Foveated Multiscale Models for Large-Scale Estimation

Paul W. Fieguth

Dept. of Systems Design Engineering
University of Waterloo

ABSTRACT

Efficient, large-scale estimation methods such as nested dissection or multiscale estimation rely on a divide-and-conquer strategy, in which a statistical problem is conditionally broken into smaller pieces. This conditional decorrelation is not possible for arbitrarily large problems due to issues of computational complexity and numerical stability. Given the growing interest in global-scale remote sensing problems (or even three-dimensional problems), in this summary we develop a class of estimators with more promising asymptotic computational properties.

1. Introduction

Heightened environmental awareness and concerns have led to an explosion in the quantity of remotely-sensed data, leading to more ever-larger problems requiring statistical estimation (for example, to remove irregularities in sampling or to detect anomalous behaviour).

Most efficient, large-scale estimation methods (e.g., nested dissection[5, 6] or multiscale estimation[1]) rely on some sort of divide-and-conquer strategy: a state vector is found which conditionally breaks the problem into smaller pieces. As the size of the underlying problem grows (Fig. 1), this first conditional division becomes increasingly problematic:

- The overall computationally complexity grows as the cube of the length of this first state.
- More significantly, the covariance associated with the state vector becomes more poorly conditioned as the state length grows, such that solutions become numerically unstable beyond a certain size.

Given the current interest in global-scale problems, the above asymptotic weaknesses motivate alternative methods for large-scale estimation; this paper presents one such alternative.

Research of this paper supported in part by the Natural Science & Engineering Research Council of Canada, and by the Office of Naval Research under Grant N0014-91-J-1004.

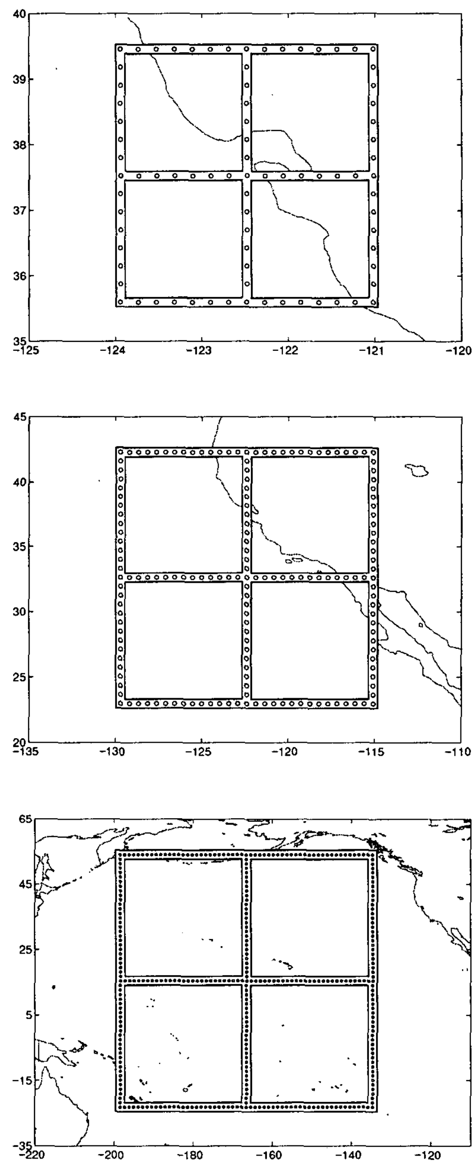


Fig. 1. For how large a domain is divide-and-conquer computationally and numerically feasible ...?

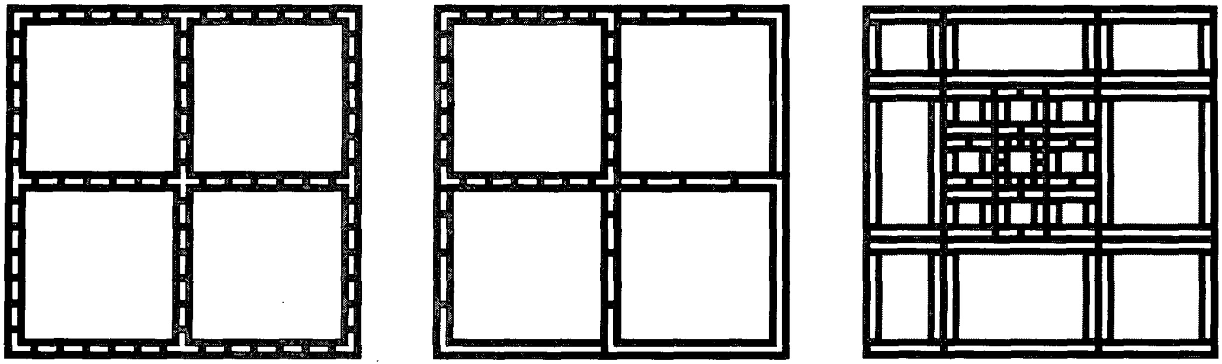


Fig. 2. How many separate models are desired for a given region? A single model (left), one model per quadrant (middle), or a large number of "foveated" models, each capturing only a small subset (right).

2. Multiple Trees

As has been discussed in an earlier paper[4], it may be possible to solve the estimation problem using $m > 1$ models (Fig. 2), where each model is responsible for a subset of the overall region; the desired estimates result from the collaging of the estimates of the individual models. The fundamental motivation behind this idea is as follows: a model designed to estimate (say) quadrant 1 does not need to conditionally decorrelate all four quadrants, rather only those statistical aspects of quadrants 2,3,4 relevant to estimating quadrant 1 need to be kept. Since computational effort scales as $\mathcal{O}(n^3)$, only a 37% reduction in state dimension is required to compensate for the $m = 4$ increase in the number of models. In addition to the computational reduction, the shortened state dimension typically leads to reduced numerical instabilities.

Arguably a *local* estimation scheme, in which estimates are based on measurements within some local vicinity, could yield similar computational and numerical benefits, however continuing to use multiscale models and estimation leads to substantial benefits:

- Most significantly, each estimate is based on *all* of the measurements; that is, the approach is *not* local. Furthermore, the global nature of the models permits data fusion with non-local measurements.
- Other advantages of the multiscale model are maintained: efficient likelihood estimation, a stochastic realization theory, and a base of existing models and applications.

With the above framework in place, three issues need to be settled for a preliminary implementation:

(a) How many models m is desirable (or "optimal")?

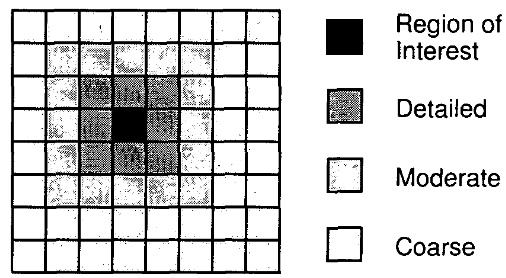


Fig. 3. The problem is broken into pieces, each chosen to estimate one particular region of interest. Each piece is surrounded by two concentric bands of decreasing statistical detail, followed by a coarse representation of the remaining domain.

(b) What is the nature of the model (i.e., what statistical information is kept)?

(c) How is the model made multiscale?

None of the above is made clear by Fig. 2(right), which presents (at best) highly qualitative answers.

The answer to (a) follows from (b), since the dependence of computational complexity on m is easily computed once the model details are known.

The answer to (b) follows from Fig. 3 which shows a stratified or foveated model, in which the $p \times p$ square region to be estimated is surrounded by several (here three) concentric regions, modeled with progressively coarser statistical fidelity. The "thickness" of the inner two concentric regions is set approximately to the correlation length of the underlying statistics. As the problem grows asymptotically, only m changes, *not* p, as opposed to the traditional multiscale approach, in which $m = 1$ and p grows with the problem size. It is this distinction which accounts for the promising computational properties of our revised approach.

(c) is discussed in the following section.

3. Multiscale Implementation

We can construct each of the m individual models are constructed as a standard quadtree, except that the statistical fidelity will vary from node to node, depending upon its distance from the region of interest. Also the tree needs to be modified for the ancestors of the region of interest — they have nine child descendents, arranged on a three-by-three grid, centered on the region of interest. The estimates over the entire domain are found by mosaicing the estimates (and error statistics) corresponding to the regions of interest for the m models.

The most obvious criticism of using m completely separate models to estimate a random field is that such an approach ignores the fact that the models may involve a great deal of duplicated effort; in particular, the degree of duplication will be m for most tree nodes at finer scales. In fact, by detecting and removing such duplication, the memory to represent the union of all required nodes on all m models is comparable to that required for traditional, single-tree, methods: although we now need to represent some nodes multiple times (e.g., at the various qualities of representation in Fig. 3), we have corresponding savings due to the absence of coarse-scale nodes with huge state dimensions (and correspondingly huge covariances). We can derive a graphical structure representing the *union* of m tree models (one per small region of interest), which removes the duplication present in the models, produces exactly the same estimates, and which leads to computationally-efficient algorithms, similar to those which exist for normal multiscale trees. A single model, with m roots, can achieve these goals.

Let $\mathcal{T}_i$ be the tree corresponding to the ith model, and let

$$\mathcal{T}_* = \mathcal{T}_1 \bigcup \cdots \bigcup \mathcal{T}_m$$

be the union of the m models (i.e., keeping only one copy of equivalent nodes); two nodes are defined as equivalent if they represent the same region of space at the same scale with the same degree of statistical fidelity.

Under these conditions the estimates computed from a single m-root tree are *identical* to the mosaicing of estimates from m separate quad-trees, but without the gross duplication in storage and computational effort.

4. Results

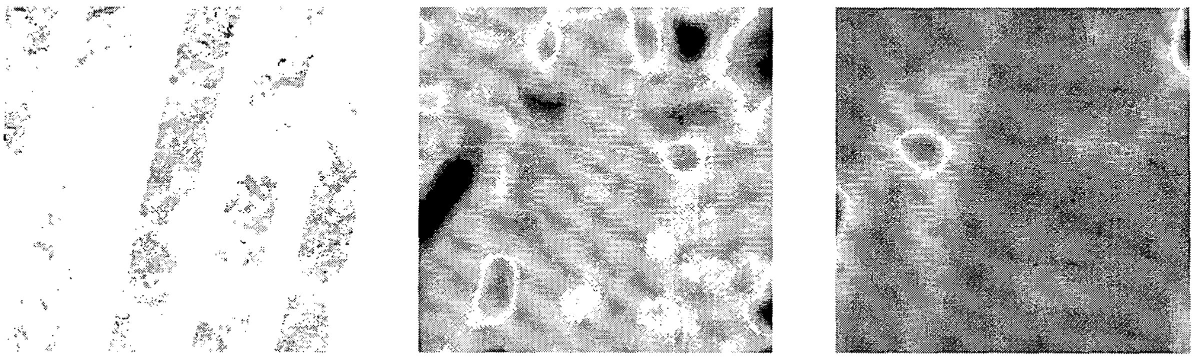
Fig. 4 shows one example of applying the multiple-tree approach to a remote-sensing problem of current interest[3] in climate-modeling and climate-change studies — estimating the surface temperature of the ocean from infrared measurements (left), which are sparse and are taken in bands. The estimates (middle) are based on 64 models within an overlapped multiscale framework[7]. Overall, 200×200 estimates and error statistics were computed in about 40 seconds of CPU time on a Sun Ultra-1. The prior model is an isotropic correlation function, Gaussian in shape, with a correlation length of five degrees.

A second application is shown in Fig. 5, which applies the multiple-tree approach to a second problem in remote-sensing — characterising the surface topography of the ocean from altimetric (height) measurements [1]. The large gap in the measurements at the top of the figure is Alaska; the estimates show an overall north-south gradient, consistent with the presence of the Kuroshio current from Japan. The prior model consistent with the ocean-height statistics has a short correlation length (three degrees), which causes the sampling nature of the measurements to appear prevalently in the estimation error variances (right panel). The estimates and error statistics were computed in about one and one half minutes of CPU time.

Both of the above applications illustrate the potential of the multiple-tree approach for extremely large estimation problems. As mentioned in the Introduction, the two aspects which limit problem size for divide-and-conquer approaches are computational complexity and numerical stability. Both of these aspects are addressed by limiting the size of the largest state dimension: we keep fixed the size of each region of interest by varying the number of models m in proportion to the overall size of the problem. Consequently numerical problems are almost independent of problem size, and computational complexity per pixel is nearly constant.

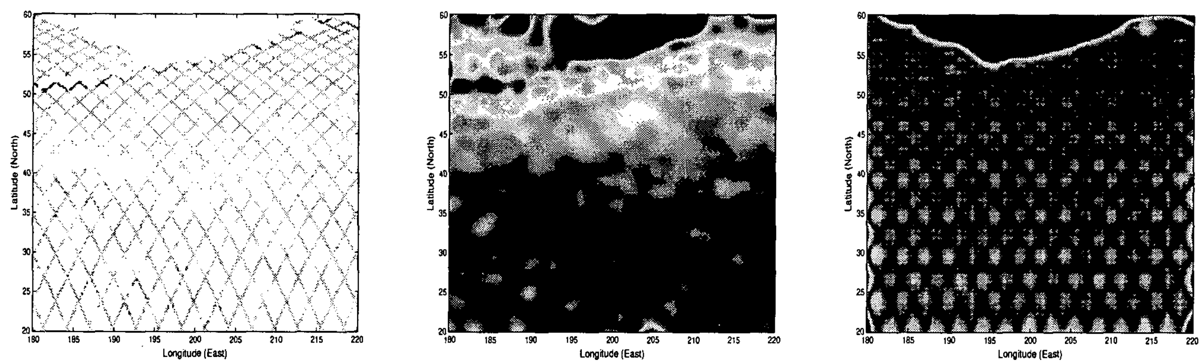
A number of issues remain for future examination. First, in order to produce estimates of a region of interest, what statistical information really needs to be preserved from the rest of the domain? The answer will clearly be prior-model dependent, but may be worth exploring for some models in widespread use.

Second, when computing estimates far away from measurements (i.e., extrapolating), the values of the estimates, although possibly statistically insignificant, may be highly sensitive to the choice of prior model. It is not clear how to properly mosaic in the face of such variations.



Infrared Measurements Temperature Estimates Estimation Error Variances

Fig. 4. Estimation results (200×200 pixels, isotropic Gaussian prior), given pointwise measurements of ocean surface temperature in the equatorial Pacific. The measurements and estimates are mean-removed, and vary over about ±1.5Kelvin.



Altimetry Measurements Surface Height Estimates Estimation Error Variances

Fig. 5. Estimation results (200×200 pixels, isotropic Gaussian prior), given pointwise measurements of ocean surface height in the northern Pacific. The measurements and estimates have a range of about ±1metre.

5. References

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

[2] 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

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

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

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

[6] 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

[7] 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

[8] W. Irving, W. Karl, A. Willsky, "A Theory for Multiscale Stochastic Realization", *33rd Conference on Decision and Control*, Orlando, FL, pp.655–662, 1994