

# **Exploratory Spatial Data Analysis - Techniques and Examples**

**Jürgen Symanzik\*,  
Utah State University, Logan, UT**

**\*e-mail: [symanzik@sunfs.math.usu.edu](mailto:symanzik@sunfs.math.usu.edu)  
WWW: <http://www.math.usu.edu/~symanzik>**

**with Louise Griffith, Rob Gillies, Di Cook, Jim Majure, Nicholas Lewin-Koh,  
Noel Cressie, Sigbert Klinke, Ed Wegman, Adi Wilhelm**

# Contents

- ESDA
  - Definition
  - Software
  - Techniques
  
- Examples
  
- Live Demo

# ESDA

## ■ Working Definition:

- Find structure (cluster, unusual observations) in large and not necessarily homogeneous spatial data sets **based on human perception using graphical methods and user interaction**
- Goal or expected outcome of exploration usually unknown in advance

# Software: XGobi/GGobi

Swayne, Cook and Buja

- Interactive environment for exploring multivariate data
  - \* Linked views allow “linked brushing”
  - \* Univariate, Bivariate and Multivariate views of the data
  - \* Grand tour
  - \* Wide variety of methods
  - \* Open source
  - \* Free
- Caveats
  - \* XGobi only on UNIX and Linux platforms
  - \* GGobi also available for PCs but not yet fully developed

# Software: ArcView

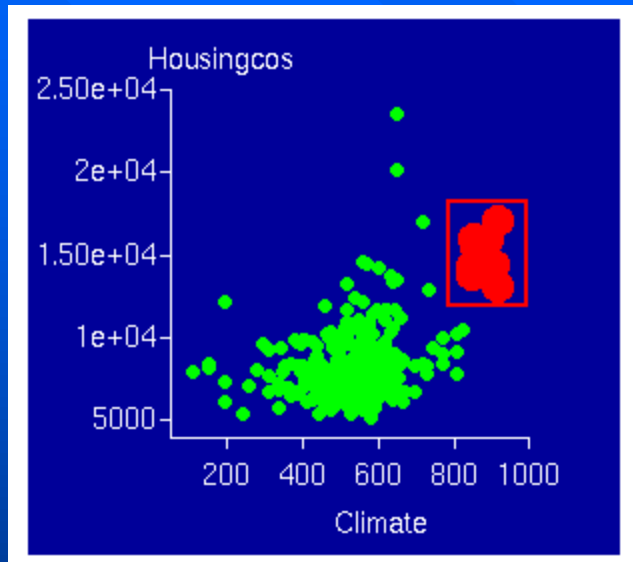
ESRI™

- Desktop GIS with wide Range of Viewing and Data Manipulation Functions
  - Editing Features
  - Query Operations
  - Map Display
  - Interactive Interface
  - High Level Internal Scripting Language

# Features of ArcView/XGobi-Link

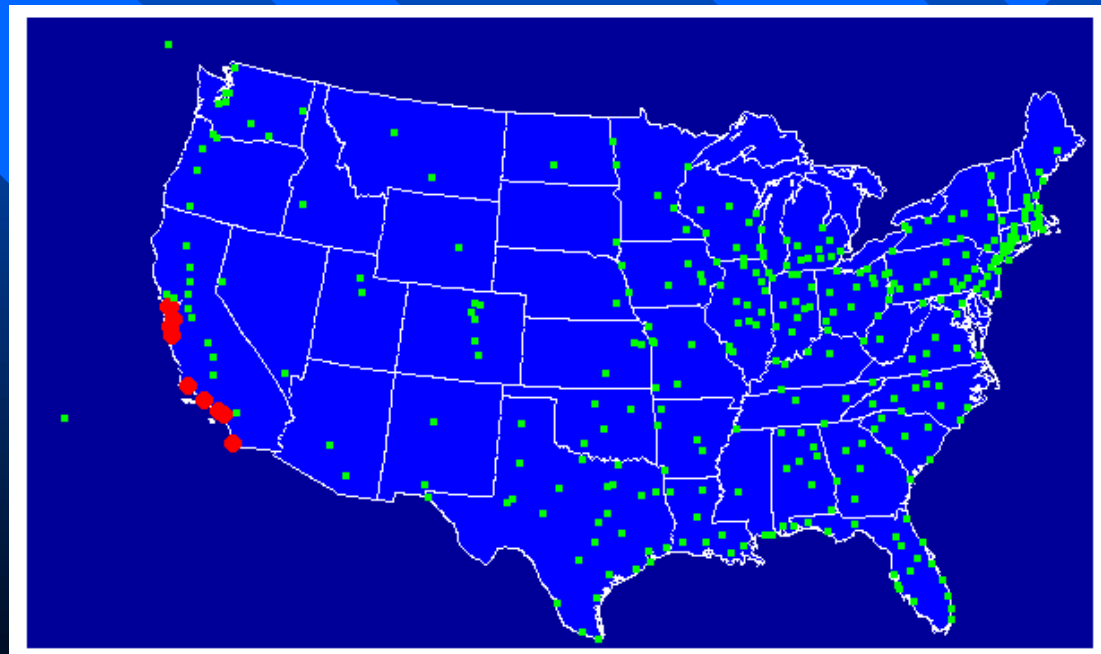
- Multivariate Link
- CDF Link
- Variogram-cloud Link
- Spatially Lagged Scatterplot Link
- Multivariate Variogram-Cloud Link

# Multivariate Link

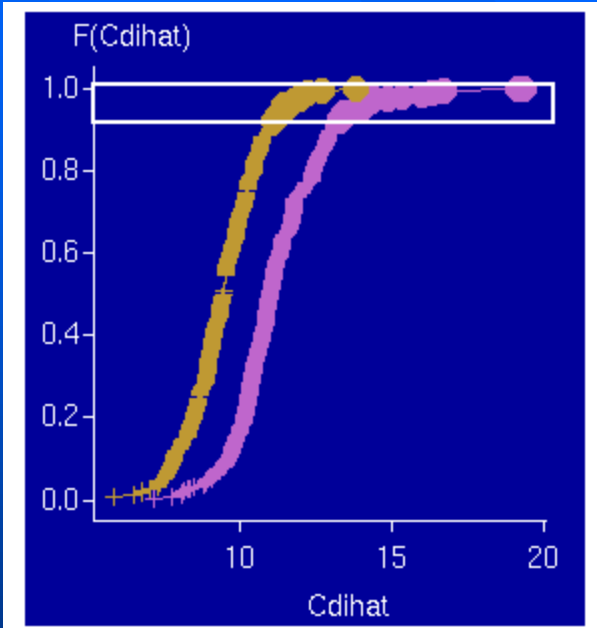


**XGobi**

**ArcView**

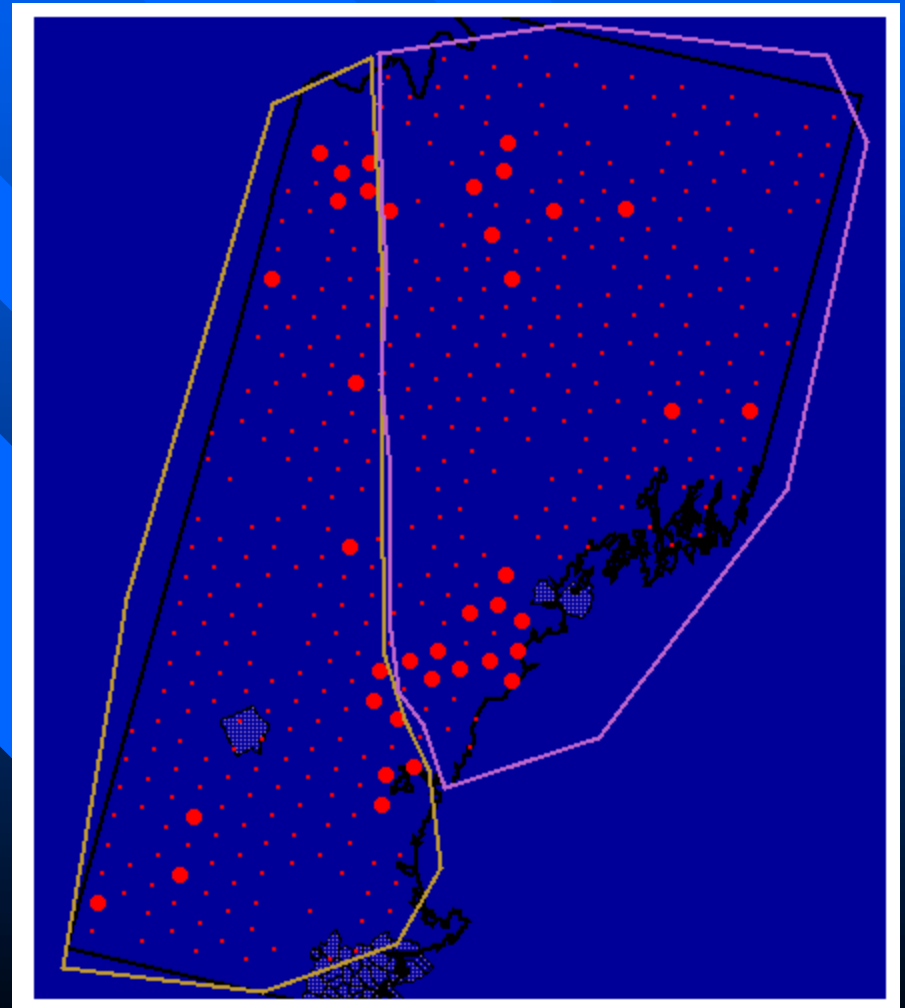


# CDF Link



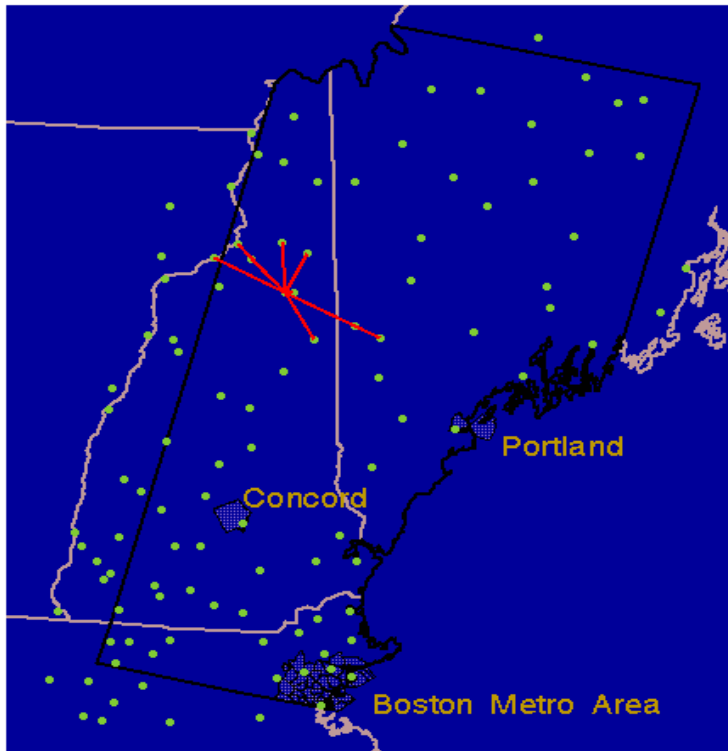
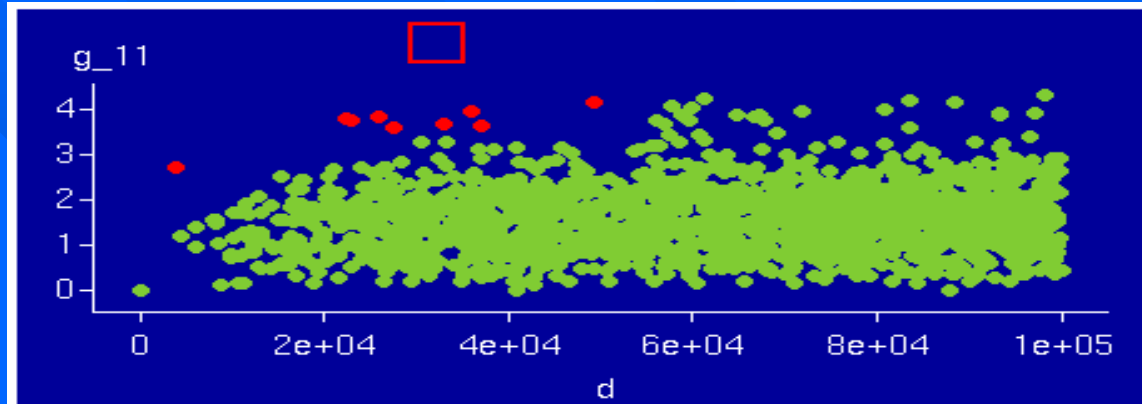
**XGobi**

**ArcView**

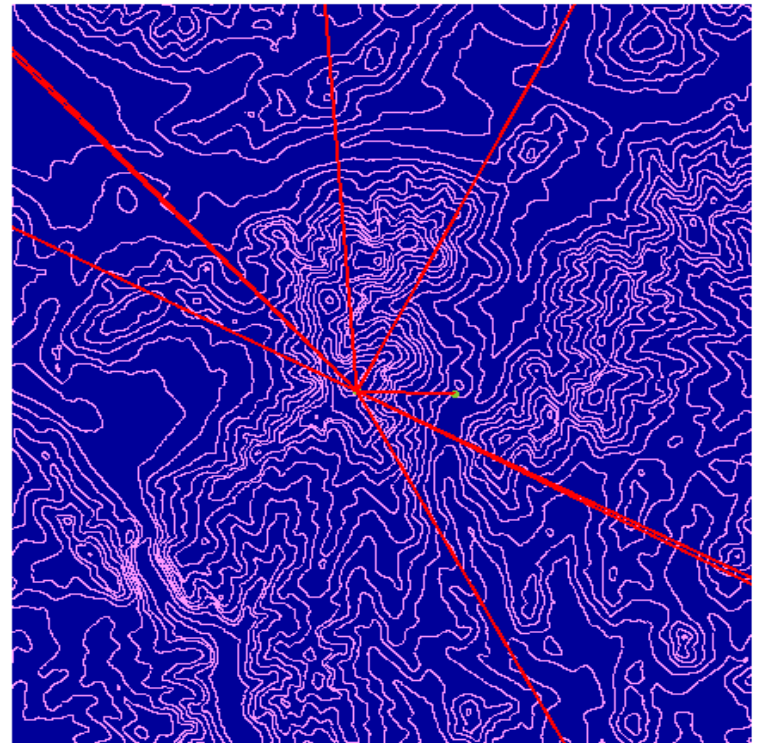


# Variogram-Cloud Link

**XGobi**

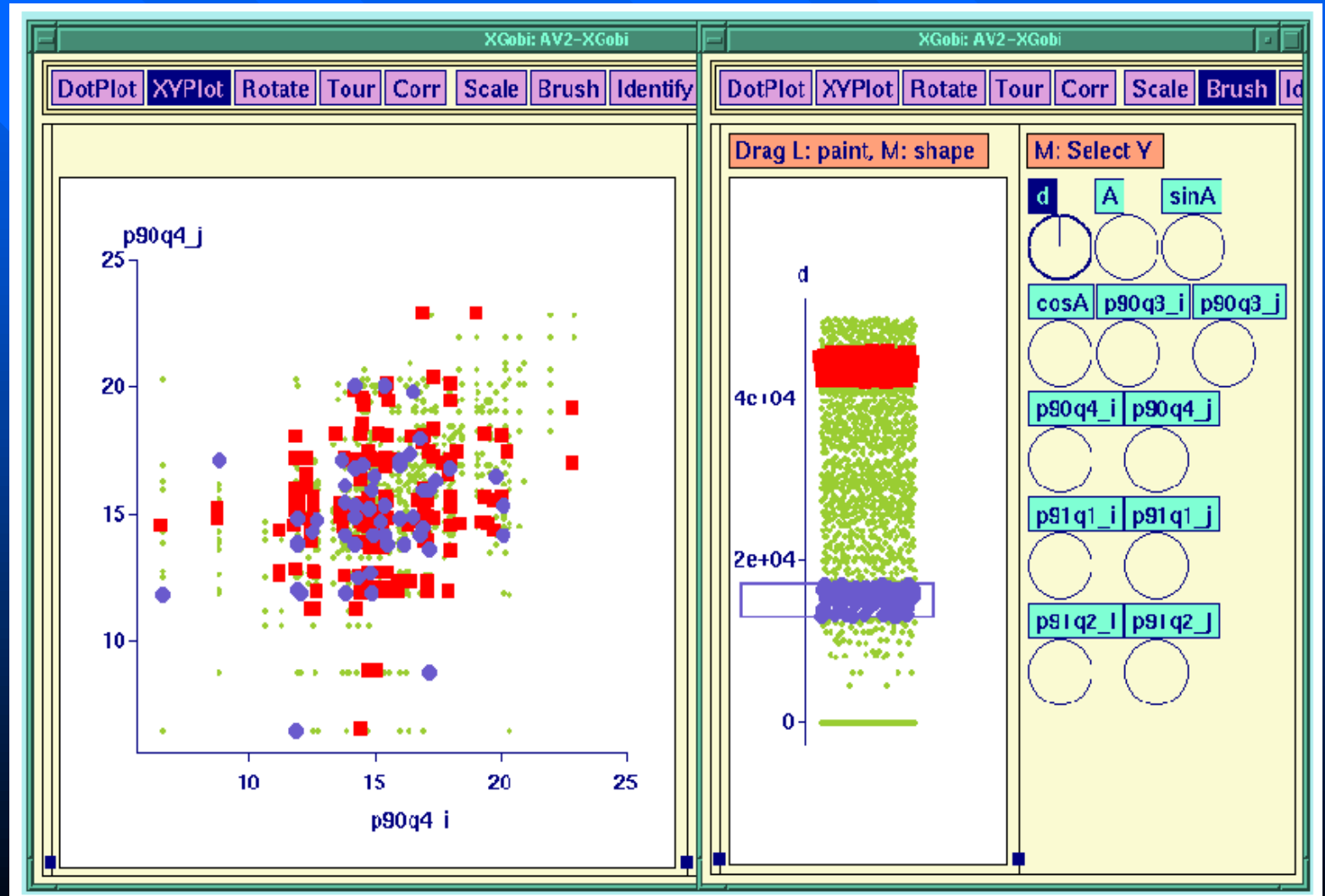


**ArcView**

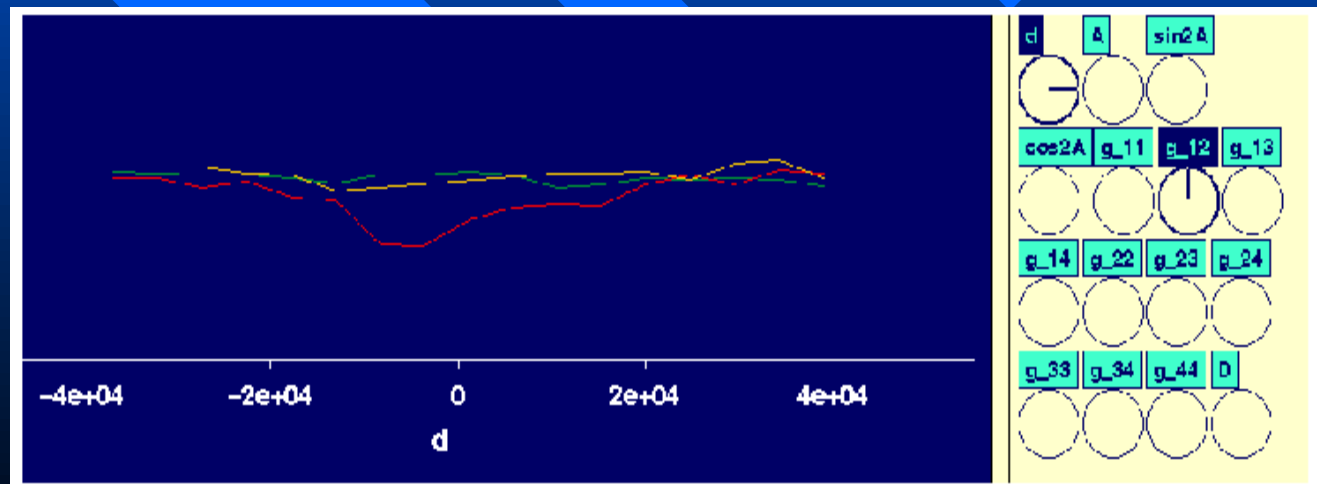
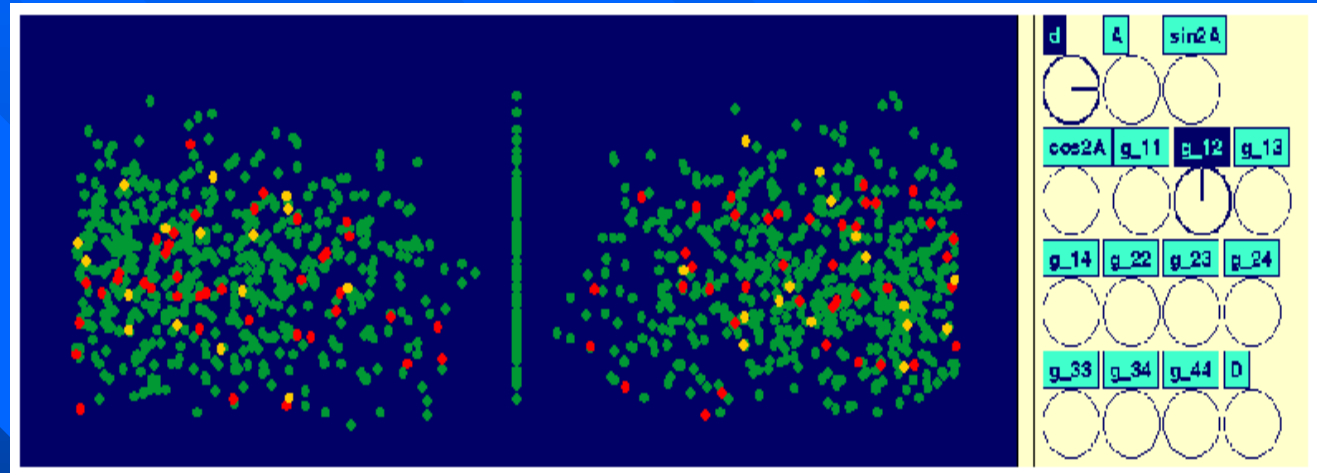
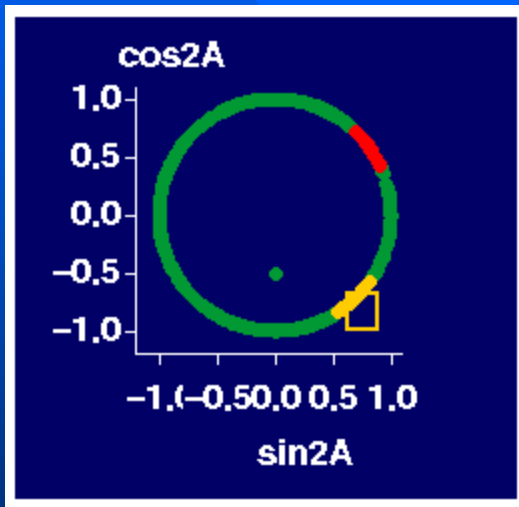


# Spatially Lagged Scatterplot Link

XGobi



# Multivariate Variogram-Cloud Link



**XGobi**

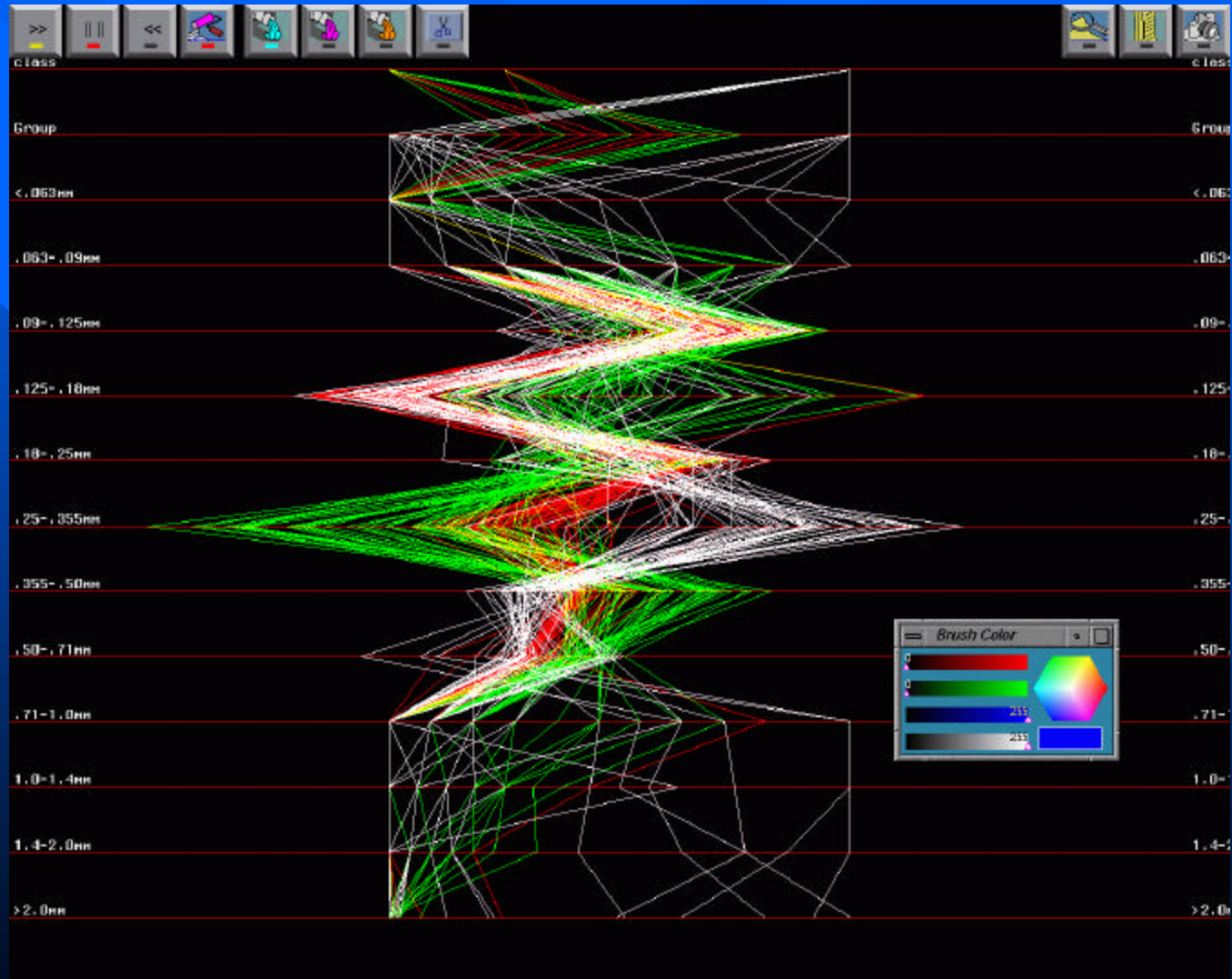
# Software: ExplorN/CrystalVision

Wegman, Luo, Carr

- Interactive environment for exploring multivariate data (similar to XGobi/GGobi)
  - \* Advanced Parallel Coordinates Displays
  - \* 3D Surfaces
  - \* Stereoscopic Displays
- Caveats
  - \* ExplorN only on SGI platforms
  - \* CrystalVision available for PCs but not yet fully developed
  - \* No interface to GIS software

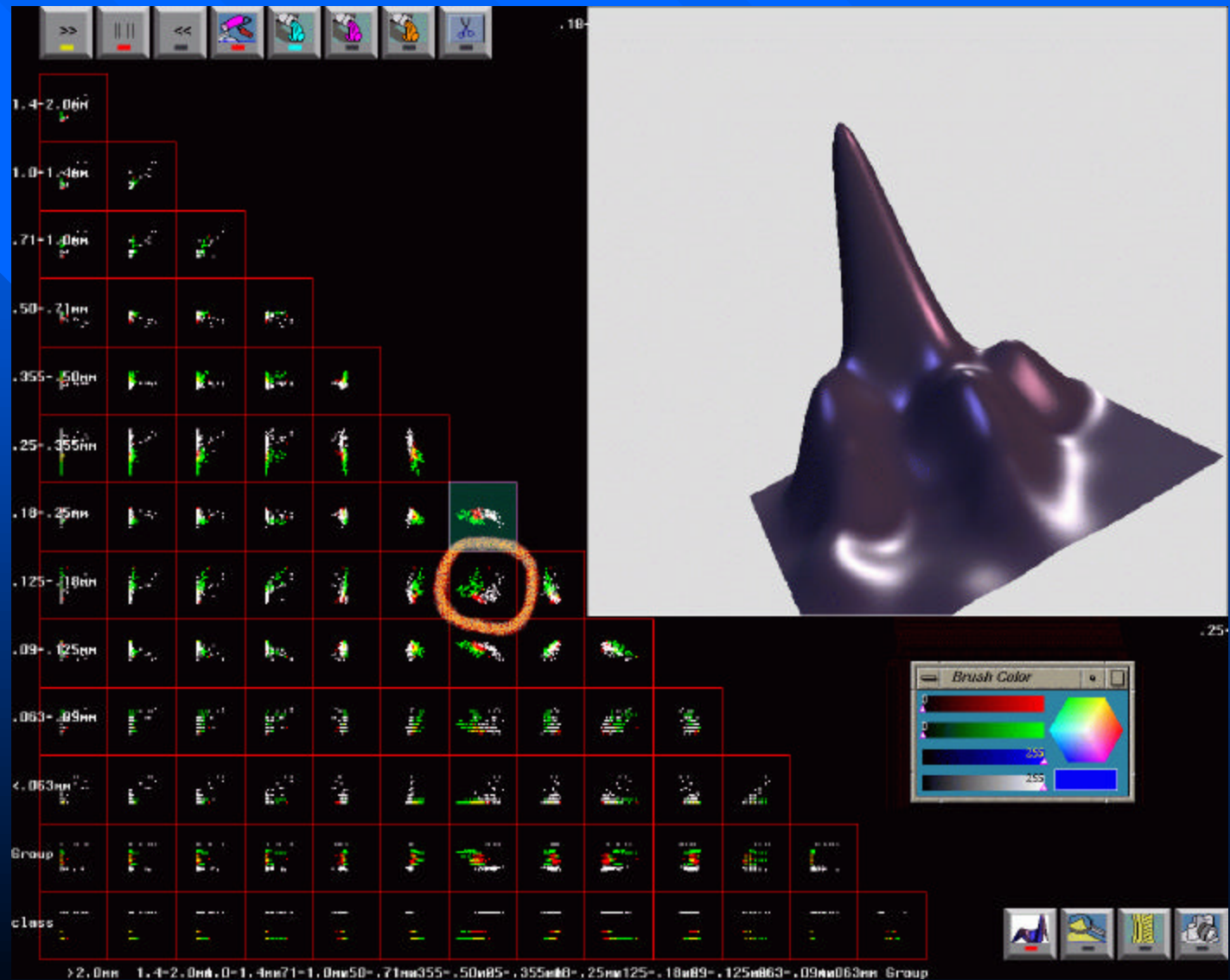
# Tools: Parallel Coordinate Plots

ExplorN



# Tools: Scatterplot Matrix & Density Plot

ExplorN



## Tools: Grand Tour

- Continuous random sequence of projections from  $n$  dimensions into 2 (or more) dimensions.

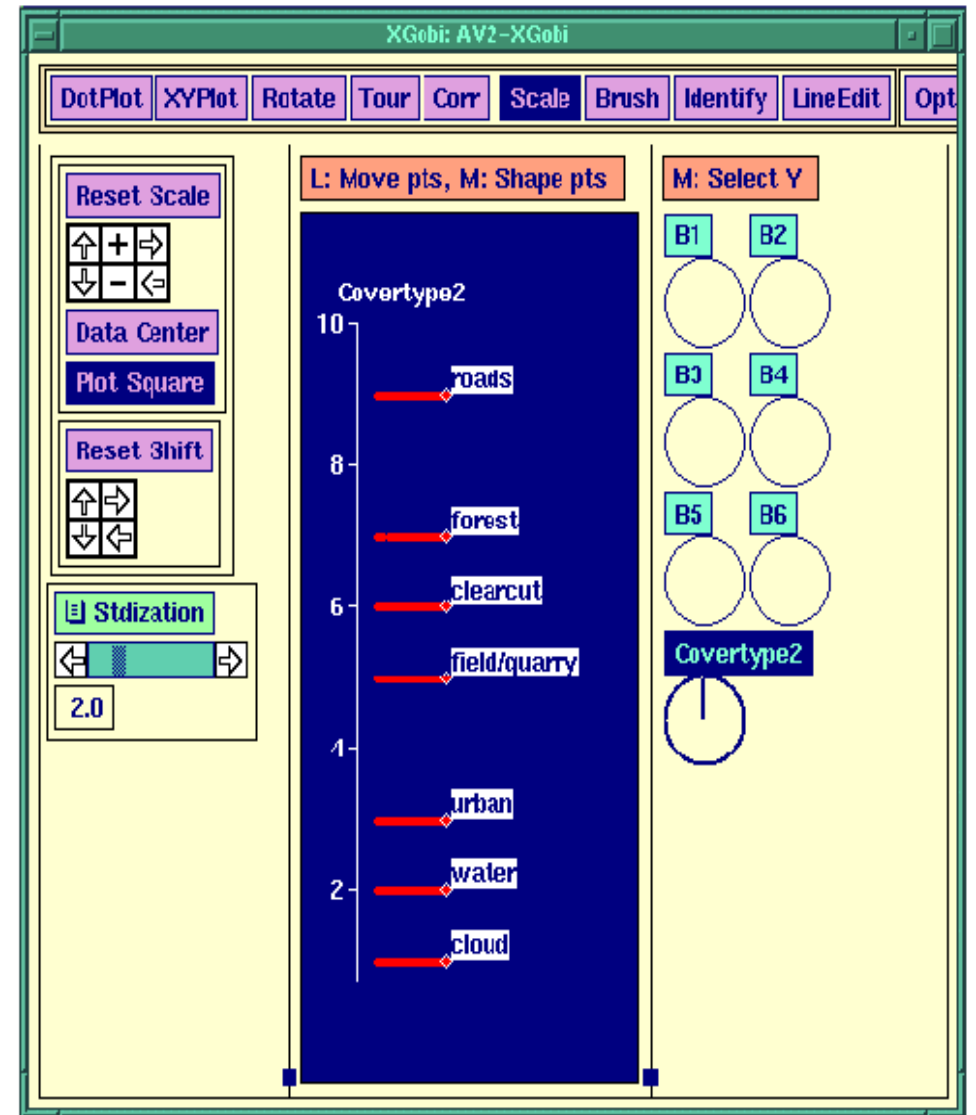
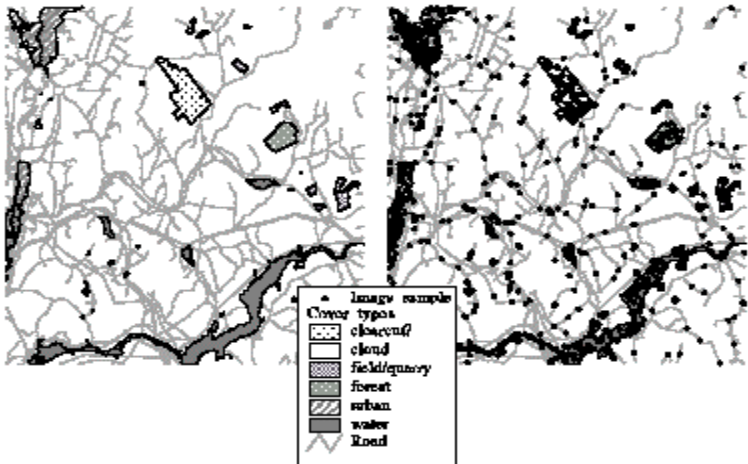
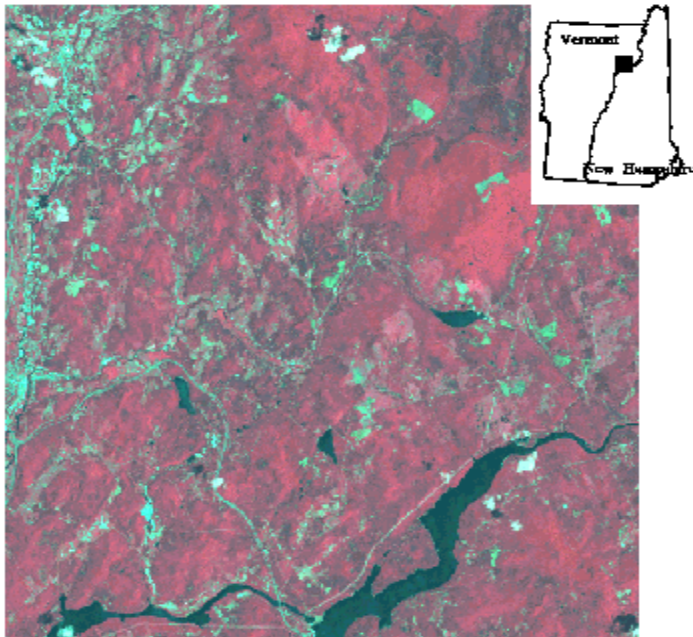
## Examples

- Remote Sensing: Vermont/New Hampshire
  - » Quarry, Water, Clouds
- Remote Sensing: Atlanta
  - » City, Forest
- Archaeological Data
  - » Sand Particles

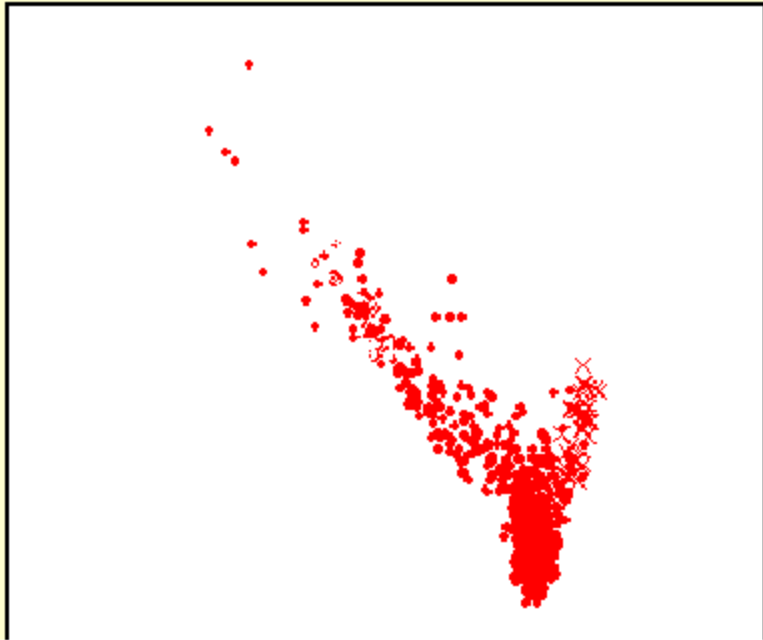
# The Vermont/New Hampshire Data

- Landsat Thematic Mapper (TM) data
- 6 Spectral Bands
- Outstanding Water Body is Connecticut River

# VT/NH: Field/Quarry/Water

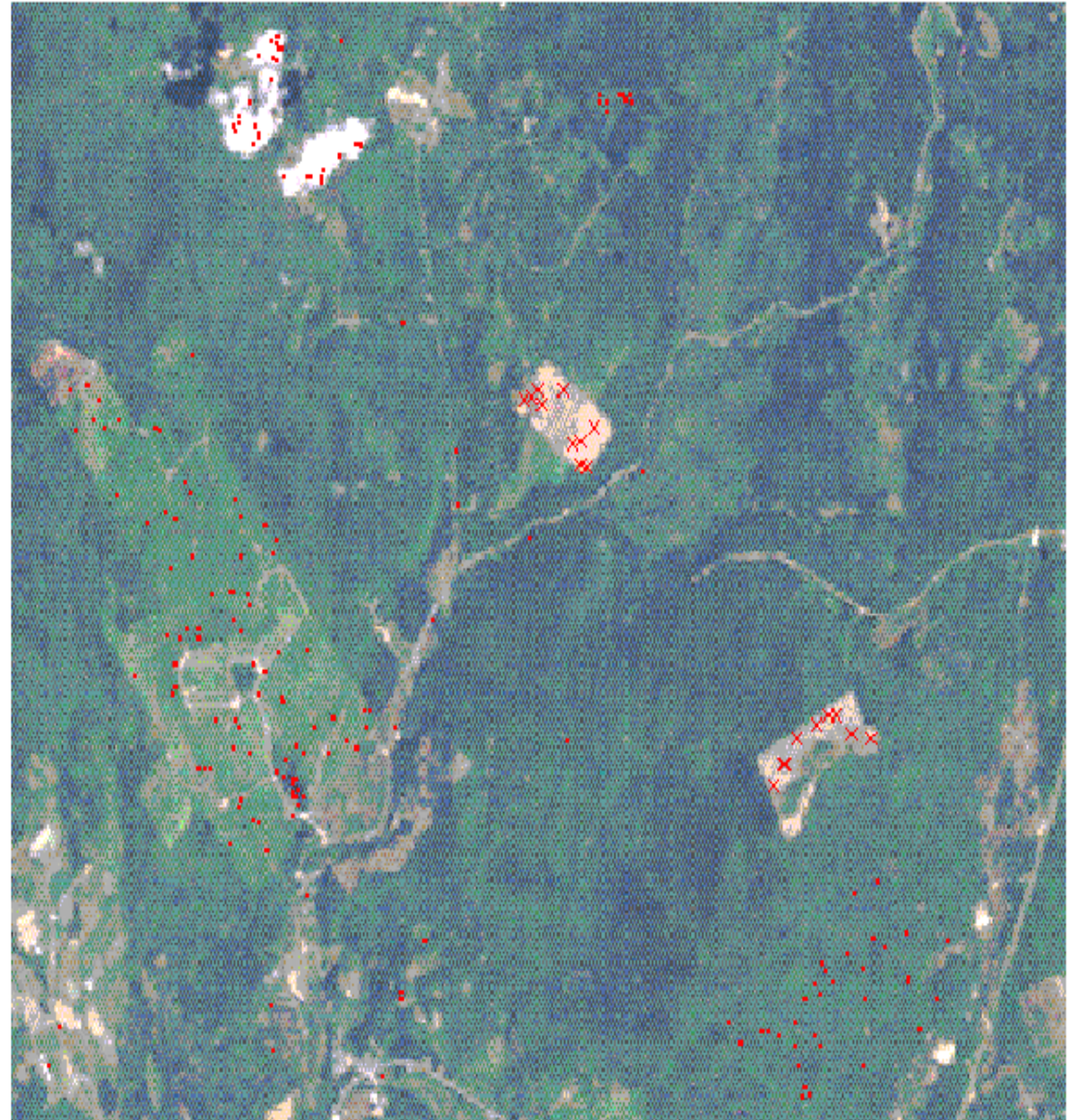


# Field/Quarry/Clouds ??

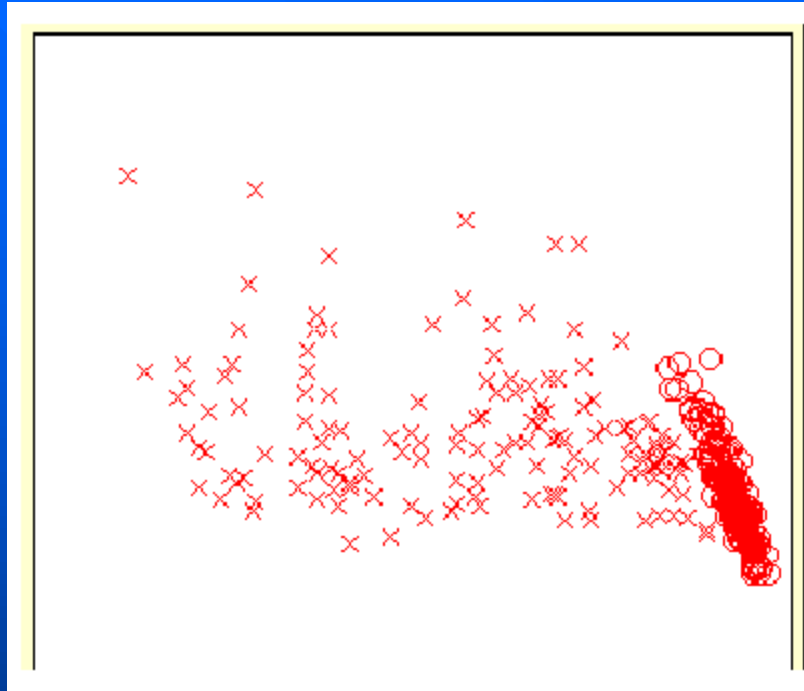


B1

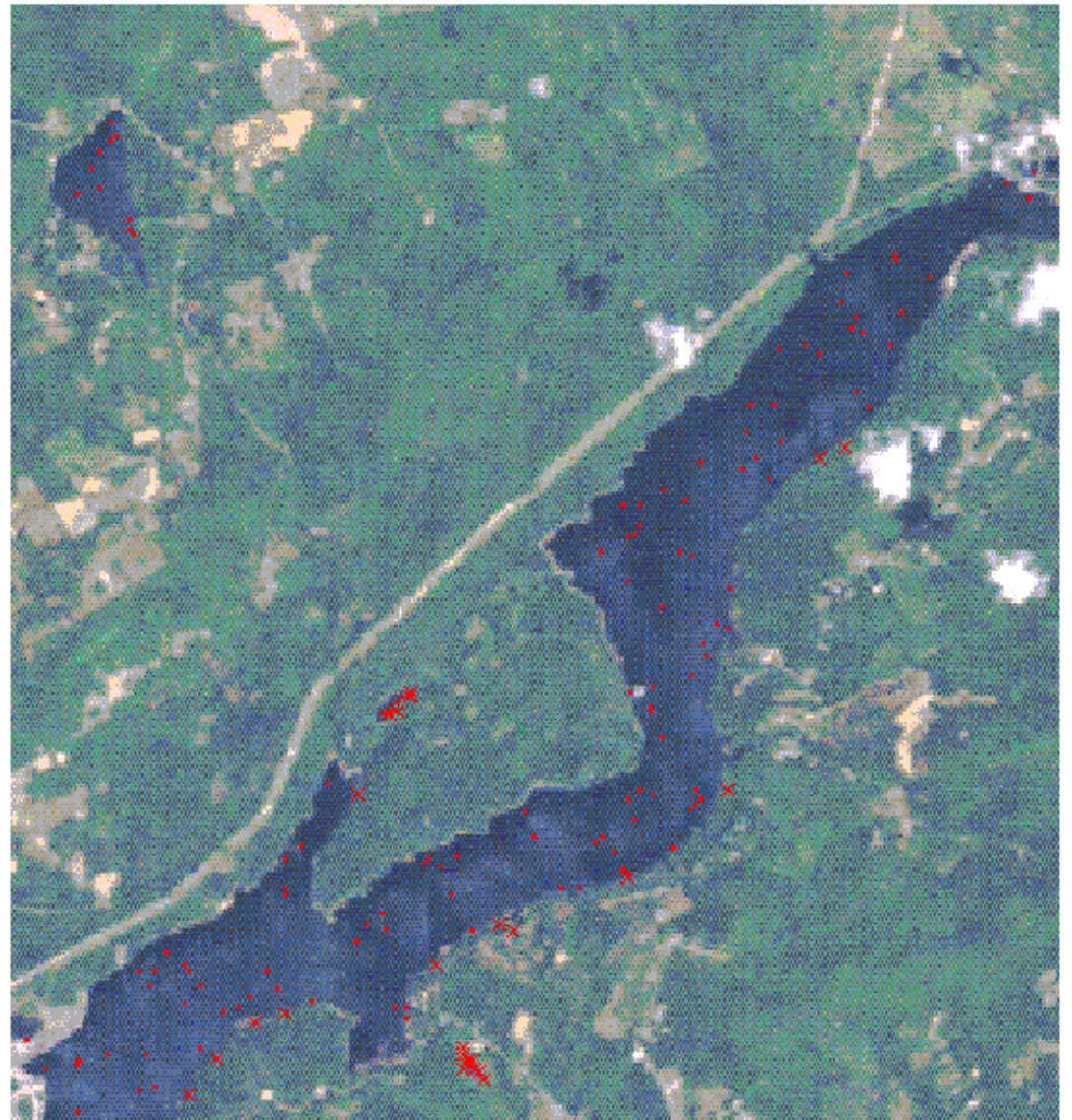
B2



# Water = Water ??



B1 B2



# The Atlanta Data

- NOAA-14 Satellite (National Oceanic and Atmospheric Administration)
- AVHRR Sensor (Advanced Very High Resolution Radiometer):
  - Band 1: Red
  - Band 2: Near Infrared
  - Band 3: Mid Infrared
  - Band 4: Long Infrared
  - Band 5: (Very) Long Infrared
- Data from “NASA’s Project Atlanta”
- 18 Days from Jan 1997 to Dec 1997
- Resolution: 1 km x 1 km per Pixel
- Main Study Area: 70 km x 46 km

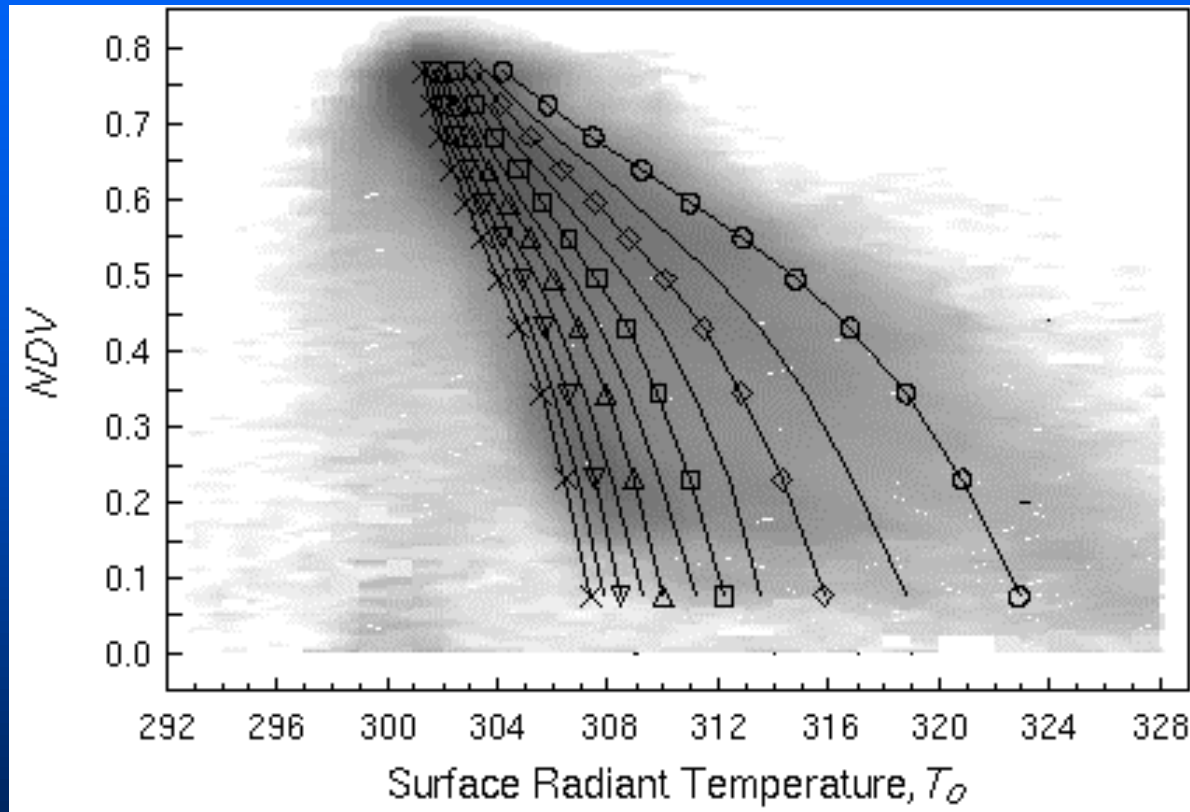
## Some Definitions

- Normalized Difference Vegetation Index:

$$NDVI = \frac{Band2 - Band1}{Band2 + Band1}$$

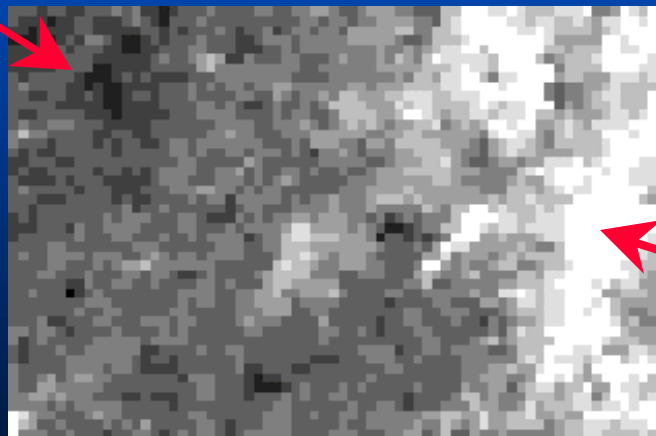
- $NDVI \sim 0.8$  for Highly Vegetated Surfaces
- $NDVI \sim 0.1$  for Bare Soil
- Surface Radiant Temperature  $T_0$ : Band 4
- Surface Moisture Availability  $M_0$

# An Example

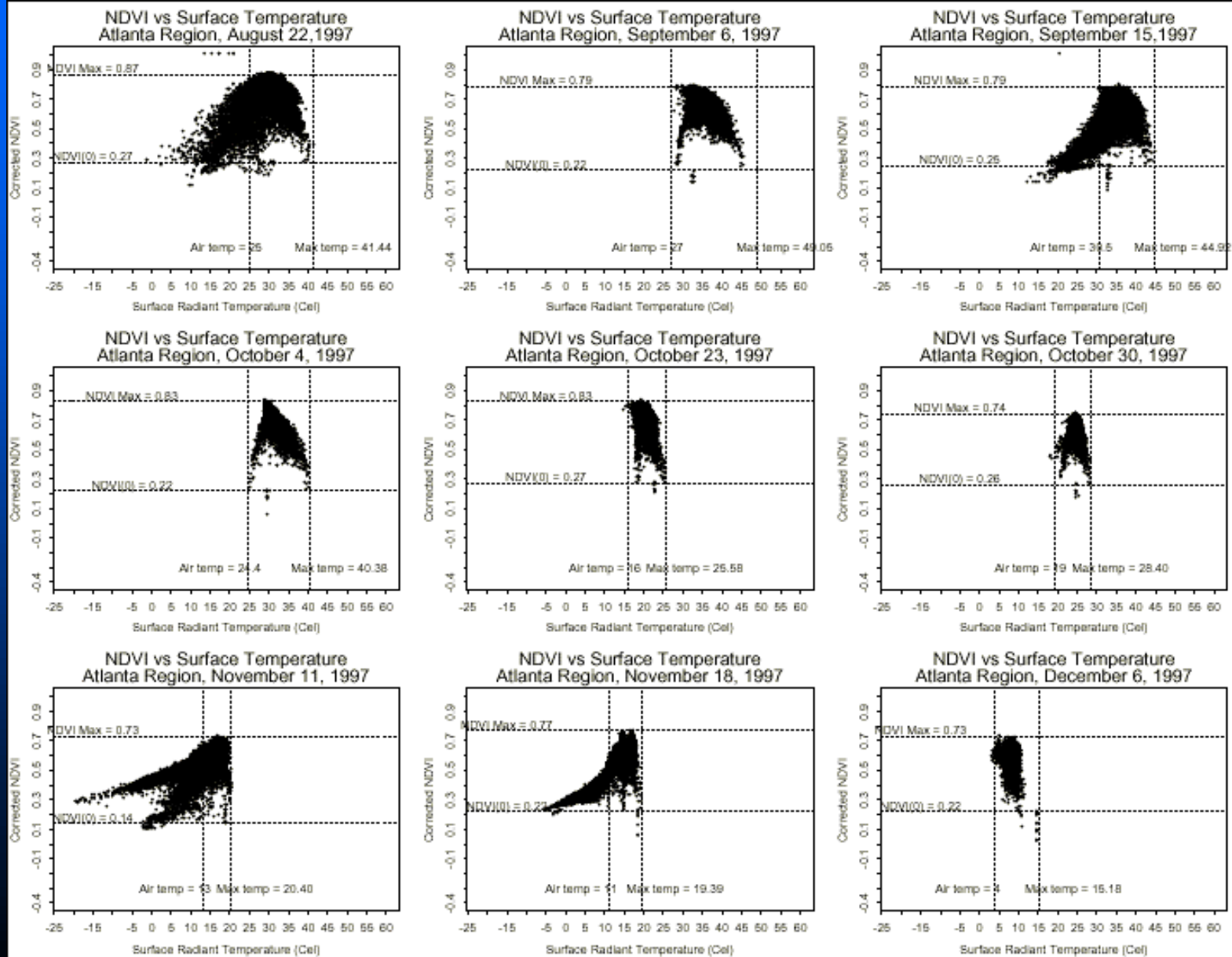


NS001-TMS derived  $T_o$ - $NDVI$  scatterplot (gray spectral scaling) at a 5 meter spatial resolution for a 7 x 3 km area of the Mahantango Watershed, Pennsylvania. 18 July 1990, 1145 LST. Isopleths representing moisture availability index,  $M_o$  are overlaid with the legend,  $o = 0.0$  ('warm' edge),  $\diamond = 0.2$ ,  $\square = 0.4$ ,  $\Delta = 0.6$ ,  $\nabla = 0.8$ , and  $\times = 1.0$  (cold edge).

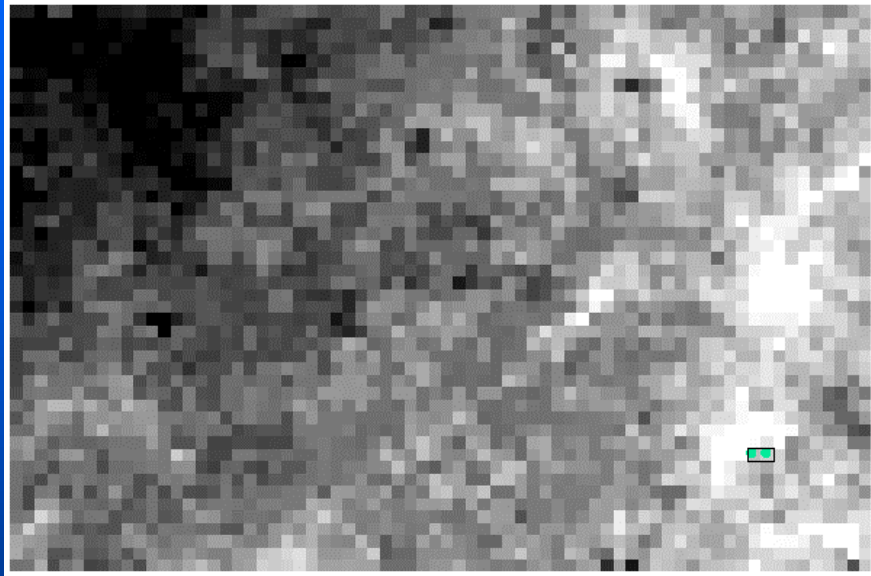
# The Main Study Area



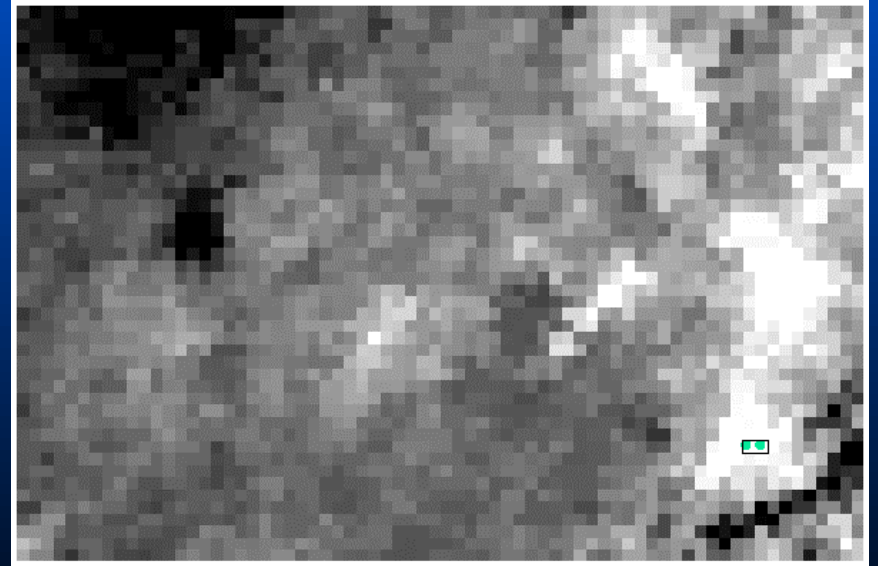
# NDVI vs Surface Temperature



# Two Months



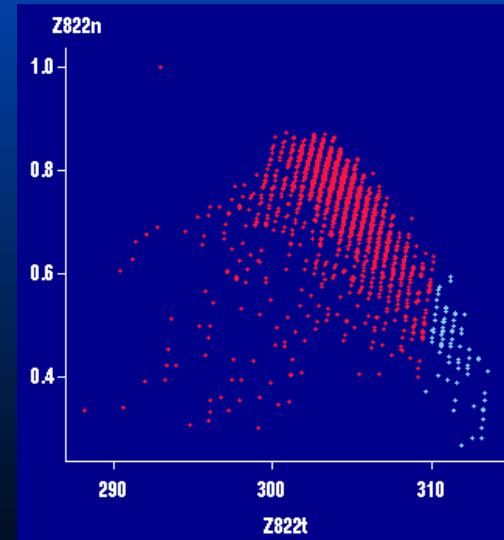
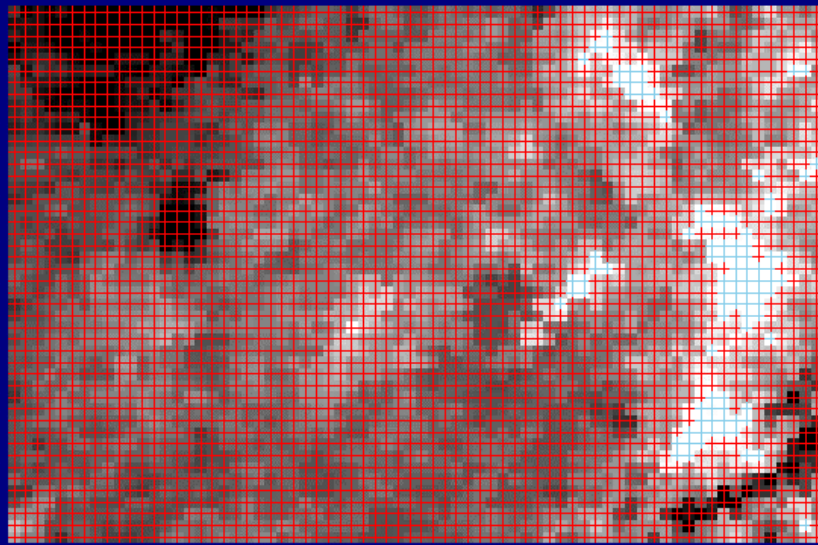
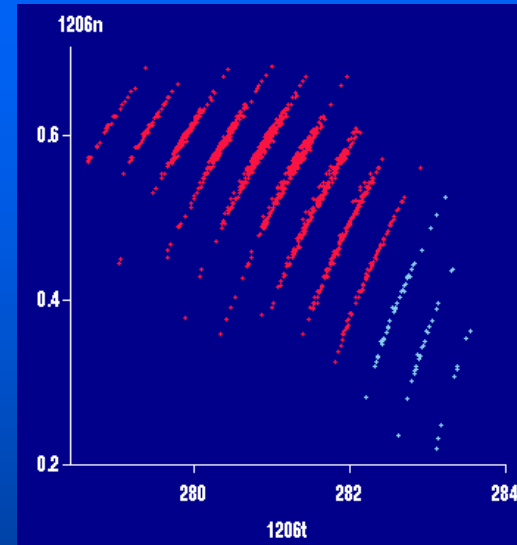
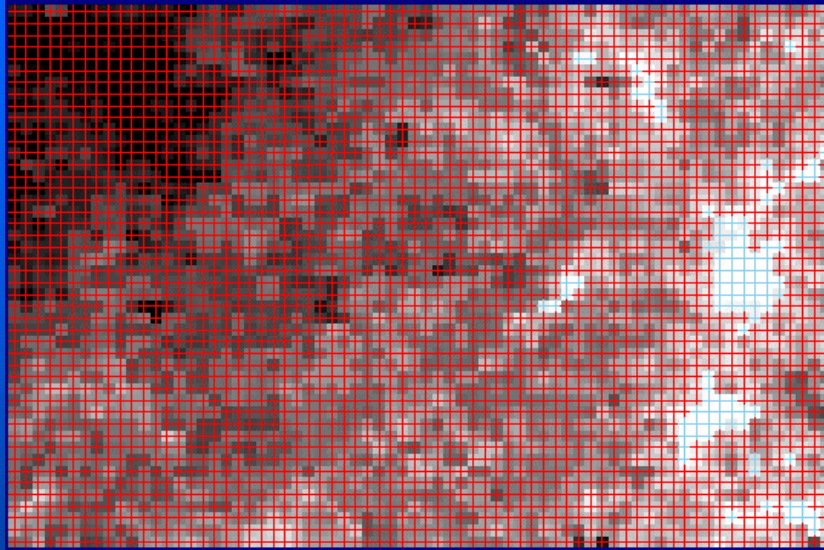
December



August

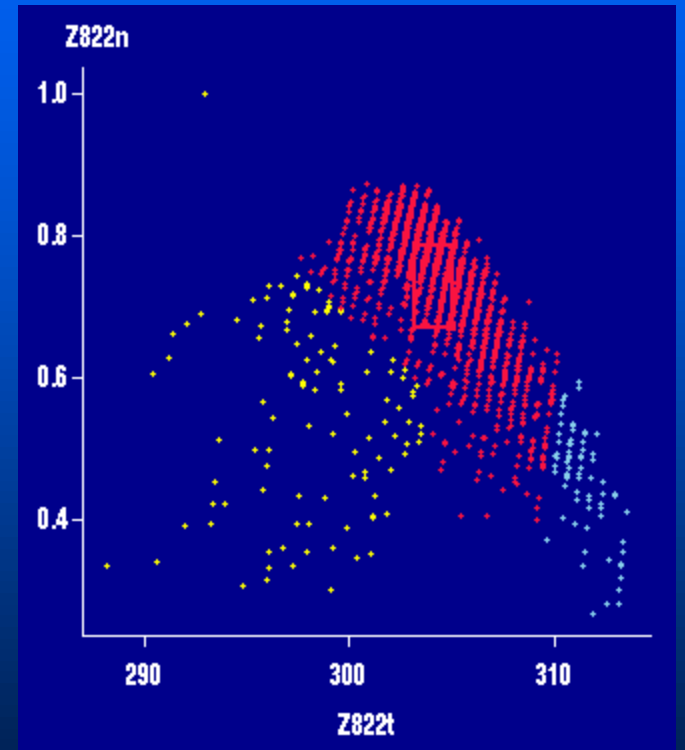
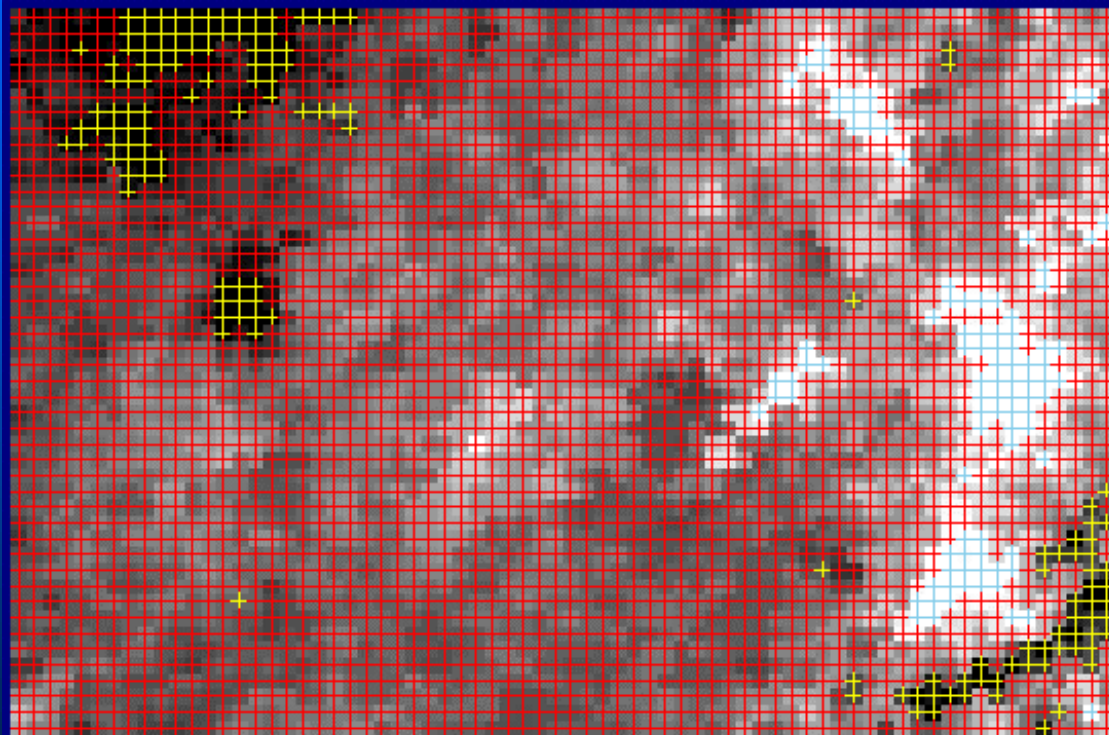
**December**

**The City**



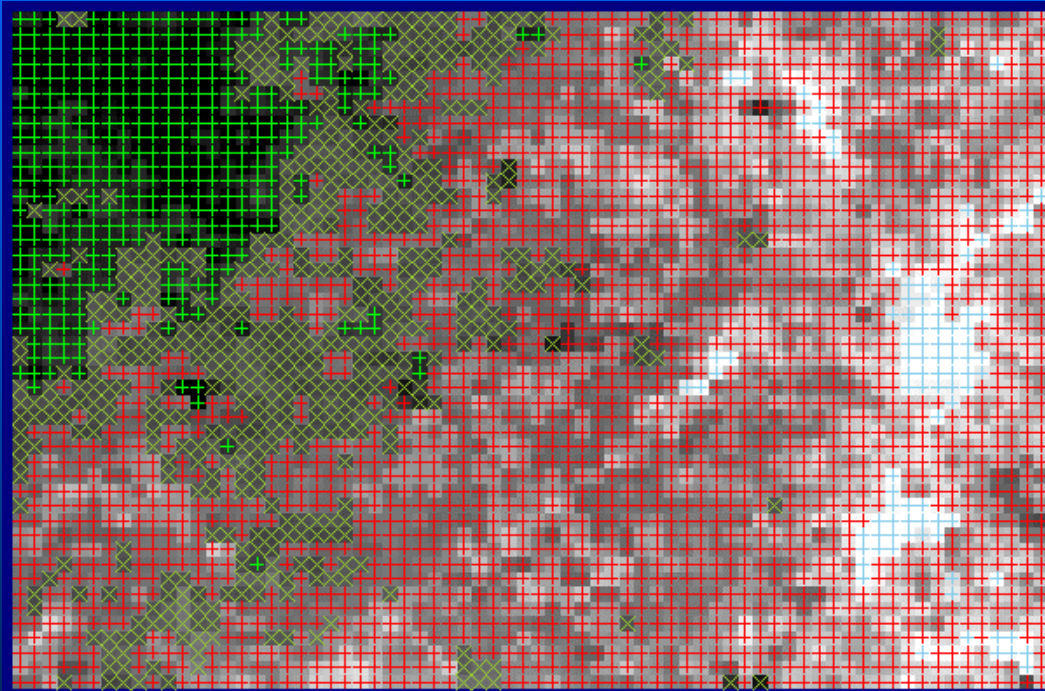
**August**

# Clouds in August

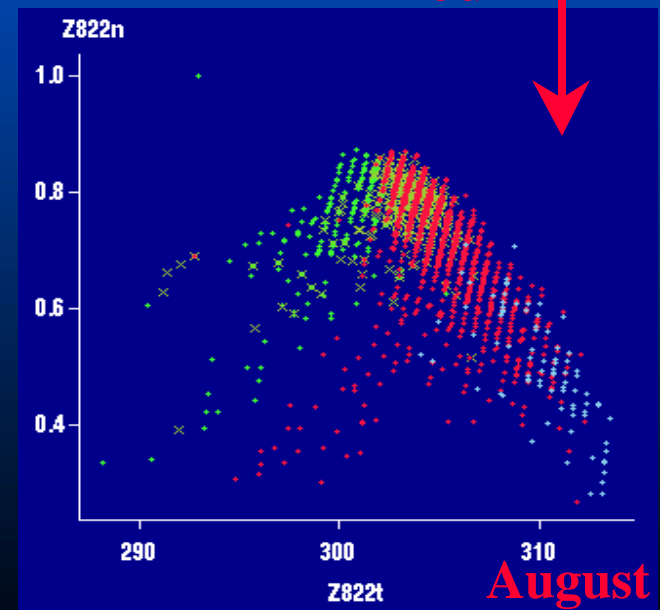
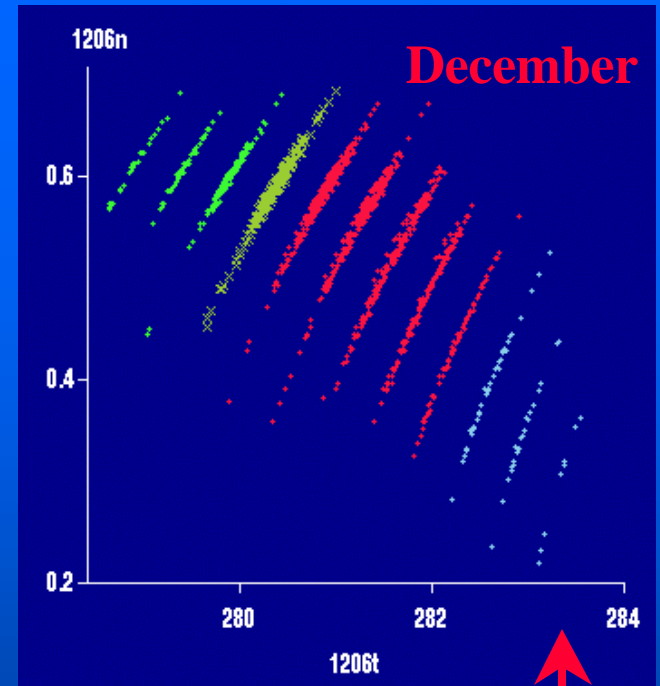


August

# Reclassification



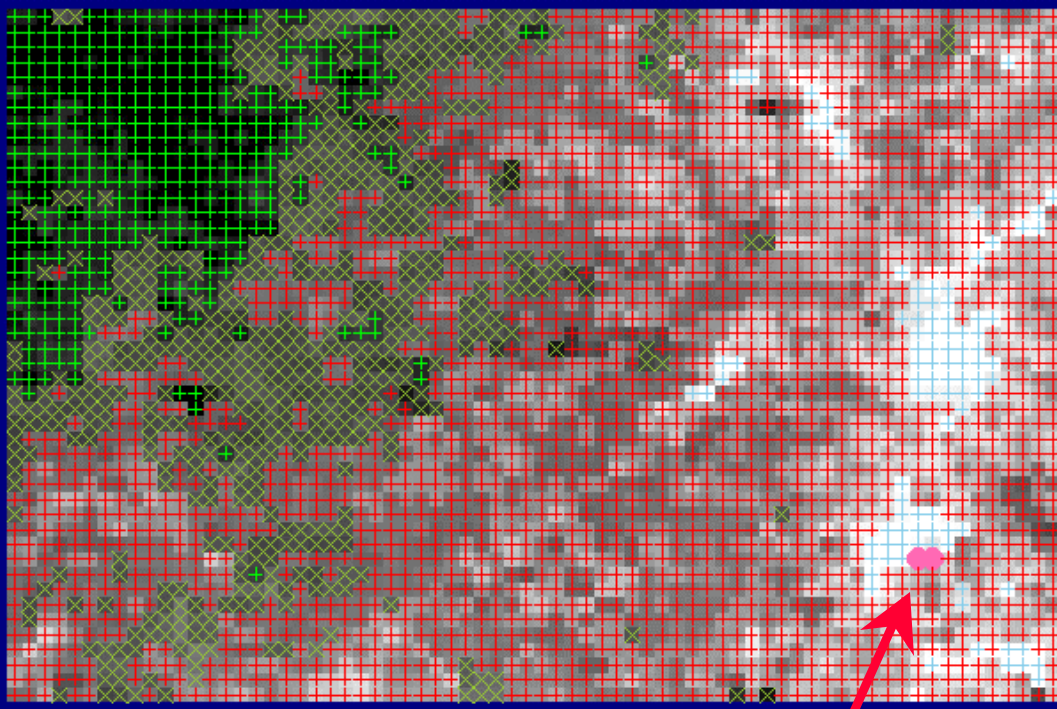
December



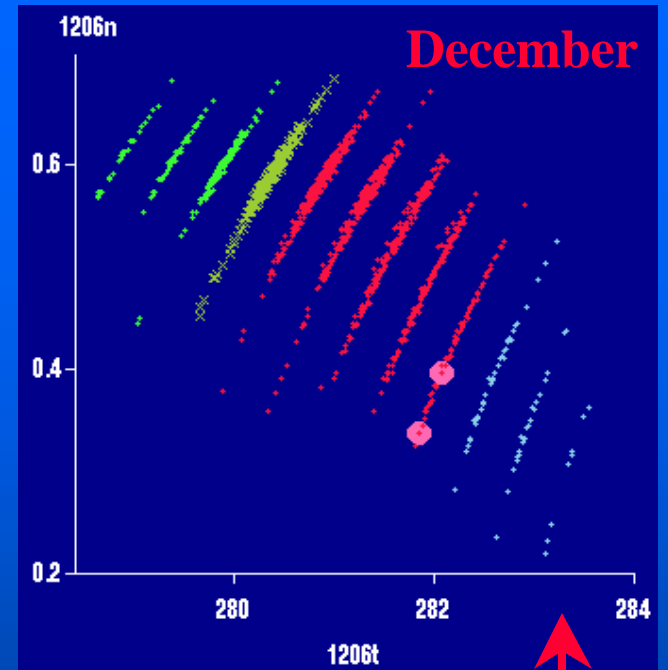
Linked



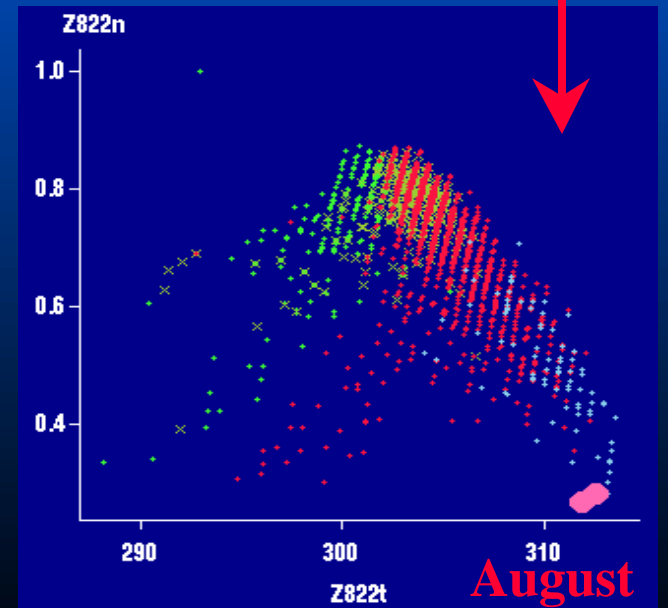
# 2 Pixels of Interest



December



Linked



# **Example: Archaeological Data**

## **Published as:**

Wilhelm, A. F. X., Wegman, E. J., Symanzik, J. (1999):  
Visual Clustering and Classification: The Oronsay Particle  
Size Data Set Revisited, Computational Statistics: Special  
Issue on Interactive Graphical Data Analysis, Vol. 14, No.  
1, 109-146.

## Oronsay Sand Particles

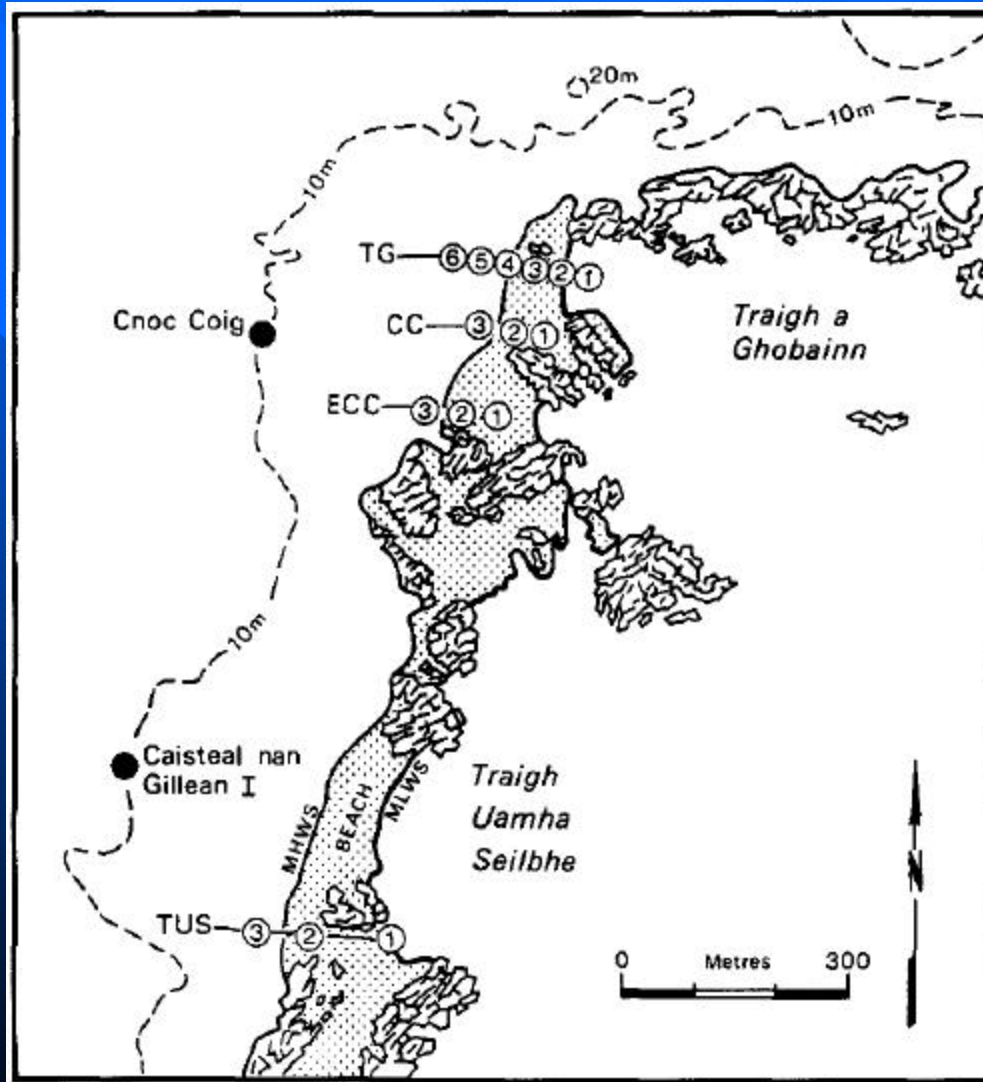
“The mesolithic shell middens on the island of Oronsay are one of the most important archeological sites in Britain. It is of considerable interest to determine their position with respect to the mesolithic coastline. If the sand below the midden were beach sand and the sand from the upper layers dune sand, this would indicate a seaward shift of the beach-dune interface.”

Flenley and Olbricht, 1993

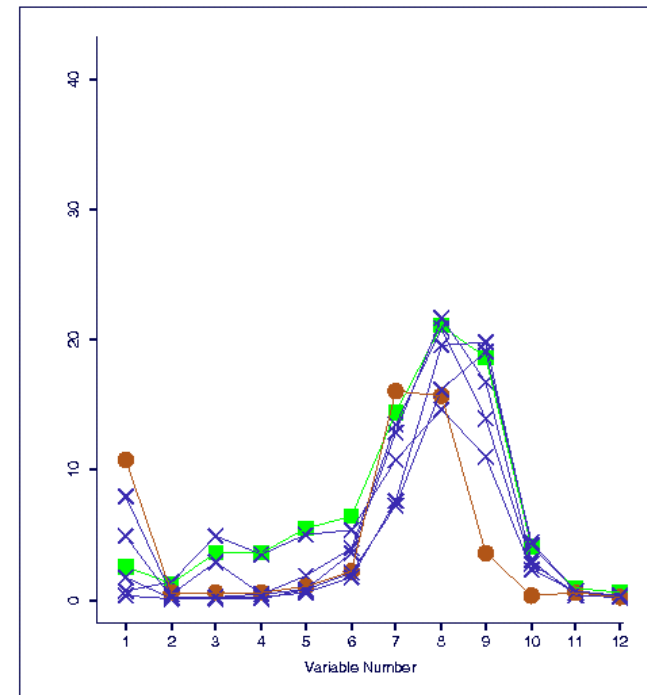
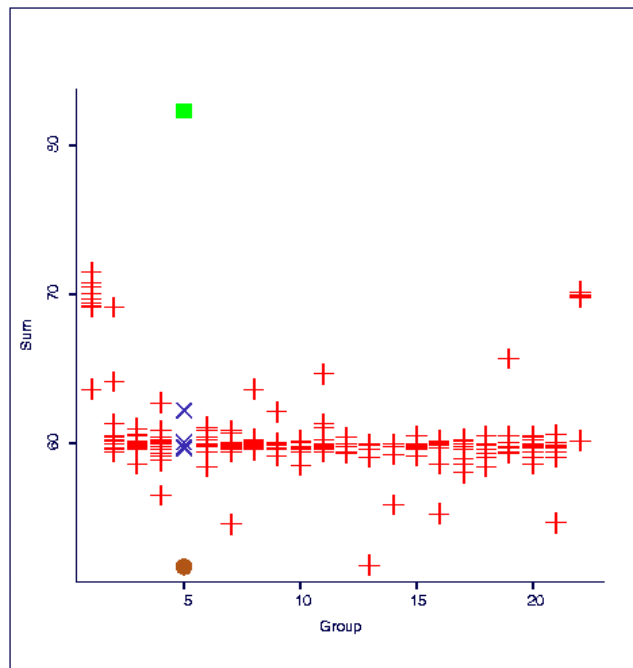
## Objective of Study

- Cluster samples of modern sand into “beach-like” or “dune-like” sand.
- Classify archeological sand samples as to whether they are beach sand or dune sand.

# Oronsay - Geography



# Oronsay - Data Problems



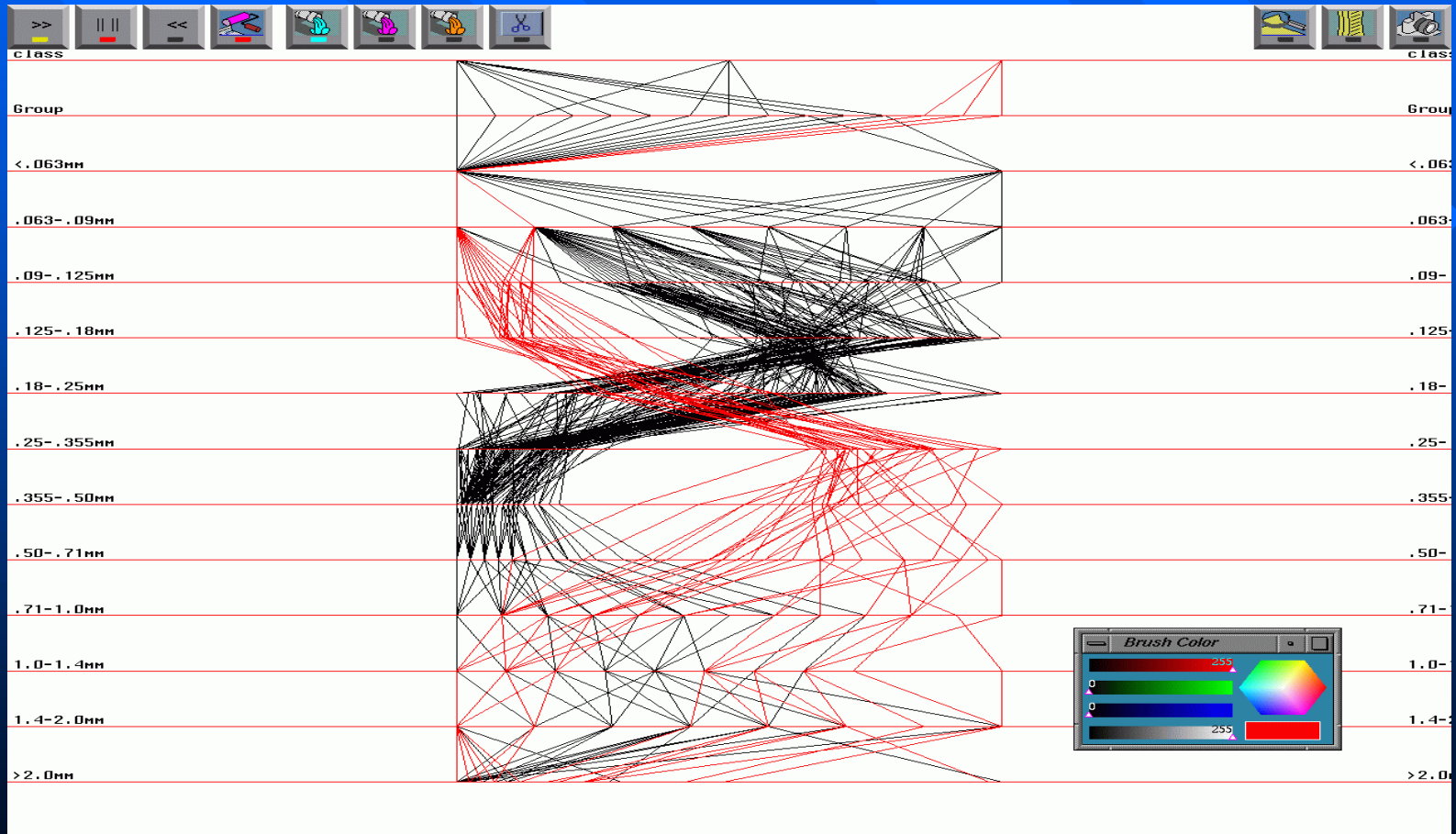
## Oronsay - Parametric Analysis

- Historical strategy is to fit parametric distributions and compare modern and archeological sands based on parameters.
- Weibull, 1933; lognormal (breakage models), log-hyperbolic, log-skew-Laplace, 1937, Barndorff-Nielsen, 1977.
- Models 2 to 4 parameters, theory developed, practice problematic.

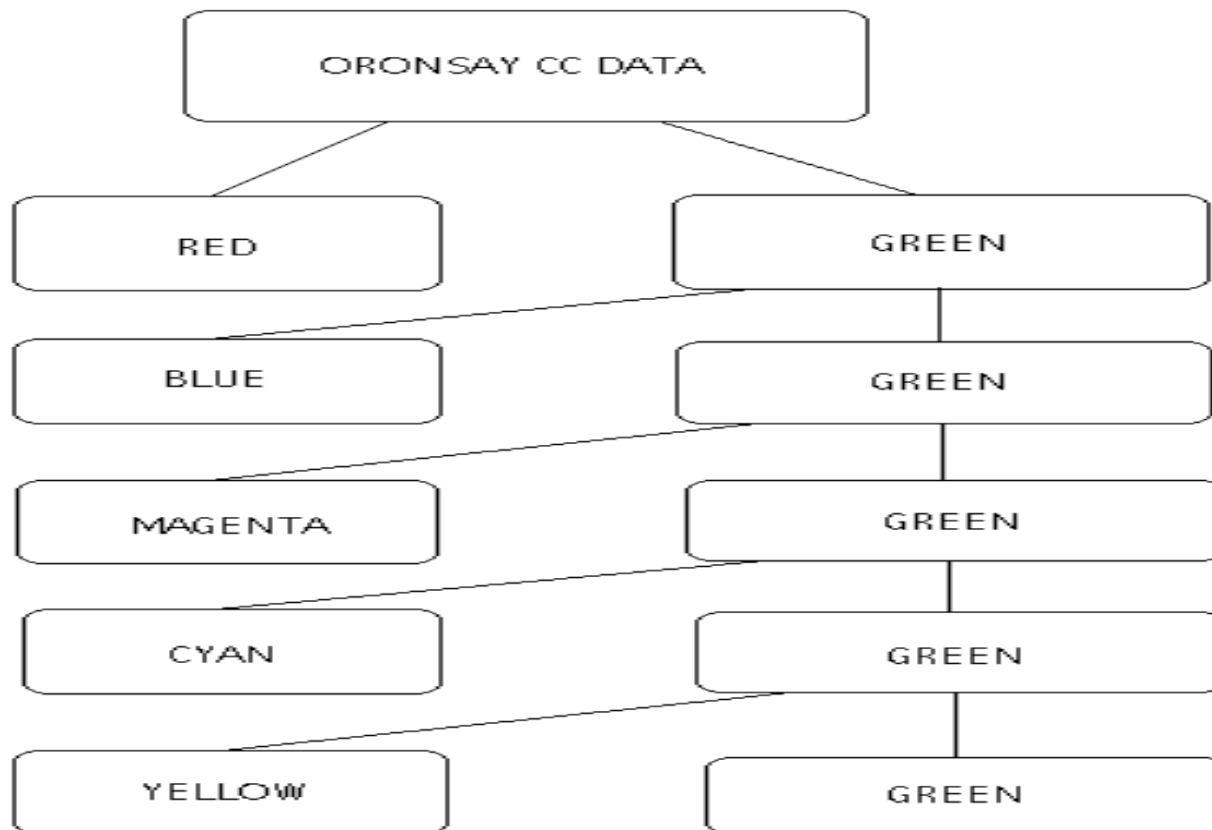
# Oronsay - Visual Approach

- Multidimensional Parallel Coordinate Display Combined with Grand Tour.
- BRUSH-TOUR strategy
  - Clusters recognized by gaps in any horizontal axis.
  - Brush existing clusters with colors.
  - Execute grand tour until new clusters appear, brush again.
  - Continue until clusters are exhausted.

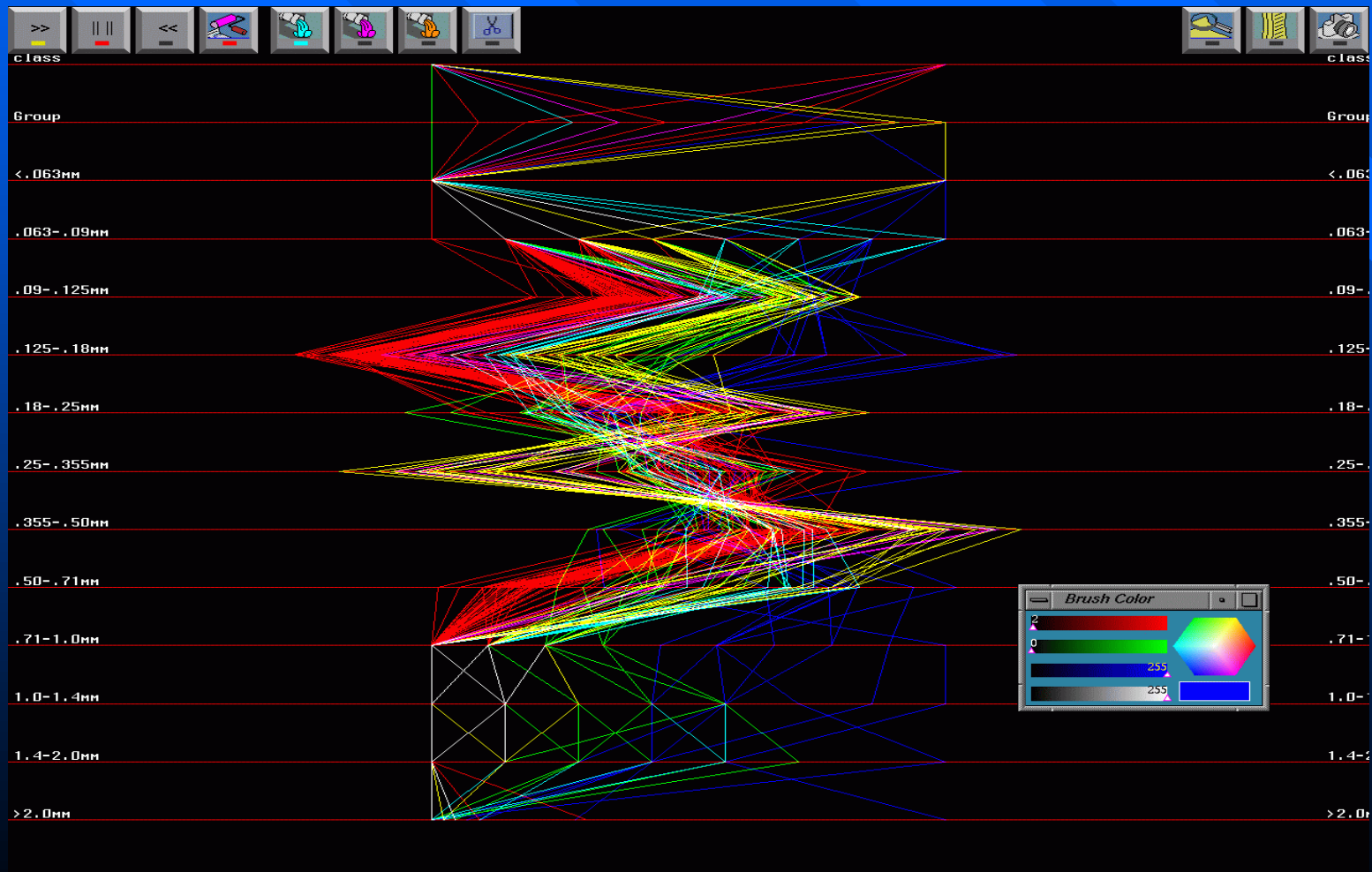
# Beach & Dune Sand



# Separation of Clusters



# Final Clustering



## Oronsay - Conclusions (1)

- Sands from the CC site and the CNG site have considerably different particle size distributions and cannot be effectively aggregated.
- Data at small and at large particle dimensions is too quantized to be used effectively.
- The visual based BRUSH-TOUR strategy is extremely effective at clustering.

## Oronsay Conclusions (2)

- Midden sands are neither modern beach sands nor modern dune sands.
- Midden sands are more similar to modern dune sands.
- This result does not support the seaward-shift-of-the-beach-dune-interface hypothesis, but suggests the middens were always in the dunes

# Live Demo



**GGobi**

## Overall Conclusion

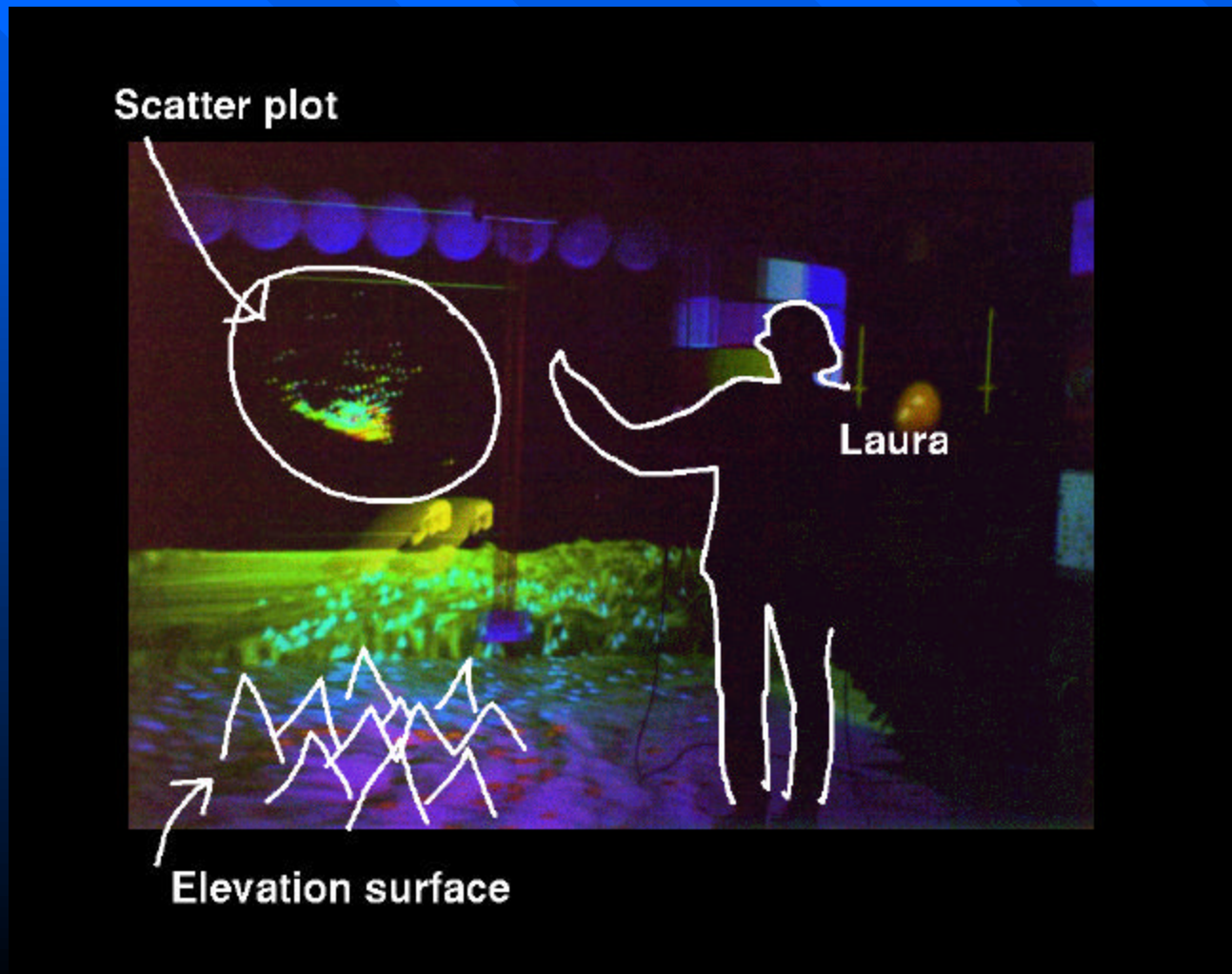
- Visual approach effective to see unexpected structure in data.
- Combination of different techniques most effective.
- Can be used for almost all types of spatial data.

## Future Work (1)

- Enhance new software (e.g., GGobi) to operate in a linked environment with GIS software.
- Allow access to data bases.

## Future Work (2)

- Use 3D environment (CAVE, MiniCAVE) for visualization and ESDA.



# Contact

- Jürgen Symanzik

- [symanzik@sunfs.math.usu.edu](mailto:symanzik@sunfs.math.usu.edu)

- Website

- [www.math.usu.edu/~symanzik](http://www.math.usu.edu/~symanzik)