Comparing and Understanding Data Distributions in EOS Applications

David Kao (NASA Ames Research Center)
Jennifer L. Dungan (NASA Ames Research Center)
Alex Pang (University of California, Santa Cruz)
Han-Wei Shen (Ohio State University)

First Year Progress Report
May 14, 2002

Goal & Technical Objectives

- Maps representing bio- and geo-physical variables result from many Earth science models. The models may generate full distributions at each location describing the spread of probable values.

- Much in the same way confidence intervals are reported for non-spatial variables, there is a need to report distributions to fully represent model output.

- The goal of this project is to develop techniques to aid in the understanding and comparison of these 2D distribution data sets.

Goal & Technical Objectives (Cont’d)

- The driving application will be NASA Office of Earth Science applications using Earth Observing System (EOS) images to derive 2D bio- and geo-physical fields representing snapshots or time-series of the Earth's surface.

- We seek answers to the following questions:
  - How applicable are current multivariate analyses and visualization techniques to this type of data sets?

  - What kind of visualization techniques need to be developed to present important aspects and features of distribution data sets?

Goal & Technical Objectives (Cont'd)

- What if the distribution data set to be analyzed is time varying?

- How would the techniques above be extended?

- What kind of time-varying features can we extract and exploit in helping the user understand data better?

- How should the analyses and visualization tools above be further expanded to allow comparisons of distribution data sets?

Technical Problem Statement

- The state of the art in data understanding using visualization tools can effectively deal with scalar and vector data sets from a broad range of applications.

- To a limited extent, there are also tools available for analyzing higher order tensor data and multivariate data sets.

- However, there are no standard visualization techniques currently available for analyzing data sets that contain a distribution at every location.

Technical Approach

- Develop a set of techniques and algorithms for visualizing, understanding and comparing distribution data sets.

- Some of these techniques will be implemented in a set of modular and extensible software tools for visualizing single distribution data sets and for pairwise comparison.

Data and NASA Relevance

- We have used three data sets on which to test our techniques:
  - a data set representing predicted vegetation amount using remotely sensed reflectances published by Dungan (1999), generated using geostatistical conditional simulation
  - a data set representing temperature and salinity from an ocean model
  - Global NDVI (Normalized Difference Vegetation Index) from NOAA based on AVHRR (Advanced Very High Resolution Radiometer) data. 3rd Generation C-level NDVI imagery from 1985 to 1997.

Data and NASA Relevance (Cont’d)

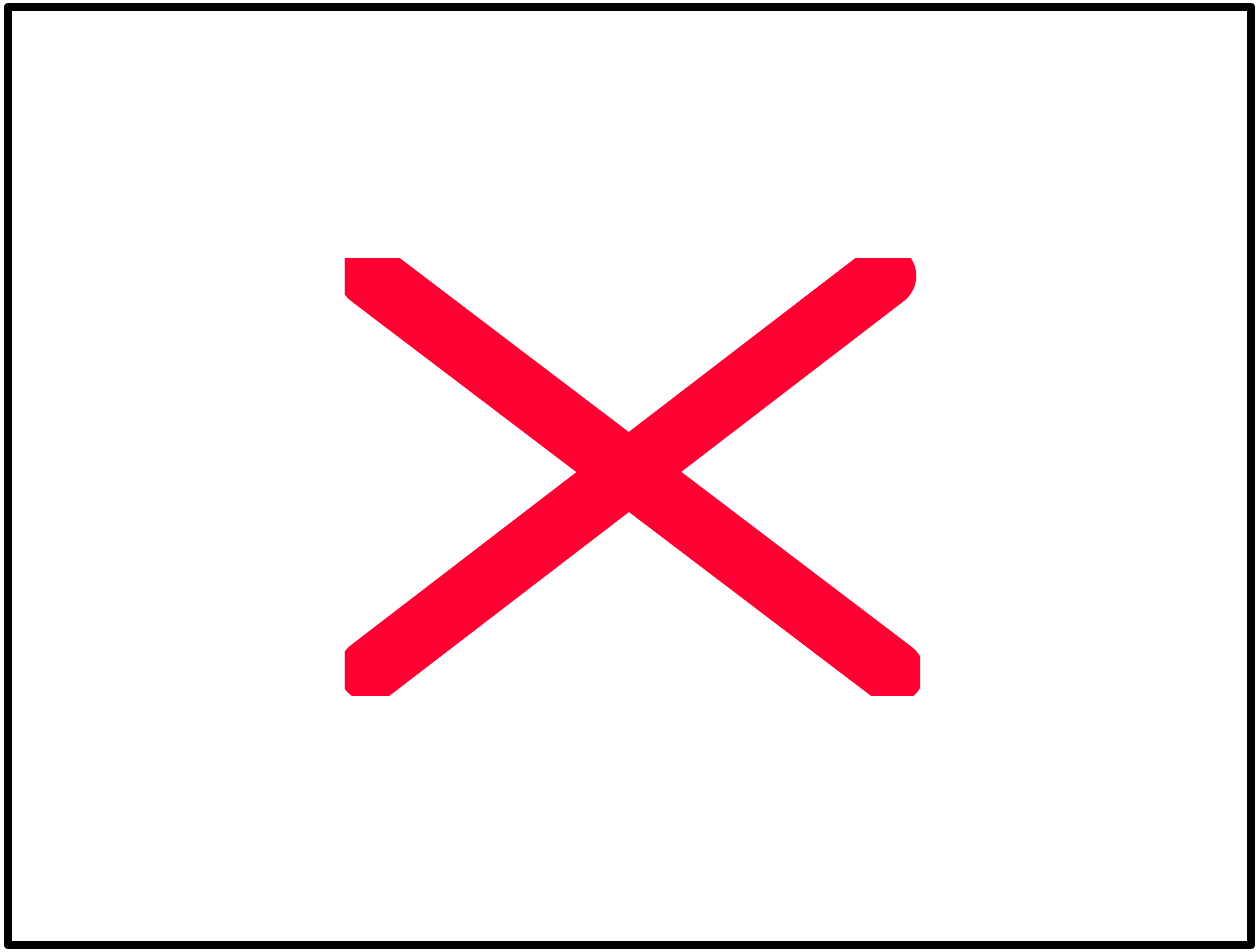
- EOS Level 3 or 4 data products will be used when they become available.

- Results of this work will be a new set of techniques and tools that can be used by the community of NASA Earth Science researchers and can be adapted for use by other research communities.

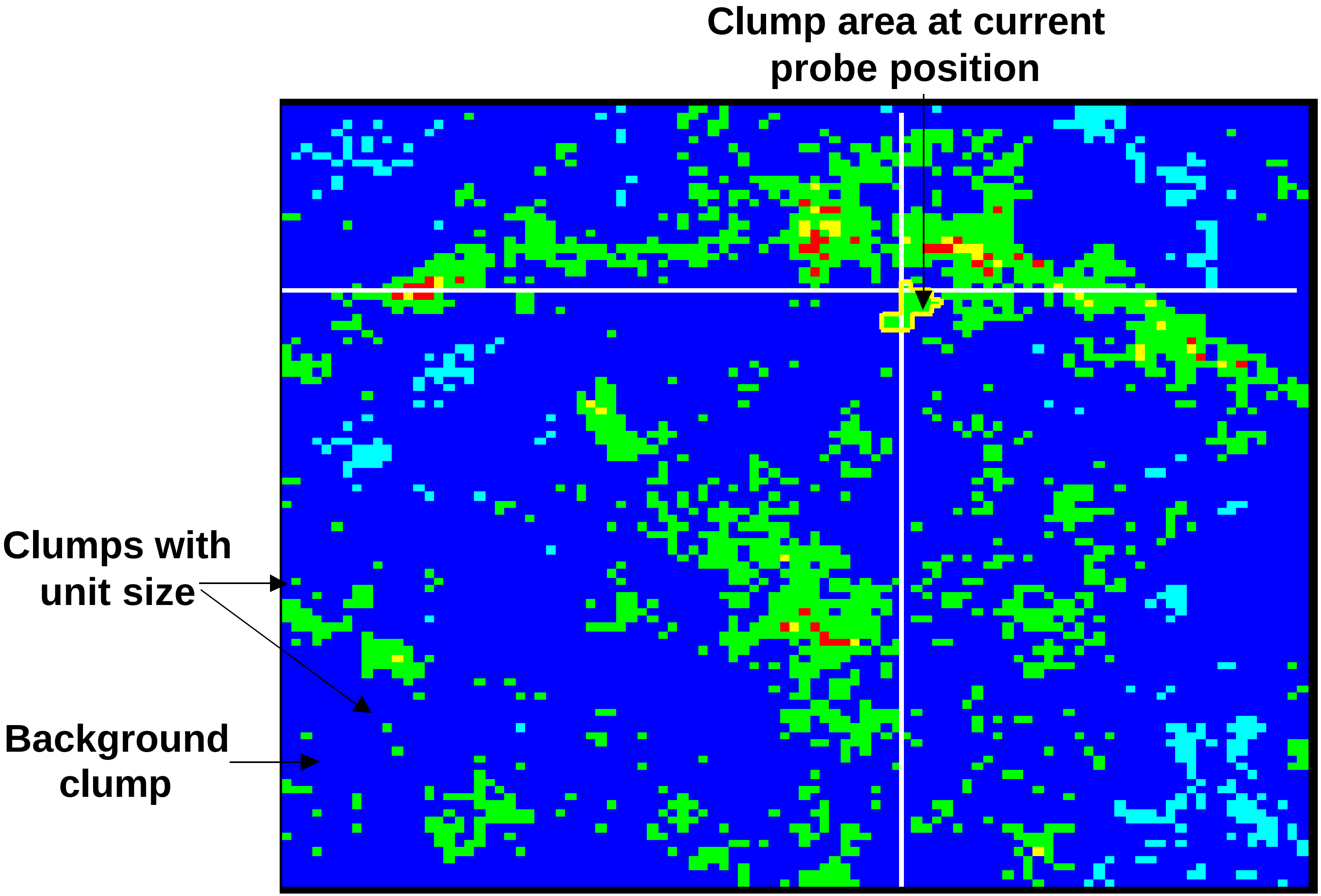
Accomplishments & Preliminary Findings

For visualizing/analyzing static distribution data:

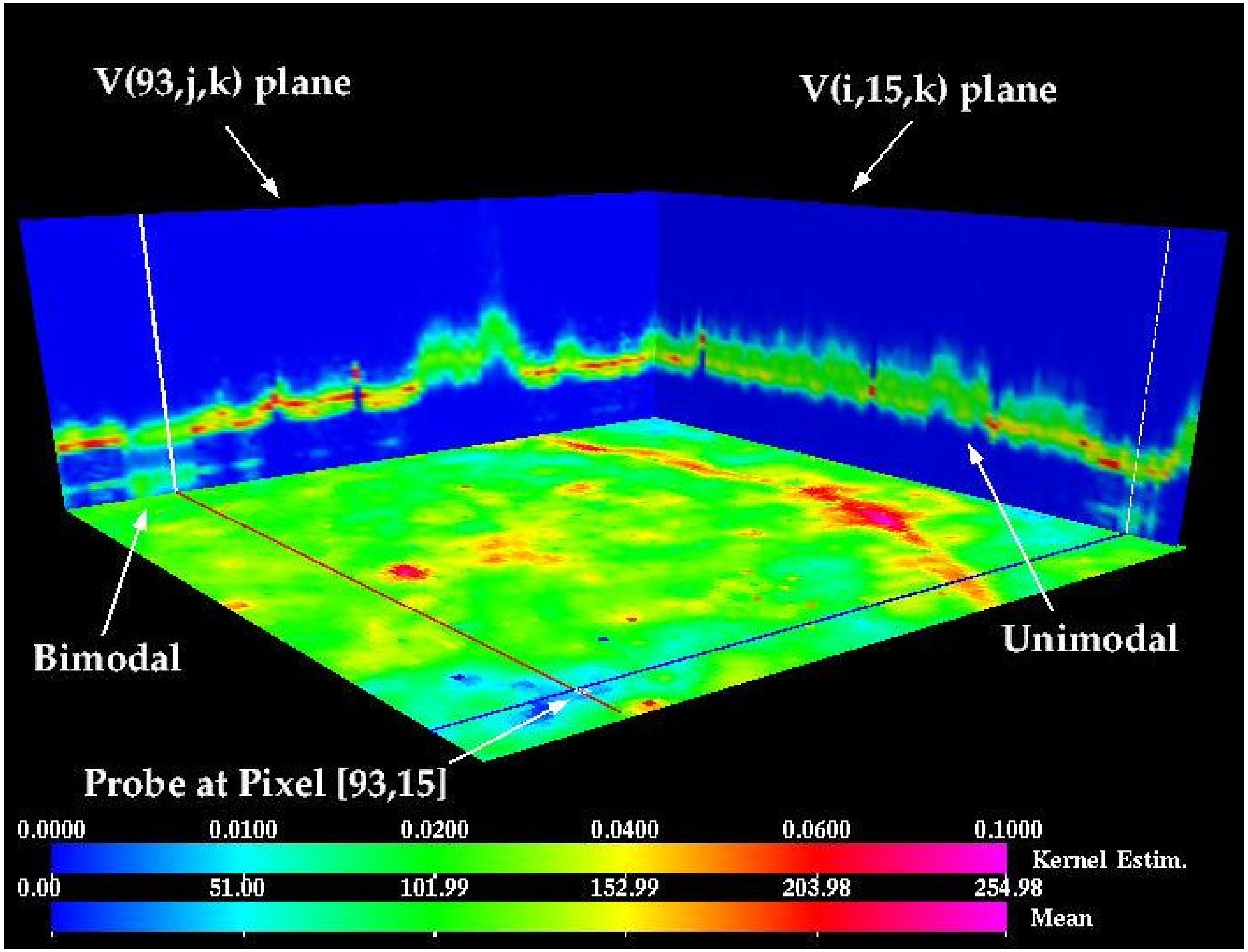
- Used a parsimonious set of statistics for summarizing central tendency, spread and shape for pixel-wise distributions
- Used value thresholds to define features in 2D and then generated distributions of the areas of these features. Feature-wise distributions are then summarized.
- Summaries were rendered using standard image visualization techniques such as contour planes, surface graphs and line bars.
- Used these techniques to capture multiple summaries in a single figure.



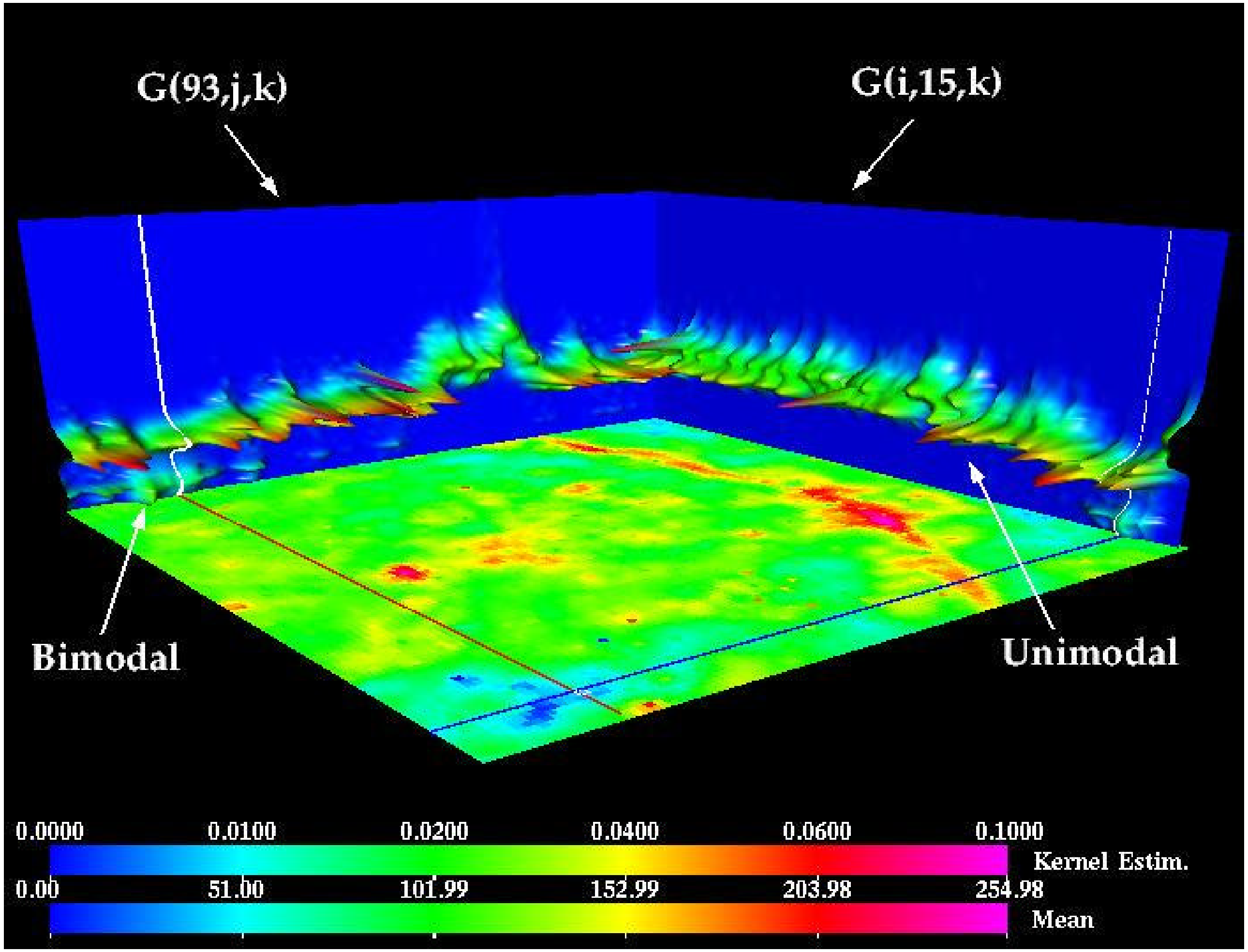
Multiple statistical summaries at each pixel, displayed simultaneously. Mean (bottom plane), kurtosis (deformed surface graph), skewness (color contour on the surface graph), and |Mean-median| (line bars).



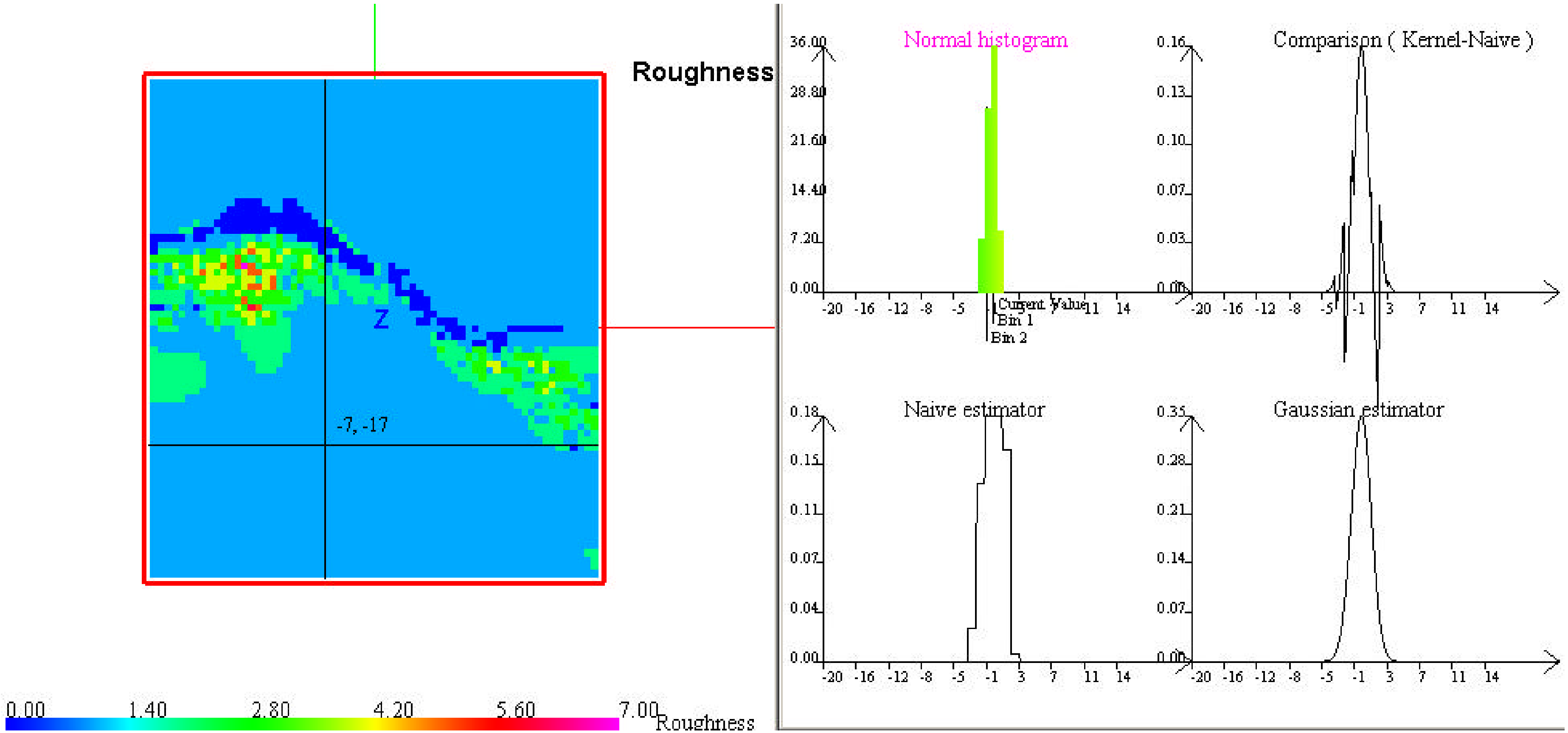
An interactive probe reports pixel-wise and feature-wise summaries at the current probe position.



Profile of distributions along a row and a column in the bottom image plane. Left wall: distribution along the blue line. Right wall: distribution along the red line.



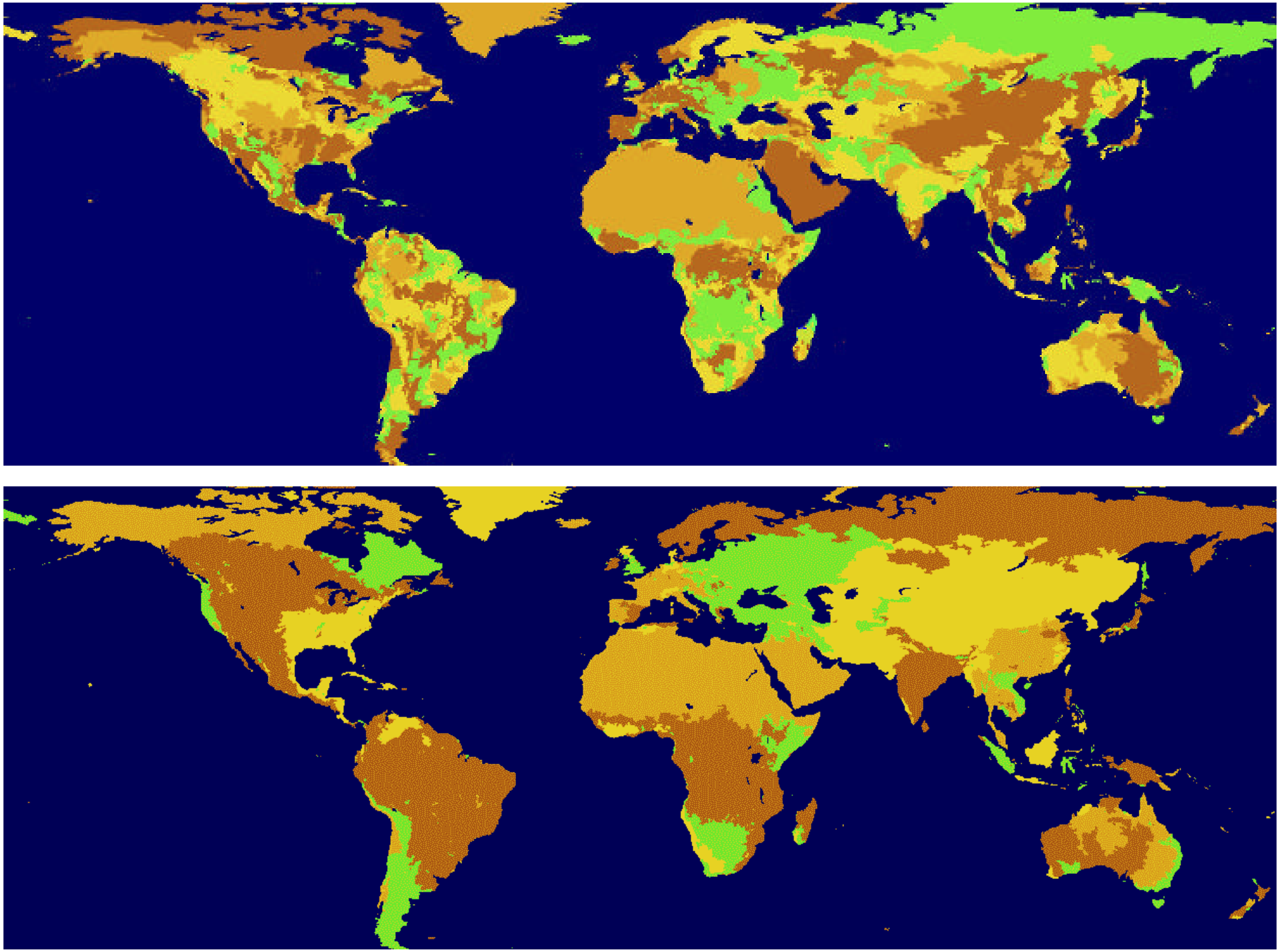
Surface graph representation of distributions along a row and a column in the bottom image plane. Left surface graph: distribution along the blue line. Right surface graph: distribution along the red line.



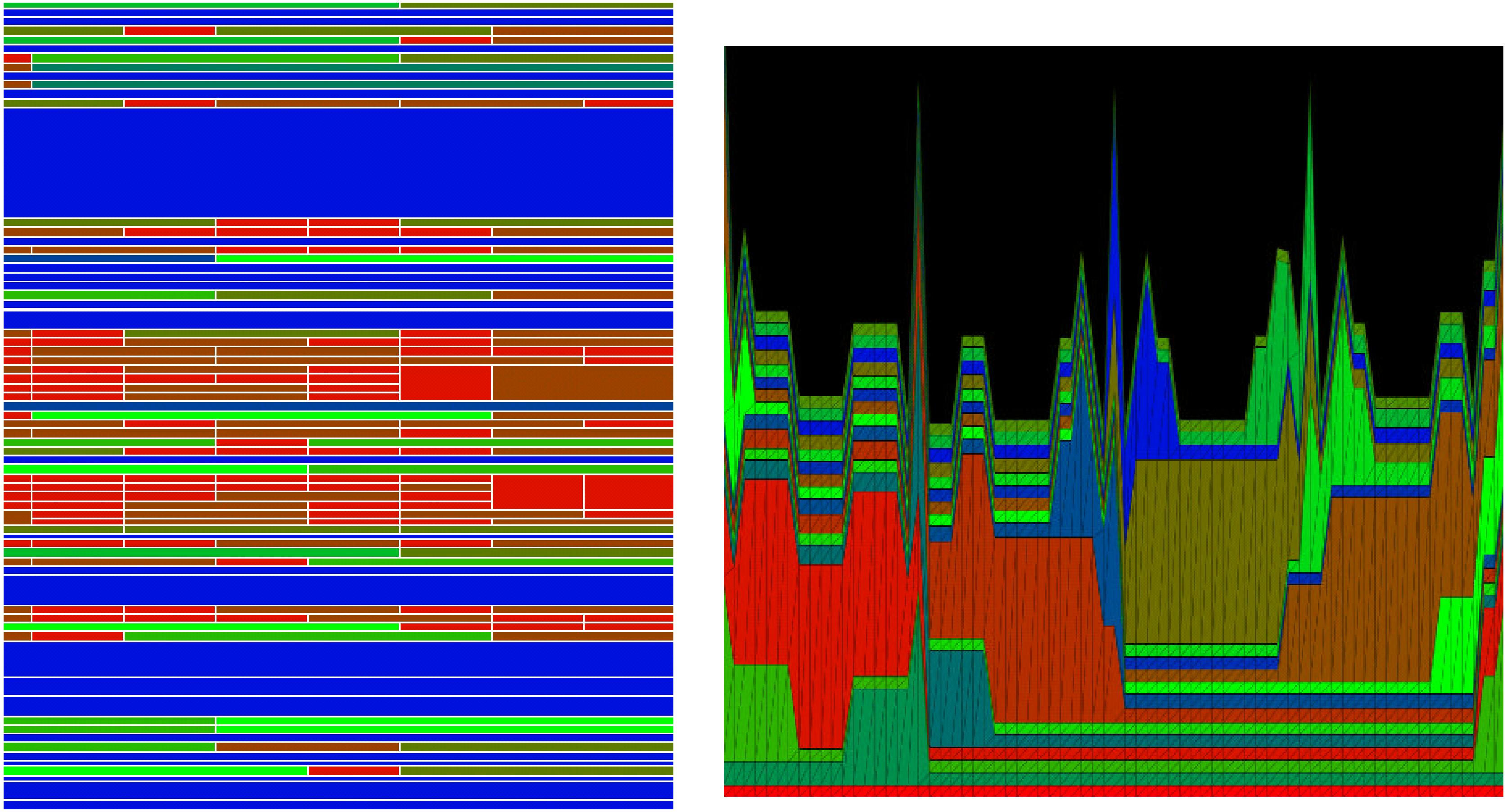
Roughness of the distributions (left). The shape descriptor of the pixel at the current probe location is represented using different density estimates (right graphs).

Accomplishments & Preliminary Findings (Continued)

- For time varying distributions:
  - Explored spatial and temporal clustering to describe the map's temporal variations and spatial correlations
  - Developed efficient data structures for fast temporal-spatial region indexing
  - Developed techniques to summarized the temporal behavior of the data in time intervals of arbitrary lengths
- Software to implement the above techniques was developed on SGI, Sun, and Windows using C++.



Hierarchical clustering of time-varying NVDI data set at two different levels of detail



Abstract views of the temporal variations in different spatial regions: Left – bands of different heights and widths represent different region sizes and temporal correlations, respectively. Right – same information represented using a ‘theme river’ scheme.

Technical Significance of Progress and Expected Impact on NASA

- Earth Observing System and related image data will continue to generate spatial products (land cover, vegetation fraction, aerosol concentration, sea surface temperature) that have associated spatial uncertainties.

- As models to describe this uncertainty mature, users will need ways to visualize and communicate this information both within the scientific community and to non-science users.

Linkable URL's

- For more information about this research, see the following url:

  http://www.cse.ucsc.edu/research/avis/nasa_is

Facilities Used and Personnel

- Computing facilities used:
  - NAS, UCSC, OSU, Ecosystem Science and Technology Branch at NASA/Ames

- Students:
  - Udeepta Bordoloi, OSU (Ph.D. student)
  - Anna Chen, UCSC (summer intern)
  - Newton Der, UCSC (summer intern)
  - Alison Luo, UCSC (Ph.D. student)

References

- Udeepta Bordoloi, Han-Wei Shen, David Kao, “Spatio-Temporal Data Visualization for EOS Data,” submitted to CODATA 2002 international conference.
- David Kao, Alison Luo, Jennifer L. Dungan, and Alex Pang, “Visualizing Spatially Varying Distribution Data”, accepted to Information Visualization '02.
- Alison Luo, David Kao, and Alex Pang, “A Methodology for Visualizing Distribution Data”, submitted to Vis2002.
- Jennifer L. Dungan, David Kao, and Alex Pang, “The Uncertainty Visualization Problem in Remote Sensing Analysis”, accepted to IGARSS'02.
- Jennifer L. Dungan, David Kao, and Alex Pang, “Definitions of Uncertainty”, white paper, http://www.cse.ucsc.edu/research/slvg/nasa_is/white.ps
- Alex Pang, “Visualizing Uncertainty in Geo-spatial Data” in proceedings of the Workshop on the Intersections between Geospatial Information and Information Technology, prepared for the National Academies committee of the Computer Science and Telecommunications Board. Arlington, VA, October 2001.
- David Kao, Jennifer Dungan, and Alex Pang, “Visualizing 2D Probability Distributions from EOS Satellite Image-Derived Data Sets: A Case Study”, in IEEE Visualization '01, pp. 457-460.