Ultra-scale Visualization Climate Data Analysis Tools (UV-CDAT)

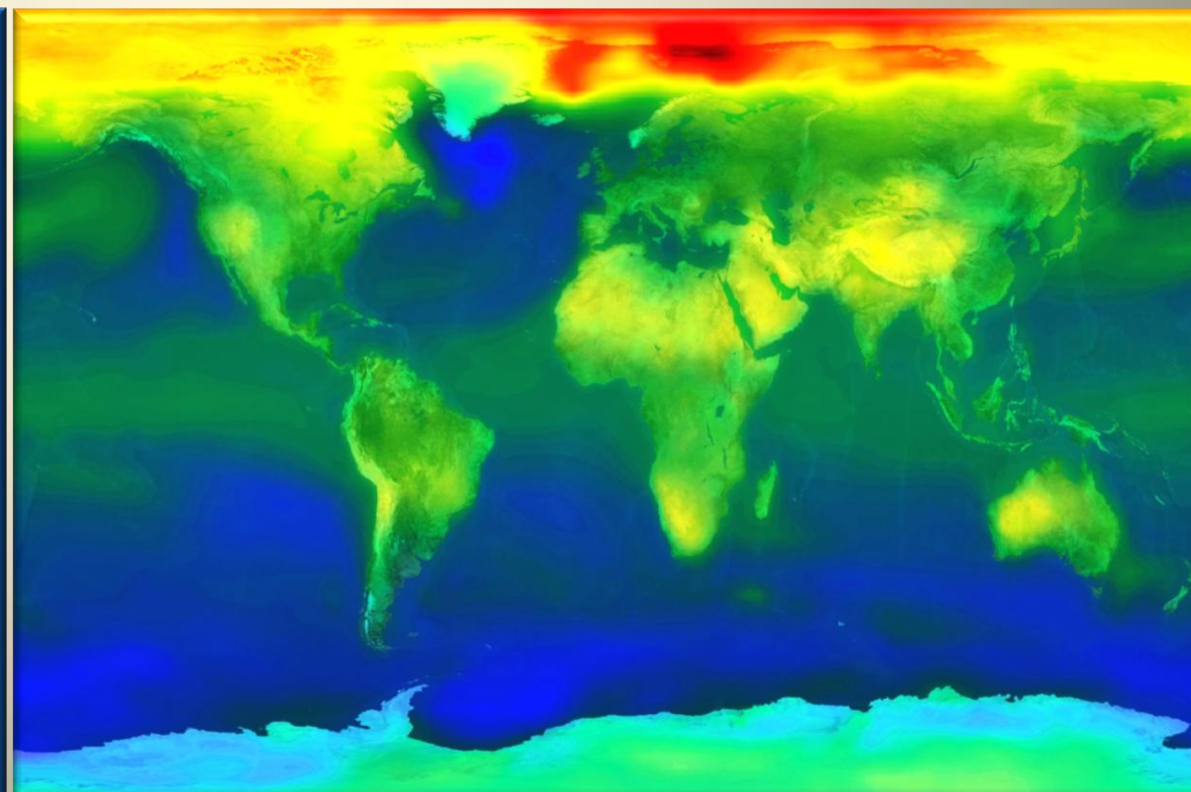
Delivering science and technology solutions to national needs in climate

UV-CDAT Team • September 8, 2011

Department of Energy • Office of Science • Biological and Environmental Research

DOE BER Climate Research

Kitware
NYU-poly
Berkeley Lab
SCI www.sci.utah.edu
Los Alamos National Laboratory EST. 1943
NASA
Oak Ridge National Laboratory
Office of Science



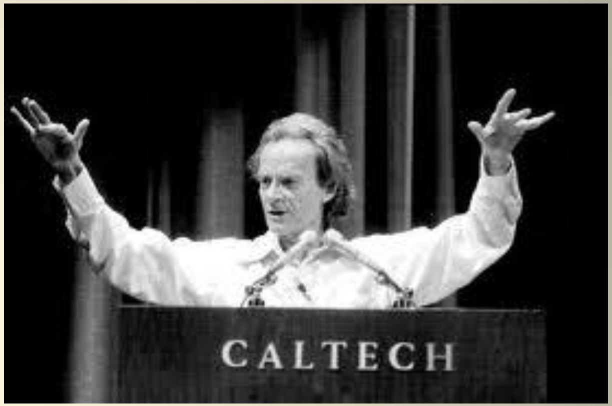
U^2V: Ultrascale Uncertainty Visualization

Richard Strelitz, LANL
E. Wes Bethel, Prabhat, LBL
Berk Gevici, Kitware, Inc
Claudio Silva, NYU Polytechnic

UV-CDAT

Motivation

What is not surrounded by uncertainty cannot be the truth.
Richard Feynman



Outline

1. Uncertainty: What is it and what are its properties
2. Uncertainty: Where does it come from in the context of simulations that cannot be run often enough to create an ensemble
3. Uncertainty: How does it propagate?
4. Uncertainty: A working definition for purposes of visualization
5. Density: A key idea borrowed from continuum mechanics
6. Uncertainty: From value to visualization and back again
7. Requirements for successful, effective visualization (of uncertainty)
8. Depicting density
9. From glyph to field via representative values that *explain* data
10. Several methods for visualizing uncertainty in 2D,3D, …
11. Future

Traits of Effective Uncertainty Visualization

- Works in 2,3,4,4+ Dimensions
- Does not interfere or distract from underlying visualization of data
- Scales with
    - data set size,
    - resolution
- Permits
    - drill down,
    - query, and
    - analysis
- Agnostic to gridding
- Solid basis in:
    - Statistics
    - Informatics
- Easily understood by Domain Scientists
- Extensible to vectors, tensors, lists, spectra

What is Uncertainty?

Uncertainty is the result of selecting one option from a set of (equally) plausible alternatives.

Our goal is to meld statistical thinking with visualization practice to allow the tracking of how uncertainty arises, propagates and affects computed results

One critical question is how uncertainty can be incorporated or extracted from very few (~1) simulation

UV-CDAT

Quantitative Uncertainty

We are all certain about one thing-that there exists something called uncertainty.

For the most part, for most people, uncertainty exists almost entirely in the range conaining the final outcome of the simulation.

For our purposes, we shall use this partial and almost begrudging consensus to define our uncertainty.

For any parameter,

$$\vartheta_{ij} = \vartheta(\vec{r}_i, t_j)$$

so the uncertainty υ_{ij} is such that:

$$P(\vartheta \in [\vartheta_{ij} \pm \upsilon_{ij}]) \geq 0.5$$

That is, the uncertainty has the same units and properties as the parameter

Origins of Uncertainty in Simulation

Simulation is overwhelmingly deterministic (except for application of Monte Carlo methods). Where is the uncertainty in arithmetic? We can't run our models more than a few times, not even for a rudimentary LHS design.

Uncertainty is certainly present in:
•Boundary and initial conditions
•Material properties
•Finite sampling

It also arises whenever a particular algorithm is used as in:
•Interpolation
•Quadrature
•Fitting

Visualizing Ultrascale Data, Cautionary

It is common knowledge that the chief block to visualizing today's ultrascale data is that the I/O channel from disk to screen is overwhelmed, massively.

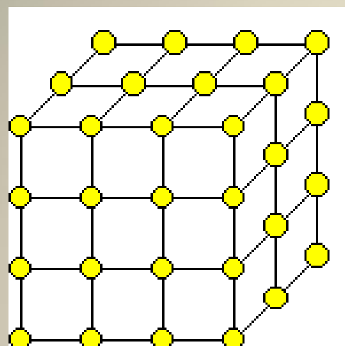
And, we have just doubled the load, adding the uncertainty at each point to the parameter in question. Are we **CRAZY**????

We assert that when there are more points than pixels, then the concept of ***point*** has neither meaning nor utility.

We cannot continue to cede the decision of how many micropolygons can be rendered and then transformed to a single colored pixel to the GPU because it was designed to the specifications of the special effects, gaming, and entertainment industries, all devoted to the creation and maintenance of illusion.

UV-CDAT

From point to field, via density



Replace notion of $\phi(\mathrm{x})$, the value at a point x

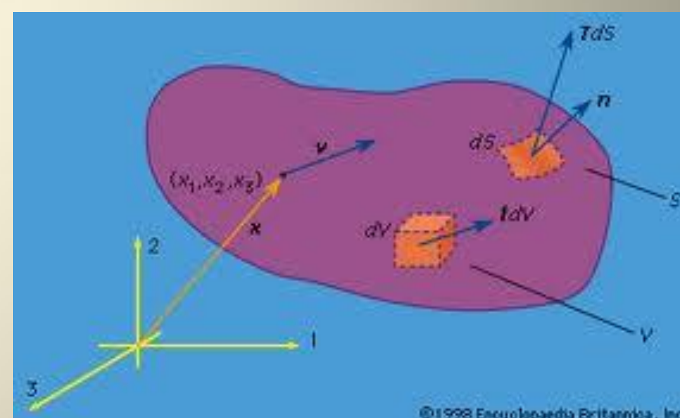
with a continuum field Φ and sampled (viewed) thru

a window or kernel $\kappa(\mathrm{x})$: $\int_{-\infty}^{\infty} \kappa(x)\,dx = 1$

$$\phi(x) = \int_{-\infty}^{\infty} \Phi(y)\kappa(x-y)\,dx$$

Now we can define ρ_Φ the density of Φ

$$\rho_\Phi = \lim_{V \to 0} \int_V \Phi\, dV \;/\; V$$



Size Matters

For a fixed volume V, find amount of φ in V

$$C_{\varphi}(V) = \int_{V} \rho_{\varphi} \, ds$$

Lead Ingots

Water

Define fixed quantity C_{φ}

Find $V_i : \int_{V_i} \rho_{\varphi} \, dV \approx C_{\varphi}$

and $V_i \bigcap V_j = \delta_{ij}$ (non-intersecting)

and $\bigcup_{i=1}^{n} V_i = D$ (space filling)

Feathers:

Each Box contains the same mass, so size changes

UV-CDAT

Basic Kalman Filter

$$\dot{\psi} = N[\psi] + F + q \qquad : x \in D, \quad 0 \le t \le T$$

$$\psi = \Psi_0 + a \qquad : x \in D, \quad t = 0$$

$$\psi = \Psi_B + b \qquad : x \in \delta D, \quad 0 \le t \le T$$

$$d = \mathfrak{M}[\psi] + \varepsilon$$

Here ψ is the state variable;

N[ ] is a nonlinear operator

F(x,t): forcing field

$\mathfrak{M}$[ ] : vector of measurement functionals

$\Psi_{[0,B]}$ initial , boundary conditions

q,a,b,ε errors in model, $\Psi_{[0,B]}$ & data

In the EnKF, we not only find ψ,

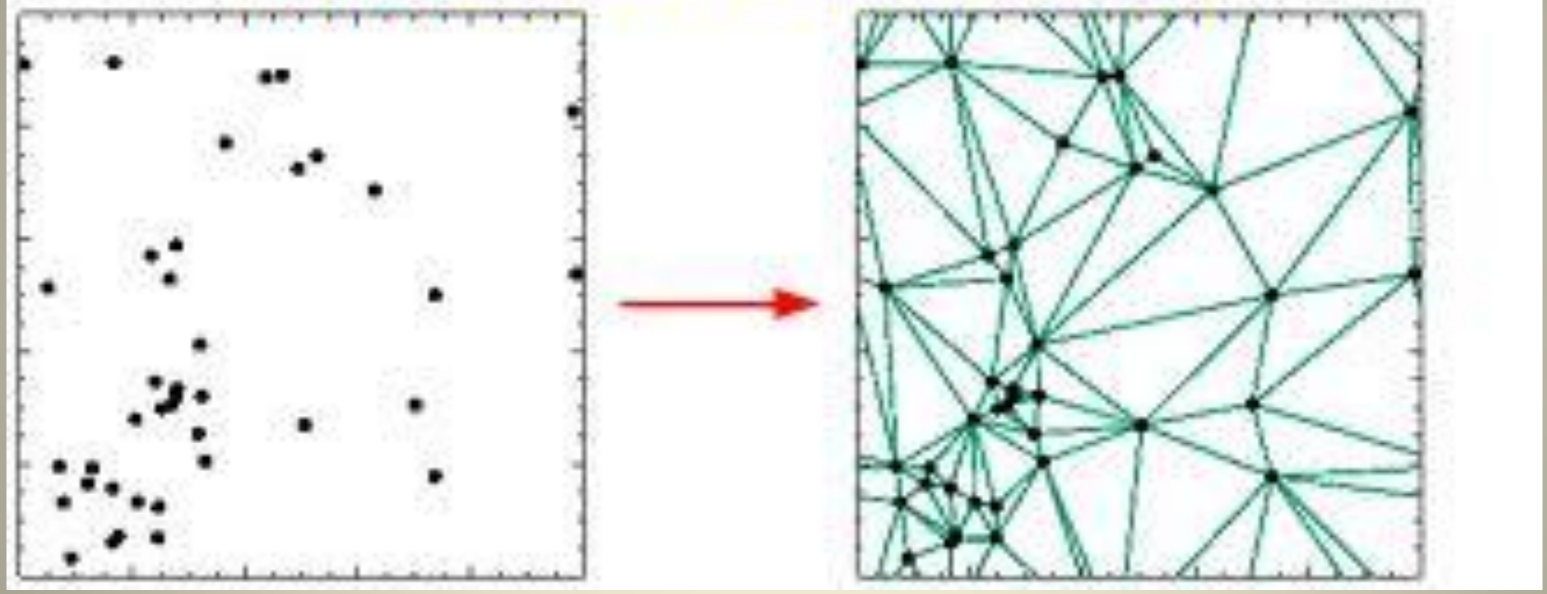
we track W_{aa} $W_{bb} W_{qq} W_{\varepsilon\varepsilon}$

weighting functions

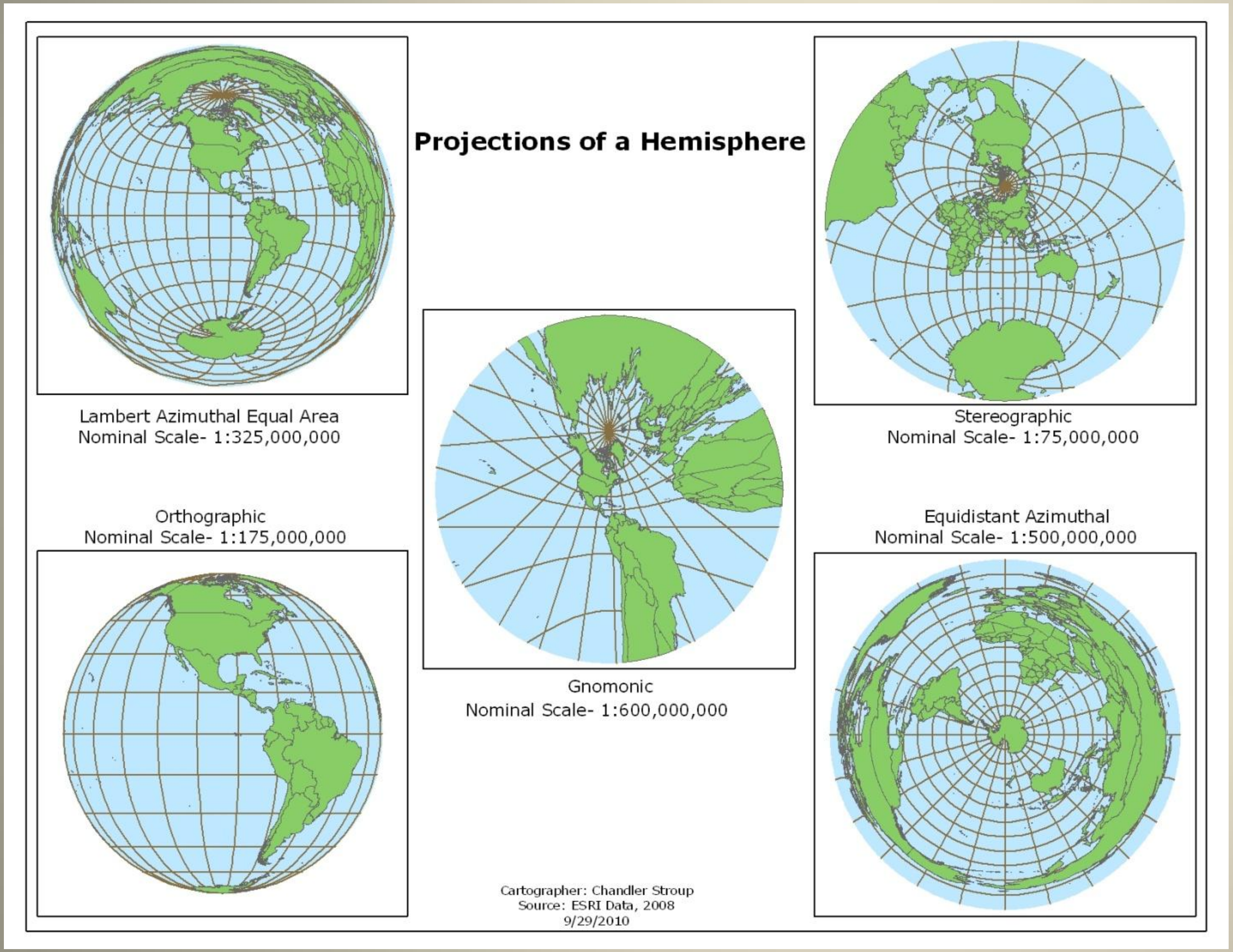
derived from $[q^T q]^{-1}$

inverse error correlation matrix

Implementation Impediments



Mapmaker's Dilemma



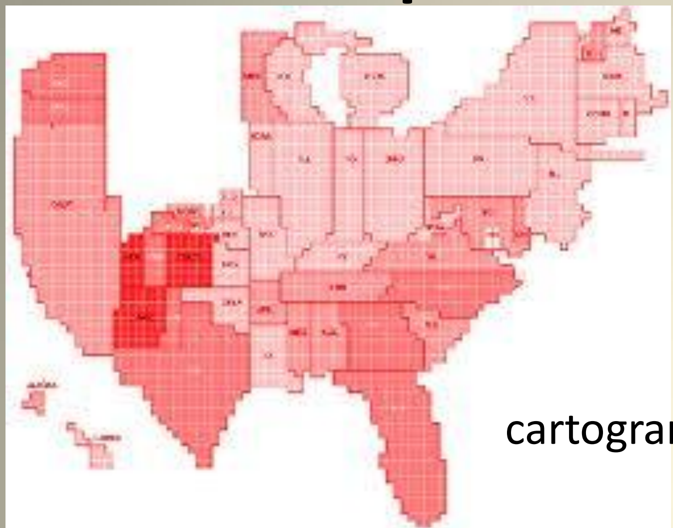
Balance distortion in:

- Area per cell
- Angle and shape
- Direction

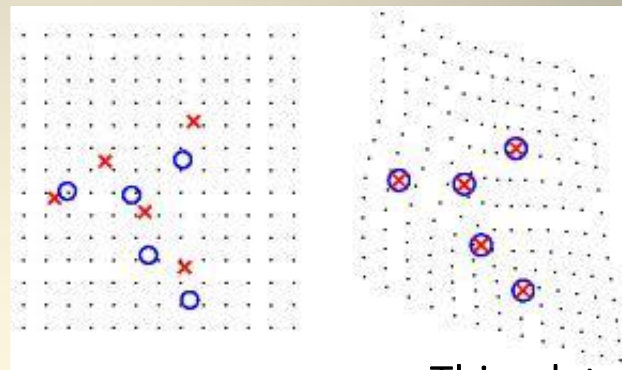
There is no projection that preserves all features

UV-CDAT

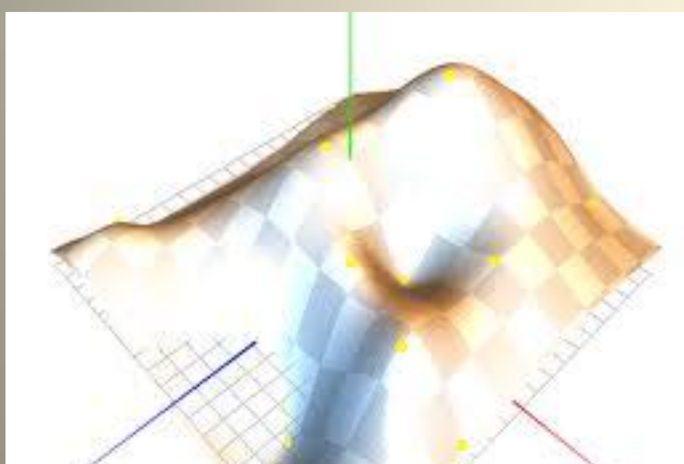
Implementation Options



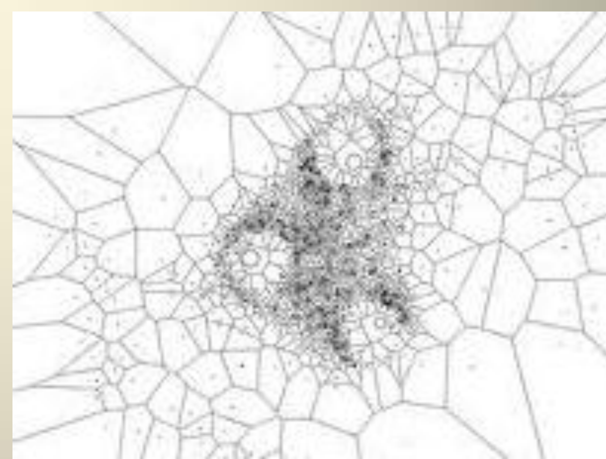
cartogram



Thin plate spline

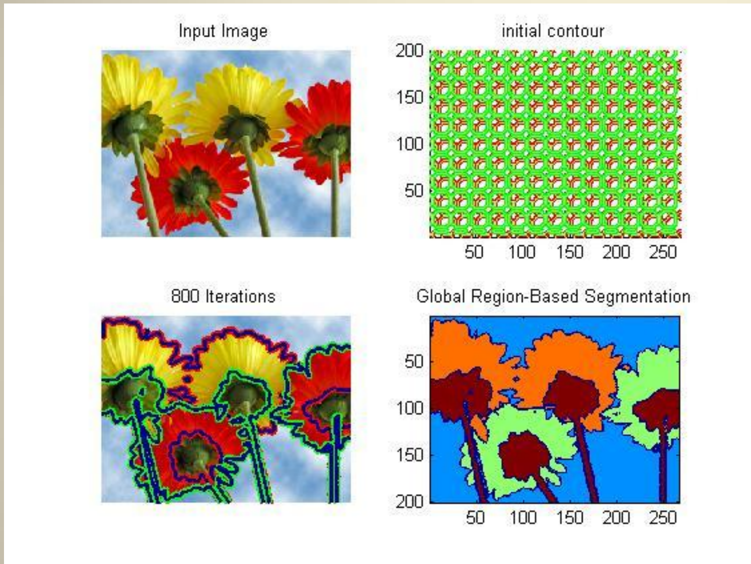


Fishnet deformation

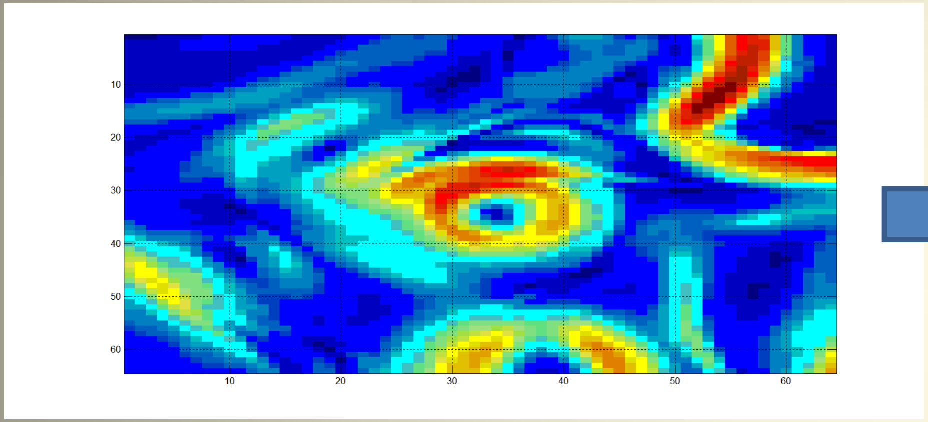


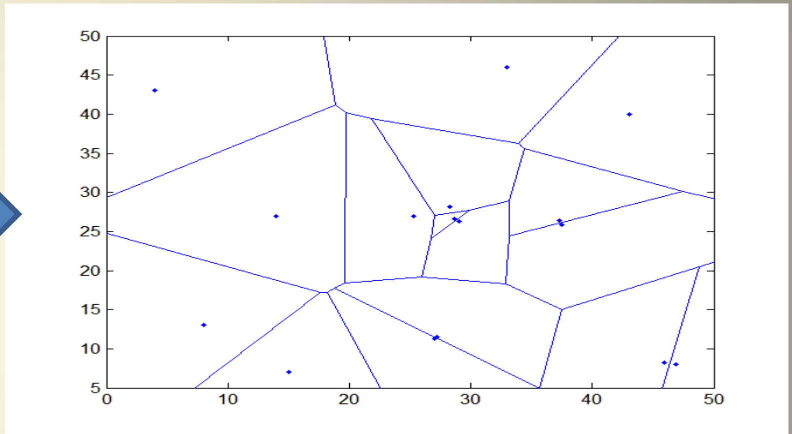
Delaunay Field Tessellation Estimator

Chan-Vese Active Contours



Implementation Impediment





If $\Sigma = \int_{D} \rho_{\varphi} dv$ = Total mass of uncertainty stuff in D,

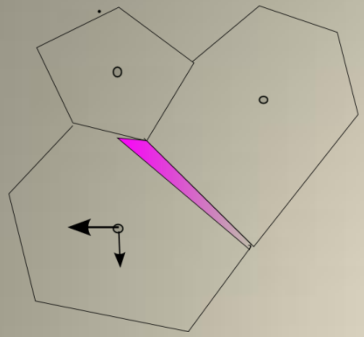
Can I find K regions V_i:

a) $C_i = \int_{V_i} \rho_{\varphi} dV = \Sigma / K = \hat{C} \quad \forall \{V_i\}$

$b) \bigcup_{i=1}^{n} V_i = D$

$c)$ Centroid of V remains in V

Iterative Refinement



For K polygons $\{V_i\}$ and points $\{p_i\}$:

Compute how area $\{A_i = \int dv\}$ and $\hat{\rho}_i = \frac{\int_V \rho\, dv}{V}$

change as $\{p_i\}$ vary $c_{ij} = \frac{\partial A_i}{\partial p_j}$

Optimal solution is found by iterative refinement. The method borrows much from force directed graph placement

Minimize:

$var(\hat{\rho}_i)$ Equalize contents

$\Sigma(\partial V_i)$ Boundary (!gerrymander)

$var(\rho - \hat{\rho})^2$ heterogeneity in V_i

UV-CDAT

Essential References

Visualization Viewpoints

Editor: Theresa-Marie Rhyne

A Next Step: Visualizing Errors and Uncertainty

Chris R. Johnson and Allen R. Sanderson

Scientific Computing and Imaging Institute, University of Utah

When was the last time you saw an isosurface with error bars or streamlines with standard deviations or volume visualizations with representations of confidence intervals? With few exceptions, visualization research has ignored the visual representation of errors and uncertainty for 3D visualizations. However, if you look at peer-reviewed science and engineering journals, you will see that the majority of 2D graphs represent error and/or uncertainty within the experimental or simulated data. Why the difference? Clearly, if it's important to represent error and uncertainty in 2D graphs, then it's equally important to represent error and uncertainty in 2D and 3D visualization.

The possible detriment caused by the failure to represent errors and uncertainties in 3D visualizations became clear to us a couple of years ago when neuro- tain representations of error and uncertainty is that the visualization research community has not made such representations a priority. To take visualization research—and its usefulness to researchers in science, engineering, and medicine—to the next level, the visualization research community needs to make visually representing errors and uncertainties the norm rather than the exception.

What's been done so far

Fortunately, a few visualization researchers have started thinking about 3D visual representations of errors and uncertainties, the sources of which can include uncertainty in

- acquisition (instrument measurement error, numer-

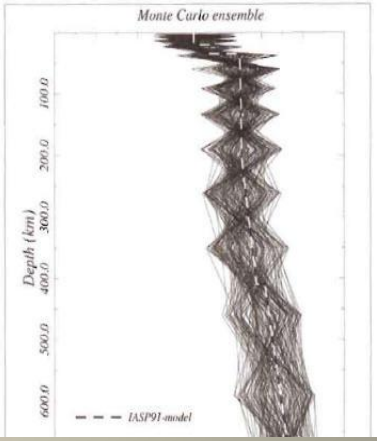
CG&A Special Edition 2003

Leading Edge

How do we understand and visualize uncertainty?

MALCOLM SAMBRIDGE, Australian National University, Canberra
CAROLINE BEGHEIN, Arizona State University, Tempe, USA
FREDERIK J. SIMONS, University College, London, U.K.
ROEL SNIEDER, Colorado School of Mines, Golden, USA

Geophysicists are often concerned with reconstructing subsurface properties using observations collected at or near the surface. For example, in seismic migration, we attempt to reconstruct subsurface geometry from surface seismic recordings, and in potential field inversion, observations are used to map electrical conductivity or density variations in geologic layers. The procedure of inferring information from indirect observations is called an inverse problem by mathematicians, and such problems are common in many areas of the physical sciences. The inverse problem of inferring the subsurface using surface observations has a corresponding forward problem, which consists of determining the data that would be recorded for a given subsurface configuration. In the seismic case, forward modeling involves a method for calculating a synthetic seismogram, for gravity data it consists of a computer code to compute gravity fields from an assumed subsurface density model. Note that forward modeling often involves assumptions about the appropriate physical relationship between unknowns (at depth) and observations on the surface, and all attempts to solve the problem at hand are limited by the accuracy of those assumptions. In the broadest sense then, exploration geophysicists have been engaged in inversion since the dawn of the profession and indeed algo-



UV-CDAT

Traits of Effective Uncertainty Visualization

- Works in 2,3,4,4+ Dimensions
- Does not interfere or distract from underlying visualization of data
- Scales with
  - data set size,
  - resolution
- Permits
  - drill down,
  - query, and
  - analysis
- Agnostic to gridding
- Solid basis in:
  - Statistics
  - Informatics
- Easily understood by Domain Scientists
- Extensible to vectors, tensors, lists, spectra

So, how did we do, admittedly according to our own scorecard

UV-CDAT

Takeaway Points & Conceptual Pillars

- Uncertainty is the side-effect of arbitrary selection from choices
- Uncertainty is measured same as the underlying parameter
- There is much to be gained and very little to be lost in replacing *value at a point* with that of *pdf in a region*.
- Density of a field should therefore replace discrete samples.
- Density and concentration can be conveyed by displaying the space needed to enclose regions containing the same amount of parameter
- The tessellation into regions of equal content is best done by DFTE
- Uncertainty can be derived from the EnKF