Foveated Multiscale Models for Large-Scale Estimation

Paul W. Fieguth
Department of Systems Design
University of Waterloo
Waterloo, Ontario, Canada

pfieguth@uwaterloo.ca
http://ocho.uwaterloo.ca/~pfieguth

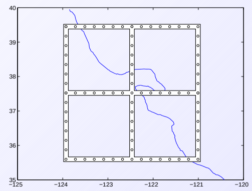
Funding:
ONR Grant N00014-91-J-1004
NSERC Canada

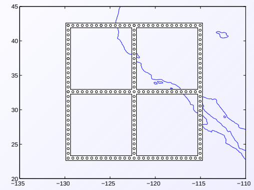
Introduction

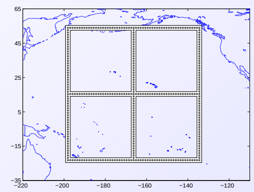
Goals:

- 3D statistical estimation:
  1. Static 3D — three space dimensions
  2. Dynamic 2D — two space & one time

Challenge:







For how large a domain (local, meso, global) is divide-and-conquer computationally and numerically feasible ...?

- Even for efficient methods, computational complexity is $\mathcal{O}(n^{2d-3})$ per pixel, for $n \times \ldots \times n$ pixels in d dimensions
- Past approximate models have not changed asymptotic complexity

Summary of Approach:

- Continue to use the divide-and-conquer, but in place of a single model with strict conditional-decorrelation assumptions, we propose multiple space-varying models with weaker assumptions.

Divide and Conquer

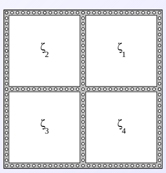
Basic Principle:

Given: 2D process ζ (e.g., ocean surface)

Find a subset ζ_o such that ζ is separated into conditionally decorrelated regions $\zeta_1, \ldots, \zeta_n$:

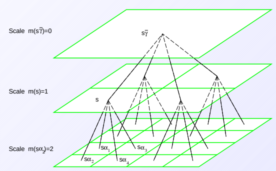
$$E[\zeta_i \zeta_j | \zeta_o] = E[\zeta_i | \zeta_o] \cdot E[\zeta_j | \zeta_o]$$

For example, a 2D first-order Markov random field can be separated into four quadrants:



Multiscale Model:

Can realize divide-and-conquer models in many forms (multiscale, nested-dissection). For purpose of discussion, we propose to realize ζ on the finest scale of a multiscale quad-tree.



Similarly we can model a 3D process on an oct-tree etc.

Tree Notation:

s Indexes nodes on the tree.
$s\bar{\gamma}$ Parent of tree node s.
0 Root node of the tree.

Tree Process:

$$x(0) \sim \mathcal{N}(x_o, P_o)$$
$$x(s) = A(s)x(s\bar{\gamma}) + B(s)w(s)$$
$$w(s) \sim \mathcal{N}(0, I)$$

That is, we define a process $x(s)$ which is autoregressive in *scale*.

Measurements:

$$y(s) = C(s)x(s) + v(s)$$
$$v(s) \sim \mathcal{N}(0, R(s))$$
$$\langle v(s)v(\bar{s})^T \rangle = \delta_{s,\bar{s}} R(s)$$

We will have only finest-scale measurements, but coarser-scale measurements are permitted.

Conditional Decorrelation:

State Model $x(s) = A(s)x(s\bar{\gamma}) + B(s)w(s)$

$w(s)$ is a white driving process
So the parent state $x(s\bar{\gamma})$ must conditionally decorrelate its children
(i.e., a Markov process in scale)

In 2D, column(s) of pixels are required to decorrelate halves of a plane.

In 3D, *plane(s)* of pixels are required for decorrelation etc.

Problems:

1. Numerical Stability:

As the number of scales ($= 1 + \log_2 n$) increases, the condition number of the covariance of boundary ζ_o deteriorates:

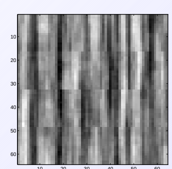
#	# States per Correlation Length					
Scales	1.0	2.0	3.0	4.0	5.0	6.0
5	10^4	10^5	10^4	10^9	10^9	10^9
6	10^3	10^7	10^7	10^9	10^9	10^{13}
7	10^5	10^6	10^{10}	10^{10}	10^{13}	
8	10^3	10^6	10^{10}			
9	10^3	10^7	10^{12}			
10	10^2	10^8			Unstable	
11	10^2	10^8			Problems	
12	10^3	10^9				

As problem size increases, ever cruder reduced-state approximations are needed to keep the problem stable. i.e., asymptotically, no degree of state reduction is stable.

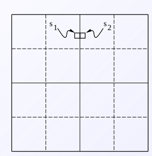
2. Statistical Fidelity:

The need to use approximate models (for computational and/or numerical stability reasons) may lead to unacceptable artifacts or statistical anomalies.

For example, a typical estimation of a random field clearly shows some of the coarser boundaries along which too little information was kept:



Problem: reduced-order models maintain an inadequate correlation across boundaries, such as between pixels s_1, s_2:



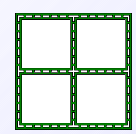
Sample table of adjacent-pixel desired / realized correlation ratios; a ratio of 1.0 is ideal:

#	# States per Correlation Length					
Scales	1.0	2.0	3.0	4.0	5.0	6.0
5	6	7	6	6	6	6
6	22	10	9	7	7	6
7	41	9	8	6	6	
8	75	10	9			
9	97	9	8			
10	120	9				
11	120	8				
12	110	8				

3. Computational Complexity:

For an $n \times \ldots \times n$ block in d dimensions, number of pixels in coarsest boundary ζ_o is $\mathcal{O}(n^{d-1})$.

Therefore, for any statistical model, the number of degrees of freedom in the boundary is $\mathcal{O}(n^{d-1})$; so any reduced-order model, e.g.,



has the *same* asymptotic dimension $\mathcal{O}(n^{d-1})$.

Therefore the estimation complexity is at least $\mathcal{O}(n^{3(d-1)})$.

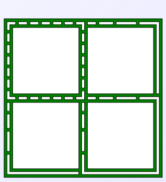
Therefore estimation complexity *per pixel* is $\mathcal{O}(n^{2d-3})$, prohibitive for even modest n, d.

Proposed Idea

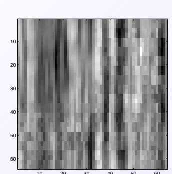
Key Points:

1. The first step of divide-and-conquer – conditional division of the whole process – is too hard.
2. Reduced-order representations of the whole boundary ζ_o are not effective approximations.

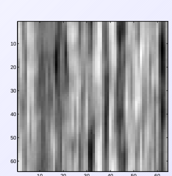
To estimate one quadrant, keeping *details* of another may be largely irrelevant; the following reduced state will perform very nearly as well:



The corresponding estimated random field ...

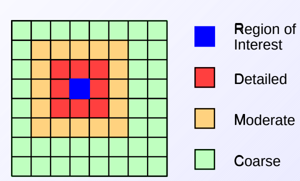


Several of these can be stitched together to produce smooth estimates of the entire field:



Using four models, one per quadrant, does not change asymptotic properties.

In general, we can consider p models, each representing $1/p$ of the problem:



If p grows as n^d then the asymptotic properties *can* be changed. If $p \propto n^d$ then the size $1/p$ of the region of interest is constant, and the size of boundary ζ_o is bounded or grows only very slowly with n, d.

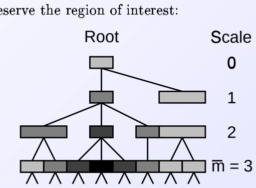
Specific Approach:

Break the problem into $2^{d\bar{m}}$ regions of interest on scale $\bar{m}$. Let t_i be the node on scale $\bar{m}$ representing the region of interest on the ith tree $\mathcal{T}_i$.

Each region of interest is surrounded by bands, represented in decreasing statistical detail.

Below scale $\bar{m}$ the tree is regular.

Above scale $\bar{m}$ the tree is adjusted to explicitly preserve the region of interest:



Many-rooted Trees

1. Numerical Stability:

Stability maintained independent of problem size:

States:	1.0	2.0	3.0	4.0	5.0	6.0
Cond. #:	10^2	10^5	10^6	10^7	10^{11}	10^{12}

2. Statistical Fidelity:

Because stability is independent of size, accurate models can be chosen, even for large problems, to set the desired tolerance on statistical fidelity:

States:	1.0	2.0	3.0	4.0	5.0	6.0
Fidel.:	400	45	6	5	1	1

3. Computational Complexity:

Fixed size of reqion of interest is $q = n^d/p$.

p models, with root node of complexity q, n^d finest-scale elements, *per pixel* complexity of

$\approx \mathcal{O}\left[\frac{pq^3+pn^d}{n^d}\right] = \mathcal{O}\left[q^2 + \frac{n^d}{q}\right]$ ...not so good.

Many of the p trees are *very* similar, suggesting that redundancies are present which can be removed.

Let $\mathcal{T}_i$ be the ith tree

Let $\mathcal{T}_i(s)$ be the subtree below node s

Let $Y(\mathcal{T})$ to be the measurements on tree $\mathcal{T}$.

On each tree $\mathcal{T}$, estimates are computed:

Upwards Pass: Compute conditional estimates

$$\hat{x}_U(s) = E[x(s)|Y(\mathcal{T}(s))]$$

based on measurements in the subtree below s.

Tree Root: $\hat{x}_D(0) = \hat{x}_U(0)$.

Downwards Pass: Compute smoothed estimates

$$\hat{x}_D(s) = f(\hat{x}_U(s), \hat{x}_D(s\bar{\gamma})).$$

based on all measurements on tree.

We characterize each node s in terms of

- $s\bar{\gamma}$ its parent
- $s\alpha_i$ its children
- $\rho(s)$ its sampling density (statistical fidelity)
- $m(s)$ its scale
- $\pi(s)$ its spatial extent.

Nodes $s_i \equiv s_j$ are equivalent iff

$$\rho(s_i) = \rho(s_j), \pi(s_i) = \pi(s_j), m(s_i) = m(s_j)$$

Let $\mathcal{T}_*$ be the union of the p models,

$$\mathcal{T}_* = \mathcal{T}_1 \bigcup \ldots \bigcup \mathcal{T}_p,$$

that is, with redundant equivalent nodes removed. $\mathcal{T}_*$ will have p root nodes.

Claim: Under the following conditions, the estimates on $\mathcal{T}_*$ are identical to the mosaic of estimates from $\mathcal{T}_1, \ldots, \mathcal{T}_p$.

Upwards Pass:

For any $s_i \in \mathcal{T}_i, s_j \in \mathcal{T}_j$,

$$s_i \equiv s_j \Longrightarrow \mathcal{T}_i(s) \equiv \mathcal{T}_j(s) \Longrightarrow \hat{x}_U(s_i) = \hat{x}_U(s_j)$$

That is, all of the estimates from the p upwards passes will be present on combined tree $\mathcal{T}_*$.

Tree Root: Let the p roots be $0_1, \ldots, 0_p$.

We let $\hat{x}_D(0_i) = \hat{x}_U(0_i)$.

Downwards Pass:

On scales $\geq \bar{m}$, $\mathcal{T}_*$ is regular and the downwards pass is unchanged.

On scales $\leq \bar{m}$, the downwards pass on tree $\mathcal{T}_i$ proceeds only between ancestors $t_i\bar{\gamma}^k$ of the region of interest.

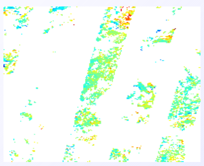
On $\mathcal{T}_*$ a node s may have multiple parents. The desired parent for the downwards pass is the one with the same region:

$$\hat{x}_D(s) = f(\hat{x}_U(s), \hat{x}_D(s')) \ni \begin{array}{l} m(s') = m(s) - 1, \\ \pi(s') = \pi(s) \end{array}$$

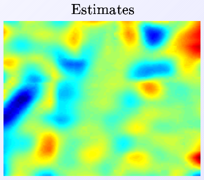
Results

Application 1: Temperature assimilation

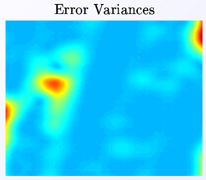
Measurements



Estimates

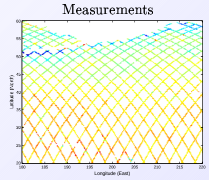


Error Variances

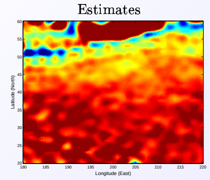


Application 2: Altimetry estimation

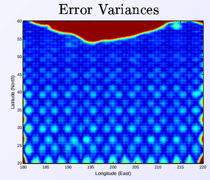
Measurements



Estimates



Error Variances



References

(Reprints Available for References in Green)

P. Fieguth, W. Karl, A. Willsky, C. Wunsch, "Multiresolution Optimal Interpolation ... of TOPEX/POSEIDON Satellite Altimetry", *IEEE GRS* (33) #2, pp.280–292, 1995

P. Fieguth, D. Menemenlis, C. Wunsch, A. Willsky, "Adaptation of a Fast Optimal Interpolation Algorithm to the Mapping of Oceanographic Data", *JGR* (102) #C5, pp.10573–10584, 1997

P. Fieguth, M. Allen, M. Murray, "Hierarchical Methods for Global-Scale Estimation Problems," *Canadian Conf. on Elec. & Comp. Eng.*, Waterloo, Canada, 1998

P. Fieguth, "Multipole-Motivated Reduced-State Estimation," *IEEE ICIP'98*, Chicago, pp.MP5.08-1–4, 1998

A. George, *Computer solution of large sparse positive definite systems,* Prentice-Hall, 1981.

G. Golub, R. Plemmons, "Large-Scale Geodetic Least-Squares Adjustment by Dissection and Orthogonal Decomposition," in *Large Scale Matrix Problems* (Björk, Plemmons, Schneider eds.), North Holland, New York, pp.3–28, 1981

W. Irving, P. Fieguth, A. Willsky, "An Overlapping Tree Approach to Multiscale Stochastic Modeling and Estimation", *IEEE Trans. on Image Processing* (6) #11, pp.1517–1529, 1997