Recent Advances in Geographic Information Science

Michael F. Goodchild
University of California
Santa Barbara

Outline

- Geographic information science
  - motivation, history, content
- Uncertainty
  - recent ideas
- An integrated portal
  - technical feasibility
  - recent progress

1985

- GIS well established
  - a nascent software industry
  - texts
    - Burrough, *Principles of GIS*
    - MacDougall, *Computer Programming for Spatial Problems*
  - a scattering of courses
    - UWO circa 1976
  - various things could be achieved by computer processing of spatial data
    - measurement and analysis
    - production and editing
    - map-making

...but some big questions

- What to teach?
  - training in software?
  - education in principles?
    - what were those principles?
- What to research?
  - algorithms and data structures to do it "faster, better, cheaper"

The analogy to statistics

- A branch of mathematics dating from well before the advent of computers or calculators
  - theory, numerical analysis predated computation
- Where is the equivalent theoretical framework for GIS?
  - computation predated the development of theory
- GIS is to *x* as the statistical packages are to statistics
  - what is *x*?
- "A spatial analytic perspective on GIS", *IJGIS* 1: 327-334, 1988

The NCGIA research agenda

- Discussions initiated by Ron Abler, 1986-1987
- The 1987 solicitation

1. Spatial analysis and spatial statistics
2. Languages of spatial relations
3. Visualization
4. Artificial intelligence and expert systems
5. Social and institutional issues

SDH, Zurich, August 1990

- Goodchild keynote
  - why "spatial data handling"?
  - are we the UPS of GIS?
- "Spatial information science"
  - NCGIA as a multidisciplinary enterprise
  - what disciplines can contribute to a basic science of geographic information?
    - spatial statistics
    - spatial databases
    - computational geometry
    - spatial cognition
  - "Geographic information science", *IJGIS* 6(1): 31-45 (1992)

Consensus-building: UCGIS

- An organization to represent the growing GIScience community
  - building the community
- Opening UCGIS Assembly, Columbus, June 1996
- What is the research agenda of GIScience?
  - white papers, discussion, vote by institutional members

The UCGIS research agenda (1996, revised 1998)

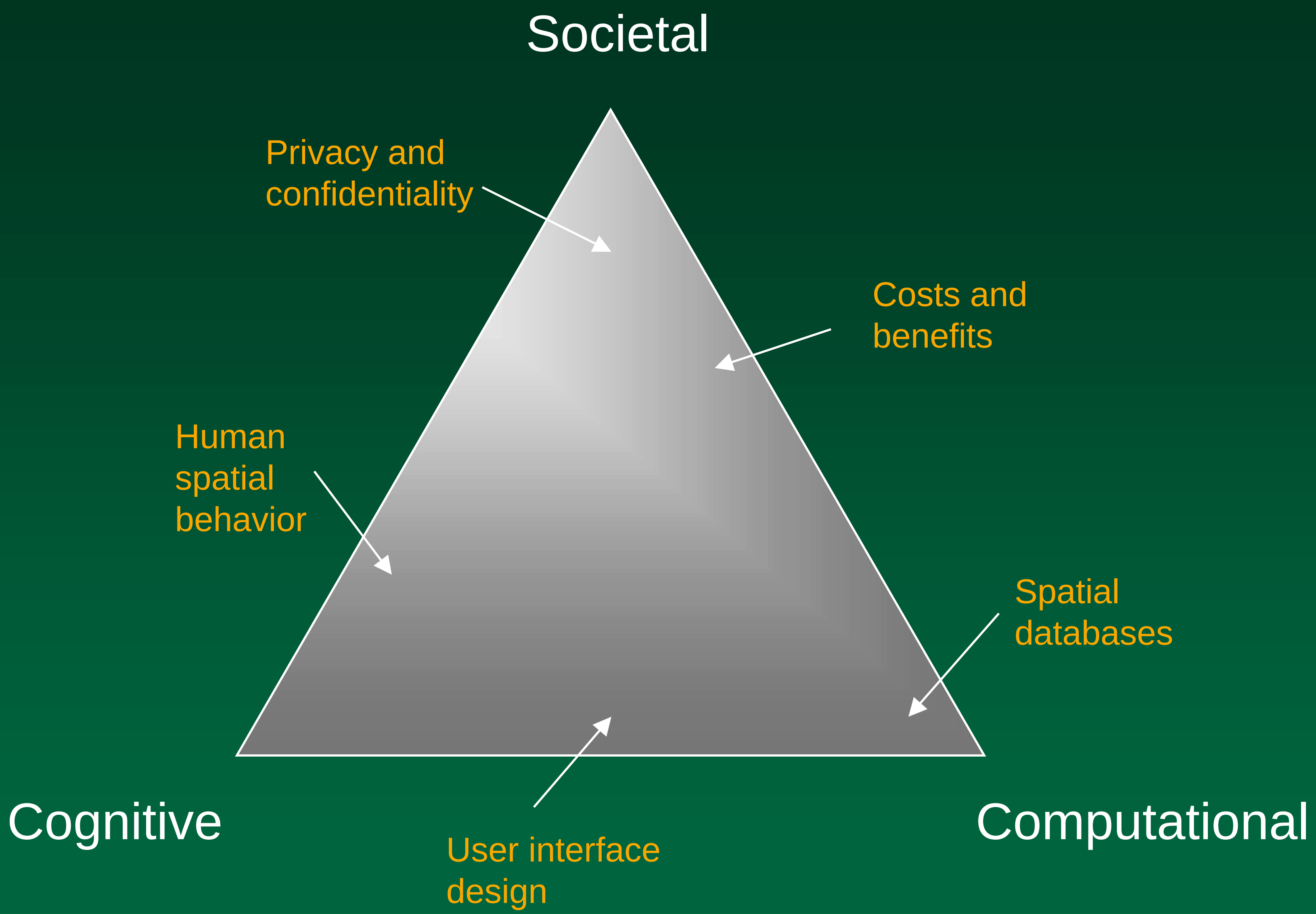
- Cognition
- Extensions to representation
- Acquisition and integration
- Distributed and mobile computing
- Interoperability
- Scale
- Uncertainty
- Spatial analysis
- Future of the spatial information infrastructure
- GIS and society

Other agendas

- Socially focused
  - Rhind 1988
- Computationally focused
  - NSF Digital Government Initiative
    - National Center for Supercomputer Applications, OGC
  - National Research Council Computer Science and Telecommunications Board
    - "IT Roadmap for a Geospatial Future", 2003

Varenius: a top-down perspective

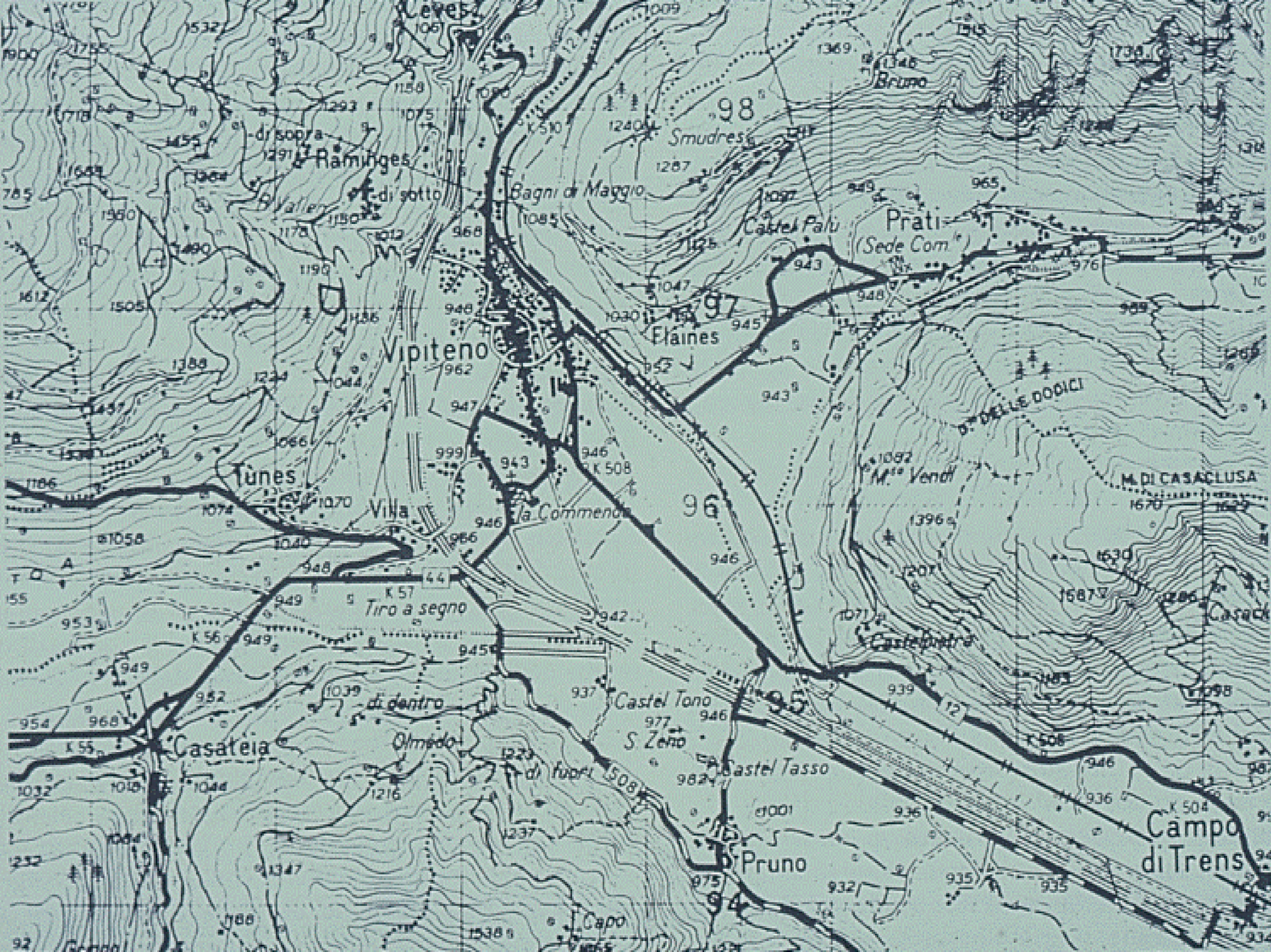
- NCGIA funded by NSF as an 8-year project, 1988-1996
- Varenius: NCGIA's Project to Advance GIScience
  - 1996-1999
- A three-vertex research agenda
  - the cognitive vertex
    - human-centric
  - the computational vertex
    - computer science
  - the societal vertex
    - social science


Societal
Privacy and confidentiality
Costs and benefits
Human spatial behavior
Spatial databases
Cognitive
Computational
User interface design

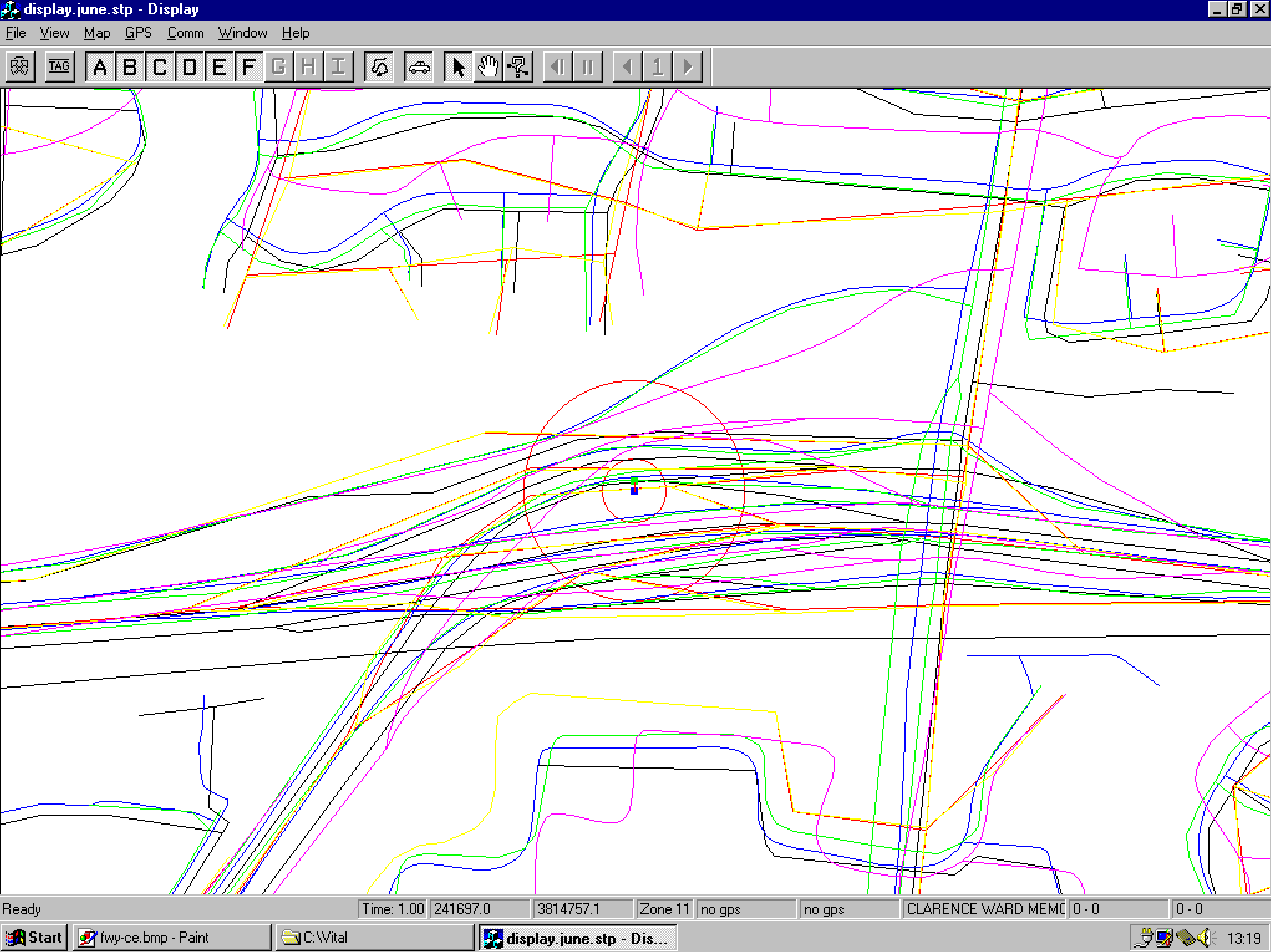
Digital representations of geographic reality

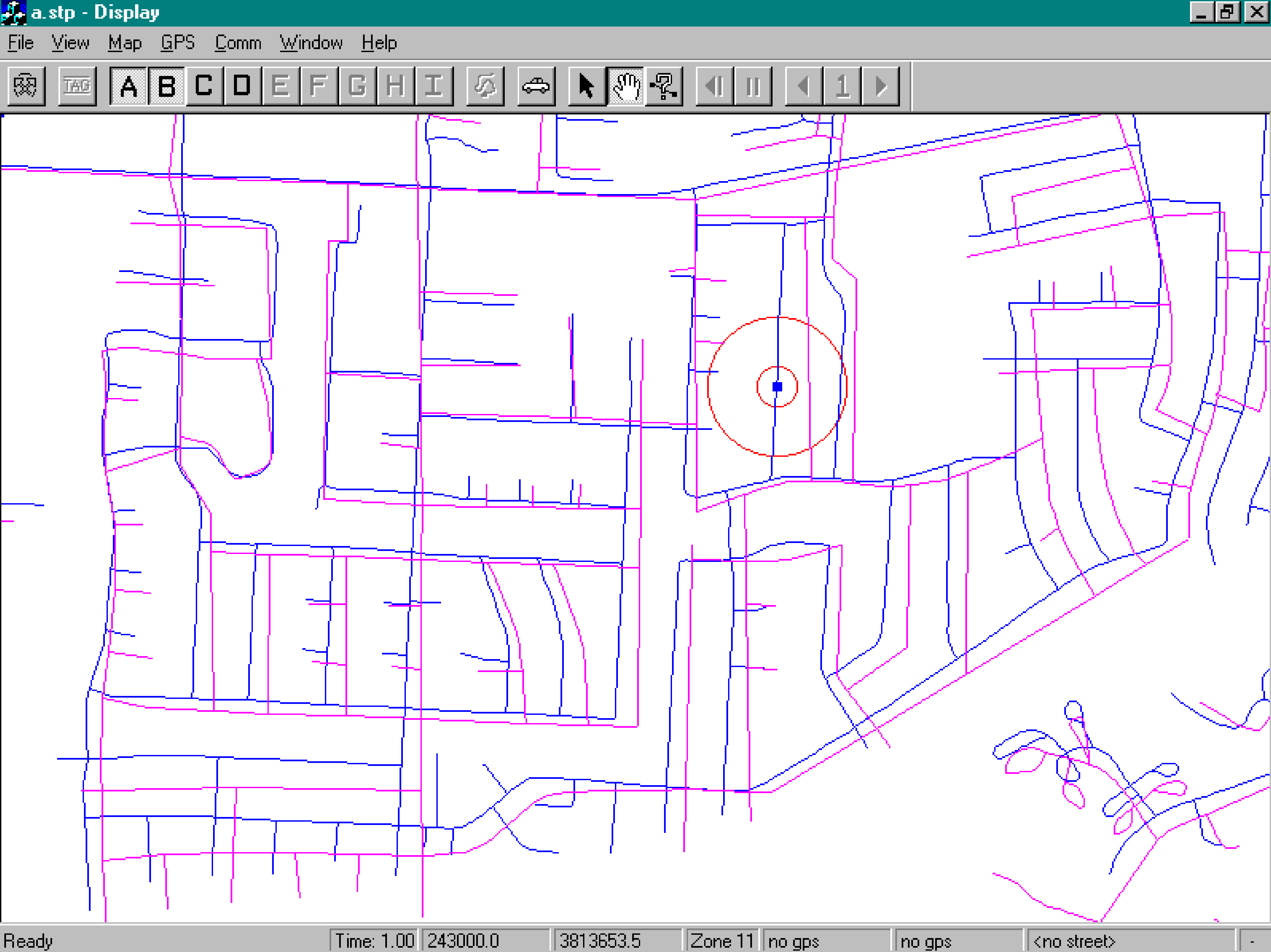
- Position as a measurement problem
- The world is infinitely complex
- The binary alphabet
  - hard limits?
  - uncertainty as difference
  - in the mind of the user
- Cartographic sources
  - the power of maps

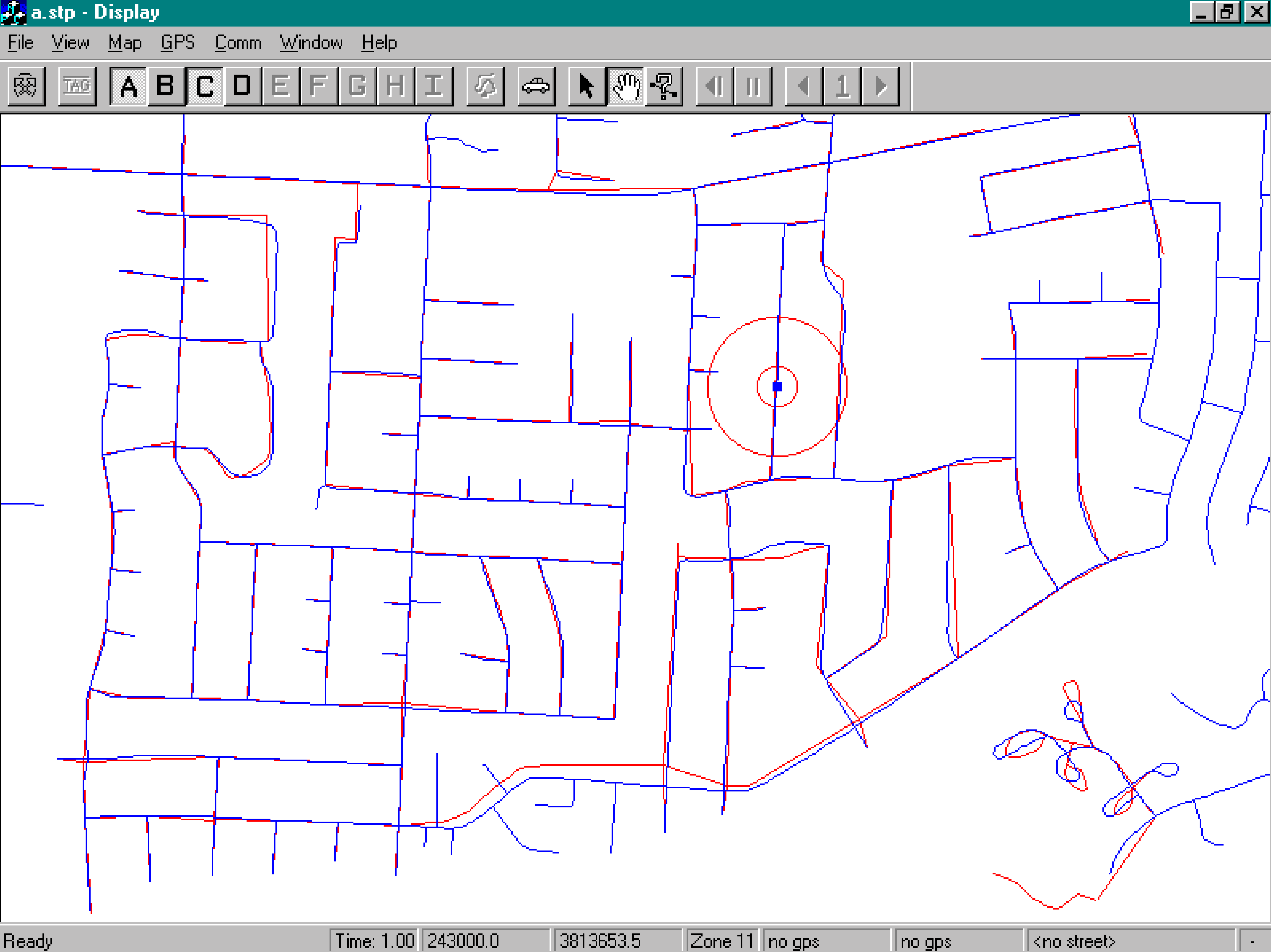

Tschöts
Braunhof
Schmuders
Ramings
Flains
Moos
Wiesen
STERZING
Thuins
942
1068
Deutschhaus
Custoza Hgl.
Sprechenstein
Wendelhof
Thumburg
Gasteig
Eisack
Elisabeth
Reifenstein
Elzenbaum
Freienfeld



Vipiteno
Raminges
di sopra
di sotto
Bagni di Maggio
Smudres
Castel Palù
Prati
(Sede Com.le)
Bruno
Flaines
Tunes
Villa
la Commenda
Tiro a segno
Casateia
di dentro
Olmedo
di fuori
Castel Tono
S. Zeno
Castel Tasso
Pruno
Campo di Trens
Castelpietra
M.te Vendi
M. DI CASACLUSA
Capo

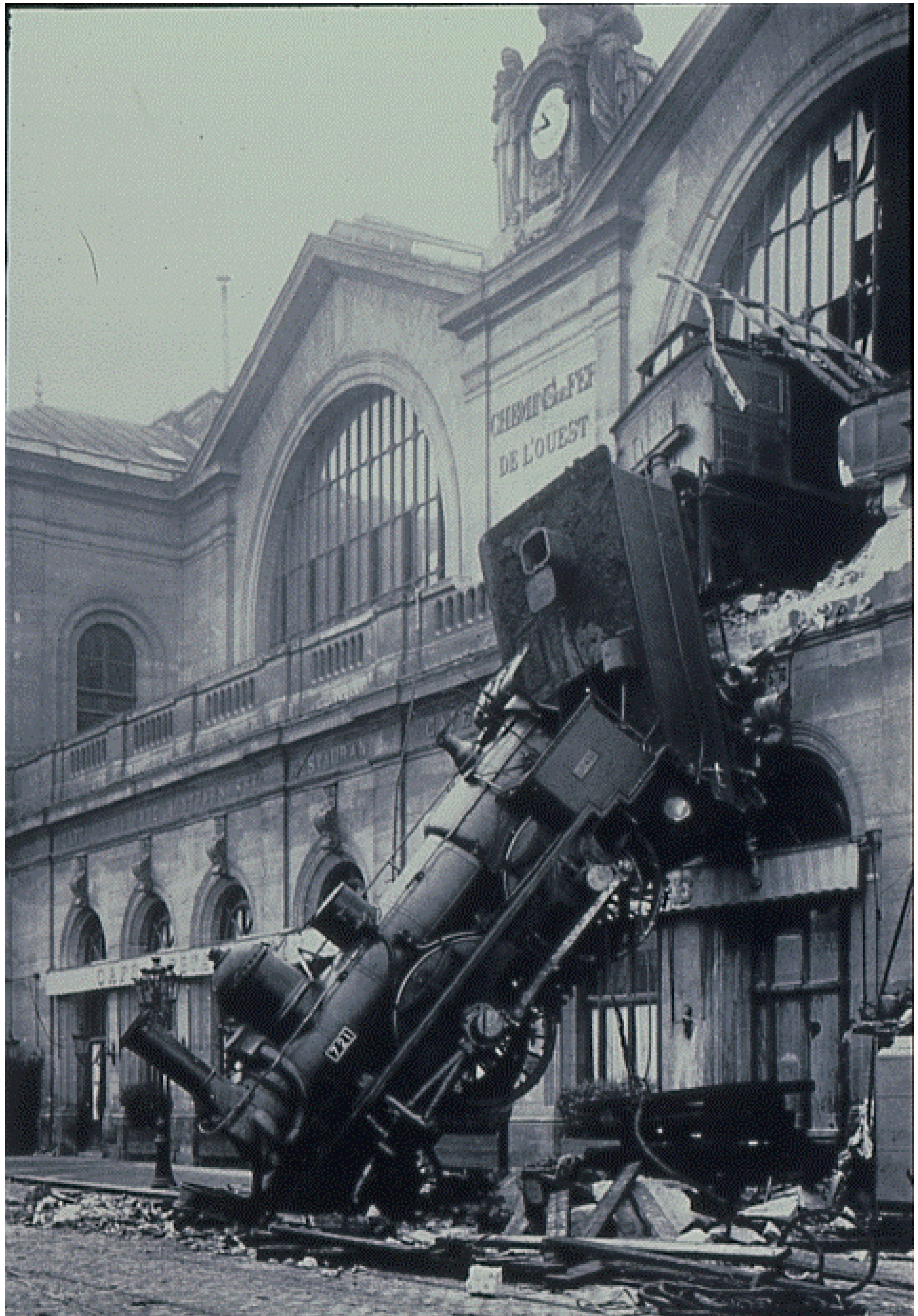

display.june.stp - Display
File View Map GPS Comm Window Help
TAG
A B C D E F G H I
1
Ready
Time: 1.00
241697.0
3814757.1
Zone 11
no gps
no gps
CLARENCE WARD MEMC
0 - 0
0 - 0
Start
fwy-ce.bmp - Paint
C:\Vital
display.june.stp - Dis...
13:19


a.stp - Display
File View Map GPS Comm Window Help
TAG
A B C D E F G H I
1
Ready
Time: 1.00
243000.0
3813653.5
Zone 11
no gps
no gps
<no street>


a.stp - Display
File
View
Map
GPS
Comm
Window
Help
TAG
A
B
C
D
E
F
G
H
I
1
Ready
Time: 1.00
243000.0
3813653.5
Zone 11
no gps
no gps
<no street>

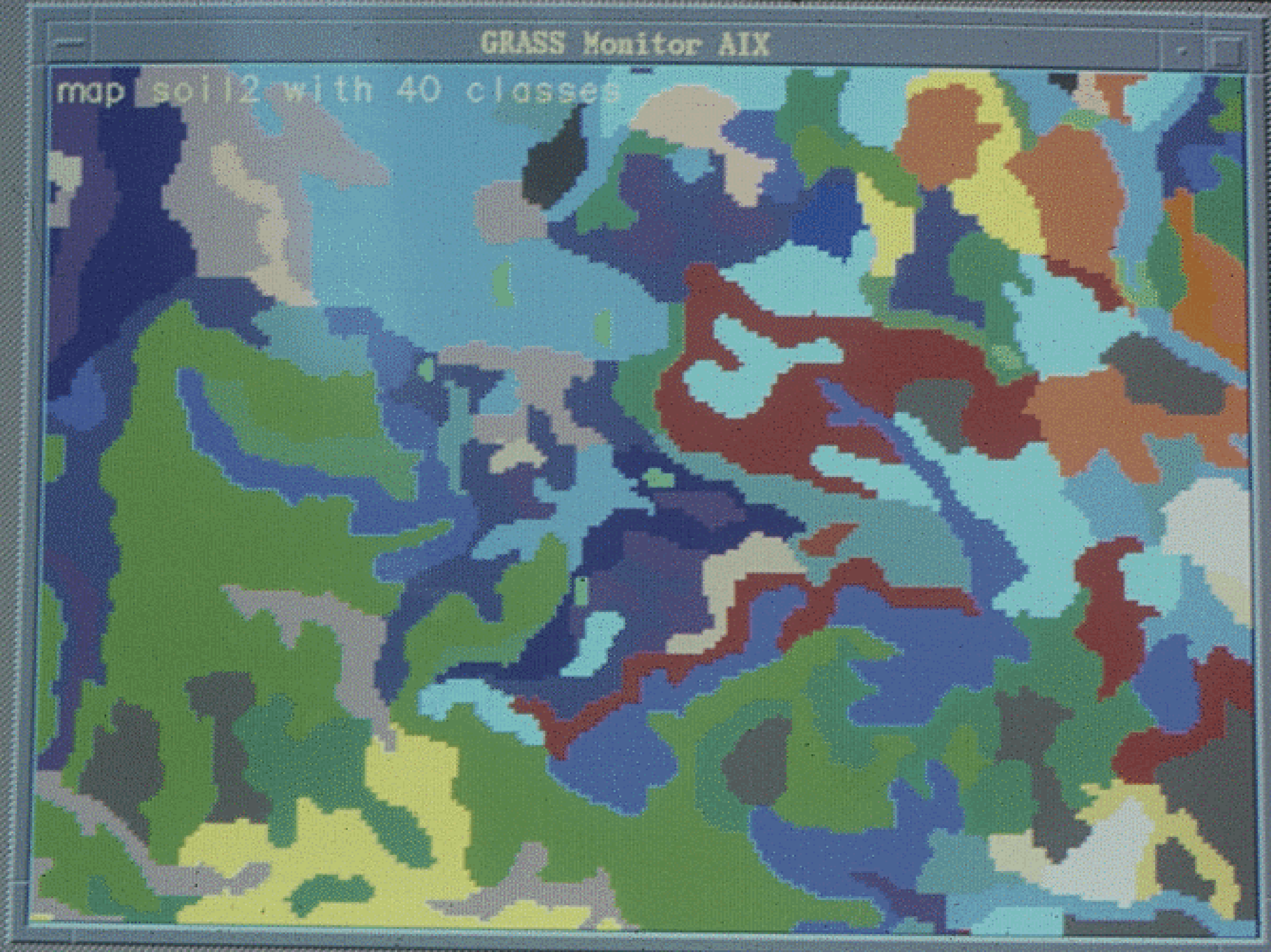
The GIS world

- Science
- Public policy
- Litigation
  - regulation by procedure
- Military, intelligence
- General user community


DE L'OUEST

Error modeling in GIS

- $\boldsymbol{Z} = \boldsymbol{z} + \varepsilon$
- Geostatistics
  - conditional simulation
  - co-Kriging
- Nominal fields


GRASS Monitor AIX
map soil2 with 40 classes

The area class map

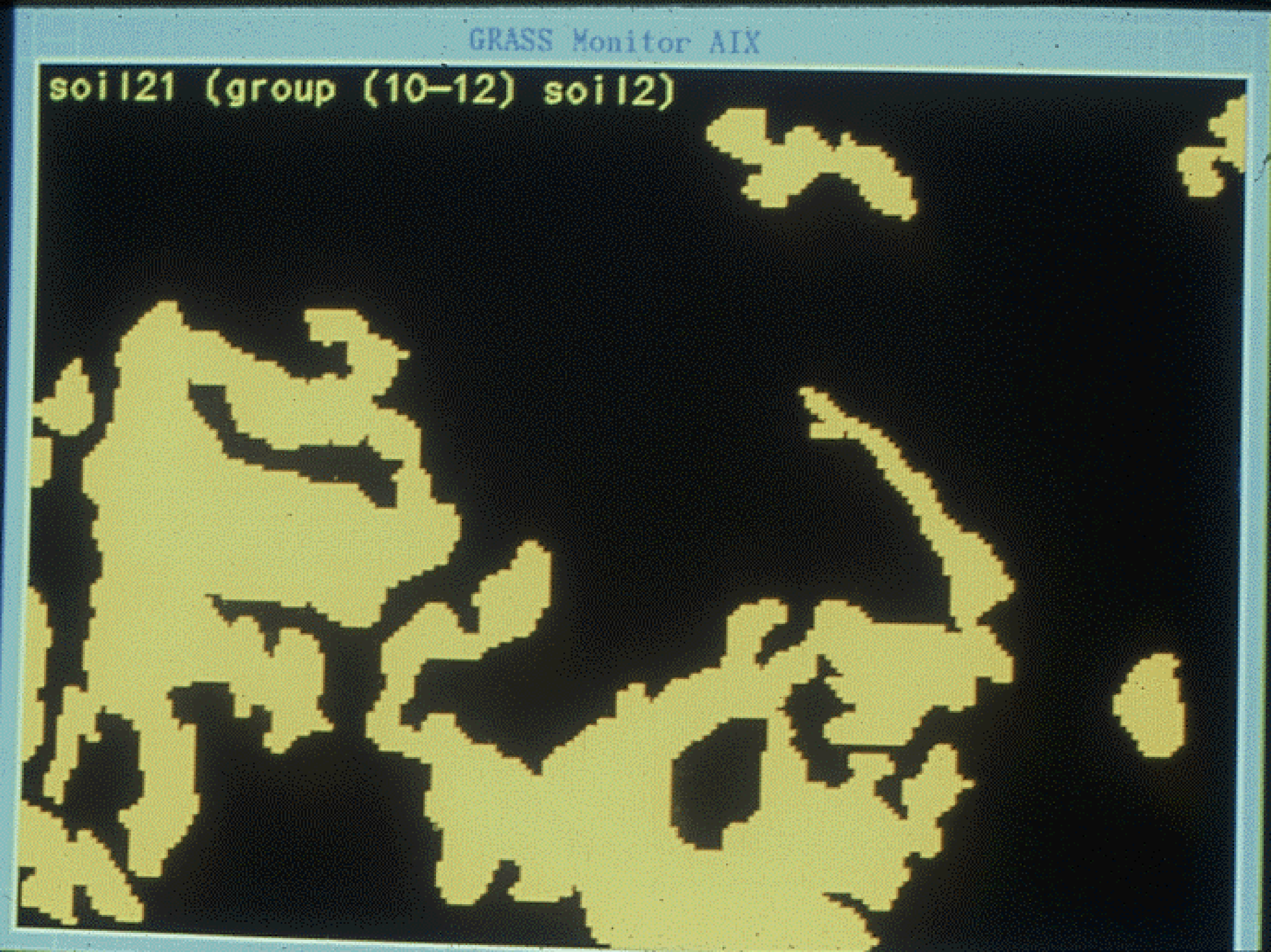
- Field of nominal values *c***(x),** *n*>1
  - spatially autocorrelated
  - count of *i*,*i* joins greater than expected
- Collection of discrete objects
  - nodes, edges, areas

The confusion matrix

- Useful descriptive device
  - quality control
- Comparing classifiers, observers, scales, accuracies
- Use all map (measure or count areas), sample of areas, points, pixels
- No information on joint distributions
  - little value for error modeling, uncertainty analysis

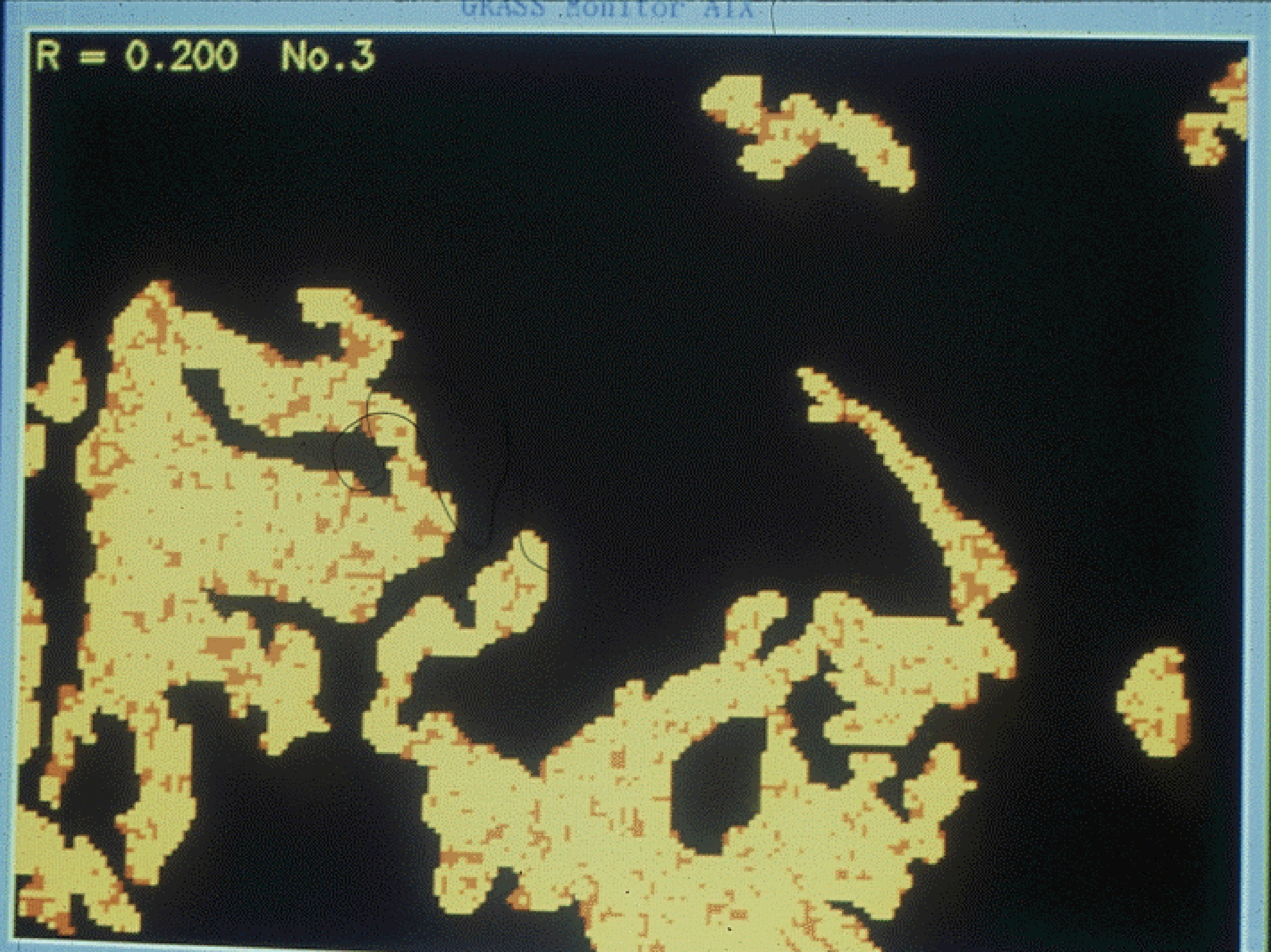
Spatial dependence in outcomes

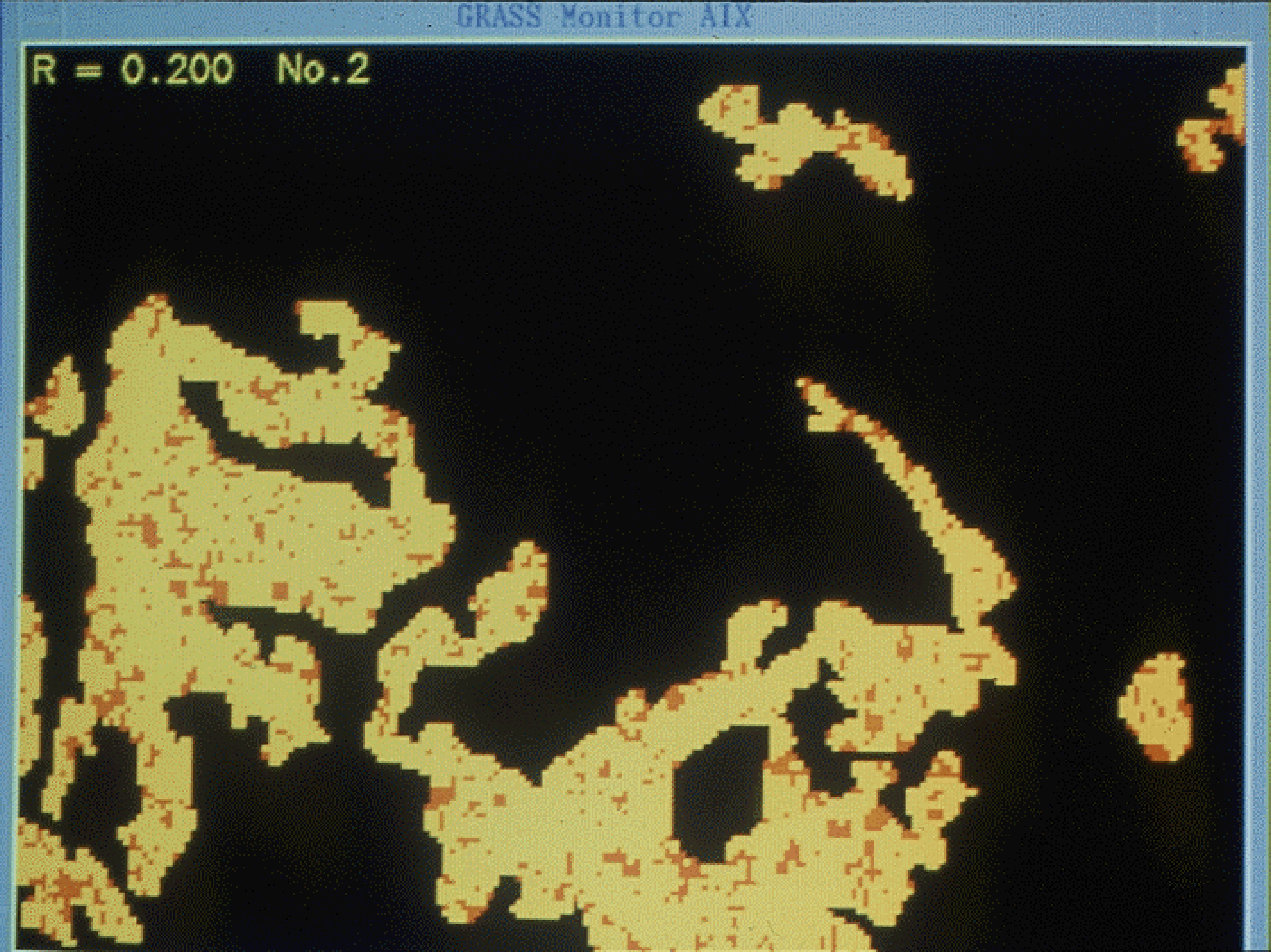
- Independent outcomes
  - zero spatial dependence between pixels
  - perfect positive spatial dependence within pixels
  - implies pixel size is meaningful
- Induce spatial dependence
  - range >> pixel size
  - spatial dependence falls smoothly
  - independent of pixel size

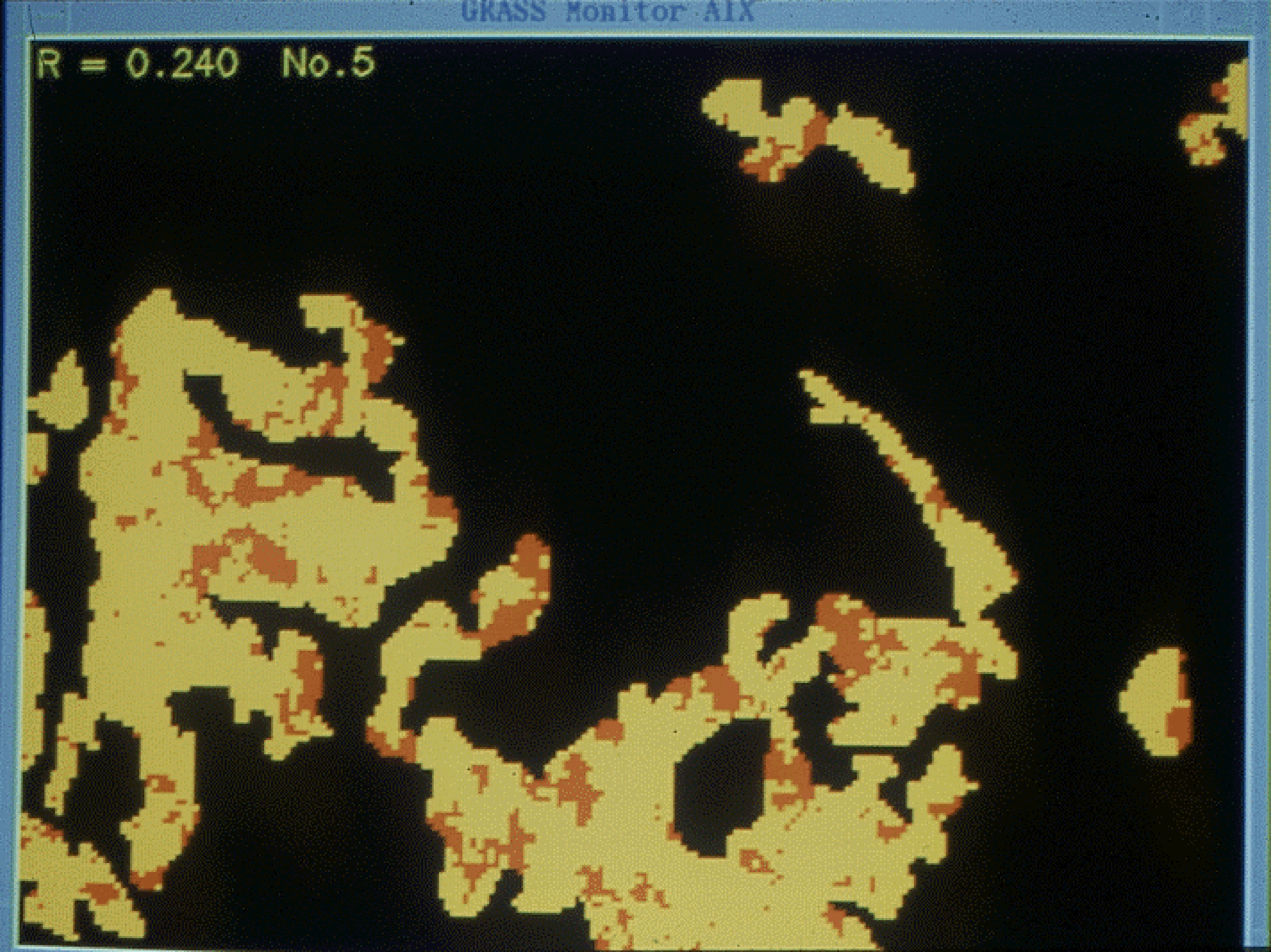

GRASS Monitor AIX
soil21 (group (10-12) soil2)

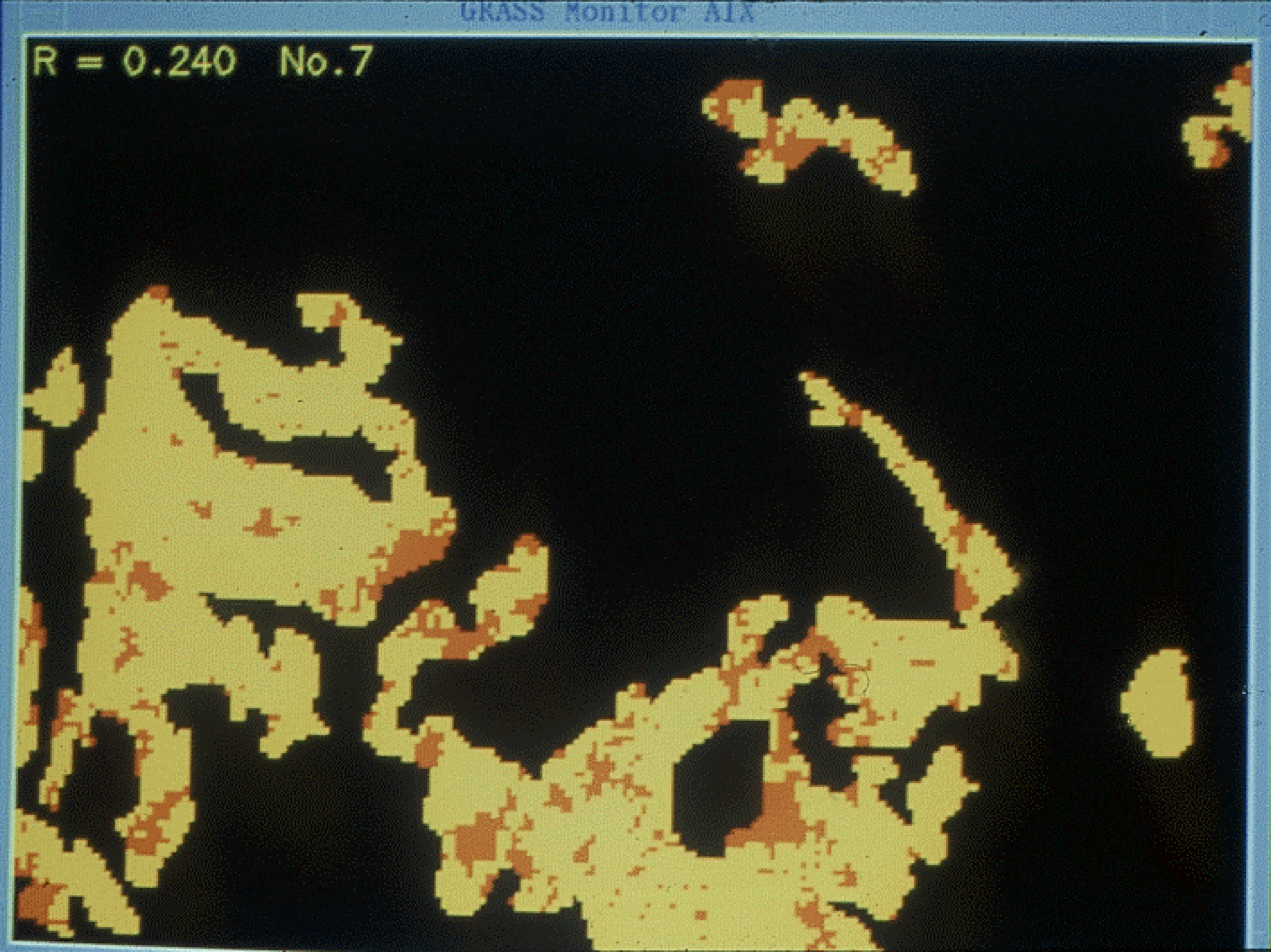
Model

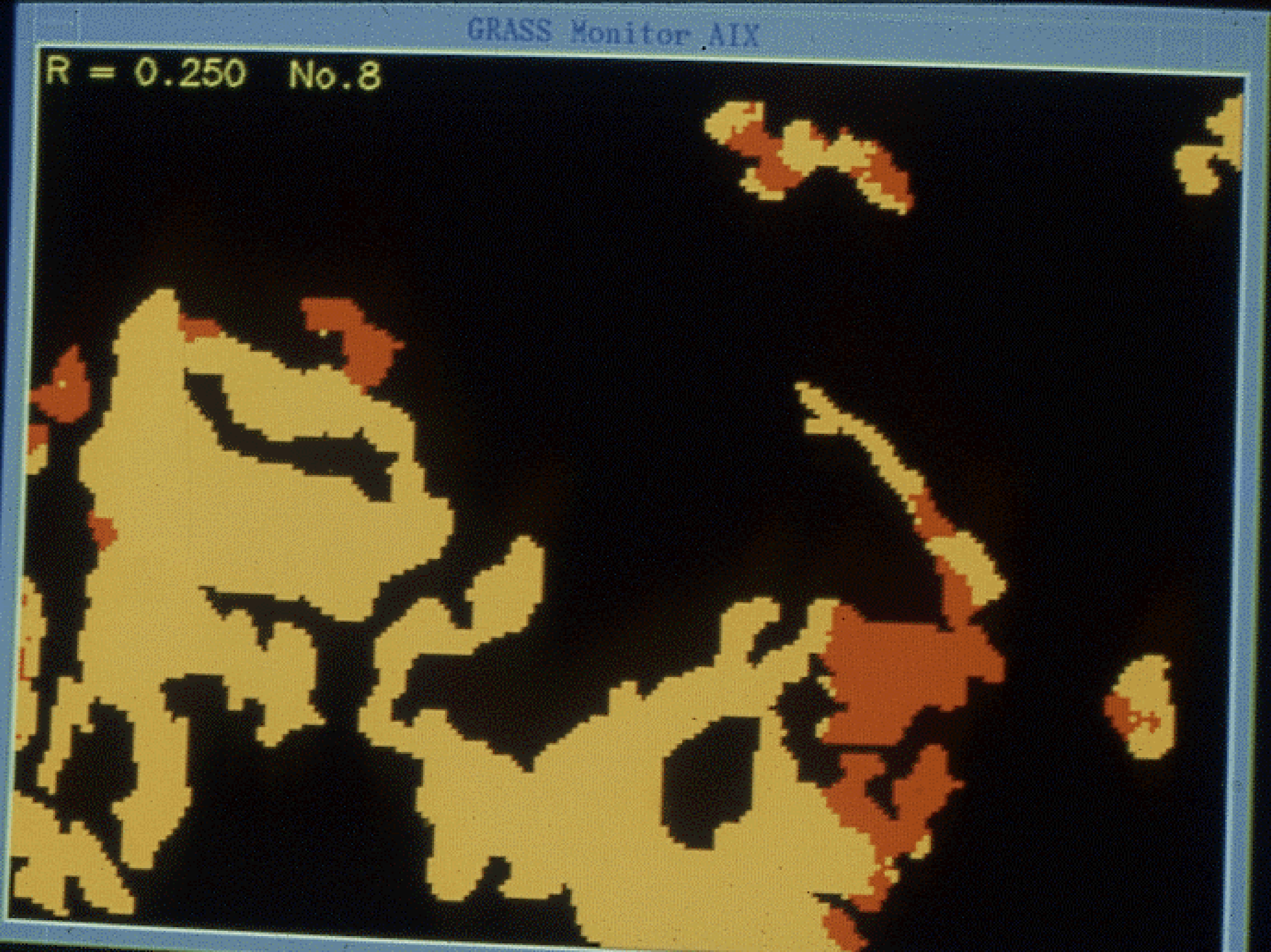
- $\{p_1, p_2, \ldots, p_n\}$
- correlation in neighboring cell outcomes
- posterior probabilities equal to priors
- 80% sand, 20% inclusions of clay
- no knowledge of correlations

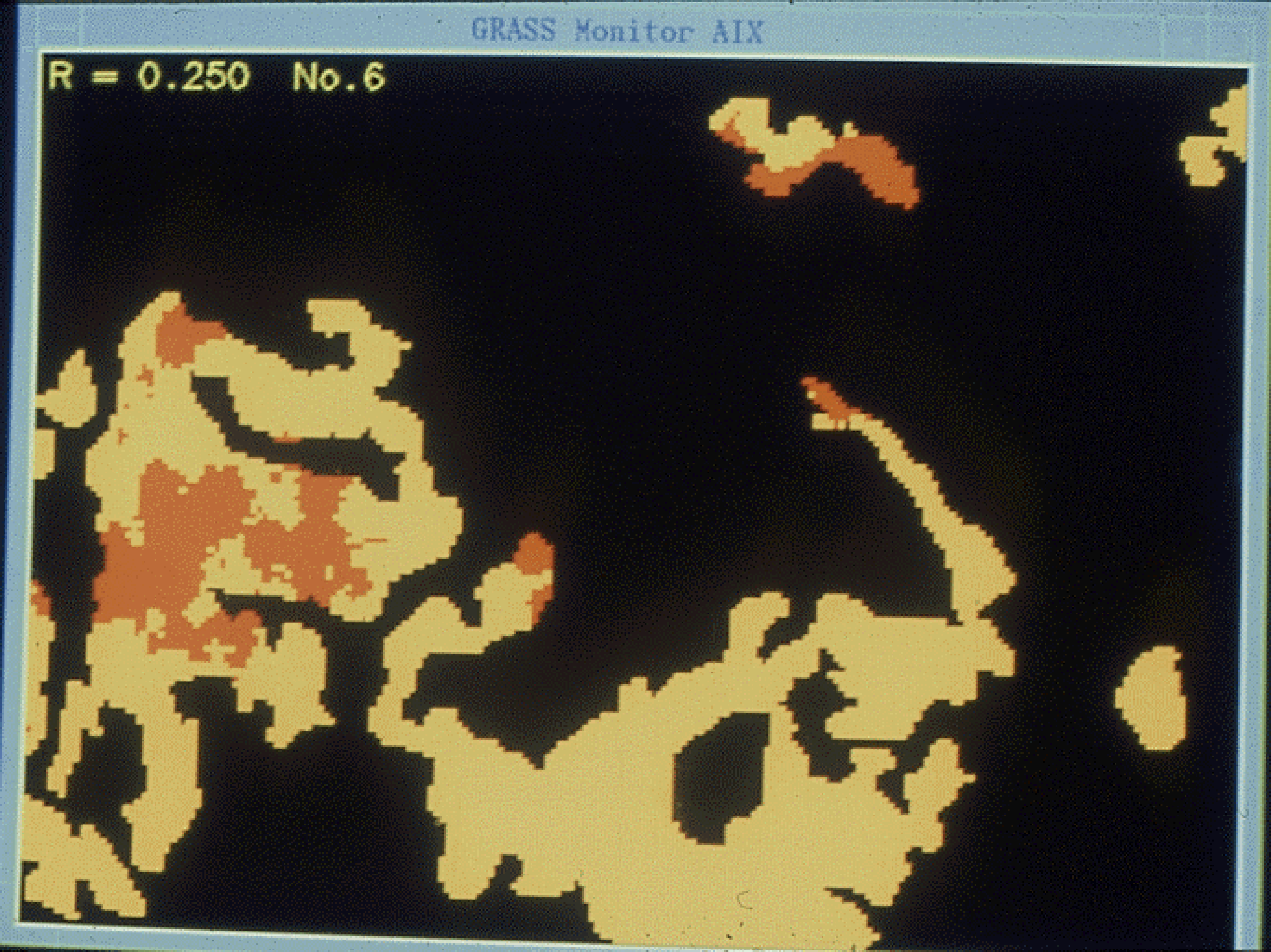

GRASS Monitor AIX
R = 0.200 No.3


GRASS Monitor AIX
R = 0.200 No.2


GRASS Monitor AIX
R = 0.240 No.5


GRASS Monitor AIX
R = 0.240 No.7


GRASS Monitor AIX
R = 0.250 No.8

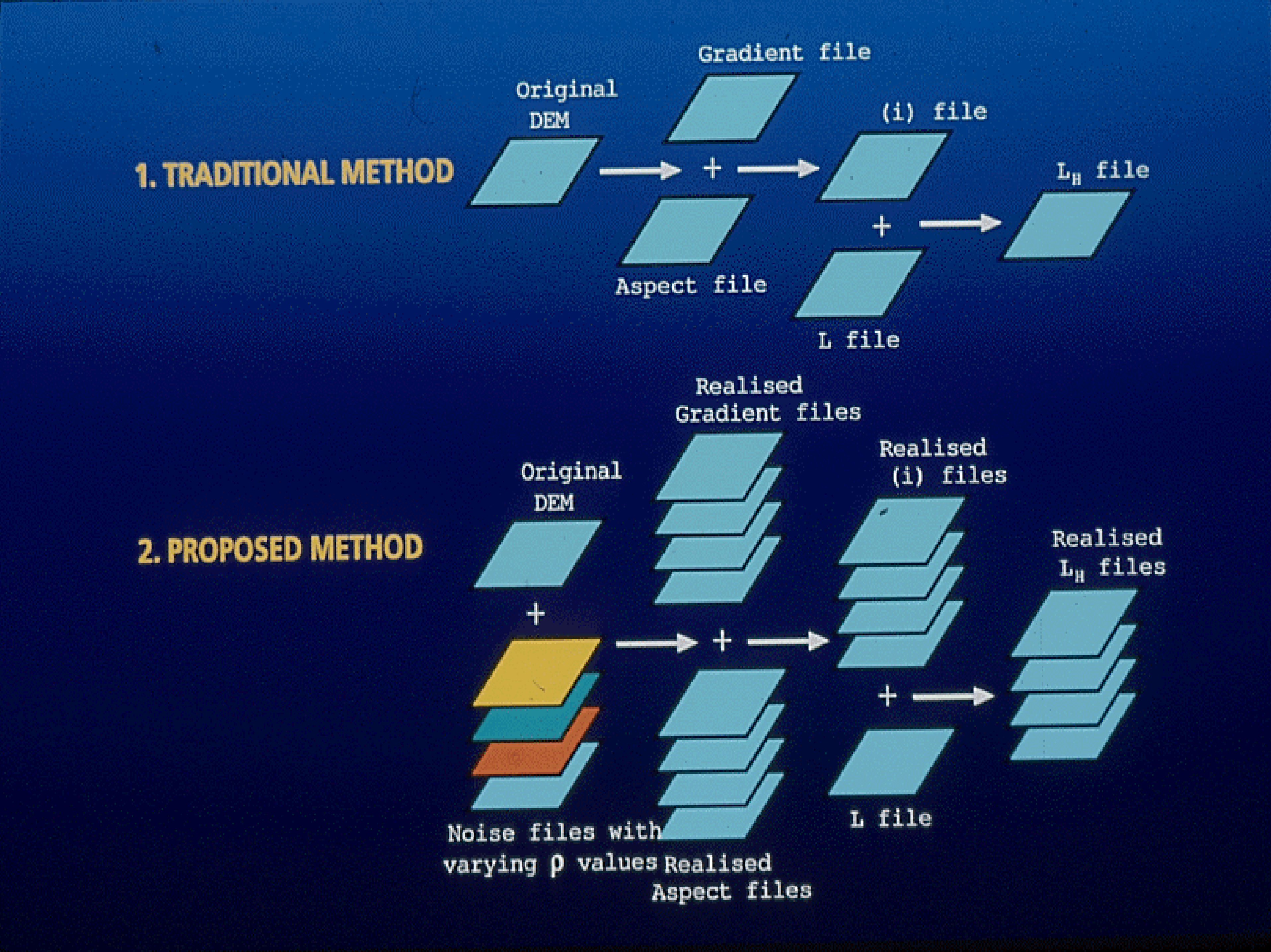

GRASS Monitor AIX
R = 0.250 No.6

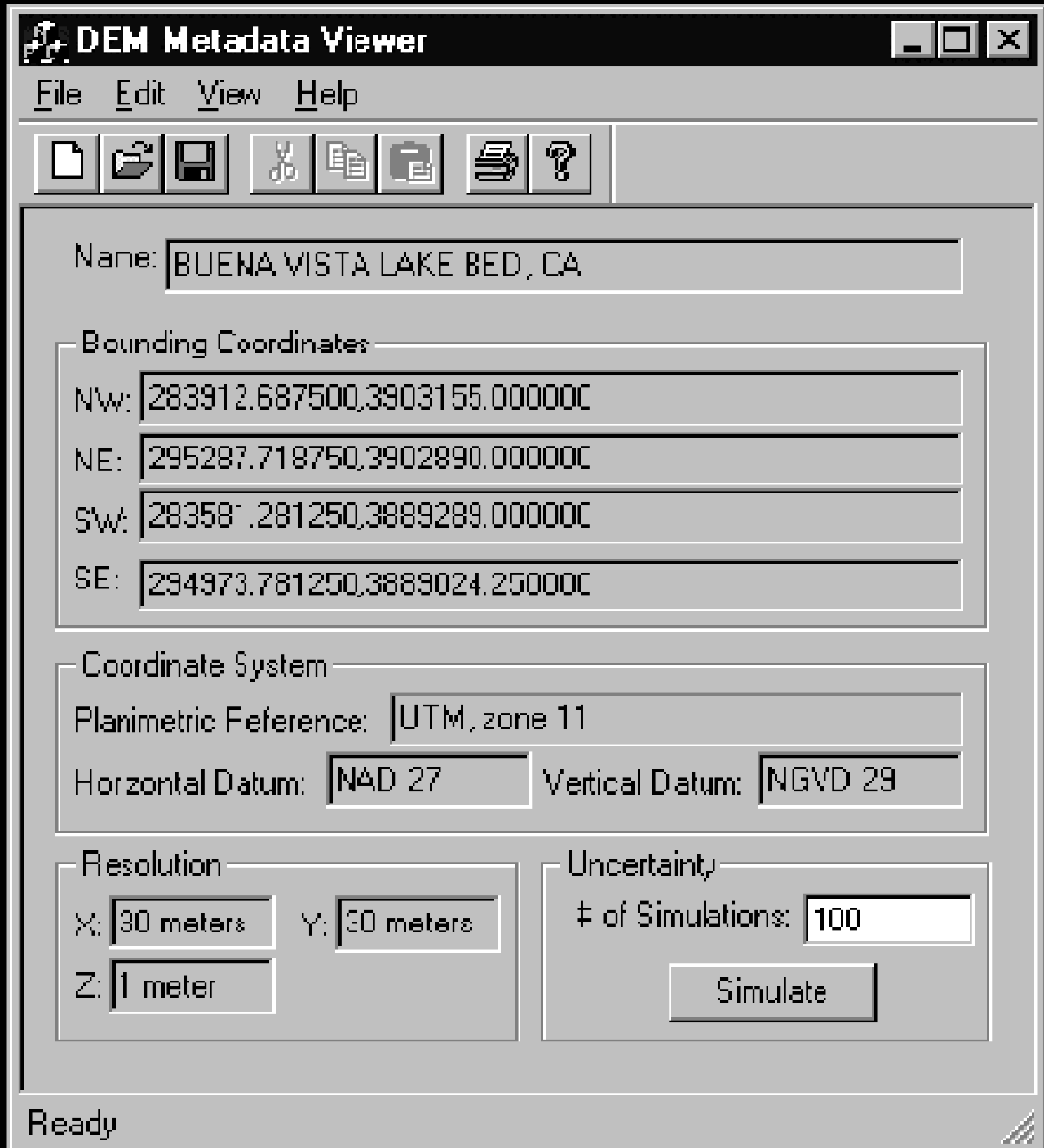
Estimates of Area and Associated Uncertainties for Various Levels of Spatial Dependence

ρ	Sample Size	Mean Estimate	Standard Error
0.000	(calculated)	6010	35
0.200	10	6139	156
0.240	10	6153	210
0.245	10	6158	268
0.249	10	6254	395
0.250	10	6514	753
LIMIT	(calculated)	6010	3005

Communication of uncertainty

- Producer to user
  - abilities
- Metadata standards
  - parameters of complex models
- Assertion:
  - all knowledge of uncertainty can be expressed in a suitable simulation model
  - equally likely realizations

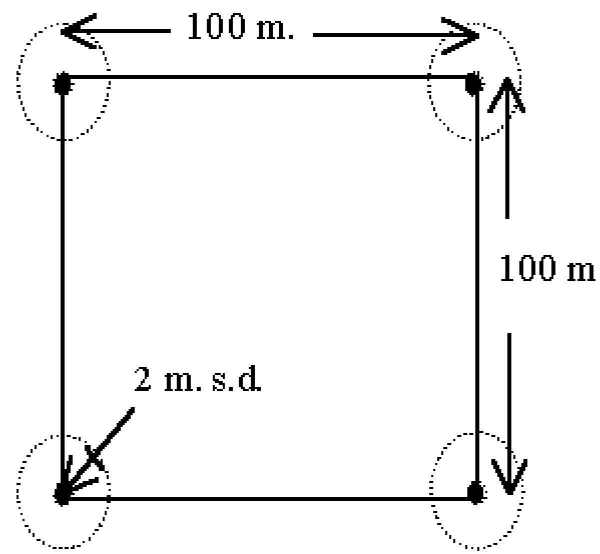

1. TRADITIONAL METHOD
Original DEM
Gradient file
Aspect file
(i) file
L file
L_H file
2. PROPOSED METHOD
Original DEM
Noise files with varying ρ values
Realised Gradient files
Realised Aspect files
Realised (i) files
L file
Realised L_H files


DEM Metadata Viewer
File Edit View Help
Name: BUENA VISTA LAKE BED, CA
Bounding Coordinates
NW: 283912.687500,3903155.00000C
NE: 295287.718750,3902890.00000C
SW: 28358ˈ.281250,3889289.00000C
SE: 294973.781250,3889024.25000C
Coordinate System
Planimetric Feference: UTM, zone 11
Horzontal Datum: NAD 27
Vertical Datum: NGVD 29
Resolution
X: 30 meters
Y: 30 meters
Z: 1 meter
Uncertainty
# of Simulations: 100
Simulate
Ready

Online Java demonstration

http://www.ncgia.ucsb.edu/~ashton/demos/propagate.html

Quadrilateral parcel defined by four surveyed points. The data consist of 4 (x,y) locations, and indicate they form a square 100m on a side, with an area of 10,000 square meters.

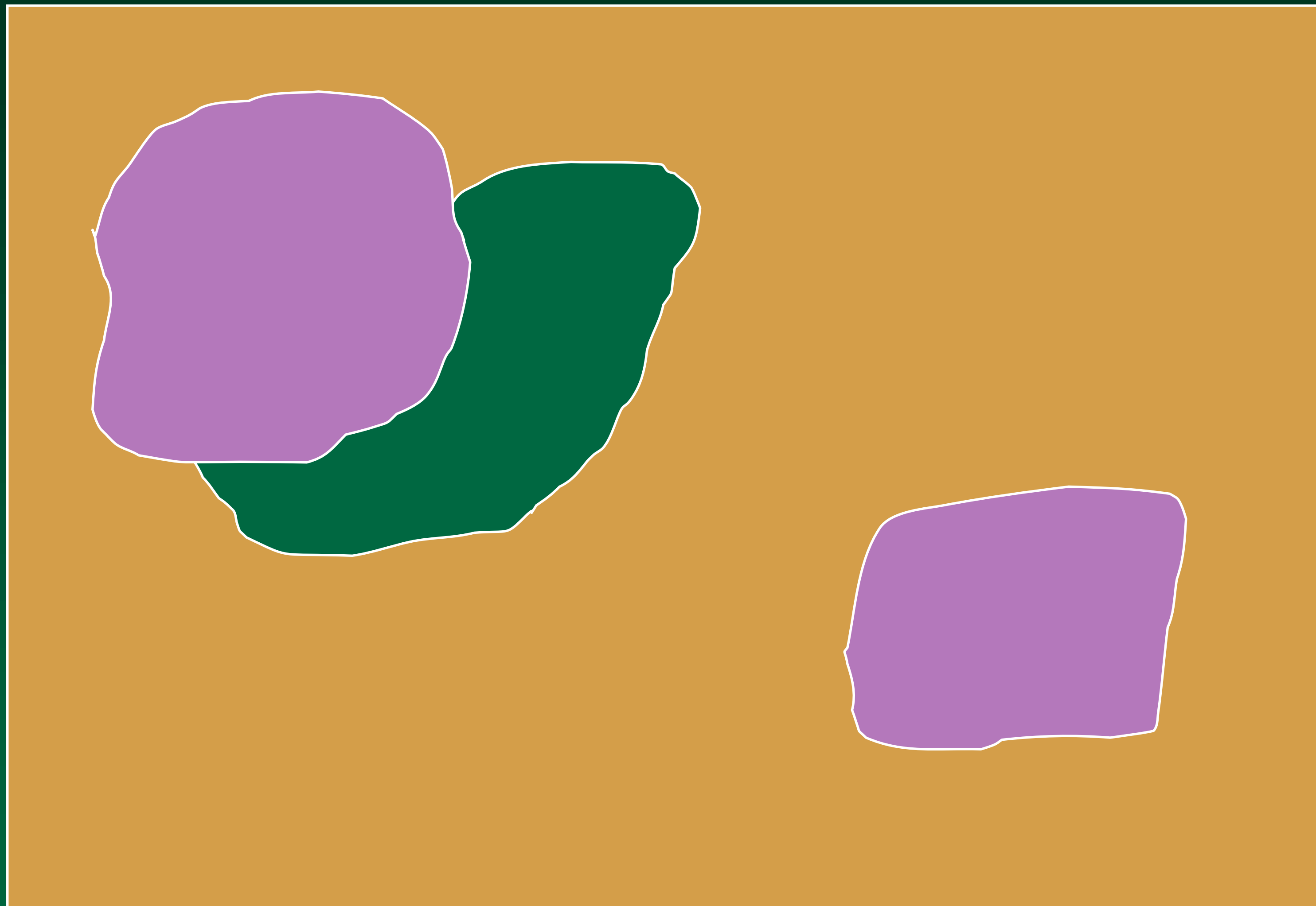


Uncertainty in point location is characterized by a Gaussian distribution, mean of zero and s.d. of 2 meters.

Question: What is the uncertainty associated with the area of the land parcel, given the positional uncertainty information?

n>2 classes

- Sequential assignment
  - class 1, not 1
  - class 2, not 2
  - …class *n*-1, *n*

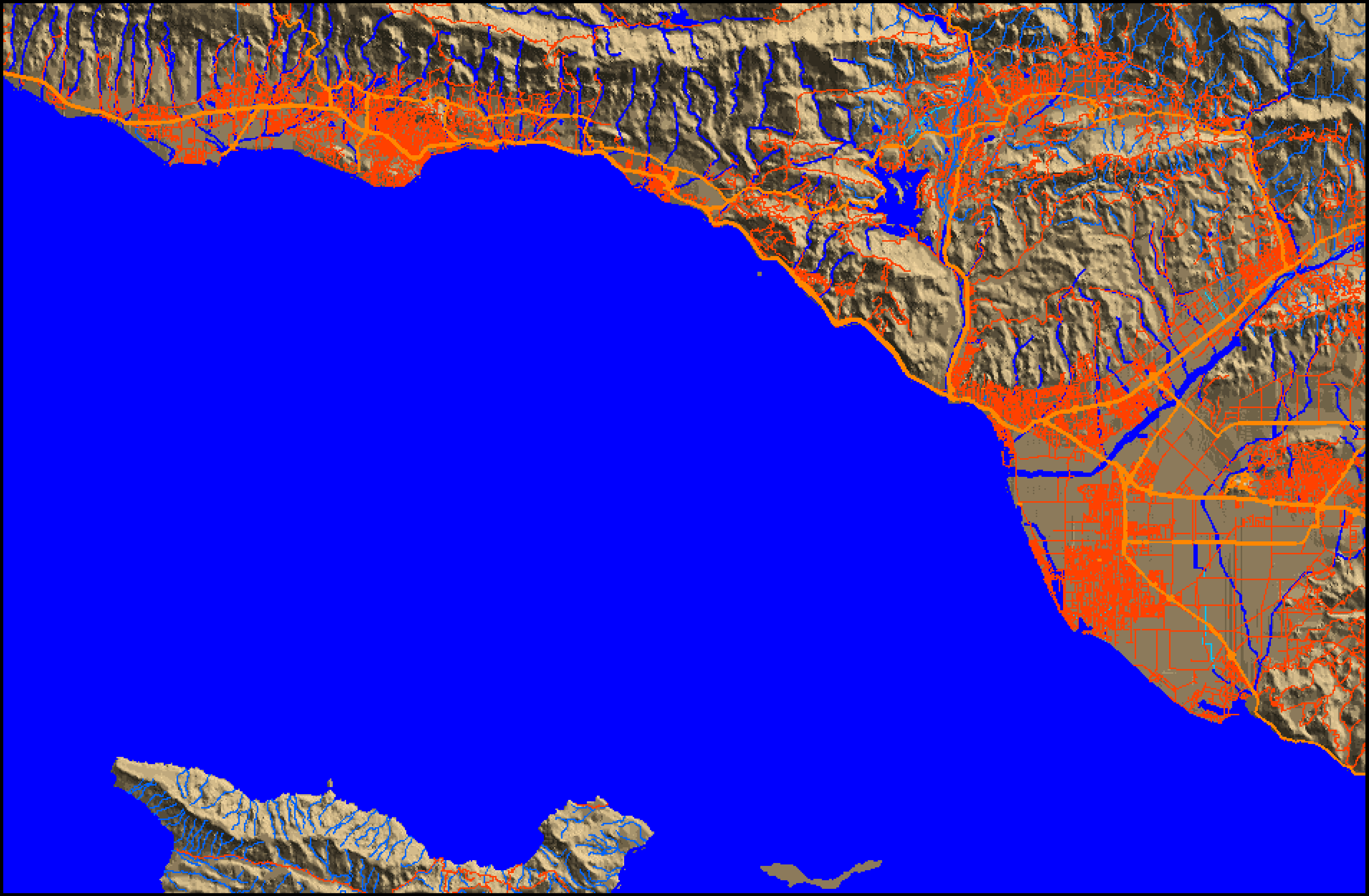


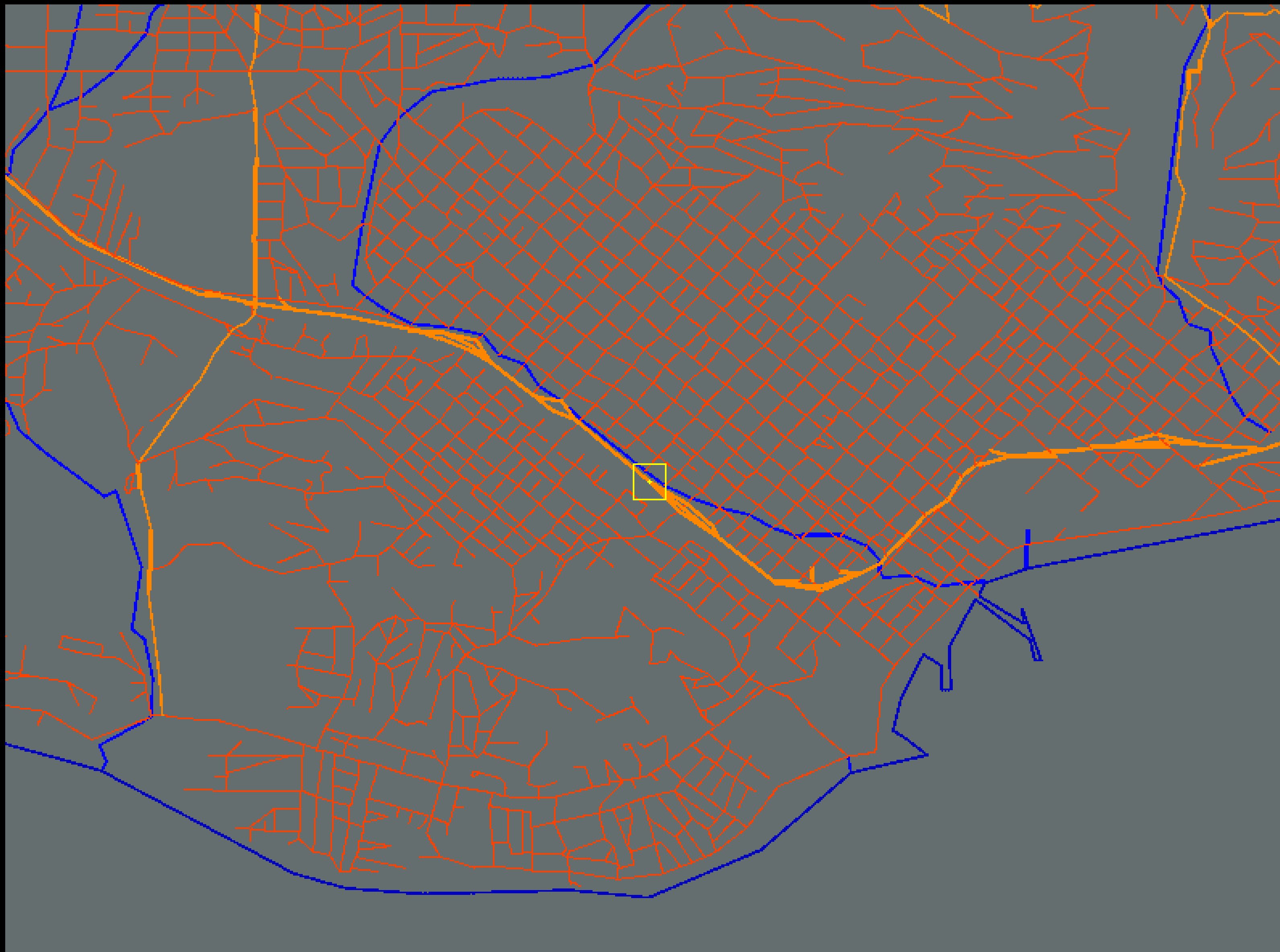
Process-based interpretation

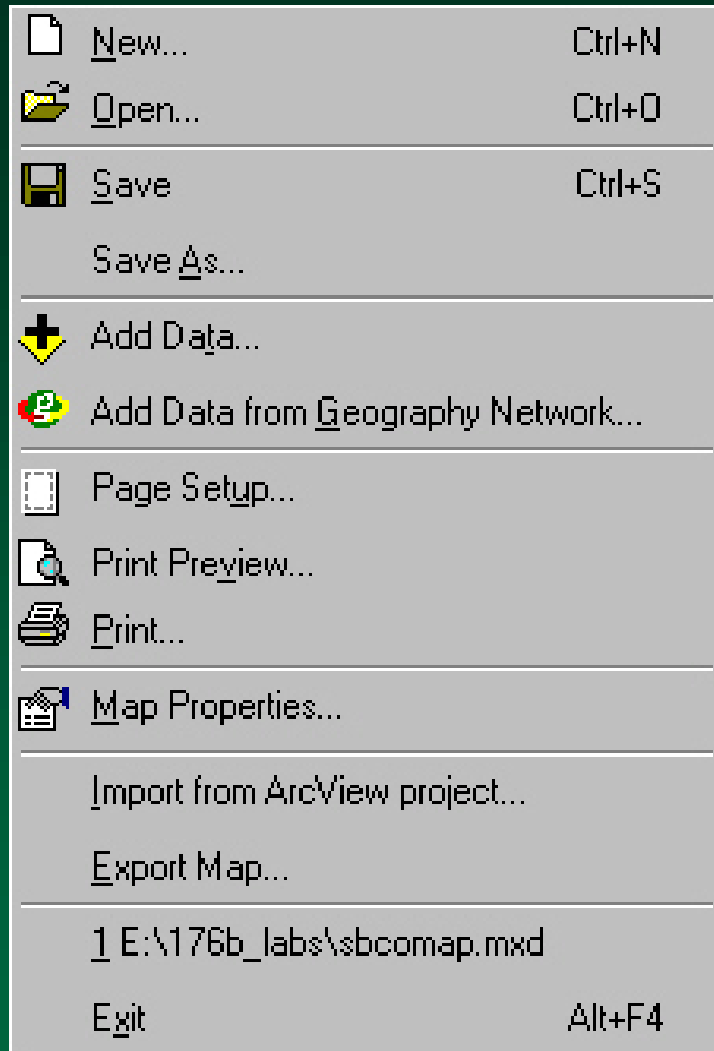
- Class *i* antecedent to class *i*-1
  - e.g. agriculture invaded by urban
  - e.g. grassland invaded by forest
  - shape of boundary between class *i* and class *i*-1 determined by class *i*
  - some applications have inherently ordered classes
    - but in this model all classes are ordered

A grand challenge of GIS

- To create useful, comprehensive digital representations of the enormous complexity of the Earth’s surface in the limited space of a digital store, using a binary alphabet
- An integrated, coherent organization of geographic information






New... Ctrl+N
Open... Ctrl+O
Save Ctrl+S
Save As...
Add Data...
Add Data from Geography Network...
Page Setup...
Print Preview...
Print...
Map Properties...
Import from ArcView project...
Export Map...
1 E:\176b_labs\sbcomap.mxd
Exit Alt+F4

geography network

ACCESS A WORLD OF INFORMATION

Challenge

SEARCH & VIEW

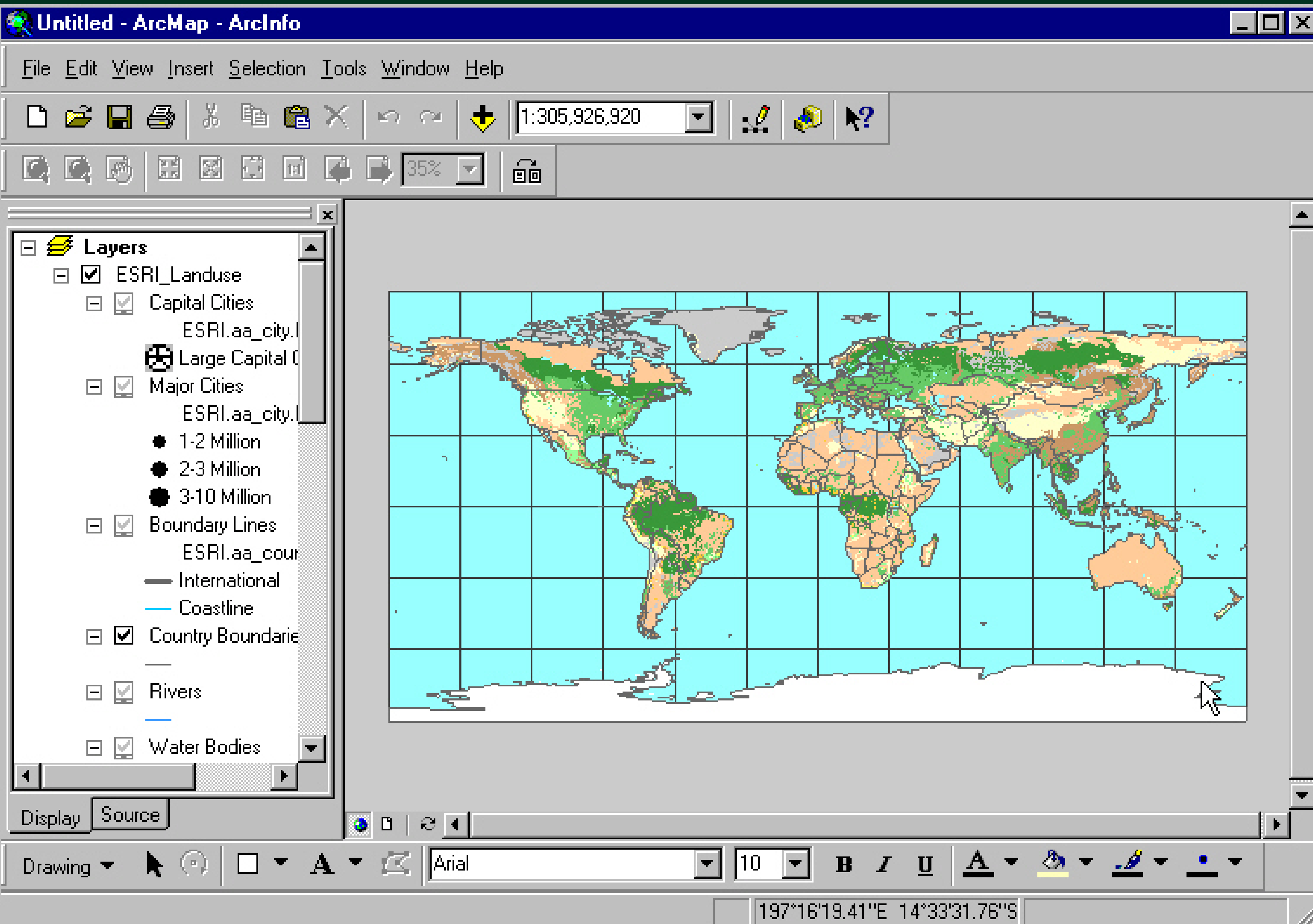
use the **Geography Network Explorer** to search and view maps and other geographic content over the Internet

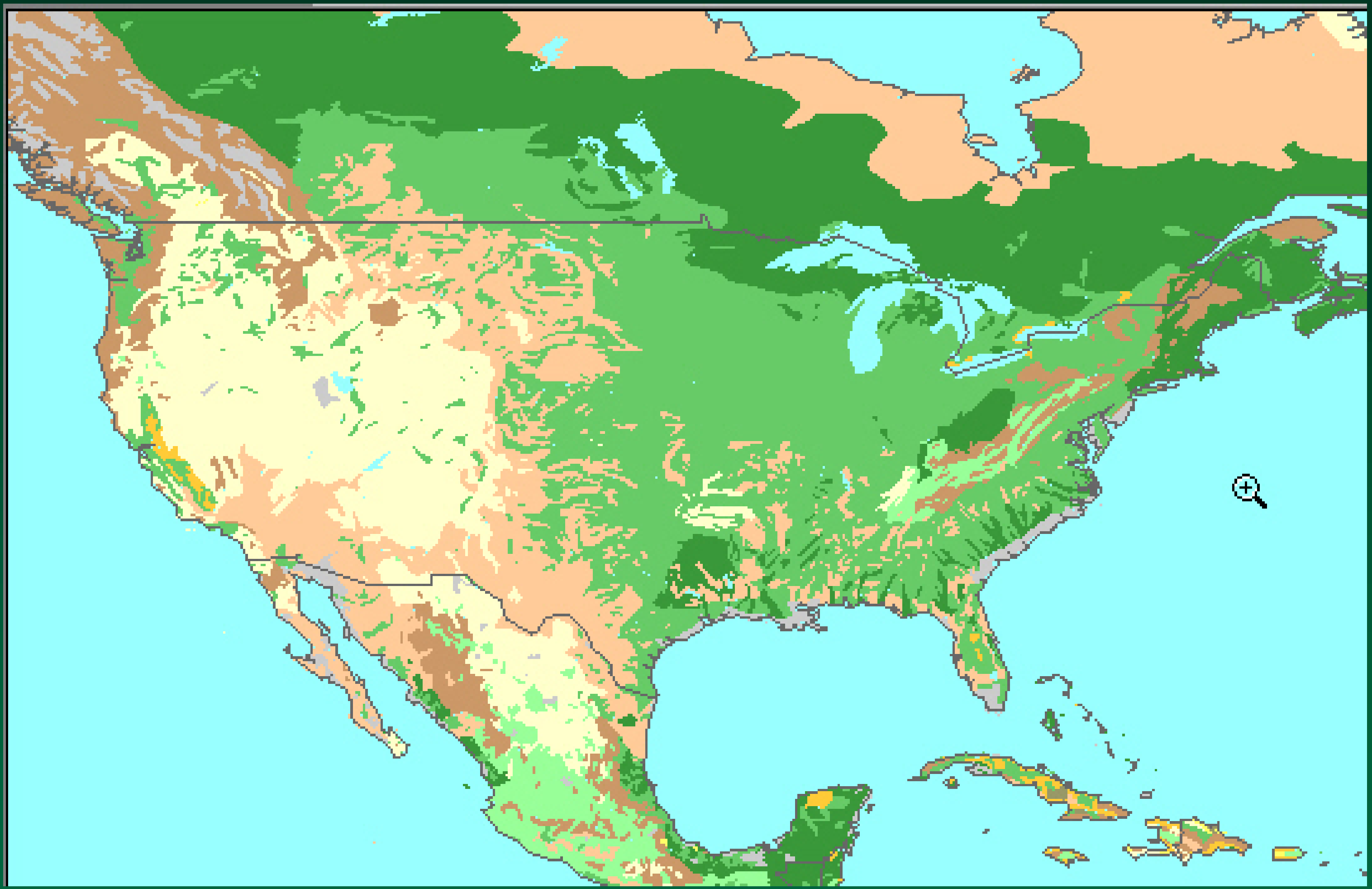
The Geography Network is a global community of data providers who are committed to making geographic content available. This content is published from many sites around the world, providing you immediate access to the latest maps, data, and related services. This portal to the Geography Network enables you to discover this content and share your own.

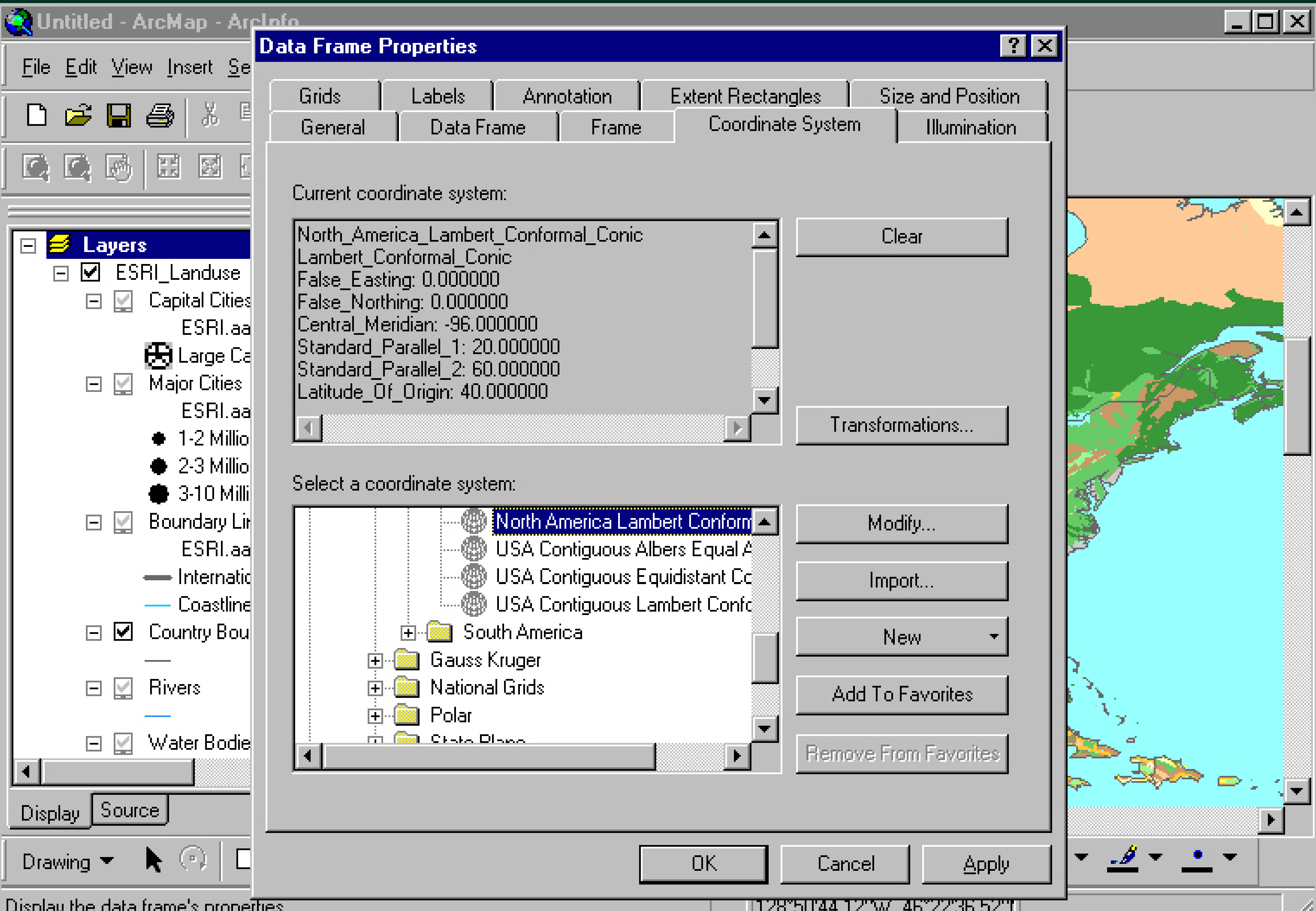
Featured Content

U.S. Census TIGER 2000

- View Live Maps
- Download Data Sets
- Build Custom Apps
- Publish Your Content
- Find Useful Tools
- Share Your Ideas


Untitled - ArcMap - ArcInfo
File Edit View Insert Selection Tools Window Help
1:305,926,920
35%
Layers
ESRI_Landuse
Capital Cities
ESRI.aa_city.l
Large Capital (
Major Cities
ESRI.aa_city.l
1-2 Million
2-3 Million
3-10 Million
Boundary Lines
ESRI.aa_cour
International
Coastline
Country Boundarie
Rivers
Water Bodies
Display
Source
Drawing
Arial
10
B I U
197°16'19.41"E 14°33'31.76"S




Untitled - ArcMap - ArcInfo
File Edit View Insert Se
Data Frame Properties
Grids
Labels
Annotation
Extent Rectangles
Size and Position
General
Data Frame
Frame
Coordinate System
Illumination
Current coordinate system:
North_America_Lambert_Conformal_Conic
Lambert_Conformal_Conic
False_Easting: 0.000000
False_Northing: 0.000000
Central_Meridian: -96.000000
Standard_Parallel_1: 20.000000
Standard_Parallel_2: 60.000000
Latitude_Of_Origin: 40.000000
Clear
Transformations...
Select a coordinate system:
North America Lambert Conform
USA Contiguous Albers Equal A
USA Contiguous Equidistant Co
USA Contiguous Lambert Confo
South America
Gauss Kruger
National Grids
Polar
Modify...
Import...
New
Add To Favorites
Remove From Favorites
OK
Cancel
Apply
Layers
ESRI_Landuse
Capital Cities
ESRI.aa
Large Ca
Major Cities
ESRI.aa
1-2 Millio
2-3 Millio
3-10 Milli
Boundary Li
ESRI.aa
Internatic
Coastline
Country Bou
Rivers
Water Bodie
Display
Source
Drawing
Display the data frame's properties
128°50'44.12"W 46°22'36.52"

A virtual Earth

- A representation of form
  - distributed, seamless, vertically integrated
- Representations of process
  - dynamic simulation models
  - integrated with the data
- Integrated with visualization, analysis clients

“Imagine, for example, a young child going to a Digital Earth exhibit at a local museum. After donning a head-mounted display, she sees Earth as it appears from space. Using a data glove, she zooms in, using higher and higher levels of resolution, to see continents, then regions, countries, cities, and finally individual houses, trees, and other natural and man-made objects. Having found an area of the planet she is interested in exploring, she takes the equivalent of a ‘magic carpet ride’ through a 3-D visualization of the terrain.”

Is Digital Earth feasible?

- 500,000,000,000 sq km
  - 5 million at 10km resolution
  - 500,000,000,000,000 at 1m resolution

[illegible]

The LS ratio

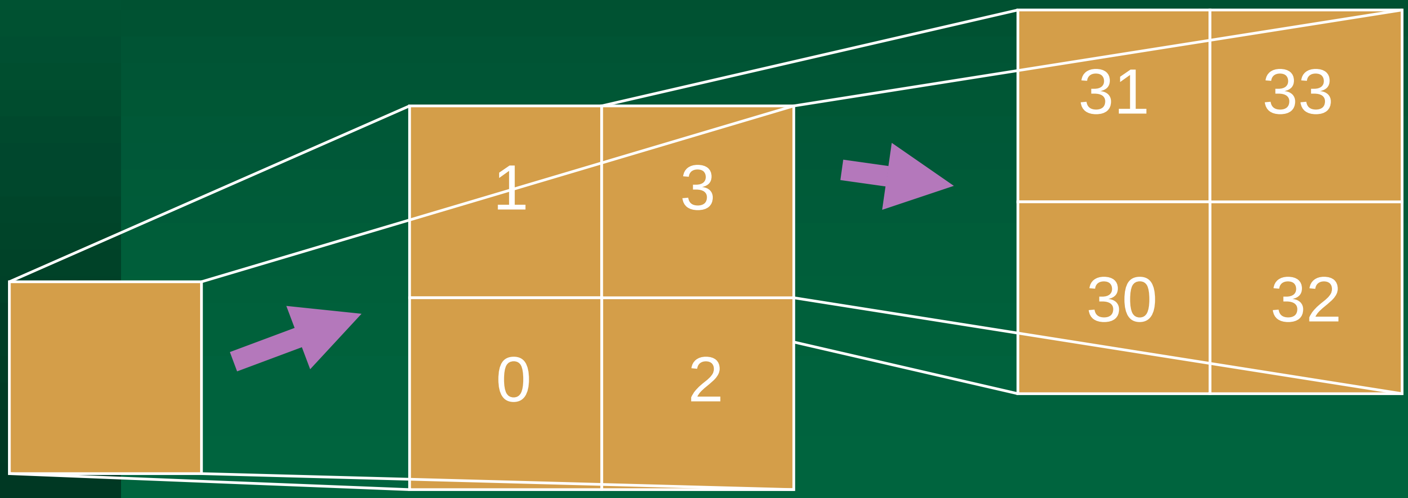
- Computer screen - 1000
- Digital camera - 1500
- Remotely sensed scene - 3000
- Paper map - 5000
- Dimensionless
- $Log_{10}L/S$ in range 3-4
- Human eye - 10,000

A data structure for DE

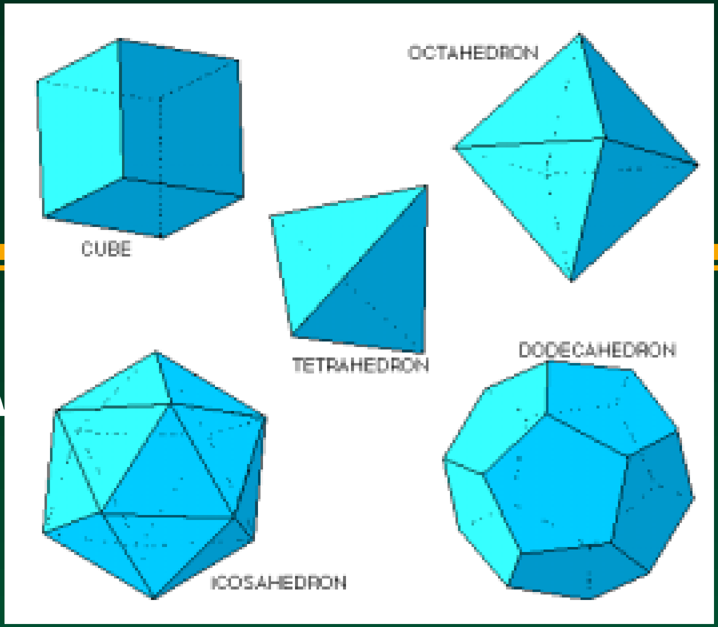
- To support smooth zooming over 4 orders of magnitude resolution
  - from 10km to 1m
  - maintaining LS ratio
- Vertically integrated
  - multiple layers

The quadtree

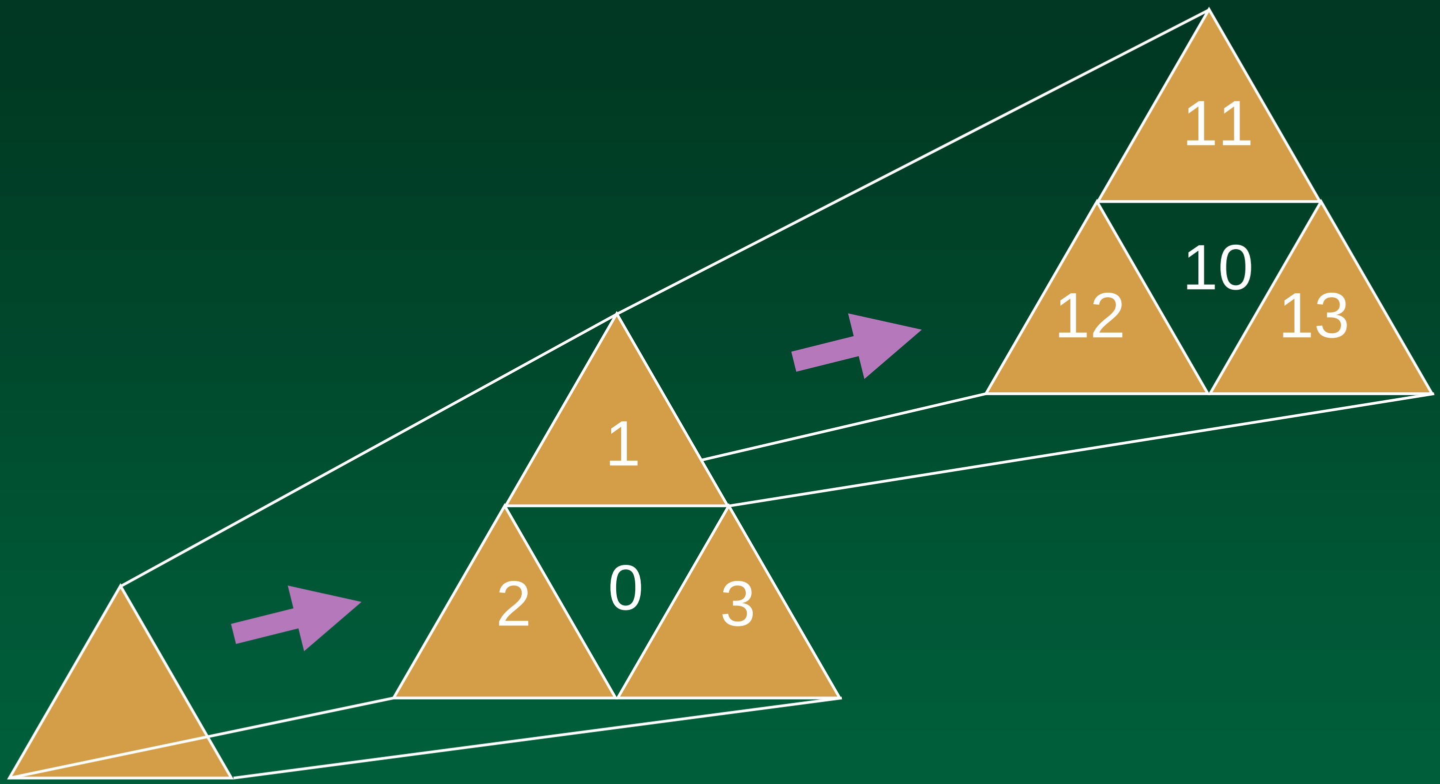
- Recursive subdivision
  - variable depth depending on local detail



Grids on the globe



- Impossible to tile a curv squares
- Five Platonic solids
  - tetrahedron: 4 triangles
  - cube: 6 squares
  - octahedron: 8 triangles
  - dodecahedron: 12 pentagons
  - icosahedron: 20 triangles



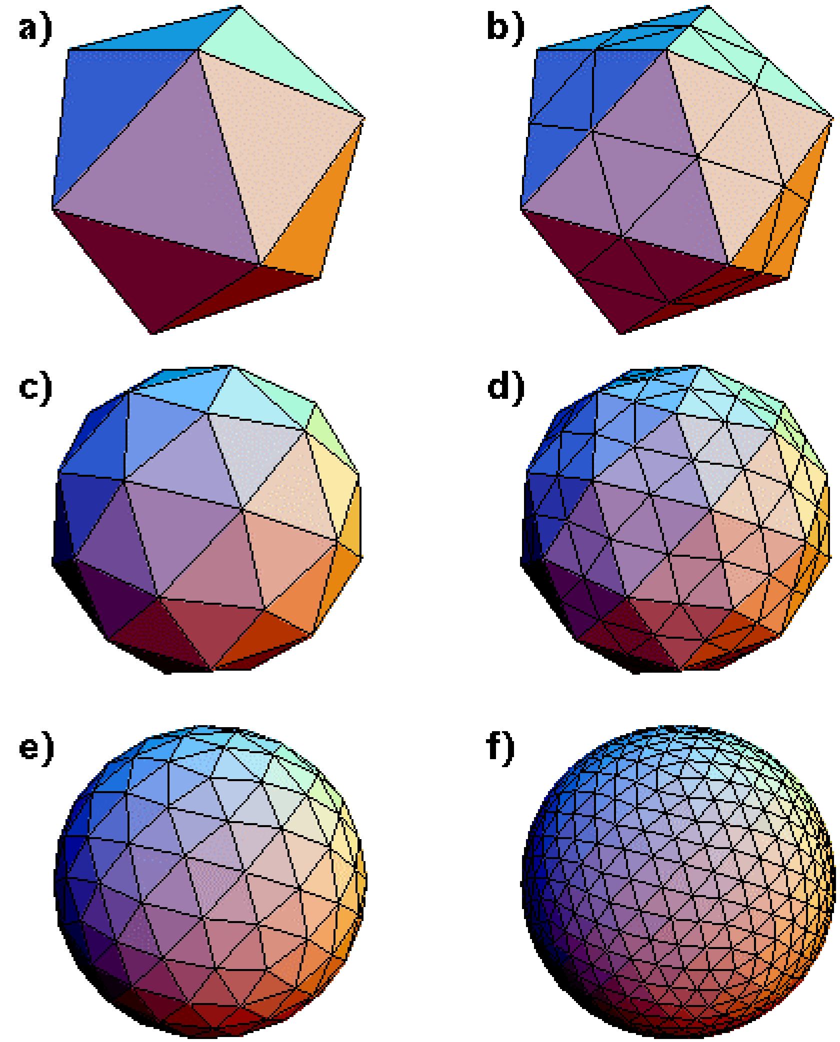
Octahedron: 1 base 8 digit plus unlimited base 4 digits

Discrete global grid based on the Icosahedron (20 triangles, 1:4 recursive subdivision)

Ross Heikes and David Randall, Colorado State University

Construction of a simple Icosahedral grid

a) Suppose we have an icosahedron inscribed inside of a unit sphere.
b) Bisecting each edge forms 30 new vertices, and partitions each equilateral face into four pieces.
c) Project the new vertices onto the unit sphere.
d) Bisect and partition again.
e) Project again.
f) And so on.... The result is a sequence of polyhedrons that increasingly approximate the sphere.



Comparison of Criteria for the Assessment of Global Grids

Criteria in Goodchild (1994)	Criteria in Kimerling et al. (1999) (Goodchild's Numbers given in parentheses)
1. Each area contains one point	Areal cells constitute a complete tiling of the globe, exhaustively covering the globe without overlapping. (3,7)
2. Areas are equal in size	Areal cells have equal areas. This minimizes the confounding effects of area variation in analysis, and provides equal probabilities for sampling designs. (2)
3. Areas exhaustively cover the domain	Areal cells have the same topology (same number of edges and vertices). (9, 14)
4. Areas are equal in shape	Areal cells have the same shape. ideally a regular spherical polygon with edges that are great circles. (4)
5. Points form a hierarchy preserving some property for m < n points	Areal cells are compact. (10)
6. Areas form a hierarchy preserving some property for m < n areas	Edges of cells are straight in a projection. (8)
7. The domain is the globe (sphere, spheroid)	The midpoint of an arc connecting two adjacent cells coincides with the midpoint of the edge between the two cells.
8. Edges of areas are straight on some projection	The points and areal cells of the various resolution grids which constitute the grid system form a hierarchy which displays a high degree of regularity. (5,6)
9. Areas have the same number of edges	A single areal cell contains only one grid reference point.(1)
10. Areas are compact	Grid reference points are maximally central within areal cells. (11)
11. Points are maximally central within areas	Grid reference points are equidistant from their neighbors. (12)
12. Points are equidistant	Grid reference points and areal cells display regularities and other properties which allow them to be addressed in an efficient manner.
13. Edges are areas of equal length	The grid system has a simple relationship to latitude and longitude.
14. Addresses of points and areas are regular and reflect other properties	The grid system contains grids of any arbitrary defined spatial resolution. (5,6)

Imagery courtesy of WorldSat Intl. Inc.

GeoFusion

Imagery courtesy of WorldSat Intl. Inc.

GeoFusion

Concluding messages

- GIS raises fundamental questions in several disciplines
  - geography, computer science, cognitive science
- Uncertainty is endemic, must be addressed explicitly
- Integrated digital representations of Earth are now feasible