the supposition possesses three motivations:

1. The effect of the assumption is to add new measurements only, thus the Cramer-Rao bound will still provide a *strict* lower bound on the estimation error variance for the multiscale model parameters.

2. The problem becomes analytically tractable.

3. Recall that on a densely-sampled quadtree (i.e., $q = 4$) 75 percent of the observations are located on the finest scale, so it does not necessarily follow that the introduction of coarser-scale, noiseless measurements will significantly weaken the Cramer-Rao bound on the estimation of multiscale model parameters.

Let us first consider the case in which the finest scale of the tree is densely sampled with noisy measurements; the sparsely sampled case will be touched upon briefly after the Cramer-Rao bound derivation. The introduction of noiseless coarser measurements leads to an easily diagonalized determinant, making the CRB calculations trivial. Let

$$\nu(s) = y(s) - y(s\bar{\gamma}), \tag{4.19}$$

then the statistics on ν follow simply from our assumption:

$$\nu(s) \sim \mathcal{N}(0, B^2\gamma^{2i} + \sigma_i^2) \sim \mathcal{N}(0, \beta\zeta^i + \sigma_i^2) \quad \sigma_i = \begin{cases} 0 & i < M-2 \\ \sigma & i = M-2 \end{cases} \quad i = m(s) - 1 \tag{4.20}$$

where σ was defined in (4.16), and where B, γ are reparameterized in terms of β, ζ for convenience.

Since noiseless measurements are available at all nodes except on the finest scale, the random variables $\boldsymbol{\nu} = \{\nu(s)\}$ are all independent; i.e., Λ_ν is diagonal. So

$$P_{\boldsymbol{\nu}|\beta,\zeta} = \prod_{i=0}^{M-2} \frac{1}{(2\pi)^{\frac{q^{i+1}}{2}} \prod_{m(s)=i+1} \sqrt{\beta\zeta^i + \sigma_i^2}} \exp\left(- \sum_{m(s)=i+1} \frac{\nu(s)^2}{2(\beta\zeta^i + \sigma_i^2)} \right) \tag{4.21}$$

$$\ln P_{\boldsymbol{\nu}|\beta,\zeta} = -\ln\Big(\prod_{i=0}^{M-2} (2\pi)^{\frac{q^{i+1}}{2}}\Big) - \sum_{i=0}^{M-2} \sum_{m(s)=i+1} \tfrac{1}{2} \ln(\beta\zeta^i + \sigma_i^2) - \sum_{i=0}^{M-2} \sum_{m(s)=i+1} \frac{(\nu(s)^2}{2(\beta\zeta^i + \sigma_i^2)} \tag{4.22}$$

From here the required derivatives are easily found, leading to the following Cramer-Rao bounds:

$$\frac{1}{E\left[\tilde{\beta}_{\mathrm{ML}}^2\right]} \leq -E\left[\frac{\partial^2 \ln P_{\boldsymbol{\nu}|\beta,\zeta}}{\partial\beta^2}\right] = \sum_{i=0}^{M-2} \frac{\zeta^{2i}}{(\beta\zeta^i + \sigma_i^2)^2} \frac{q^{i+1}}{2} \tag{4.23}$$

$$\frac{1}{E\left[\tilde{\beta}_{\mathrm{ML}}\tilde{\zeta}_{\mathrm{ML}}\right]} \leq -E\left[\frac{\partial^2 \ln P_{\boldsymbol{\nu}|\beta,\zeta}}{\partial\zeta^2}\right] = \sum_{i=0}^{M-2} \frac{\zeta^{2(i-1)}(i\beta)^2}{(\beta\zeta^i + \sigma_i^2)^2} \frac{q^{i+1}}{2} \tag{4.24}$$

$$\frac{1}{E\left[\tilde{\zeta}_{\mathrm{ML}}^2\right]} \leq -E\left[\frac{\partial^2 \ln P_{\boldsymbol{\nu}|\beta,\zeta}}{\partial\beta\partial\zeta}\right] = \sum_{i=0}^{M-2} \frac{\zeta^i\zeta^{i-1}i\beta}{(\beta\zeta^i + \sigma_i^2)^2} \frac{q^{i+1}}{2} \tag{4.25}$$

where $\tilde{\beta}_{\mathrm{ML}}$ represents the error in a maximum likelihood estimation of β, similarly $\tilde{\zeta}_{\mathrm{ML}}$. It should be pointed out that the CRB derivation itself is not novel – it is very similar to that of Worneli[113] – the novelty of this discussion is the assumption of the previous page that made this CRB derivation applicable to our multiscale problem at all.

As stated earlier, it is implicit in equations (4.23)-(4.25) that measurements are available at all nodes on the finest scale. Suppose now that measurements are available at only a subset of the finest-scale nodes. Then the covariance $\Lambda_{\boldsymbol{\nu}}$ will be diagonal under the assumption of the availability of zero-noise measurements at all nodes on the tree which are ancestors of those nodes on the finest scale at which measurements are available. More specifically, let $\{s_1, \ldots, s_N\}$ be the collection of nodes on the finest scale where noisy measurements are available. Then let

$$Q(i) = \left\| \left\{ s_1 \bar{\gamma}^{M-i-2}, \ldots, s_N \bar{\gamma}^{M-i-2} \right\} \right\| \tag{4.26}$$

where $\|\mathcal{S}\|$ counts the number of *different* members in set $\mathcal{S}$. Then $Q(i)$ represents the number of noiseless "measurements" that need to be assumed on scale $(i+1)$ in order to diagonalize $\Lambda_{\boldsymbol{\nu}}$; for example, when the finest scale is densely sampled, $Q(i) = q^{(i+1)}$. The three CRB equations (4.23)-(4.25) are then modified by replacing q^{i+1} in each sum with $Q(i)$; for example

$$\frac{1}{E\left[\tilde{\beta}^2_{\mathrm{ML}}\right]} \leq -E\left[\frac{\partial^2 \ln P_{\boldsymbol{\nu}|\beta,\zeta}}{\partial \beta^2}\right] = \sum_{i=0}^{M-2} \frac{\zeta^{2i}}{(\beta\zeta^i + \sigma_i^2)^2} \frac{Q(i)}{2} \tag{4.27}$$

Finally there is the issue of estimating the prior covariance P_o at the root node of the tree. Two possibilities may be considered:

1. The root node covariance is some unknown P_o, not known to be related to the process noise parameters β, ζ. In this case, the noiseless measurement $y(0)$ at the root node of the tree is a sufficient statistic for estimating P_o, giving the usual CRB:

$$P_{y(0)|P_o} = \frac{1}{\sqrt{2\pi}\sqrt{P_o}} \exp\left(-\frac{y(0)^2}{2P_o}\right) \tag{4.28}$$

$$\frac{1}{E\left[\tilde{P}^2_{o_{\mathrm{ML}}}\right]} \leq -E\left[\frac{\partial^2 P_{y(0)|P_o}}{\partial P_o^2}\right] = \frac{1}{2P_o^2} \tag{4.29}$$

That is, the variance of the error in the estimate of P_o is bounded by $2P_o^2$.

2. The root node covariance is given by the extension of the process noise covariances, i.e.,

$$P_o = \beta\zeta^{-1} \tag{4.30}$$

then (4.20) remains valid at the root node, i.e., $i = -1$. The changes to the CRB are thus accomplished by indexing all expressions in i from -1; e.g., (4.23) becomes

$$\frac{1}{E\left[\tilde{\beta}^2_{\mathrm{ML}}\right]} \leq -E\left[\frac{\partial^2 \ln P_{\boldsymbol{\nu}|\beta,\zeta}}{\partial\beta^2}\right] = \sum_{i=-1}^{M-2} \frac{\zeta^{2i}}{(\beta\zeta^i + \sigma_i^2)^2}\frac{q^{i+1}}{2} \tag{4.31}$$

Except in certain unusual circumstances (e.g., very small ζ, small M or small q), the difference between (4.31) and (4.23) will be vanishingly small.

4.3.2 CRB - Synthetic Data Tests

Three tests were performed, in which data was synthesized on a multiscale tree using a model with known parameters:

Basic Tree Parameters: 7 Level Quad-Tree ($M = 7, q = 4$)

$$\beta = (48cm)^2$$

$$\zeta = (0.5)^2$$

A non-linear optimization routine was used to maximize the multiscale likelihood expression (4.3), leading to an estimate of the model parameters.

The three tests are as follows:

1. No noise is added to the measurements on the finest scale. The multiscale likelihood calculation algorithm is given noiseless measurements on *all* scales – consistent with the assumption made in the derivation of the Cramer-Rao bound.

2. No noise is added to the measurements on the finest scale. Although the CRB is still based on noiseless measurements on all scales, the likelihood calculation algorithm is given measurements on the finest scale only which are noiseless.

3. Gaussian white noise, variance $(2\text{cm})^2$, is added to the synthesized data. As in test 2, the likelihood calculation algorithm is given measurements on the finest scale only.

Each of the above tests were applied to 200 sets of synthesized sample paths; the results corresponding to these tests are shown in Table 4.5. In the table, covariance matrices always refer to the covariance of the joint vector $[\beta \;\; \zeta]^T$, and ρ refers to the correlation coefficient of the errors in $\hat{\beta}$ and $\hat{\zeta}$.

The conditions of Test 3, i.e., noisy measurements on a single scale, are similar to those which we typically encounter in practice and thus represent the conditions of greatest interest. There are two basic conclusions to be drawn from these results:

- The Cramer-Rao bound developed above was based on the assumption that noiseless measurements are available on all scales above the finest. Although this assumption might have appeared to destroy the usefulness of the bound (i.e., make it excessively weak), the results of the table show the bound to be relatively tight. Indeed in Test 3, a realistic test case, the empirical variances exceeded the lower bound by at most 38%.

- The large correlation coefficient, predicted by the CRB and observed empirically, confirm the behavior originally observed in Figure 4-2 which we had set out to explain, i.e., the fact that the likelihood value deviates very little from its maximum over a wide range of values of B, γ.

4.3.3 CRB - Oceanographic Data Tests

The previous subsection has provided an empirical verification of the CRB. In this subsection we would like to explore the identification of the parameters in the same

Test 1:	Finest scale measurement noise variances: $\sigma^2 = 0$ Scales on which meas. available to m.s. likelihood calculator: ALL		
	Empirical:	Error Covariance	$\begin{bmatrix} 9.57{\cdot}10^{4} & -2.12{\cdot}10^{0} \\ -2.12{\cdot}10^{0} & 4.86{\cdot}10^{-5} \end{bmatrix}$
		Correl. Coeff.	$\rho = -0.98$
	CRB:	Covariance Bound	$\begin{bmatrix} 9.92{\cdot}10^{4} & -2.26{\cdot}10^{0} \\ -2.26{\cdot}10^{0} & 5.26{\cdot}10^{-5} \end{bmatrix}$
		Correl. Coeff.	$\rho = -0.99$

Test 2:	Finest scale measurement noise variances: $\sigma^2 = 0$ Scales on which meas. available to m.s. likelihood calculator: Finest		
	Empirical:	Error Covariance	$\begin{bmatrix} 1.41{\cdot}10^{5} & -3.10{\cdot}10^{0} \\ -3.10{\cdot}10^{0} & 7.11{\cdot}10^{-5} \end{bmatrix}$
		Correl. Coeff.	$\rho = -0.98$
	CRB:	Covariance Bound	$\begin{bmatrix} 9.92{\cdot}10^{4} & -2.26{\cdot}10^{0} \\ -2.26{\cdot}10^{0} & 5.26{\cdot}10^{-5} \end{bmatrix}$
		Correl. Coeff.	$\rho = -0.99$

Test 3:	Finest scale measurement noise variances: $\sigma^2 = (2cm)^2$ Scales on which meas. available to m.s. likelihood calculator: Finest		
	Empirical:	Error Covariance	$\begin{bmatrix} 1.94{\cdot}10^{5} & -4.76{\cdot}10^{0} \\ -4.76{\cdot}10^{0} & 1.24{\cdot}10^{-4} \end{bmatrix}$
		Correl. Coeff.	$\rho = -0.97$
	CRB:	Covariance Bound	$\begin{bmatrix} 1.45{\cdot}10^{5} & -3.73{\cdot}10^{0} \\ -3.73{\cdot}10^{0} & 1.00{\cdot}10^{-4} \end{bmatrix}$
		Correl. Coeff.	$\rho = -0.98$

Table 4.5: This table compares the Cramer-Rao bound and empirically determined estimation error covariances for three different test scenarios. The empirical covariances are subject to errors due to insufficient Monte-Carlo runs and due to quantization effects of the likelihood optimization technique. The significant aspects of the table are the close agreement between CRB and empirical covariances, and the large correlation coefficient ρ, which confirms the behavior originally observed in Figure 4-2.

multiscale model (4.15),(4.16), but applied to oceanographic data. This identification serves two purposes:

1. To estimate the values of B, γ in (4.15). The estimated values provide an independent test of the model values deduced in Chapter 3 using power spectral matching.

2. As a model validation exercise. If the ocean surface is reasonably described by (4.16), then the empirical covariance of $[\beta, \zeta]^T$ should be similar to the CRB.

Empirical tests were performed using about 12 months of Topex / Poseidon data (measurements of the ocean height) on a 9-level quadtree. The satellite measurements are sparse, and thus only a subset of the finest scale nodes of the tree possess measurements (10000 measurements distributed among 65000 nodes). The satellite, its measurements, and the measurement sampling pattern have been discussed in Chapter 3.

We propose to identify the parameters of (4.15) for two different circumstances:

1. The measurements placed on the finest scale of the tree make up a ten day set of direct satellite measurements of ocean height having an approximate measurement noise variance of $(5\text{cm})^2$.

2. The measurements placed on the finest scale of the tree make up a ten day set of measurements of the *ocean anomaly* – the difference between the ocean height at some time t and the average ocean height. The measurement noise variance is approximated as $(5\text{cm})^2$.

Under each of these two circumstances, a nonlinear optimization routine was used to deduce the empirical values of β, ζ for each of 37 ten-day data sets. The covariance of $\left[\hat{\beta}\ \hat{\zeta}\right]^T$ is computed as the sample covariance. The results are tabulated in the "empirical" sections of Table 4.6.

Next, for each of the two circumstances a Cramer-Rao bound on the estimation error covariance was determined; as was discussed above and demonstrated in (4.27),

the CRB calculation was modified to account for the sparsity of the finest-scale measurements. In each test, the empirically determined values for β, ζ were used in the Cramer-Rao bound computation, since this bound depends upon β, ζ. Results corresponding to each of the above tests are shown in the "CRB" sections of Table 4.6. As in the previous subsection, the covariance matrices in the table refer to the covariance of the joint vector $[\beta \ \ \zeta]^T$, and ρ refers to the correlation coefficient of the errors in $\hat{\beta}$ and $\hat{\zeta}$.

Several conclusions may be drawn from the results of Table 4.6:

1. The estimated model values under test 1,

$$B = 30.5\text{cm} \qquad \mu = 1 - 2\log_2 \gamma \simeq 1.8 \tag{4.32}$$

   agree well with the model values determined by power spectral matching in Chapter 3.

2. Despite the fact that we are applying a simple, idealized multiscale model to an enormously complicated time-varying natural system, the Cramer-Rao bound and the empirical parameter covariances are of comparable order, lending considerable credence to the use of the multiscale model.

4.3.4 Anisotropic Model Identification

We would like to end this chapter with one final model identification example, which simultaneously addresses a problem of oceanographic interest, and demonstrates the ability to identify anisotropic fractal dimensions of higher order processes (as claimed at the end of Section 4.2).

Consider the multiscale model (4.15),(4.16). On a tree of order q, this model implies a certain kind of isotropic behavior: the statistics of the evolution from a node s to any of its children $s\alpha_i$ is the same for all i, regardless of the relative parent-child orientation. An assertion of such an isotropic model may be inappropriate for the ocean surface: the rotation of the earth produces a fundamental asymmetry between

Test 1:	Use Topex/Poseidon altimetry data as input. Number of tree scales: $M = 8$.		
	Empirical:	Error Covariance	$\begin{bmatrix} 2.57{\cdot}10^{4} & -3.67{\cdot}10^{0} \\ -3.67{\cdot}10^{0} & 5.86{\cdot}10^{-4} \end{bmatrix}$
		Correl. Coeff.	$\rho = -0.95$
		Estimated Offset:	$\hat{\beta} = (30.5cm)^2$
		Estimated Slope:	$\hat{\zeta} = (0.772)^2$
	CRB:	Variance Offset:	$\beta = (30cm)^2$
		Variance Slope:	$\zeta = (0.77)^2$
		Measurement Noise:	$\sigma^2 = (5cm)^2$
		Covariance Bound	$\begin{bmatrix} 4.11{\cdot}10^{3} & -4.40{\cdot}10^{-1} \\ -4.40{\cdot}10^{-1} & 4.85{\cdot}10^{-5} \end{bmatrix}$
		Correl. Coeff.	$\rho = -0.985$

Test 2:	Use Topex/Poseidon anomaly data as input. Number of tree scales: $M = 8$.		
	Empirical:	Error Covariance	$\begin{bmatrix} 4.74{\cdot}10^{0} & -3.67{\cdot}10^{-2} \\ -3.67{\cdot}10^{-2} & 4.25{\cdot}10^{-4} \end{bmatrix}$
		Correl. Coeff.	$\rho = -0.82$
		Estimated Offset:	$\hat{\beta} = (4.66cm)^2$
		Estimated Slope:	$\hat{\zeta} = (1.07)^2$
	CRB:	Variance Offset:	$\beta = (5cm)^2$
		Variance Slope:	$\zeta = (1.0)^2$
		Measurement Noise:	$\sigma^2 = (5cm)^2$
		Covariance Bound	$\begin{bmatrix} 3.12{\cdot}10^{0} & -2.02{\cdot}10^{-2} \\ -2.02{\cdot}10^{-2} & 1.34{\cdot}10^{-4} \end{bmatrix}$
		Correl. Coeff.	$\rho = -0.985$

Table 4.6: This table compares the Cramer-Rao bound and empirically determined estimation error covariances based on two oceanographic data sets: ocean height and anomaly height (the difference between the instantaneous height and the time-averaged ocean height).

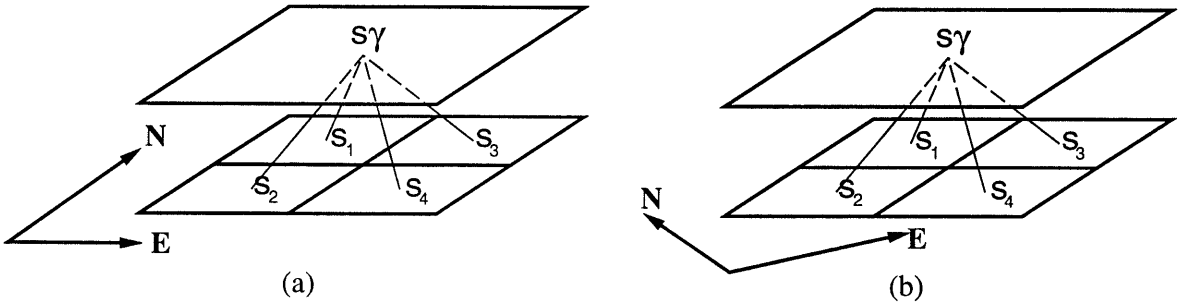


Figure 4-3: Two possible tree orientations and associated node labels for investigating anisotropic spectra using a quad-tree.

the east-west and north-south directions; this asymmetry is realized in the action of the the Coriolis force. We can investigate the presence of such spectral asymmetries by using a modified quad-tree model which contains separate parameters for the statistics along two orthogonal directions:

$$x(s) = 1 \cdot x(s\bar{\gamma}) + B_1 \gamma_1^{m(s)} w(s) \qquad \begin{matrix} x(0) \sim \mathcal{N}(0, P_o) \\ w(s) \sim \mathcal{N}(0,1) \end{matrix} \qquad s \in \{s_1, s_4\} \quad (4.33)$$

$$x(s) = 1 \cdot x(s\bar{\gamma}) + B_2 \gamma_2^{m(s)} w(s) \qquad \begin{matrix} x(0) \sim \mathcal{N}(0, P_o) \\ w(s) \sim \mathcal{N}(0,1) \end{matrix} \qquad s \in \{s_2, s_3\} \quad (4.34)$$

Figure 4-3 shows two possible orientations of the quad-tree and the associated node labels $s_1, \ldots, s_4$.

Figure 4-3(a) shows the usual tree orientation which was used throughout Chapter 3. The model equations (4.33) and (4.34) parameterize the spectra along the orthogonal NW-SE and SW-NE directions; we do not expect the Coriolis force to induce a difference in behavior between these two directions. The resulting identified model parameters $(B_1, \gamma_1, B_2, \gamma_2)$ for each of 37 ten-day data sets are shown in the top half of Figure 4-4; indeed, the NW-SE and SW-NE spectral parameters are essentially indistinguishable.

On the other hand, Figure 4-3(b) represents a rotation of the usual orientation by 45^o, allowing the model (4.33),(4.34) to parameterize the spectra along the orthogonal N-S and W-E directions; in this case we certainly expect the Coriolis effect to act differently upon the two directions. The resulting identified model parameters $(B_1, \gamma_1, B_2, \gamma_2)$ are shown in the bottom half of Figure 4-4 – a clear separation between

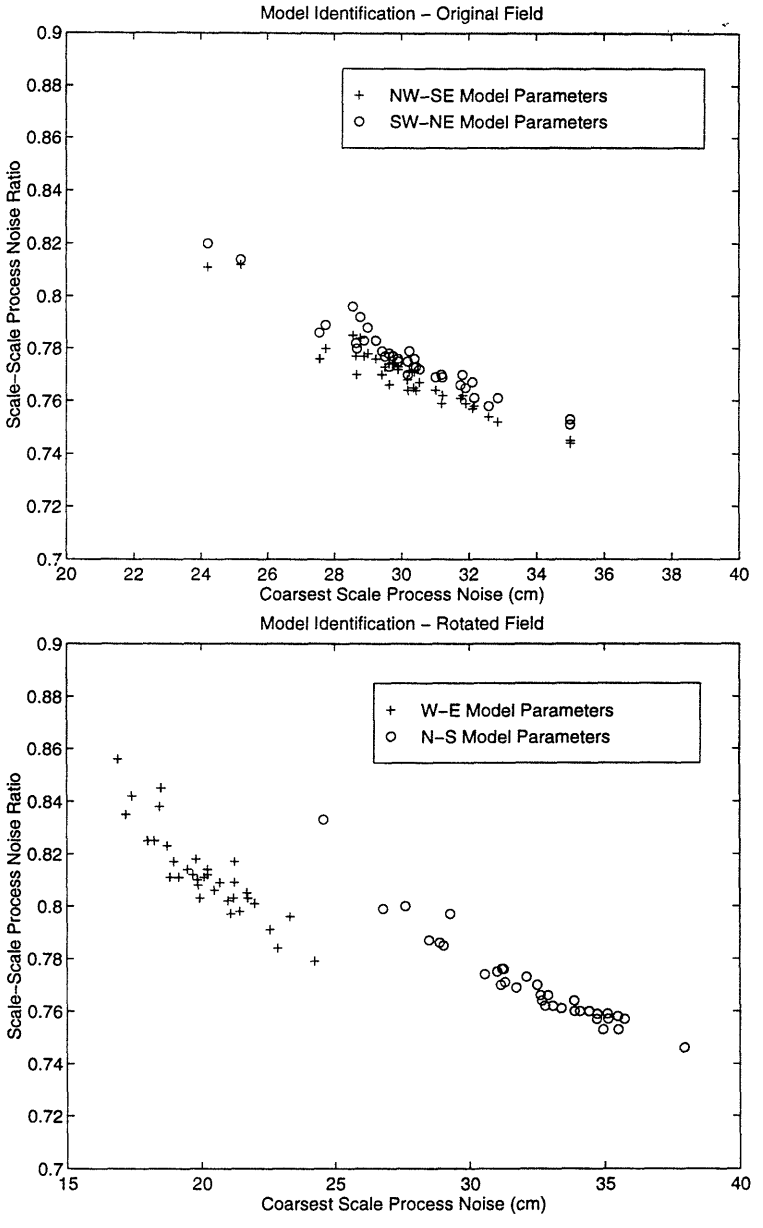


Figure 4-4: Two examples of isotropic oceanographic spectrum identification. In each example one set of spectral parameters $(B_1, \gamma_1, B_2, \gamma_2)$ was identified for each of 37 ten-day TOPEX/POSEIDON data sets.
Top Figure: spectral identification results along NW-SE and SW-NE directions.
Bottom Figure: spectral identification results along W-E and N-S directions.

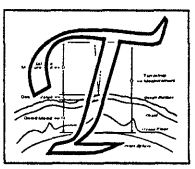
the N-S and W-E parameters is observed.

It is possible to postulate effects other than the Coriolis force which might induce such spectral anisotropies;[1] however this does not affect the basic result of this section: multiscale models such as (4.33),(4.34), coupled with the multiscale likelihood function evaluator, are capable of assessing and quantifying certain anisotropies in multidimensional processes.

[1]For example, the nature in which the curved surface of the earth is sampled onto the rectangular grid which we process is also nonisotropic. Furthermore the predominant W→E circulation in this portion of the ocean introduces an asymmetry.

Chapter 5

Geoid Surface Estimation



his chapter continues to investigate the oceanographic estimation problem considered in Chapter 3; however, whereas in Chapter 3 we considered the geoid as given (albeit with a spatially varying error variance), in this chapter we will *jointly* estimate the ocean elevation and the geoid. It must be emphasized that the results of this chapter are of a preliminary nature; additional work, particularly in the area of model validation, will be required in order to properly address the joint estimation problem. Furthermore some of the arguments in this chapter are not yet adequately justified for this work to be considered complete.

5.1 Introduction

As this chapter represents, in a sense, a continuation of the ocean elevation estimation study of Chapter 3, for basic introductory and background material the reader is referred to Section 3.1.

There are at least two significant motivations for studying and estimating the geoid:

1. Given the availability of accurate global ocean elevation data, which measures the sum of the geoid N and the ocean height ζ relative to the geoid, an improved estimate of the N immediately implies an improved estimate of ζ.

2. The geoid is correlated with features inside the earth: the thickness of the crust, density variations within the crust, and the nature of the boundary between the crust and the mantle. An improved estimate of the geoid leads to an improved understanding of the earth's interior.

Although the geoid can be measured pointwise on the surface of the earth, it is essentially impossible to densely sample the geoid on a global scale in this manner. High-resolution global data sets are available which contain measurements strongly correlated with the geoid, however the fine-scale estimation of the geoid remains a challenge because of the difficulty of separating the geoid from other signals in the data sets. For example, a detailed knowledge of the shape of the earth's surface is insufficient to determine the gravitational effects of its crust because the space-varying density of the crust is inadequately known, and because even the thickness of the crust (i.e., the spatially-varying degree to which it projects into the mantle) is unknown. Similarly the separation of the geoid N and the relative ocean height ζ from satellite altimetric measurements is a difficult task. There are certain consistency relationships which must hold for ζ, which in principle allows the separation of the satellite measurements into its geoidal and oceanic components, however the consistency constraints involve sophisticated climatological models, difficult to incorporate into an estimator.

Nevertheless, we have a strong reason to believe that our multiscale approach will be able to deduce improved estimates of the geoid: the measurement residuals of Figure 3-16 are strongly correlated with fine-scale topographic and bathymetric features – precisely those features too small (i.e., local) to be captured in the OSU91A geoid[87] used in Chapter 3. That is, since the measurement residuals are correlated with certain aspects of the OSU91A error, in principle a part of that error is estimatable from satellite measurements.

Motivated by Figure 3-16, we propose to jointly estimate the geoid N and the relative ocean height ζ from TOPEX / POSEIDON data and from the OSU91A geoid estimates. This proceeds in three stages:

1. We need to determinate an appropriate joint multiscale model. We will continue to use the prior model for ζ from Chapter 3, and a similar prior model for N will be developed. The multiscale measurement model will need to relate N, ζ to two data sets:

   (a) TOPEX / POSEIDON altimetric measurements of $N + \zeta$

   (b) The OSU91A estimates of N

   The key quantity to be specified is the error variance of the OSU91A data set; this variance will be denoted by $\phi(x, y)$.[1]

2. We believe that $\phi(x, y)$ is strongly correlated with the measurement residuals $\nu(x, y)$ of Section 3.4. We do not, however, wish to infer ϕ directly from ν, since doing so asserts, without justification, that the value of each residual can be attributed to the OSU91A error.

   Instead, we wish to independently determine a function $\Delta G(x, y)$ which is correlated with the OSU91A error. We will demonstrate how two such candidate functions may be determined from measurements of the earth's surface (or topography) $T(x, y)$.

3. Once we have determined $\nu(x, y)$ and $\Delta G(x, y)$, we will correlate these functions to infer $\phi(x, y)$. Once we have determined ϕ our joint multiscale model is complete and can be used for joint estimation.

The above stages follow the logical flow of the chapter:

Section 5.2 - Background and previous work
Section 5.3 - Joint multiscale model definition
Section 5.4.1 - Computation of residuals $\nu(x, y)$
Section 5.4.2 - Computation of function $\Delta G(x, y)$

[1] Measurements taken over a planet are usually indexed in terms of some spherical coordinate system (e.g., latitude and longitude). On the other hand, when only a small fraction of the planet is being studied we may assume the region to be flat and to index measurements and estimates in terms of a Cartesian (x, y) system. This latter approach will be used throughout this chapter.

Section 5.4.3 - Computation of $\phi(x, y)$

Section 5.5 - Joint estimation results

5.2 Past Efforts

To be sure, there have been a great number of efforts to estimate the geoid[30, 71, 79, 82, 86, 87] including efforts to jointly estimate the geoid and the ocean elevation[21, 23, 24, 72, 80, 101, 116]. Although the sheer number of efforts is large, they differ primarily in terms of details of geodetic and oceanographic models, parameters, and corrections. There is a significant common thread to most of these efforts; the goal of this section is to briefly review some of this common ground in order to put the work of this chapter into context.

There are four types of measurements which are widely used in estimating the geoid:

1. Oceanographic Altimetry: altimetric measurements of the ocean elevation measure

$$Y_S(x, y, t) = N(x, y) + \zeta(x, y, t) + \text{noise} \tag{5.1}$$

Since the range of ζ is small (about one meter), frequently satellite altimetric measurements have been considered noisy measurements of the geoid, $Y_s \approx N$, with ζ being part of the "noise". Until recently, long-wavelength errors in the estimation of the satellite orbit made this data of limited use when estimating long-wavelength components of the geoid.

2. Orbit Tracking: the trajectory of a satellite orbiting the earth is perturbed from an ellipsoidal shape by the undulations of the gravitational field felt by the satellite. The transfer function of the gravitational field at the surface of a flat earth to that at altitude z is given by the upwards continuation function

$$e^{-z\left(w_x^2+w_y^2\right)^{-\frac{1}{2}}} \tag{5.2}$$

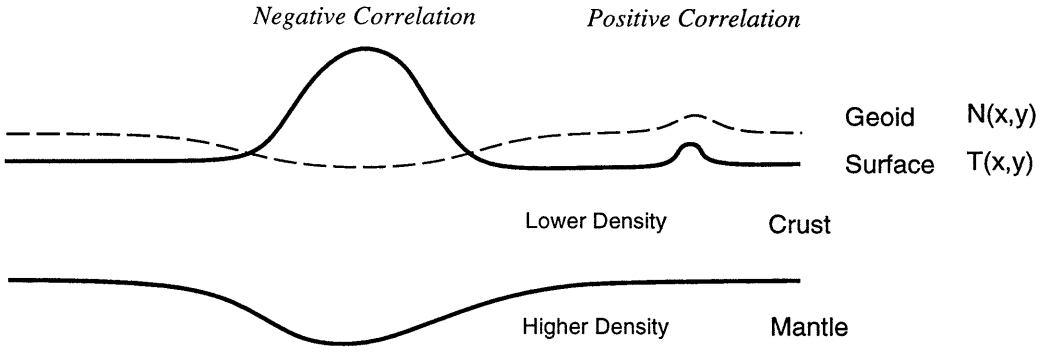


Figure 5-1: A simple example of positive and negative geoid-topography correlation.

(where w_x, w_y are the spatial frequencies along the x and y directions). For typical satellites in orbit at altitudes $z > 300$km, only relatively long wavelengths (> 2000km) of the geoid may be estimated from orbital tracking data.

3. Surface Topography: A detailed knowledge of the shape of the earth's surface, $T(x, y)$, is insufficient to determine the gravitational effects of its crust. For example, consider a single mountain on the earth's surface; even assuming that its density is uniform and known, the gravitational field in the vicinity of the mountain is difficult to predict because of the unknown response of the underlying crust to the mass of the mountain (see sketch in Figure 5-1). In many cases, a low density crustal root projects into the mantle under the mountain[17, 41], such that the gravitational perturbation near the mountain may be one of *repulsion*, rather than attraction.

   Consequently the geoid may be negatively correlated with large structures, such as mountains, but positively correlated with smaller hills, whose mass is too small to generate a crustal root. The correlation between the topography and the geoid, expressed as a function of wavenumber, is known as the *coherence*[42, 50]. For high wavenumbers the coherence is positive; for low wavenumbers the coherence is small and possibly negative.

4. Pointwise Measurements: direct measurements of the geoid are inadequate in scope to serve as the basis for a global estimation of the geoid. These measurements have primarily served as a validation points for geoids estimated by other

means.

All of the cited geoid estimation efforts performed least-squares estimation of a vector of unknowns using brute-force matrix inversion techniques (i.e., by directly computing (2.9),(2.10)). Such an approach has clearly constrained researchers to estimating relatively small numbers of unknowns. This constraint has motivated the use of a global set of basis functions in which to represent the geoid: a spherical harmonic basis is almost universally used for this purpose. That is, the geoid is represented as[42]

$$N(\lambda,\phi) = K \sum_{l=2}^{J} \sum_{m=-n}^{n} \left(\bar{C}_{l,m} \cos(m\lambda) + \bar{S}_{l,m} \sin(m\lambda) \right) \bar{P}_{l,m}(\sin\phi) \tag{5.3}$$

where K is some constant, $\bar{P}_{l,m}$ represents the fully normalized Legendre function[42], λ represents the longitude, and ϕ counts degrees of latitude from the north pole. J is chosen relatively small ($J < 50$) such that the vector of coefficients $\{\bar{C}_{l,m}, \bar{S}_{l,m}\}$ may be estimated by matrix inversion.

To be sure, higher order estimates of the geoid have been produced; for example, the geoid estimates used in Chapter 3 were based on a spherical harmonic expansion to $J = 360$. The low and high order coefficients of high order geoid estimates are typically computed separately (where a natural separation is frequently provided by the measurement data):

1. The low order coefficients (e.g., up to degree and order $J \approx 50$) are estimated using straightforward least-squares matrix inversion techniques. For example, these low order coefficients may be estimated based on satellite orbital tracking data[79, 82] (which is accurate at low orders).

2. The remaining, higher order, coefficients are computed as spherical harmonic coefficients of the measurement data without any statistically-based processing. For example, these coefficients may be computed as the spherical harmonic integrals of topographic data[79, 82] (measurements of $T(x,y)$).

For low-order terms, the complete estimation error covariance is available, whereas

no error statistics are available for the remaining higher order terms.

Prior models are chosen primarily to regularize the low order component of the estimation problem (that is, to improve the conditioning of the matrix inversion), rather than to assert prior statistical knowledge. The selected prior model is frequently the weakest possible one which provides adequate regularization. Three prior models are commonly used for the geoid:

1. No prior may be used[79, 80]. In low-order estimation problems there are typically far more measurement data than unknown coefficients.

2. A prior model[72, 101] based on Kaula's law[50], which states that

$$\mathrm{var}(\bar{C}_{l,m}) \propto \frac{1}{l^4} \qquad \mathrm{var}(\bar{S}_{l,m}) \propto \frac{1}{l^4} \tag{5.4}$$

3. A prior model[21, 23, 82, 86] is chosen based on the estimation error covariance matrix of previous geoid estimates.

A number of efforts[21, 23, 24, 72, 80, 101, 116] have jointly estimated the geoid and the relative ocean height. The relative ocean height ζ is written as a low-order (typically $J < 20$) spherical harmonic expansion as in (5.3), and one of the following three prior models is asserted:

1. No prior is used[80, 101]: the prior model chosen for the geoid is assumed adequate for regularization purposes.

2. A prior model based on empirical power spectral laws, such as in ([33]), is chosen.

3. A prior model is chosen[21, 23, 72] based on the error covariance matrix of previous ocean height estimates, in particular that of Levitus[58].

The past efforts referenced in this section all used full matrix inversion to compute least squares estimates; such an approach seriously constrains the number of coefficients to be estimated, and hence the resolution of the estimated field. Furthermore

the use of global spherical harmonic functions introduces further problems[21, 23] (e.g., statistical questions regarding the meaningfulness of global estimates of ocean height on a planet whose surface is 30% land).

In the next section we will explore an alternative means to perform joint geoid / ocean height estimation using our multiscale estimation framework. The efficiency of this framework allows large numbers of coefficients to be estimated, allowing us, in principle, to compute high resolution pointwise estimates without resorting to global basis functions.

5.3 Multiscale Model Selection

As in Chapter 3, our physical region of interest will continue to be the region in the north-east Pacific ocean, as shown in Figure 3-2. We will continue to use the same multiscale model for the ocean surface ζ, selected in Chapter 3:

$$\zeta(s) = \zeta(s\bar{\gamma}) + 35\text{cm} \cdot 2^{-m(s)/2} w(s) \tag{5.5}$$

a model which implicitly assumes the ocean surface to be static over time. We will use Kaula's law[50] as the motivation for the multiscale prior model of the geoid. Recall that Kaula's law asserts that

$$\text{var}(\bar{C}_{l,m}) \propto \frac{1}{l^4} \qquad \text{var}(\bar{S}_{l,m}) \propto \frac{1}{l^4} \tag{5.6}$$

Let

$$\begin{aligned} N(\lambda,\phi) &= K \sum_{l=2}^{J} \sum_{m=-n}^{n} \left(\bar{C}_{l,m} \cos(m\lambda) + \bar{S}_{l,m} \sin(m\lambda) \right) \bar{P}_{l,m}(\sin\phi) \\ &= K \sum_{l=2}^{J} N_l(\lambda,\phi) \end{aligned} \tag{5.7}$$

Then the variance of the total power in N_l is given by

$$\text{var}\left(\int\int N_l^2 \sin(\phi)\, d\phi\, d\lambda\right) \propto \sum_{m=-l}^{l} l^{-4} \propto l^{-3} \tag{5.8}$$

That is, if we consider the order l to be a rough measure of scale, then Kaula's law predicts a $1/f^3$ prior spectrum. The above argument can be made more precise by computing the exact one-dimensional wavenumber spectrum associated with the spherical harmonic coefficient variances[118]:

$$S(k) = \sum_{l=k}^{J} \sigma_l^2 (2l+1) \frac{\Gamma(l-k+1)}{4\pi\Gamma(l+k+1)} \left[\frac{\Gamma((l+k+1)/2)2^m}{\Gamma((l\pm k)/2+1)\sqrt{\pi}} cos(\frac{\pi}{2}(l+k))\right]^2 \tag{5.9}$$

where Γ represents the Gamma function. A numerical solution for (5.9), given $\sigma_l^2 \propto l^{-4}$, predicts a $1/f^{3.2}$ spectrum. Such a power-law form for the prior model of the geoid suggests a multiscale model of the form

$$N(s) = N(s\bar{\gamma}) + B_N 2^{m(s)(1-\mu_N)/2} w(s) \tag{5.10}$$

where $\mu_N = 3.2$.

The likelihood techniques of Chapter 4 may be applied to the model of (5.10) as a validation of our proposed value of μ_N. The optimization of the likelihood, given samples of the OSU91A geoid at the TOPEX/POSEIDON measurements locations, as a function of B_N, μ_N yields

$$B_N \simeq 3600\text{cm} \qquad \mu_N \simeq 3.23 \tag{5.11}$$

which agrees well with the value of μ_N from Kaula's law. Consequently we propose the following model for the geoid:

$$N(s) = N(s\bar{\gamma}) + 3600\text{cm} \cdot 2^{-1.1m(s)} w(s) \tag{5.12}$$

Two data sets will be used as measurements: the set of OSU91A geoid estimates

Y_N, and TOPEX/POSEIDON altimetric data Y_S. Under the assumption that the noise in the altimetric data has a variance of $(5\text{cm})^2$[32], the multiscale measurement model becomes

$$\begin{bmatrix} Y_S \\ Y_N \end{bmatrix}(s) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \zeta \\ N \end{bmatrix}(s) + \begin{bmatrix} 5 & 0 \\ 0 & \sqrt{\phi(s)} \end{bmatrix} v(s) \qquad m(s) = M - 1 \quad (5.13)$$

The function ϕ describes the space-varying error variance in the OSU91A geoid. To be sure, the creators of the OSU91A geoid have computed ϕ, but only at very low resolutions, inadequate to capture the high-resolution variability implied by Figure 3-16. We did perform an approximate evaluation of function ϕ in Section 3.4 based on the OSU91A geoid gradient; a more elaborate determination of ϕ forms the topic of the next section. Once ϕ has been determined, the specification of the multiscale model is complete, permitting joint estimation of the geoid and ocean height.

5.4 Determination of ϕ

In Section 3.4, the geoid error variance ϕ was estimated via the observation that the measurement residuals, resulting from a multiscale estimation of the ocean elevation, were strongly correlated with the slope of the OSU91A geoid. That the OSU91A geoid slope was chosen as the basis of comparison with the residuals was somewhat ad-hoc; we should be able to do better by looking more closely at the error in the OSU91A geoid estimates.

The OSU91A geoid estimates may be written as

$$Y_N = N + \tilde{N} = N + \tilde{N}_c + \tilde{N}_o \quad (5.14)$$

That is, the OSU91A geoid Y_N possess two sources of error:

- Errors of commission $\tilde{N}_c$ - errors in the coefficients in the geoid estimates.
- Errors of omission $\tilde{N}_o$ - errors due to the truncation of the spherical harmonic expression at some limit J.

We propose to model the geoid error term $\tilde{N}_o$ and to ignore the error associated with $\tilde{N}_c$:

1. Much of the commission error occurs at long wavelengths[87], making this error difficult to estimate from the relatively small area of study in the north Pacific.

2. The distribution of *large* measurement residuals from Chapter 3 (Figure 3-16) is strongly correlated with *local* bathymetric features (i.e., those features too small to be captured by the OSU91A geoid and hence contributing to $\tilde{N}_o$).

Now recall from our discussion in the introduction that there remain three steps to determine ϕ:

1. Compute the measurement residuals under the assumption that the OSU91A geoid is exact; this is the subject of Subsection 5.4.1.

2. Independently compute a function ΔG which is believed to be correlated with the OSU91A geoid error; this is the subject of Subsection 5.4.2.

3. Once the residuals ν and the function ΔG have been calculated, we will correlate ν with ΔG to infer ϕ; this is the subject of Subsection 5.4.3.

5.4.1 Determination of Oceanographic Residuals

Our goal in this section is to compute statistically significant residuals, computed ignoring errors in the OSU91A geoid; i.e., the residuals are based on the multiscale model

$$\zeta(s) = \zeta(s\bar{\gamma}) + 35\text{cm} \cdot 2^{-m(s)/2} w(s) \tag{5.15}$$

$$Y_\zeta(s) = \zeta(s) + v(s) \qquad E\left[v^2(s)\right] = (5\text{cm})^2 \tag{5.16}$$

where the data set Y_ζ is computed as $Y_\zeta = Y_S - Y_N$, and where the normalized measurement residuals are computed as

$$\nu(s) = \frac{Y_S(s) - Y_N(s) - \hat{\zeta}(s)}{\sqrt{\tilde{P}(s) + R(s)}} \tag{5.17}$$

A simple set of residuals were computed in Section 3.4, however these were computed based on a single ten-day set of data, whereas over one year of data is currently available.

Recall that our multiscale prior model for the ocean surface (5.15) assumes the ocean surface to be static, consequently it is not obvious how best to make use of satellite data taken over long periods of time (i.e., periods sufficiently long such that the static assumption may be poor). There exist at least two reasonable alternatives:

1. One year of satellite measurements corresponds to 37 repeats cycles; that is, for our region of interest in the north Pacific, 37 independent measurements are taken at each of the roughly 20,000 sample points. By averaging the 37 measurements at each of these sample points, we obtain a new data set $\bar{Y}_\zeta$ containing "measurements" of the mean ocean surface $\bar{\zeta}$. The variance of the "measurement" $\bar{Y}_\zeta$ will be

$$\text{var}\left(\bar{Y}_\zeta(x,y)\right) = \frac{(5\text{cm})^2}{37} + (\text{pointwise ocean surface variance over one year})(x,y) \tag{5.18}$$

that is, computing the pointwise average of the satellite measurements will reduce the 5cm white measurement error by a factor of $\sqrt{37}$, however additional variance is added by the intrinsic variability of the ocean surface over time (over one year).

2. Rather than averaging the measurements of the 37 repeat cycles, we can compute a set of residuals for each of the 10-day repeat cycles, and then average the residuals. The variance of the data set $Y_\zeta(x,y) = Y_S(x,y) - Y_N(x,y)$ for each of the repeat cycles is

$$\text{var}\left(Y_\zeta(x,y)\right) = (5\text{cm})^2 + (\text{pointwise ocean surface variance over ten days})(x,y) \tag{5.19}$$

The choice between these two alternatives is dictated by the size of the variances in

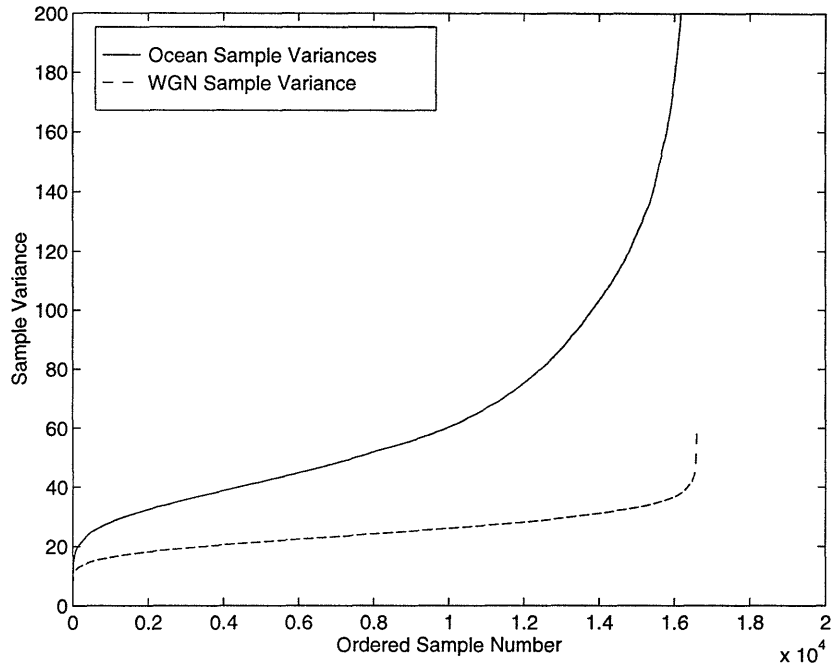


Figure 5-2: A sample variance was computed at each TOPEX/POSEIDON measurement location based on one year of data (36 samples). The solid line shows the sorted sample variances; the dashed line shows the distribution of sample variances computed from 36 samples of 5cm Gaussian noise.

(5.18), (5.19). We have chosen the latter approach – to divide the satellite measurements into ten day periods and to compute measurement residuals for each period – based on the following observations:

- Figure 5-2 plots the sorted pointwise sample variances of the ocean measurements over a period of one year; it is clear that the variance of (5.18) is well in excess of $(5\text{cm})^2$.

- From Figure 3-6 the ten-day variability of the ocean in the north Pacific is known to be small (considerably less than 5cm). Consequently the variance of (5.19) is on the order of $(5\text{cm})^2$.

A set of normalized residuals $\nu_i(x, y)$ was computed for each of 37 repeat cycles, $1 \leq i \leq 37$ as per (5.17). For each repeat cycle, multiscale estimates were computed on a 512 x 512 grid based on the average of 17 multiscale trees (each with an origin shifted relative to the others to attenuate artifacts due to the multiscale boundaries) using the multiscale model of (5.5).

Now since the ocean surface ζ is close to static for the ten-day duration of repeat cycle i (i.e., the spatial RMS variability of ζ over the ten days is considerably less than 5cm); i.e., $\zeta \approx \zeta_i$ for the duration of repeat cycle i; then

$$\nu_i(x,y) = \frac{Y_{S,i}(x,y) - Y_N(x,y) - \hat{\zeta}_i(x,y)}{\sqrt{\tilde{P}_i(x,y) + R(x,y)}} \tag{5.20}$$

$$= \frac{\zeta_i(x,y) - \tilde{N}_o(x,y) - \hat{\zeta}_i(x,y)}{\sqrt{\tilde{P}_i(x,y) + R(x,y)}} \tag{5.21}$$

$$= \alpha \tilde{N}_o(x,y) + \eta_i(x,y) \tag{5.22}$$

where

$$\alpha = \frac{-1}{\sqrt{\tilde{P}_i(x,y) + R(x,y)}} \qquad \eta_i(x,y) = \frac{\zeta_i(x,y) - \hat{\zeta}_i(x,y)}{\sqrt{\tilde{P}_i(x,y) + R(x,y)}} \tag{5.23}$$

These equations are defined only at those (x,y) corresponding to sample points of the satellite. What (5.22) implies is that the residuals contain one component proportional to the error in the OSU91A geoid and a separate random component

$$E[\eta_i(x,y)] = 0 \qquad \text{var}(\eta_i(x,y)) \approx 1 \tag{5.24}$$

Certainly the $\nu_i(x,y)$ at a given (x,y) for different values of i may be correlated, however to capture such correlations requires a multiscale model which is able to capture the evolution of the ocean surface over time; the development of such models remains an active area of research. Motivated by (5.22) it is reasonable to propose to average the residuals over the repeat cycles

$$\bar{\nu}(x,y) = \left| \sum_i^W \nu_i(x,y)/W \right| \tag{5.25}$$

where, based on (5.22) we expect that for W sufficiently large,

$$\bar{\nu}(x,y) \propto \tilde{N}_o \tag{5.26}$$

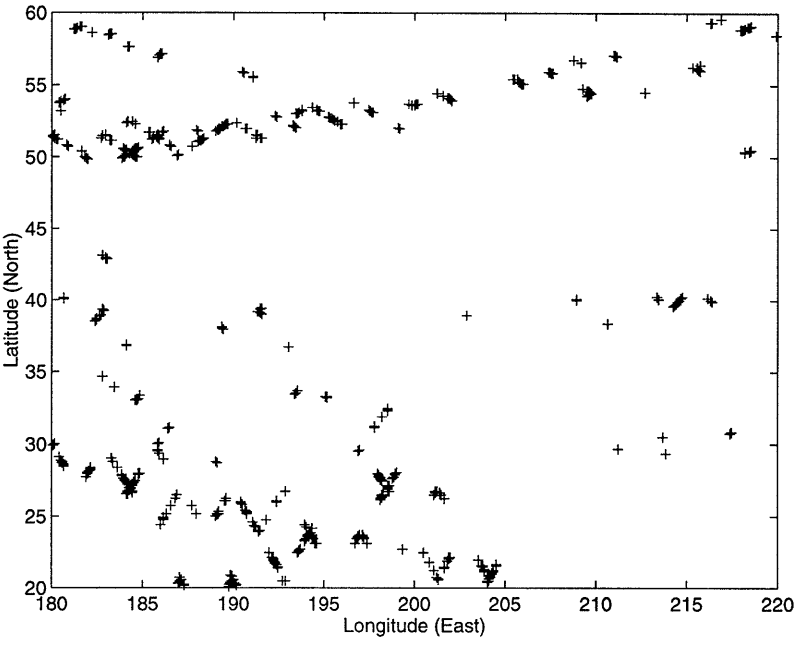


Figure 5-3: The figure shows the distribution of those measurement residuals (each marked with a '+' sign) in excess of 1.5σ.

The resulting normalized, averaged residuals which are in excess of 1.5 are shown in Figure 5-3. In the next section we will independently compute two functions ΔG_{ANOM}, ΔG_{TOPOG} which we believe are correlated with high wavenumber errors in the OSU91A geoid estimates. The need for such an independent function ΔG stems from the following: without knowing anything about the geoid, we have no way of discriminating between the geoid-related and random components of the residuals $\bar{\nu}$; e.g., a single, large residual may be due to an underwater mountain or due to an anomalously large satellite measurement error.

If such a function ΔG can be found, why not compute it and let $\phi = (\Delta G)^2$? Proposed forms of the function ΔG (for example the function based on the OSU91A geoid slope in Section 3.4) are correlated with the OSU91A geoid error, but the quantitative form of the correlation between ΔG and ϕ is unknown; it is the residuals process $\bar{\nu}$ which will establish the form of this correlation (in Section 5.4.3).

5.4.2 Computation of $\Delta G(x, y)$

We wish to find a function $\Delta G(x, y)$ which is correlated with the OSU91A geoid error, i.e., correlated with $\tilde{N}_o$. We propose to compute ΔG as the high-pass filtered geoid-anomaly associated with the topography of the earth; this approach is reasonable, since the high wavenumber features of the geoid are almost entirely dependent on *local* topographic features (which follows from (5.28),(5.29) — two transfer functions which relate the topography and the geoid); consequently we propose to ignore the subtleties of global topographical data processing[30, 107] and to assume a locally flat earth.

Consider a two-dimensional density anomaly $\Delta\rho_z(x, y)$ at a depth z below the surface of the earth.[2] The anomalous gravitational field, at depth z, associated with this density plane is

$$\Delta\gamma_z(x, y) = 2\pi G\Delta\rho_z(x, y) \tag{5.27}$$

where G is the gravitational constant. The corresponding anomalous gravitational field at the surface of the earth is related to that at depth z by means of the upwards continuity relationship[54]:

$$\Delta\gamma_o(x, y) = \Delta\gamma_z(x, y) * \frac{z}{2\pi(x^2 + y^2 + z^2)^{3/2}} \tag{5.28}$$

where $*$ represents the two-dimensional convolution operator. Finally, the anomalous gravitational field at the surface of the earth is related to the corresponding geoid anomaly[3] at the earth's surface by means of Stokes' relationship[54]:

$$\Delta G(x, y) = \Delta\gamma_o(x, y) * \frac{1}{2\pi g_o(x^2 + y^2)^{1/2}} \tag{5.29}$$

[2]We ignore the curvature of the earth in our computations – a reasonable assumption for local gravimetric processing.

[3]The geoid anomaly is a surface of constant anomalous gravitational potential, where the anomalous potential is that potential field associated with the density anomaly function $\Delta\rho_z(x, y)$. The geoid anomaly surface is typically chosen to have a mean height of sea-level.

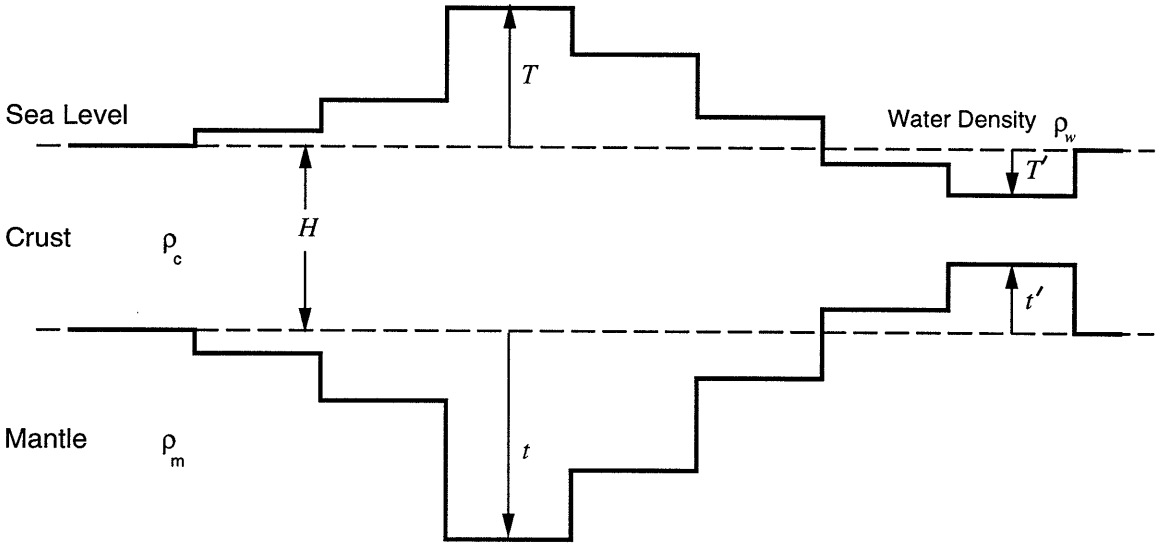


Figure 5-4: Airy-Heiskanen compensation for topographical data: mountains on the crust have corresponding "roots" which project into the mantle, allowing the mountains to "float" on the mantle; the opposite effect occurs over water.

where g_o is the gravitational acceleration at the earth's surface. In principle we can apply (5.28) and (5.29) to the problem of computing a geoid anomaly associated with an anomalous density distribution $\Delta\rho(x, y, z)$ inside the earth: we discretize this density distribution into a set of density anomaly planes $\{\Delta\rho_{z_i}(x, y)\}$ and approximate the geoid anomaly as the sum of the geoid anomalies associated with each of the $\Delta\rho_{z_i}(x, y)$.

For this chapter we will define the anomalous density distribution $\Delta\rho(x, y, z)$ as the difference between the true density distribution of the earth and the density distribution of a planet with a constant surface height and a constant crustal thickness (the dashed lines of Figure 5-4). The "true" density distribution of the earth is, of course, unknown. Instead, we will be approximating the shape of the crust based on high resolution (5 minute) data of the earth's topography[77] for the north Pacific; the thickness of the crust follows from a common assumption[82] regarding the response of the mantle-crust interface to surface topography and is given by the Airy-Heiskanen compensation method[42] (see Figure 5-4). Under this compensation, the crust is divided up into vertical columns; the thickness of the crust in each column is chosen so that the amount of mantle-mass displaced by the projection of the crust into the

mantle equals the mass of the crust above sea-level (i.e., we are satisfying a buoyancy constraint: the crust "floats" on the higher-density mantle). With respect to the notation shown in Figure 5-4, the Airy-Heiskanen compensation implies the following:

$$t(\rho_m - \rho_c) = T\rho_c \qquad \text{(on land)} \tag{5.30}$$

$$t'(\rho_m - \rho_c) = T'(\rho_c - \rho_w) \qquad \text{(on water)} \tag{5.31}$$

Typical parameter choices are[42]

$$H = 30\text{km}, \rho_m = 3.27\text{g/cm}, \rho_c = 2.67\text{g/cm}, \rho_w \approx 1\text{g/cm} \tag{5.32}$$

where H represents the mean thickness of the crust, and where the subscripts m, c, w represent the mantle, crust, and water respectively.

We computed the geoid associated with the topographical data as follows:

1. The topographical data as provided is regularly sampled in latitude and longitude. Consequently pixels at high latitudes have a much smaller size than those nearer the equator; in principle, such a change of scale would imply that a non-isotropic transfer function would be required to compute (5.28) or (5.29); the use of an isotropic transfer function (required in order to use FFT techniques) would introduce distortions. To minimize such distortions, we have regularly resampled the field in latitude and sin(latitude)·longitude.

2. From the resampled topographical data and the Airy-Heiskanen compensation method we determine the depth of the crust.

3. The crust is stratified into 1000m layers; for each layer we compute the upwards-continuation and Stokes' convolutions (in the frequency domain using FFT techniques), yielding the geoid anomaly associated with that layer. The total geoid anomaly associated with the topographical data is equal to the sum of the geoid anomalies by layer.

The effect of truncating the spherical harmonic expansion of the geoid at degree and order 360 is emulated by applying a high-pass filter to the computed geoid anomaly. The high-pass filter is implemented as the difference between the original field and the field after low-pass filtering. To avoid possible ringing effects, the low-pass filter is implemented in the time-domain as the convolution with a Gaussian kernel having an effective cutoff wavelength of one degree. The resulting high-passed geoid anomaly, denoted by $\Delta G_{\mathrm{ANOM}}(x, y)$, is plotted in Figure 5-5.

The computation of ΔG_{ANOM} has been somewhat involved. If we assume the topography to be perfectly correlated with the geoid at high wavenumbers (i.e., ignored the effects of Airy-Heiskanen compensation of Figure 5-4 and ignoring the effects of (5.28), (5.29)) we can propose a function $\Delta G_{\mathrm{TOPOG}}(x, y)$ which is computed by applying a high pass filter (the same filter as in the previous paragraph) directly to the topography data $T(x, y)$. The resulting function is plotted in Figure 5-6.

By way of comparison, Figure 5-7 plots ΔG_{OSU}, the gradient of the OSU91A geoid estimates; this gradient was used (with some justification, see Section 3.2) as a rough measure of the OSU91A error in Chapter 3.

Figures 5-5, 5-6, and 5-7 show three possible candidates for the function ΔG. It is possible to identify the basic features of the topography – the Aleutian trench and the Hawaiian island chain – from ΔG_{ANOM} (Figure 5-5). Furthermore there are a fair number of more detailed features present here which are not visible in the geoid slope map of Figure 5-7; this is encouraging, since by definition ΔG_{ANOM} is meant to capture fine scale features not resolved by the OSU91A estimates.

$\Delta G_{\mathrm{TOPOG}}$ (Figure 5-6) possesses the finest structures of the three functions. Since ΔG_{ANOM} is essentially a low-pass filtered version of Figure 5-6 (due to the nature of the upwards-continuation transfer function), it should come as no surprise that $\Delta G_{\mathrm{TOPOG}}$ shows the greatest amount of detail, including many seamounts not at all visible in the crude picture of Figure 5-7.

Section 5.4.1 computed the normalized measurement residuals, $\bar{v}(x, y)$, which contain a random component plus a component correlated with the OSU91A geoid error. This section has independently computed two functions ΔG_{ANOM}, $\Delta G_{\mathrm{TOPOG}}$, which

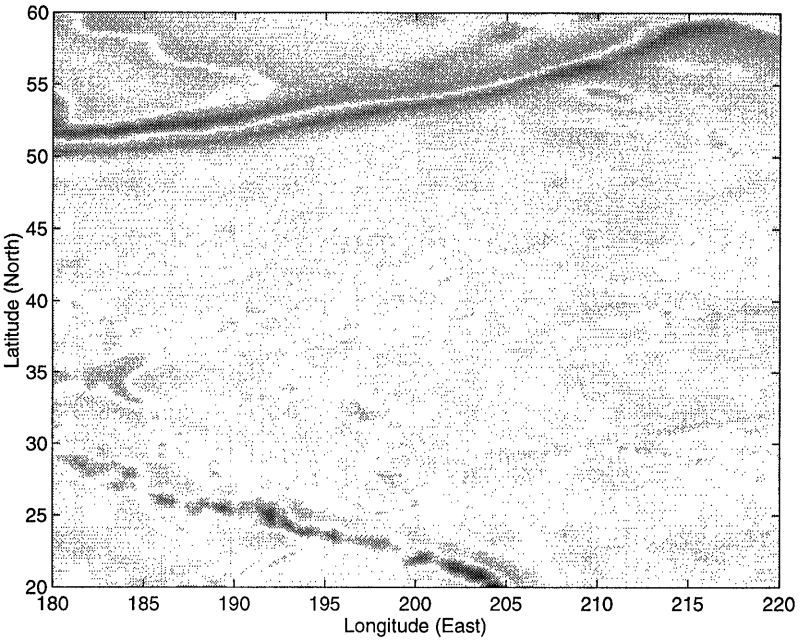


Figure 5-5: A plot of the absolute value of high-pass-filtered geoid anomaly estimates computed from topographical data. Darker regions represent greater values.

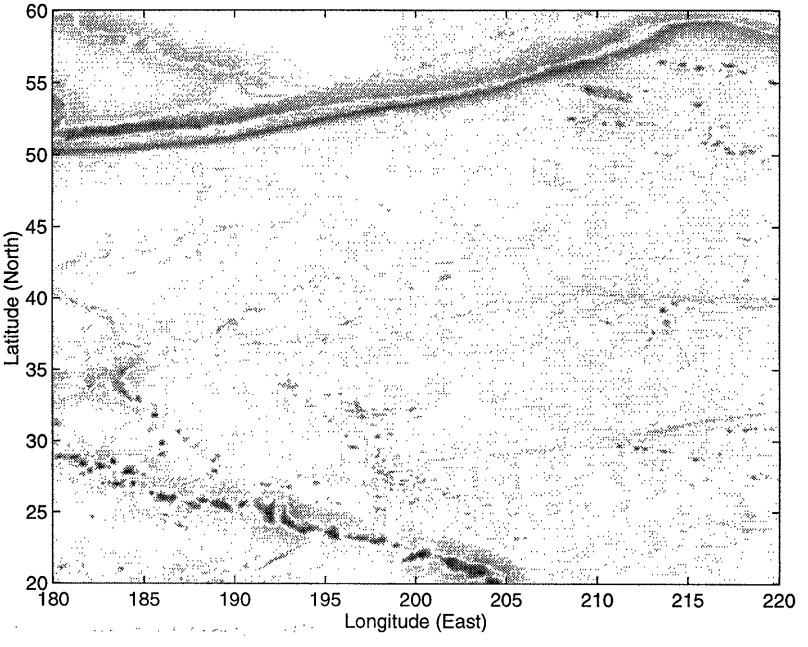


Figure 5-6: A plot of the absolute value of high-pass-filtered bathymetry. Darker regions represent greater values.

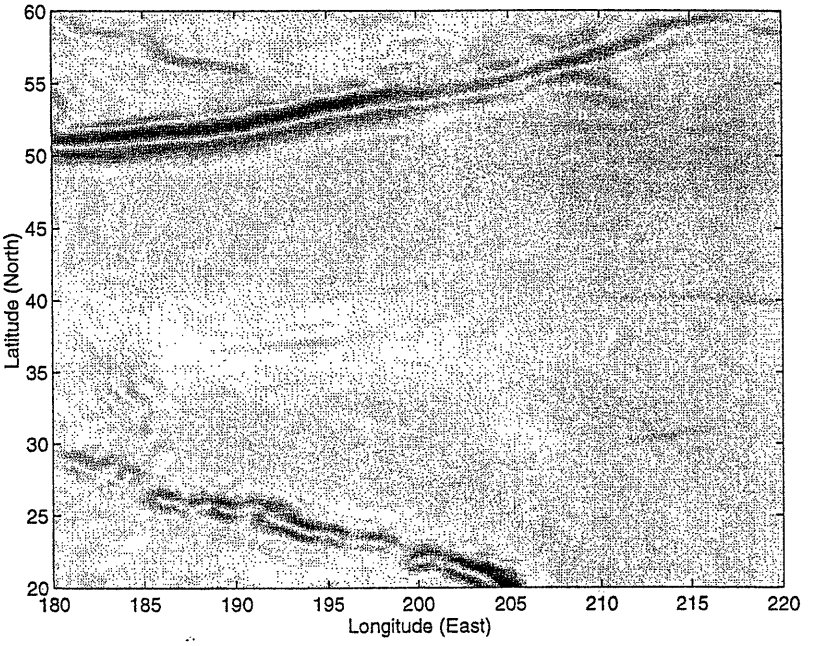


Figure 5-7: A plot of the geoid slope based on the OSU91A estimates. Darker regions represent higher slope.

we expect to be correlated with the OSU91A geoid error. The remaining task, to be addressed in the next section, is to assess the correlation between $\bar{\nu}(x, y)$ and ΔG in order to propose ϕ: the space-dependent error variance of the OSU91A model in (5.13).

5.4.3 Selection of ϕ

The chief difficulty in establishing a correlation between the normalized residuals $\bar{\nu}(x, y)$ of Section 5.4.1 and functions ΔG_{ANOM}, ΔG_{TOPOG}, believed to be correlated with the OSU91A error, stems from uncertainties in what it is that these functions represent. For example, consider a zero-crossing of ΔG_{ANOM} located between a large positive peak and a proximate large negative peak (e.g., point $\circ$ in Figure 5-9). To be sure, we do *not* expect the geoid error to be zero at that point, since the precise position of the zero crossing in ΔG is subject to error (due to errors in the topographical data $T(x, y)$, due to approximations made in applying the upwards-continuation and Stokes functions, and due to the choice of a high pass filter which

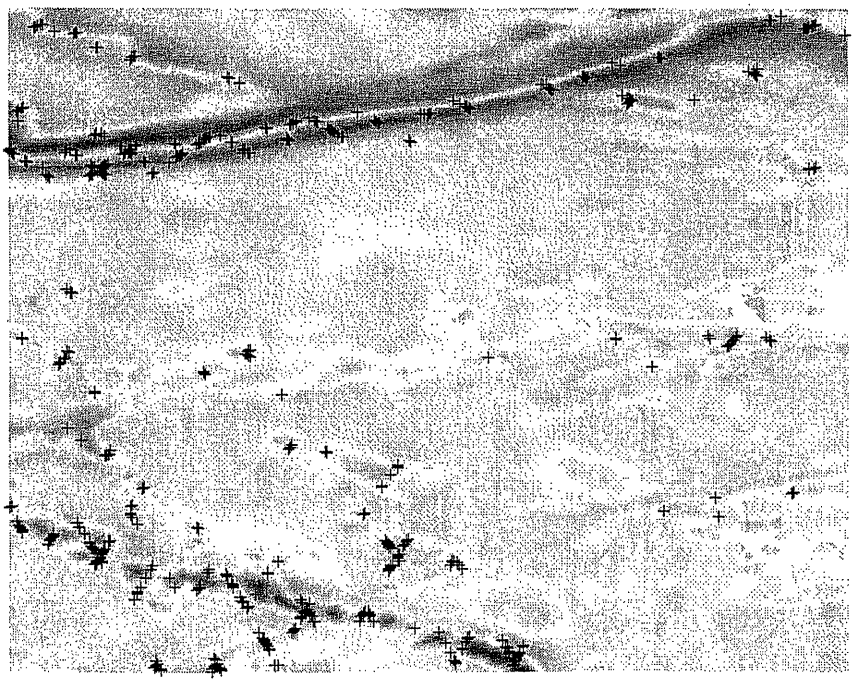


Figure 5-8: A plot showing the distribution of those measurement residuals (each marked with a '+' sign) in excess of 1.5σ superimposed on ΔG_{ANOM} — a function which we believe to be strongly correlated with errors in the OSU91A geoid estimates. Darker regions represent larger values of ΔG_{ANOM}.

only approximates the effect of spherical-harmonic series truncation); in fact, we would expect a relatively *large* geoid error *variance* at such points. This expectation is well supported by Figure 5-8: the 19 or so residuals in the top left corner of the figure all lie on or very near to such a zero crossing (the white line threaded between two darker regions); a number of other such examples may be found throughout the figure.

In response to this observation, we propose to apply a local maximum filter to ΔG_{ANOM} to raise such zero crossings to the largest nearby values in ΔG_{ANOM}:

$$\overline{\Delta G}_{\mathrm{ANOM}}(x,y) = \max\{|\Delta G_{\mathrm{ANOM}}(x',y')| \ni | x-x' |< \delta, | y-y' |< \delta\} \quad (5.33)$$

where $\delta \approx \frac{1}{4}$degree was chosen empirically as the typical size of the region separating large peaks of opposite sign. The effect of such filtering is illustrated in Figure 5-9. The same maximum filtering was applied to $\Delta G_{\mathrm{TOPOG}}$.

We are interested in computing ϕ; specifically, we are interested in finding a linear

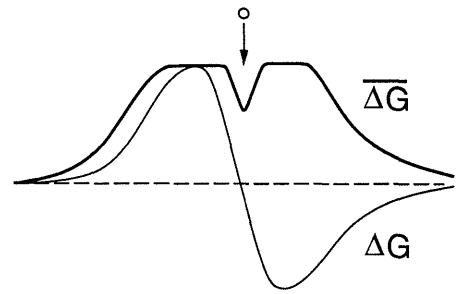


Figure 5-9: The effect of applying a maximum filter to a function ΔG. The intent is to increase the value at zero crossings such as $\circ$, which are expected to correspond to regions of significant OSU91A geoid error variance.

relationship between $\overline{\Delta G}$ and ϕ:

$$\phi(x,y) \approx \kappa + \lambda \overline{\Delta G}(x,y) \tag{5.34}$$

Consider some scalar value g; under the assumption that $\overline{\Delta G}$ is roughly proportional to the OSU91A geoid error, then

$$\text{At all } (x,y) \ni \overline{\Delta G}(x,y) \approx g \longrightarrow \phi(x,y) \approx \kappa + \lambda g \tag{5.35}$$

Next, let

$$V = \left\{ \bar{\nu}(x,y) \ni \overline{\Delta G}(x,y) \approx g \right\} \tag{5.36}$$

represent a set of normalized, averaged residuals from (5.25). The OSU91A error $\tilde{N}_o$ is modeled as a random process with variance ϕ, thus from (5.22) we expect each residual in V to satisfy

$$\left(\tilde{P} + R\right) \text{var}(\bar{\nu}) \approx \phi + \frac{1}{37} \tag{5.37}$$

We can estimate this variance by examining the RMS value of the residuals in V; i.e., by computing the RMS value of binned residuals, where all of the residuals in the bin are associated with approximately the same value of $\overline{\Delta G}$. Figures 5-10 and 5-11 characterize this RMS relationship between the averaged, normalized residuals (5.25)

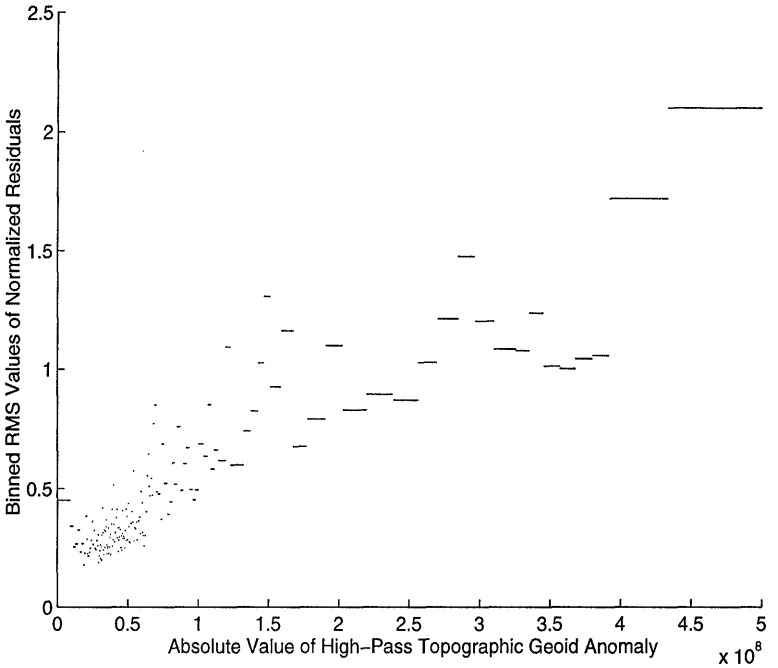


Figure 5-10: This figure plots the RMS value of binned normalized, averaged measurement residuals $\bar{\nu}$ as a function of $\overline{\Delta G}_{\mathrm{ANOM}}$ — the function obtained by gravimetrically processing topographical data, followed by high-pass and maximum filtering. Each bin consists of 100 residuals, all of which correspond to nearly the same value of $\overline{\Delta G}$.

and the functions $\overline{\Delta G}_{\mathrm{ANOM}}$, $\overline{\Delta G}_{\mathrm{TOPOG}}$ (5.33).

The RMS values in the figure were computed as follows:

- form the following set of two-vectors

$$\mathcal{Q} = \left\{ \left[\overline{\Delta G}_{\mathrm{ANOM}}(x,y), \quad \bar{\nu}(x,y) \right] \right\} \tag{5.38}$$

  over all points (x, y) where $\bar{\nu}$ is defined (i.e., at all of the satellite sampling locations);

- sort all of the elements of $\mathcal{Q}$ in ascending order of $\overline{\Delta G}_{\mathrm{ANOM}}$;

- now divide this sorted list into successive groups, each containing 100 two-vectors;

- compute the RMS of the 100 residuals contained in each of these groups.

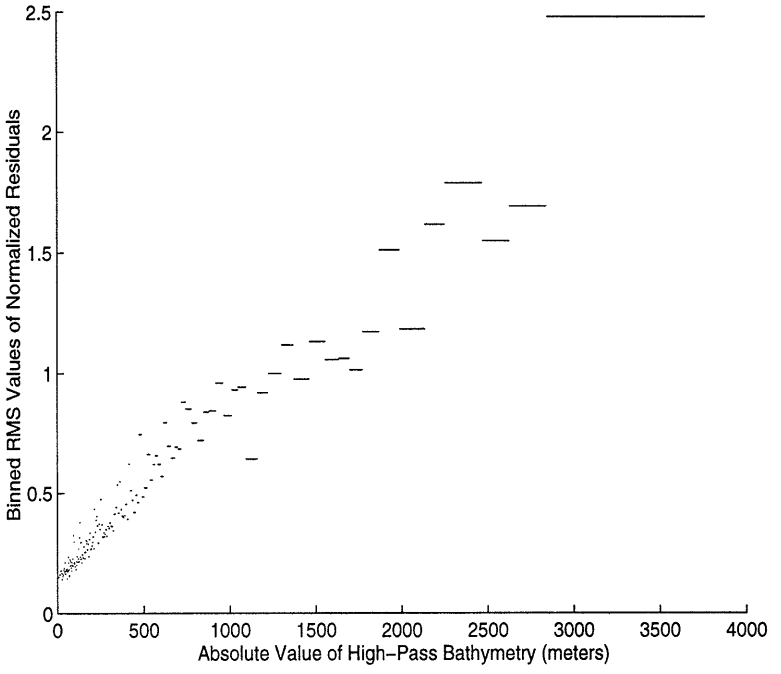


Figure 5-11: This figure plots the RMS value of binned normalized, averaged measurement residuals $\bar{\nu}$ as a function of $\overline{\Delta G}_{\mathrm{ANOM}}$ — the function obtained after high-pass and maximum filtering. Each bin consists of 100 residuals, all of which correspond to nearly the same value of $\overline{\Delta G}$.

Each of the horizontal lines in Figure 5-10 represents the RMS value of the residuals of one group; the horizontal extent of the line goes from the minimum value to the maximum value of $\overline{\Delta G}_{\mathrm{ANOM}}$ in each group. Figure 5-11 contains a similar plot, but based on $\overline{\Delta G}_{\mathrm{TOPOG}}$ instead of $\overline{\Delta G}_{\mathrm{ANOM}}$. Several observations should be made:

- Consider the extrapolation of Figure 5-10 or Figure 5-11 leftward to the vertical axis; i.e., to the point $\overline{\Delta G}_{\mathrm{ANOM}} = 0$ or $\overline{\Delta G}_{\mathrm{TOPOG}} = 0$ respectively. Now $\overline{\Delta G}_{\mathrm{ANOM}} \approx 0$ implies an anticipation of a very small OSU91A error variance; consequently from (5.22) we expect that any residuals $\nu_i(x_j, y_j)$, computed from measurements taken over repeat cycle i, for which $\overline{\Delta G}_{\mathrm{ANOM}}(x_j, y_j) \simeq 0$ should satisfy

$$E\left[\nu_i(x_j, y_j)\right] \simeq 0 \qquad \mathrm{var}\left(\nu_i(x_j, y_j)\right) \simeq 1 \tag{5.39}$$

furthermore implying that the averaged residuals (5.25) satisfy

$$E\left[\bar{\nu}(x_j, y_j)\right] \simeq 0 \qquad \text{var}\left(\bar{\nu}(x_j, y_j)\right) \simeq \frac{1}{37} \tag{5.40}$$

Then the RMS value of such residuals will be

$$\left(\sum_j \bar{\nu}(x_j, y_j)^2\right)^{\frac{1}{2}} \simeq \sqrt{\frac{1}{37}} \simeq 0.164 \tag{5.41}$$

which is very close to the y-intercept suggested by both figures.

What does this agreement serve to validate? The value deduced in (5.41) relies on the fact that the normalized residuals are in fact correctly normalized to unit variance in (5.39), which relies upon the accuracy of the sum $(R + \tilde{P})$. In other words, (5.41) and Figure 5-11 validate the selected value of the measurement error variance R and the accuracy of the estimation error variance $\tilde{P}$.

- $\overline{\Delta G}_{\text{ANOM}}$ and $\overline{\Delta G}_{\text{TOPOG}}$, measured along the horizontal axis in the figures, were constructed with the intent of being proportional to the error in the OSU91A geoid. Furthermore, from (5.22) we expect that the averaged, normalized residuals (5.25), measured along the vertical axis in the figure, be proportional to the error in the OSU91A geoid. However the residuals $\bar{\nu}$ and the $\overline{\Delta G}$ functions were computed completely independently, thus it is encouraging that both Figure 5-10 and Figure 5-11 admit relatively straight-line fits. It should be pointed out that the geoid-slope based correlation determined in Chapter 3 (Figure 3-17) was not such a straight line.

- The relationship between $\bar{\nu}$ and ΔG_{TOPOG} in Figure 5-11 is straighter and more convincing than that based on ΔG_{ANOM} in Figure 5-10. We offer two possible explanations for this observation:

  1. The true coherence between the topography and the geoid is greater than that associated with the Airy-Heiskanen compensation of Figure 5-4.

2. Some of the approximations made in the computation of $\overline{\Delta G}_{\text{ANOM}}$ (i.e., the FFT implementation of the upward-continuation (5.28) and Stokes (5.29) transfer functions) may be inappropriate.

A more careful assessment of the above explanations is a suggested matter for future inquiry.

From Figures 5-10 and 5-11 we propose the following straight-line fits:

$$\bar{\nu} \simeq 0.16 + \overline{\Delta G}_{\text{ANOM}}/(16 \cdot 10^7) \tag{5.42}$$

$$\bar{\nu} \simeq 0.16 + \overline{\Delta G}_{\text{TOPOG}}/1500 \tag{5.43}$$

We could use these two relationships to directly estimate the geoid. For example, from (5.43) we can infer the following "measurements" of the OSU91A geoid error:

$$\overline{\Delta G}_{\text{TOPOG}} = N - Y_N + v_{\text{TOPOG}} \tag{5.44}$$

The exploration of such an estimator is proposed as an avenue of future research. For the remainder of this chapter, we will continue to use the model of (5.13), which requires the determination of ϕ. From (5.22) we can compute ϕ as

$$\phi \simeq (R + \tilde{P})(\bar{\nu} - 0.16)^2 \tag{5.45}$$

$$\simeq 35 \left(\frac{\overline{\Delta G}_{\text{ANOM}}}{16 \cdot 10^7} \right)^2 \tag{5.46}$$

or, alternatively,

$$\phi \simeq 35 \left(\frac{\overline{\Delta G}_{\text{TOPOG}}}{1500} \right)^2 \tag{5.47}$$

where the constant 35 represents an empirical average over all residual locations of the quantity $(R + \tilde{P})$.

5.5 Joint Estimation Results

Having established a form for ϕ in the previous section, the computation of joint estimates follows in a straightforward manner from the multiscale model (5.5), (5.12), (5.13). Once again we may consider whether we are better off taking as measurements the average of multiple repeat cycles of satellite data, or just a single ten-day cycle. The answer here remains unchanged from the answer argued in Section 5.4.1:

- Estimate a geoid based on measurements from a single ten-day repeat cycle, computed as the average of 13 multiscale trees, each having a tree origin shifted with respect to the others.
- Perform the joint estimation for each of ten repeat cycles.
- Compute the final geoid estimates and estimation error variances as the average of the ten estimates and error variances respectively, one from each repeat cycle.

Due to the large dynamic range of the geoid and the relatively small size of the OSU91A geoid error, Figure 5-12 plots the *difference*

$$N_{\text{estimated}} - N_{\text{OSU91A}} \tag{5.48}$$

i.e., the estimated correction to the OSU91A geoid. The error variances corresponding to Figure 5-12(a) are plotted in Figure 5-13. A number of observations are in order:

- Some blockiness is evident in the estimated corrections, particularly along the northern (upper) edge. The blockiness is this region is due to the presence of land (i.e., an absence of measurements).
- The satellite tracks are visible in the estimated corrections. Our multiscale estimator is trying to estimate *high-resolution* features of the geoid based in part on sparsely sampled altimetric measurements. Those high-resolution features which lie on the satellite tracks may be resolved; other features which lie wholly unsampled between the satellite tracks are not observed and cannot be esti-

mated. Thus a certain degree of feature clustering about the satellite paths is to be expected.

- The dominant features of the estimated corrections are consistent with our expectations based on the north-Pacific topography. The largest corrections are predicted along the Hawaiian island chain (lower left) and the Aleutian archipelago (top).

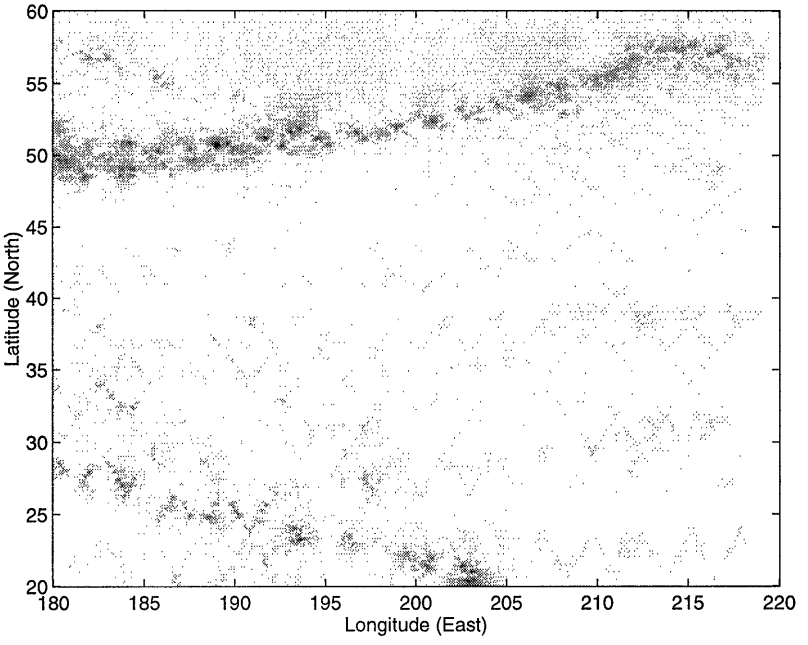
5.6 Conclusions

This chapter has considered the problem of estimating the error in the OSU91A geoid by developing a joint multiscale model for both the ocean height and the geoid. The separability of the satellite signal into oceanic and geodetic components was motivated by the distribution of measurement residuals in Chapter 3, which formed the basis for the work of this chapter. Guided by an understanding of the omission errors in the OSU91A geoid estimates, we postulated two functions which were expected to correlate with the geoid errors. Correlating these functions with the measurement residuals completed the determination of the estimation of the space-varying geoid error variances.

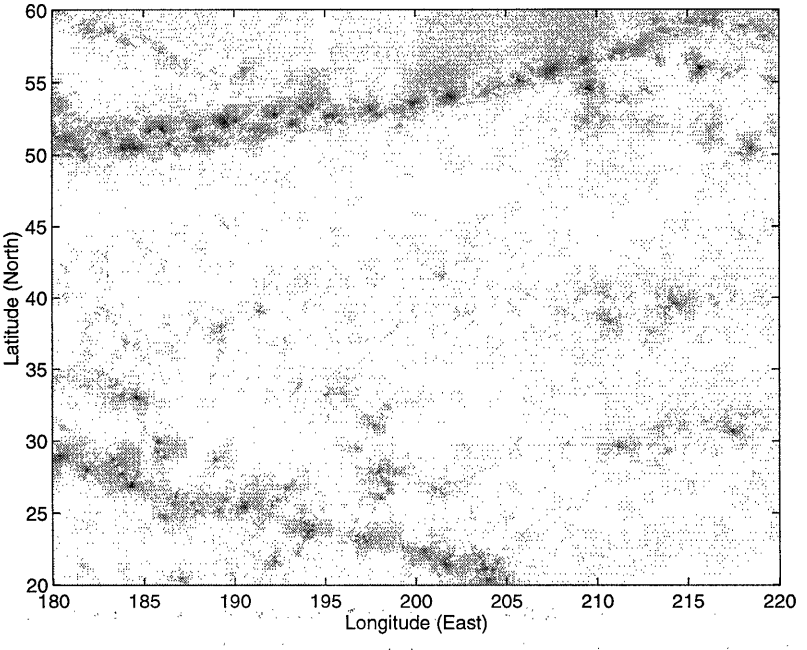
As was indicated in the first paragraph of this chapter, the results which have been presented are only preliminary. There are several matters which should be investigated before the revised estimates of the geoid in Section 5.5 can be accepted as reasonable:

- We need to look more closely at the assumptions implicit in the calculation of $\overline{\Delta G}_{\mathrm{ANOM}}$. In particular, why was the determined function $\overline{\Delta G}_{\mathrm{ANOM}}$ a relatively poor indicator of the measurement residuals in Figure 5-10?

- We need to validate the multiscale model (5.12) which was asserted for the geoid. This validation involves looking at more than just the two coefficients in the selected $1/f$-like model; we need to assess the appropriateness of the $1/f$-like model in representing the geoid.



(a)



(b)

Figure 5-12: The estimated corrections to the OSU91A geoid. The two plots differ in the function ϕ chosen to model the geoid error variance:
(a) ϕ based on $\overline{\Delta G}_{\mathrm{ANOM}}$ (5.46).
(b) ϕ based on $\overline{\Delta G}_{\mathrm{TOPOG}}$ (5.47).

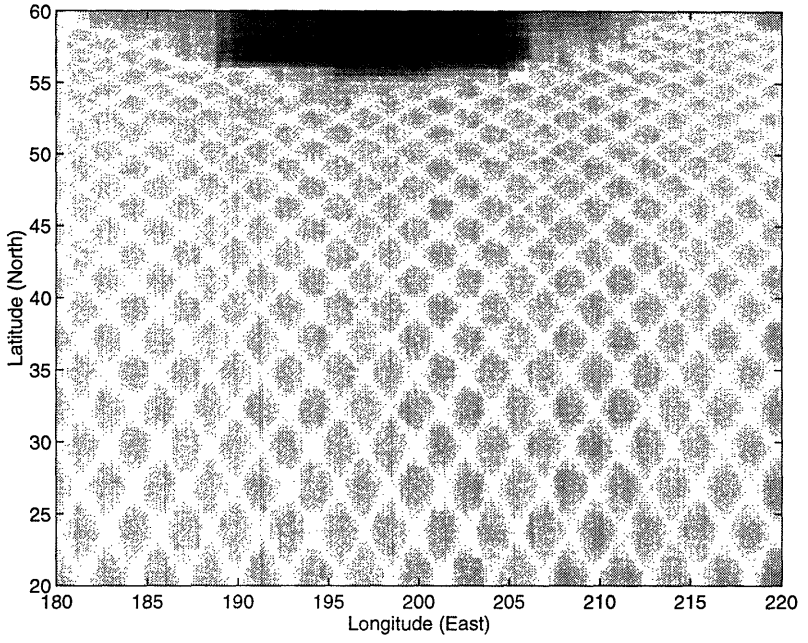


Figure 5-13: Estimation error variances corresponding to the geoid estimates of Figure 5-12(a). Darker regions represent greater uncertainty. The standard deviation Near the satellite paths, the standard deviation of the error is $\sigma \approx 4$cm; between the satellite paths $\sigma \approx 2$m; at the top edge of the figure (over Alaska), $\sigma \approx 10$m.

Given that these validation have not yet been performed, it is somewhat premature to embark on a variety of extensions to the joint estimation research of this chapter. Nevertheless, there are a variety of research directions which could be followed once a joint ocean / geoid estimator has been fully developed and validated. Indeed, the flexibility of the multiscale framework suggests a number of interesting avenues for future exploration:

- We are computing estimates based on a static prior model of the ocean surface. A multiscale model which incorporates some understanding of the time evolution of the ocean surface should be capable of affecting a greater separation of the oceanic and geodetic components of the altimetric measurements.

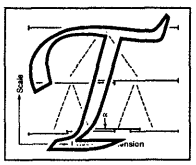
- A variety of sensitivity analyses are possible; e.g., the sensitivity of the estimated OSU91A geoid errors to the power law in the prior model of the geoid. The efficiency of the multiscale framework allows a large number of parameter settings to be examined.

- It is possible to perform an independent validation of the geoid estimates. Complex ocean circulation models have been developed which can assess the self-consistency of an ocean height field ζ. Such methods could be used to compare the relative consistencies of ζ estimated in Chapter 3 and the ζ estimated jointly in this chapter.

- It may be possible to jointly estimate ζ, N *and* T (where T represents the surface topography). Although T is fairly well known over land, the shape of the ocean floor is still subject to considerable uncertainties. In principle such joint estimation is possible; for example, "bumps" that consistently appear in the satellite measurements with no counterpart in the topographic data may suggest the presence of an unknown seamount. Although it is not clear how to go about developing a multiscale model for such joint estimation problems, this problem may well represent an intriguing direction of future research.

Chapter 6

Overlapping Multiscale Trees



his chapter introduces and develops the concept of *overlapping* multiscale models. These models, which represent something of a conceptual departure from other models developed for our multiscale framework, were developed with the specific interest of producing *smooth* estimates (i.e., without "blocky" artifacts). Section 6.1 introduces and motivates the overlapping tree concept. Section 6.2 develops a simple example in some detail, considering both multiscale modeling and multiscale estimation using overlapping trees; the example is presented both in order to develop an intuition for the overlapping tree concept, and in order to identify the important issues that must be confronted to turn the technique into a systematic, useful tool. Section 6.3 outlines the process of selecting a particular overlapping model and describes how to derive the projection operators associated with that choice of model. Finally, in Section 6.4, we show three examples of the effectiveness of the overlapping approach to estimation.

6.1 Introduction

In spite of the success of our multiscale approach to estimation with regard to computational efficiency, mean-square estimation error, and ability to supply error covariance information, the multiscale approach, as developed up to this point in the thesis, has a characteristic that would appear to limit its utility in certain appli-

cations. Specifically, estimates based on the types of multiscale models previously proposed may exhibit a visually distracting blockiness [63].

While various interpretations of and ways to overcome this blockiness have been developed, discussed, and shown to be more than adequate in particular applications, none of these offers a completely satisfactory resolution of this issue in general. For example an interpretation is provided in [63] in which the authors argue correctly that in many applications the construction of fine-scale estimates is not supported by the quality of available data; instead, in such cases only coarser-scale estimates are statistically significant. In these applications, one should be suspicious of *any* fine-scale estimate of the field in question, and any corresponding blockiness has a complete lack of statistical significance. However, in some applications such as the problem of estimating the ocean surface height in Chapter 3, multiscale-based estimates are subsequently used in a manner that requires the calculation of surface gradients and normals; in these cases, there is an essential need for having smooth estimates, so that the gradients and normals can be calculated meaningfully. Similarly in certain applications of computer vision, such as investigation of surface reconstruction in Chapter 7, there exist aesthetic reasons for desiring smooth estimates.

Although estimate blockiness can be eliminated by simple post-processing (e.g., the application of a low pass filter), the resulting increase in smoothness and visual appeal comes at a price. In particular, such post-processing can render less clear the proper interpretation of error covariance information provided by the smoothing algorithm, and it limits the resolution of fine-scale details in the post-processed estimate, since the added smoothness is achieved by spatial blurring.

The preceding remarks suggest that for applications in which the computational efficiency of the multiscale framework is desired, but where blockiness is unacceptable, we have considerable motivation for seeking a new approach to both multiscale modeling and estimation. There are several interpretations regarding the "blocky" artifacts present in some multiscale estimates:

1. The multiscale prior model for the tree is nonstationary.

2. Fine scale pixels separated by multiscale boundaries (in particular coarse scale boundaries) are inadequately correlated.

3. The descendants of a given node (e.g., the q descendants of the root node) evolve independently, leading to artifacts at their boundaries.

Each of these interpretations motivates a different solution:

1. The nonstationarity of the multiscale estimates is due to the boundaries between tree nodes, the positions of which are determined by some (typically arbitrary) origin for the tree. One can attenuate this boundary effect by computing estimates as an average over an ensemble of W tree-origins:

$$\hat{z}(x,y) = \frac{1}{W}\sum_{i=1}^{W} \hat{z}(s_i(x,y)) \tag{6.1}$$

where the finest-scale node $s_i(x,y)$ on tree $i, 1 \leq i \leq W$, corresponds to the spatial position (x,y); i.e., (6.1) is strictly an average over an ensemble of tree origins and bears no resemblance to spatial filtering. This technique was applied in the oceanographic and geodesic applications of Chapter 3 and Chapter 5.

The computational complexity for determining $\hat{z}(x,y)$ from W trees is $\mathcal{O}(Wk^3K)$, where k represents the size of each multiscale state vector, and where K counts the number of pixels being estimated; i.e., a complexity per pixel of $\mathcal{O}(Wk^3)$. The idea, then, is to select a relatively small value of k and a fixed value of W independent of K. However the empirical observation that large values of W may be required for smooth estimates from sparsely sampled fields and the fact that we are unable to calculate error statistics for $\hat{z}(x,y)$ in a consistent manner motivate an exploration of other techniques for artifact attenuation.

2. The correlation between the descendants of a common parent can increase as the information (i.e., the state vector dimension, k) of the parent increases. For example, a multiscale model of a smooth surface which passes height *and*

gradient information from parent to child nodes may be expected to exhibit fewer artifacts than a surface model which passes only height information.

In an extreme case of such an approach, discussed in Section 2.5.1, the state vector at the parent contains a dense representation of the boundaries around each of its four descendants. This approach can, in principle, exactly model any first-order Gauss-Markov process on the multiscale tree and has been successfully applied[65] to a problem of texture discrimination. A less extreme approach, discussed in Section 2.5.2, leads to both optimal and reduced order multiscale models for Gaussian processes.

3. That the children of a common parent evolve independently is due to the fact that the children represent completely disjoint subsets of the process of interest (typically the process residing on the finest scale of the tree).

   To insist that the children of a common parent represent disjoint subsets of a process would seem to be reasonable, or perhaps even necessary; it is possible, however, to develop a novel "overlapping" framework, in which the regions represented by the child nodes of a common parent may overlap, rather than being strictly disjoint.

In this chapter we will consider a novel approach that blends the latter two approaches just discussed to yield the desired artifact attenuation. Our approach simultaneously achieves three objectives:

1. It yields low-dimensional multiscale models that are quite faithful to the prespecified random field covariance structure to be realized, and thus admits an extremely efficient, optimal (or nearly optimal) estimation algorithm;

2. The resulting estimation algorithm retains one of the most important advantages of the multiscale estimation framework, namely the efficient computation of estimation error covariances;

3. Both the multiscale models and the corresponding estimation algorithm eliminate the blockiness associated with previously developed multiscale models and

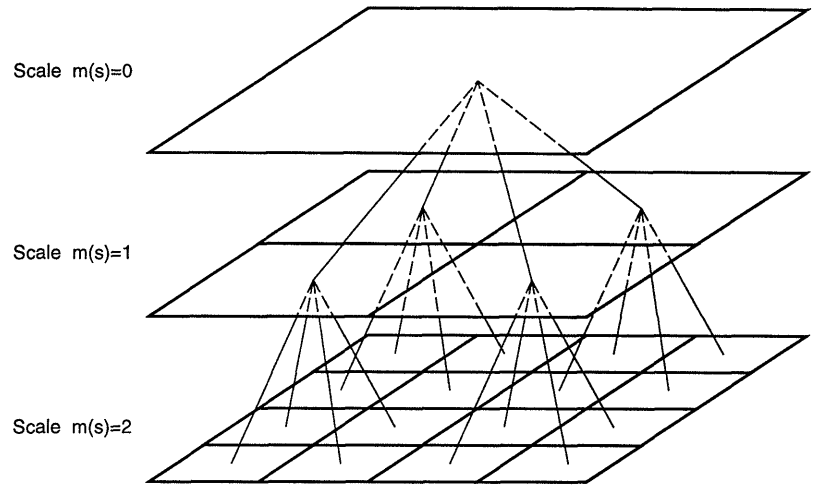


Figure 6-1: Illustration of the first three levels of a quad-tree, which is useful for the indexing of multiscale representations of two-dimensional random fields.

estimates.

In contrast to standard multiscale processing[27, 63], which achieves objectives one and two, and to standard multiscale processing with simple post-processing[63], which achieves objective one and partially achieves objective three, our approach is able to accomplish all three objectives.

6.1.1 Higher Order Models

Recall the following important *Markov* property of our multiscale models: if $x(s)$ is the value of the state at node s, then conditioned on the value of $x(s)$, the states of the q children of s are independent. In light of the Markov property, the role of the state at any node in a multiscale process is to store enough information about the process in the parent to decorrelate the q subsets of the process corresponding to the q children. It is this decorrelation which can lead both to efficient estimation algorithms and to the source of the problem with blockiness.

We can clearly see the connection between the decorrelating role of state information and the blockiness problem by considering Figures 6-1 and 6-2. Consider the upper-left and upper-right quadrants of the image domain depicted in Figure 6-1: the two quadrants are separated at the coarsest level of the tree, therefore all of the

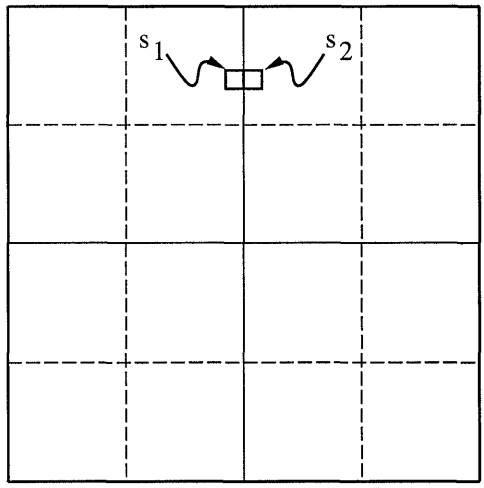


Figure 6-2: Two nodes, s_1 and s_2, neighbors in physical space, but distantly separated in tree space.

correlation between any two finer scale pixels in the two quadrants, such as s_1 and s_2 in Figure 6-2, must be completely captured in their common ancestor, namely the root node 0 at the coarsest scale of the tree. In this sense, the pixels s_1 and s_2 may be close physically, but they are separated considerably in terms of the distance to their nearest common ancestor node. We refer to this latter distance as so-called *tree-distance*; with respect to tree distance, pixels s_1 and s_2 are far apart. For fields having a significant level of regularity or smoothness we may expect high correlations between such spatially close neighbors. Such correlations imply a high state dimension $x(s_0)$, in essence to keep track of all of the correlations across quadrant boundaries (basically leading to a MRF-like model, as in Section 2.5.1).

One way to reduce this high dimensionality is to identify and retain only the principal sources of correlation across boundaries at each level on the tree. Keeping only these principal sources effectively achieves maximal decorrelation of descendants with minimal dimension of state variables. Indeed, as discussed in Section 2.5.2, a systematic procedure has been developed[47] to identify the needed principal sources of decorrelating information, and thus to build multiscale models of any desired fidelity. The procedure is based on the application of canonical correlation analysis; given a maximum model order and a finest-scale, desired correlation structure, an appropriate multiscale model can be built. While this approach by itself can yield

low-dimensional models of sufficient fidelity for many applications (such as texture discrimination [65] or problems such as that in [63] where only coarse-scale estimation is meaningful), it cannot overcome the blockiness problem. In particular, neglecting even a small amount of correlation at a coarse level of the tree can cause noticeable irregularities across boundaries such as that separating s_1 and s_2 in Figure 6-2.

6.1.2 Overlapping Trees

As was alluded to earlier, there is a means by which to correlate pixels such as s_1 and s_2 in Figure 6-2 without resorting to high-order multiscale models: we discard the standard assumption that distinct nodes at a given level of our tree correspond to disjoint portions of the image domain. Instead we construct models in which distinct tree nodes correspond to *overlapping* portions of the image domain. As a consequence of this idea, a given physical image pixel may now correspond to *several* tree nodes at the finest scale. In this way, we remove the hard boundaries between image-domain pixels such as s_1 and s_2 in Figure 6-2, because now multiple tree nodes may contribute to each of these pixels, thus reducing the effective tree distance[1] between the two sets of nodes corresponding to these pixels and spreading the correlation that must be captured among a set of nodes. For obvious reasons, we refer to these multiscale models as overlapped-tree models.

We can use these overlapped-tree models for both modeling and estimation, as depicted abstractly in Figure 6-3. In both of these contexts, we start with a knowledge of the correlation structure P of some random field x. Corresponding to this random field x, we devise a so-called *lifted-domain* version x_l, where this lifted-domain field lives at the finest-scale of an overlapped-tree multiscale representation of x. The mapping from x to x_l is denoted by $x_l = G_x x$, where we emphasize that this operator

[1]The notion of an *effective tree distance* has not been carefully defined. Essentially, on an un-overlapped tree the tree distance $d(s_1, s_2) = (M-1)$, for s_1 and s_2 in Figure 6-2. Now suppose we have an overlapped tree, where the nodes $\{s_{1,i}\}$ correspond to pixel 1, and similarly $\{s_{2,j}\}$ correspond to pixel 2. Then $d(s_{1,i}, s_{2,j}) \leq (M-1)$, and typically $\text{Average}_{i,j}(d(s_{1,i}, s_{2,j})) \ll (M-1)$. This last expression is one possible effective tree distance measure; in any event it helps to explain why we expect multiscale artifacts to be eliminated by the overlapping approach.

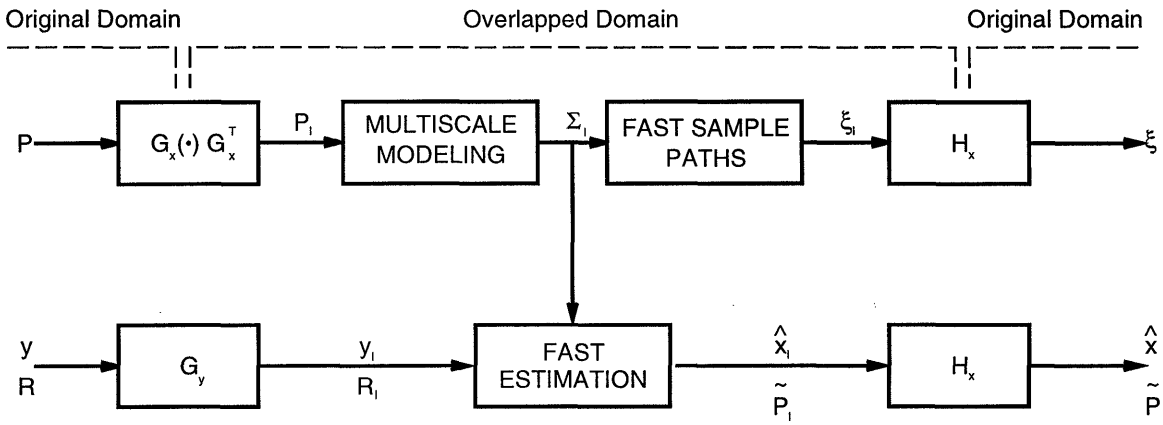


Figure 6-3: An abstract, high level, view of our overlapped approach to multiscale-based modeling and least-squares estimation. Fast multiscale estimation and sample-path generation (producing possibly blocky ξ_l and $\hat{x}_l$ respectively) are accomplished in the overlapped domain. G_x projects the statistics of x into the overlapped domain; G_y projects measurements y into the domain; H_x, which possesses certain smoothness properties, projects the estimates $\hat{x}_l$ back out of the overlapped domain.

G_x is one-to-many: the lifted-domain field x_l has more pixels than the image-domain field x. To map back from x_l to x, we devise an operator H_x having two important properties:

1. the field $H_x x_l$ has exactly, or nearly exactly, the same correlation structure as x;

2. the field $H_x x_l$ is guaranteed to have the desired level of smoothness.

The boxes in the top half of Figure 6-3 depict the use of our overlapped-tree models to efficiently generate sample paths of a random field having the prespecified correlation structure P. Given P, a low-order multiscale model is built to approximately realize the correlation structure of the overlapped field x_l; we denote this correlation by P_l, where $P_l = G_x P G_x^T$. Because of the low order of this multiscale model, sample paths can be generated in a computationally efficient manner, and by post-processing these sample paths with the smoothing operator H_x, we obtain sample paths of a random field that are guaranteed to be smooth and that approximately have the desired correlation P. We will address the technical problem of constructing both the tree model *and* the lifting and interpolation operators G_x and H_x; we will

construct these to meet the objectives of low-dimensional states on the tree, accurate approximation of the desired second-order statistics of the field x, and the generation of fields without blocky artifacts.

The bottom half of Figure 6-3 depicts the application of our overlapped-tree models to optimally estimating the value of a random field x, given noisy observations y. For this purpose, we devise an operator G_y that plays a role directly analogous to the role of G_x: the operator G_y lifts the actual observations y of the random field to yield lifted-domain observations $y_l = G_y y$ of the random field x_l. These observations are then processed by our efficient multiscale tree algorithm to produce an estimate $\hat{x}_l$ which is then projected back to yield $\hat{x} = H_x \hat{x}_l$, the desired estimate of the random field. The low dimensionality of the multiscale model allows the estimation calculations to be carried out in an extremely efficient manner, and the properties of the operator H_x guarantee that the resulting estimates will have the desired level of smoothness. We address the technical problem of justifying the optimality, or near optimality of this estimation procedure, and we also demonstrate that estimation error covariance information can be generated in an efficient and meaningful way.

6.1.3 Computational Complexity

To make the computational tradeoffs explicit, this subsection will present expressions for the computational complexity of both the multiscale modeling[2] and estimation algorithms. These expressions will highlight one of the strengths of the multiscale framework and will also point to some of the challenges.

There are three multiscale model parameters of fundamental interest in our discussion:

(i) the number of pixels K in the image domain,

(ii) the number of finest-scale nodes N in the multiscale model,

(iii) the maximal dimension k of any state vector $x(s)$ in the multiscale model.

[2]By "modeling" we refer to the simulation of sample paths of the modeled process. This is achieved by just simulating the multiscale recursion (B.1) based on initial conditions (B.4).

In previous chapters, where the trees have not been overlapped, N has always been identically equal to K. On the other hand, the use of overlapping trees leads to values of N that are greater than K; we relate the two by $K = \eta N$, where $0 < \eta < 1$ is a measure of the degree of overlap, with smaller η corresponding both to more overlap and potentially greater smoothness.

The two-sweep structure of our estimation algorithm (summarized in Appendix B) implies that each node of the tree is visited exactly twice, where the computations at each node involve a number of floating point operations proportional to the cube of the state vector at the given node. Thus the application of the estimation algorithm requires a total of $\mathcal{O}(k^3 N)$ floating point operations; that is, a computational effort per pixel of

$$\mathcal{O}\left(\frac{k^3}{\eta}\right) \tag{6.2}$$

or, if the algorithm is fully parallelized on each scale,

$$\mathcal{O}\left(\frac{k^3 \log(K/\eta)}{K}\right) \tag{6.3}$$

Similarly, the modeling algorithm visits each node only once, where the computations at each node involve a number of floating point operations proportional to the square of the state vector at the given node. Thus, we see that the application of the modeling algorithm requires a total of $\mathcal{O}(k^2 N)$ floating point operations; that is, a computational effort per pixel of

$$\mathcal{O}\left(\frac{k^2}{\eta}\right) \tag{6.4}$$

or, if the algorithm is fully parallelized on each scale,

$$\mathcal{O}\left(\frac{k^2 \log(K/\eta)}{K}\right) \tag{6.5}$$

The point is that we can achieve dramatic computational efficiencies as long as the

maximal dimension k of the state model and the amount of overlap (as measured by $1/\eta$) are not too large. As we will illustrate, the overlapping tree procedure will allow us to meet these criteria.

6.2 Formulation of the Problems of Modeling and Estimation with Overlapping Trees

In this section, we identify the precise technical problems to be solved in order to develop our new approach to multiscale modeling and estimation. Because our overlapped-tree approach represents something of a conceptual departure from other approaches, we make the ideas more concrete through the use of a simple example.

6.2.1 Modeling of Random Fields with Overlapped Tree Processes

Let x be a zero-mean random field, written for simplicity as a vector, and having covariance P. In this section we consider the problem of generating sample values of x; more specifically, we consider the problem of generating sample functions of a zero-mean random field with covariance equal to P or close enough to P so as to capture its significant statistical characteristics.

From a computational point of view, this simulation problem poses nontrivial challenges and has been the focus of a considerable amount of research in the signal and image processing communities. One notable case in which computationally efficient techniques do exist is for generation of sample functions of stationary random fields, defined on regularly sampled toroidal lattices, in which case the 2-D FFT can be used to diagonalize the field's covariance matrix. However, for most other types of fields, the generation of sample paths is computationally quite complex. For example, sample paths may be simulated using the following three step procedure:

(i) compute the square root $P^{1/2}$ of the covariance matrix,

(ii) generate a vector w of unit variance uncorrelated random variables,

(iii) compute the sample path as $x = P^{1/2}w$.

Although conceptually straightforward, there is a considerable challenge in computing the matrix square root $P^{1/2}$, requiring in general $\mathcal{O}(K^3)$ calculations for a random field of K points. Similar computational difficulties are encountered with iterative generation methods, such as those for Markov random fields, which can frequently require an exorbitant number of iterations, especially to capture significant large-scale correlations. On the other hand, as discussed in Section 6.1.3, the simulation of a random field having a multiscale model is extremely fast; thus we are led to consider more completely the issues involved in an overlapping-tree approach to simulation. We conveniently decompose the approach into three steps (Figure 6-3 depicts these steps graphically).

In the first step, we must specify the matrix G_x which serves to lift the random field x into another random field x_l via

$$x_l = G_x x. \tag{6.6}$$

This lifted-domain field x_l corresponds to the finest scale of a tree process, and acts as a particular, redundant, overlapped representation of x, having more pixels than the original field. The matrix G_x is not chosen arbitrarily; it has a considerable amount of sparse structure, as we discuss in greater detail in Section 6.3; furthermore, G_x is chosen such that it has a left inverse H_x,

$$H_x G_x = I, \tag{6.7}$$

satisfying certain smoothness properties to be discussed shortly.

In the second step, we combine our knowledge of the covariance P and matrix G_x to build a low-dimensional multiscale model whose finest-scale statistics Σ_l are an approximation to P_l, the statistics of x_l. As a consequence of (6.6), the covariance of

x_l is given by

$$P_l = G_x P G_x^T. \tag{6.8}$$

We denote by ξ_l (see Figure 6-3) the random field living at the finest scale of this multiscale tree; i.e., the covariance of ξ_l is Σ_l.

The last step in generating a sample function of a random field qualitatively similar to x is to apply the operator H_x to ξ_l, thus yielding

$$\xi = H_x \xi_l. \tag{6.9}$$

This random field ξ is guaranteed to be smooth, by our assumed smoothness properties for H_x; also, ξ will have approximately the same statistics as x, since from (6.7),

$$\text{cov}(\xi) = H_x \Sigma_l H_x^T \approx H_x P_l H_x^T = H_x G_x P G_x^T H_x^T = P = \text{cov}(x), \tag{6.10}$$

Thus, the design problem confronting us is that of specifying the operators G_x and H_x and then constructing the multiscale model for ξ_l, so that the following properties hold:

(i) G_x and H_x are sparse and local,

(ii) H_x achieves the desired smoothness,

(iii) the multiscale model is of sufficiently low dimension that simulation can be done efficiently,

(iv) the approximation in (6.10) is sufficiently accurate so as to lead to sample functions capturing the desired statistical characteristics.

The method we use here to construct the multiscale models is the canonical-correlations-based stochastic realization method, described in detail in [48] and summarized in Section 2.5.2. The focus of attention in the remainder of this and the next

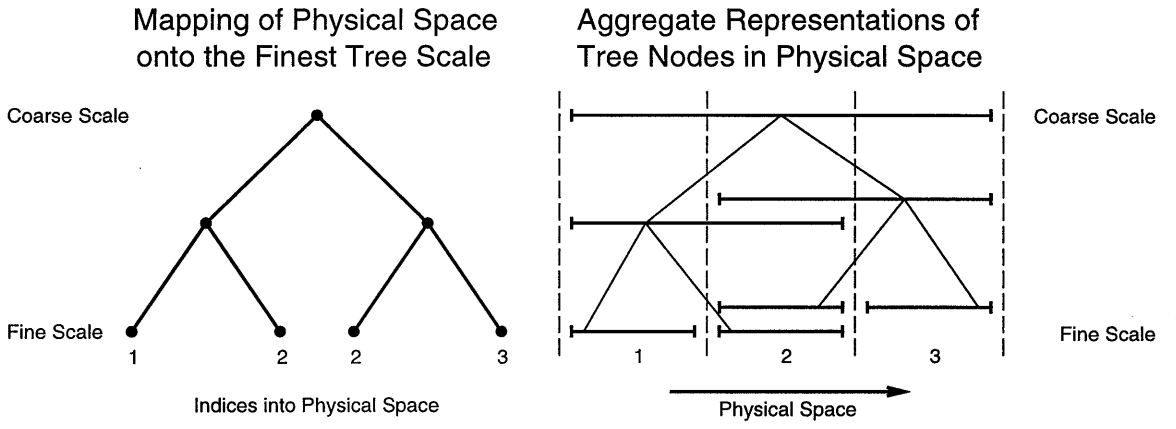


Figure 6-4: Illustration of an overlapping-tree representation of a process of length three showing both the dyadic tree (left) on which the representation is based, and a depiction (right) of the representation of each tree node. The bar $\vdash$ associated with each tree node represents the subset of the finest-scale process aggregated by that node.

section is on the design of G_x and H_x. Following that, computational examples will demonstrate that our approach does indeed achieve objectives (i)-(iv).

To introduce the basic issues involved in specifying G_x and H_x, let us consider a very simple 1-D example of a random process of length 3. Collecting the process values into a vector $x^T = [x_1, x_2, x_3]^T$, suppose that the covariance of x is given by

$$x \equiv \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad E\left[xx^T\right] = P \equiv \begin{bmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{bmatrix} \tag{6.11}$$

Let us index our multiscale model on a dyadic tree that has four finest-scale nodes, thereby providing only a minimal amount of redundancy; Figure 6-4 displays an example of such a tree. On the right side of the figure, we depict the tree with an indication of the subsets of real, physical points (i.e., subsets of $\{1, 2, 3\}$) to which each node corresponds. Thus, the top node corresponds to all three points (i.e., $\{1, 2, 3\}$) and the two nodes at the second level correspond to $\{1, 2\}$ and $\{2, 3\}$ respectively. At the bottom level there is a single node corresponding to process element x_1 and another for x_3, but there are *two* nodes corresponding to x_2: in the lifted domain on the tree, process element x_2 is lifted to have two finest-scale tree nodes. Thus

if we order the four fine-scale nodes from left to right (as shown in the left half of Figure 6-4), we are led to the following choice of G_x:

$$G_x \equiv \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{6.12}$$

which implies that

$$P_l \equiv G_x P G_x^T = \begin{bmatrix} 1 & 0.5 & 0.5 & 0 \\ 0.5 & 1 & 1 & 0.5 \\ 0.5 & 1 & 1 & 0.5 \\ 0 & 0.5 & 0.5 & 1 \end{bmatrix} \tag{6.13}$$

This example illustrates the basic constraints that we place on any *lifting* matrix G_x:

1. It consists entirely of zeros and ones.
2. Each column has at least one nonzero entry.
3. Each row has exactly one nonzero entry.

These conditions ensure the following basic properties:

1. Every position in the original domain corresponds to at least one position in the overlapped domain.
2. Every position in the overlapped domain corresponds to exactly one position in the original domain

Thus, the lifting process is local and is in fact trivial to compute once G_x has been specified. The specification of G_x is typically carried out implicitly in terms of the overlapping structure of the tree: any given overlap structure uniquely specifies a corresponding G_x. We can certainly imagine more general lifting schemes, allowing for example for fine-scale lifted nodes that are associated with more than one real

data point; however, we will find our restricted lifting scheme to be sufficient for our purposes.

Turning to the smoothing operator, we note that the inverse relation (6.7) between H_x and G_x, together with our imposed constraints on the structure of G_x, lead to an important constraint on the structure of H_x. In particular, the value of any element of $\xi = H_x \xi_l$, the projection of the simulated process back into the original domain, is equal to a weighted average of the corresponding elements in ξ_l (that is, each row of H_x must sum to one). For example, with G_x as in (6.12), the possible choices for H_x are of the form

$$H_x = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & a & b & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{6.14}$$

where $a + b = 1$. Here a and b can be thought of as the weights being placed on the value of the two nodes corresponding to physical index 2 in order to specify x_2. For example, an equal weighting $a = b = \frac{1}{2}$ would intuitively lead to the greatest smoothness in the correlation structure of ξ. On the other hand, it is important to emphasize that the averaging implied by (6.14) is *not* at all the same as spatial averaging, since we average only over those tree points corresponding to the *same* point in real space. A general method for choosing H_x will be presented in Section 6.3.

6.2.2 Estimation of Random Fields with Overlapped Tree Processes

We now turn to the problem depicted in the bottom half of Figure 6-3. The objective is to exploit the efficiency of the multiscale estimation algorithm to perform optimal or near-optimal estimation of a random field x, while avoiding blocky artifacts.

Suppose that we have noisy measurements of x

$$y = Cx + v \qquad v \sim \mathcal{N}(0, R) \tag{6.15}$$

such that two conditions hold:

(i) each component of y represents a measurement of an individual pixel, so that each row of C has only one nonzero entry,

(ii) the noise on these measurements are uncorrelated with each other, so that the covariance R of v is diagonal.

From Section 2.1 we know that the optimal estimate is given by $\hat{x} = Ly$, where L is given by (2.9), assuming that x has prior covariance P. However for a K-pixel field the calculation of L is generally $\mathcal{O}(K^3)$ and the calculation of the product Ly is $\mathcal{O}(KM)$ where M is the number of measurements. Virtually the only case in which this computational load can be reduced to a practical level is when the field x is stationary and we have dense, regularly sampled measurements of identical quality (implying that C and R are both multiples of the identity) such that FFT methods can be applied (reducing the computational load to $\mathcal{O}(KlogK)$). However in other cases the $\mathcal{O}(K^3)$ computational load cannot be reduced in this manner, compelling us to turn to iterative methods for the computation of $\hat{x}$. Not only can these iterative methods be slow, but they also do not yield error covariance information.

We are thus motivated to consider the estimation approach illustrated in the bottom half of Figure 6-3. To develop this approach we will extend the results from our approach to modeling. In particular, we will need the lifting and projection operators G_x and H_x for our random field, as well as a multiscale model for ξ_l, such that $H_x\xi_l$ is an adequate approximation of the field x; the issues related to determining these were discussed in the preceding section. In addition, specific to the estimation problem, we need to define a lifting operator G_y for the measurements:

$$y_l = G_y y \tag{6.16}$$

and a lifted measurement model

$$y_l = C_l x_l + v_l \tag{6.17}$$

which involves the specification of C_l and of the covariance of v_l. The definition of this measurement model allows us to carry out estimation as a three-step procedure:

1. Compute the lifted measurements as per (6.16).

2. Apply the multiscale estimation algorithm to estimate x_l based on y_l (which satisfies (6.17)).

3. Apply H_x to the resulting estimate, thereby yielding a near-optimal estimate of x based on y.

For step (2.) to be feasible, the components of y_l must represent observations of individual fine-scale tree nodes, where the observation noises (the elements of v_l) are uncorrelated. In other words, each row of C_l must have only one non-zero entry and the measurement covariance R_l of v_l must be diagonal. To clearly see the requirements for the success of step (3.), let us write the multiscale estimation solution in input-output form:

$$\hat{x}_l = L_l y_l = \Sigma_l C_l^T \left(C_l \Sigma_l C_l^T + R_l \right)^{-1} y_l \tag{6.18}$$

Combining (6.18) with (6.16), we see that our step (3.) objective of satisfying $\hat{x} \approx H_x \hat{x}_l$ is equivalent to satisfying

$$PC^T \left(CPC^T + R \right)^{-1} = L \approx H_x L_l G_y = H_x \Sigma_l C_l^T \left(C_l \Sigma_l C_l^T + R_l \right)^{-1} G_y \tag{6.19}$$

Assuming that G_x, H_x, and the multiscale model (which specifies Σ_l) have been chosen, the remaining quantities to be specified are G_y, C_l, and R_l. Given the value of G_x, which implicitly associates each pixel with a set of fine-scale nodes, the most natural choice for C_l is one that yields lifted measurements at each fine-scale tree node associated with any pixel having a measurement. For example, let us consider again the three-point process and dyadic tree lifting illustrated in Figure 6-4; let us

suppose we have measurements of x_1 and x_2, namely

$$y \equiv \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \quad C \equiv \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \end{bmatrix} \quad R \equiv \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix} \tag{6.20}$$

Then, in our lifted domain we should have *three* measurements, one measurement corresponding to the single node associated with x_1, and measurements corresponding to each of the two nodes associated with x_2. That is,

$$C_l \equiv \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{6.21}$$

(for reasons of *notational* convenience in defining G_y, a dummy measurement is added here to the last row to make C_l diagonal;[3] the reader is free to ignore this last row).

The question at this point is how to create three measurement values on the tree when only two real measurements are available. The answer here, and in our general procedure, is that we simple *copy* the measurement value at any pixel to all fine-scale nodes associated with that pixel. In our example,

$$G_y \equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad y_l = G_y y \equiv \begin{bmatrix} y_1 \\ y_2 \\ y_2 \\ 0 \end{bmatrix} \tag{6.22}$$

At first glance, this procedure appears to create a significant problem: for the multi-scale estimation algorithm to work we require that the measurements at distinct nodes have uncorrelated errors. With y_l and C_l defined as in (6.22) and (6.21) this uncorrelatedness certainly does not hold, since two of the "measurements" are identical. This problem turns out to not be serious; we simply *model* these two measurements

[3] In Section 6.3 and in Appendix E, we shall find that having a measurement at *every* finest-scale node will make the precise description of operator G_y enormously simpler. Consequently the measurement matrix C_l is padded with zero-rows (i.e., dummy measurements) to make it diagonal. It must be stressed that this is purely a notational matter and has no consequences on the theory or practical implementation of overlapping trees.

as being measurements of the states at their corresponding nodes with uncorrelated measurement errors. This approach appears to have created another difficulty: by modeling y_l in this way we appear to be saying that we have more information than we actually do; in particular in our example we now have two independent measurements of the nodes corresponding to x_2. To compensate for this we need to ensure that the total information represented by these two measurements is the same as that represented by the single real measurement; this is accomplished simply by doubling the corresponding measurement noise variances in our model for each of the replicated measurements, i.e., we define

$$R_l \equiv \begin{bmatrix} 1 \cdot 3 & 0 & 0 & 0 \\ 0 & 2 \cdot 4 & 0 & 0 \\ 0 & 0 & 2 \cdot 4 & 0 \\ 0 & 0 & 0 & * \end{bmatrix} \tag{6.23}$$

(where $*$ corresponds to the dummy measurement of (6.21); the value of $*$ is irrelevant).

In the next section we show that if $\Sigma_l = P_l$, i.e., if no approximation is made in the canonical correlations-based realization of x_l on the tree, then one can indeed choose G_x, H_x, G_y, C_l, and R_l so that the resulting estimate $\hat{x} = H_x \hat{x}_l$ is *exactly* equal to the optimal estimate of x based on y (i.e., so that (6.19) is an equality) and the same is true of the resulting error covariances. The example we have given illustrates one very simple choice for these matrices, and as we will see there is actually considerable flexibility in their choice. Note also that what this result implies is that *any* suboptimality in using the procedure we describe can be traced *completely* to the approximation in building a low-dimensional approximate model for x_l, allowing explicit control of the complexity-accuracy tradeoff. Of course that tradeoff may be different for different choices of G_x, H_x etc; e.g., for different amounts of smoothness in the resulting fields and estimates; we will provide some illustrations of these tradeoffs in Section 6.4.

6.2.3 Optimal Estimation Through Lifting and Projection

In this section we present a set of conditions on G_x, G_y, H_x, C_l, and R_l that will guarantee that if an exact covariance model is used in the lifted domain then we can indeed compute the optimal estimate using the procedure depicted in Figure 6-3. Specifically, let x be a random field with covariance P and let $y = Cx + v$ be noisy measurements of individual pixels of this field. That is, C is a weighted selection matrix, i.e., each row of C has exactly one nonzero entry, and each column has at most one nonzero entry[4] and R, the covariance of v, is diagonal.

Motivated by the discussion in the preceding subsections we first begin by specifying the lifting matrix G_x. As in the simple example we restrict ourselves to choices of G_x that meet three criteria:

1. G_x consists entirely of zeros and ones;

2. each column of G_x has at least one nonzero entry;

3. each row of G_x has exactly one nonzero entry.

We then must choose H_x such that $H_x G_x = I$; while this requirement severely constrains H_x, it does leave some remaining degrees of freedom.[5] In particular, as we have seen, the choice of G_x is directly related to the overlapping structure that we have chosen, which in turn specifies which fine-scale tree nodes correspond to which real pixels. H_x then performs a weighted averaging among each set of tree nodes that correspond to each individual pixel, where there is flexibility in the choice of these weights. Thus there is considerable freedom in the choice of G_x and H_x. Furthermore the resulting matrices are quite sparse. On the other hand, for 2-D problems of practical interest these matrices will be extremely large, and thus any structure that can be imposed or discerned about the sparsity in G_x and H_x will be of considerable

[4]These conditions are equivalent to saying that each measurement is of a distinct pixel and any pixel has at most one measurement associated with it. The latter assumption is for simplicity only – i.e., if there are multiple measurements of a single pixel, since R is diagonal we can replace these by a single measurement obtained as the weighted average of the redundant measurements where each measurement is weighted proportionally.

[5]In fact, the number of degrees of freedom equals $N - K$.

benefit. Section 6.3 will describe how these matrices can be specified in an implicit manner that achieves a considerable reduction in storage requirements.

Once G_x and H_x have been chosen, we next turn our attention to the choice of G_y and C_l. In particular our real measurements are $y = Cx + v$ while our lifted measurements are computed as $y_l = G_y y$ and modeled as $y_l = C_l x_l + v_l$, where $x_l = G_x x$. Thus we have two expressions for how the real random field x affects the lifted measurements y_l, namely $C_l G_x x$ and $G_y C x$. A logical requirement on C_l and G_y then is to require these two expressions to be equal for any x; i.e.,

$$C_l G_x = G_y C \tag{6.24}$$

Thus, since C has full row rank, once the value C_l is determined the value of G_y is automatically determined.

We now construct an appropriate matrix for C_l exactly as we did for our example. Specifically, we assume that for each real pixel measurement we have an analogous measurement for *each* of the tree nodes corresponding to that real pixel. Furthermore, for reasons of convenience, we will pad C_l with zero rows (dummy measurements) to make C_l diagonal; that is, y_l has the same dimension as x_l. Thus if the jth component of y is $y_j = \alpha_j x_i + \text{noise}$ (where x_i is a component of x) then y_l will have measurements of the form

$$(y_l)_n = \alpha_j (x_l)_n + \text{noise} \tag{6.25}$$

for each n such that finest-scale node $(x_l)_n$ corresponds to the real pixel x_i. Since C is a weighted selection matrix, so is C_l. Since C has full row rank, from (6.24) we see that

$$G_y = C_l G_x C^T (C C^T)^{-1}. \tag{6.26}$$

While this expression for G_y is correct, it obscures the simple, sparse structure of the matrix. However, once we note that $(CC^T)^{-1}$ is a diagonal matrix and that the diag-

onal weights in C_l are the same as those in C, it follows fairly easily that G_y is a lifting matrix, just as G_x is (in fact, because dummy measurements were added to C_l to make it diagonal, G_y is just a particular subset of columns of G_x). This structural property of G_y is consistent with our simple example: we define lifted-domain observations to exist for those lifted-domain nodes where corresponding original-domain observations exist, and then we assign overlapped-domain values to these measurements by simply replicating the appropriate original-domain measurement values.

The construction of R_l is facilitated by defining $g(j)$ to be the number of ones in the jth column of G_y; in other words, $g(j)$ is the number of times that the jth original-domain measurement is replicated. We then define R_l, the covariance of the measurement noise vector v_l, to be a diagonal matrix with the ith diagonal entry $R_l(i,i)$ given by

$$R_l(i,i) \;=\; g(j)R(j,j), \tag{6.27}$$

where j is the unique index for which $G_y(i,j) = 1$, (i.e., finest-scale node i corresponds to pixel j). This choice for R_l, which is exactly what was done in our simple example, provides the observation covariance amplification required in the lifted domain to counter the apparent increase in information caused by the replication of the measurements by G_y. We note that this choice for R_l leads to the following important identity:

$$G_y^T R_l^{-1} G_y \;=\; R^{-1} \tag{6.28}$$

We have the following Proposition.

Proposition: Let x be a random field with covariance P and let $y = Cx + v$ be a set of measurements with C a weighted selection matrix and R, the covariance of v, diagonal. Suppose we then choose G_x, H_x, G_y, C_l and R_l as just described. Furthermore, define

$$L \;=\; PC^T\left(CPC^T + R\right)^{-1} \tag{6.29}$$

$$L_l = P_l C_l^T \left(C_l P_l C_l^T + R_l \right)^{-1} \tag{6.30}$$

where P_l is defined in (6.8). Then the optimal estimates, computed as

$$\hat{x} = Ly \tag{6.31}$$

are the *same* as those estimates computed as the projection of the optimal lifted-domain estimates; that is

$$\hat{x} = H_x \hat{x}_l = H_x L_l G_y y \tag{6.32}$$

Moreover, if $\tilde{P}$ denotes the estimation error covariance in estimating x based on y, and $\tilde{P}_l$ the estimation error covariance in estimating x_l based on y_l, then

$$\tilde{P} = H_x \tilde{P}_l H_x \tag{6.33}$$

Two proofs for this proposition may be found in Appendix D: the first proof succinctly demonstrates the correctness of the above proposition, the other proof is much longer but tries to provide a deeper understanding of the structures possessed by G_x, G_y, H_x, and R_l.

6.3 Specification of the Overlapping Framework

In this section we describe an implicit, compact, and efficient method for specifying the operators G_x, G_y and H_x directly from the overlapping structure that is chosen in order to achieve our desired objective of producing random fields and estimates with some desired level of smoothness. The intent of this section is to provide an accessible, reasonably intuitive, notion of the means of specifying these operators; the discussion in Appendix E parallels the subject matter of this section, but in greater generality and detail.

As we have already discussed and will see in our examples, as η decreases (i.e.,

as the amount of overlap increases) the resulting simulated or estimated fields (i.e., those fields in the original domain after projection by H_x) become smoother, but the computational complexity of the resulting simulation and estimation algorithms increases. For simplicity in exposition we focus on the most important case, namely the one in which the tree and overlap structure are regular. Specifically, we focus here on regular overlapping q-adic trees representing 1-D random fields, where each node above the finest scale possesses q descendants. The overlap structure is presumed to be spatially stationary; that is, for any two nodes s_1 and s_2 on the same scale of the tree, the manner in which their descendants overlap is the same. Our focus on the one-dimensional case is strictly for reasons of clarity; Appendix E extends the results of this section to regular trees in d dimensional space, where each node above the finest scale possesses

$$q = q_1 \cdot q_2 \cdot \ldots \cdot q_d \tag{6.34}$$

descendants; that is, the descendants of any node form a hyper-grid with q_i descendants along the edge of the grid parallel to dimension i.

While it is certainly possible to develop the lifting and projection procedures for non-regular trees or non-regular overlapping schemes, we have never found this necessary in any applications which we have considered (including the nonstationary example in Section 6.4). Furthermore, with the regularity assumption comes a significant simplification in that the complete specification of G_x and H_x can be accomplished by determining a total of $(M-1)$ parameters as follows (M is the number of scales in the tree). Recall that each node on the multiscale tree is associated with a connected interval of points in the original domain. On each scale m, the size of each of these connected intervals (one interval per tree node) is the same; we let this size be denoted by w_m (see Figure 6-5). In addition, the intervals associated with neighboring nodes (i.e., successive children of a common parent) may be overlapping; the amount of overlap between neighboring multiscale nodes on scale m is denoted

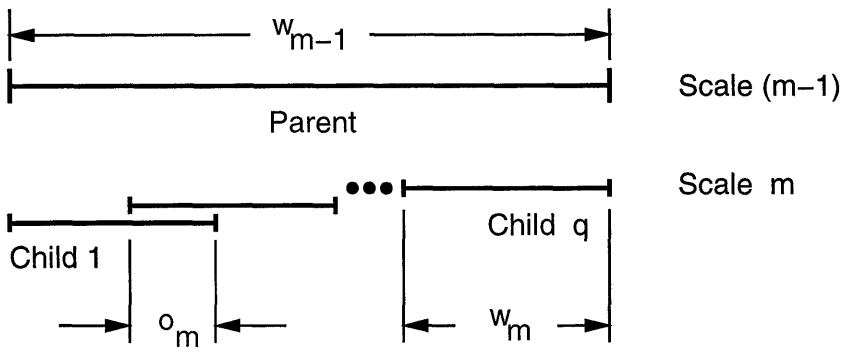


Figure 6-5: Basic overlapping-tree notation: o_m represents the degree of overlap between the regions represented by sibling multiscale nodes on scale m; w_m represents the width of the region represented by each node on scale m.

by $o_m \geq 0$. Further, we insist that neighboring nodes do not completely overlap:

$$0 \leq o_m < w_m \tag{6.35}$$

Finally, from Figure 6-5, the $\{w_m\}$ and $\{o_m\}$ values are related by the following recursion:

$$w_{m-1} = qw_m - (q-1)o_m \tag{6.36}$$

The $M-1$ parameters $\mathcal{O} = (o_1, \ldots, o_{M-1})$ uniquely characterize the overlapping nature of the tree.

Now suppose that we wish to represent a one-dimensional random field having a length of K pixels on an overlapping tree having M scales, using a given overlapping structure characterized by $\mathcal{O}$. For such a tree to represent our random field, two constraints (in addition to (6.35),(6.36)) must be satisfied by the overlap parameters:

1. Each multiscale node on the finest level of the tree must correspond to a single pixel:

$$w_{M-1} = 1 \tag{6.37}$$

2. The root node of the multiscale tree must be associated with the entire random

field:

$$w_0 = K \tag{6.38}$$

From these constraints it follows that M and the $\mathcal{O}$ may not be chosen arbitrarily (upper and lower limits for M are specified in Appendix E).

Even subject to the constraints (6.35)–(6.38), each process of length K pixels still admits a large class of possible overlapping tree structures. We have frequently chosen the following heuristic: the fractional overlap should be close to a constant function of scale; i.e.,

$$\frac{o_m}{w_m} \simeq \psi \tag{6.39}$$

Appendix E outlines a recursive process by which to select $\mathcal{O}$ to simultaneously satisfy this heuristic and (6.35)–(6.38). Our interest is by no means limited to overlapping structures satisfying (6.39) however; in Chapter 7 we will make use of other possibilities.

Now suppose that M,$\{o_m\}$ have been specified (i.e., the characterization of the overlapping tree is complete); the unique value of G_x determined by the choice of M,$\{o_m\}$ may be found as follows. Consider a node s on the finest scale of the tree; we can describe s as a descendant of the root node in the following manner:[6]

$$s = 0\alpha_{j_1}\alpha_{j_2}\ldots\alpha_{j_{M-1}} \qquad j_m \in \{1,\ldots,q\} \tag{6.40}$$

where we orient the tree such that $t\alpha_1$ is the leftmost child of node t and $t\alpha_q$ is the rightmost child. The effect of (6.40) is to code node s as a $(M-1)$-tuple $(j_1,\ldots,j_{M-1})$.

Knowing the ancestry of s, the position of s on the multiscale tree is readily

[6]Recall that $t\alpha_i$ is the ith descendant of multiscale tree node t.

calculated. Let

$$\bar{p} = \sum_{m=1}^{M-1} (j_m - 1) q^{M-m-1} \tag{6.41}$$

then s is the $\bar{p}$th node on the finest scale of the tree, measured from the left end of the tree. Furthermore, knowing $M, \mathcal{O}$, we can also infer the point in physical space with which s is associated. Define

$$p = \sum_{m=1}^{M-1} (j_m - 1)(w_m - o_m), \tag{6.42}$$

then s is associated with the pth element of the one-dimensional process which we are representing on the tree. It follows then that

$$G_x(\bar{p}, p) = 1 \text{ and } G_x(\bar{p}, \rho) = 0, \ \forall \rho \neq p \tag{6.43}$$

Given s (and its associated coding (6.40)) it is straightforward to calculate $\bar{p}, p$; by repeating this calculation for all s on the finest scale (that is, for all $j_i \in \{1, 2\}, 1 \leq i < M$), G_x is constructed implicitly.

The construction of G_y follows immediately from that of G_x: G_y just consists of the subset of columns of G_x at which we have measurements. Recall that G_y copies measurements from the original domain into the lifted domain. Consequently, if the ith measurement y_i measures the jth pixel x_j of the one-dimensional process, i.e.,

$$y_i = \beta x_j + v_i \tag{6.44}$$

then the ith column of G_y is set equal to the jth column of G_x.

The construction of H_x, on the other hand, is not fully determined by M,$\mathcal{O}$; in fact, H_x must only satisfy $H_x G_x = I$. The matrix H_x determines the interpolation of elements from the lifted domain back into the original domain. We desire that this interpolation be purely an *ensemble* average; that is, H_x must perform *no* spatial averaging of any kind — each element of $H_x \hat{x}_l$ is based only on a linear combination

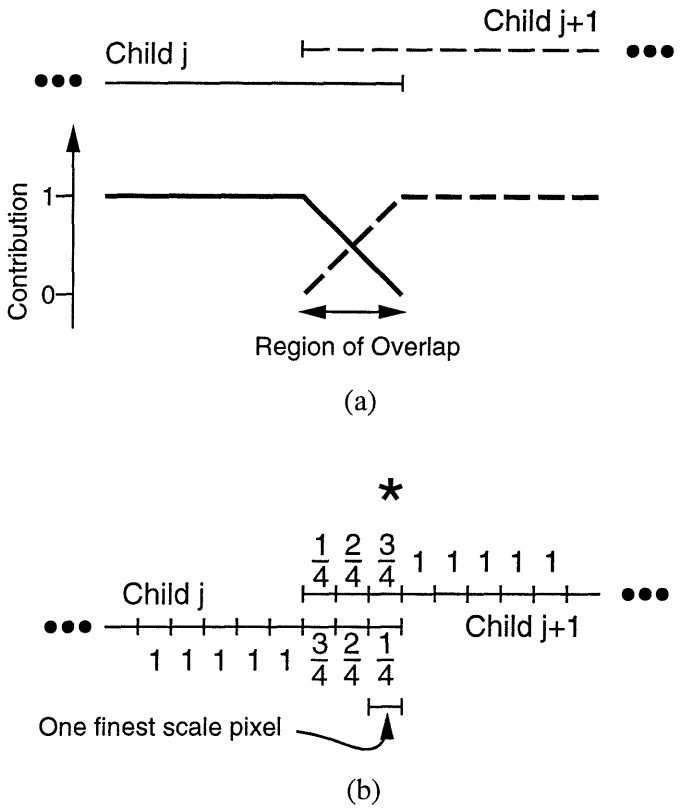


Figure 6-6: Two overlapping nodes: the set of relative contributions to each finest-scale pixel must sum to one. The contributions are tapered linearly over the region of overlap. Figure (a) shows this tapering pictorially; Figure (b) provides a specific example for two nodes which overlap by three pixels.

of the corresponding elements in $\hat{x}_l$. For H_x to eschew any spatial mixing, it follows that

$$G_x(i,j) = 0 \longrightarrow H_x(j,i) = 0 \tag{6.45}$$

i.e., the distribution of non-zero elements in H_x is the same as in G_x^T. Note that the choice $H_x = G_x^+$, the Moore-Penrose inverse of G_x[2], satisfies (6.45) and also $H_x G_x = I$. A smoother selection of H_x, outlined below, is also possible.

Consider two nodes on some scale m, such as the two child nodes $j, (j+1)$ shown in Figure 6-6(b); now consider some pixel that lies within the overlapping regions of these two nodes (e.g., the pixel marked $\star$ in the figure). We need to specify the contributions of child j and child $(j+1)$ (and their descendants) in determining the

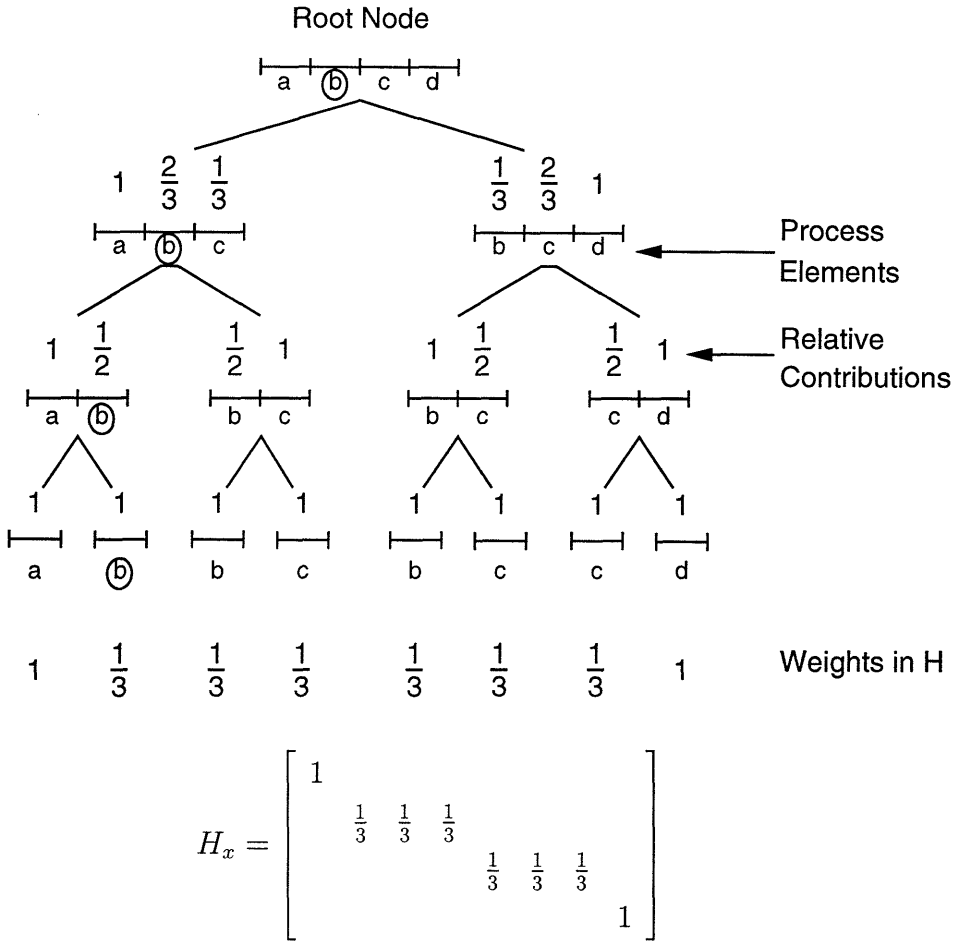
value of pixel $\star$; for example, in Figure 6-6(b), child $(j+1)$ (and its descendants) will have a contribution three times that of child j. In order to maintain a total contribution of unity at each pixel, we will normalize the contributions at each pixel to sum to one; these normalized values will be referred to as *relative* contributions.

We propose to achieve smoothness in H_x as follows:

- The relative contributions of a node to its associated pixels should be a smooth function of the pixel's spatial position.
- The relative contributions of a node should taper towards zero as one approaches an overlapped end of the interval associated with the node (e.g., pixel $\star$ represents the end of the interval associated with child j in Figure 6-6(b)),

Figure 6-6(a) sketches one possible choice of relative contributions consistent with the points just listed: we propose to taper the relative contributions linearly across the region of overlap.

The previous paragraph outlined a procedure for determining the relative contributions of two overlapping nodes. Suppose this procedure has been applied to all neighboring nodes on all scales; how can we determine H_x? Consider a node s on the finest scale, and define $\bar{p}, p$ as in (6.41),(6.42). $H_x(p, \bar{p})$ controls the participation of the value at node s in determining $\hat{x}_p$ — the estimate of the pth component of the original domain vector x. The participation of node s on the finest scale is determined as the product of all relative contributions associated with position p on all ancestors of s. This construction is illustrated in Figure 6-7; the figure illustrates an overlapping tree representation of a process having four points: (a, b, c, d). Process points b and c have a triply-redundant representation on the finest scale of the tree; points a and d each correspond to only a single finest scale tree node. Now consider finest scale node $s = ⓑ$ (second from the left end of the tree). The participation of s in determining the value of b is given by the product of the relative contributions to b of all ancestors of s — i.e., the numerical values above each ⓑ in Figure 6-7. Thus the participation of s is equal to $1 \cdot \frac{1}{3} \cdot \frac{2}{2} = \frac{1}{3}$; so the weight in H_x associated with s is $\frac{1}{3}$. The weights in H_x corresponding to each of the finest-scale nodes are shown in



$$H_x = \begin{bmatrix} 1 & & & & & & & \\ & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & & & & \\ & & & & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \\ & & & & & & & 1 \end{bmatrix}$$

Figure 6-7: An example of the construction of H_x. A four-level tree is used to represent a process having four points (a, b, c, d). The process points associated with a multiscale node are indicated below the node. The relative contributions of each node to its associated process points are indicated above each node. Products of these relative contributions determine the elements of H_x.

Figure 6-7.

For all but the smallest estimation problems, a dense representation of the G_x, G_y, and H_x matrices is completely impractical. The observation that each row of G_x, G_y and each column of H_x contains only one non-zero entry suggests that a sparse representation based on storing only these non-zero entries might be adequately compact. However for large multidimensional problems even this sparse representation may be very large (indeed, the combined number of of non-zero entries in G_x, G_y and H_x may exceed the number of values in the entire multiscale tree). As has been discussed in this section (and demonstrated more rigorously in Appendix E) the overlapping

k =	-2	-1	0	1	2
2		-0.0085	0.0139	-0.0058	
1	-0.0008	-0.1164	0.2498	-0.1405	0.0091
l = 0	-0.0517	0.5508		0.5508	-0.0517
-1	0.0091	-0.1405	0.2498	-0.1164	-0.0008
-2		-0.0058	0.0139	-0.0085	

Table 6.1: Coefficients $\{h_{k,l}\}$ of the Markov random field "wood" model[53].

structure parameterization $\{M, \{o_m\}\}$, in which there are only M parameters, forms a sparse and implicit representation of G_x, G_y and H_x. We have found the construction of G_x, G_y and H_x from the overlap parameters $\{o_m\}$ to be so rapid that we have exclusively used this latter representation in our software.

6.4 Experimental Results

The overlapping tree framework has given us a powerful new tool with which to perform estimation. In this section we will demonstrate three applications of this tool to problems of texture estimation.

An anisotropic Markov random field will form the basis of all three of our estimation examples. The Markov random field model takes the form

$$z(x,y) = \sum_{k,l \in \mathcal{D}} h_{k,l} z(x-k, y-l) + v(x,y) \tag{6.46}$$

where $v()$ is a Gaussian noise process having the following correlation structure:

$$E\left[v(x,y)v(x+k,y+l)\right] = \begin{cases} \sigma^2 & k=l=0 \\ -\sigma^2 h_{k,l} & (k,l) \in \mathcal{D} \\ 0 & (k,l) \notin \mathcal{D} \end{cases} \tag{6.47}$$

$$E\left[v(x,y)z(x+k,y+l)\right] = \begin{cases} \sigma^2 & k=l=0 \\ 0 & \text{Otherwise} \end{cases} \tag{6.48}$$

where we will use $\sigma^2 = 1$ throughout this section. We will assume that the Markov random field lies on a toroidal lattice; consequently the correlation structure becomes

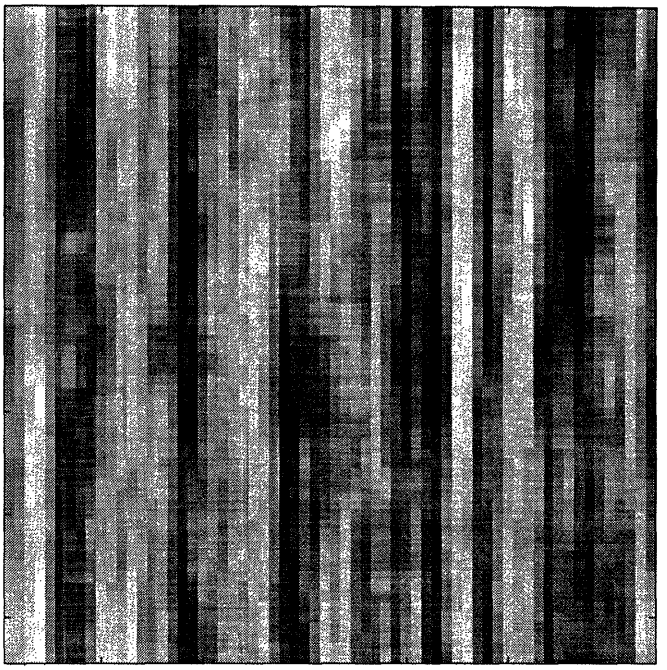


Figure 6-8: Original "wood" texture, 64×64 samples, simulated using Markov Random Field techniques.

circulant, such that efficient FFT techniques may be applied if the field is densely and uniformly sampled. The specific choice of coefficients $\{h_{k,l}\}$ to be used in our examples are those of the "wood" texture[53], tabulated in Table 6.1. The texture shown in Figure 6-8 shows a 64 by 64 pixel sample path, generated using FFT techniques, based on the "wood" texture coefficients. The image possesses an obvious grain – that is, a much stronger correlation in the vertical direction than in the horizontal. The long correlation length in the vertical direction is of particular interest: it is such correlations which non-overlapped multiscale trees find difficult to preserve, even using relatively high order models[61].

Based on the overlapping-tree construction (i.e., the determination of $\mathcal{O}, G_x, G_y, H_x$) outlined in the previous section, there are three basic parameters to be specified in developing an overlapping tree model:

1. The order q of the tree structure. For the two dimensional problem considered here, $q = 4 = 2^2$ is the most natural.

2. The number of scales M in the overlapped tree

3. The order k of the multiscale model

Other parameters, for example the parameterization $\mathcal{O}$ of the overlapping tree, follow from M and q per our discussion in the previous section.

In each of the examples of this section we will be using a canonical correlations technique (from Section 2.5.2) to develop multiscale models of any specified order. The development of a multiscale model proceeds as follows:

1. The prior model of the estimation problem (here a Markov random field model, as in (6.46)) determines the correlation structure P of the field to be estimated.

2. The overlapping tree structure determines the projection operator G_x; together G_x and P specify the correlation structure of the finest-scale nodes on the overlapped multiscale tree.

3. Using canonical correlations, determine the q sets of linear functionals that must be kept at each node s in order to conditionally decorrelate the q children of s. In each case, keep only the k most significant components.

4. Express the linear functionals at each node as a deterministic function of the the linear functionals kept at its parent plus some unknown component. This decomposition precisely describes the $A(s)$ and $B(s)$ matrices of the multiscale model.

This procedure is described in greater detail in [48].

6.4.1 Densely Sampled Field, Homogeneous Model

The original texture shown in Figure 6-8 was corrupted to 0dB SNR by white Gaussian noise, and estimated in three different ways:

1. Using an optimal FFT technique (Figure 6-9)

2. Using a non-overlapped multiscale tree (i.e., a tree having seven scales) with a multiscale model of order $k = 40$ (Figure 6-10)

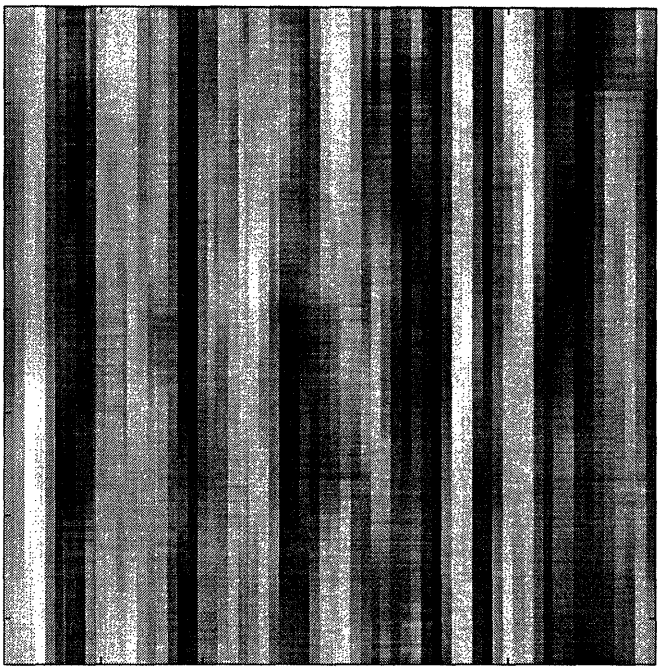


Figure 6-9: Estimated texture using optimal FFT techniques, based on noisy measurements of Figure 6-8.

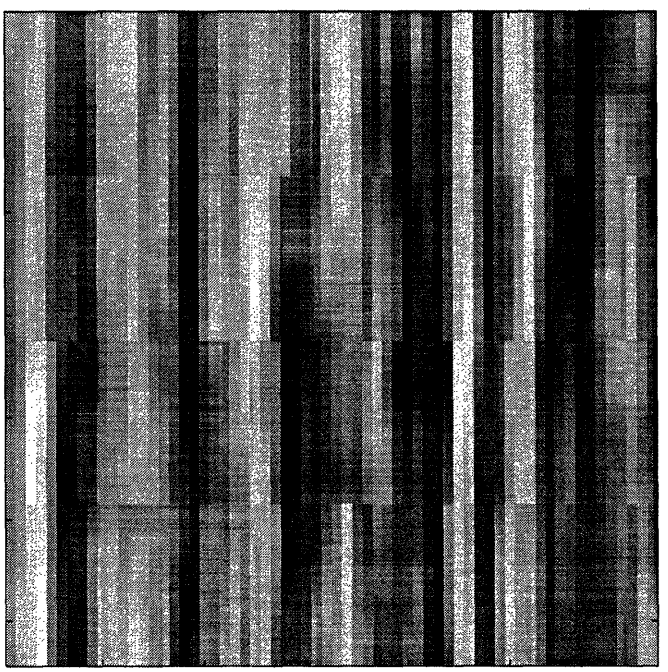


Figure 6-10: Estimated texture using a multiscale tree model, but without using an overlapping tree. Note the artifacts across the boundaries of the image quadrants.

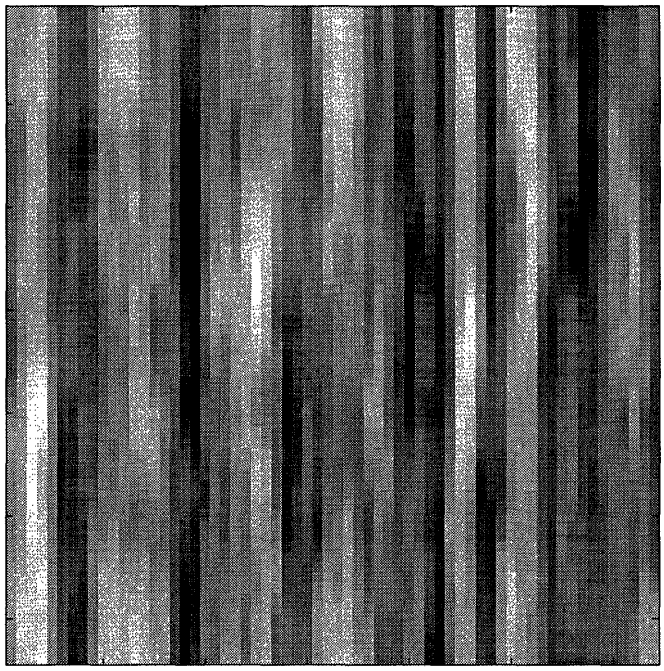


Figure 6-11: Estimated texture using a multiscale tree model applied to an overlapped tree. The computational burden of this estimator is the same as that in Figure 6-10.

3. Using an overlapped multiscale tree having nine scales, and a multiscale model of order $k = 16$ (Figure 6-11)

The model orders of the two multiscale techniques were chosen such that the total computational burden is the same for both; that is, Figures 6-10 and 6-11 each represent an equivalent computational effort. The RMS error of each of the three estimates is compared in Table 6.2. Since the FFT method implements the optimal least-squares estimator, its mean square error (MSE) is the lowest of the three estimators; however given that the measured process (i.e., the noisy version of Figure 6-8) possessed a MSE of 1.0, the reduction in MSE offered by the tree-based estimators is actually a large fraction ($> 97\%$) of the optimal reduction.

An interesting comparison is provided by Figures 6-10 and 6-11: the choice between a high order model ($k = 40$) versus a lower order model ($k = 16$) coupled with an overlapped tree. Specifically, we are interested in the presence of artifacts, particularly anomalous breaks in the vertically-correlated bands of the Markov random field across multiscale tree boundaries. Despite its relatively high model order, the recon-

Estimator	Figure #	MSE
Noisy Field:		1.0000
FFT (Optimum):	6-9	0.1130
MS, No Overlap:	6-10	0.1253
MS, Overlapped:	6-11	0.1311

Table 6.2: Performance comparison of the random field estimators of Figures 6-9, 6-10, 6-11.

struction computed by the non-overlapped tree in Figure 6-10 shows obvious breaks in the vertical bands, whereas these breaks are eliminated by using the low-order overlapping tree model estimates in Figure 6-11. It must also be emphasized that the smoothing operation of H_x in the overlapping tree has not blurred the estimates in Figure 6-11: H_x does not perform any spatial smoothing and is capable of preserving discontinuities and decorrelations.

It is also interesting to compare the overlapping tree estimates (Figure 6-11) with the estimates produced using FFT methods (Figure 6-9). Although the FFT method is optimal in a MSE sense, a lower MSE value does not necessarily imply a better reconstruction from an aesthetic point of view (indeed, this point is made admirably clear by Figures 6-10 and 6-11). In the eyes of the thesis author the FFT reconstruction seems excessively smooth compared to the original process in Figure 6-8, such that the overlapping reconstruction in fact appears to be closer in nature to the original wood texture.

Although the FFT technique is both efficient and optimal, it suffers from a limited applicability to special circumstances; in particular, each of the following cases preclude the use of the FFT, but may be solved using our multiscale method:

- irregularly sampled measurements,
- spatially varying measurement noise,
- spatially varying prior model.

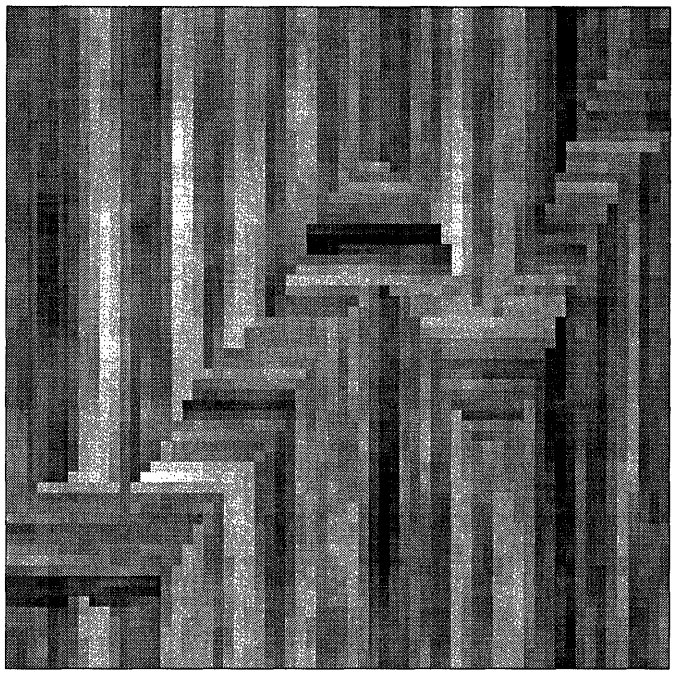


Figure 6-12: This figure shows a sample path of an inhomogeneous Markov random field, where each pixel belongs to a horizontally or vertically correlated texture.

For those very special estimation problems such as the one of this subsection, the FFT approach may be the method of choice. The next two subsections will consider more general estimation problems for which the FFT approach cannot be used.

6.4.2 Densely Sampled Field, Heterogeneous Model

In this section we consider an estimation problem for which FFT techniques are inapplicable: the computation of estimates for a random field having a nonstationary prior model. Figure 6-12 shows a sample path of the nonstationary model. The 64x64 pixels of the process were divided into groups g_1 and g_2: g_1 contains the pixels in the upper left and lower right of the image, and g_2 contains the pixels in the diagonal band running through the center of the image. The prior model for the pixels in g_1 is the "wood" MRF model of Table 6.1; the prior model for the pixels in g_2 uses the the same coefficients in Table 6.1, but with the whole table rotated by 90 degrees. The cross correlation between groups g_1 and g_2 is zero.

The choice of such a nonstationary prior model, as opposed to the simple prior

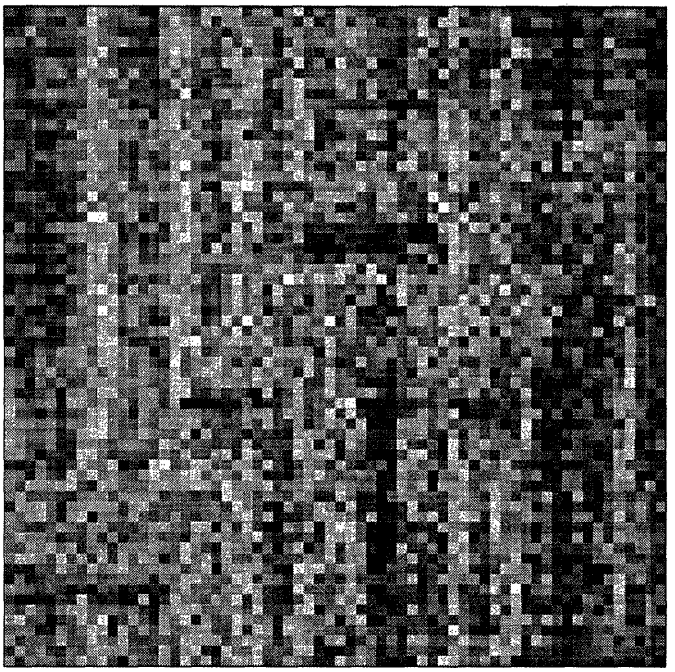


Figure 6-13: Observations of Figure 6-12 in the presence of 0dB white, Gaussian, noise.

model of the previous example, just implies a change in the prior statistics on the finest scale of the multiscale tree. Otherwise there is no essential difference, and the multiscale model development and estimation procedure proceed unaffected.

Figure 6-13 shows a noisy version of the original sample path, corrupted by white Gaussian noise to 0dB; Figure 6-14 shows the corresponding multiscale reconstruction based on an overlapping multiscale model of order $k = 32$. Two observations should be made:

- As mentioned in the previous example, the smoothing operation H_x of the overlapping framework has not at all blurred the edge between the two prior models — the edge stands out distinctly.

- Essentially no artifacts are visible along the correlated bands in either orientation.

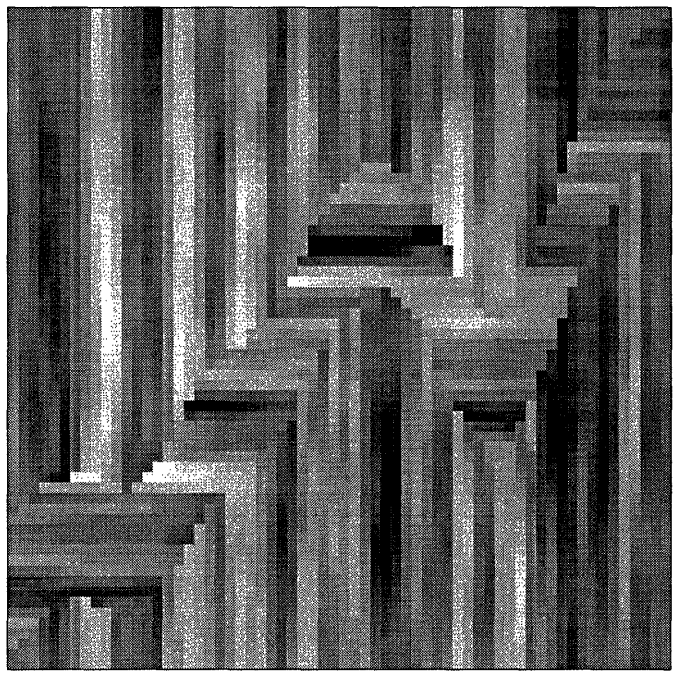


Figure 6-14: The estimated texture using an inhomogeneous overlapped multiscale model, based on the measurements of Figure 6-13 and given the correct prior texture model at each pixel.

6.4.3 Locally Sampled Field, Homogeneous Model

In this last example we will consider one further estimation problem in which FFT techniques would be inapplicable: the computation of a set of estimates given a stationary prior model, but with measurements available at only a small subset of the pixels. Figure 6-15 shows a sample path of the "wood" texture, based on the MRF coefficients of Table 6.1; the small ellipse in the figure indicates the subset of pixels which will be measured. Figure 6-16 shows the measurements of Figure 6-15: no noise was added to the measurements, however since a measurement error variance of zero is not permitted in the particular implementation of the multiscale estimator used here,[7] a measurement noise variance of 10^{-4} was specified.

Being given measurements at a subset of the image pixels, as opposed to a dense set of measurements as in the previous two examples, just implies a trivial change in the measurement projection operator G_y and, consequently, in the multiscale mea-

[7] Although such an estimator *could* be implemented.

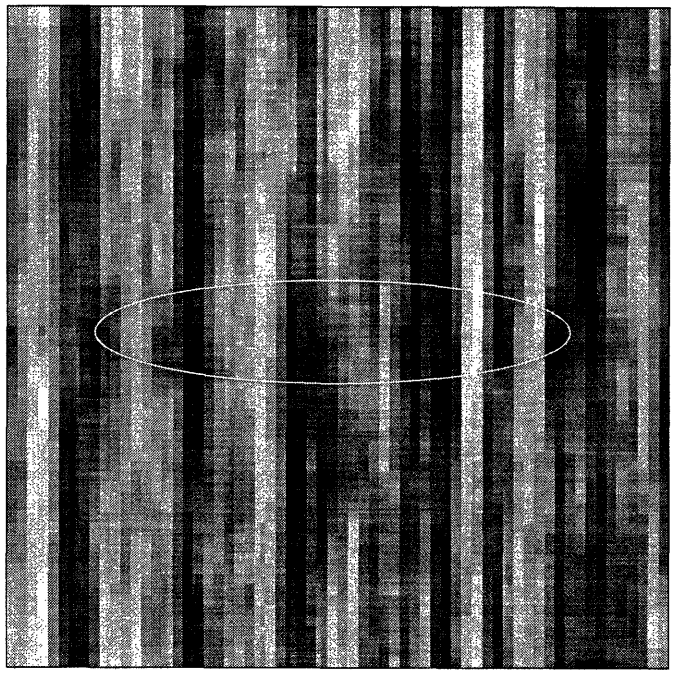


Figure 6-15: Original "wood" texture, 64 × 64 samples, simulated using Markov Random Field techniques. The small ellipse indicates the set of pixels to be measured.

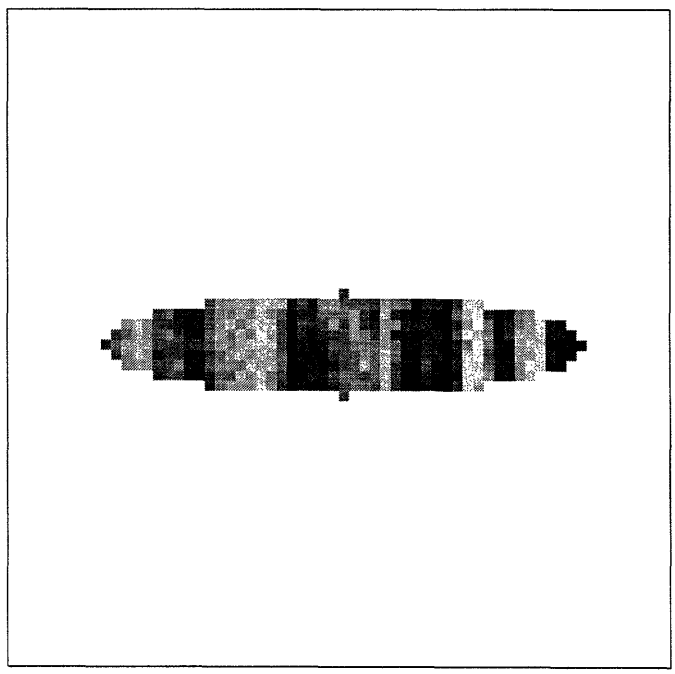


Figure 6-16: Noiseless observations of a small subset of Figure 6-15.

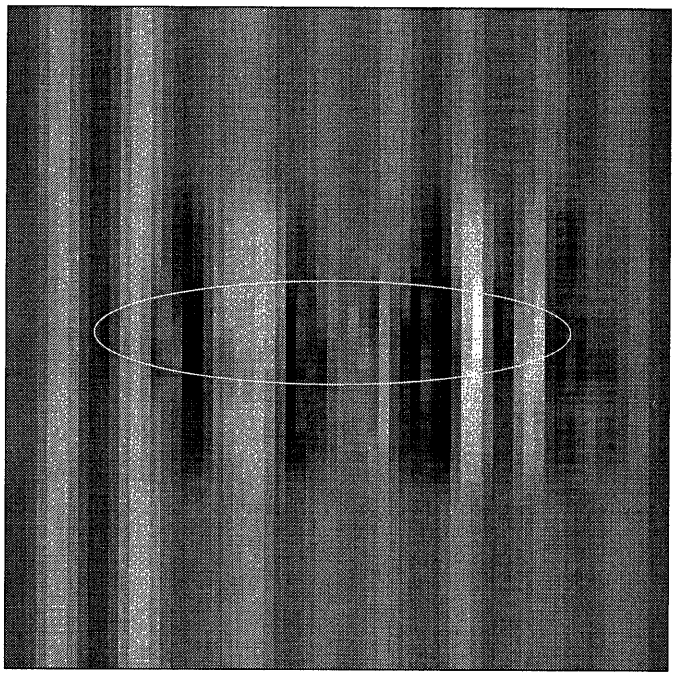


Figure 6-17: The estimated texture using an overlapped multiscale model, based on the measurements of Figure 6-16. Despite the use of a multiscale estimator, the estimates evolve smoothly from the region in which measurements are present to the surrounding area without measurements.

surement matrices on the finest scale of the tree. Otherwise there is no essential difference, and the multiscale model development and estimation procedure proceed unaffected. It is rather significant to note, however, that while the multiscale framework is readily adapted to the loss of measurements, a change from dense to sparse sampling immediately makes FFT-based approaches inapplicable.

Figure 6-17 shows the multiscale reconstruction based on the limited set of measurements given in Figure 6-16. Again, two observations should be made:

- The multiscale estimator does capture the coarse features of the original texture of Figure 6-15 outside of the measured region. Even certain aspects of the vertical bands to the left and right of the measured region are properly captured.

- Despite the fact that we are using a multiscale estimator, the estimated texture evolves smoothly as we move away from the measured pixels.

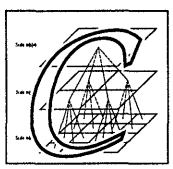
6.5 Conclusions

This chapter has presented and developed an overlapping tree approach to modeling and estimation on a multiscale tree. We defined three projection operators G_x, G_y and H_x, which relate the original or physical domain to the finest scale of the overlapped tree. If a multiscale model is chosen which is faithful to the statistics of the process on the finest level of the overlapped tree, then under certain sufficient conditions the estimates which are computed by the overlapping method can be shown to be optimal. Coupling this overlapping framework with a multiscale stochastic realization technique based on canonical correlations yields a powerful estimation and modeling tool which offers tradeoffs between estimate smoothness, statistical fidelity, and computational effort.

The examples at the end of the chapter demonstrated some of the potential for the overlapping multiscale framework. Three examples were presented, two of which could not be solved by FFT methods; in each case the overlapping multiscale tree computed estimates free from obvious artifacts. Chapter 7 will continue the study of overlapping multiscale models, although with a different sort of prior model, and will present a further set of estimation examples based on overlapping multiscale trees.

Chapter 7

Surface Reconstruction



hapter 6 discussed the development of an overlapping framework, capable of computing smooth and aesthetically appealing estimates on a multiscale tree. This capability motivates an exploration of the application of our multiscale framework to a problem in computer vision: this chapter will develop multiscale models for surface reconstruction. The chapter is organized as follows: Section 7.2 reviews variational model development for surface reconstruction; Section 7.3 details the construction of multiscale models corresponding to a given variational counterpart; and Section 7.4 presents several experimental results.

7.1 Introduction

The problem of surface reconstruction[8, 31, 40, 44, 103, 111] has been a topic of considerable interest in the field of computer vision for some time, involving the estimation of an unknown surface based on a set of noisy measurements of some function of the surface and its derivatives and based on a prior model for the surface (generally necessary to regularize the problem). Variational methods[18, 111] have enjoyed considerable success in dealing with surface reconstruction problems, both as an analytical means of formulating the problem, and as a means of determining a method of solution. However, as discussed in Section 2.2, the solution of the associated Euler-Lagrange[18] partial differential equations (PDEs) may represent a significant

computational task, and the calculation of reconstruction error covariances[51, 98] are, for all practical purposes, completely infeasible for large problems.

This chapter is motivated by a class of problems for which FFT techniques are inapplicable, and for which the solution of the Euler-Lagrange equations via variational approaches is impractical, but for which the multiscale framework of this thesis, and more specifically, the overlapping multiscale framework of the previous chapter, may yield significant improvements. We will be taking advantage of a dual interpretation of variational problems (Section 2.2): variational problems with quadratic costs may be interpreted as Gaussian least-square statistical estimation problems. Specifically, as has been noted by others[99], the membrane and thin-plate variational models commonly used in surface reconstruction allow interpretations as $1/f^2$ prior statistical models, for which we can develop multiscale counterparts.

In fact, the method of this chapter allows a broader class of surface models than just thin-plate and membrane. The flexibility of the multiresolution framework allows us to define a far richer class of surface reconstruction models and algorithms corresponding to different prior models which have either more complicated or *no* variational counterparts, but that admit the same efficient solutions.

7.2 Background

7.2.1 Notation

The general surface reconstruction problem[103] involves estimating the shape of a surface given a discrete (and possibly sparse) set of noisy observations of some function of the surface and/or its gradients. The surface of interest is a two-dimensional function $z(x, y)$ presumed twice differentiable everywhere; denote by

$$p(x,y) = z_x(x,y) = \frac{\partial z(x,y)}{\partial x} \qquad q(x,y) = z_y(x,y) = \frac{\partial z(x,y)}{\partial y} \tag{7.1}$$

the gradients of the surface at each point. Normally we shall refer simply to z, p, q etc; the dependence on x and y will be implicit. We are interested in least squares

problems of the following type:

$$\hat{z}(Y) = \min_z \left\{ (Y - \mathcal{C}(z))^T \mathcal{R}^{-1} (Y - \mathcal{C}(z)) + \mathcal{Z}(z) \right\} \tag{7.2}$$

$$Y = \mathcal{C}(z) + V \tag{7.3}$$

where Y is a discrete vector of observations, corrupted by white noise V, and where V has covariance $\mathcal{R}$. $\mathcal{C}$ is the measurement function, and $\mathcal{Z}$ represents a prior model for the surface z. The specific nature of the measurement function $\mathcal{C}$ will typically depend on the specific surface reconstruction problem of interest; for example we may directly observe a sparse subset of the surface heights,

$$\mathcal{C}(z) = \begin{bmatrix} z(x_1, y_1) \\ z(x_2, y_2) \\ \vdots \end{bmatrix} \tag{7.4}$$

or we may just observe a subset of the surface gradients,

$$\mathcal{C}(z) = \begin{bmatrix} p(x_1, y_1) \\ q(x_1, y_1) \\ p(x_2, y_2) \\ q(x_2, y_2) \\ \vdots \end{bmatrix} \tag{7.5}$$

7.2.2 Surface Prior Models

A least-squares solution for the surface z given the measurements alone is typically ill-posed, i.e., there is not a single optimum solution for the surface. In order to guarantee a unique solution the problem is regularized by asserting a prior model $\mathcal{Z}$ for the unknown surface, typically reflecting our prior knowledge about the surface to be reconstructed or, equivalently, asserting certain desired smoothness properties for the reconstructed surface.

One of the simplest and most common surface prior models is to assert a smooth-

ness constraint[44]:

$$\mathcal{Z}_p(z) = \alpha \int\int \left\{ p_x^2 + p_y^2 + q_x^2 + q_y^2 \right\} dxdy \tag{7.6}$$

This constraint is also referred to as a "thin plate" term, in that (7.6) represents the potential energy in an isotropic thin plate[18].

An alternative function, representing the potential energy contained in a stretched membrane[18], punishes variations from $p = 0, q = 0$:

$$\mathcal{Z}_m(z) = \beta \int\int \left\{ p^2 + q^2 \right\} dxdy \tag{7.7}$$

Combining (7.6),(7.7) yields a variational formulation familiar to researchers in computer vision[44]:

$$\hat{z}(Y) = \min_z \left\{ (Y - \mathcal{C}(z))^T \mathcal{R}^{-1} (Y - \mathcal{C}(z)) + \iint \alpha \left\{ p_x^2 + p_y^2 + q_x^2 + q_y^2 \right\} + \beta \left\{ p^2 + q^2 \right\} dxdy \right\} \tag{7.8}$$

There are two common interpretations of this formulation:

- The regularization term of (7.8) is a special case of the class of two-dimensional generalized spline functionals[103, 109]

$$| z |_m^2 = \int\int_{R^2} \sum_{i=0}^{m} \binom{m}{i} \left(\frac{\partial^m z}{\partial x^i \partial y^{m-i}} \right)^2 dxdy \tag{7.9}$$

  for the cases $m = 1, 2$, where the value of m determines the order of continuity in the solutions for z.

- The regularization term of (7.8) represents a deformable sheet[111] or a stiff surface, being acted upon by forces (i.e., the observations), where the resulting deformation is a function of the specific stiffness properties of the sheet. In particular, a membrane term $\mathcal{Z}_m$ minimizes the surface "area" (like a rubber sheet), whereas the thin plate term $\mathcal{Z}_p$ minimizes the surface curvature (like a

steel plate).

7.2.3 Explicit Estimation of Surface Gradients

One final addition to the variational formulation is appropriate. In many surface reconstruction applications, the gradients of the surface play a central role (the most notable example being the shape-from-shading problem[44, 45]). It is frequently of interest to estimate the gradients explicitly, rather than to infer them implicitly as a function of the estimated surface.

For example, consider a problem in which the measurements are functions of the surface gradients only, not the surface heights themselves. In such problems, we may be motivated to use the following common variational equation[44]:

$$\min_{p,q} \left\{ (Y_g - \mathcal{C}(p,q))^T \mathcal{R}_g^{-1} (Y_g - \mathcal{C}(p,q)) + \int\int_{R^2} \alpha \left(p_{\mathrm{x}}^2 + p_{\mathrm{y}}^2 + q_{\mathrm{x}}^2 + q_{\mathrm{y}}^2 \right) + \beta \left(p^2 + q^2 \right) dxdy \right\} \tag{7.10}$$

where Y_g represents gradient-dependent observations, and where the measurement function $\mathcal{C}_g(p,q)$ explicitly depends upon gradient terms only. In general, the estimates of p, q resulting from such a variational equation will not correspond to the gradients of *any* surface z – this is the well known integrability problem[31, 43, 55]. Our multiscale surface reconstruction model (to be outlined in Section 7.3), will similarly be estimating the surface gradients explicitly, hence the relevance of the following discussion.

In order for p, q to be gradients of a surface, the consistency constraint

$$\oint_{\mathcal{L}} (pdx + qdy) = 0 \tag{7.11}$$

must hold over all closed paths $\mathcal{L}$ in the plane[40]. In other words,

$$p_{\mathrm{y}} = q_{\mathrm{x}} \tag{7.12}$$

at all points in the plane, leading to the revised variational problem[40]

$$\min_{p,q} \Big\{ (Y_g - \mathcal{C}_g(p,q))^T \mathcal{R}_g^{-1} (Y_g - \mathcal{C}_g(p,q)) + \int\int_{R^2} \alpha \left(p_{\mathrm{x}}^2 + p_{\mathrm{y}}^2 + q_{\mathrm{x}}^2 + q_{\mathrm{y}}^2 \right) + \beta \left(p^2 + q^2 \right) + \gamma \left(p_{\mathrm{y}} - q_{\mathrm{x}} \right)^2 dxdy \Big\} \quad (7.13)$$

The inclusion of the additional penalty term in the above expression does not guarantee that (7.11) is exactly satisfied, although using a large value for γ will in general result in nearly consistent (p, q) fields. Even if this consistency relationship is exactly satisfied, however, we still have the non-trivial problem of computing z from the (p, q) fields. On the other hand, this problem *and* the consistency problem may be avoided by explicitly reconstructing z as well as (p, q) through a simple surface-gradient consistency penalty

$$\int\int_{R^2} \left(z_{\mathrm{x}} - p \right)^2 + \left(z_{\mathrm{y}} - q \right)^2 dxdy \quad (7.14)$$

This leads to the following variational problem, where for generality, we allow both direct measurements $Y_s = \mathcal{C}_s(z) + V_s$, as well as gradient measurements Y_g:

$$\min_{z,p,q} \Big\{ (Y_s - \mathcal{C}_s(z))^T \mathcal{R}_s^{-1} (Y_s - \mathcal{C}_s(z)) + (Y_g - \mathcal{C}_g(p,q))^T \mathcal{R}_g^{-1} (Y_g - \mathcal{C}_g(p,q)) + \int\int_{R^2} \alpha \left(p_{\mathrm{x}}^2 + p_{\mathrm{y}}^2 + q_{\mathrm{x}}^2 + q_{\mathrm{y}}^2 \right) + \beta \left(p^2 + q^2 \right) + \gamma \left(\left(z_{\mathrm{x}} - p \right)^2 + \left(z_{\mathrm{y}} - q \right)^2 \right) dxdy \Big\} \quad (7.15)$$

This variational problem, which is similar to one introduced in [40], forms the point of departure for our analysis and in particular for the development of efficient multi-scale counterparts to problems of this type. We will restrict our attention to *linear* measurements of the surface and its gradients, as in [102, 103].

7.2.4 Euler-Lagrange Equations

After a specific variational expression has been selected, the solution for the optimal estimated surface $\hat{z}(Y)$ can be characterized in a straightforward manner. For

example, the Euler-Lagrange equation[18, 103] corresponding to (7.8) is

$$\frac{\partial}{\partial z}(Y - \mathcal{C}(z))\mathcal{R}^{-1}(Y - \mathcal{C}(z))^T - \alpha\left\{\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2}\right\} + \beta\left\{\frac{\partial^4 z}{\partial x^4} + 2\frac{\partial^4 z}{\partial x^2 y^2} + \frac{\partial^4 z}{\partial y^4}\right\} = 0 \tag{7.16}$$

This is an elliptic PDE which, after specifying appropriate boundary conditions, may be solved numerically by discretizing the PDE and applying numerical solution techniques. In special cases extremely efficient FFT techniques may be applied. Typically, however, FFT techniques are *not* applicable, leading researchers to propose a variety of other techniques[44, 83, 111]. These techniques successfully estimate the surface by solving the PDE in a comparatively efficient manner, however they are unable to produce estimation error statistics for reasons outlined in Sections 2.1 and 2.2. Instead, we propose to replace the variational model by a similar multiscale one, allowing efficient computation of the surface estimates *and* estimation error variances.

7.3 Multiscale Model Development for Variational Problems

This section begins by developing multiscale counterparts to each of the terms in a variational cost function such as (7.15). Next, once a tentative multiscale model has been established, the model is adapted for use in the overlapping framework of Chapter 6.

7.3.1 Equivalent Estimation Problem

Consider the following variational problem, a variation of (7.15):

$$\min_{z,p,q} \left\{(Y - \mathcal{C}(z,p,q))^T\mathcal{R}^{-1}(Y - \mathcal{C}(z,p,q)) + \int\int_{R^2}\left[\mathcal{L}_1(z,p,q)^T\mathcal{L}_1(z,p,q) + \mathcal{L}_2(z,p,q)^T\mathcal{L}_2(z,p,q)\right]dxdy\right\} \tag{7.17}$$

where $\mathcal{L}_1, \mathcal{L}_2$ are column vectors of linear functionals of z, p, q, and where Y may contain both height and gradient measurements (e.g., the Y_s, Y_g of (7.15)) corresponding to the linear measurement functional $\mathcal{C}$.

Let $Z^T = [\ldots, z(x_i, y_j), \ldots]^T$ represent a vector of samples of $z(x, y)$; similarly define P, Q. Then, based on our discussion in Section 2.2, (7.17) may be discretized as

$$\min_{Z,P,Q} \left\{ \left(Y - C \begin{bmatrix} Z \\ P \\ Q \end{bmatrix} \right)^T \mathcal{R}^{-1} \left(Y - C \begin{bmatrix} Z \\ P \\ Q \end{bmatrix} \right) + \left\{ L_1 \begin{bmatrix} Z \\ P \\ Q \end{bmatrix} \right\}^T \left\{ L_1 \begin{bmatrix} Z \\ P \\ Q \end{bmatrix} \right\} + \left\{ L_2 \begin{bmatrix} Z \\ P \\ Q \end{bmatrix} \right\}^T \left\{ L_2 \begin{bmatrix} Z \\ P \\ Q \end{bmatrix} \right\} \right\} \tag{7.18}$$

where L_1, L_2, C are matrices representing discrete approximations of the linear functionals $\mathcal{L}_1, \mathcal{L}_2, \mathcal{C}$ respectively.

From Section 2.2 we know that optimization problems such as (7.18) have counterparts as statistical estimation problems; furthermore from the prior–measurement duality discussed in Section 2.3 we know that there is a certain degree of flexibility in the choice of the measurement and prior models. Taking advantage of this flexibility, we propose to consider the following estimation problem:

$$\begin{bmatrix} Y \\ 0 \end{bmatrix} = \begin{bmatrix} C \\ L_1 \end{bmatrix} \begin{bmatrix} Z \\ P \\ Q \end{bmatrix} + \bar{v} \qquad \bar{v} \sim \mathcal{N}\left(0, \begin{bmatrix} \mathcal{R} & 0 \\ 0 & I \end{bmatrix}\right) \tag{7.19}$$

$$L_2 \begin{bmatrix} Z \\ P \\ Q \end{bmatrix} = \bar{w} \qquad \bar{w} \sim \mathcal{N}(0, I) \tag{7.20}$$

The estimation problem (7.19),(7.20) is the same as the original estimation problem in (7.18).

The next three subsections discuss the determination of appropriate values of C, L_1, L_2, followed by the synthesis of the complete multiscale model appropriate for surface estimation.

7.3.2 Measurement Model

In our multiscale model the nodes at the finest scale correspond to the scale at which we both have measurements and wish to perform surface reconstruction. Thus the measurements (2.28) in our multiscale model are defined only at the finest scale, i.e., $C(s) = 0$ except for nodes at the finest scale (and even at that scale $C(s)$ may be zero at some nodes if the surface measurements are sparse). For example, in Section 7.4 we will focus primarily on direct surface measurements. In such cases, if $z(s)$ is the first component of $x(s)$, then our measurement will take the form of (2.28) with $C(s) = [C_1(s), 0, \ldots]$ where $C_1(s) = 1$ at those finest scale nodes at which we have direct measurements of the surface.

7.3.3 Quadratic penalties on state derivatives, e.g., z_x^2

The basic idea for this case, first discussed in [14, 63, 99], was outlined in Section 2.5.4. Essentially, a variational cost

$$\int\int \left[z_x^2 + z_y^2\right] dxdy \tag{7.21}$$

is similar to a Brownian motion prior model[63], that is, a process with a $1/f^2$ spectrum. Consequently we propose the following multiscale prior model which possesses a similar $1/f^2$-like power spectrum:

$$z(s) = 1 \cdot z(s\bar{\gamma}) + B_o 2^{-m(s)/2} w(s) \tag{7.22}$$

7.3.4 Quadratic penalties on linear combinations of state variables:

The other components of the prior model involve quadratic costs of linear functions of variables, e.g.,

$$\int\int \left((z_x - p)^2 + (z_y - q)^2\right) dxdy \tag{7.23}$$

which asserts the relationship between the surface and its gradients. As discussed in Section 2.3, we can (and do) choose to interpret such terms as part of the measurement model. For example, the penalty $\int\int(z_x - p)^2$ has the direct statistical interpretation as a measurement of the form

$$0 = z_x(x,y) - p(x,y) + v(x,y) \tag{7.24}$$

where $v(x,y)$ is a spatially white "measurement noise" with a variance reflecting the weight placed on the $(z_x - p)$ penalty in the variational problem.

"Measurements" of the form (7.24) are not in the most convenient form for multiscale implementation. However observe that

$$(z_x - p)^2 + (z_y - q)^2 = \tfrac{1}{2}(z_x + z_y - p - q)^2 + \tfrac{1}{2}(z_x - z_y - p + q)^2 \tag{7.25}$$

which leads to the following types of measurements:

$$0 = z_x(x,y) + z_y(x,y) - p(x,y) - q(x,y) + v_1(x,y) \tag{7.26}$$

$$0 = z_x(x,y) - z_y(x,y) - p(x,y) + q(x,y) + v_2(x,y) \tag{7.27}$$

As we describe next, these measurements are readily captured in our quadtree structure.

Now in order to incorporate a discretized version of (7.26) or (7.27) into our multiscale framework we must define appropriate approximations to the derivatives $z_x(x,y)$ and $z_y(x,y)$. In particular, we will want to define such an approximation at all scales. Since the concept of the gradient of a $1/f^2$-like surface is ill-defined at best, we have some flexibility in how we choose to do this, and we have taken advantage of this flexibility to specify an approximation that leads to a very simple model. Consider Figure 7-1, in which we have portrayed a parent pixel $s\bar{\gamma}$ and its four descendants s_1, s_2, s_3, s_4. We consider each node on the tree to represent a particular point on the surface, chosen to be the point in the center of the region aggregated by each node: the points are marked as a filled circle • for each of the descendent nodes,

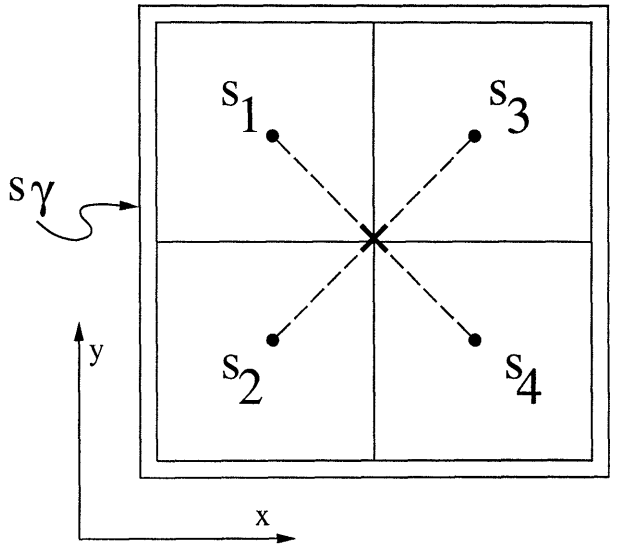


Figure 7-1: A set of five node labels for gradient discussion purposes. s_1, s_2, s_3, s_4 are each children of coarser node $s\bar{\gamma}$. The $\times$ indicates the point represented by node $s\bar{\gamma}$; the solid circles $\bullet$ specify the points represented by nodes $s_1, \ldots, s_4$.

and as a cross $\times$ for node $s\bar{\gamma}$. Measured in units of finest-scale pixels, the point $\bullet$ in each of the child nodes is separated from $\times$ by a distance $\sqrt{2} \cdot 2^{m(s_1)-M}$. One possible approximation to $z_x \pm z_y$ is to define the surface gradients in terms of values at the four child nodes; e.g.,

$$z_x(s_3) + z_y(s_3) \simeq 2^{m(s_3)-M+1} \left(z(s_3) - z(s_2)\right) \tag{7.28}$$

where $2^{m(s_3)-M+1}$ measures the separation between nodes s_2 and s_3 along the x and y directions. This approach has the undesirable property of leading to a higher-order model than an alternative choice. Instead, we define our approximation to gradients at each child node in terms of the *parent* value as well:

$$z_x(s_3) + z_y(s_3) \simeq 2^{m(s_3)-M} \left(-z(s\bar{\gamma}) + z(s_3)\right) \tag{7.29}$$

$$z_x(s_4) - z_y(s_4) \simeq 2^{m(s_4)-M} \left(-z(s\bar{\gamma}) + z(s_4)\right) \tag{7.30}$$

with analogous definitions at nodes s_1, s_2.

Consequently, if we wish to view (7.26) as a "measurement" we are led to discrete

measurements of the form

$$0 = 2^{m(s_3)-M}\left(-z(s\bar{\gamma}) + z(s_3)\right) - p(s_3) - q(s_3) + v(s_3) \tag{7.31}$$

$$0 = 2^{m(s_2)-M}\left(-z(s\bar{\gamma}) + z(s_2)\right) + p(s_2) + q(s_2) + v(s_2) \tag{7.32}$$

which we will interpret as a measurement at node s_3 and s_2 respectively. In a similar fashion we can defined measurements corresponding to (7.27):

$$0 = 2^{m(s_1)-M}\left(-z(s\bar{\gamma}) + z(s_1)\right) + p(s_1) - q(s_1) + v(s_1) \tag{7.33}$$

$$0 = 2^{m(s_4)-M}\left(-z(s\bar{\gamma}) + z(s_4)\right) - p(s_4) + q(s_4) + v(s_4) \tag{7.34}$$

For these "measurements" to be in the form of (2.28) all of the variables in (7.31)–(7.34) (other than the noise terms $v(s_i)$) must be in the respective state vectors $x(s_i)$. The measurements (7.31)–(7.34) require that $x(s_i)$ include the parent value $z(s\bar{\gamma})$; this is accomplished easily through state augmentation as described in the next subsection.

7.3.5 Elementary Multiscale Model Synthesis

Combining the model components described in the preceding subsections leads to the following elementary multiscale estimation problem corresponding to the deformable-sheets variational problem (7.15). Specifically, the state dynamics are given by

$$\begin{bmatrix} z \\ p \\ q \\ zp \end{bmatrix}(s) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} z \\ p \\ q \\ zp \end{bmatrix}(s\bar{\gamma}) + \begin{bmatrix} B_s 2^{-m(s)/2} & 0 & 0 \\ 0 & B_g 2^{-m(s)/2} & 0 \\ 0 & 0 & B_g 2^{-m(s)/2} \\ 0 & 0 & 0 \end{bmatrix} w(s) \tag{7.35}$$

Here the first component of the dynamics captures the thin membrane term (7.7) and the next two the thin plate penalty (7.6). The last component accomplishes the state augmentation required for the penalty (7.14) as discussed in the preceding

section: $zp(s)$ is simply $z(s\bar{\gamma})$. The measurement equation accompanying this model includes both the actual measurements (at the finest scale) as well as measurements like (7.31)–(7.34) which assert the surface-gradient consistency. For example, if we have only direct surface height measurements, the measurement equation at each of the four nodes $s_i, 1 \leq i \leq 4$ in Figure 7-1 would take the form

$$\begin{bmatrix} d(s_i) \\ \\ 0 \end{bmatrix} = \begin{bmatrix} C_1(s_i) & 0 & 0 & 0 \\ \\ -2^{m(s_i)-M} & a_i & b_i & 2^{m(s_i)-M} \end{bmatrix} \begin{bmatrix} z \\ p \\ q \\ zp \end{bmatrix} (s_i) + v(s_i) \tag{7.36}$$

where $C_1(s_i)$ (from Section 4.1) equals one at those nodes at which we have measurements (zero otherwise), and a_i, b_i are given by

$$\begin{array}{cccc} a_1 = -1 & a_2 = -1 & a_3 = 1 & a_4 = 1 \\ b_1 = 1 & b_2 = -1 & b_3 = 1 & b_4 = -1 \end{array} \tag{7.37}$$

where a_i and b_i specify the orientation of s_i with respect to $s\bar{\gamma}$ along the x and y axes respectively (see Figure 7-1).

There are several final points that we should make about this multiscale model. First, there is the issue of specifying the unknown parameters, e.g., the B_s, B_g of (7.35). This is the same type of problem as the the selection of appropriate weights α, β, γ in the original variational formulation (7.15); for example, by adjusting the relative sizes of these quantities we can control the relative importance of the thin-plate and membrane terms. However, now that we have a precise statistical interpretation of these terms, we can use that to advantage in determining these quantities. In particular B_s, B_g have an explicit physical meaning, measured in real physical units, that represent the prior statistical knowledge of the surface; e.g., if the surface varies over a range of $\pm$15cm, then $B_s = 10$cm would make a reasonable choice. In addition, based on the multiscale likelihood and parameter identification methods that our models admit (as described in Chapter 4), one can estimate the optimal values of these parameters from the data and perform model validation. Finally, while the model (7.35),(7.36) does capture many of the features of the variational problem with

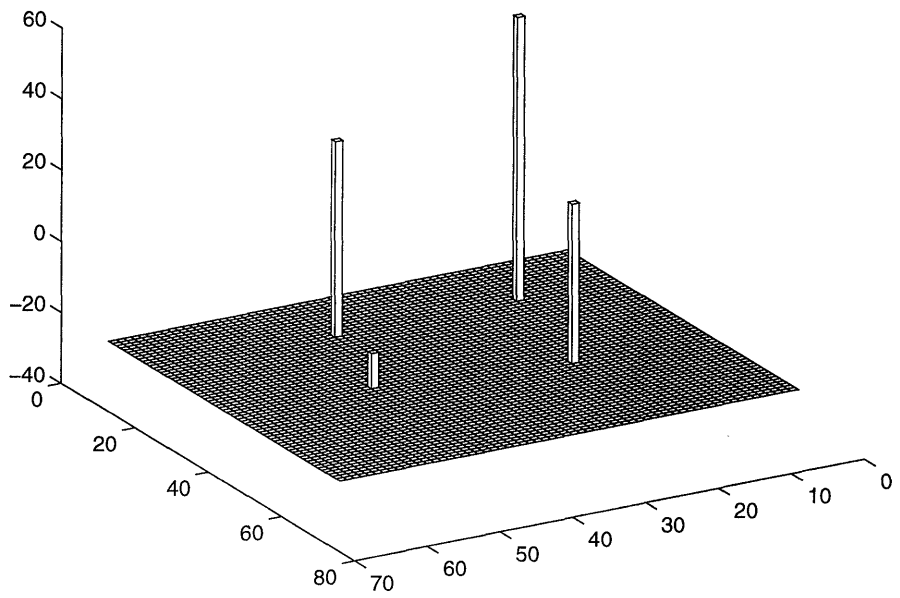


Figure 7-2: Four sparse measurements used in testing the elementary multiscale surface reconstruction model.

which we started, there is one last issue that is of consequence in some applications and which necessitates a change in the multiscale model. This is discussed next.

7.3.6 Advanced Multiscale Model Development

The straightforward application of the elementary multiscale model (7.35),(7.36) derived in Section 7.3 would be to implement it directly on a quad-tree. Figure 7-3 demonstrates the result of such an approach, based on the set of four, sparse measurements from Figure 7-2. The computation of this set of estimates *and* estimation error variances (not shown) required about 3 seconds of SPARC-10 computer time.

One striking feature of these estimates is their blockiness, or lack of smoothness, characteristic of many quad-tree algorithms. As argued in Chapter 6, in some problems the quality and quantity of information available for reconstruction may be sufficiently low that *no* statistically significant fine scale estimates can be computed. Nevertheless, in many applications there are compelling reasons for producing smooth reconstructions, something that the original variational formulations gener-

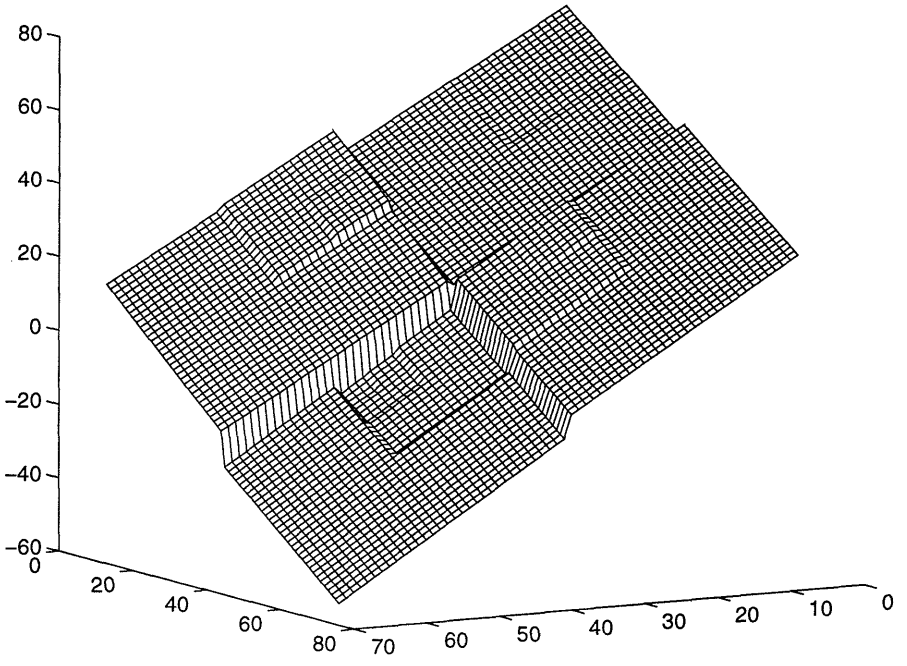


Figure 7-3: The surface estimates resulting from a straightforward implementation of the elementary multiscale model.

ally do. We propose to apply the overlapping framework of Chapter 6 to produce smoother estimates; that is, we wish to select projection operators H_x, G_x and G_y which relate the original domain and the finest scale of the overlapping tree structure.

Since the projection onto an overlapping tree involves a change in the multiscale tree structure from that used in the preceding subsection, it is also necessary to make a corresponding change in the model parameters on the tree. Because of the interpretation of thin-plate and membrane models as fractal priors, we can readily develop such a model motivated by results on hierarchical fractal surface synthesis in[93]. In particular, a key quantity in this construction is the ratio of the dimensions of the multiscale pixels in going from scale to scale. Specifically, if w_m denotes the size at scale m, then the key quantity is

$$r = \frac{w_{m+1}}{w_m} \tag{7.38}$$

which is related to the *lacunarity*[70, 93], or texture, of the synthesized surface. Then in constructing a $1/f^2$ model on such a pyramidal structure the variance of the detail

added in going from the mth to the $(m+1)$st scale is proportional to $(1-r)r^m$. In a non-overlapped tree $r = \frac{1}{2}$, resulting in variances that decrease by a factor of two from scale to scale as in (7.35). For an overlapped tree $r > \frac{1}{2}$, and thus when we use an overlapped tree with lacunarity r we modify the model (7.35), replacing noise gains $2^{-m(s)/2}$ by $2(1-r)r^{m(s)/2}$. Note also that since we are decreasing the size of the regions more slowly as we move to finer scales, our trees will necessarily require more scales than in the non-overlapped case. In particular, if we require M scales in the non-overlapped case, we will need $M/\log_2(1/r)$ scales in the overlapped case. Since there is an increase in the computational complexity as the number of tree levels increases, we have a tradeoff between the amount of overlap (corresponding to r closer to one, and to greater smoothness in the reconstructions) and computational effort.

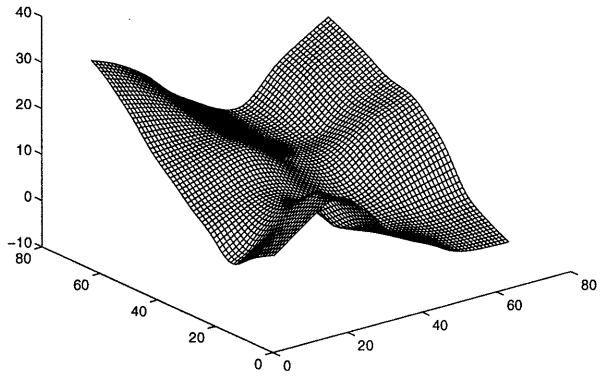
7.4 Reconstruction Examples

This section presents four surface reconstruction experiments, illustrating the application of our multiscale framework in a variety of surface-reconstruction scenarios: densely sampled surfaces, sparsely sampled surfaces, discontinuous surfaces, and surfaces not characterized by variational thin-plate/membrane priors. For each multiscale example an overlap vector[1] $\mathcal{O} = (o_1, \ldots, o_{M-1})$ will be given, which characterizes the overlapping tree structure. The means of determining the operators G_x, G_y, and H_x from $\mathcal{O}$ is summarized in Appendix E.

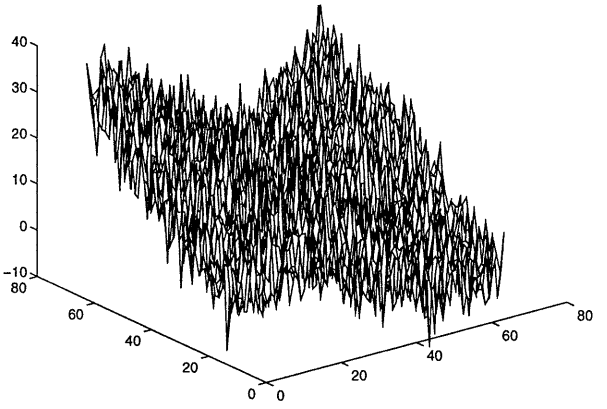
7.4.1 Densely Sampled Measurements

Figure 7-4 summarizes our experimental results for surface reconstruction of a densely sampled surface. A smooth surface of size 64×64 pixels is shown in Figure 7-4(a) and the corresponding dense measurements with added Gaussian noise (variance 25) are shown in (b). The particular surface used does not correspond exactly to either

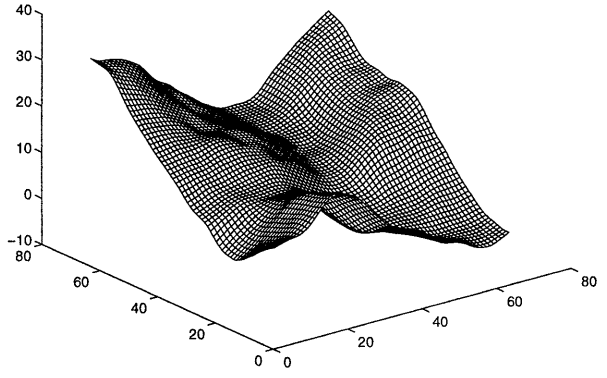
[1] o_m, defined in Figure 6-5, represents the number of pixels by which neighboring nodes on the mth scale overlap.



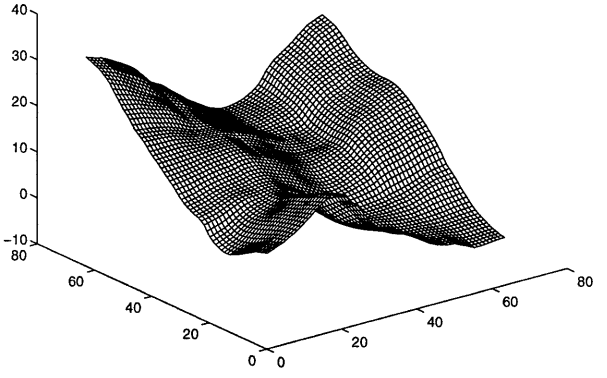
(a) Original surface to be estimated.



(b) Original surface plus noise.



(c) Multigrid Reconstruction.



(d) Multiscale reconstruction.

Figure 7-4: Dense measurement reconstruction example: a surface is reconstructed based on dense measurements with 5cm Gaussian noise. The multiscale and multigrid approaches involve the same number of computations. $B_s = 80, B_g = 0.4, \mathcal{O} = (12, 6, 4, 3, 0, 0, 0)$.

the variational or multiscale prior (and, in fact, is synthesized as a weighted blend of the two) and thus a comparison between variational and multiscale reconstruction is not biased in favor of either formulation.

Figure 7-4(c) shows the reconstructed surface using an iterative multigrid algorithm[39, 74, 102, 103, 110] to solve the variational problem based on a thin-plate prior model (i.e., $\alpha = 1, \beta = 0$ in (7.15)). The multigrid algorithm was chosen as it represents one of the most efficient methods for solving the Euler-Lagrange PDE arising in surface reconstruction variational formulations and thus provides a fair test

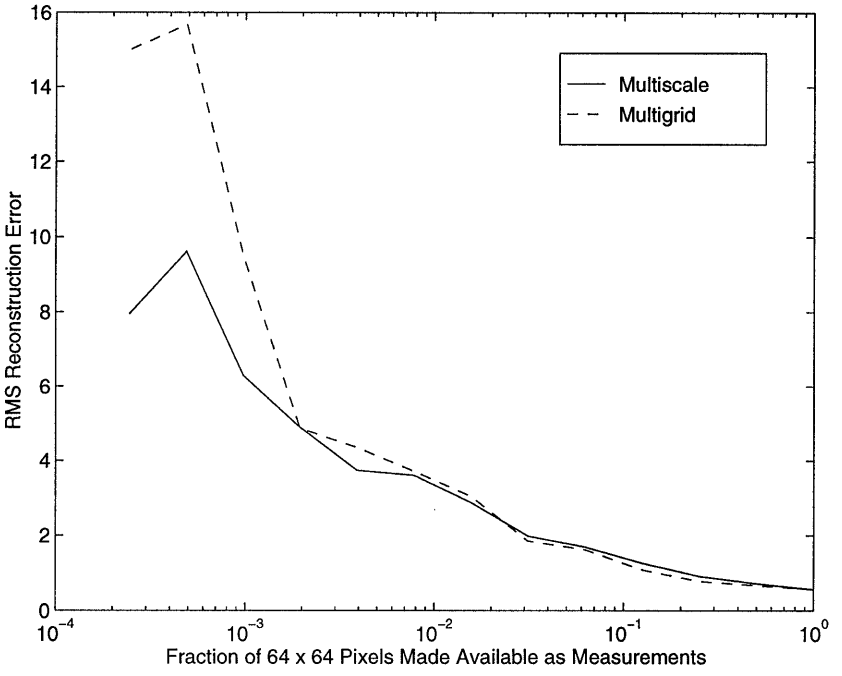
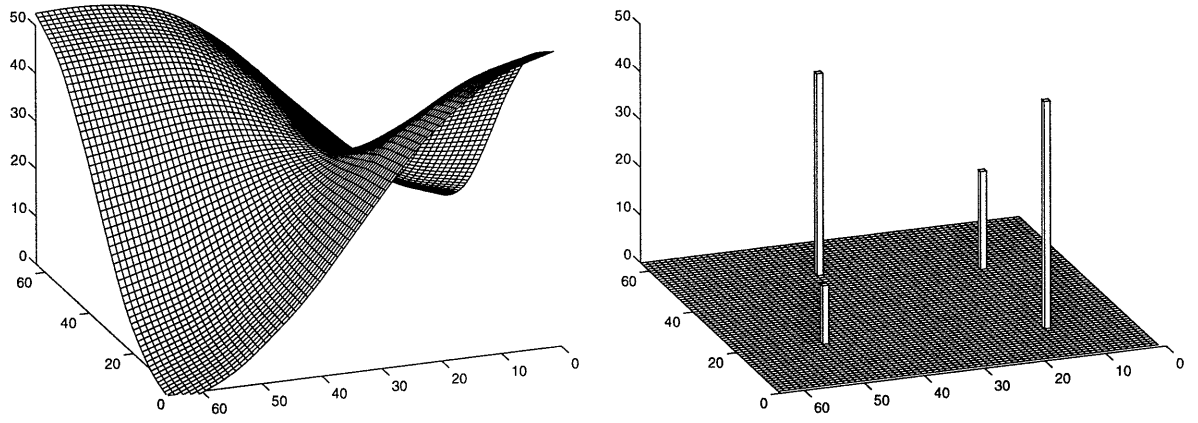


Figure 7-5: RMS error in surface estimation using multiscale and multigrid methods as a function of the proportion of noisy measurements retained.

of our claim that our non-iterative multiscale approach is superior. In particular, in all comparisons involving our multiscale approach and an iterative method (such as multigrid), the number of iterations for the iterative methods is selected such that the total computational effort of each method is the same. Figure 7-4(d) shows the reconstructed surface using our proposed multiscale algorithm computed using an overlapping tree with 8 scales (a *non*-overlapped tree for this surface would require 7 scales) and a thin-plate-like prior model (i.e., the prior model is dominated by gradient constraints by choosing $B_s \gg 16B_g$). As indicated in the figure caption, $\mathcal{O} = (12, 6, 4, 3, 0, 0, 0)$, so the regions corresponding to nodes on the first level below the root node have an overlap with a width of 12 pixels; at the next level the overlap is 6 pixels etc.

The multigrid and multiscale reconstructions of Figure 7-4 are arguably equally good – both give reasonable estimates of the original surface, and our multiscale approach shows no signs of blockiness. A more precise comparison is provided by Figure 7-5. Here we consider a set of reconstruction problems in which only a randomly sampled subset of the noisy surface measurements is used. What Figure 7-5



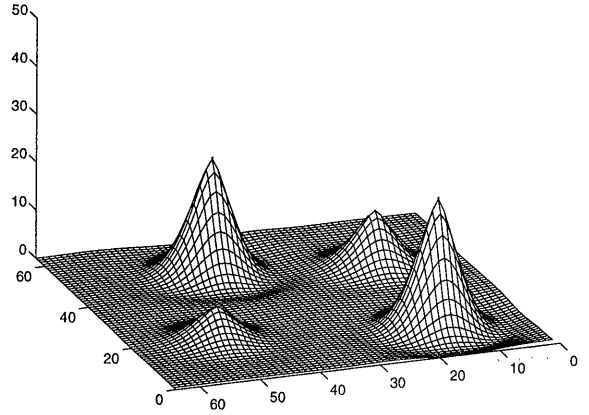
(a) Original "truth" surface. (b) Four sparse measurements.

Figure 7-6: The "truth" surface is shown in (a). Our sparse-data reconstruction examples will be based on the four surface samples shown in (b).

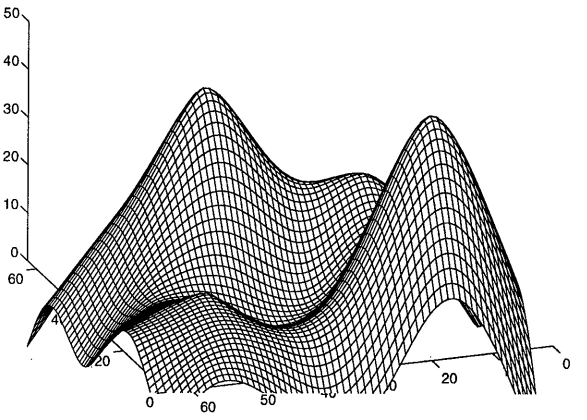
depicts is the RMS reconstruction error for both the multiscale and multigrid methods as a function of the *fraction* of noisy surface measurements used. Observe that for the same computational effort, the multiscale algorithm performs as well *or better* than the multigrid approach in an RMS sense. Thus the multiscale model which we have developed should *not* simply be viewed as an approximation to a variational equation, rather we are motivated by a certain variational form to develop a surface prior model – indeed, a conspicuously effective one – that leads to competitive reconstructions. Moreover it must be emphasized that the multiscale algorithm is computing surface estimates *and* estimation error variances in the time that multigrid computes surface estimates only.

7.4.2 Sparse Data and Multiscale Preconditioning

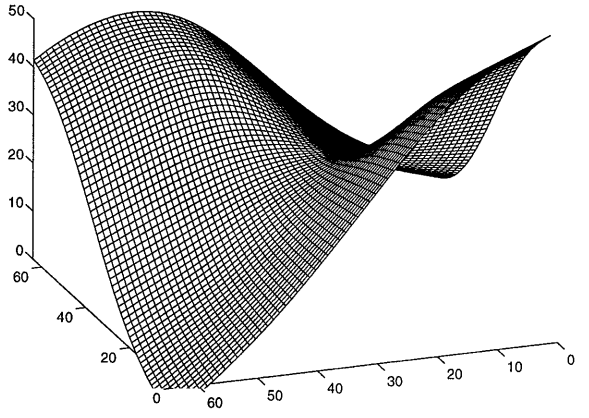
The preceding example demonstrated that our algorithm provides competitive solutions for problems in which we have either dense or (randomly sampled) sparse data. In this section we use a second sparse data example to illustrate several additional issues. Figure 7-6(a) shows the true surface we wish to estimate from the sparse set of four measurements in Figure 7-6(b). Once again, the true surface does not exactly match either the variational or multiscale priors. Also, in our reconstructions we



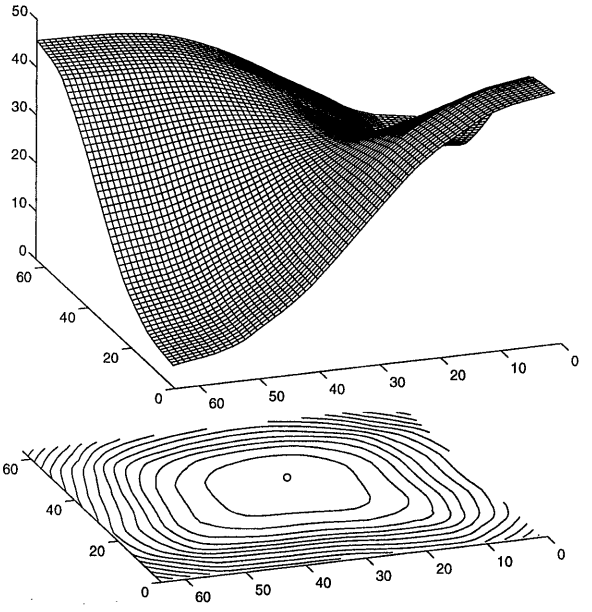
(a) Gauss-Seidel reconstruction.



(b) Conjugate-Gradient reconstruction.



(c) Multigrid reconstruction.



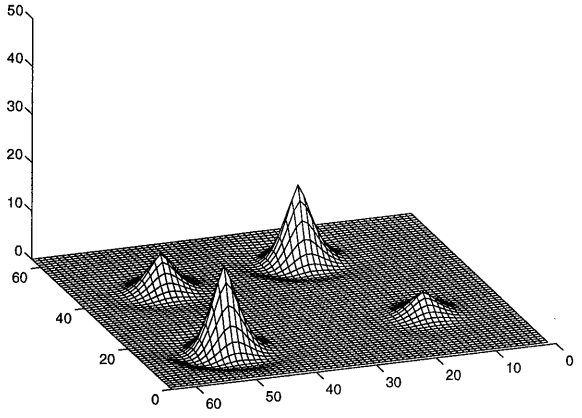
(d) Multiscale reconstruction.

Figure 7-7: Sparse measurement reconstruction examples: each reconstruction is based on the measurements of Figure 7-6. Each of the methods uses the same number of total computations, except that the multiscale approach provides *both* surface estimates and error statistics. The contours of the error variance surface are shown in (d); the minimum of the error surface is marked with an 'o'. $B_s = 80, B_g = 0.05, \mathcal{O} = (38, 29, 24, 0, 0, 0, 0, 0)$.

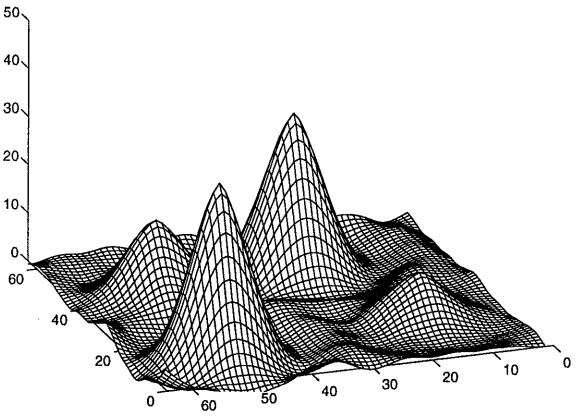
continue to use thin-plate variational priors as this model is the more challenging one for our multiscale algorithm: it asserts a high degree of smoothness – an attribute often difficult to achieve with multiscale algorithms.

Figure 7-7, (a,b,c,d) show the surfaces reconstructed from the measurements of Figure 7-6(b) by means of four different approaches: Gauss-Seidel[19], conjugate-gradient[19], multigrid, and multiscale. The Gauss-Seidel and conjugate-gradient approaches are generally not practical algorithms for the surface reconstruction problems of interest, however they are well understood and many researchers have a sense for the performance of these algorithms. Meaningful comparisons are achieved between these different algorithms by examining the surface estimated after a common amount of computational effort – the effort to produce estimates and error variances on a multiscale tree having nine scales (about 35 seconds on a Sun SPARC-10). Given the extreme sparsity of the measurements, a detailed comparison of reconstructions and truth is not particularly meaningful. Nevertheless from Figure 7-7 we can immediately see the problems associated with the Gauss-Seidel and conjugate-gradient methods, as neither is near to convergence. In contrast, both the multigrid and multiscale algorithms yield smooth estimates of arguably equal quality. In addition, the multiscale algorithm computes estimation error variances, as shown in Figure7-7(d), which provide useful information regarding the accuracy that can be expected from the reconstruction. Furthermore the multiscale algorithm in fact produces surface estimates, gradient estimates, and error variances on *all* scales, much more than just the finest scale results shown here.

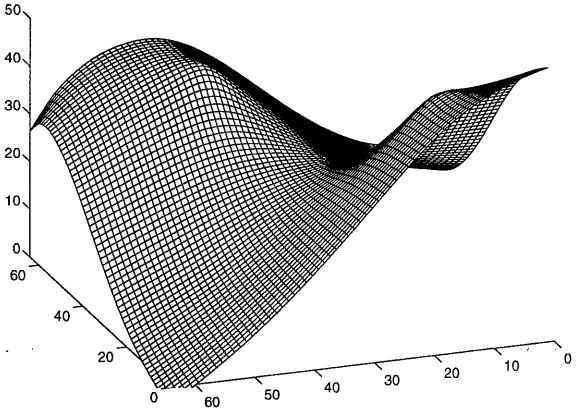
The tradeoff between the degree of smoothness in the estimates and the level of computational effort may be explored by changing the number of levels in the multiscale tree: Figure 7-8 shows a set of four surface reconstructions paralleling the approaches of Figure 7-7, but with one fourth the computational effort (the effort to produce estimates and error variances on a multiscale tree with eight scales). The multiscale-reconstructed surface of Figure 7-8(d) possesses discontinuities in its gradients, however it is able to capture certain aspects of the true surface better than its multigrid competitor of Figure 7-8(c) (compare in particular the upper-



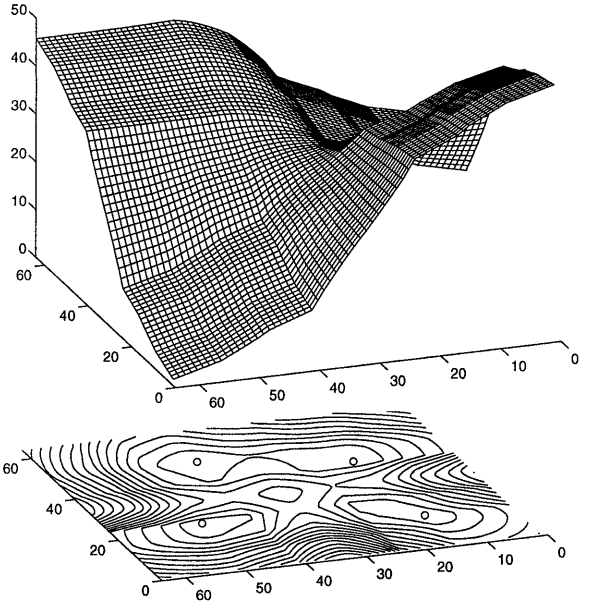
(a) Gauss-Seidel reconstruction.



(b) Conjugate-Gradient reconstruction.

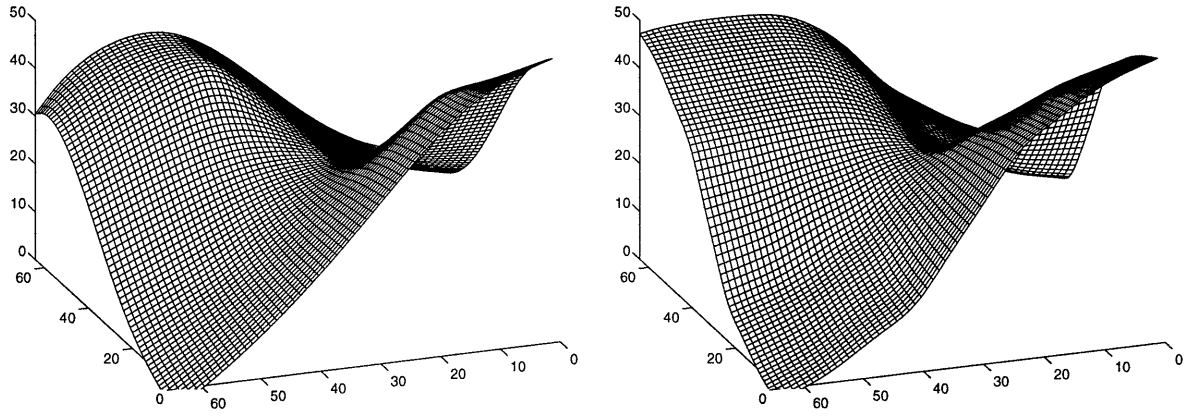


(c) Multigrid.



(d) Multiscale.

Figure 7-8: Sparse measurement reconstructions using the same four methods of Figure 7-7, but using one fourth of the computational effort in each case. Each minimum of the error surface in (d) is marked by an 'o'. $B_s = 80, B_g = 0.05, \mathcal{O} = (16, 10, 7, 0, 0, 0, 0)$.



(a) Multigrid (b) Multiscale + Multigrid.

Figure 7-9: Surface reconstructions computed by applying one multigrid iteration to the two surfaces Figure 7-8(c),(d). $B_s = 80, B_g = 0.05, \mathcal{O} = (16, 10, 7, 0, 0, 0, 0)$.

left portion of the two estimated surfaces); the Gauss-Seidel and conjugate-gradient reconstructions in Figure 7-8(a),(b) are far from convergence. The ability of the multiscale approach to capture certain elements of the surface with relatively little computational effort motivates its use as a highly sophisticated preconditioner for the solution of the Euler-Lagrange PDE. The reconstructions in Figure 7-9 are computed by the application of one multigrid iteration to each of the two surfaces in Figures 7-8 (c),(d).

Figure 7-10 presents a more quantitative comparison of preconditioning. Figure 7-11 depicts the exact thin-plate solution to the surface reconstruction problem based on the four measurements of Figure 7-6; i.e., it is to this solution that the multigrid algorithm will converge. Let $\text{rms}_{\text{MG}}(i)$ represent the rms difference between the thin-plate solution of Figure 7-11 and the reconstruction achieved after i multigrid iterations. Let $\text{rms}_{\text{MS}}(i)$ represent the rms error of the surface using multigrid, preconditioned by the multiscale estimates from a seven-level tree, such that the total computational effort is the same as that of i multigrid cycles. Figure 7-10 plots

$$\left(1 - \frac{\text{rms}_{\text{MS}}(i)}{\text{rms}_{\text{MG}}(i)}\right) \cdot 100\% \tag{7.39}$$

that is, the percentage rms improvement brought about by multiscale preconditioning.

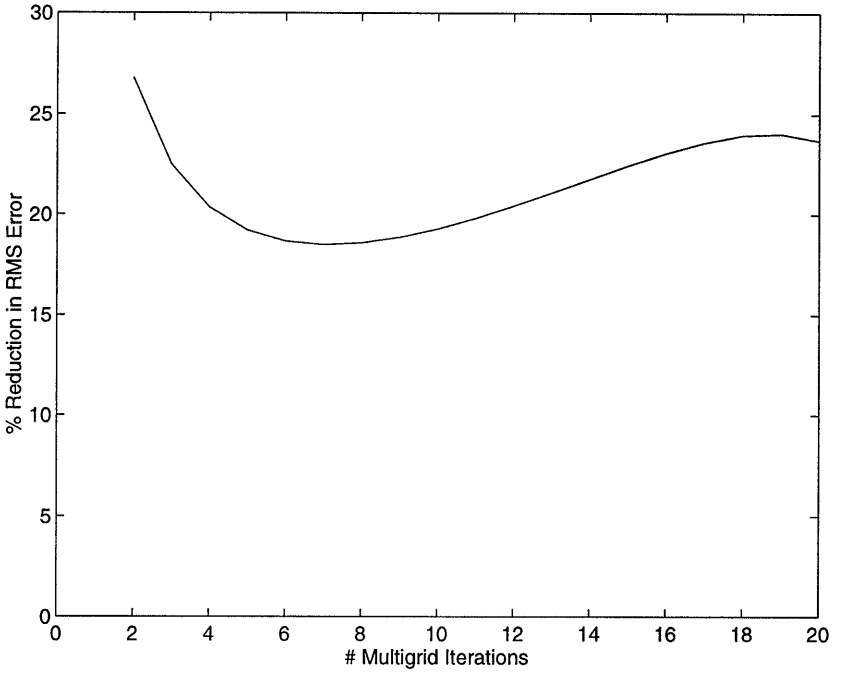


Figure 7-10: Percent reduction in RMS error from just using i iterations of multigrid, to preconditioning the surface reconstruction problem using the multiscale algorithm followed by multigrid iterations such that the total computation is equal to that of i multigrid iterations. The horizontal axis measures i.

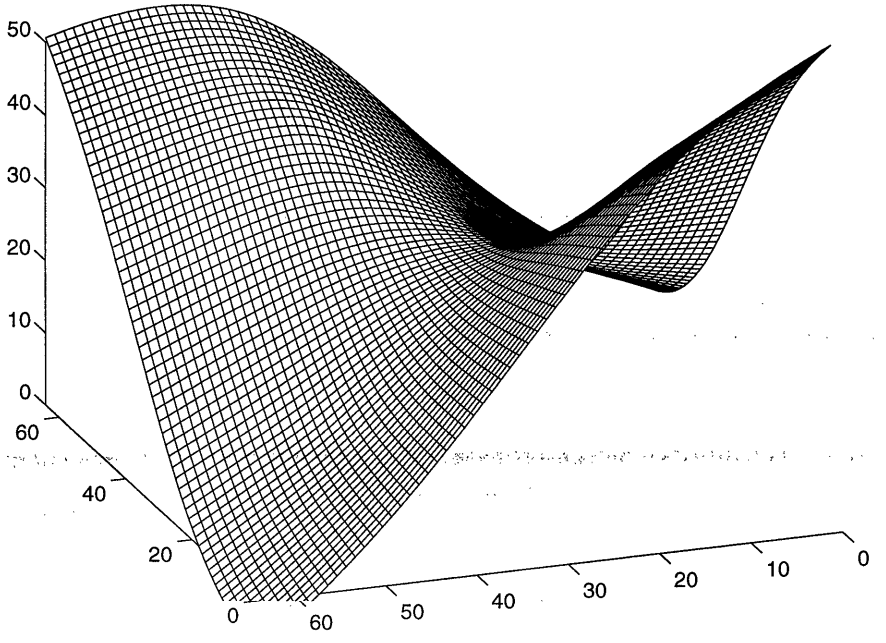


Figure 7-11: The exact thin-plate reconstruction based on the measurements in Figure 7-7 (b).

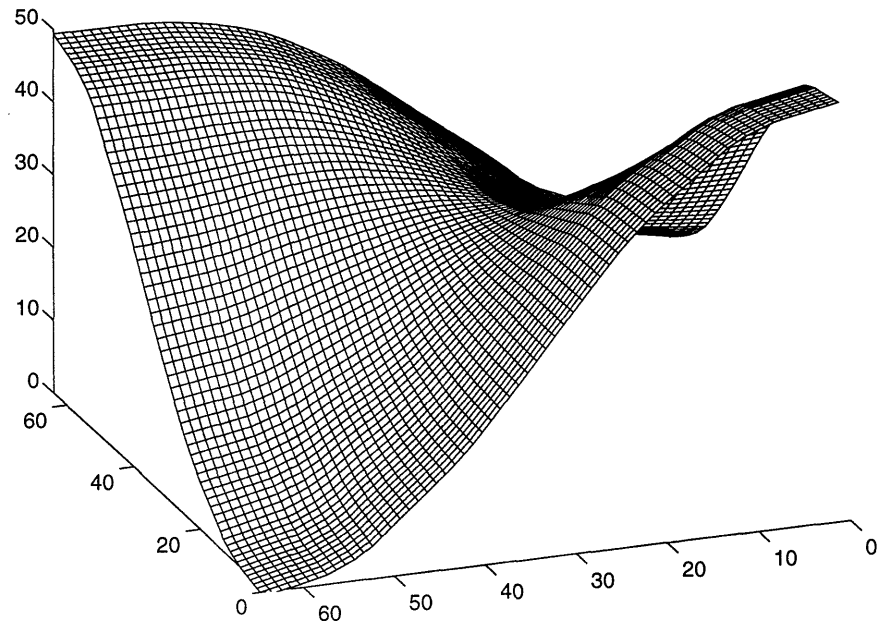


Figure 7-12: The multiscale reconstruction from the height measurements in Figure 7-7 (b), plus regularly sampled measurements (4% measurement density) of the surface gradients.

Note that this percentage is significant, averaging more than 20%.

7.4.3 Surface Reconstruction with Sparse Surface and Gradient Measurements

The examples of the preceding two sections have exclusively considered surface reconstruction based upon direct measurements of the surface height. However our multiscale model readily lends itself to measurements of both height and gradient at the finest scale as well as at coarser resolutions via trivial modifications of (7.36). Figure 7-12 presents a single example to demonstrate this potential. The figure shows the estimated surface, given the four height measurements of Figure 7-7(b), and given regularly but sparsely sampled (one sample per 5x5 block of pixels) gradients of the true surface in Figure 7-7(a).

7.4.4 Surfaces with Discontinuities

A topic of particular interest in surface reconstruction involves the estimation of surfaces possessing known discontinuities, or the estimation of unknown discontinuities. The overlapping multiscale tree model possesses a number of attributes which make it appropriate for such tasks:

- Unlike FFT-accelerated PDE methods, which require space-invariant surface models, the performance of the multiscale approach is unaffected by space-varying models (e.g., a piecewise thin-plate model broken by discontinuities).
- The smooth projection operator H_x of the overlapping tree does *not* blur the surface estimates spatially. As a result, the overlapping model is quite capable of capturing abrupt changes such as discontinuities.
- Because the multiscale estimator takes as input a statistical model for the unknown surface, not only the position of the discontinuities, but also the *statistics* of the discontinuity height may be specified.
- Moreover, since our multiscale algorithms directly produce error variance statistics, we are able to detect statistically significant anomalies in the observed data which in turn allows us to detect and locate discontinuities.

The example surface which we use in this section is shown in Figure 7-13. The 64x64 surface has four step discontinuities of height 10. The step edges are oriented diagonally so as to avoid a convenient alignment with the multiscale tree boundaries. The surface was measured by randomly sampling 30% of the surface elements and adding unit-variance Gaussian noise.

We first consider the situation in which we know the location of the discontinuities. We do not model each discontinuity line as a step of constant height, rather we assume that we have much less prior information and simply model each point along the discontinuity with independent zero-mean random variables. Specifically, in the multiscale context we propose to model the discontinuities as follows: every time a branch of the multiscale tree crosses a discontinuity we will add $\eta = 100$ to the process

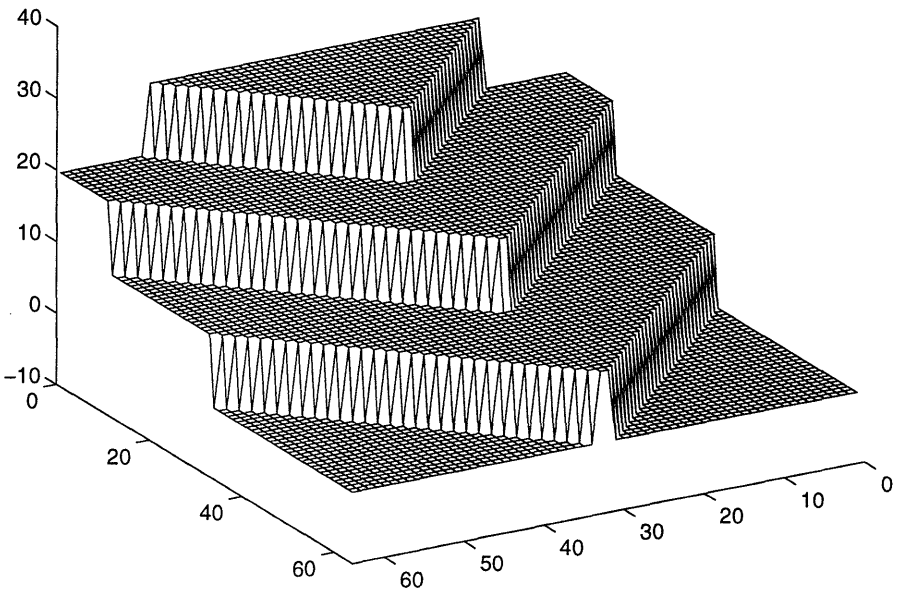


Figure 7-13: Example discontinuous surface.

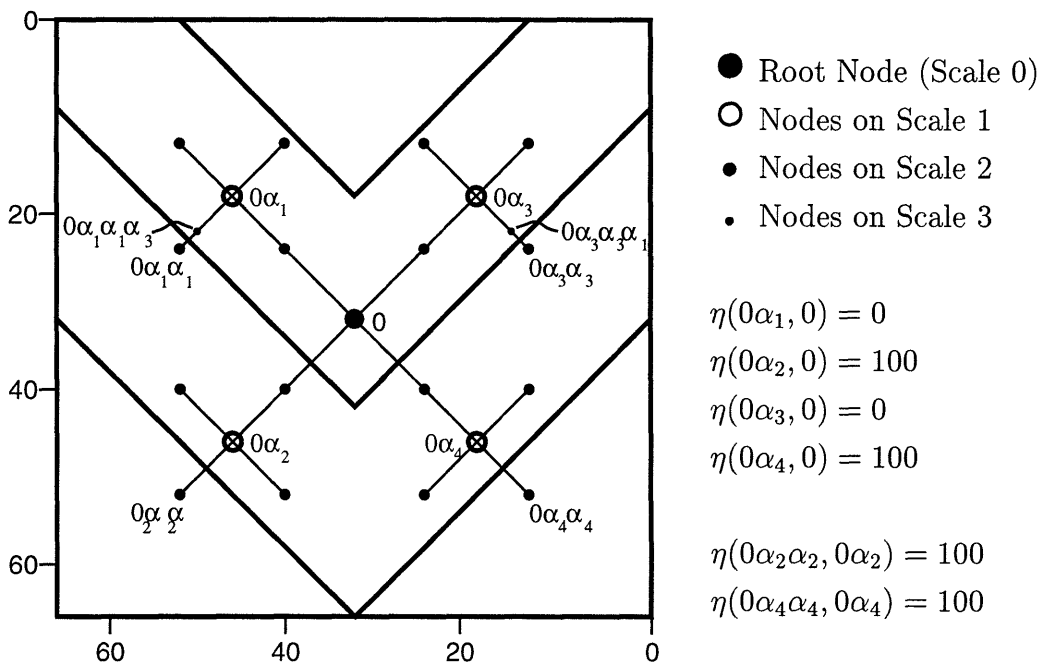


Figure 7-14: This figure demonstrates the behavior of the function $\eta(s,t)$, which measures the increase in the variance of the surface due to discontinuity crossings. The position of the circles $\bullet, \circ$ represent the coordinates $(c_x(s), c_y(s))$ associated with the labeled multiscale nodes. The thick lines represent the locations of discontinuities, consistent with Figure 7-13.

noise variance for $z(s)$ of that branch; Figure 7-14 illustrates this procedure for the example of Figure 7-13:

- Let $(c_x(s), c_y(s))$ be the coordinate of the center of the region represented by multiscale node s; Figure 7-14 shows these coordinates for the coarsest three tree scales (plus two fourth-scale nodes).
- Let $l(s, s\bar{\gamma})$ be the line segment from $(c_x(s), c_y(s))$ to $(c_x(s\bar{\gamma}), c_y(s\bar{\gamma}))$.
- Let $\eta(s, s\bar{\gamma})$ be the sum of the variances of the discontinuities crossed by $l(s, s\bar{\gamma})$. That is $\eta(s, s\bar{\gamma})$ represents the variance of $(z(s) - z(s\bar{\gamma}))$ that can be attributed to the presence of the modeled discontinuities, thus $\eta(s, s\bar{\gamma})$ represents the amount by which the process noise variance for $z(s)$ will be increased at node s.

For example, the line segment $l(0\alpha_4, 0)$ in Figure 7-14 crosses one discontinuity, hence $\eta(0\alpha_4, 0) = 1 \cdot 100$; five other such examples of $\eta()$ are listed in Figure 7-14. Note that it is certainly possible for a given point of discontinuity to be crossed more than once in traversing the multiscale tree, i.e., in following a path from coarse to finer nodes. While one can certainly imagine adding differing amounts of uncertainty at each of these stages, we have used the simple procedure here of adding a variance of 100 at each such crossing. For example, the tree branches from node $0\alpha_3$ to $0\alpha_3\alpha_3\alpha_1$ in Figure 7-14 cross the same discontinuity twice, however the full variance ($\eta = 100$) will be added each time: $\eta(0\alpha_3\alpha_3, 0\alpha_3) = 100$ *and* $\eta(0\alpha_3\alpha_3\alpha_1, 0\alpha_3\alpha_3) = 100$.

We require a modified model at each node s for which $\eta(s, s\bar{\gamma}) > 0$; that is, for those nodes s where a discontinuity lies between s and its parent $s\bar{\gamma}$. When $\eta(s, s\bar{\gamma}) > 0$, (7.35),(7.36) must be modified to reflect the increased variance of the surface and the loss (or, more precisely, the irrelevance) of gradient information across the discontinuity:

$$\begin{bmatrix} z \\ p \\ q \\ zp \end{bmatrix}(s) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} z \\ p \\ q \\ zp \end{bmatrix}(s\bar{\gamma}) + \begin{bmatrix} \left(B_s^2 2^{-m(s)} + \eta(s, s\bar{\gamma})\right)^{\frac{1}{2}} & 0 & 0 \\ 0 & (P_{og})^{\frac{1}{2}} & 0 \\ 0 & 0 & (P_{og})^{\frac{1}{2}} \end{bmatrix} w(s) \tag{7.40}$$

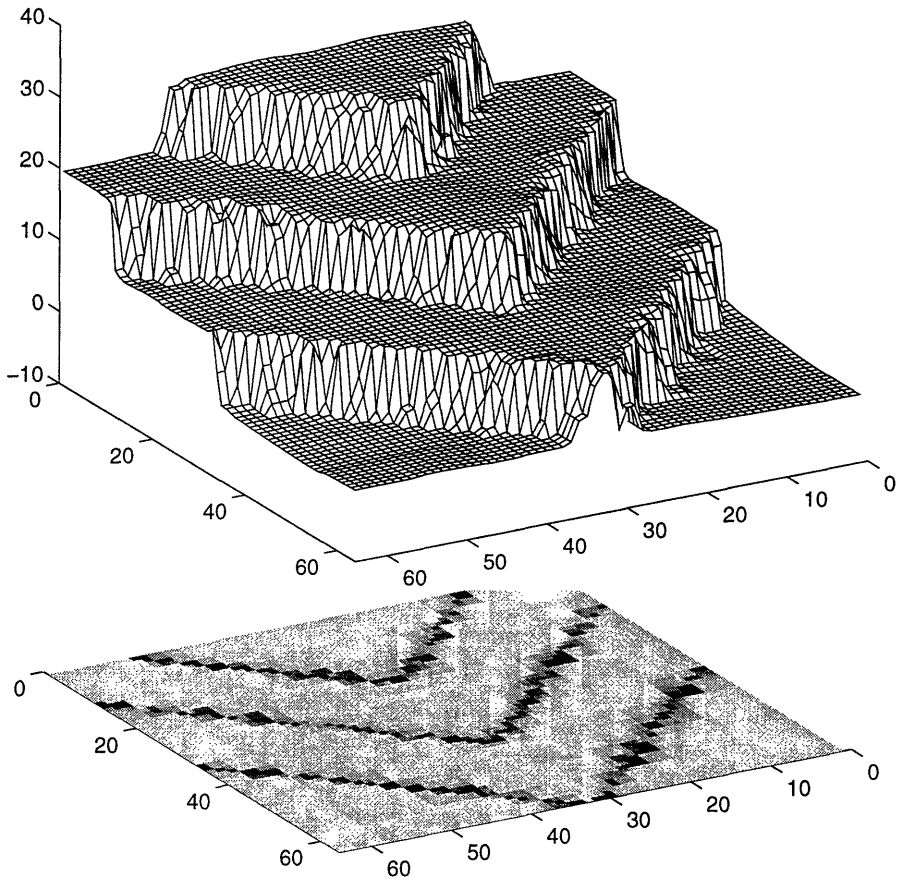


Figure 7-15: Reconstruction of the surface of Figure 7-13 and associated estimation error variances based on a knowledge of discontinuity locations and a sampling, at random, of one third of the surface pixels having unit variance Gaussian noise added. The lower half of the figure plots the estimation error variances; darker regions represent greater uncertainty. $B_s = 20, B_g = 0.2, \mathcal{O} = (26, 19, 14, 9, 0, 0, 0, 0)$.

$$d(s) = \left[C_1(s_i) \;\; 0 \;\; 0 \;\; 0 \right] \left[z \;\; p \;\; q \;\; zp \right]^T (s) + v(s) \tag{7.41}$$

where P_{og} represents the gradient prior variance at the root node. Figure 7-15 shows the estimated surface and estimation error variances, based on this model, applied to an overlapping tree having nine scales. The presence of the discontinuities is clear and well-preserved.

We now look at the case in which we have no prior information about the presence or location of discontinuities, and thus use the same multiscale prior surface model (7.35) as that use in the previous sections without any modifications. Figure 7-16

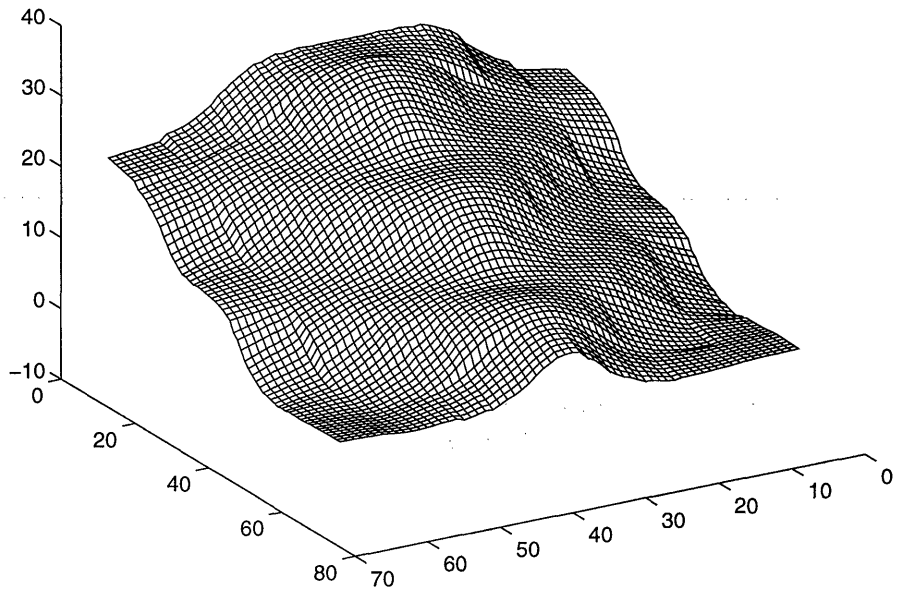


Figure 7-16: Reconstruction of the surface of Figure 7-13 based on the same surface measurements as in Figure 7-15 but using a prior model possessing no discontinuity information. $B_s = 20, B_g = 0.2, \mathcal{O} = (26, 19, 14, 9, 0, 0, 0, 0)$.

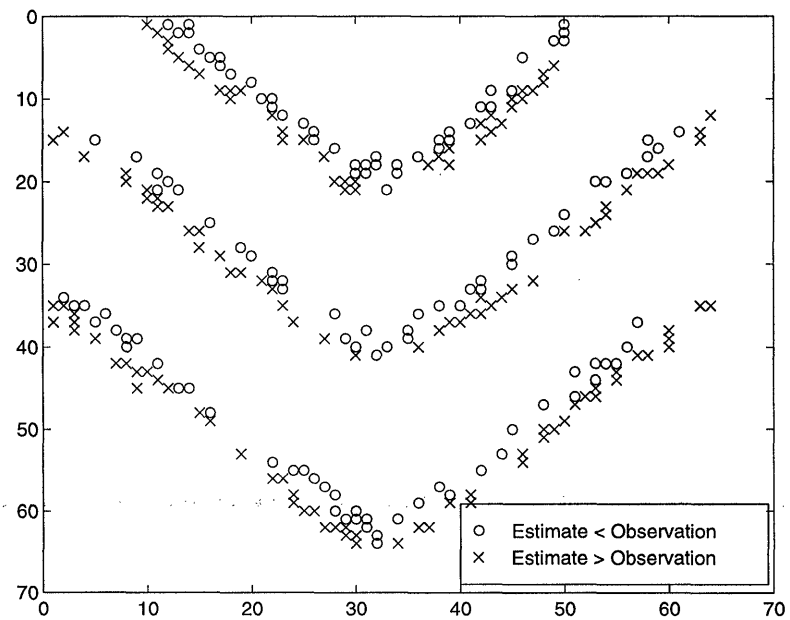


Figure 7-17: Distribution of those measurement residuals, in excess of 3σ, corresponding to the estimates of Figure 7-16.

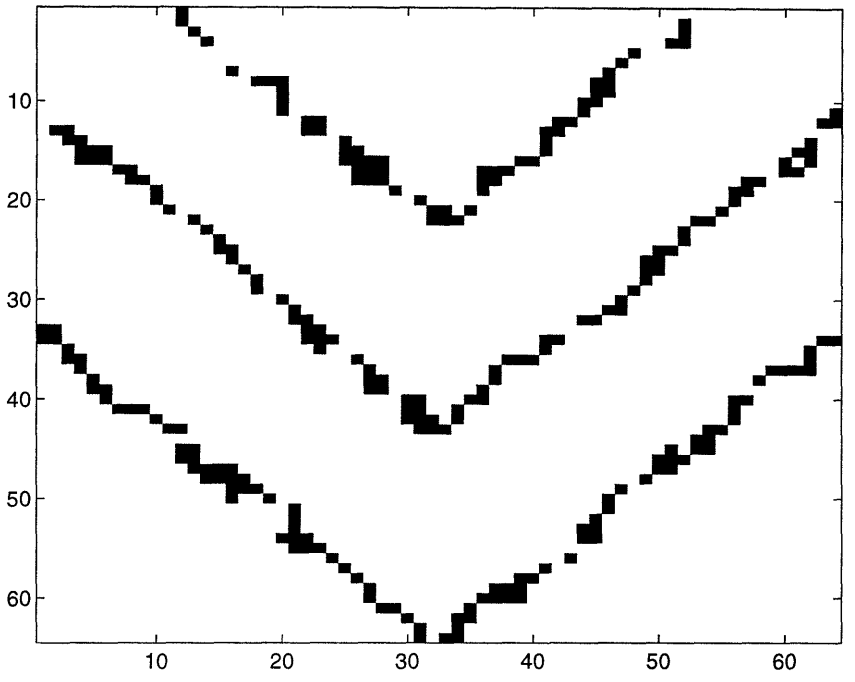


Figure 7-18: A simple estimation of surface discontinuities based on the measurement residuals of Figure 7-17.

shows the estimated surface based on the same set of measurements as was used in Figure 7-15. That the estimated surface looks like a smoothed version of Figure 7-13 comes as no surprise. However, by examining the measurement residuals (residual = measurement - estimate) one can formulate a procedure for estimating the discontinuity locations. Specifically, since estimation error variance statistics are available, we can determine a set of statistically significant residuals (e.g., those in excess of three standard deviations). Aggregations of such residuals outline regions in which the prior model may be inappropriate; for example, a thin-plate model is inappropriate in the vicinity of a surface discontinuity. Figure 7-17 shows the location of all residuals with magnitudes in excess of three times the expected standard deviations; the implied location of the discontinuities is obvious. As a simple example, Figure 7-18 estimates the discontinuity locations as the locus of points whose nearest two residuals differ by least 5. Coupling Figure 7-18 with prior information of discontinuities (e.g., piecewise straight lines) would lead to a set of discontinuity estimates, which could in principle be incorporated back into a prior model such as (7.40) to give a result such as in Figure 7-15.

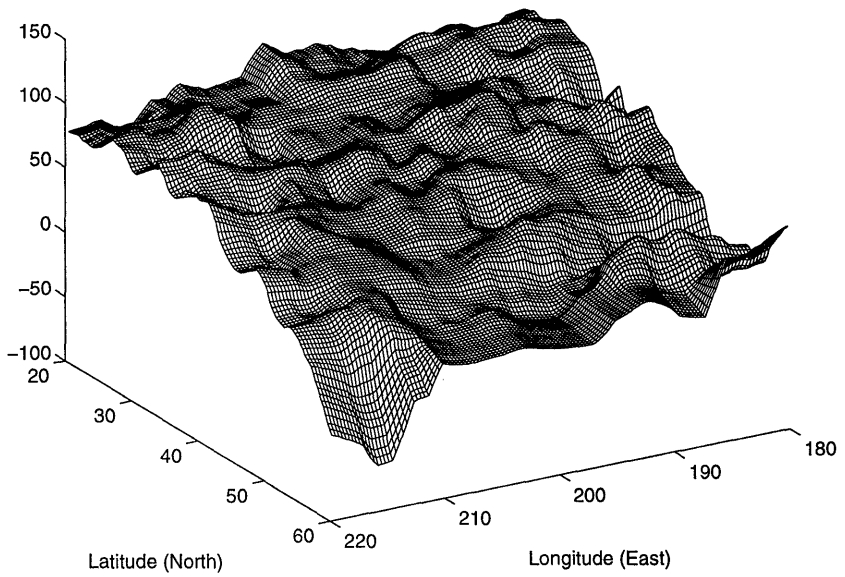


Figure 7-19: A plot of the estimated ocean height, viewed from the north-east. $B_s =$ 35cm, $B_g = 0.5$. $\mathcal{O} = (16, 10, 5, 3, 1, 1, 0, 0)$.

7.4.5 Non-Variational Priors

While the multiscale surface reconstruction models used to this point are motivated by a certain variational thin-plate/membrane model (7.15), the results in the preceding sections demonstrate that our multiscale prior stands on its own as an equally valid prior model to those used in variational formulations. Moreover, one of the strengths of the multiscale framework is its flexibility. In particular, there are many surface statistical models which do not correspond to a thin-plate/membrane prior model, and for which a variational optimization expression may be difficult to write and much more difficult to solve. Due to the flexibility of our framework, many of these surface models may be readily realized in a multiscale setting.

One such example is the class of $1/f^{\mu}$ prior models for $\mu \neq 2$ (and possibly non-integer). Consider, for example, our remote-sensing problem of interest from Chapter 3. The oceanographic estimation problem is of interest for two reasons:

- The prior model for the ocean is based on a $1/f^{\mu}$-like prior, where $\mu \approx 2$ (recall that in Chapter 4, values for μ were identified using multiscale likelihood

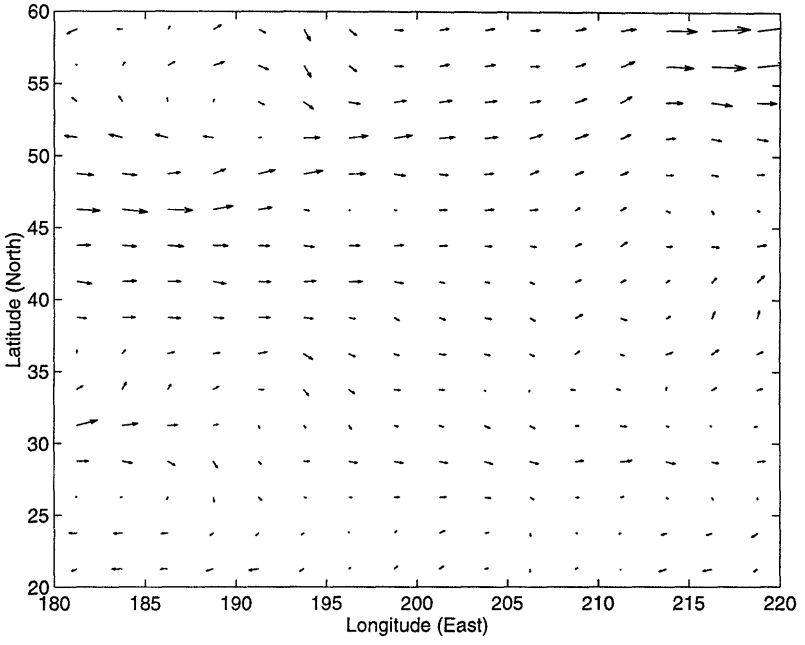


Figure 7-20: The circulation field implied by the surface estimates of Figure 7-19.

techniques, and these values did not exactly equal 2).

- There is significant interest in being able to estimate ocean surface gradients (and hence current flow) directly. Using the joint surface-gradient estimation models of this chapter would allow gradients estimates to be computed on all scales. In particular, the direction of the ocean circulation at any point is (due to Coriolis effects) at right angles to the gradient of the ocean surface.

Recall the oceanographic prior model from Chapter 3:

$$x(s) = x(s\bar{\gamma}) + 35\text{cm} \cdot 2^{m(s)(1-\mu)/2} \tag{7.42}$$

where $\mu \approx 2$. The above oceanographic prior model is not incompatible with the models explored in this chapter; by setting $B_s = 35$cm and by selecting a reasonable scale-to-scale variance for the gradient, e.g., $B_g = 0.5$, we can use the model (7.35) to jointly estimate the surface height and gradients from altimetric data. Figure 7-19 shows a set of ocean height estimates for the north-east Pacific; the circulation pattern, inferred from the estimated gradient field, is shown in Figure 7-20. The

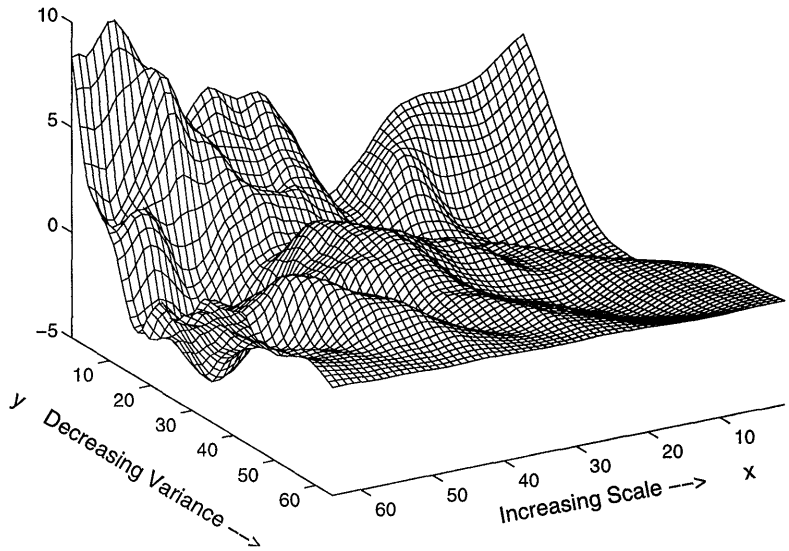


Figure 7-21: An instance of a zero-mean surface, with spatially dependent variance and feature scale.

results of Figure 7-20 are reasonable: the dominant circulation in this area of the Pacific is the eastward flow of the Kuroshio[97]. Deviations from a uniform flow are to be expected due to interference from land and due to eddies.

Another example which illustrates the flexibility of multiscale surface reconstruction is shown in Figure 7-21. The figure shows a zero-mean surface $z(x, y)$, where the surface variance is a decreasing function of y:

$$\text{var}\,(z(x,y)) = 10\frac{64-y}{64} \tag{7.43}$$

and the scale length (i.e., the size of surface features) decreases with x:

$$\text{Feature Size}\,(z(x,y)) = 40 - 0.5x \quad (pixels) \tag{7.44}$$

A multiscale model capable of approximating such a prior is

$$z(s) = z(s\bar{\gamma}) + B(s)w(s) \tag{7.45}$$

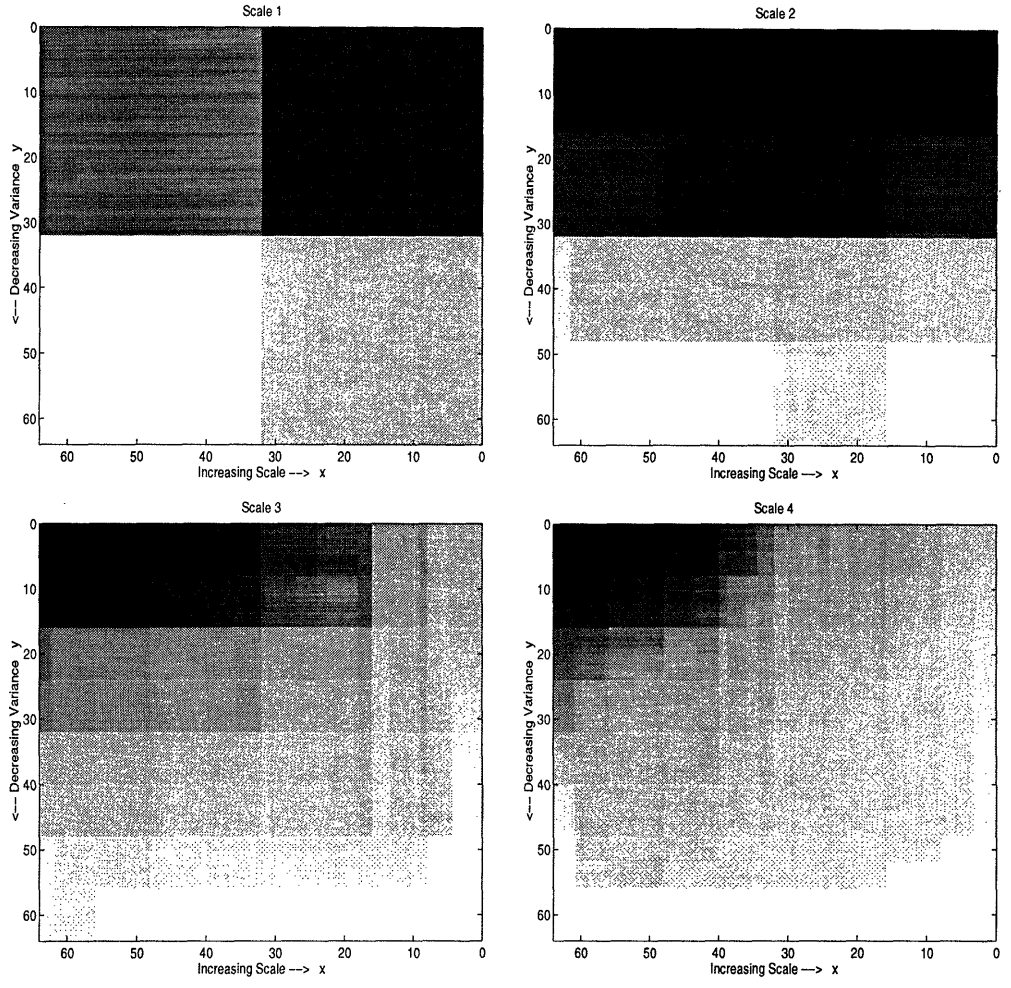


Figure 7-22: Each of the four plots shows the variance $B^2(s)$ for one scale. Darker regions represent larger variances. Observe how the variance is concentrated in the upper-right (large-feature portion of surface) at coarse scales, moving to the upper-left (small-feature portion of surface) at finer scales.

The $B(s)$ are chosen to satisfy the variance and scale properties of the surface. In particular, suppose that a finest-scale node s corresponds to position (x, y), then the $B(s)$ should satisfy

$$B^2(s) + B^2(s\bar{\gamma}) + B^2(s\bar{\gamma}^2) + \ldots + B^2(s\bar{\gamma}^{m(s)-1}) \approx 10^{\frac{64-y}{64}} \tag{7.46}$$

$$\arg_m \max B^2(s\bar{\gamma}^m) \approx \log_2(40 - 0.5x) \tag{7.47}$$

Figure 7-22 plots the spatial distribution of the scale-to-scale variance for four tree scales. Observe how the variance is consistently small near the bottom of the surface,

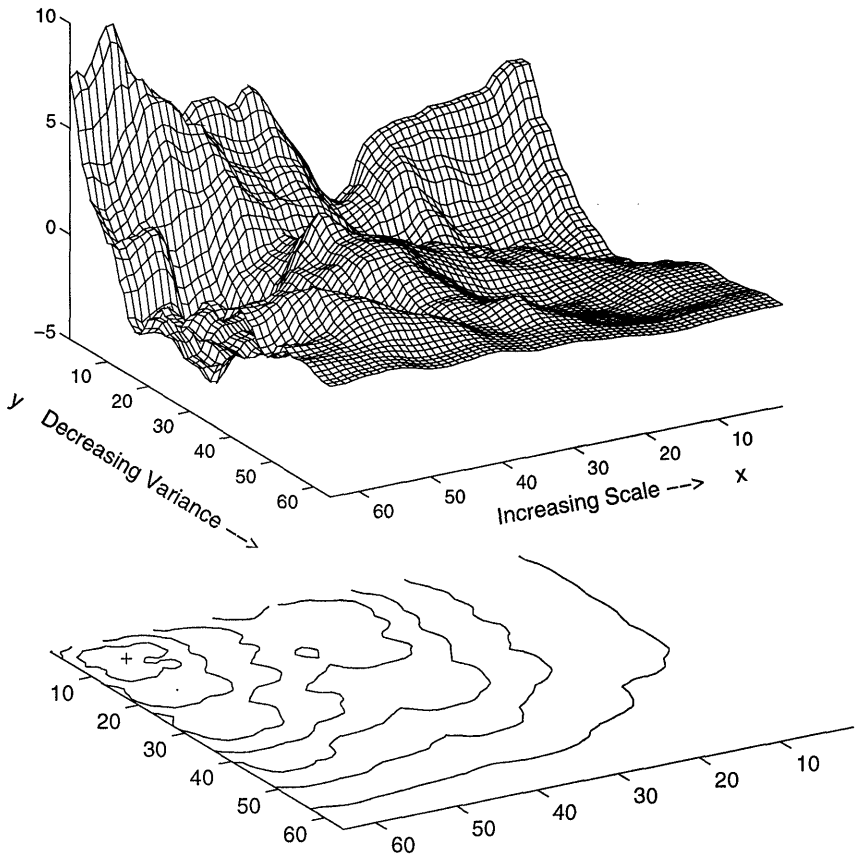


Figure 7-23: The estimation of the surface of Figure 7-21 based on noisy measurements (unit variance). The lower half of the figure plots the corresponding estimation error variances. The peak of the error variance surface is marked with a '+'. $\mathcal{O} = (12, 6, 4, 3, 2, 1, 1, 0)$.

consistent with (7.46), and how the scale-to-scale variance shifts from right to left as we move to finer scales, consistent with (7.47).

Figure 7-23 shows the reconstructed surface, given dense measurements of Figure 7-21 with modest added noise (variance 1). Particularly significant are the estimation error variances, which reflect not only the expected dependence along the y axis, but also exhibits a decrease in estimation error variance as the feature size increases (along x).

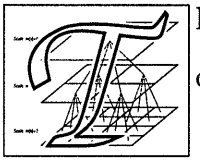
7.5 Conclusions

This chapter has described and illustrated a multiresolution methodology for surface reconstruction. By using the dual interpretation of variational formulations as estimation problems and the relationship between standard variational penalties such as thin-plate and membrane models and fractal priors, we were able to define multiresolution estimation problems that possess very similar interpretations and yield reconstructions of equal or better quality at lower computational burden than efficient iterative methods for solving the variational formulations. Furthermore, with this same computational effort, our multiscale method also produces reconstructions at multiple resolutions as well as estimation error variances, a task extremely difficult to accomplish within a variational setting. We have illustrated the potential value of these error statistics in detecting and localizing surface discontinuities. In addition, we have also shown that our multiscale estimator can serve as an effective preconditioner for the solution of variational formulations of the surface reconstruction problem.

The breadth of uses for the multiscale framework is broader than just as an approximation of certain variational problems however. The variational formulation which motivated the multiscale model developed in this chapter does not represent "truth", rather it is a convenient form of mathematical expression. Similarly, the multiscale formulation offers not only a computationally attractive alternative, but also offers a flexible setting in which to construct surface prior models directly, including many meaningful ones that have no simple variational counterparts.

Chapter 8

Contributions and Future Research



his final chapter reviews the primary contributions of this thesis and outlines possible avenues for continued research.

8.1 Thesis Contributions

Chapters 3 and 5 demonstrated the application of the multiscale framework to very large scale estimation problems of current interest in oceanography and geodesy. For both problems, existing models from the literature and empirical observations motivate the selection of a $1/f$-like multiscale model class.

With this selection of multiscale model, the multiscale estimation algorithm can compute ocean elevation estimates on a 512x512 grid *with* error variances in about 1 minute on a Sparc-10 workstation. This level of computational efficiency makes practical the estimation of the ocean surface over regions as large as ocean basins.

An investigation of the measurement residuals shows them to be strongly correlated to bathymetric (and hence gravimetric) features. This observation motivates the *joint* estimation of the ocean height and the gravitational equipotential (geoid) surface based on the satellite measurements of ocean height. A close look at the existing spherical harmonic model for the geoid and at the statistics of the measure-

ment residuals allows us to estimate the space-dependent variance of the geoid which leads to the desired joint multiscale model and to a computationally efficient joint estimator.

The contribution of these chapters goes beyond the solution of an oceanographic problem of interest using multiscale methods. Although the multiscale framework has been considered a promising tool for the large, irregularly sampled, estimation problems which occur in a variety of scientific disciplines and particularly in remote sensing, research into our multiscale framework has not been disseminated in this scientific community. These chapters are meant to begin this dissemination, and are written specifically with the remote sensing community in mind.

Chapter 4 contributed a deeper understanding of multiscale likelihood and parameter estimation techniques in the form of a multiscale estimator for the fractal dimension of fractal Brownian motion processes, and the determination of a Cramer-Rao bound on the maximum-likelihood estimation of parameters in a commonly used $1/f$-like class of multiscale models.

To estimate the fractal dimension $(2-H)$ of fBm, a multiscale model was chosen which represents an fBm-like process based on a Haar wavelet decomposition. Using the multiscale likelihood function algorithm, an estimator was developed for H and compared with existing estimators in the literature. Our estimator competes favorably in performance, while at the same time offering greater flexibility: the fractal dimension of sparsely sampled processes or of multidimensional processes may also be estimated.

The repeated use of the $1/f$-like multiscale model in Chapters 3, 5, 7 and in [63] motivates the development of a technique to estimate the parameters in the model, and motivates an understanding of the bounds in making such estimates. Such an understanding is provided by the Cramer-Rao bound which is derived for this multiscale model class. The bound is derived under the assumption of the availability of noiseless measurements on all nodes above the finest scale. Despite such an assumption the bound is demonstrated, via Monte-Carlo simulations, to be fairly tight (the variance bounds are at most 40% below their empirical counterparts).

Chapter 6 contributed the method of overlapping trees: a new approach to multiscale estimation. The use of overlapping trees offers a considerable reduction in multiscale artifacts at a modest computational effort. A simple multiscale modeling and estimation example is considered in some detail to motivate and illustrate the overlapping approach. Under certain theoretical conditions, an extended deductive derivation of which is presented in Appendix D, the estimates produced by the overlapping approach can be shown to be optimal. Three projection operators are defined which project measurements and estimates between the original domain of interest and the overlapping domain in which the multiscale tree is constructed. A means for constructing an overlapping tree structure and for determining the associated projection operators is developed. The ability of the overlapping technique to reduce or eliminate artifacts is demonstrated via the multiscale estimation of strongly correlated, anisotropic textures.

Chapter 7 contributed a development of multiscale counterparts to a certain class of variational cost functions commonly used for surface reconstruction, extending the results of [63]. Many computer vision problems are formulated in terms of variational equations which provide a convenient, albeit frequently computationally demanding, approach for estimating images; Chapter 7 presents an alternative approach using our multiscale framework. Whereas the results of previous multiscale approaches may have been less than compelling to the computer vision community due to the presence of artifacts, the coupling of the overlapping framework with a multiscale surface model gives rise to computationally efficient and aesthetically smooth reconstructed surfaces.

8.2 Topics for Future Research

There are many interesting directions for continued research. A few of the more promising directions, organized by topic, are listed in the following subsections.

8.2.1 Ocean Surface Estimation

While perhaps not a research extension *per se*, one useful extension of the oceanographic research of Chapter 3 would be to actually *use* the multiscale algorithm to produce regular estimates of the ocean surface and to make these available to interested researchers. Some fundamental open questions do arise, however, once ocean surface estimation is attempted on a grand scale:

- The north-Pacific subset of the planet may be approximated as being flat, and therefore is simply represented by a quadtree structure. A whole globe, on the other hand, is not usually meaningfully approximated by any flat plane. Thus, once ocean surface estimation is attempted on a global scale, it is unclear what choice of tree structure would be appropriate.

- A system as complex as the earth's oceans is unlikely to be meaningfully characterized by a single model, based on some sort of globally-averaged power spectrum. The space-dependent nature of the earth's ocean should be reflected in a space-varying multiscale model (similar in spirit to the Kuroshio example in Chapter 3). It remains to be determined how such a global model should be parameterized and identified.

8.2.2 Likelihood Methods

The canonical $1/f$-like multiscale model used throughout this thesis,

$$x(s) = x(s\bar{\gamma}) + B2^{m(s)(1-\mu)/2}w(s) \tag{8.1}$$

is parameterized in terms of two values B, μ having intuitive interpretations: B controls the low-frequency "power" in the process, and μ is related to its fractal dimension (i.e., B and μ essentially control the offset and slope of the power spectrum of the finest-scale process of x, plotted in the log-log domain). It is not obvious, however, that this intuitive parameterization is the most appropriate. Recall from Chapter 4 that the error in estimating B is very strongly correlated with the error in estimating

μ; that is, one component in B, μ space (that component *along* the contour lines in Figure 4-2) is estimated very poorly, whereas the perpendicular component is estimated rather well. The contour lines in Figure 4-2 obey

$$B2^{\tau(1-\mu)/2} = \Theta \tag{8.2}$$

very closely (to about one part in 10^4 over a wide range of values of μ) for some particular value of τ. An alternative parameterization of (8.1), for example in terms of (B, Θ), may be more meaningful (and almost certainly better numerically conditioned) for multiscale likelihood purposes than (B, μ).

There is a second, related topic. Suppose we have observations $\bar{y}$ of a physical process $\bar{x}$. Furthermore suppose that the prior statistics of $\bar{x}$ are unknown, but that the periodogram of $\bar{y}$ is observed to have $1/f$-like behavior, and that the measurement noise in $\bar{y}$ is known to be well approximated as Gaussian. Motivated by the observed $1/f$-like behavior we consider a multiscale model of the form

$$x(s) = x(s\bar{\gamma}) + B2^{m(s)(1-\mu)/2} w(s) \tag{8.3}$$

$$y(s) = x(s) + v(s) \tag{8.4}$$

on a multiscale tree having M scales, where the finest-scale measurement process is a subsampled version of $\bar{y}$. Now suppose that we use multiscale likelihood techniques to estimate B and μ in (8.3). We have found certain processes $\bar{x}$ such that the estimated values $\hat{B}, \hat{\mu}$ can be a *strong* function of M (e.g., if we let $\bar{y}$ be sparsely sampled measurements of the geoid, then changing the number of scales from 9 to 10 can induce $\hat{B}$ to change by two orders of magnitude, although the corresponding change in $\hat{\mu}$ is much less — a factor of two or so). The cause of this discrepancy is unknown, however it again suggests that the B, μ parameterization of (8.3) may be inadequate.

8.2.3 Joint Geoid – Ocean Surface Estimation

The development of a joint estimator for the ocean surface and the geoid is only the first of many possible scientific matters to be addressed. Indeed, there are a number of scientific investigations which could follow upon the work of Chapter 5:

- **Independent validation:** a revised geoid estimate N implies a revised set of ocean height measurements Y_ζ and a revised set of estimates $\hat{\zeta}$. Sophisticated global climate models have been developed which can assess the consistency of a set of estimates $\hat{\zeta}$. Such a consistency assessment would provide a solid, scientific, validation of our joint estimates.

- **Likelihood experiments:** many oceanographic statistical quantities, such as the power law of the ocean surface or of the geoid, are still an active area of research. The availability of the efficient multiscale likelihood function calculator may allow us to compute certain statistical parameters of interest.

- **Sensitivity Tests:** the assertion of a prior statistical model for the geoid serves not only to regularize the estimation problem, but also to contribute statistical information about the geoid. It is unclear, however, to what degree the estimated geoid is sensitive to the choice of a particular prior model (e.g., Kaula's law). The computational efficiency of the multiscale estimation procedure would allow us to deduce such sensitivities by observing the dependence of geoid estimates to a variety of prior models.

8.2.4 Overlapping Models

The demonstrated potential of the overlapping framework for attenuating or eliminating artifacts in multiscale estimates motivates further research of a variety of aspects of the framework, four of which are listed below:

1. There is a considerable amount of flexibility inherent in the overlapping framework (i.e., there are a great variety of possible overlapping trees). Two heuristics

were proposed in order to make practical the selection of a particular overlapping tree:

(a) Given an image of K pixels and a multiscale tree having M scales, what is the "best" choice of $\mathcal{O} = \{o_m\}$? The heuristic proposed in Chapter 6 was to set

$$\frac{o_m}{w_m} \simeq \psi = \text{a constant function of scale} \tag{8.5}$$

However in preparing the surface reconstruction examples of Chapter 7, it was found that the coarse scales required proportionately more smoothing (i.e., greater overlap) than the fine scales; for example, improved surface reconstructions were achieved using

$$\frac{o_m}{w_m} > \psi, \text{ small } m \qquad\qquad \frac{o_m}{w_m} < \psi, \text{ large } m \tag{8.6}$$

Is it possible to develop a more general set of rules to determine $\mathcal{O}$?

(b) M and $\mathcal{O}$ uniquely define the projection G_x, however many degrees of freedom remain to be specified in H_x. What is the "best" choice of H_x subject to $H_x G_x^T = I$? The proposed heuristic was to linearly taper the relative weights of H_x across the interval of overlap between two sibling nodes (as in Figure 6-6, Section 6.3). Certainly other kinds of tapering (e.g., quadratic) are possible, however these have not yet been explored; an exploration of such other choices of H_x and their associated smoothness properties is a recommended direction for future research.

2. In the proposition of Chapter 6 (and in the associated proofs in Appendix D) we investigated the conditions to be satisfied in order for an overlapped estimator to produce the optimal estimates. Throughout the discussion we assumed that the desired correlation structure $P_l = G_x P G_x^T$ was realized in the multiscale model. A precise realization is possible with sufficiently high order models, however in the interest of computational efficiency we have always used lower

order models which are unable to precisely realize P_l. A compelling desire to use such low order models leads to two questions:

(a) Can we develop bounds for the deviation of the estimates from optimality $(\hat{x} - \hat{x}_{\text{opt}})$ as a function of the deviation of the realized model Σ_l from its desired form $G_x P G_x^T$?

(b) Given that the realized correlation structure is an approximation to the ideal one, i.e., $\Sigma_l \approx G_x P G_x^T$, what is the optimal choice of R_l — the measurement error covariance in the lifted domain? One possibility is considered here.

Suppose that some element $x(i)$ has a redundant representation of n elements in x_l: $x_l(i_1), \ldots, x_l(i_n)$. Furthermore, suppose that we have a measurement $y = x(i) + v$ with noise variance r; this single measurement is copied into n measurements in the lifted domain:

$$y = x_l(i_p) + v_l(p) \qquad 1 \leq p \leq n \tag{8.7}$$

From Chapter 6, the overlapping measurement noise variances are set using the heuristic

$$\text{cov}(v_l(p)) = r_l(p) = n \cdot r \tag{8.8}$$

It was shown in Chapter 6 that if $\Sigma_l = G_x P G_x^T$, then by using this heuristic one can compute the optimal estimates. The matter is less clear if $\Sigma_l \neq G_x P G_x^T$:

- Should the $r_l(p)$ of (8.8) all be set to the same value?
- If all of the $r_l(p)$ are set to the same value, which value is most appropriate?

In the following we take a preliminary look at the latter question.

The heuristic (8.8) was motivated by the following:

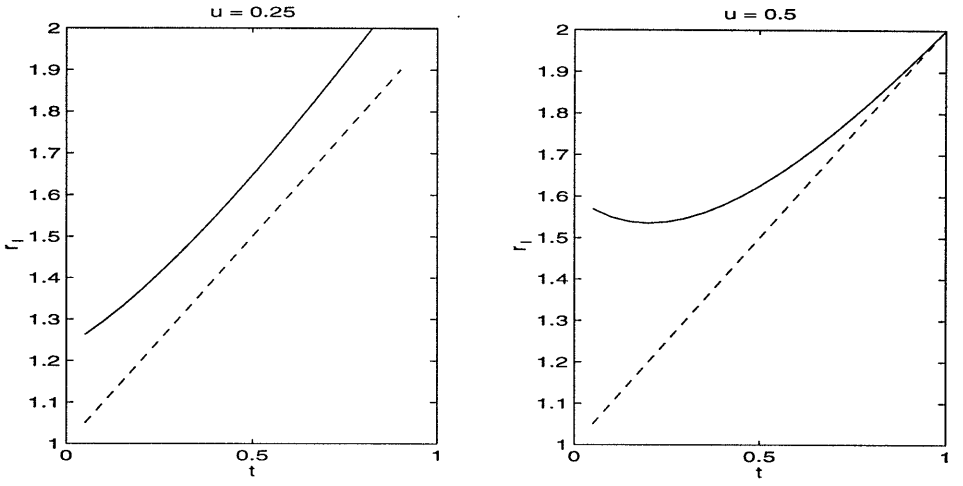


Figure 8-1: The values t, u refer to (8.12). The solid line depicts the optimal value of r_l, in the sense of optimizing (8.14); the dashed line depicts the value of r_l based on the heuristic (8.9). In both cases, r_l refers to the measurement variance associated with each of the doubly-redundant elements in (8.12). The method of Chapter 6 would just have set $r_l = 2$, regardless of u, t.

> A single measurement of noise variance r is the same as n independent measurements, each having noise variance $n \cdot r$, but each measuring the same unknown.

The crucial point is that all of the measurements are measuring the *same* unknown; i.e., $x_l(i_1), \ldots, x_l(i_n)$ are perfectly correlated; i.e., the set $\{x_l(i_1), \ldots, x_l(i_n)\}$ possesses one effective degree of freedom.

Now suppose a correlation structure $\Sigma_l = \text{cov}(x_l)$ is realized. We propose the following heuristic for r_l:

$$r_l = n \cdot r/\text{effective degrees of freedom in}\{x_l(i_1), \ldots, x_l(i_n)\} \quad (8.9)$$

$$= n \cdot r/d_{eff} \quad (8.10)$$

We can test this proposition on the example of Section 6.2:

$$P_l = \begin{bmatrix} 1 & 0.5 & 0.5 & 0 \\ 0.5 & 1 & 1 & 0.5 \\ 0.5 & 1 & 1 & 0.5 \\ 0 & 0.5 & 0.5 & 1 \end{bmatrix} \qquad G_y \equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad r = 1 \quad (8.11)$$

where P_l represents the desired statistics for x_l in the lifted domain. Suppose that the following statistics are actually realized for x_l:

$$\Sigma_l = \begin{bmatrix} 1 & 0.5 & u & 0 \\ 0.5 & 1 & t & u \\ u & t & 1 & 0.5 \\ 0 & u & 0.5 & 1 \end{bmatrix} \tag{8.12}$$

The covariance matrix of the redundant elements and the associated effective degrees of freedom are given by

$$\Lambda = \begin{bmatrix} 1 & t \\ t & 1 \end{bmatrix} \qquad d_{eff} = \sum_i \sum_j \Lambda^{-1}_{i,j} = \frac{2}{1+t} \tag{8.13}$$

Let x_{opt} represent the optimal least-squares estimate of x given y. Let $\hat{x}_l$ represent the estimate produced by the multiscale estimator based on prior model Σ_l and measurements $y_l = G_y y$. Then Figure 8-1 compares the optimal value of r_l, the value of r_l that minimizes the mean square error

$$(x_{opt} - H_x \hat{x}_l)^T (x_{opt} - H_x \hat{x}_l) \tag{8.14}$$

with the value of r_l based on our heuristic (8.9). Our prediction is clearly not exact, however it captures much of the behavior of the optimal solution; in any event our prediction is much better than just setting $r_l = 2$, the procedure recommended by Chapter 6.

3. The availability of a technique to attenuate artifacts and to produce *smooth* estimates permits us to consider applications that *rely* upon such smoothness. For example, Luettgen[62] considered solving the relaxation problem

$$Tx = y \tag{8.15}$$

that is, solving for x given T and y; T incorporates smoothness constraints (i.e., T possesses a differential-like nature). T is a very large matrix, so solving

$$x = T^{-1}y \tag{8.16}$$

is impractical. However recall that the multiscale estimator can efficiently solve

$$x = T_{ms}^{-1}y \tag{8.17}$$

for special T_{ms} which possesses a certain multiscale structure. Luettgen considered finding a multiscale model T_{ms} similar to T and iterating

$$\hat{x}_k = \hat{x}_{k-1} - T_{\text{ms}}^{-1}\left(T\hat{x}_{k-1} - y\right) \tag{8.18}$$

If the eigenvalues of $(I - T_{ms}^{-1}T)$ all lie inside the unit circle then the iteration (8.18) is guaranteed to converge. However the iteration was found to be unstable: the overlapping framework had not been developed, so $T_{ms}^{-1}y$ possessed certain artifacts (similar to those of Figure 7-3). These artifacts, essentially discontinuities, were amplified in the next iteration by the differential nature of T, leading to even greater artifacts etc. — the iteration is unstable. It would be interesting to return to this problem, now using an overlapping T_{ms}, perhaps investigating the stability of the iteration (8.18) as a function of the smoothness of T_{ms} (i.e., as a function of the number of overlapping scales M).

4. Finally, the overlapping framework proposed in this thesis can be viewed as a special case of a much broader class of modeling and estimation problems solved by multiscale means in some projected domain. For example, consider the generalized projection operators

$$x_l = \mathcal{G}_x(x) \qquad y_l = \mathcal{G}_y(y) \qquad \hat{x} = \mathcal{H}_x(\hat{x}_l) \tag{8.19}$$

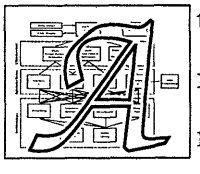
Research into the multiscale framework before the work of this thesis essentially assumed $\mathcal{G}_x = \mathcal{H}_x = I$, $\mathcal{G}_y = I$. Similarly Chapter 6 of this thesis develops the special case that $\mathcal{G}_x$ is one-to-many, $\mathcal{G}_y$ is one-to-many and projects onto the finest scale only, and that $\mathcal{H}_x$ is many-to-one. The development of more general projection schemes, and an analysis of their properties, may lead to more powerful and more insightful multiscale implementations.

8.2.5 Surface Reconstruction

The success of the development of a multiscale counterpart to variational models for surface reconstruction motivates a search for other common variational models for which suitable multiscale representations can be found. To be sure, the results of Chapter 7 apply to a broader class of variational problems than just those associated with the surface reconstruction problem. On the other hand, only those variational problems made up of a sum of quadratic penalties on linear functionals of measurement and state elements have linear, Gaussian statistical counterparts, so the extension of Chapter 7 to a broader class of variational problems will be challenging.

Appendix A

General Multiscale Engine



t an early stage in this thesis, the need developed for a software implementation of the multiscale estimation and likelihood calculation algorithms. These algorithms did not originate in this thesis and have been implemented by others, however an investigation of these implementations showed each of them to be rather specifically targeted to a particular estimation problem.

At that point in my thesis, no particular assumptions had yet been made regarding the trees and models which might be of ultimate interest, and so the development of such targeted code was inappropriate. On the other hand, estimation code having no built-in assumptions tends to execute very slowly and is complicated to use, since the tree and its model need to be specified in arduous detail. What I developed instead was a *completely* general piece of code which could, in a sense, adapt itself to certain assumptions at compile-time, thus allowing the program to remain computationally efficient and competitive with more targeted implementations. This philosophy has proven to be extremely effective, in that a *single* piece of multiscale code, written near the beginning of this research, has been capable of being applied to all of the estimation problems of this thesis — from simple $1/f$-like scalar models to overlapped canonical-correlations based Markov random field models.

The purpose of this appendix, then, is to give a high level view of the multiscale framework implementation (i.e., the philosophy of the previous paragraph). It is hoped that this discussion may make the nature of the implementations in Chapters 3

through 7 a little more concrete.

Figure A-1 presents the highest level view of the software architecture. There are four levels to the hierarchy:

1. Implemented in MATLAB, this level contains commands issued directly by the user. A typical set of commands might be

```
Load(Topex_Poseidon)
Format_Matrices(Topex_Poseidon,Data,Data_Variance)
Multiscale_Estimate(Data,Data_Variance,Estimates,Estimates_Variance)
Contour_Plot(Estimates)
```

   It is significant to note that the overlapping-tree parameters (Chapter 6) are relevant only at this top level. That is, once a multiscale tree with an associated multiscale model and measurements has been defined (based on the overlap parameters as described in Appendix E), pieces of code lower in the hierarchy are just aware of some given multiscale tree; on lower levels of the software hierarchy, no distinction is made between overlapped and non-overlapped trees.

2. Each call from the MATLAB environment (such as `Multiscale_Estimate()` from above) to any of the multiscale algorithms is made via a function written in the "C" language. Direct access of the multiscale algorithms to the MATLAB environment seems inconceivable (or at least impractical) since MATLAB supports only two-dimensional matrices; the greater flexibility of the "C" data types makes them a much more practical choice.

   Each "C" function parameterizes the multiscale tree structure, model, and observations into a small set of parameters which are supplied by the user from MATLAB. A variety of "C" functions have been implemented, supporting a range of tree structures and multiscale models (e.g., scalar vs. vector states, isotropic vs. heterogeneous models).

3. At present four interfaces have been developed, each of which supports differing degrees of multiscale-tree generality. The interfaces are not pieces of code,

rather they are sets of "macros" ("C" language compile-time definitions) which redefine the behavior of the multiscale engine, asserting various assumptions of the multiscale tree as appropriate.

The assumptions inherent in the four interfaces, and an example of the function invocation for the multiscale estimator in each case, are shown in Figures A-2 through A-5.

4. The lowest level of the hierarchy contains a set of tree initialization routines (which may depend on the interface used), a small matrix library, and completely general implementations of the multiscale estimator, likelihood calculator, error cross-covariance calculator, and sample path generator.

   All operations that depend in any way on the multiscale tree structure or model are accomplished via the macros defined in the interfaces. In particular, each node on the tree is parameterized by its scale m, and by its index s within that scale, and *all* quantities (e.g., the A, B, C, R matrices) are indexed with respect to these parameters. For example, if at node m, s on the tree the dynamics matrix at the parent of m, s is required, then the matrix is referenced as

$$A(m-1, \text{Parent}(s)) \tag{A.1}$$

   where the nature of the macro "Parent" will vary from simple to complex, depending upon the simplicity of the tree (and hence upon the selected interface); furthermore, the invocation "A()" is itself a macro. The interpretation by the compiler of (A.1) under the different interfaces should serve to make this point clear:

   - In interface 1 the dynamics matrix is constant over the entire tree; the tree itself is assumed to be two-dimensional and regular. Thus (A.1) would be converted at compile-time into

$$A(m-1, \text{Parent}(s)) \overset{\text{Compiler}}{\Longrightarrow} a \tag{A.2}$$

where a is the actual, constant, matrix passed by the user (see Figure A-2).

- In interface 2, the dynamics matrix is permitted to vary with scale and with the cardinality of each descendant, but not arbitrarily over the tree:

$$A(m-1,\text{Parent}(s)) \overset{\text{Compiler}}{\Longrightarrow} \tag{A.3}$$
$$a[m-1]\left[(s\%dy[m-1]+(\frac{s}{\text{sizey}[m]}\%dx[m-1])*dy[m-1])\right]$$

where % represents the modulus operator, $dx[m], dy[m]$ represent the number of offspring of each node on scale m in the x, y directions respectively, and sizey$[m]$ represents the total number of tree nodes along the y direction on scale m.

- Finally, in interface three, the dynamics matrix is permitted to vary arbitrarily over the entire tree. In fact, as shown in Figure A-4, the matrix is specified via a user supplied function $(*a)$:

$$A(m-1,\text{Parent}(s)) \overset{\text{Compiler}}{\Longrightarrow} \tag{A.4}$$
$$(*a)\left(m-1,\frac{s\%\text{sizey}[m]}{dy[m-1]}+\frac{s/\text{sizey}[m]}{dx[m-1]}\text{sizey}[m-1]\right)$$

That each of the above three possibilities, from trivial to completely general, can be captured by the same code statement attests to the power of this approach. An enormous number of such "macro" substitutions occur throughout the code; the above examples represent by a single macro, however we hope that this discussion has been sufficient to elucidate the nature of such macro substitutions. The significant point, of course, is that all of these interpretations are accomplished at compile-time; the effect of such interface-dependent compilation is to leave us with a conspicuously efficient run-time program under interface 1, and a powerfully general program under interface 3 or 4.

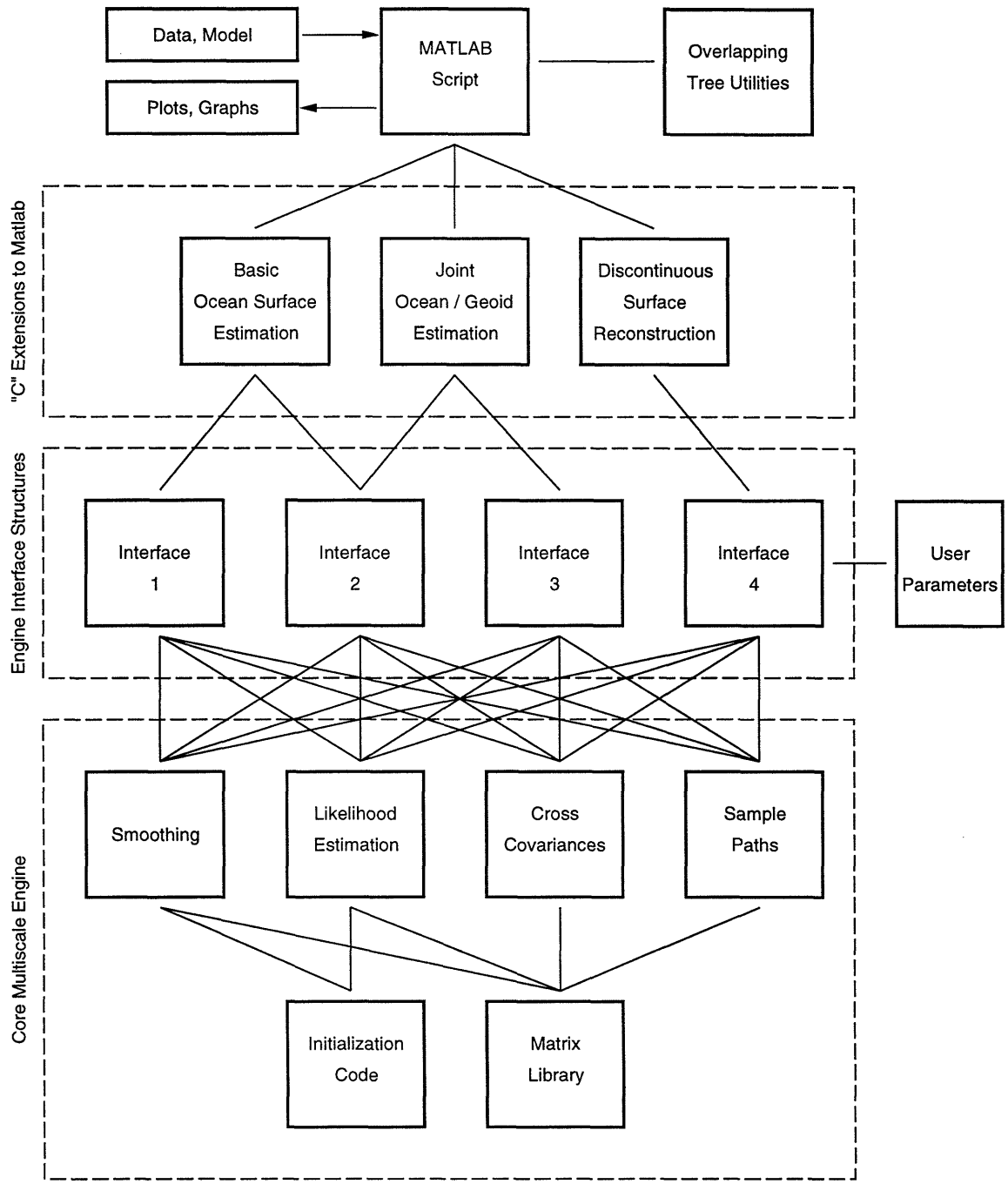


Figure A-1: A high level overview of the multiscale engine structure, and the interface to it from a high level application such as Matlab.

```
/*************************************************************
 *
 * interface 1 - Observations permitted at finest scale only
 *               Constant C matrix over finest scale
 *               Tree structure completely homogeneous and 2D
 *               Constant A matrix over entire tree
 *               Process noise B matrix constant over each scale
 */

#define Smoother smth1(                                              \
              int x_size,        /* state vector length          */  \
              int y_size,        /* observation vector length    */  \
              double *x_prior,   /* prior values at coarse scale */  \
              double **x_est,    /* multiscale estimated values  */  \
              double *y_obs,     /* finest scale measurements    */  \
              double *r_obs,     /* measurement covariances      */  \
              double **p_cond,   /* multiscale error variances   */  \
              double **p_serr,   /* mult. smoothed error varianc */  \
              double *po,        /* root node apriori variance   */  \
              double *a,         /* A matrix                     */  \
              double **bb,       /* B*B' matrices                */  \
              double *c,         /* C matrix at finest level     */  \
              int scales,        /* number of scales             */  \
              int *dx,           /* subgrid elements along x     */  \
              int *dy )          /* subgrid elements along y     */
```

Figure A-2: Interface 1 assumptions and estimator parameters.

```
/*************************************************************
 *
 * interface 2 - Measurements permitted on all scales
 *               C matrix must be constant over each scale
 *               Tree structure is homogeneous on each scale, 2D
 *               A matrix may vary with scale and orientation:
 *                  dx*dy A matrices specified per scale
 *                  (one A matrix per descendant orientation)
 */

#define Smoother smth2(                                              \
              int *x_size,      /* state vector length          */   \
              int *y_size,      /* observation vector length    */   \
              double *x_prior,  /* prior values at coarse scale */   \
              double **x_est,   /* multiscale estimated values  */   \
              double **y_obs,   /* state measurements           */   \
              double **r_obs,   /* measurement covariances      */   \
              double **p_cond,  /* multiscale error variances   */   \
              double **p_serr,  /* mult. smoothed error varianc */   \
              double *po,       /* root node apriori variance   */   \
              double ***a,      /* A matrices                   */   \
              double **bb,      /* B*B' matrices                */   \
              double **c,       /* C matrices over scales       */   \
              int scales,       /* number of scales             */   \
              int *dx,          /* subgrid elements along x     */   \
              int *dy )         /* subgrid elements along y     */
```

Figure A-3: Interface 2 assumptions and estimator parameters.

```
/*************************************************************
 *
 * interface 3 - Measurements permitted on all tree scales.
 *               Tree structure is homogeneous on each scale, 2D.
 *               The A, B, C matrices may vary arbitrarily over
 *                 the tree, except that the state length and
 *                 the observation length be fixed on each scale.
 *                 These matrices are specified as external
 *                 functions.
 */

#define Smoother smth3(                                              \
              int *x_size,      /* state vector length          */   \
              int *y_size,      /* observation vector length    */   \
              double *x_prior,  /* prior value at coarse scale  */   \
              double **x_est,   /* multiscale estimated values  */   \
              double **y_obs,   /* state measurements           */   \
              double **r_obs,   /* measurement covariances      */   \
              double **p_cond,  /* multiscale error variances   */   \
              double **p_serr,  /* mult. smoothed error varianc */   \
              double *po,       /* root node apriori variance   */   \
              double *(*a)(),   /* A matrix function            */   \
              double *(*bb)(),  /* B*B' matrix function         */   \
              double *(*c)(),   /* C matrix function            */   \
              int scales,       /* number of scales             */   \
              int *dx,          /* subgrid elements along x     */   \
              int *dy )         /* subgrid elements along y     */
```

Figure A-4: Interface 3 assumptions and estimator parameters.

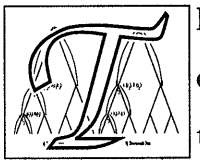
```
/*************************************************************
 *
 * interface 4 - Measurements permitted on all tree scales.
 *               Tree structure is homogeneous on each scale, 2D.
 *               All system matrices may vary arbitrarily over
 *                 the tree, except that the state length must
 *                 be fixed on each scale.
 *               The following parameters are supplied as
 *                 external functions:
 *                   Measurement state length
 *                   Observation matrix C
 *                   Measurement
 *                   Measurement covariance
 *                   Dynamics matrix A
 *                   Process noise matrix B
 */

#define Smoother smth4(                                              \
              int *x_size,      /* state vector length          */   \
              int (*ysiz)(),    /* observation vector length    */   \
              double *x_prior,  /* prior value at coarse scale  */   \
              double **x_est,   /* multiscale estimated values  */   \
              double *(*y)(),   /* state measurements           */   \
              double *(*r)(),   /* measurement covariances      */   \
              double **p_cond,  /* multiscale error variances   */   \
              double **p_serr,  /* mult. smoothed error varianc */   \
              double *po,       /* root node apriori variance   */   \
              double *(*a)(),   /* A matrix function            */   \
              double *(*bb)(),  /* B*B' matrix function         */   \
              double *(*c)(),   /* C matrix function            */   \
              int scales,       /* number of scales             */   \
              int *dx,          /* subgrid elements along x     */   \
              int *dy )         /* subgrid elements along y     */
```

Figure A-5: Interface 4 assumptions and estimator parameters.

Appendix B

Multiscale Estimation Equations



his appendix describes the algorithm that implements our multiscale estimation scheme. The description below is complete but terse; interested readers are referred to [14, 15, 62] for a more thorough development. Similarly, interested readers are referred to Appendix C and [65] for a description of the multiscale likelihood calculation algorithm, and to [64] for the derivation of the smoothing error model which enables the calculation of off-diagonal entries in the estimation error covariance matrix.

The multiscale smoother is basically the same as the Rauch-Tung-Striebel smoother operating in one dimension (along scale) with the addition of a merge operation that combines the information of multiple child nodes into one parent node (upwards pass), and the addition of a split operation which distributes information from a parent node to its multiple child nodes (downwards pass).

A certain amount of notation is required in order to describe the relative positions of state nodes on a tree; Figure B-1 shows the various relations:

s is an abstract index for identifying nodes on the tree

$\bar{\gamma}$ is the raising operator; i.e., $s\bar{\gamma}$ is the parent of s

$\bar{\delta}$ is the sibling operator; i.e., $s\bar{\delta}$ is the sibling node next to s

α is the lowering operator; i.e., $s\alpha_n$ is the n^{th} child of s

q is the order of the tree; i.e., the number of descendants of each parent

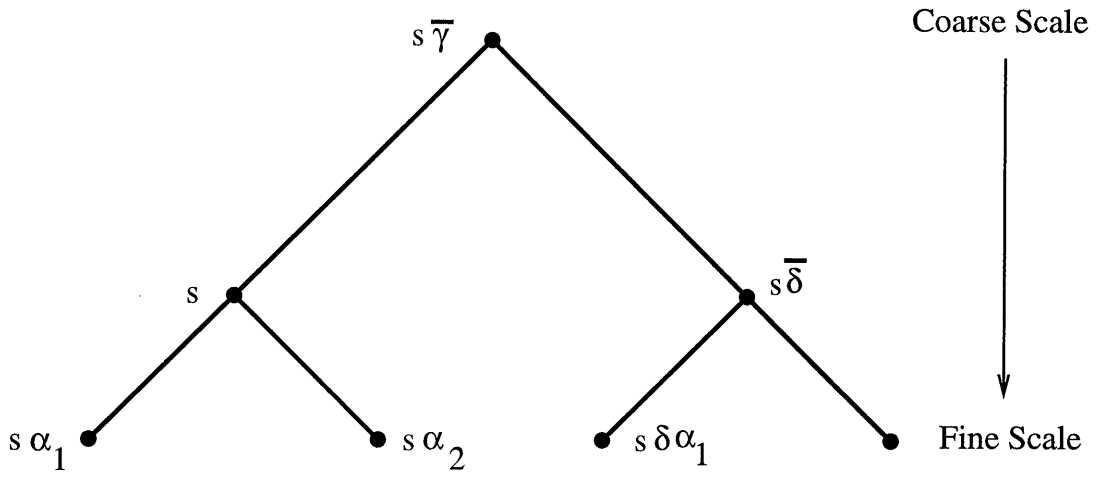


Figure B-1: Simple multiscale tree demonstrating node nomenclature

M is the number of scales on the tree

Note that operators can be cascaded, e.g., $s\bar{\gamma}\bar{\delta}\alpha_2$. The terms "upwards" and "downwards" are used with respect to the tree of Figure B-1; that is "upwards" implies a movement towards coarser scales, and "downwards" towards finer scales.

The tree process and observation relations are described as follows:

$$x(s) = A(s)x(s\bar{\gamma}) + B(s)w(s) \tag{B.1}$$

$$y(s) = C(s)x(s) + v(s) \tag{B.2}$$

where the process noise satisfies

$$E[w(s)] = 0 \qquad E\left[w(s)w(t)^T\right] = I\delta_{s,t} \tag{B.3}$$

and with a prior covariance at the root node

$$E[x(0)] = 0 \qquad E\left[x(0)x(0)^T\right] = P_o \tag{B.4}$$

From [15], corresponding to any choice of a downwards model in Equation B.1, we have the following upward model:

$$\begin{aligned} x(s\bar{\gamma}) &= F(s)x(s) + \bar{w}(s) && \text{(B.5)} \\ y(s) &= C(s)x(s) + v(s) && \text{(B.6)} \\ F(s) &= P_{s\bar{\gamma}}A^T(s)P_s^{-1} && \text{(B.7)} \\ E\left[\bar{w}(s)\bar{w}^T(s)\right] &= P_{s\bar{\gamma}}\left(I - A^T(s)P_s^{-1}A(s)P_{s\bar{\gamma}}\right) && \text{(B.8)} \\ &= Q(s) && \text{(B.9)} \end{aligned}$$

P_s is the *prior* variance of the state $x(s)$. To make the estimator equations more compact, additional notation is required at this point:

$$\begin{aligned} Y_s &= \{y(\sigma) \mid \sigma \text{ is a descendant of } s\} && \text{(B.10)} \\ \hat{x}(\sigma \mid s) &= E\left[x(\sigma) \mid \sigma \in Y_s \cup y(s)\right] && \text{(B.11)} \\ \hat{x}(\sigma \mid s+) &= E\left[x(\sigma) \mid \sigma \in Y_s\right] && \text{(B.12)} \\ \tilde{P}(\sigma \mid s) &= \mathrm{Cov}\left[x(\sigma) - \hat{x}(\sigma \mid s)\right] && \text{(B.13)} \\ \tilde{P}(\sigma \mid s+) &= \mathrm{Cov}\left[x(\sigma) - \hat{x}(\sigma \mid s+)\right] && \text{(B.14)} \end{aligned}$$

The algorithm now proceeds in three steps, outlined below.

1. Initialization

At each finest-scale node s, assign the following prior values:

$$\begin{aligned} \hat{x}(s \mid s+) &= 0 && \text{(B.15)} \\ \tilde{P}(s \mid s+) &= P_s && \text{(B.16)} \end{aligned}$$

2. Upward Sweep

The upward sweep operates much like a Kalman filter operating along scale with the addition of a merge step. The Kalman filter update step is performed at all nodes:

$$\hat{x}(s \mid s) = \hat{x}(s \mid s+) + K(s)\left[y(s) - C(s)\hat{x}(s \mid s+)\right] \tag{B.17}$$

$$\tilde{P}(s \mid s) = \left[I - K(s)C(s)\right]\tilde{P}(s \mid s+) \tag{B.18}$$

$$K(s) = \tilde{P}(s \mid s+)C^T(s)V^{-1}(s) \tag{B.19}$$

$$V(s) = C(s)\tilde{P}(s \mid s+)C^T(s) + R(s) \tag{B.20}$$

The Kalman filter prediction step is applied at all nodes except for leaf nodes (which were initialized as outlined above):

$$\hat{x}(s \mid s\alpha_i) = F(s\alpha_i)\hat{x}(s\alpha_i \mid s\alpha_i) \tag{B.21}$$

$$\tilde{P}(s \mid s\alpha_i) = F(s\alpha_i)\tilde{P}(s\alpha_i \mid s\alpha_i)F^T(s\alpha_i) + Q(s\alpha_i) \tag{B.22}$$

Finally, at all nodes except leaf nodes, the merge step combines predicted estimates from offspring $(1 \ldots q)$ into a single prediction to be used in the update step:

$$\hat{x}(s \mid s+) = \tilde{P}(s \mid s+)\sum_{i=1}^{q}\tilde{P}^{-1}(s \mid s\alpha_i)\hat{x}(s \mid s\alpha_i) \tag{B.23}$$

$$\tilde{P}(s \mid s+) = \left[(1-q)P_s^{-1} + \sum_{i=1}^{q}\tilde{P}^{-1}(s \mid s\alpha_i)\right]^{-1} \tag{B.24}$$

3. Downward Sweep

The termination of the upward sweep gives the smoothed estimate $\hat{x}(0) = \hat{x}(0 \mid 0)$ at the root node. The remainder of the smoothed estimates are found by propagating information back down the tree:

$$\hat{x}(s) = \hat{x}(s \mid s) + J(s)\left[\hat{x}(s\bar{\gamma}) - \hat{x}(s\bar{\gamma} \mid s)\right] \tag{B.25}$$

$$\tilde{P}(s) = \tilde{P}(s \mid s) + J(s)\left[\tilde{P}(s\bar{\gamma}) - \tilde{P}(s\bar{\gamma} \mid s)\right]J^T(s) \tag{B.26}$$

$$J(s) = \tilde{P}(s \mid s)F^T(s)\tilde{P}^{-1}(s\bar{\gamma} \mid s) \tag{B.27}$$

The smoothed measurements are given by $\hat{x}(s)$; the corresponding estimation error variances are given by $\tilde{P}(s)$. Cross covariances are not computed explicitly, rather

the means for their computation is implicit (although by no means obvious) in the above algorithm. The multiscale form of the smoothing error[64] is as follows:

$$\tilde{x}(s) = \tilde{P}(s \mid s)F^T(s)\tilde{P}^{-1}(s\bar{\gamma} \mid s)\tilde{x}(s\bar{\gamma}) + \check{w}(s) \tag{B.28}$$

$$\tilde{x}(s) = x(s) - \hat{x}(s) \tag{B.29}$$

where $\check{w}(s)$ represents white noise.

Appendix C

Multiscale Likelihood Calculation Equations

hereas Appendix B outlined the multiscale estimation algorithm, this appendix gives a parallel description of the likelihood calculation scheme. The description below is sufficiently complete for implementation purposes, however it is terse and it does not present any insights into the equations themselves; interested readers are referred to [62, 65] for a more thorough development.

Let θ represent the multiscale model parameters (e.g., the model parameters $A(s), B(s), C(s), R(s), P_o$, and possibly other parameters which specify the structure of the tree such as q), and let Y represent a stacked vector of multiscale measurements. Then the likelihood function $\mathcal{L}(\theta)$ is given by

$$\mathcal{L}(\theta) = \log p_{y|\theta}(Y|\theta) \tag{C.1}$$

$$= -\frac{1}{2}\log|\Lambda_y| - \frac{1}{2}Y^T\Lambda_y^{-1}Y - \frac{\|Y\|}{2}\log 2\pi \tag{C.2}$$

where $\|Y\|$ counts the number of elements in Y, and where Λ_y represents the covariance of random vector y. The difficulty in determining $\mathcal{L}(\theta)$ directly from (C.2) is that Λ_y is a full matrix, making the direct computation of Λ_y^{-1} or $|\Lambda_y|$ a computationally difficult task, requiring $\mathcal{O}(\|Y\|^3)$ computations. The matrix Λ_y does possess a great

deal of structure, inherited from the multiscale tree, and so a much more enlightened approach is possible.

Specifically, by *whitening* the measurement residuals, the computation of the likelihood function becomes trivial. We require an invertible transformation between the measurements $\{y(s)\}$ and a set of measurement residuals $\{\nu(s)\}$ such that the covariance Λ_ν is diagonal. Given the construction of such a sequence, the likelihood function is readily calculated

$$\mathcal{L}(\theta) = -\frac{1}{2}\sum_{s\in\mathcal{O}}\left\{|\ \nu(s)\log 2\pi + \log|\Lambda_{\nu(s)}| + \nu^T(s)\Lambda^{-1}_{\nu(s)}\nu(s)\right\} \tag{C.3}$$

The multiscale likelihood calculator operates in a manner similar to that of the multiscale smoother, except that the tree is traversed in a "depth first" manner, rather than uniformly by scale. The reader is referred to the beginning of Appendix B for a summary of the required multiscale notation. Although the equations below will use unfamiliar superscript/subscript conventions, we choose not to define these conventions because their definitions are complicated and not necessary for the implementation of the algorithm.

The tree process and observation relations are described as follows:

$$x(s) = A(s)x(s\bar{\gamma}) + B(s)w(s) \tag{C.4}$$

$$y(s) = C(s)x(s) + v(s) \tag{C.5}$$

where the random variables $\{w(s)\}$ and $\{v(s)\}$ are jointly normal, such that

$$w(s) \sim \mathcal{N}(0,I) \qquad E\left[w(s)w(t)^T\right] = I\delta_{s,t} \qquad x(0) \sim \mathcal{N}(0,P_o) \tag{C.6}$$

$$v(s) \sim \mathcal{N}(0,R(s)) \qquad E\left[v(s)v(t)^T\right] = I\delta_{s,t} \tag{C.7}$$

Corresponding to (C.4) we have an upward model described by $F(s)$, $Q(s)$:

$$F(s) = P_{s\bar{\gamma}}A^T(s)P_s^{-1} \tag{C.8}$$

$$Q(s) = P_{s\bar{\gamma}} \left(I - A^T(s) P_s^{-1} A(s) P_{s\bar{\gamma}} \right) \tag{C.9}$$

P_s is the *prior* variance of the state $x(s)$:

$$P_s = A(s) P_{s\bar{\gamma}} A(s)^T + B(s) B^T(s) \tag{C.10}$$

The whitening algorithm now proceeds in three steps, outlined below.

1. Initialization

At each leaf node s, assign the following prior values:

$$\hat{x}(s \mid Y_s^{\alpha_q}) = 0 \tag{C.11}$$

$$\tilde{P}(s \mid Y_s^{\alpha_q}) = P_s \tag{C.12}$$

2. Upward Sweep

Kalman filter update step:

$$\hat{x}(s \mid Y_s) = \hat{x}(s \mid Y_s^{\alpha_q}) + K(s) \left[y(s) - C(s) \hat{x}(s \mid Y_s^{\alpha_q}) \right] \tag{C.13}$$

$$K(s) = \tilde{P}(s \mid Y_s^{\alpha_q}) C^T(s) \left[C(s) \tilde{P}(s \mid Y_s^{\alpha_q}) C^T(s) + R(s) \right]^{-1} \tag{C.14}$$

$$\tilde{P}(s \mid Y_s) = \left[I - K(s) C(s) \right] \tilde{P}(s \mid Y_s^{\alpha_q}) \tag{C.15}$$

Kalman filter prediction step (applied at all nodes except for leaf nodes):

$$\hat{x}(s \mid Y_s \alpha_i) = F(s\alpha_i) \hat{x}(s\alpha_i \mid Y_s \alpha_i) \tag{C.16}$$

$$\tilde{P}(s \mid Y_s \alpha_i) = F(s\alpha_i) \tilde{P}(s\alpha_i \mid Y_s \alpha_i) F^T(s\alpha_i) + Q(s\alpha_i) \tag{C.17}$$

Finally, at all nodes except leaf nodes, the merge step combines predicted estimates from offspring $(1 \ldots i, 1 \leq i \leq q)$ into a set of conditional predictions:

$$\hat{x}(s \mid Y_s^{\alpha_i}) = \tilde{P}(s \mid Y_s^{\alpha_i}) \sum_{j=1}^{i} \tilde{P}^{-1}(s \mid Y_s \alpha_q) \hat{x}(s \mid Y_s \alpha_j) \tag{C.18}$$

$$\tilde{P}(s \mid Y_s^{\alpha_i}) = \left[(1-i) P_s^{-1} + \sum_{j=1}^{i} \tilde{P}^{-1}(s \mid Y_s \alpha_j) \right]^{-1} \tag{C.19}$$

3. Downward Sweep

The termination of the upward sweep gives the conditional estimate $\hat{x}(s \mid Y_s^{\alpha_i})$ and conditional error smoothed estimate $\tilde{P}(s \mid Y_s^{\alpha_i})$ at each node s. The downward sweep is initialized at the root node:

$$\hat{x}(0 \mid \bar{Y}_0) = 0 \tag{C.20}$$

$$\tilde{P}(0 \mid \bar{Y}_0) = P_o \tag{C.21}$$

$$\hat{x}(0 \mid \bar{Y}_0, Y_0^{\alpha_i}) = \hat{x}(0 \mid Y_0^{\alpha_i}) \qquad 1 \le i \le q \tag{C.22}$$

$$\tilde{P}(0 \mid \bar{Y}_0, Y_0^{\alpha_i}) = \tilde{P}(0 \mid Y_0^{\alpha_i}) \qquad 1 \le i \le q \tag{C.23}$$

The essence of the downward sweep is captured by the following:

$$\hat{x}(s\alpha_i \mid \bar{Y}_s\alpha_i) = \begin{cases} A(s\alpha_i)\hat{x}(s \mid \bar{Y}_s) & i = 1 \\ A(s\alpha_i)\hat{x}(s \mid \bar{Y}_s, Y_s^{\alpha_{i-1}}) & 2 \le i \le q \end{cases} \tag{C.24}$$

$$\tilde{P}(s\alpha_i \mid \bar{Y}_s\alpha_i) = \begin{cases} A(s\alpha_i)\tilde{P}(s \mid \bar{Y}_s)A^T(s\alpha_i) & i = 1 \\ A(s\alpha_i)\tilde{P}(s \mid \bar{Y}_s, Y_s^{\alpha_{i-1}})A^T(s\alpha_i) + B(s\alpha_i)B^T(s\alpha_i) & 2 \le i \le q \end{cases} \tag{C.25}$$

Finally, at each node on the tree below the root we merge the information from the upward and downward stages:

$$\hat{x}(s \mid \bar{Y}_s, Y_s^{\alpha_i}) = \tilde{P}(s \mid \bar{Y}_s, Y_s^{\alpha_i}) \left[\tilde{P}^{-1}(s \mid Y_s^{\alpha_i})\hat{x}(s \mid Y_s^{\alpha_i}) + \tilde{P}^{-1}(s \mid \bar{Y}_s)\hat{x}(s \mid \bar{Y}_s)\right] \tag{C.26}$$

$$\tilde{P}(s \mid \bar{Y}_s, Y_s^{\alpha_i}) = \left[\tilde{P}^{-1}(s \mid Y_s^{\alpha_i}) + \tilde{P}^{-1}(s \mid \bar{Y}_s) - P_s^{-1}\right]^{-1} \tag{C.27}$$

for $1 \le i \le q$.

The computation of the whitened residuals follows trivially from the results of the above merge step:

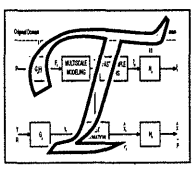
$$\nu(s) = y(s) - C(s)\hat{x}(s \mid \bar{Y}_s, Y_s^{\alpha_q}) \tag{C.28}$$

$$\Lambda_{\nu(s)} = C(s)\tilde{P}(s \mid \bar{Y}_s, Y_s^{\alpha_q})C^T(s) + R(s) \tag{C.29}$$

The computation of $\nu(s), \Lambda_{\nu(s)}$ then leads directly to the likelihood function via (C.3).

Appendix D

Proofs of Overlap Propositions



his appendix extends in greater detail some of the analytical results of Chapter 6. Section D.1 presents a proof of the propositions in Chapter 6 that the estimates produced using an overlapping framework are optimal under certain conditions. Section D.2 presents a somewhat more involved deductive proof that leads to the same conclusions, but which adds certain insights not found in the proof of the propositions in Section D.1.

D.1 Sufficiency Proof

This section[1] establishes the validity of (6.32) and (6.33), thereby proving the Proposition in Section 6.2.3. Our proof is facilitated by the following identity:

$$(CPC^T + R)^{-1} = G_y^T (C_l P_l C_l^T + R_l)^{-1} G_y \tag{D.1}$$

Proof of (6.32):

$$\begin{aligned} L &= PC^T (CPC^T + R)^{-1} \\ &= PC^T G_y^T \left(C_l P_l C_l^T + R_l \right)^{-1} G_y \end{aligned}$$

[1] A large fraction of the results of Section D.1 are due to the efforts of W. Irving. The section is included here for completeness; the essential results of this appendix, in Section D.2, are the work of the thesis author alone.

$$
\begin{aligned}
&= PG_x^T C_l^T \left(C_l P_l C_l^T + R_l\right)^{-1} G_y \\
&= H_x P_l C_l^T \left(C_l P_l C_l^T + R_l\right)^{-1} G_y \\
&= H_x L_l G_y
\end{aligned}
\tag{D.2}
$$

Proof of (6.33):

$$
\begin{aligned}
\tilde{P} &= P - LCP \\
&= H_x P_l H_x^T - H_x L_l G_y C P \\
&= H_x P_l H_x^T - H_x L_l C_l G_x P \\
&= H_x \left(P_l - L_l C_l P_l\right) H_x^T \\
&= H_x \tilde{P}_l H_x^T.
\end{aligned}
\tag{D.3}
$$

D.2 Constructive Proof of Overlapping Conjectures

This section provides an alternative proof to that of Section D.1. The proof is included for two reasons:

1. Its assumptions are slightly weaker than those of the proposition of Section 6.2.3, and may lead to interesting generalizations.

2. The work of this section actually preceded most of the development of Chapter 6; several of the insights in Chapter 6 (in particular, the structure of R_l) were acquired in the process of developing this proof. Thus, while the proof of Section D.1 is succinct and more easily understood, it did not play as significant a role in the development of the sufficient conditions of Section 6.2.3 and in the definition of the overlapping framework.

Proposition:

Given an estimation problem:

$y = Cx + v \qquad E[x] = 0, \ \text{cov}(x) = P, \ E[v] = 0, \ \text{cov}(v) = R$

P, R nonsingular, R diagonal

C a weighted selection matrix

Then, given projection matrices G_x, H_x, G_y to/from the lifted domain,

$y_l = G_y y, \ x_l = G_x x, \ \hat{\hat{x}} = H_x \hat{x}_l$

such that the projection matrices satisfy

G_x, G_y have full column rank

G_y is a one-to-many operator, the elements of G_y are zero or one

and which define the lifted-domain estimation problem

$y_l = C_l x_l + v_l \qquad E[x_l] = 0, \ \text{cov}(x) = G_x P G_x^T, \ E[v_l] = 0, \ \text{cov}(v) = R_l$

P, R nonsingular, R diagonal

C_l a weighted selection matrix

Then the selection of H_x, C_l, R_l as described in Section 6.2.2 is sufficient to ensure that the optimal estimate,

$\hat{x} = Ly$

and the projection of the optimal estimate from the lifted domain,

$\hat{\hat{x}} = H_x \hat{x}_l = H_x L_l y_l$

both yield the same, optimal result:

$\hat{x} = \hat{\hat{x}}$

We will first present the estimation procedure of interest in the original domain in order to introduce the appropriate nomenclature, followed by a general description of the redundant estimation domain and a development of its properties.

Pseudoinverse Properties

The following development will require the use of Moore-Penrose pseudoinverse matrices[2], certain properties of which are summarized below. Let A, B be real matrices, and represent their pseudoinverses by A^+, B^+ respectively; let $Nu(A)$ represent the nullspace of matrix A; then

$$A^+A = (A^+A)^T, AA^+ = (AA^+)^T \tag{D.4}$$

$$\left(A^+\right)^T = \left(A^T\right)^+ \tag{D.5}$$

$$AA^+A = A \tag{D.6}$$

$$A^+AA^+ = A^+ \tag{D.7}$$

$$Nu(A^T) = \{0\} \Rightarrow AA^+ = I \tag{D.8}$$

$$Nu(A) = \{0\} \Rightarrow A^+A = I \tag{D.9}$$

$$Nu(A) = \{0\} \Rightarrow A^+ = (A^TA)^{-1}A^T \tag{D.10}$$

$$Nu(A^T) = \{0\} \Rightarrow A^+ = A^T(AA^T)^{-1} \tag{D.11}$$

$$A \in \mathcal{R}^{m\times n}, B \in \mathcal{R}^{n\times p}, \; rank(A) = rank(B) = n \Rightarrow (AB)^+ = B^+A^+ \tag{D.12}$$

Finally, suppose that

$$y = Cx + v \qquad E[v] = 0, \;\; \mathrm{cov}(v) = I \tag{D.13}$$

where C is any real matrix. Then the least-squares minimum-norm solution for x is given by[2]

$$\hat{x} = C^+y \tag{D.14}$$

Estimation in Original Domain

We would like to perform least-squares estimation of a random vector x, governed by a prior model P, and based on observations y having measurement error covariance R; i.e.,

$$y = Cx + v \qquad \begin{array}{l} x \sim \mathcal{N}(0, P) \\ v \sim \mathcal{N}(0, R) \end{array} \tag{D.15}$$

where $\mathcal{N}$ denotes a Gaussian distribution. The problems of interest are those in which C is a selection matrix (i.e., each row of C is a multiple of a unit vector), the prior covariance P is nonsingular, and in which the measurement covariance R is diagonal and nonsingular.

We are interested in projecting the above estimation problem into a "lifted" domain, in which the state and measurement vectors possess a redundant representation, yielding a singular estimation problem.

The interest in such a domain stems from the results of Chapters 6 and 7, which suggest that desirable smoothness properties may be obtained by applying the multiscale estimation framework directly in the lifted domain.

We will be needing a fairly general estimation formulation (specifically, one tolerating measurement and prior model singularities) once this problem is projected into the lifted domain. To anticipate this need we rewrite the estimation problem (D.15) more generally (although still assuming that P,R are invertible) as the following equivalent maximum likelihood statement by incorporating prior knowledge as additional measurements:

$$\begin{bmatrix} y \\ 0 \end{bmatrix} = \begin{bmatrix} C \\ I \end{bmatrix} x + \bar{v} \qquad \bar{v} \sim \mathcal{N}\left(0, \begin{bmatrix} R & \\ & P \end{bmatrix}\right) \tag{D.16}$$

If we define L such that

$$LL^T = \begin{bmatrix} R & \\ & P \end{bmatrix} \tag{D.17}$$

then the following estimation problem is equivalent to (D.16),

$$L^{-1}\begin{bmatrix} y \\ 0 \end{bmatrix} = L^{-1}\begin{bmatrix} C \\ I \end{bmatrix} x + \bar{\bar{v}} \qquad \bar{\bar{v}} \sim \mathcal{N}(0, I) \tag{D.18}$$

Since the estimation problem (D.18) has a diagonal, unit-variance measurement error covariance, then from (D.14) the least-squares solution is straightforward:

$$\hat{x} = \left\{ L^{-1}\begin{bmatrix} C \\ I \end{bmatrix} \right\}^{+} L^{-1}\begin{bmatrix} y \\ 0 \end{bmatrix} \tag{D.19}$$

Since

$$\left\{ L^{-1}\begin{bmatrix} C \\ I \end{bmatrix} \right\}^{+} \tag{D.20}$$

has full column rank, by using property (D.10) the above result reduces to the expected classical least-squares estimation formula:

$$\hat{x} = \left(\left\{ L^{-1}\begin{bmatrix} C \\ I \end{bmatrix} \right\}^{T} \left\{ L^{-1}\begin{bmatrix} C \\ I \end{bmatrix} \right\} \right)^{-1} \left\{ L^{-1}\begin{bmatrix} C \\ I \end{bmatrix} \right\}^{T} L^{-1}\begin{bmatrix} y \\ 0 \end{bmatrix} \tag{D.21}$$

$$= \left(C^T R^{-1} C + P^{-1} \right)^{-1} C^T R^{-1} y \tag{D.22}$$

Estimation in Lifted Domain

We define the "lifted" domain as a domain possessing a redundant representation of the elements of x. The specific distribution and usefulness of the added redundancies will, of course, be highly dependent on the intended application. The nature of the redundancy is parameterized in terms of the following projection operations between the lifted and original domains:

- A projection of the unknown process into the lifted domain: $x_l = G_x x$
- A projection of measurements into the lifted domain: $y_l = G_y y$
- A projection of estimates from the lifted domain: $\hat{x} = H_x \hat{x}_l$

The subscript l, as in x_l, denotes vectors and matrices of the lifted domain. There are two assumptions imposed on these projections:

1. Each row of G_y copies a measurement from one element of y into one element of y_l in the lifted domain; i.e., each row of G_y is a row vector with one element equal to one, and the remaining elements equal to zero.

2. G_y has full column rank.

3. G_x has full column rank.

Using these projection matrices and the estimator from (D.16) we can express the estimation problem in the lifted domain:

$$\begin{bmatrix} y_l \\ 0 \end{bmatrix} = \begin{bmatrix} G_y & \\ & G_x \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} + \bar{\eta} \tag{D.23}$$

$$= \begin{bmatrix} G_y & \\ & G_x \end{bmatrix} \begin{bmatrix} C \\ I \end{bmatrix} x + \bar{\eta} \tag{D.24}$$

$$= \begin{bmatrix} G_y & \\ & G_x \end{bmatrix} \begin{bmatrix} C \\ I \end{bmatrix} G_x^+ G_x x + \bar{\eta} \tag{D.25}$$

$$= \begin{bmatrix} G_y C G_x^+ \\ G_x G_x^+ \end{bmatrix} x_l + \bar{\eta} \qquad \bar{\eta} \sim \mathcal{N}\left(\begin{bmatrix} G_y R G_y^T & \\ & G_x P G_x^T \end{bmatrix} \right) \tag{D.26}$$

where, by (D.9), $G_x^+ G_x = I$. We propose to determine an estimator for the problem of (D.26) in a manner analogous to (D.17). Let

$$L_l = \left\{ \begin{bmatrix} G_y & \\ & G_x \end{bmatrix} L \right\} \tag{D.27}$$

Premultiplying (D.26) by L_l^+:

$$L_l^+ \begin{bmatrix} y_l \\ 0 \end{bmatrix} = L_l^+ \begin{bmatrix} G_y C G_x^+ \\ G_x G_x^+ \end{bmatrix} x_l + L_l^+ \bar{\eta} \tag{D.28}$$

This operation is invertible, since (D.23) is recoverable from (D.28); specifically,

$$L_l L_l^+ \begin{bmatrix} y_l \\ 0 \end{bmatrix} = \begin{bmatrix} G_y G_y^+ G_y y \\ 0 \end{bmatrix} = \begin{bmatrix} G_y y \\ 0 \end{bmatrix} = \begin{bmatrix} y_l \\ 0 \end{bmatrix} \tag{D.29}$$

where, by (D.9), $G_y G_y^+ = I$. Since L is invertible and G_x, G_y have full column rank, by (D.12)

$$L_l^+ = L^{-1} \begin{bmatrix} G_y^+ & \\ & G_x^+ \end{bmatrix} \tag{D.30}$$

Thus the covariance of $L^+\bar{\eta}$ in (D.28) possesses straightforward statistics:

$$\bar{\bar{\eta}} = L_l^+ \bar{\eta} \sim \mathcal{N}\left(0, L_l^+ \begin{bmatrix} G_y R G_y^T & \\ & G_x P G_x^T \end{bmatrix} L_l^{+T}\right) \tag{D.31}$$

$$\sim \mathcal{N}\left(0, L^{-1} \begin{bmatrix} G_y^+ & \\ & G_x^+ \end{bmatrix} \begin{bmatrix} G_y R G_y^T & \\ & G_x P G_x^T \end{bmatrix} \begin{bmatrix} G_y^{+T} & \\ & G_x^{+T} \end{bmatrix} L^{-1}\right) \tag{D.32}$$

$$\sim \mathcal{N}(0, I) \tag{D.33}$$

where, by (D.9), $G_y^+ G_y = G_y^T G_y^{+T} = I$ and $G_x^+ G_x = G_x^T G_x^{+T} = I$. By (D.14) the estimator for x_l immediately follows as

$$\hat{x}_l = \left\{ L_l^+ \begin{bmatrix} G_y C G_x^+ \\ G_x G_x^+ \end{bmatrix} \right\}^+ L_l^+ \begin{bmatrix} y_l \\ 0 \end{bmatrix} \tag{D.34}$$

$$= \left\{ \left\{ \begin{bmatrix} G_y & \\ & G_x \end{bmatrix} L \right\}^+ \begin{bmatrix} G_y C G_x^+ \\ G_x G_x^+ \end{bmatrix} \right\}^+ \left\{ \begin{bmatrix} G_y & \\ & G_x \end{bmatrix} L \right\}^+ \begin{bmatrix} y_l \\ 0 \end{bmatrix} \tag{D.35}$$

$$= \left\{ L^{-1} \begin{bmatrix} G_y^+ G_y C G_x^+ \\ G_x^+ G_x G_x^+ \end{bmatrix} \right\}^+ L^{-1} \begin{bmatrix} G_y^+ y_l \\ 0 \end{bmatrix} \tag{D.36}$$

$$= \left\{ L^{-1} \begin{bmatrix} C G_x^+ \\ G_x^+ \end{bmatrix} \right\}^+ L^{-1} \begin{bmatrix} G_y^+ y_l \\ 0 \end{bmatrix} \tag{D.37}$$

Once the estimates $\hat{x}_l$ have been computed in the lifted domain, an associated set of estimates in the original domain are determined by the projection

$$\hat{\hat{x}} = H_x \hat{x}_l \tag{D.38}$$

Our goal is to be able to apply our multiscale framework (Section 2.4, Appendix B) to the estimation problem in the lifted domain. The multiscale framework leads to an extremely efficient estimation algorithm if certain constraints are met. Specifically,

the algorithm is capable of efficiently solving the estimation problem

$$\begin{bmatrix} y_l \\ 0 \end{bmatrix} = \begin{bmatrix} C_l \\ I_l \end{bmatrix} x_l + w \qquad\qquad w \sim \mathcal{N}\left(0, \begin{bmatrix} R_l & \\ & P_l \end{bmatrix}\right) \tag{D.39}$$

under certain conditions for C_l and R_l:

1. The observation matrix C_l is a weighted selection matrix; that is, each row of C_l contains exactly one non-zero element.

2. The observation error covariance matrix R_l is diagonal.

From Section 2.5.2 multiscale techniques exist to model arbitrary prior covariance functions (at least in principle), so for the time being we place no restrictions on P_l and presume that $P_l = G_x P G_x^T$ can be modeled exactly.

Our goal is to determine conditions on the projection matrices G_y, G_x, H_x such that the estimation problem (D.36) in the lifted domain maps into the class of problems which we can solve using multiscale techniques.

Condition 1: $\hat{x} = H_x \hat{x}_l$

That is, we wish that $\hat{x}$, the optimal estimate of x in the original domain, equals $\hat{\hat{x}}$, the projection (D.38) of the optimal estimate of x_l from the lifted domain. Let us apply the projection H_x to the estimator for $\hat{x}_l$ from (D.37):

$$\begin{aligned} \hat{\hat{x}} &= H_x \hat{x}_l & \text{(D.40)} \\ &= H_x \left\{ L^{-1} \begin{bmatrix} C G_x^+ \\ G_x^+ \end{bmatrix} \right\}^+ L^{-1} \begin{bmatrix} G_y^+ y_l \\ 0 \end{bmatrix} & \text{(D.41)} \end{aligned}$$

then, since L is invertible and G_x^+ has full row rank,

$$\begin{aligned} \hat{\hat{x}} &= H_x G_x \left\{ L^{-1} \begin{bmatrix} C \\ I \end{bmatrix} \right\}^+ L^{-1} \begin{bmatrix} G_y^+ G_y y \\ 0 \end{bmatrix} & \text{(D.42)} \\ &= H_x G_x \hat{x} & \text{(D.43)} \end{aligned}$$

from which it follows that

$$H_x G_x = I \tag{D.44}$$

is sufficient to satisfy the given constraint.

Condition 2: R_l is diagonal

The value of R_l from (D.39) enters into the lifted domain estimator (D.36) only implicitly. Recall that R_l represents the measurement error covariance of y_l; i.e., the statistics of the noise term $\bar{\eta}$ in (D.26) are

$$\bar{\eta} \sim \mathcal{N}\left(0, \begin{bmatrix} \mathrm{cov}(y_l) & \\ & \mathrm{cov}(x_l) \end{bmatrix}\right) \tag{D.45}$$

$$\sim \mathcal{N}\left(0, \begin{bmatrix} R_l & \\ & G_x P G_x^T \end{bmatrix}\right) \tag{D.46}$$

The estimate produced by (D.36) is the same for any R_l such that $\bar{\eta}$ satisfies (D.33); i.e., when

$$L_l^+ \begin{bmatrix} R_l & \\ & G_x P G_x^T \end{bmatrix} L_l^{+T} = I \tag{D.47}$$

Substituting the expression for L_l^+ from (D.30),

$$L^{-1} \begin{bmatrix} G_y^+ R_l G_y^{+T} & \\ & G_x^+ G_x P G_x^T G_x^{+T} \end{bmatrix} L^{-1} = I \tag{D.48}$$

$$\begin{bmatrix} G_y^+ R_l G_y^{+T} & \\ & P \end{bmatrix} = \begin{bmatrix} R & \\ & P \end{bmatrix} \tag{D.49}$$

which is true whenever $G_y^+ R_l G_y^{+T} = R$. A sufficient condition for R_l to satisfy (D.48) is that

$$R_l = G_y R G_y^T + \sum_i \alpha_i \alpha_i^T \qquad \alpha_i \in Nu\left(G_y^+\right) \tag{D.50}$$

where the symmetrical form $\alpha_i \alpha_i^T$ is required to maintain the symmetrical form of R_l.

Label the columns of G_y as follows:

$$G_y = [g_1 \; g_2 \ldots g_N] \tag{D.51}$$

and let r_{jj} represent the j^{th} diagonal element of R; thus

$$G_y R G_y^T = \sum_j g_j r_{jj} g_j^T \tag{D.52}$$

Recall that we have restricted ourselves to projection matrices G_y such that each row of G_y is a unit vector; consequently each column of G_y may be written as

$$g_j = \sum_{i=1}^{M_j} e_{f(j,i)} \tag{D.53}$$

where e_i represents the i^{th} unit vector (i.e., a column vector with a '1' in the ith position and zeros everywhere else), and where

$$f(j,i) = f(k,l) \quad \text{iff } j = k, \; i = l \tag{D.54}$$

implying that the single observation $y[j]$ is mapped to observations

$$y_l[f(j,1)], \; y_l[f(j,2)], \; \ldots, \; y_l[f(j,M_j)] \tag{D.55}$$

in the lifted domain. The pseudoinverse G_y^+ and its nullspace follow immediately from (D.53):

$$G_y^+ = \begin{bmatrix} g_1^T/M_1 \\ g_2^T/M_2 \\ \cdots \\ g_N^T/M_N \end{bmatrix} \tag{D.56}$$

$$Nu\left(G_y^+\right) = span\left(\left\{(e_{f(j,a)} - e_{f(j,b)}) \mid 1 \le j \le N, 1 \le a \ne b \le M_j\right\}\right) \tag{D.57}$$

Based on this nullspace, each of the products $g_j r_{jj} g_j^T$ can be diagonalized as follows:

$$D_j = g_j r_{jj} g_j^T + \sum_{l=1}^{M_j-1} \sum_{m=l+1}^{M_j} r_{jj}(e_{f(j,l)} - e_{f(j,m)})(e_{f(j,l)} - e_{f(j,m)})^T \tag{D.58}$$

where D_j is diagonal, and each of the nonzero diagonal entries take on the value $M_j r_{jj}$. Consequently the following expression results in a diagonal R_l:

$$R_l = G_y R G_y^T + \sum_{j=1}^{N} \sum_{l=1}^{M_j-1} \sum_{m=l+1}^{M_j} r_{jj}(e_{f(j,l)} - e_{f(j,m)})(e_{f(j,l)} - e_{f(j,m)})^T \tag{D.59}$$

where each diagonal entry is equal to the multiplicity M_j of the corresponding observation, multiplied by the original error variance r_{jj}.

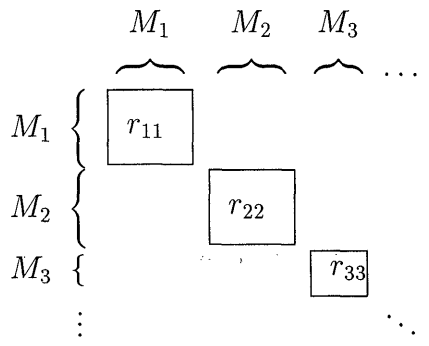
A more intuitive version of the above argument goes as follows. Without loss of generality we can reorder our observation vector y_l in the lifted domain, such that the columns of G_y possess the following regular structure:

$$g_1 = e_1 + \ldots e_{M_1} \tag{D.60}$$

$$g_2 = e_{M_1+1} + \ldots e_{M_1+M_2} \tag{D.61}$$

$$\vdots$$

Then $G_y R G_y^T$ is a block diagonal matrix such as the following



(D.62)

where $\boxed{r}$ denotes a square matrix with all elements equal to r. It is easy to see that

$$g_1 r_{11} g_1^T + \sum_{m=1}^{M_1-1} \sum_{n=m+1}^{M_1} r_{11}(e_m - e_n)(e_m - e_n)^T \tag{D.63}$$

is diagonal, where the nonzero diagonal elements are equal to $M_1 r_{11}$. Applying the above notion to each of the blocks in (D.62) yields the expected diagonal form for R_l.

Condition 3: C_l is a selection matrix

Let us define the region of support of a vector (or matrix) to be the set of indices (or index pairs) associated with its non-zero values. Let us write G_y, C, and G_x^+ as follows:

$$G_y = \begin{bmatrix} g_1 \; g_2 \ldots \end{bmatrix} \qquad C = \begin{bmatrix} \delta_1 e_{i_1}^T \\ \delta_2 e_{i_2}^T \\ \vdots \end{bmatrix} \qquad G_x^+ = \begin{bmatrix} h_1^T \\ h_2^T \\ \vdots \end{bmatrix} \tag{D.64}$$

Without loss of generality, we assume that $i_a = i_b$ iff $a = b$, that is, the rows of C are all different – there are no repeated measurements of the same element of x (if repeated measurements were present they could be collapsed into a single measurement).

Under the following two conditions,

1. $P_l = G_x P G_x^T$
2. $R_l = G_y R G_y^T + \alpha_1 \alpha_1^T + \alpha_2 \alpha_2^T + \ldots \qquad \alpha_i \in Nu\left(G_y^+\right)$
   (i.e., the condition from (D.48))

we can determine the estimate produced by the multiscale estimator. From the above conditions, the noise covariance of (D.39) satisfies

$$L_l^+ w \sim \mathcal{N}(0, I) \tag{D.65}$$

Then, paralleling the development from (D.28) through (D.36),

$$\hat{x}_l = \left\{ L_l^+ \begin{bmatrix} C_l \\ I_l \end{bmatrix} \right\}^+ L_l^+ \begin{bmatrix} y_l \\ 0 \end{bmatrix} \tag{D.66}$$

$$= \left\{ \left(\begin{bmatrix} G_y & \\ & G_x \end{bmatrix} L \right)^+ \begin{bmatrix} C_l \\ I_l \end{bmatrix} \right\}^+ \left(\begin{bmatrix} G_y & \\ & G_x \end{bmatrix} L \right)^+ \begin{bmatrix} y_l \\ 0 \end{bmatrix} \tag{D.67}$$

$$= \left\{ L^{-1} \begin{bmatrix} G_y^+ & \\ & G_x^+ \end{bmatrix} \begin{bmatrix} C_l \\ I_l \end{bmatrix} \right\}^+ L^{-1} \begin{bmatrix} G_y^+ & \\ & G_x^+ \end{bmatrix} \begin{bmatrix} y_l \\ 0 \end{bmatrix} \tag{D.68}$$

$$= \left\{ L^{-1} \begin{bmatrix} G_y^+ C_l \\ G_x^+ \end{bmatrix} \right\}^+ L^{-1} \begin{bmatrix} G_y^+ y_l \\ 0 \end{bmatrix} \tag{D.69}$$

The difference between this estimator and the desired form in the original domain (D.36) is due to the undetermined observation matrix C_l in the lifted domain. The two expressions (D.36),(D.69) are equivalent under the following sufficient condition:

$$G_y^+ C_l = G_y^+ G_y C G_x^+ \tag{D.70}$$

Condition (D.70) is satisfied by observation matrices having the form

$$C_l = G_y C G_x^+ + \sum_m \beta_m \chi_m^T \qquad \beta_m \in Nu(G_y^+) \tag{D.71}$$

We need to determine conditions on G_x such that the set of satisfying C_l matrices in (D.71) include a selection matrix.

Consider the following product:

$$G_y C G_x^+ = G_y \begin{bmatrix} \delta_1 e_{i_1}^T \\ \delta_2 e_{i_2}^T \\ \vdots \end{bmatrix} G_x^+ = \sum_n g_n \cdot \delta_n \cdot h_{i_n}^T = \sum_n \xi_n \tag{D.72}$$

Since the regions of support of all of the columns of G_y are disjoint, it follows that the regions of support of all of the ξ_n matrices are disjoint. Thus the product $G_y C G_x^+$ can be transformed into a selection matrix as in (D.71) only if the same is true for each of the ξ_n individually.

For C_l to be a selection matrix, each row of C_l contains exactly one non-zero entry. Consider $\zeta_n = \xi_n + \sum_m \beta_m \chi_m^T$, where $\beta_m \in Nu(G_y^+)$; there are three cases to consider:

1. The size of the region of support of g_n is smaller than that of h_{i_n}.

   Since $g_n \notin Nu(G_y^+)$, ζ_n can have no columns equal to zero, and thus contains at least $size(support(h_{i_n}))$ non-zero elements; however for ζ_n to be a selection matrix, it can have at most $size(support(g_n))$ non-zero elements, leading to a contradiction.

2. The size of the region of support of g_n is greater than that of h_{i_n}.

   A selection matrix ζ_n having more rows than columns must either have repeated measurements (linearly dependent rows) or zero rows; neither of these possibilities present intrinsic difficulties, but these cases are not of particular interest, and we disallow them for the sake of brevity.

3. The size of the region of support of g_n is equal to that of h_{i_n} and equals M_n.

   Suppose that

$$g_n = \sum_{m=1}^{M_n} e_{f(n,m)} \qquad h_{i_n} = \sum_{m=1}^{M_n} \gamma_{n,m} e_{\rho(n,m)} \tag{D.73}$$

   where the $\gamma_{n,m}$ are a set of scalar weights determining the non-zero elements of h_{i_n}. Then it is easy to see that

$$M_n e_{f(n,m)} - g_n \in Nu(G_y^+) \qquad 1 \leq m \leq M_n \tag{D.74}$$

   and thus one possible choice of ζ_n becomes

$$\zeta_n = \sum_{m=1}^{M_n} M_n \delta_n \gamma_{n,m} e_{f(n,m)} e_{\rho(n,m)}^T \tag{D.75}$$

It should be noted that the above result does not imply that the regions of support of the rows of G_x^+ need to be disjoint.

There is one special case of interest which is the case considered in all of the applications of the overlapping framework throughout this thesis:

- The elements of C_l must equal their counterparts in C; i.e., the elements of ζ_n are either zero or δ_n. From (D.75) this implies that

$$\gamma_{n,m} = \frac{1}{M_n} \tag{D.76}$$

- The regions of support of the columns of G_x (i.e., the rows of G_x^+) are disjoint.

It follows immediately from (D.73), (D.76) that

$$h_{i_n} = \sum_{m=1}^{M_n} e_{\rho(n,m)}/M_n \tag{D.77}$$

which leads to a simple form for G_x^+, specifically

$$G_x^+ = \left[h_1^{+T} \;\; h_2^{+T} \;\; \ldots \right] \tag{D.78}$$

$$h_{i_n}^{+T} = \textstyle\sum_{m=1}^{M_n} e_{\rho(n,m)} \tag{D.79}$$

and a corresponding form for G_x:

$$G_x = \begin{bmatrix} h_1^T \\ h_2^T \\ \vdots \end{bmatrix} \tag{D.80}$$

for which satisfying matrices $H_x G_x = I$ (from Condition 1) are straightforward to calculate and visualize (the topic of Appendix E).

Estimation Error Covariance Calculation

Consider the estimation problem,

$$\begin{bmatrix} y \\ 0 \end{bmatrix} = \begin{bmatrix} C \\ I \end{bmatrix} x + v \tag{D.81}$$

and a proposed unbiased estimator,

$$\hat{x} = A \begin{bmatrix} y \\ 0 \end{bmatrix} \tag{D.82}$$

Since the estimator is unbiased,

$$E[x - \hat{x}] = E\left[x - A \begin{bmatrix} C \\ I \end{bmatrix} x + v\right] \tag{D.83}$$

$$= \left(I - A \begin{bmatrix} C \\ I \end{bmatrix}\right) x = 0 \tag{D.84}$$

that is,

$$A \begin{bmatrix} C \\ I \end{bmatrix} = I \tag{D.85}$$

The estimation error variance is given by

$$\tilde{P} = E\left[(\hat{x} - x)(\hat{x} - x)^T\right] \tag{D.86}$$

$$= E\left[\left(A \begin{bmatrix} y \\ 0 \end{bmatrix} - x\right)\left(A \begin{bmatrix} y \\ 0 \end{bmatrix} - x\right)^T\right] \tag{D.87}$$

$$= E\left[\left(\left(A \begin{bmatrix} C \\ I \end{bmatrix} - I\right) x + Av\right)\left(\left(A \begin{bmatrix} C \\ I \end{bmatrix} - I\right) x + Av\right)^T\right] \tag{D.88}$$

$$= AE\left[vv^T\right] A^T \tag{D.89}$$

where the last step follows from (D.85).

Now consider the application of (D.89) to our estimators in the original and lifted domains. From (D.19) the estimator in the original domain is characterized by

$$A = \left\{L^{-1} \begin{bmatrix} C \\ I \end{bmatrix}\right\}^+ L^{-1} \tag{D.90}$$

When all of the three earlier conditions (D.44), (D.59), (D.75) are satisfied, then the

estimator in the lifted domain, from (D.35), is characterized by

$$A_l = \left\{ \left\{ \begin{bmatrix} G_y & \\ & G_x \end{bmatrix} L \right\}^+ \begin{bmatrix} G_y C G_x^+ \\ G_x G_x^+ \end{bmatrix} \right\}^+ \left\{ \begin{bmatrix} G_y & \\ & G_x \end{bmatrix} L \right\}^+ \tag{D.91}$$

$$= \left\{ L^{-1} \begin{bmatrix} G_y^+ & \\ & G_x^+ \end{bmatrix} \begin{bmatrix} G_y C \\ G_x \end{bmatrix} G_x^+ \right\}^+ L^{-1} \begin{bmatrix} G_y^+ & \\ & G_x^+ \end{bmatrix} \tag{D.92}$$

$$= G_x \left\{ L^{-1} \begin{bmatrix} C \\ I_l \end{bmatrix} \right\}^+ L^{-1} \begin{bmatrix} G_y^+ & \\ & G_x^+ \end{bmatrix} \tag{D.93}$$

The above forms A, A_l for the estimators allows us to establish a relationship between the estimation error covariances in the original and lifted domains:

$$\tilde{P} = A \begin{bmatrix} R & \\ & P \end{bmatrix} A^T \tag{D.94}$$

$$= \left\{ L^{-1} \begin{bmatrix} C \\ I \end{bmatrix} \right\}^+ L^{-1} \begin{bmatrix} R & \\ & P \end{bmatrix} L^{-1} \left\{ L^{-1} \begin{bmatrix} C \\ I \end{bmatrix} \right\}^{+T} \tag{D.95}$$

$$= H_x G_x \left\{ L^{-1} \begin{bmatrix} C \\ I \end{bmatrix} \right\}^+ L^{-1} \begin{bmatrix} G_y^+ & \\ & G_x^+ \end{bmatrix} \begin{bmatrix} R_l & \\ & P_l \end{bmatrix} \begin{bmatrix} G_y^+ & \\ & G_x^+ \end{bmatrix}^T L^{-1} \left\{ L^{-1} \begin{bmatrix} C \\ I \end{bmatrix} \right\}^{+T} G_x^T H_x^T \tag{D.96}$$

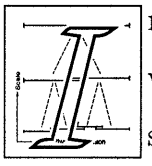
$$= H_x A_l \begin{bmatrix} R_l & \\ & P_l \end{bmatrix} A_l^T H_x^T \tag{D.97}$$

$$= H_x \tilde{P}_l H_x^T \tag{D.98}$$

That is, the estimation error covariance in the original domain is just the projected version of the error covariance determined by the estimator in the lifted domain.

Appendix E

Overlapping Framework Details



n this appendix we will make explicit two aspects of the overlapping framework: Section E.1 details how an overlapping tree structure $\mathcal{O}$ may be selected, and Section E.2 details how the projection operators G_x, G_y, H_x may be determined from $\mathcal{O}$.

Throughout this discussion we will make the assumption that the overlapping tree is regular: the tree lies in d-dimensional space, has M scales, and each node above the finest scale has $q = q_1 \cdot q_2 \cdot \ldots \cdot q_d$ offspring.[1] An example of a two-dimensional regular tree is shown in Figure E-1.

E.1 Overlap Structure Determination

As is suggested pictorially by Figure E-1, each of the nodes on the multiscale tree is associated with a bounded d-dimensional cube[2] in the original domain. Now consider the notation described in Figure E-2. The regularity of the tree implies that the cubes associated with the multiscale nodes on the same scale all have the same shape and size; for those cubes associated with nodes on scale m, we denote by $w_{m,i}$ the length of those cube edges parallel to dimension i. Next, the cubes associated with the children

[1] That is, the offspring of any node s (except for nodes on the finest scale) form a d-dimensional regular finite lattice having q_i lattice points along dimension i.

[2] That is, the logical extension of a cube to d-dimensional space: a solid, regular shape, such that any two faces of the cube are either parallel or oriented at right angles to each other.

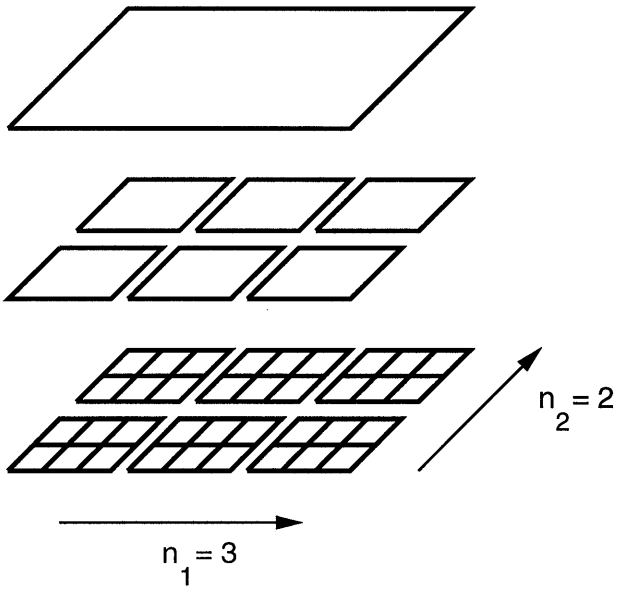


Figure E-1: A simple example of a two-dimensional regular tree.

of a common parent may overlap. Consider two multiscale nodes on scale m which have a common parent, and which are neighbors along dimension i,[3] then the amount of overlap, measured along dimension i, between the two cubes associated with these two nodes is denoted by $o_{m,i}$. It immediately follows that

$$w_{m-1,i} = q_i w_{m,i} - (q_i - 1) o_{m,i} \tag{E.1}$$

(a recursion which is made clear by Figure E-2). Furthermore, let

$$\psi_{m,i} = \frac{o_{m,i}}{w_{m,i}} \tag{E.2}$$

that is, $\psi_{m,i}$ represents the fractional overlap of neighboring nodes along dimension i on scale m.

Suppose we are given an "image" hypercube (i.e., an ensemble of pixels on a finite, regular lattice) in d-dimensions over which we wish to define our overlapping

[3]Suppose two nodes, s_1 and s_2, have a common parent $s = s_1\bar{\gamma} = s_2\bar{\gamma}$. The descendants of s form a d-dimensional lattice; s_1 and s_2 are neighbors along dimension i if they occupy positions in this lattice separated by one unit along dimension i and separated by zero units along all other dimensions.

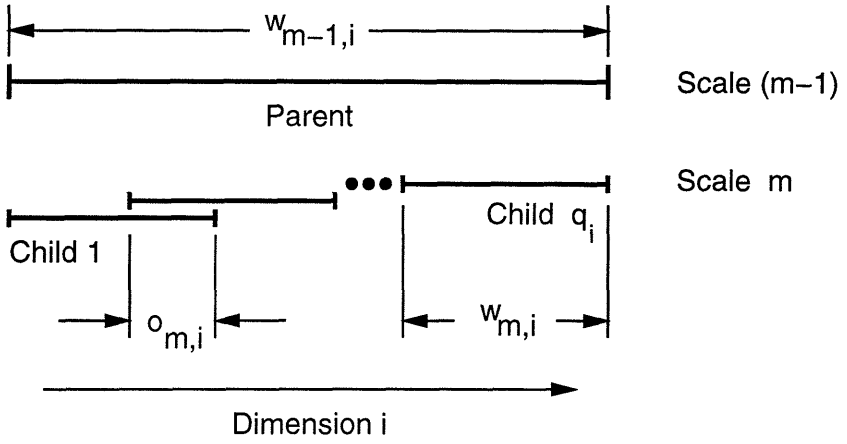


Figure E-2: The basic overlap notation required for the purpose of this appendix.

tree model, such that the image has N_i pixels along edges parallel to the axis of dimension i. Furthermore, suppose that we are also given an M-level multiscale tree having order

$$q = q_1 \cdot \ldots \cdot q_d, \tag{E.3}$$

We need to solve the following problem: determine a suitable overlapping framework, parameterized by $\mathcal{O} = \{o_{m,i}\}$,[4] which establishes a mapping between the multiscale tree and the image hypercube such that

1. The coarsest scale of the multiscale tree aggregates the whole image:

$$w_{0,i} = N_i \qquad 1 \leq i \leq d \tag{E.4}$$

2. The nodes on the finest scale of the multiscale tree correspond to individual image pixels:

$$w_{M-1,i} = 1 \qquad 1 \leq i \leq d \tag{E.5}$$

3. The fractional overlap (E.2) is nearly a constant function of scale for each di-

[4]If the overlap is not a function of dimension, i.e., $o_{m,i} = o_m$, then the parameterization simplifies to $\mathcal{O} = \{o_m\}$, which was the parameterization used in Chapter 7.

mension:

$$\psi_{m,i} \simeq \bar{\psi}_i \qquad 1 \leq m \leq M-1, \quad 1 \leq i \leq d \tag{E.6}$$

The number of scales in the multiscale tree, M, regulates the tradeoff between smoothness and computational effort. M may be freely chosen within the following constraints:

- The multiscale tree must have a sufficient number of scales to be able to represent the whole image cube:[5]

$$M \geq \max_i \left\lceil 1 + \log_{q_i} N_i \right\rceil \tag{E.7}$$

- The image regions aggregated by neighboring nodes cannot completely overlap, i.e., $o_{m,i} < w_{m,i}$; thus

$$w_{m-1,i} \geq w_{m,i} + q_i - 1 \tag{E.8}$$

$$w_{0,i} = N_i \geq 1 + (M-1)(q_i - 1) \tag{E.9}$$

$$M \leq \min_i \left(\frac{N_i - 1}{q_i - 1} + 1 \right) \tag{E.10}$$

Given a value of M satisfying (E.7),(E.10) it is relatively easy to determine $\{o_{m,i}\}$ to satisfy (E.4)-(E.6):[6]

$$\left.\begin{array}{l} \text{Let} \quad w_{M-1,i} = 0 \\ \text{For} \quad m = M-1, M-2, \ldots, 1 \text{ in decreasing order :} \\ \qquad \hat{o}_i = \left\lfloor \dfrac{q_i w_{m,i} - w_{m,i}\left(\frac{N_i}{w_{m,i}}\right)^{\frac{1}{m}}}{q_i - 1} \right\rfloor \\ \qquad o_{m,i} = \min\left\{(w_{m,i} - 1), \max\left[0, \hat{o}_i\right]\right\} \\ \qquad w_{m-1,i} = q_i w_{m,i} - (q_i - 1) o_{m,i} \end{array}\right\} 1 \leq i \leq d \tag{E.11}$$

That is, ideally $\bar{\psi}_i = N_i^{1/(M-1)}$; this ideal value for $\bar{\psi}_i$ is asserted in the expression for

[5] $\lceil x \rceil$ represents the smallest integer equal to or greater than x.

[6] Similarly, $\lfloor x \rfloor$ represents the greatest integer equal to or less than x.

$\hat{o}_i$; the approximation comes about due to the constraint that $o_{m,i}$ be integer.

The parameters $\mathcal{O} = \{o_{m,i}\}$ are a parameterization of the overlapping tree, sufficient to guarantee a unique tree structure if the regularity assumptions are met. The next section will describe the determination of the projection operators from $\mathcal{O}$.

E.2 Projection Operator Determination

This section will specify explicit definitions of G_x, G_y and H_x. We will first consider the case of one-dimensional trees, from which the extension to d-dimensional trees will be straightforward due to the assumed regularity of the trees. The material of this section is similar to that of Section 6.3, however there are two differences:

1. The notation is extended to allow d-dimensional trees.

2. The definition of H_x is more general and explicit.

Let $\mathcal{O} = (o_1, o_2, \ldots, o_{M-1})$ represent the overlapping tree structure for a one-dimensional tree with order q. Furthermore, suppose that the observation matrix in the original domain is the following weighted selection matrix:

$$C = \begin{bmatrix} \beta_1 e_{c_1}^T \\ \vdots \\ \beta_K e_{c_K}^T \end{bmatrix} \tag{E.12}$$

where e_i represents the ith unit vector.

Consider any node s on the finest scale of the tree; we can explicitly write s as a descendant of the root node as

$$s = 0\alpha_{j_1}\alpha_{j_2}\ldots\alpha_{j_{M-1}} \qquad j_m \in \{1, \ldots, q\} \tag{E.13}$$

i.e., the vector $\boldsymbol{j} = (j_1, \ldots, j_{M-1})$ uniquely parameterizes s. Let

$$\bar{p}(\boldsymbol{j}, q) = \sum_{m=1}^{M-1} (j_m - 1) q^{M-m-1} \tag{E.14}$$

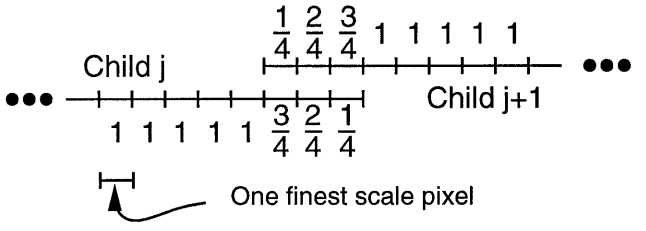


Figure E-3: The relative weights applied to two neighboring, overlapping regions should vary linearly over the overlapping portion, as shown.

$$p(\boldsymbol{j}) = \sum_{m=1}^{M-1} (j_m - 1)(w_m - o_m) \tag{E.15}$$

where the explicit parameterization in terms of $\boldsymbol{j}$ and q (as opposed to (6.41) and (6.42)) will be necessary when we consider the extension to d-dimensional trees. From (E.14) and (E.15), $\bar{p}$ represents the position of s along the finest scale of the tree, and node s corresponds to the pth element of the original one-dimensional process being represented by the overlapping tree (i.e., $\bar{p}$ indexes the elements of x_l, and p indexes the elements of x). With these interpretations of $\bar{p}, p$ in mind, the projection matrix G_x is given by

$$G_x(a,b) = \begin{cases} 1 & \text{if } \exists \boldsymbol{j} \ni \bar{p}(\boldsymbol{j},q) = a, p(\boldsymbol{j}) = b \\ 0 & \text{otherwise} \end{cases} \tag{E.16}$$

and the corresponding measurement projection operator G_y is given by

$$G_y(a,b) = G_x(a,c_b) \qquad \begin{array}{l} 1 \le a \le q^{M-1} \\ 1 \le b \le K \end{array} \tag{E.17}$$

that is, G_y is just a reordered subset of the columns of G_x, with one column selected per measurement.

The smooth projection operator H_x used throughout this thesis is one in which the relative weights, applied to neighboring overlapping tree nodes, taper linearly across the overlapping region, as sketched in Figure E-3. A precise definition of H_x is as follows. Let

$$\delta_{M-1} = 0 \qquad \delta_{m-1} = (j_m - 1)(w_m - o_m) + \delta_m \tag{E.18}$$

$$h_m(\boldsymbol{j},q)=\begin{cases}1 & j_m=1; \quad \delta_m < w_m - o_m \\ \frac{w-\delta_m}{o_m+1} & j_m=1; \quad \delta_m \geq w_m - o_m \\ 1 & 1<j_m<q; \quad o_m \leq \delta_m < w_m - o_m \\ \frac{w-\delta_m}{o_m+1} & 1<j_m<q; \quad \delta_m \geq \max(o_m, w_m - o_m) \\ \frac{\delta_m+1}{o_m+1} & 1<j_m<q; \quad \delta_m < \min(o_m, w_m - o_m) \\ \frac{w_m-o_m}{o_m+1} & 1<j_m<q; \quad w_m - o_m \leq \delta_m < o_m \\ 1 & j_m=q; \quad \delta_m \geq o_m \\ \frac{\delta_m+1}{o_m+1} & j_m=q; \quad \delta_m < o_m \end{cases} \tag{E.19}$$

(the above expression for h_m is essentially a mathematical description of the weights in Figure E-3) then the smoothing matrix H_x is given by

$$H_x(b,a)=\begin{cases}\prod_{m=1}^{M-1} h_m(\boldsymbol{j},q) & \text{if } \exists \boldsymbol{j} \ni \bar{p}(\boldsymbol{j},q)=a, p(\boldsymbol{j})=b \\ 0 & \text{otherwise}\end{cases} \tag{E.20}$$

This completes the specifications of the projection operators for the case of a regular one-dimensional tree. The extension of these results to the case of d-dimensional trees is straightforward when such trees are viewed as a product of d one-dimensional trees. As in (E.13), we describe each node s on the finest scale explicitly as a descendant of the root note:

$$s = 0\alpha_{j_{1,1}j_{1,2}\cdots j_{1,d}}\alpha_{j_{2,1}\cdots j_{2,d}}\cdots\alpha_{j_{M-1,1}\cdots j_{M-1,d}} \qquad j_{m,i}\in\{1,\ldots,q_i\} \tag{E.21}$$

where $\alpha_{j_{m,1}\cdots j_{m,d}}$ represents a d-dimensional descendant operator, specifying the descendant at position $(j_{m,1},\ldots,j_{m,d})$ in the d-dimensional finite lattice of descendants of each node; i.e., the operator specifies the $j_{m,i}$th position along each dimension i. Now consider the projection of the descent (E.21) onto dimension i; then (E.21) would look like the following sequence of one-dimensional descent steps

$$\alpha_{j_{1,i}}\alpha_{j_{2,i}}\cdots\alpha_{j_{M-1,i}} \tag{E.22}$$

Since this holds for each dimension $1\leq i\leq d$, we see that a node s on the finest scale of the d-dimensional tree can be indexed as a d-tuple of one-dimensional tree indices.

Specifically, define the following $(M-1)$-tuples:

$$\begin{aligned} \boldsymbol{j}_1 &= (j_{1,1}, j_{2,1}, \ldots, j_{M-1,1}) \\ &\vdots \\ \boldsymbol{j}_d &= (j_{1,d}, j_{2,d}, \ldots, j_{M-1,d}) \end{aligned} \tag{E.23}$$

We can use these vectors to define the operators G_x, H_x, and G_y. G_x and H_x are now projection operators from one d-dimensional space to another d-dimensional space, thus G_x and H_x are no longer matrices; instead, $G_x(\boldsymbol{a}, \boldsymbol{b})$ and $H_x(\boldsymbol{b}, \boldsymbol{a})$ are real functions of two discrete, d-dimensional vectors $\boldsymbol{a}, \boldsymbol{b}$:

$$G_x(\boldsymbol{a}, \boldsymbol{b}) = \begin{cases} 1 & \text{if } \exists\{\boldsymbol{j}_i\} \ni \begin{matrix} \{\bar{p}(\boldsymbol{j}_1, q_1), \ldots, \bar{p}(\boldsymbol{j}_d, q_d)\} = \boldsymbol{a} \\ \{p(\boldsymbol{j}_1), \ldots, p(\boldsymbol{j}_d)\} = \boldsymbol{b} \end{matrix} \\ 0 & \text{otherwise} \end{cases} \tag{E.24}$$

$$H_x(\boldsymbol{b}, \boldsymbol{a}) = \begin{cases} \prod_{i=1}^{M-1} h_m(\boldsymbol{j}_1, q_1) \cdot \ldots \cdot h_m(\boldsymbol{j}_d, q_d) & \text{if } \exists\{\boldsymbol{j}_i\} \ni \begin{matrix} \{\bar{p}(\boldsymbol{j}_1, q_1), \ldots \bar{p}(\boldsymbol{j}_d, q_d)\} = \boldsymbol{a} \\ \{p(\boldsymbol{j}_1), \ldots p(\boldsymbol{j}_d)\} = \boldsymbol{b} \end{matrix} \\ 0 & \text{otherwise} \end{cases} \tag{E.25}$$

Furthermore, under the assumption that the d-dimensional observation operator is a weighted selection operator (where the operator is a mapping from a d-dimensional lattice to a one-dimensional vector of observations), the operator can be written in the following product form:

$$C = \begin{bmatrix} \beta_1 e^T_{c_{1,1}} \\ \vdots \\ \beta_K e^T_{c_{K,1}} \end{bmatrix} \times \begin{bmatrix} e^T_{c_{1,2}} \\ \vdots \\ e^T_{c_{K,2}} \end{bmatrix} \times \ldots \times \begin{bmatrix} e^T_{c_{1,d}} \\ \vdots \\ e^T_{c_{K,d}} \end{bmatrix} \tag{E.26}$$

The operator G_y, viewed as a projection from a one-dimensional vector of observations into d-dimensional space, is then defined as

$$G_y(\boldsymbol{a}, b) = G_x(\boldsymbol{a}, (c_{b,1}, c_{b,2}, \ldots, c_{b,d})) \qquad \begin{matrix} \forall \boldsymbol{a} \in [1 \ldots q_1] \times \ldots \times [1 \ldots q_d] \\ \forall 1 \leq b \leq K \end{matrix} \tag{E.27}$$

Bibliography

[1] H. Akaike. "Markovian Representation of Stochastic Processes by Canonical Variables", *SIAM Journal of Control* (13) #1, pp.162–173, 1975.

[2] A. Albert, *Regression and the Moore-Penrose Pseudoinverse*, Academic Press, New York, 1972

[3] T. Anderson, *The Statistical Analysis of Time Series*, Wiley & Sons, New York, 1971

[4] B. Anderson, J. Moore, *Optimal Filtering*, Prentice-Hall, New Jersey, 1979

[5] Ballard, Hinton, Sejnowski, "Parallel Computation in Vision Problems", *Nature* (306) #5938, pp.21–26, 1983

[6] M. Basseville, A. Benveniste, K. Chou, S. Golden, R. Nikoukhah, A. Willsky, "Modeling and estimations of multiresolution stochastic processes.", *IEEE Transactions on Information Theory* (38) #2, pp.766–784, 1992

[7] J. Besag, "Spatial Interaction and the Statistical Analysis of Lattice Systems", *J. Royal Society, Series E* (36), pp.192–236, 1974

[8] M. Bertero, T. Poggio, V. Torre, "Ill-Posed Problems in Early Vision", *Proc. IEEE* (76) #8, pp.869–889, 1988

[9] A. Brandt, "Multi-Level Adaptive Solutions to Boundary-Value Problems", *Mathematics of Computation* (31) #138, pp.333–390, 1977

[10] M. Brooks, "Surface Normals from Closed Paths", *Proc. 6th IJCAI Tokyo*, pp.98–101, 1979

[11] M. Brooks, B. Horn, "Shape and Source from Shading", *MIT A.I. Tech. Report 820*, 1985

[12] D. Chelton, M. Schlax, D. Witter, J. Richman, "Geosat Altimeter Observations of the Surface Circulation of the Southern Ocean", *J. Geophys. Research* (95), pp.17877–17903, 1990

[13] K.C. Chou, A.S. Willsky, A. Benveniste, and M. Basseville. "Recursive and iterative estimation algorithms for multiresolution stochastic processes." *Proc. IEEE Conference on Decision and Control*, 1989.

[14] K. Chou, "A Stochastic Modeling Approach to Multiscale Signal Processing", PhD Thesis, Dept. of EECS, Massachusetts Institute of Technology, 1991

[15] K. Chou, A. Willsky, A. Benveniste, "Multiscale Recursive Estimation, Data Fusion, and Regularization", *IEEE Trans. on Automatic Control* (39) #3, pp.464–478, 1994

[16] K. Chou, A. Willsky, A. Benveniste, "Multiscale Systems, Kalman Filters, and Riccati Equations", *IEEE Trans. on Automatic Control* (39) #3, pp.479–492, 1994

[17] A. Cook, *Gravity and the Earth*, Wykeham Science Series, Wykeham Publications, London, 1969

[18] R. Courant, D. Hilbert, *Methods of Mathematical Physics V1*, Interscience Publishers, New York, 1953

[19] Dahlquist, Bjorck, *Numerical Methods*, Prentice-Hall, New Jersey, 1974

[20] R. Daley, *Atmospheric Data Analysis*, Cambridge University Press, New York, 1991

[21] H. Denker, R. Rapp, "Geodetic and Oceanographic Results from the Analysis of 1 Year of Geosat Data", *J. Geophys. Research* (95) #C8, pp.13151–13168, 1990

[22] H. Derin, P. Kelly, "Discrete-Index Markov-Type Random Processes", *Proc. IEEE* (77) #10, pp.1485–1510, 1989

[23] T. Engelis, "On the Simultaneous Improvement of a Satellite Orbit and Determination of Sea Surface Topography using Altimetric Data", *Manus. Geod* (13), pp.180–190, 1988

[24] T. Engelis, P. Knudsen, "Orbit Improvement and Determination of the Ocean Geoid and Topography from 17 Days of Seasat Data", *Manus. Geod.* (14), pp.193–200, 1989

[25] W. Fieguth, "Multi-Input Quasi-Linearization", Ph.D. Thesis, Dept. of Electrical Engineering, University of New Brunswick, 1967

[26] P. Fieguth, A. Willsky, W. Karl, "Multiresolution Stochastic Imaging of Satellite Oceanographic Altimetric Data." *Proc. IEEE International Conference on Image Processing* (II), pp.1–5, 1994.

[27] P. Fieguth, W. Karl, A. Willsky, C. Wunsch, "Multiresolution Optimal Interpolation and Statistical Analysis of TOPEX/POSEIDON Satellite Altimetry", *IEEE Trans. Geoscience and Remote Sensing* (33) #2, pp.280–292, 1995

[28] P. Fieguth, "Application of Multiscale Estimation to Large Scale Multidimensional Imaging and Remote Sensing Problems", PhD Thesis, Dept. of EECS, Massachusetts Institute of Technology, 1995

[29] P. Flandrin, "Wavelet Analysis and Synthesis of Fractional Brownian Motion", *IEEE Trans. Information Theory* (38) #2, pp.910–917, 1992

[30] R. Forsberg, "Modeling the Fine Structure of the Geoid: Methods, Data Requirements, and Some Results", *Surveys in Geophysics* (14) #4-5, pp.403–418, 1993

[31] R. Franktot, R. Chellappa, "A Method for Enforcing Integrability in Shape from Shading Algorithms", *IEEE PAMI* (10) #4, pp.439–451, 1989

[32] L. Fu, E. Christensen, C. Yamarone, M. Lefebvre, Y. Menard, M. Dorrer, P. Escudier, "TOPEX/POSEIDON mission overview", *J. Geophys. Res.* (99) #C12, pp.24369–24381, 1994,

[33] P. Gaspar, C. Wunsch, "Estimates from Altimeter Data of Baryotropic Rossby Waves in the Northwestern Atlantic Ocean", *J. Physical Oceanography* (19)#12, pp.1821–1844, 1989

[34] S. Gelfand, S. Mitter, "On Sampling Methods and Annealing Algorithms" in *Markov Random Fields - Theory and Application* (R. Chellappa, A. Jain ed.s), pp. 499–515, 1993

[35] S. Geman, D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images", *IEEE Trans. PAMI* (6) #6, pp. 721–741, 1984

[36] A. George, J. Liu, *Computer Solution of Large and Sparse Positive Definite Systems*, Prentice Hall, Englewood Cliffs, NJ, 1981

[37] M. Ghil, P. Malanotti-Rizzoli, "Data Assimilation in Meteorology and Oceanography", *Advances in Geophysics* (33), pp.141–266, 1991

[38] J. Goodman, A. Sokal, "Multigrid Monte Carlo method. Conceptual foundations", *Physical Review D* (40) #6, pp. 2035–2071, 1989

[39] W. Hackbusch, *Multi-Grid Methods and Applications*, Springer-Verlag, Berlin, 1985

[40] J. Harris, "The Coupled Depth/Slope Approach to Surface Reconstruction", MIT S.M. Thesis, 1986

[41] W. Heiskanen, F. Meinesz, *The Earth and Its Gravity*, McGraw Hill, New York, 1958

[42] W. Heiskanen, H. Moritz, *Physical Geodesy*, W.H. Freemand & Co., San Francisco, 1967

[43] B. Horn, M. Brooks, "Integrability of Surface Gradients", *MIT A.I. Memo 813*, 1985

[44] B. Horn, *Robot Vision*, MIT Press, 1986

[45] B. Horn, "Height and Gradient from Shading", *Int'l Journal of Comput. Vision* (5) #1, pp.37–46, 1990

[46] Ikeuchi, B. Horn, "Parallelizable Shape from Shading", *Artificial Intelligence* (17) #1-3, pp.141–186, 1981

[47] W. Irving, W. Karl, A. Willsky, "A Theory for Multiscale Stochastic Realization", *33rd Conference on Decision and Control*, Orlando, FL, pp.655–662, 1994

[48] W. Irving, A. Willsky, "A Theory for Multiscale Stochastic Realization of Gaussian Random Processes and Fields", In Preparation

[49] L. Kaplan, C.C. Kuo, "Fractal Estimation from Noisy Data via Discrete Fractional Gaussian Noise (DFGN) and the Haar Basis", *IEEE Trans. S.P.* (41) #12, pp.3554–3562, 1993

[50] W. Kaula, *Theory of Satellite Geodesy*, Blaisdell Publishing Co., Waltham MA, 1966

[51] D. Keren, M. Werman, "Probabilistic Analysis of Regularization", *IEEE PAMI* (15) #10, pp.982–995, 1993

[52] C. King, D. Stammer, C. Wunsch, "The CMPO/MIT Topex/Poseidon Altimetric Data Set", *MIT CGCS Report #30*, 1994

[53] B. Kosko (editor) *Neural Networks for Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, pp.37–61, 1992

[54] W. Larimore, S. Thomas, S. Baumgartner, "Statistical Modeling of Gravity Disturbances using Measured Anomaly Data", *TASC Tech. Report TR-776-1*, Reading MA, 1977

[55] K. Lee, C. Kuo, "Shape from Shading with a Linear Triangular Element Surface Model", *IEEE PAMI* (15) #8, pp.815–822, 1993

[56] K. Lee, C. Kuo, "Surface Reconstruction from Photometric Stereo Images", *JOSA-A* (10) #5, pp.855–868, 1993

[57] D. Lee, J. Shiau, "Thin Plate Splines with Discontinuities and Fast Algorithms for Their Computation", *SIAM J. on Scientific Computing* (15) #6, pp.1311–1330, 1994

[58] S. Levitus, "Professional Paper 13", *Climatological Atlas of the World Ocean*, NOAA

[59] B. Levy, "Non-causal estimation for Markov random fields", In *Proc. of the Int'l Symposium MTNS-89, Vol 1* (M. Kaachoek, J. Schuppen, A. Ran ed.s), Birkhauser-Verlag, 1990

[60] L. Ljung, *System identification : theory for the user*, Prentice-Hall, Englewood Cliffs, NJ, 1987

[61] M. Luettgen, W. Karl, A. Willsky, R. Tenney, "Multiscale Representations of Markov Random Fields", *IEEE Trans. Signal Processing* (41) #12, pp.3377–3396, 1993

[62] M. Luettgen, "Image Processing with Multiscale Stochastic Models", PhD Thesis, Dept. of EECS, Massachusetts Institute of Technology, 1993

[63] M. Luettgen, W. Karl, A. Willsky, "Efficient Multiscale Regularization with Applications to the Computation of Optical Flow", *IEEE Trans. on Image Processing* (3) #1, pp.41–64, 1994.

[64] M. Luettgen, A. Willsky, "Multiscale Smoothing Error Models", *IEEE Trans. on Automatic Control* (40) #1, 1995

[65] M. Luettgen, A. S. Willsky, "Likelihood Calculation for a Class of Multiscale Stochastic Models, with Application to Texture Discrimination.", *IEEE Trans. Image Processing* (4) #2, pp.194–207, 1995

[66] T. Lundahl, W. Ohley, S. Kay, R. Siffert, "Fractional Brownian motion: A maximum likelihood estimator and its application to image texture", *IEEE Trans. Med. Im.* (5), pp.152–161, 1986

[67] P. Malanotte-Rizzoli, "Data Assimilation: Fundamentals, Global and Mediterranean Examples", *Ocean Processes in Climate Dynamics: Global and Mediterranean Examples* (P. Malanotte-Rizzoli, A. Robinson ed.s), NATO Asi Series (419), 1994

[68] S. Mallat, "A theory of multiresolution signal decomposition: The wavelet representation", *IEEE Trans. PAMI* (11) #7, pp.674–693, 1989

[69] B. Mandelbrot, J. van Ness, "Fractional Brownian motions, fractional noises and applications", *SIAM Review* (10), pp.422–437, 1968

[70] B. Mandelbrot, The Fractal Geometry of Nature, W.H. Freeman & Co., San Francisco, 1983

[71] J. Marsh et al., "The GEM-T2 Gravitational Model", *J. Geophysical Research* (95) #B13, pp.22043–22071, 1990

[72] J. Marsh et al., "Dynamic Sea Surface Topography, Gravity, and Improved Orbit Accuracies from the Direct Evaluation of Seasat Altimeter Data", *J. Geophys. Research* (95) #C8, pp.13129–13150, 1990

[73] P. Marth et al., "Prelaunch Performance of the NASA Altimeter for the TOPEX/POSEIDON Project", *IEEE Trans. on Geoscience and Remote Sensing* (31) #2, pp.315–332, 1993

[74] S. McCormick, *Multilevel Adaptive Mtehods for Partial Differential Equations*, SIAM, Philadelphia, 1989

[75] F. Monaldo, "Topex Ionospheric Height Correction Precision Estimated from Prelaunch Test Results", *IEEE Trans. Geoscience and Remote Sensing* (31) #2, pp.371–375, 1993

[76] R. Nash, S. Jordan, "Statistical Geodesy – An Engineering Perspective", *Proc. IEEE* (66) #5, pp.532–550, 1978

[77] "Worldwide Gridded Bathymetry DBDB5", Prepared by the Defense Mapping Agency, National Geophysical Data Center, 1985

[78] R. Nerem et al., "Expected Orbit Determination Performance for the Topex/Poseidon Mission", *IEEE Trans. Geoscience and Remote Sensing* (31) #2, pp.333–354, 1993

[79] R. Nerem et al., "Gravity model development for Topex/Poseidon: Joint gravity model 1 and 2", *J. Geophys. Res* (99) #C12, pp.24421–24447, 1994

[80] R. Nerem et al, "Ocean Dynamic Topography from Satellite Altimetry based on the GEM-T3 Gravity Model", *Manus. Geod.* (19), pp.346–366, 1994

[81] J. Nystuen, C. Andrade, "Tracking Mesoscale Ocean Features in the Caribbean Sea Using Geosat Altimetry", *J. Geophys. Research* (98)#C5, pp.8389–8394, 1993

[82] N. Pavlis, R. Rapp, "The Development of an Isostatic Gravitational Model to Degree 360 and its use in Global Gravity Modeling", *Geophys. J. International* (100), pp.369–378, 1990

[83] S. Peleg, G. Ron, "Nonlinear Multiresolution: A Shape from Shading Example", *IEEE PAMI* (12) #12, pp.1206–1210, 1990

[84] "Shape Information from Shading: A Theory of Human Perception", *IEEE 2nd Int'l Conf. on Comput. Vision*, pp.404–413, 1988

[85] T. Poggio, "Regularization Approaches in Early Vision Problems", *MIT A.I. Memo 773*, 1984

[86] R. Rapp, N. Pavlis, "The Development and Analysis of Geopotential Coefficient Models to Spherical Harmonic Degree 360", *J. Geophys. Research* (95) #B13, pp.21885–21911, 1990

[87] R. Rapp, Y. Wang, N. Pavlis, "The Ohio State 1991 Geopotential and Sea Surface Topography Harmonic Coefficient Models", Report #410, Dept. of Geodetic Science and Surveying, Ohio State University, 1991

[88] R. Rapp, "Geoid Undulation Accuracy", *IEEE Trans. Geoscience and Remote Sensing* (31) #2, pp.365–370, 1993

[89] H. Rauch, F. Tung, C. Striebel, "Maximum Likelihood Estimates of Linear Dynamic Systems",*AIAA Journal*, (3) #8, 1965

[90] R. Ray, "Global Ocean Tide Models on the Eve of Topex/Poseidon", *IEEE Trans. Geoscience and Remote Sensing* (31) #2, pp.355–364, 1993

[91] B. Ripley, *Spatial Statistics*, Wiley, 1991

[92] E. Rodriguez, J. Martin, "Correlation Properties of Ocean Altimeter Returns", *IEEE Trans. Geoscience and Remote Sensing* (32) #3, pp.553–561, 1994

[93] D. Saupe, "Algorithms for Random Fractals", in *The Science of Fractal Images* (Peitgen, Saupe Ed.s), Springer Verlag, 1988

[94] T. Schlatter, G. Branstator, I. Ihiel, "Testing a Global Multivariate Statistical Objective Analysis Scheme with Observed Data", *Mon. Weather Review* (104), pp.765–783, 1976

[95] D. Stammer, C. Wunsch, "Preliminary assessment of the accuracy and precision of TOPEX/POSEIDON altimeter data with respect to the large scale ocean circulation", *J. Geophys. Res.*, 1994, in press

[96] T. Strat, *A Numerical Method for Shape from Shading from a Single Image*, MIT S.M. Thesis, 1979

[97] *Kuroshio: Physical Aspects of the Japan Current* (H. Stommel, K. Yoshida ed.s), University of Washington Press, 1972

[98] R. Szeliski, "Regularization uses Fractal Priors", *Proc. National Conference on Artificial Intelligence*, pp.749–754, 1987

[99] R. Szeliski, *Bayesian Modeling of Uncertainty in Low-level Vision*, Kluwer Academic, 1989

[100] R. Szeliski, "Fast Surface Interpolation Using Hierarchical Basis Functions", *IEEE PAMI* (12) #6, pp.513–528, 1990

[101] B. Tapley et al., "Determination of the General Circulation of the Oceans from a Joint Gravity Field Solution", *Geophys. Research Letters* (15) #10, pp.1109–1112, 1988

[102] D. Terzopoulos, "Multilevel Computer Processes for Visual Surface Reconstruction", *Computer Vision, Graphics, Image Processing* (24) pp.52–96, 1983

[103] D. Terzopoulos, "Regularization of Inverse Visual Problems Involving Discontinuities", *IEEE PAMI* (8) #4, pp.413–424, 1986

[104] A. Tewfik, M. Kim, "Correlation Structure of the Discrete Wavelet Coefficients of Fractional Brownian Motion", *IEEE Trans. Information Theory* (38) #2, pp.904–909, 1992

[105] A. Tewfik, M. Deriche, "Maximum Likelihood Estimation of the Parameters of Discrete Fractionally Differenced Gaussian Noise Processes", *IEEE Trans. Signal Processing* (41) #10, pp.2977–2989, 1993

[106] R. Tokmakian, P. Challenor, "Observations in the Canary Basis in and Azores Frontal Region Using Geosat Data", *J. Geophys. Research* (98)#C3, pp.4761–4773, 1993

[107] C. Tscherning, R. Forsberg, "Harmonic Continuation and Gridding Effects on Geoid Height Prediction", *Bulletin Geodesique* (66) #1, pp.41–53

[108] R. Voss, "Fractals in Nature", in *The Science of Fractal Images* (Peitgen, Saupe Ed.s), Springer Verlag, 1988

[109] G. Wabha, *Spline Models for Observational Data*, Series in Applied Mathematics #59, SIAM, 1990

[110] P. Wesseling, *An Introduction to Multigrid Methods*, Wiley & Sons, London, 1991

[111] G. Whitten, "Scale Space Tracking and Deformable Sheet Models for Computational Vision", *IEEE PAMI* (15) #7, 1993

[112] K. Wilson, "Problems in Physics with Many Scales of Length", *Scientific American* (241), pp. 158–179, 1979

[113] G. Wornell, "Synthesis, Analysis, and Processing of Fractal Signals", PhD Thesis, Dept. of EECS, Massachusetts Institute of Technology, 1991

[114] G. Wornell, A. Oppenheim, "Estimation of fractal signals from noisy measurements using wavelets", *IEEE Trans. Signal Processing* (40), pp.611–623, 1992

[115] G. Wornell, "Wavelet-Based Representation for the $1/f$ Family of Fractal Processes", *Proc. IEEE*, Sept. 1993

[116] C. Wunsch, E. Gaposchkin, "On Using Satellite Altimetry to Determine the General Circulation of the Oceans with Application to Geoid Improvement", *Reviews of Geophysics and Space Physics* (18) #4, pp.725–745, 1980

[117] C. Wunsch, "Sampling Characteristics of Satellite Orbits", *J. of Atmospheric and Oceanic Tech.* (6)#6, pp.891–907, 1989

[118] C. Wunsch, D. Stammer, "The Global Frequency-Wavenumber Spectrum of Oceanic Variability Estimated from TOPEX / POSEIDON Altimetric Measurements", In preparation.

[119] D. Young, *Iterative Solution of Large Linear Systems*, Academic Press, New York, 1971

[120] Q. Zheng, R. Chellappa, "Estimation of Illuminant Direction, Albedo, and Shape from Shading", *IEEE PAMI* (13) #7, pp.680–702, 1991