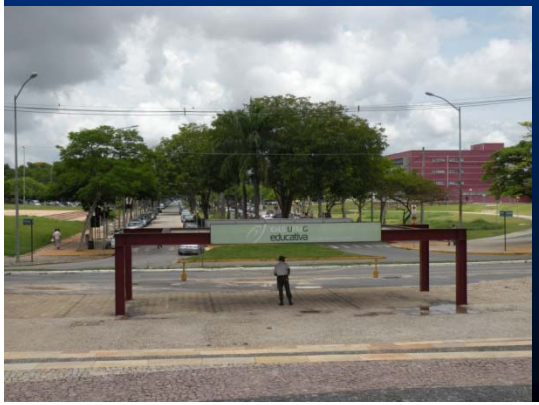
Interactive and Dynamic Statistical Graphics – Special Applications

Jürgen Symanzik

Utah State University, Logan, UT

***e-mail: symanzik@math.usu.edu**

WWW: http://www.math.usu.edu/~symanzik



Examples

- Circular-Spatial Data
- Spatial Archaeological Data

Example 1: Circular-Spatial Data

References:

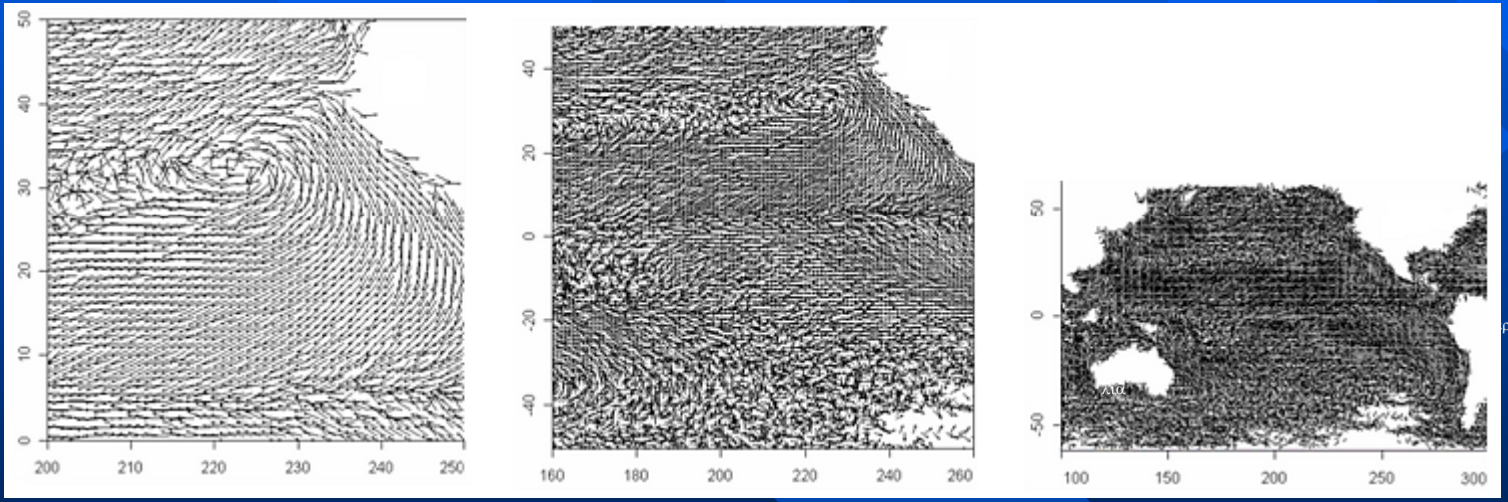
Morphet, W. J. (2009): Simulation, Kriging, and Visualization of Circular-Spatial Data, Dissertation, Utah State University, Logan, Utah.

Morphet, W. J., Symanzik, J. (2009): The Circular Dataimage, a Graph for High–Resolution Circular–Spatial Data, International Journal of Digital Earth, Accepted.

Traditional Plots for Directional Data (1)

What happens as scale and data density increase?

Vortex direction? Less Intelligible Not Intelligible



Traditional Arrow Plot at Increasing Scales

Traditional Plots for Directional Data (2)

Will a heat map work?

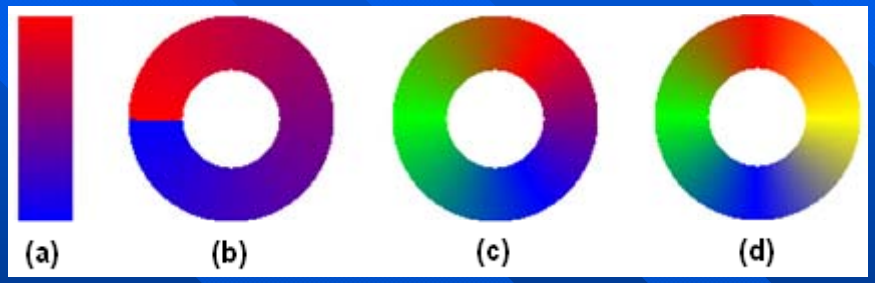
- Images using linear color scale are color discontinuous at the cross over point from 360 deg to 0 deg

Heat Map for Direction of Rocket Nozzle Combustion Flow.

How Can the Color Discontinuity be Fixed?

Transform the linear color scale into a circular scale

(a) Start with a linear color scale
(b) Wrap on a circle
(c) Insert a matching color gradient (RGB going CCW) at the location of color discontinuity
(d) Insert an additional matching gradient to align 4 main colors to 4 main directions



Evolution of the Yellow-Red-Green-Blue (YRGB) Color Wheel.

Color Wheel for Directional Data

A sequence of 2 or more 2-color gradients with the same color where the color gradients are connected

What happens to a heat map when the color wheel is used?

We Call it a "Circular Dataimage"

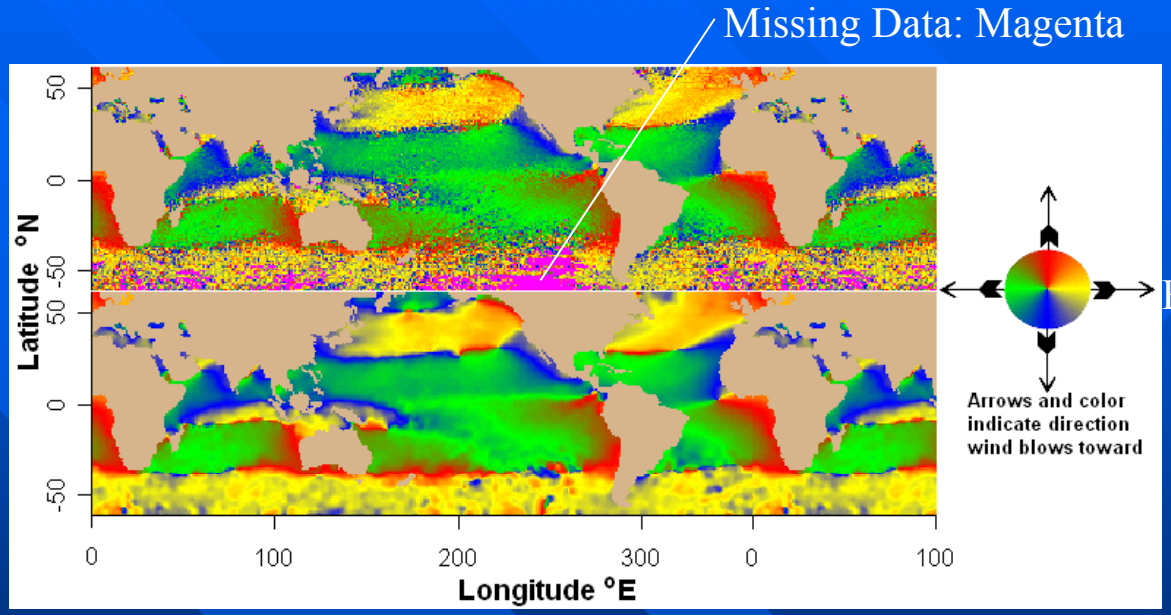
Definition

A heat map of circular-spatial data with direction coded as the color on a color wheel

Example

Noisy

Smoothed



Circular Dataimage of the Ocean Wind Data.

Advantages

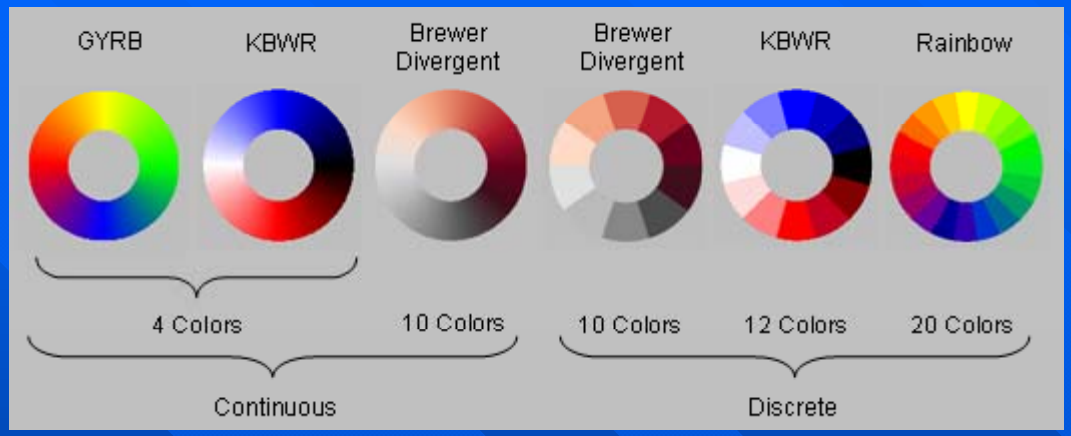
- High resolution continuous image – Each pixel gives direction, an arrow takes many pixels
- Color continuous at any direction

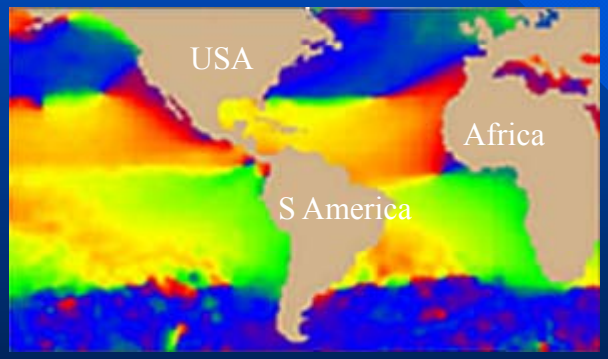
Usability Study

- Directional structure intelligible on wide range of scales and with noise
- Missing data and outliers easily spotted

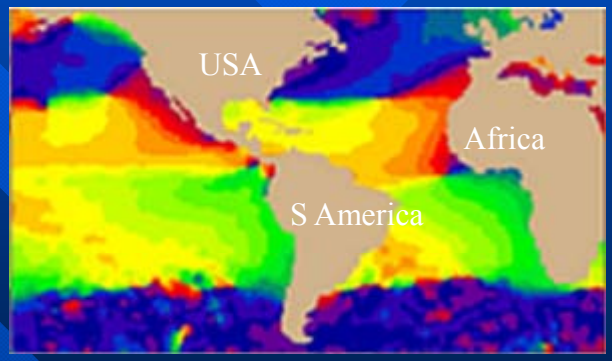
Variations (1)

Many color wheels possible to meet different needs





Continuous color wheels produce maximum detail

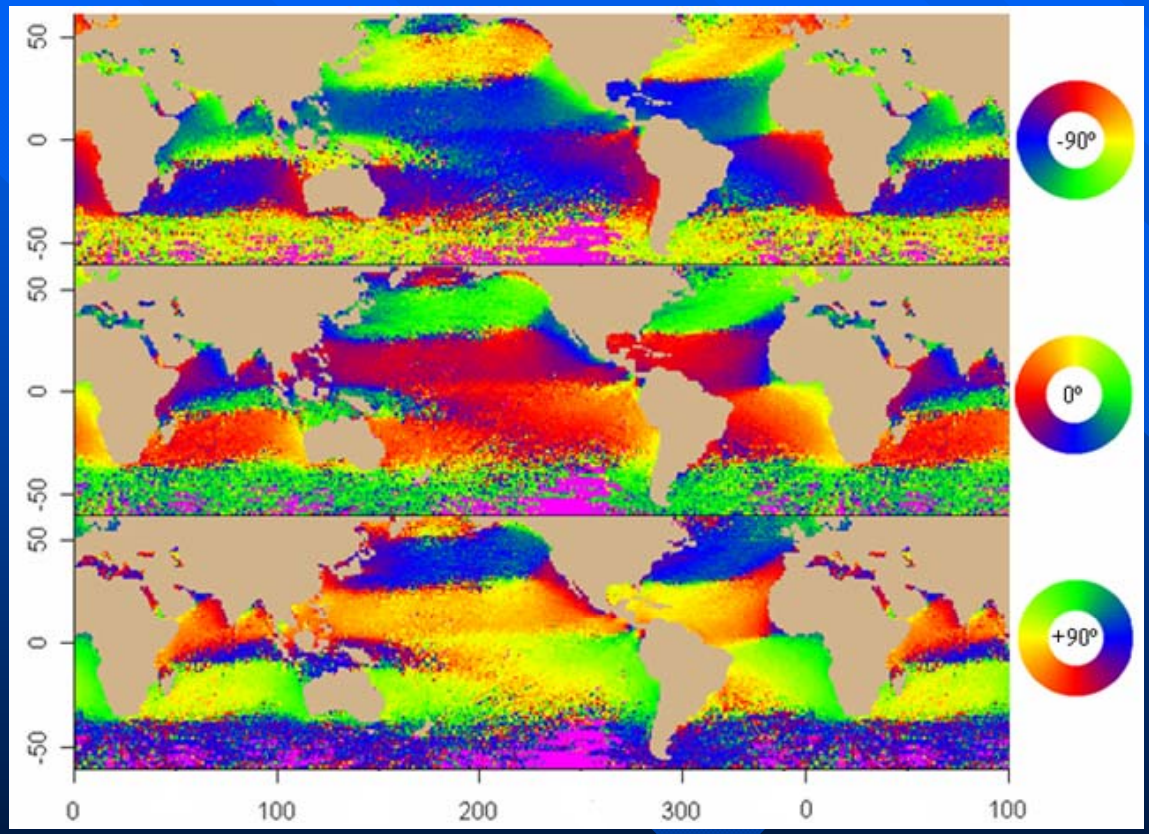


Binned color wheels distinguish similar areas & allow visual quantification

Effects of Changing Main Colors and Binning.

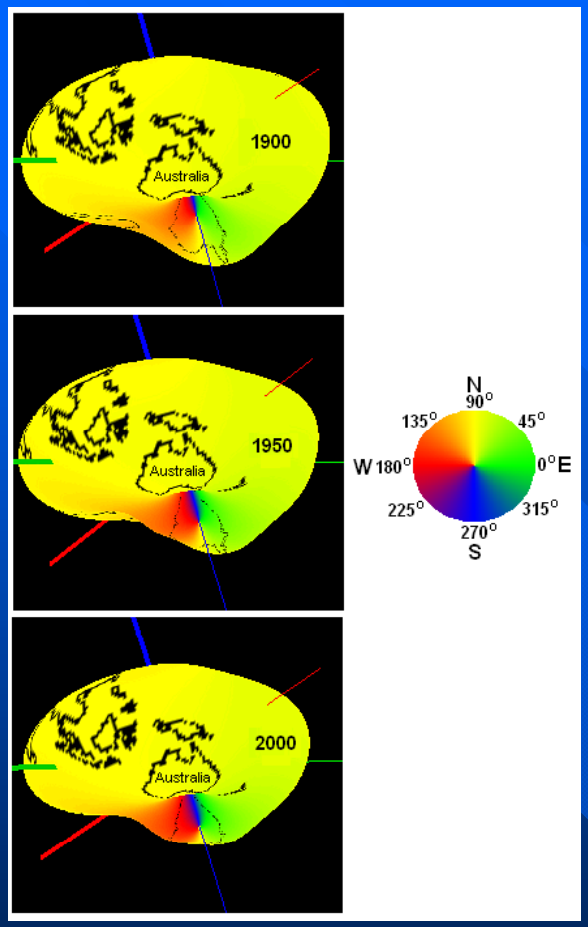
Variations (2)

Rotation of the color wheel may improve the perception of apparent structure in an area of interest

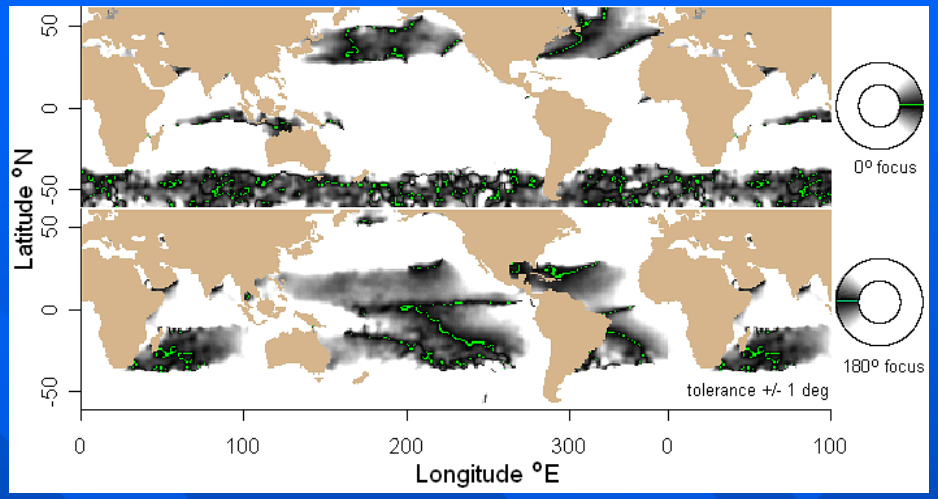


Effects of Color Wheel Rotation.

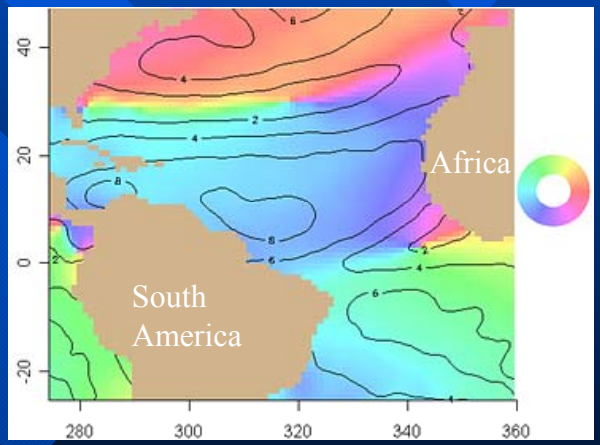
Variations (3)



Circular Dataimage Plotted on 3D Polar Plot - Earth main magnetic H field with direction as color and magnitude as radius.



Focus Plot Highlights Specific Direction



Overplot Contour Curves of Magnitude

Live Demo (1)

-  R Demo

- R Library CircSpatial available from CRAN:
  http://cran.r-project.org/web/packages/CircSpatial/index.html

Example 2: Archaeological Data

Reference:

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

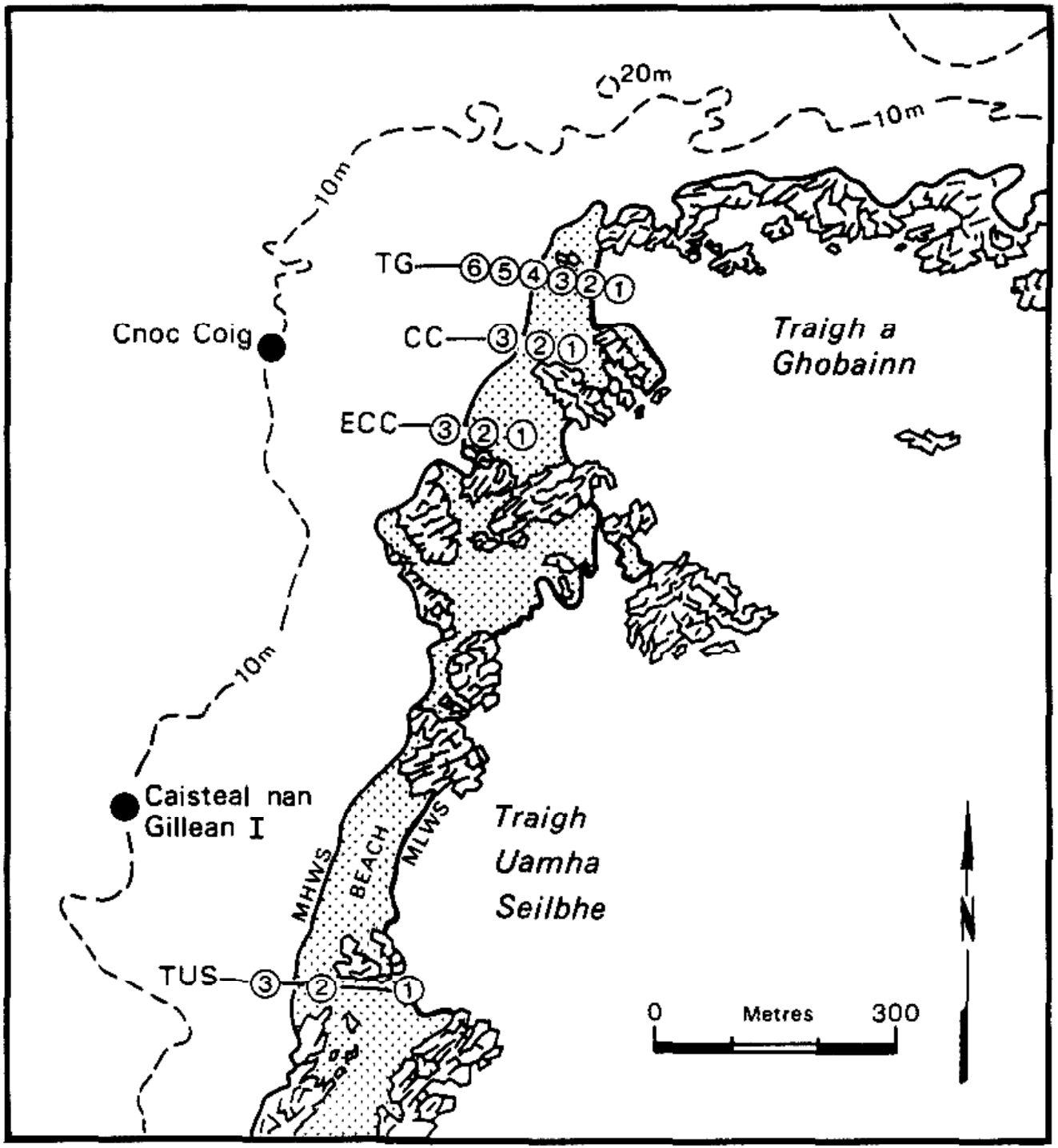
Objective of Study

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

Oronsay Data

- 226 Sand samples from different sites
- Data obtained from sieving 60g or 70g of sand through a stack of 11 sieves of sizes 0.063, 0.09, 0.125, 0.18, 0.255, 0.355, 0.5, 0.71, 1.0, 1.4 and 2.0 mm
- => 12 classes of particle sizes
- Extreme classes [0, 0.063) and [2.0, ∞) might be unreliable

Oronsay - Geography



Oronsay - Legend

Group	# Obs	Classification	Site Code
1	10	Sands above CC Midden	CCJ6N (layers 1–10)
22	7	CC Shell Midden	CCJ6N (layers 11–17)
2	14	Sands below CC Midden	CCH17
3	18	Sands below CC Midden	CCJ6
4	13	CC Soil Pit	CCSP1
5	7	CNG Shell Midden	CNGIE (layers 1–5)
21	8	Sands below CNG Midden	CNGIE (layers 6–15)
6	15	CC Mid Beach	CC1
7	15	CC Upper Beach	CC2
8	14	CC Upper Dune	CC3
9	10	CC Lower Beach	TG1
10	10	CC Mid Beach	TG2
11	10	CC Upper Beach	TG3
12	5	CC Base of Dune	TG4A
13	5	CC Face of Dune	TG4B
14	5	CC Top of Dune	TG4C
15	10	CC Mid Beach	EC–C1
16	10	CC Upper Beach	EC–C2
17	10	CC Upper Beach	EC–C3
18	10	CNG Lower Beach	TUS1
19	10	CNG Upper Beach	TUS2
20	10	CNG Dune	TUS3

Table 1: Group number, number of observations per group, classification, and site code. CC stands for *Cnoc Coig* and CNG stands for *Caisteal nan Gillean*. Site codes for the "modern" sand samples directly relate to locations shown in Figure 1.

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.

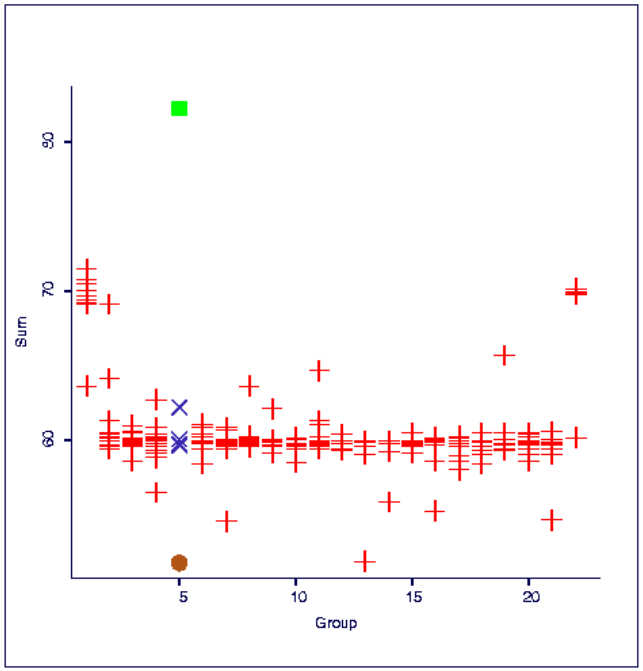
Oronsay - Visual Approach

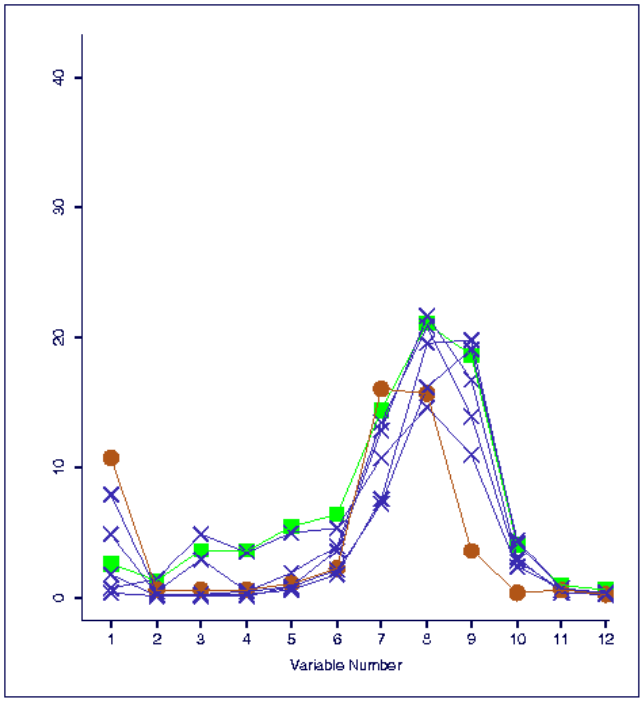
- Use Grand Tour (in GGobi)
- BRUSH-TOUR strategy
  - Clusters recognized by gaps in any projection.
  - Brush existing clusters with colors.
  - Execute grand tour until new clusters appear, brush again.
  - Continue until clusters are exhausted.
- Alternatively: Brush in Parallel Coordinate Plots

Live Demo (2)

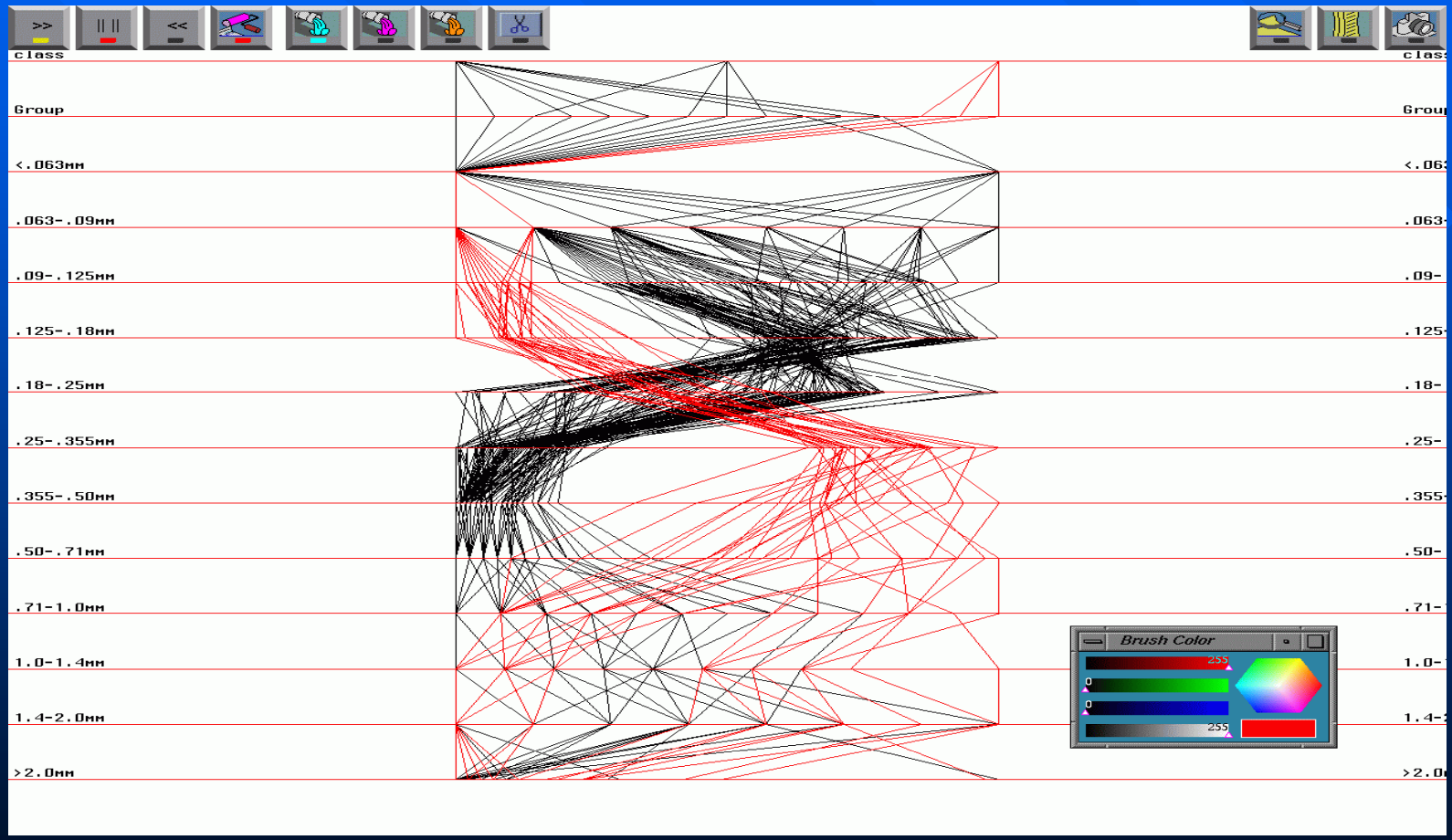
-  GGobi

Oronsay - Data Problems

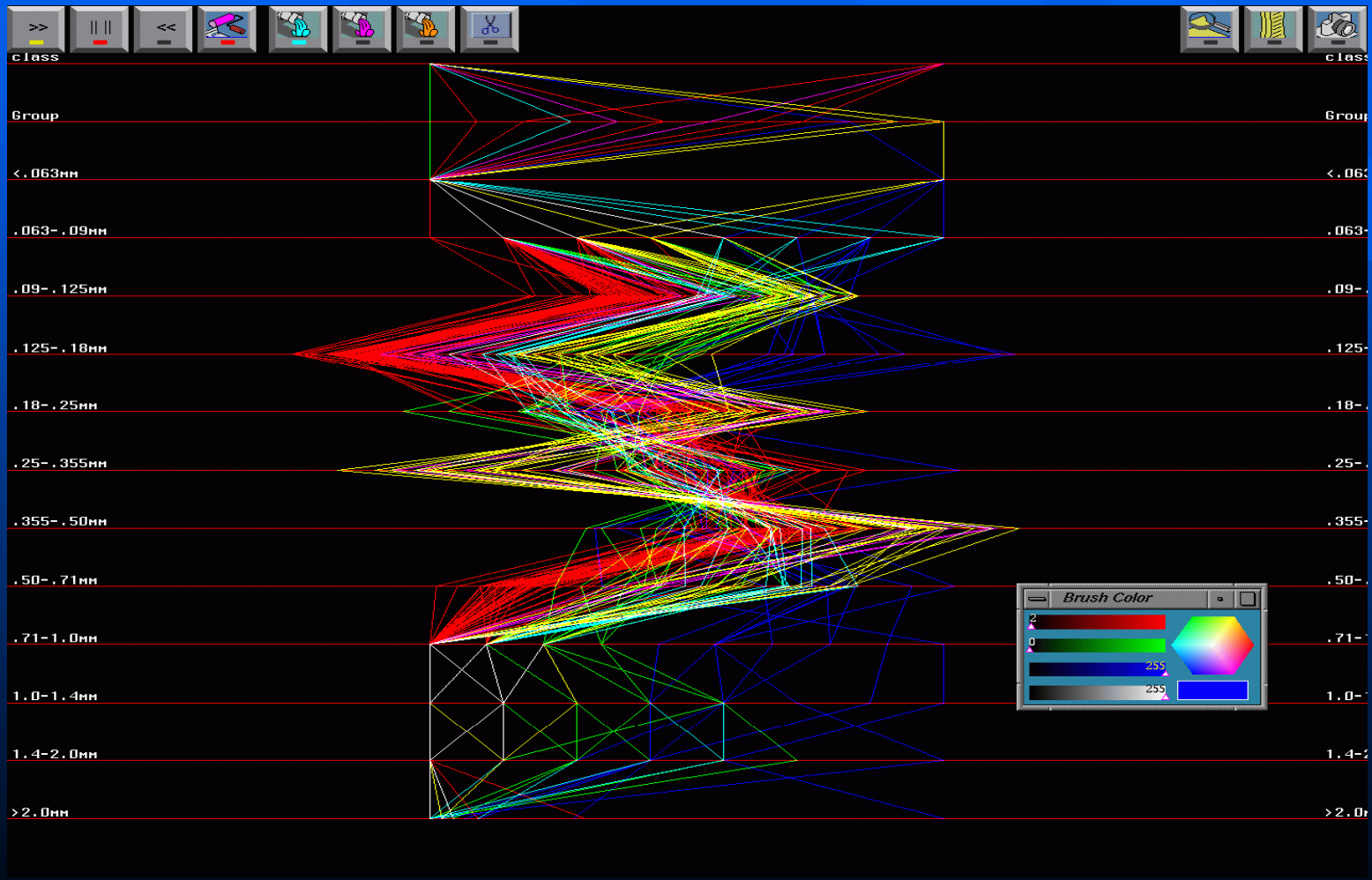




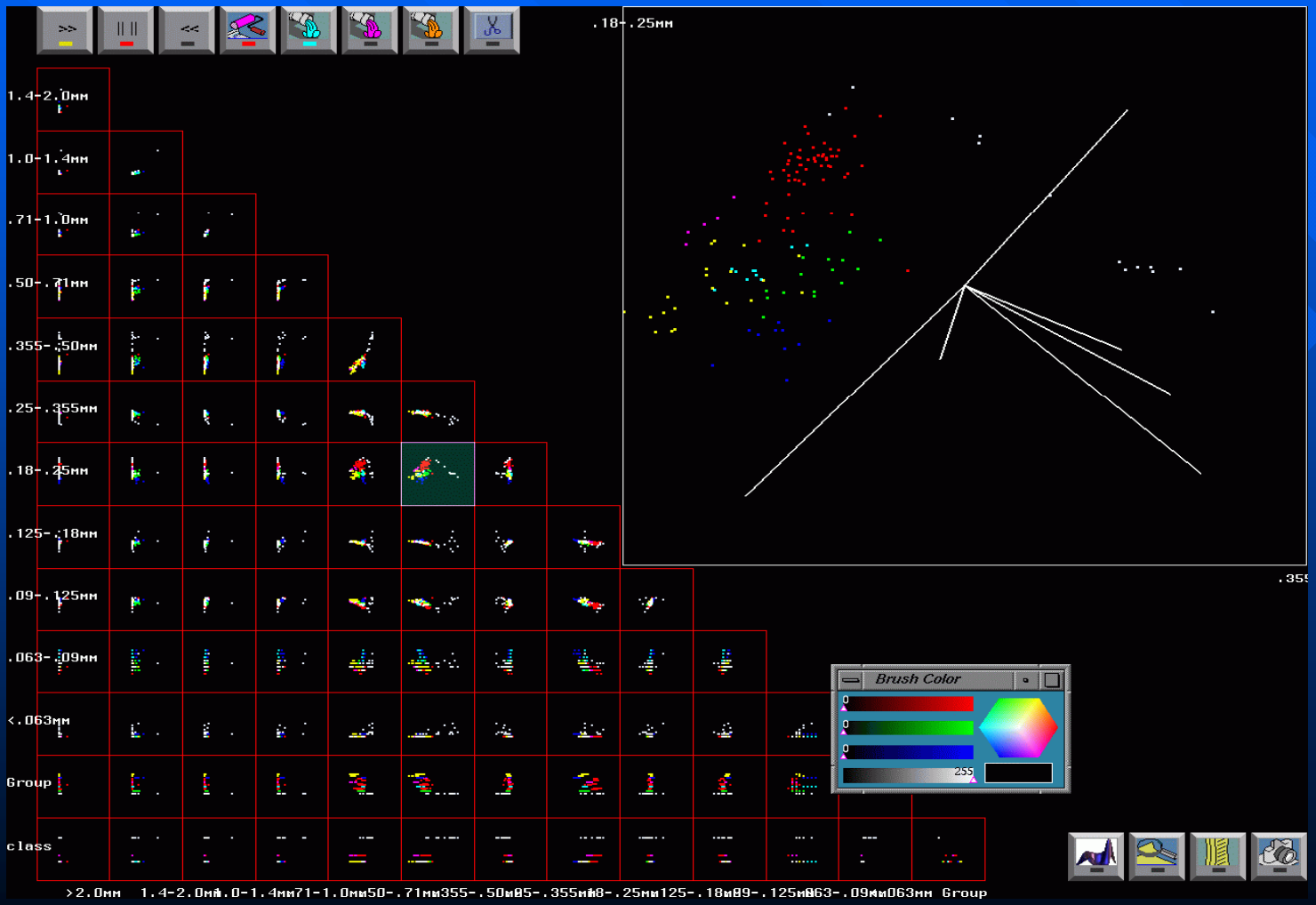
Beach & Dune Sand (in ExplorN)



Final Clustering (in ExplorN)



Scatterplots & Projection (in ExplorN)



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

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 data (another major application: Remote Sensing).

Questions ???