Revisiting Wavelet Compression for Large-Scale Climate Data using JPEG 2000 and Ensuring Data Precision

John Patchett (patchett@lanl.gov) Presenter
October 24, 2011

Jonathan Woodring (woodring@lanl.gov)
Susan Mniszewski (smm@lanl.gov)
Christopher Brislawn (brislawn@lanl.gov)
David DeMarle (dave.demarle@kitware.com)
James Ahrens (ahrens@lanl.gov)

Operated by Los Alamos National Security, LLC for the U.S. Department of Energy's NNSA

Introduction: Use Case

- Los Alamos scientist (New Mexico)
  - Computational Ocean Modeler
- Data at Oak Ridge (Tennessee)
  - Parallel Ocean Program (POP)
  - 3600x2400x42 structured grid
  - Typically wants scalar fields (temp, salt, u, v)
- Limited Bandwidth (1MB/s)
  - We don't control arbitrary endpoints
  - Bandwidth issues for various situations that are beyond our control.

Related Work

- Large Data
  - Creates data movement issues
- Distance Visualization
  - Has bandwidth limited channels exacerbating data movement problems
- In situ analysis
  - A response to data movement
- Compression
  - Attempt to have less data movement

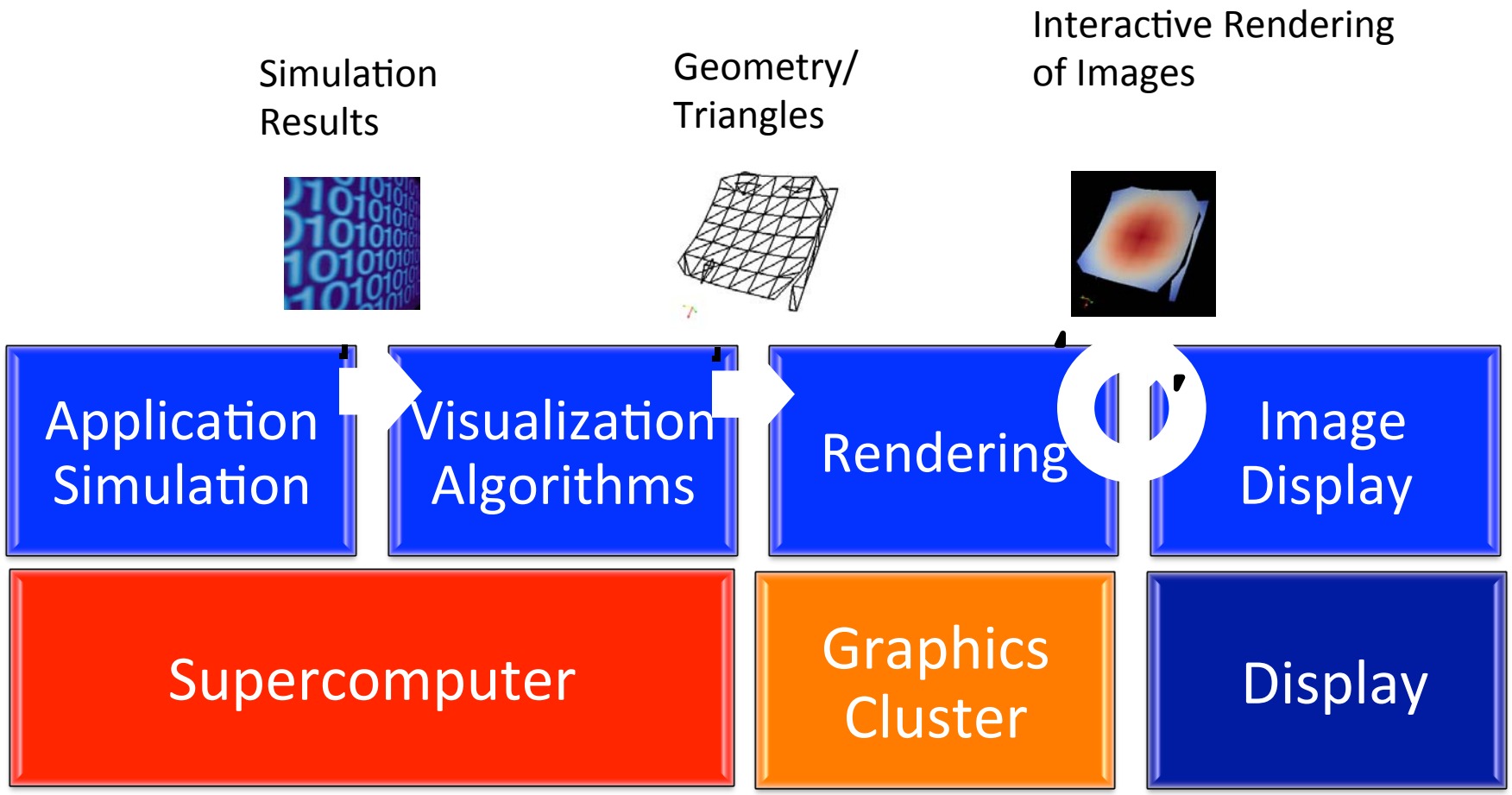
Related Work: Data Compression

- Wavelets have been used extensively for data reduction and multiscale visualization
- Non-wavelet techniques have been developed by vis community for compression, quantization, and multiresolution
- We leverage the signal processing and data compression communities by using JPEG 2000

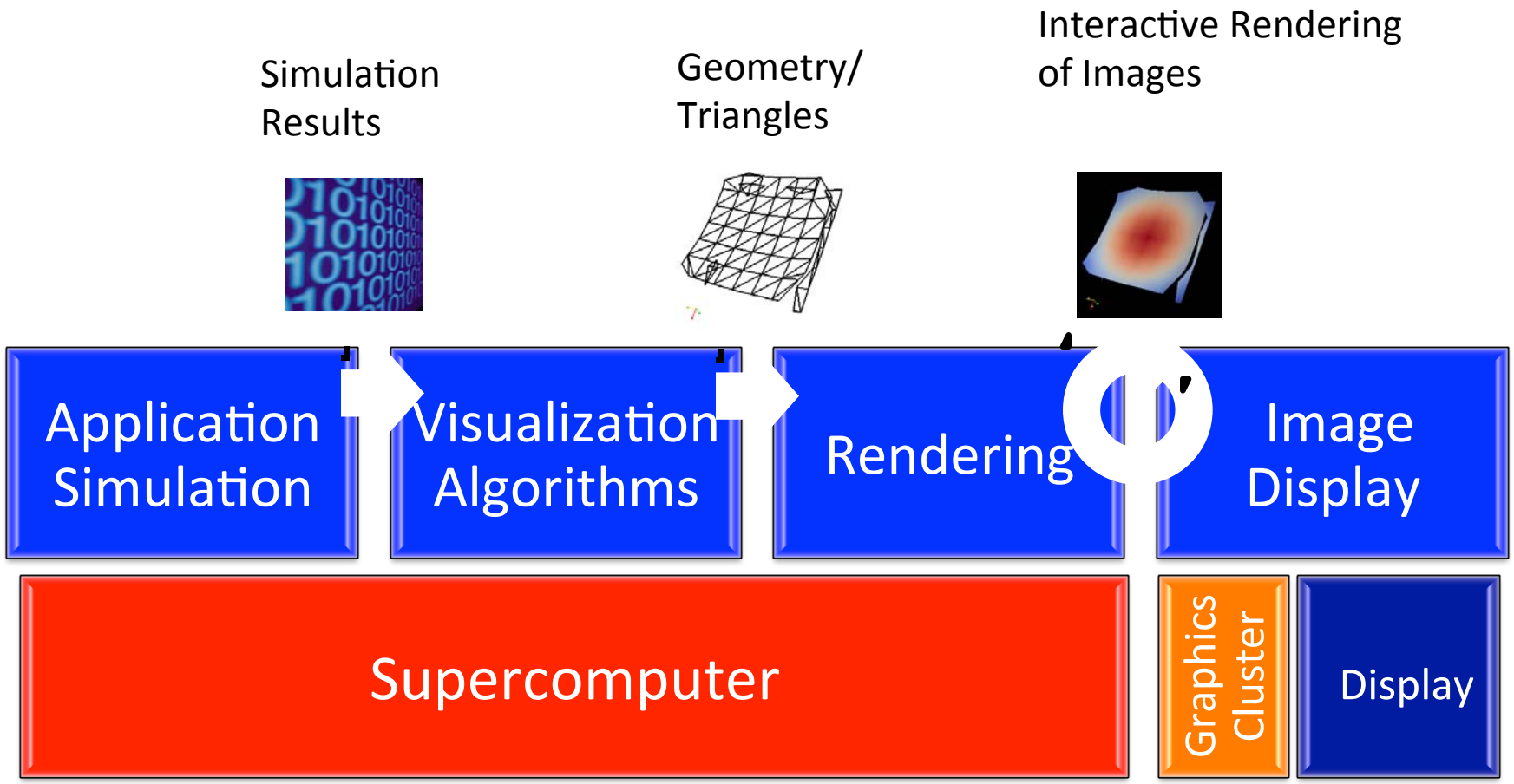
Related Work: Wavelets

- J. Clyne, P. Mininni, A. Norton, and M. Rast. Interactive desktop analysis of high resolution simulations: Application to turbulent plume dynamics and current sheet formation. *New Journal of Physics*, 9(8):301–301, 2007.
- S. Muraki. Approximation and rendering of volume data using wavelet transforms. *Proceedings of the 3rd conference on Visualization '92*, pages 21–28, 1992. ACM ID: 949694.
- J. Woodring and H. Shen. Multiscale time activity data exploration via temporal clustering visualization spreadsheet. *IEEE Transactions on Visualization and Computer Graphics*, 15(1):123–137, 2009.
- Z. Zhu, R. Machiraju, B. Fry, and R. Moorhead. Wavelet-based multiresolutional representation of computational field simulation datasets. In *Visualization '97., Proceedings*, pages 151–158. IEEE, Oct. 1997.
- S. Guthe, M. Wand, J. Gonser, and W. Strasser. Interactive rendering of large volume data sets. In *Visualization Conference, IEEE*, pages 50–60, Los Alamitos, CA, USA, 2002. IEEE Computer Society.
- I. Ihm and S. Park. Wavelet-based 3D compression scheme for very large volume data. In *Graphics Interface*, pages 107–116, 1998.
- T. Kim and Y. Shin. An efficient wavelet-based compression method for volume rendering. In *Computer Graphics and Applications, 1999. Proceedings. Seventh Pacific Conference on*, pages 147–156, 1999.
- F. Rodler. Wavelet based 3D compression with fast random access for very large volume data. In *Computer Graphics and Applications, 1999. Proceedings. Seventh Pacific Conference on*, pages 108–117, 1999.
- A. Trott, R. Moorhead, and J. McGinley. Wavelets applied to lossless compression and progressive transmission of floating point data in 3-D curvilinear grids. In *Visualization '96. Proceedings.*, pages 385–388. IEEE, Nov. 1996.
- C. Wang, J. Gao, L. Li, and H. Shen. A multiresolution volume rendering framework for Large-Scale Time-Varying data visualization. In *Volume Graphics, 2005. Fourth International Workshop on*, pages 11–223, 2005.
- R. Westermann. A multiresolution framework for volume rendering. *Proceedings of the 1994 symposium on Volume visualization*, page 51–58, 1994. ACM ID: 197963.

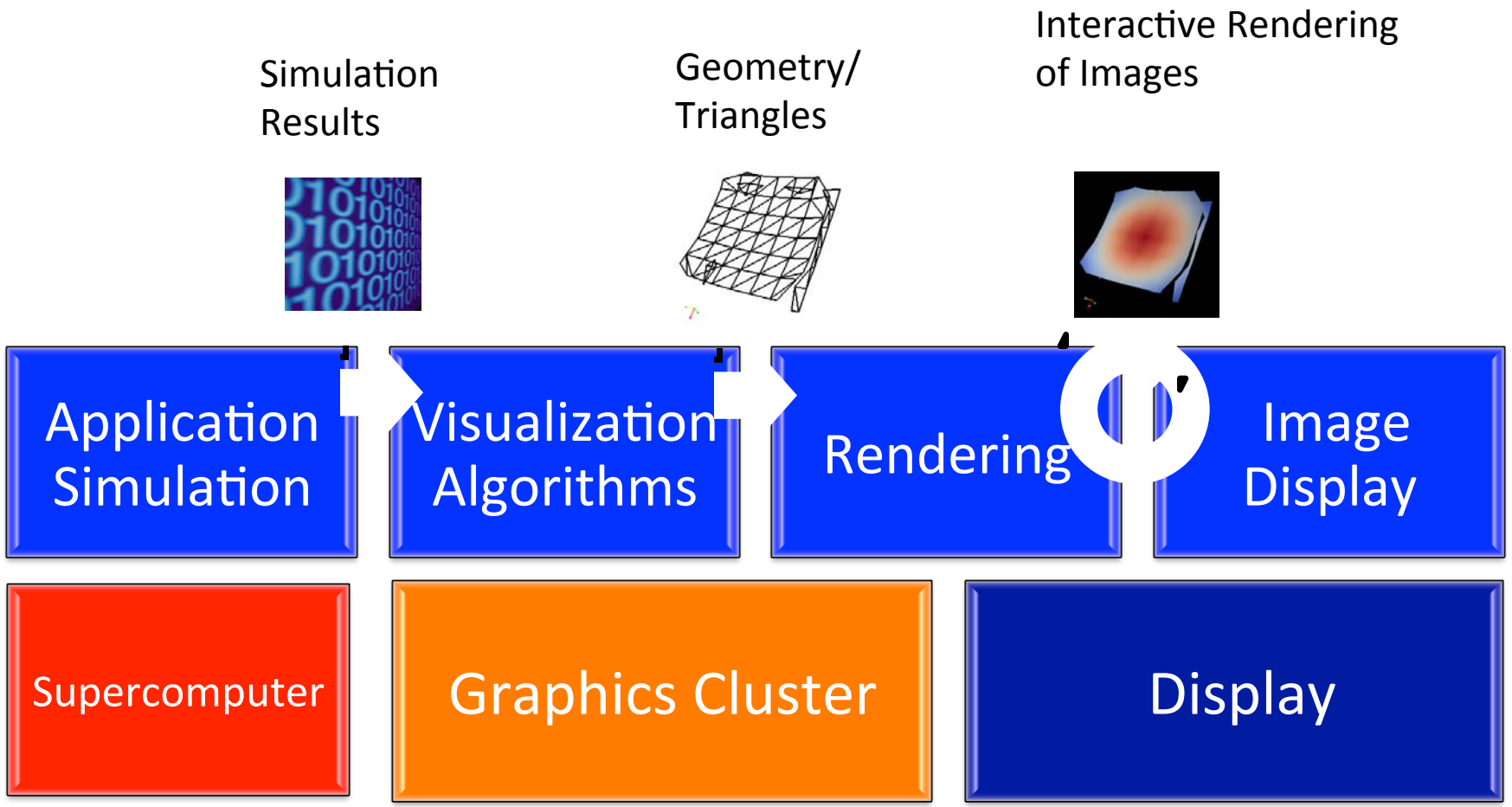
Related Work: Send Geometry



Related Work: Send Imagery



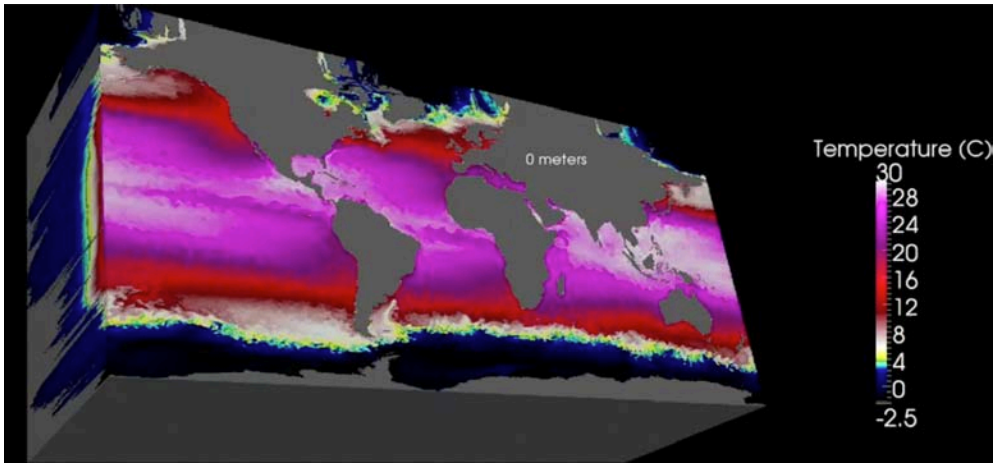
Our Approach: Send Data



Our Approach: Data Compression with Quantified Accuracy

- In visualization and image processing, data compression and the resulting error has been measured as average difference
  - concerned with reducing visual quality differences
- Compression directly in-situ on simulation data as a data reduction mechanism
  - our research focus is to quantify the maximum/L-infinity norm (rather than average/L2 norm) data quality for scientific analysis
  - Provide a solution that automatically compresses simulation data with accuracy guarantees
- (Simulation Data – Compressed Representation) provides an accuracy metric

Our Approach: Compression-Decompression Pipeline



Original Floating Point Data

Masking (Global Average)

Quantization (PCM)

Compression (JPEG2000)

Network

Decompression (JPEG2000)

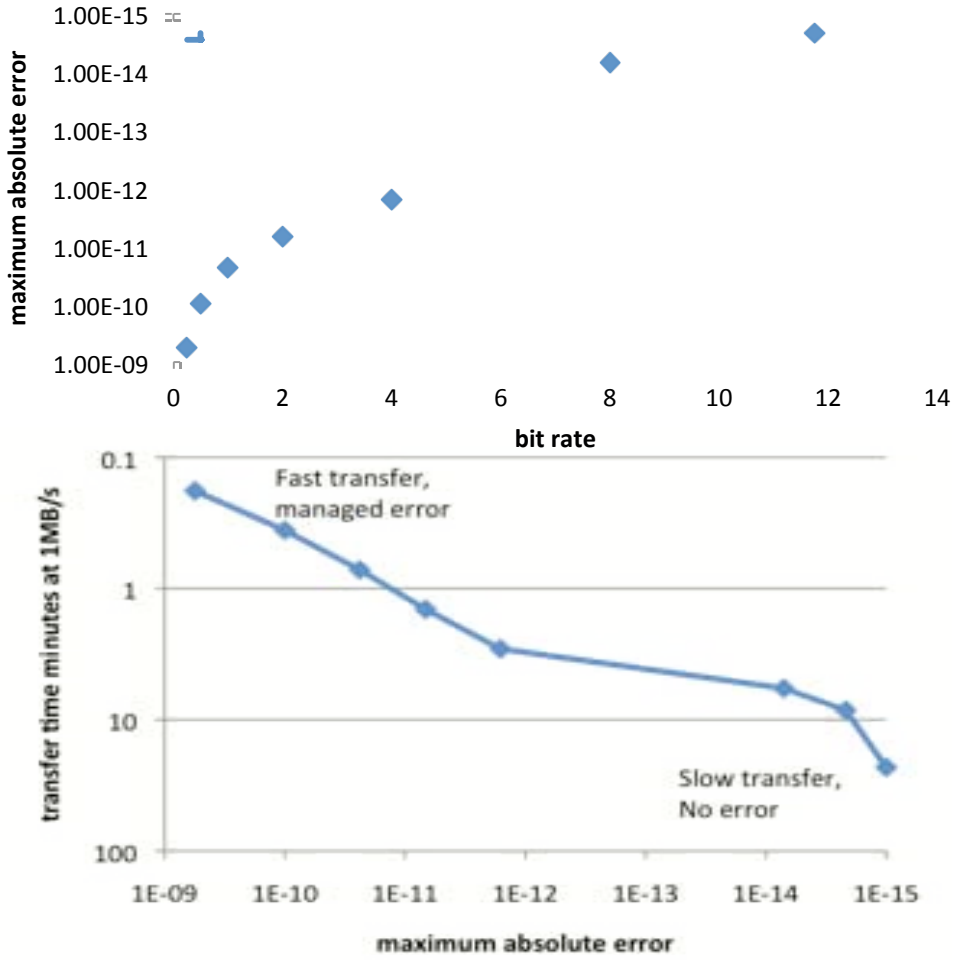
Dequantization (PCM)

UnMasking

Reconstructed Floating Point Data

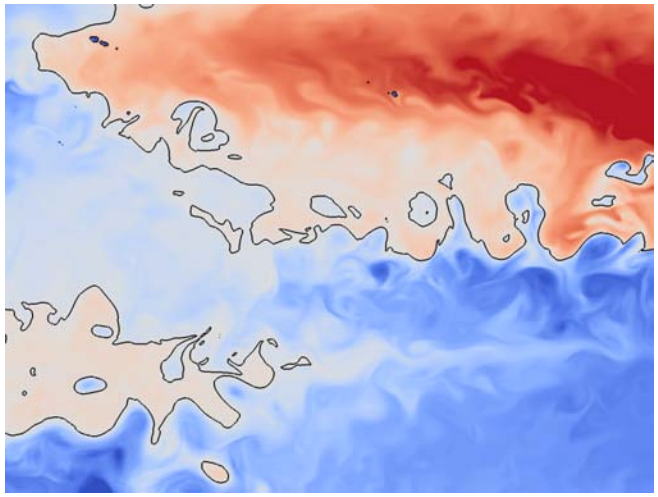
Quantify the Maximum Error (L-infinity norm) so the Scientist Knows the Data Precision

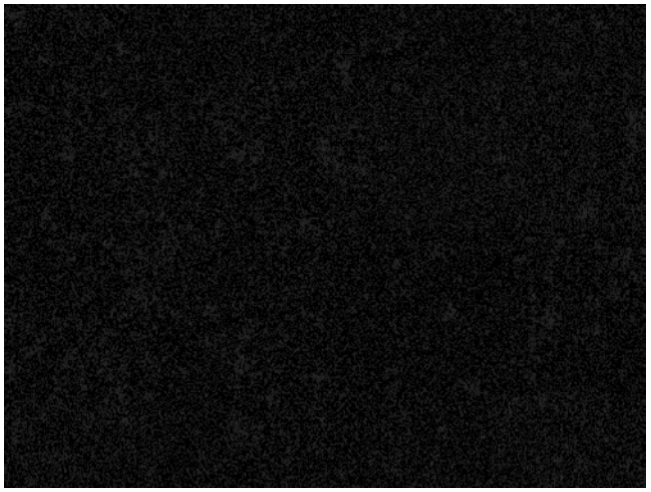
- We measure the maximum point error so there is a guarantee that the data are accurate to *x* decimal places
- The user can trade read I/O time vs. data accuracy in a quantifiable manner



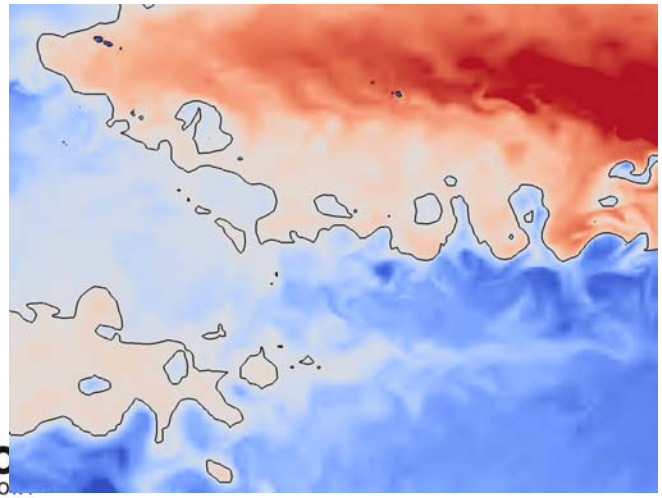
Results: Data Differencing

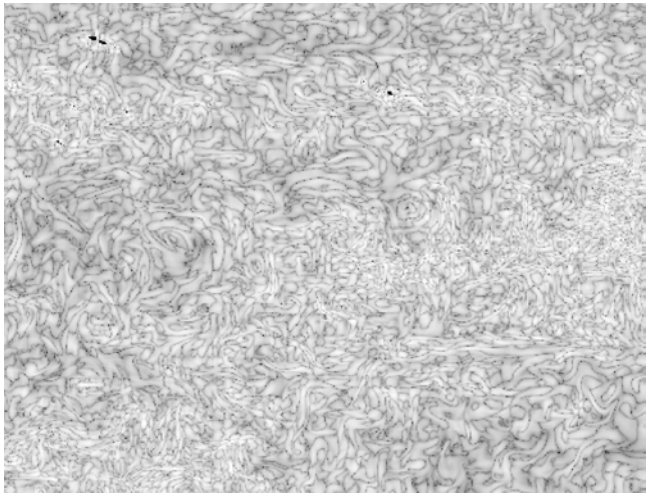
Bitrate = 8
Max Error
=1.49e-09
345 MB





Bitrate = .25
Max Error
=0.000303
11 MB

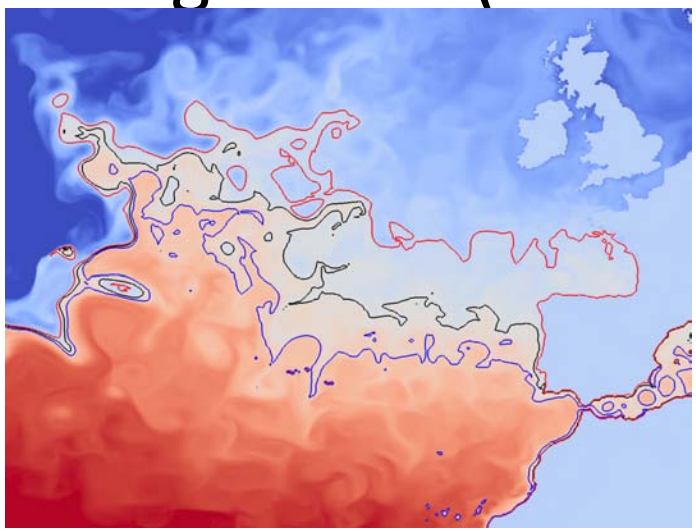


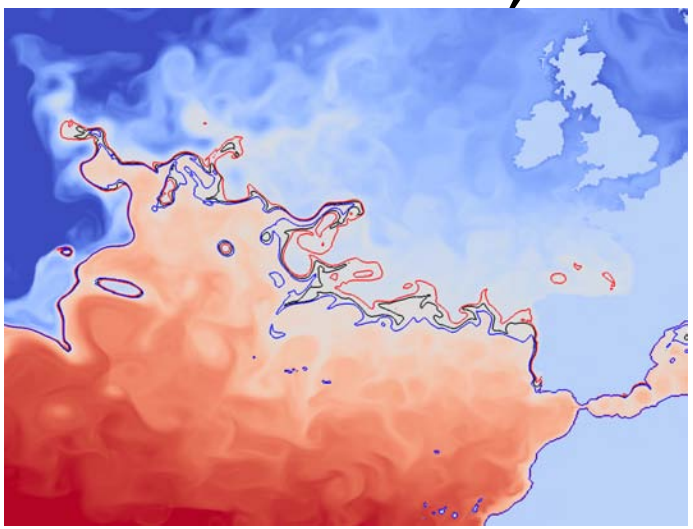


Los Alamos
NATIONAL LABORATORY
EST.1943

Isovalues on Compressed Simulation Data with Bounding Error - (32 bits, 3200x2400x42, 1.4 GB)

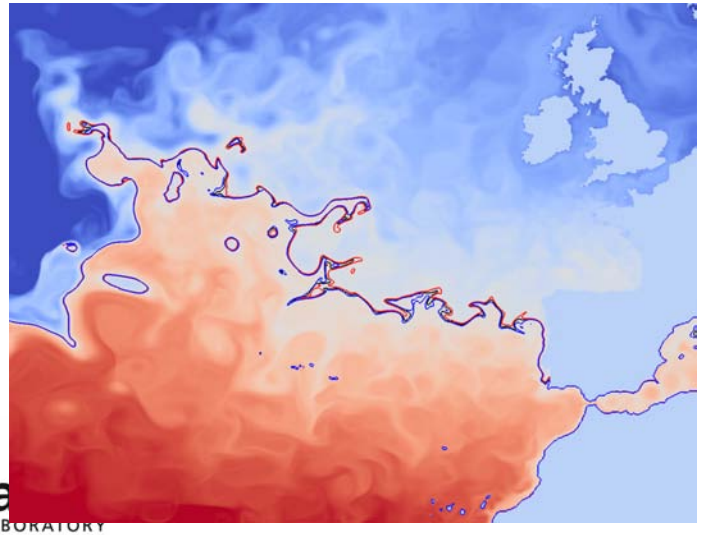
0.25 bits
10.8 MB

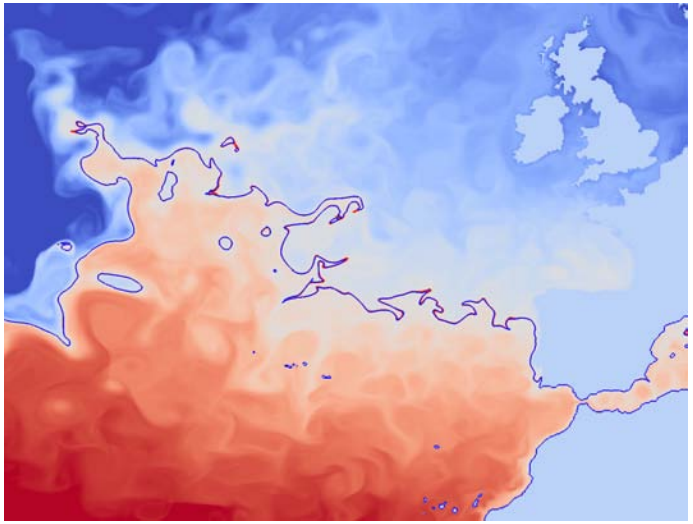




0.5 bits
21.6 MB

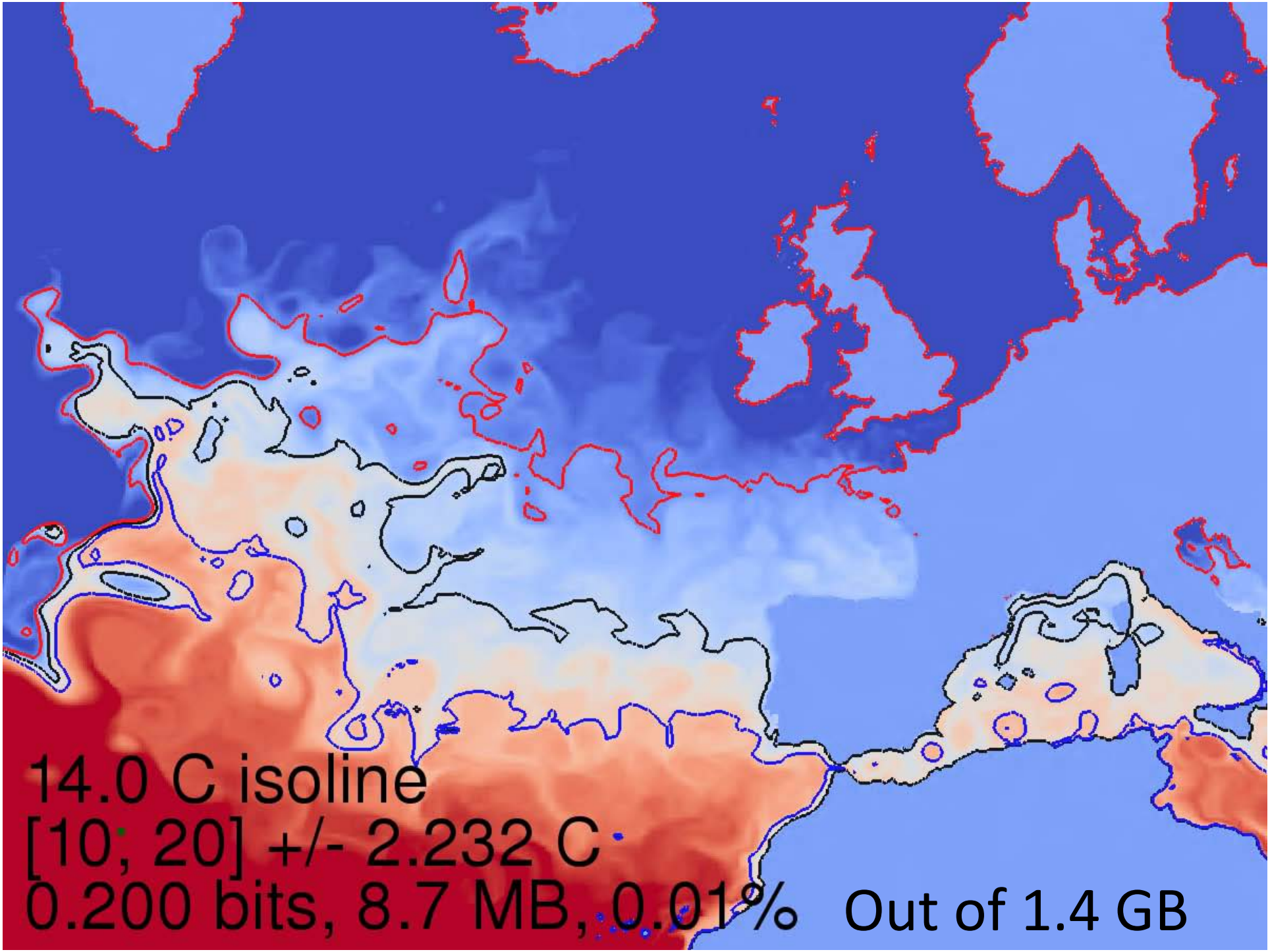
1.0 bits
43.3 MB





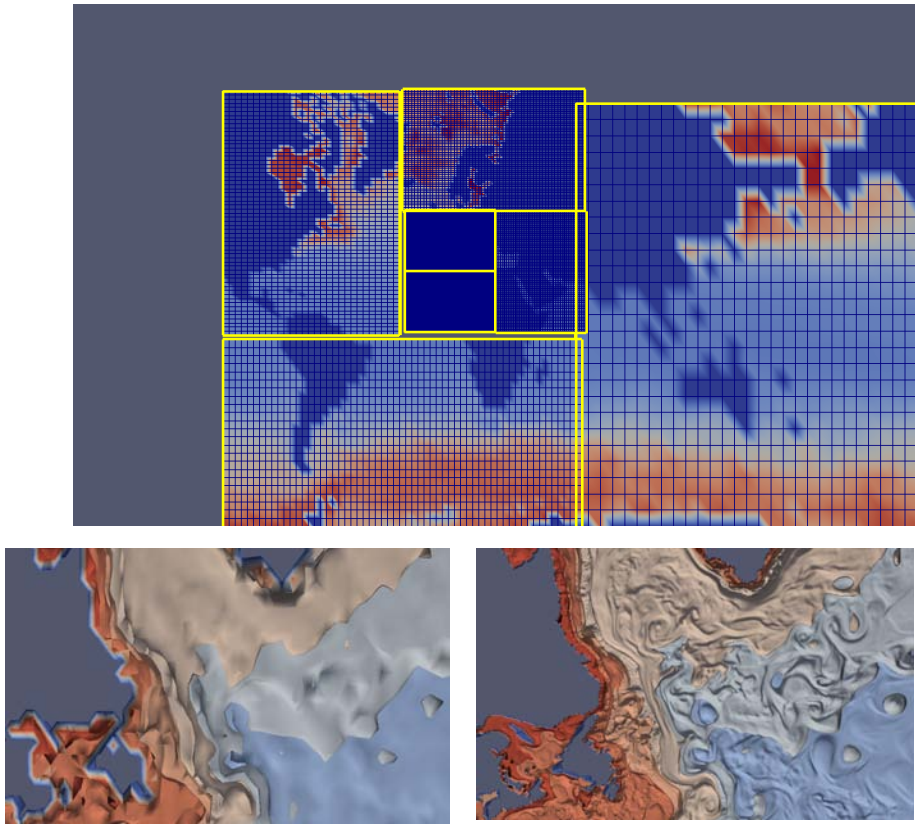
2.0 bits
86.5 MB

Los Alamos
NATIONAL LABORATORY
EST.1943

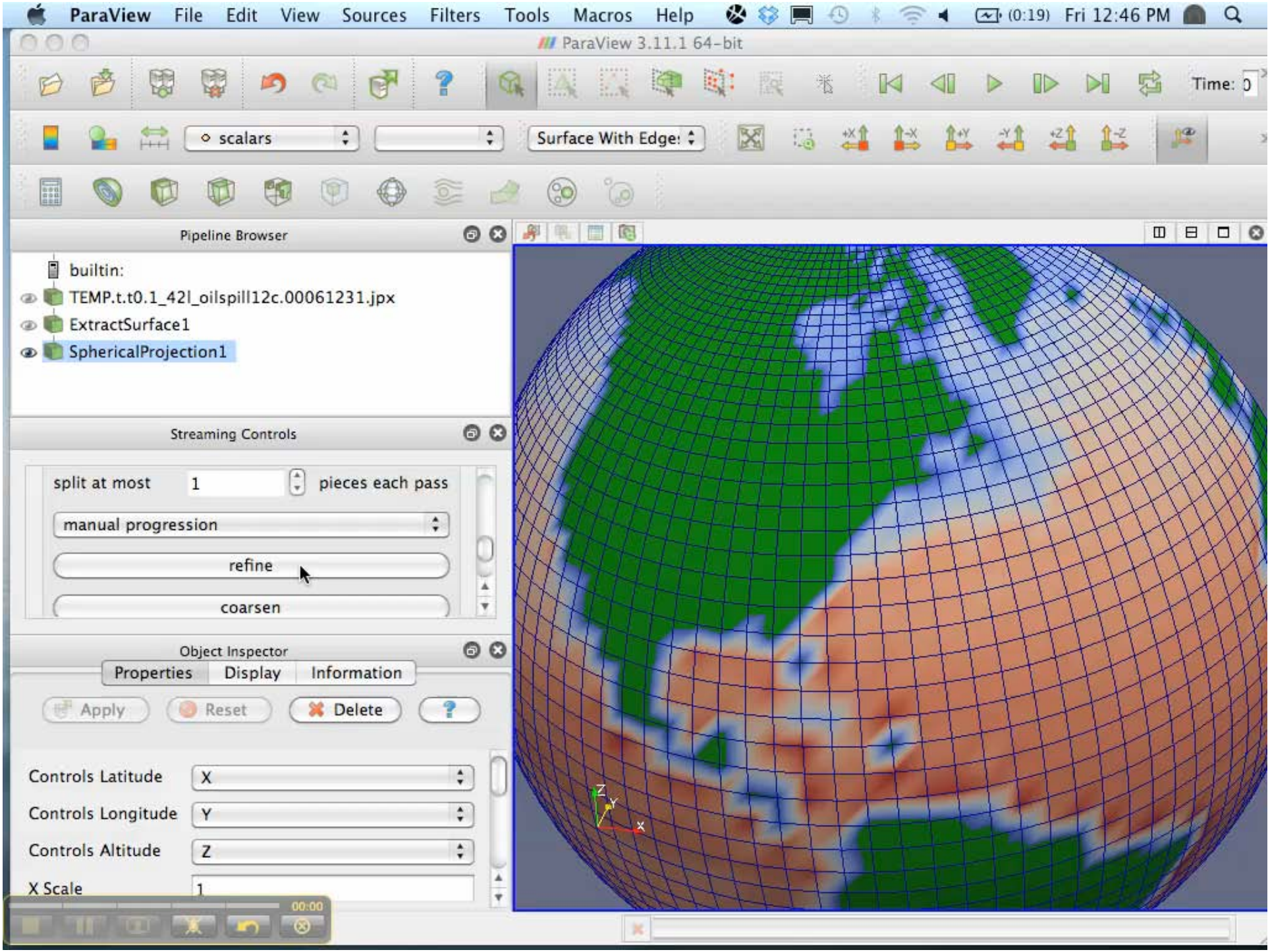

14.0 C isoline
[10; 20] +/- 2.232 C
0.200 bits, 8.7 MB, 0.01% Out of 1.4 GB

Multi-resolution Compression and Streaming

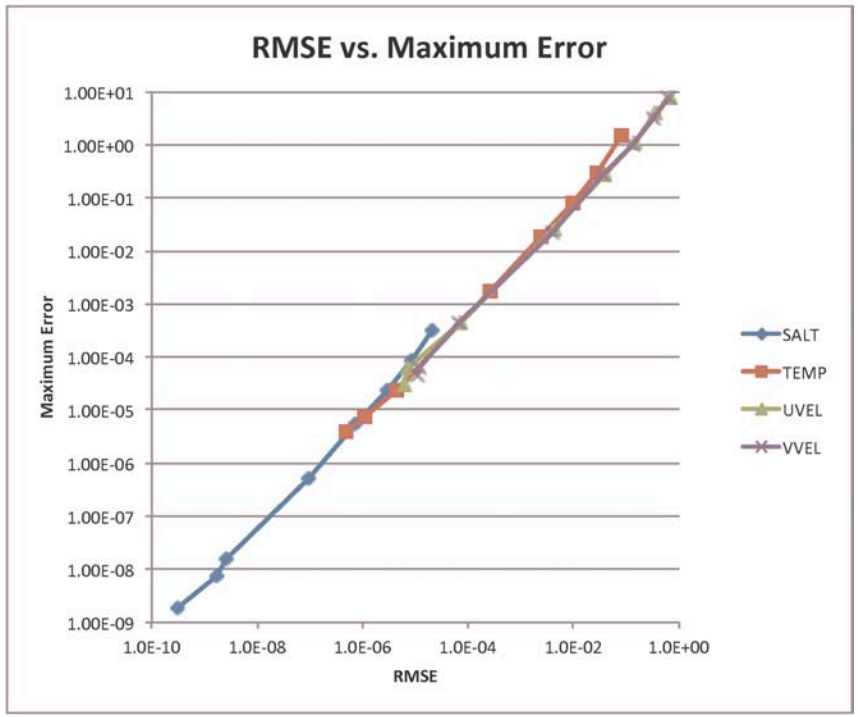
- A multi-resolution representation of simulation data is created using spatial compression or sampling
- View in a multi-resolution visualization and analysis tool
- Mat Maltrud, Climate Scientist, LANL: "This new distance visualization technology will increase our productivity by significantly reducing the amount of time spent in transferring and analyzing our remote data."

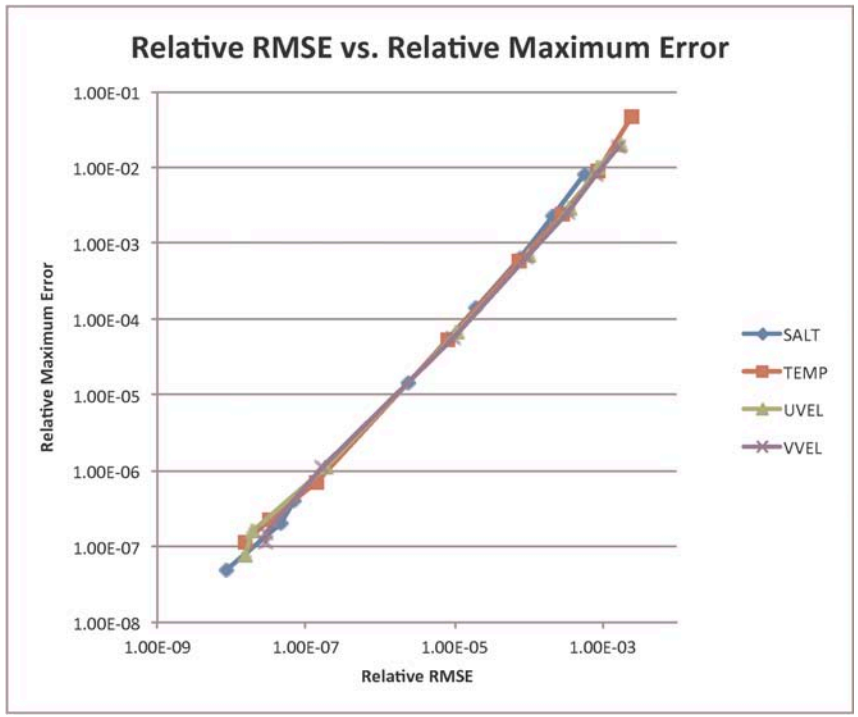


Images from multi-resolution streaming ParaView


ParaView File Edit View Sources Filters Tools Macros Help
ParaView 3.11.1 64-bit
Time: 0
scalars
Surface With Edge:
Pipeline Browser
builtin:
TEMP.t.t0.1_42l_oilspill12c.00061231.jpx
ExtractSurface1
SphericalProjection1
Streaming Controls
split at most 1 pieces each pass
manual progression
refine
coarsen
Object Inspector
Properties Display Information
Apply Reset Delete
Controls Latitude X
Controls Longitude Y
Controls Altitude Z
X Scale 1

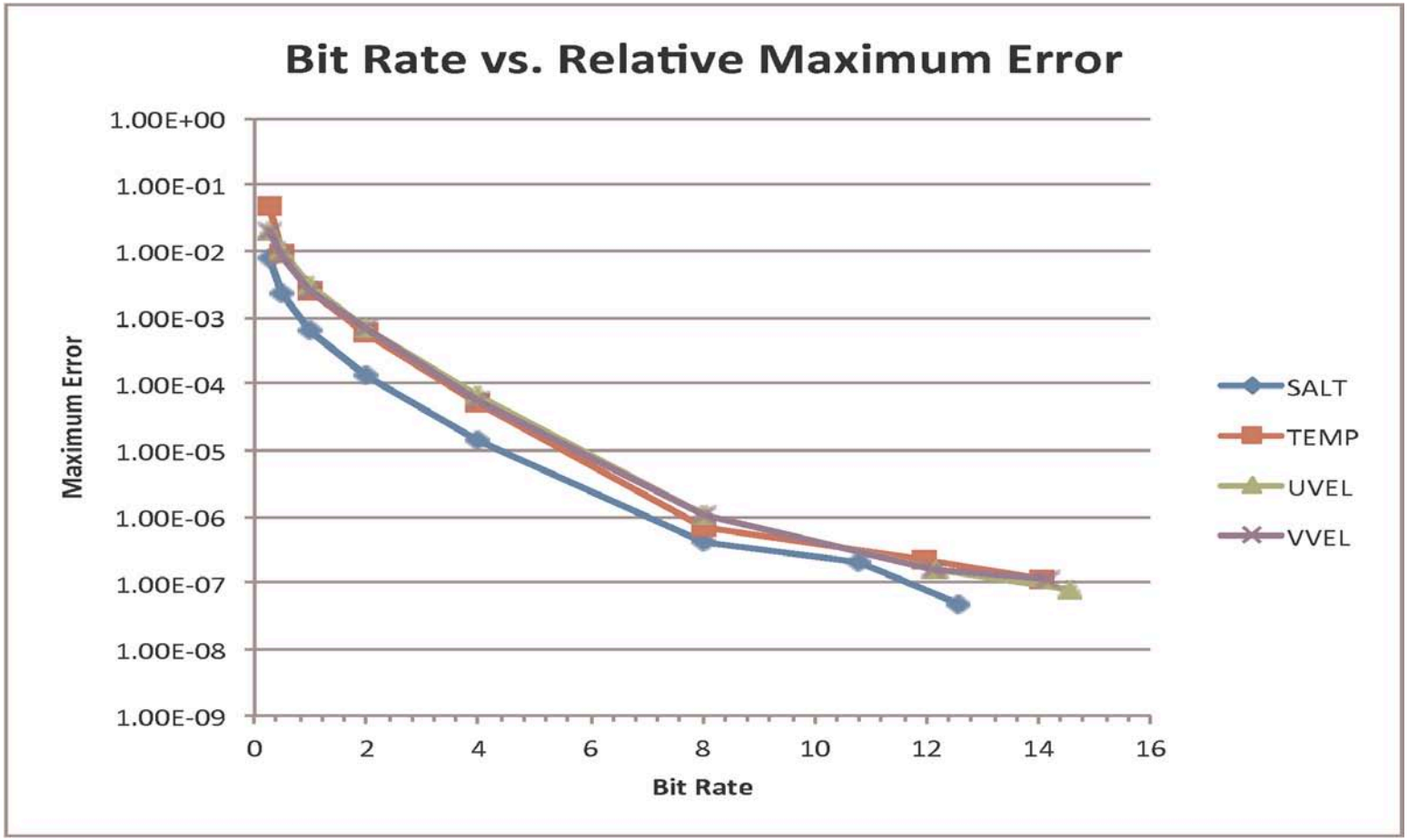
RMSE vs. Relative Max Error





Los Alamos NATIONAL LABORATORY EST.1943

Relationship between bit rate and Maximum Error



Los Alamos NATIONAL LABORATORY EST.1943

Acknowledgements

This work was supported fully by the DOE Office of Science, Advanced Scientific Computing Research (ASCR), program manager Lucy Nowell.

Operated by Los Alamos National Security, LLC for the U.S. Department of Energy's NNSA