This report is very disappointing.

What kind of software are you using?

Stone-Age Space-Age Syndrome

	Stone-age data	Space-age data
Stone-age analysis	⊕	⊕
Space-age analysis		⊕

- Project MARMAP SYSTEM PARTNERSHIP

  Multiscale Advanced Raster Map Analysis System
- GEOGRAPHICAL SURVEILLANCE

  Hot Spot Detection, Delineation, and Prioritization
- A TREE-STRUCTURED SATSCAN APPROACH

DETECTION AND DELINEATION OF CRITICAL AREAS USING ECHELONS AND SPATIAL SCAN STATISTICS WITH SYNOPTIC CELLULAR DATA

G. P. Patil[1], J. Bishop[2], W. L. Myers[3], C. Taillie[1], R. Vraney[1], and Denice Wardrop[2]

[1]Center for Statistical Ecology and Environmental Statistics

[2]Cooperartive Wetlands Center

[3]Office of Remote Sensing for Earth Resources

The Pennsylvania State University

Special Thematic Conference on
Healthy Ecosystems/Healthy People:
Linkages Between Biodiversity,
Ecosystem Health and Human Health

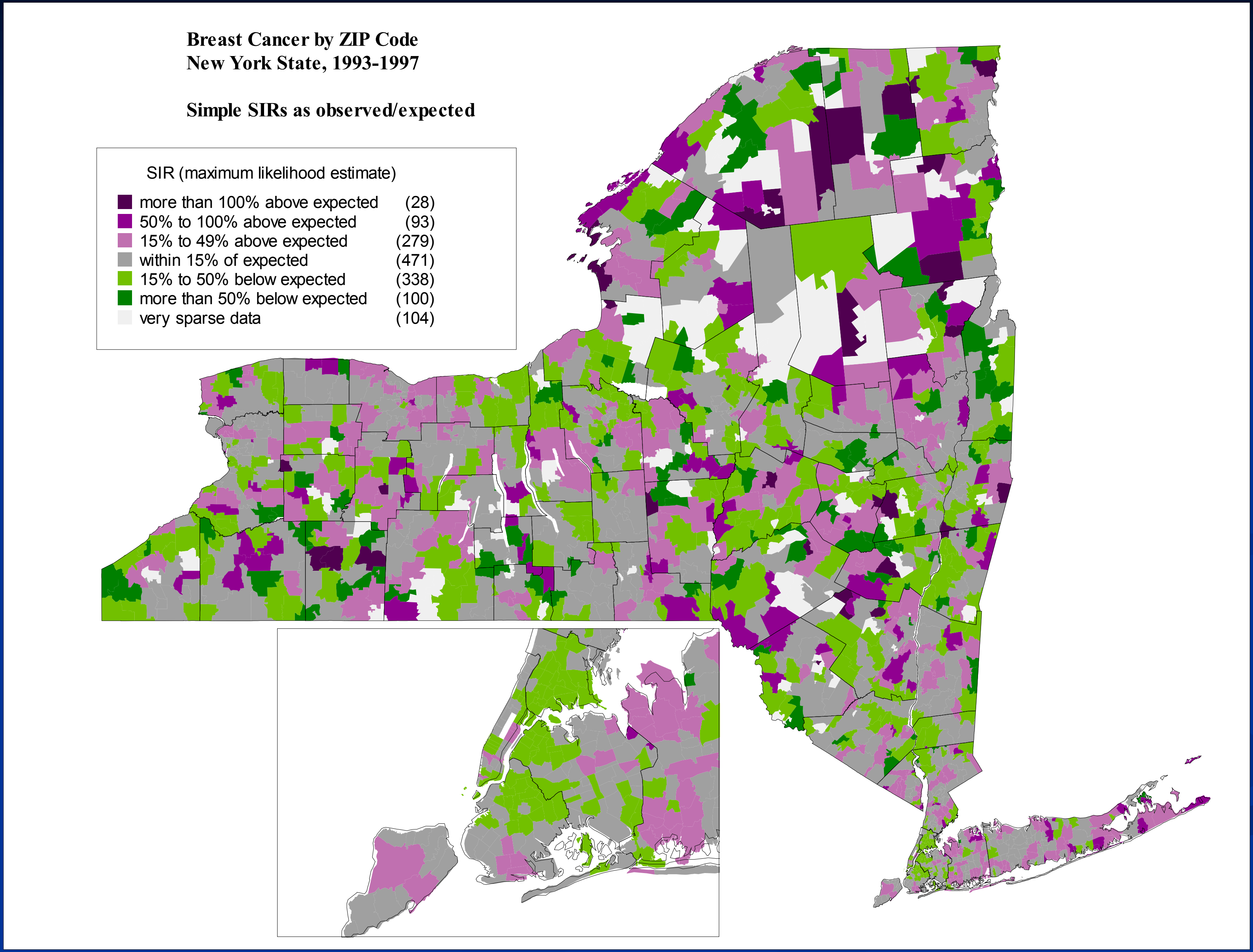
Invited Paper
to appear in
Environmental and Ecological Statistics:
Special Issue on
Multiscale Advanced Raster Map Analysis System

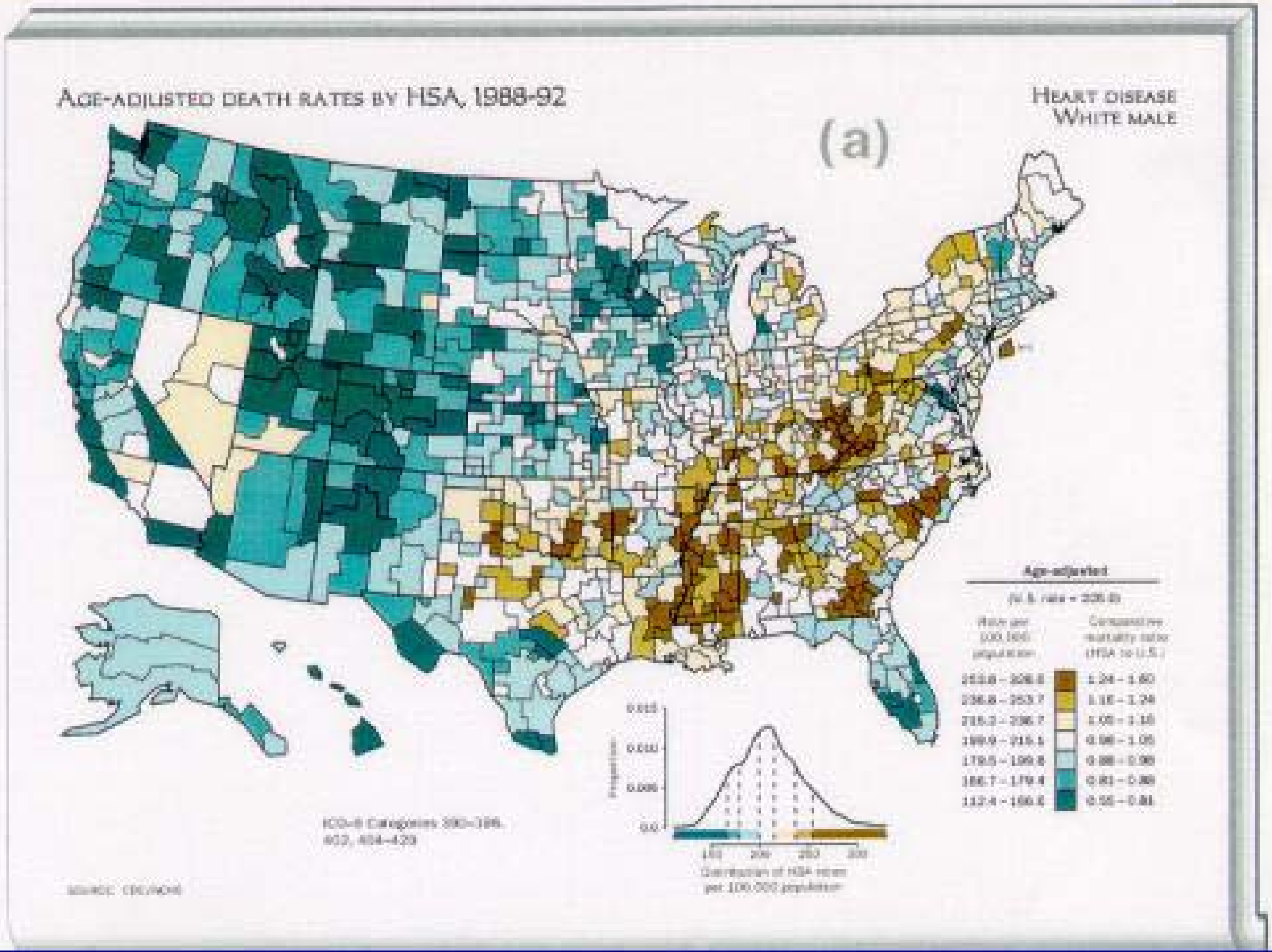
Areas of Application

- Biodiversity, species-rich, and species-poor areas
- Water resources at watershed scales
- Power lines and their effects
- Networks of water distribution systems, subway systems, and road transport systems
- Urban and regional planning
- Disease epidemiology
- Medical imaging
- Reconnaissance
- Astronomy
- Archaeology

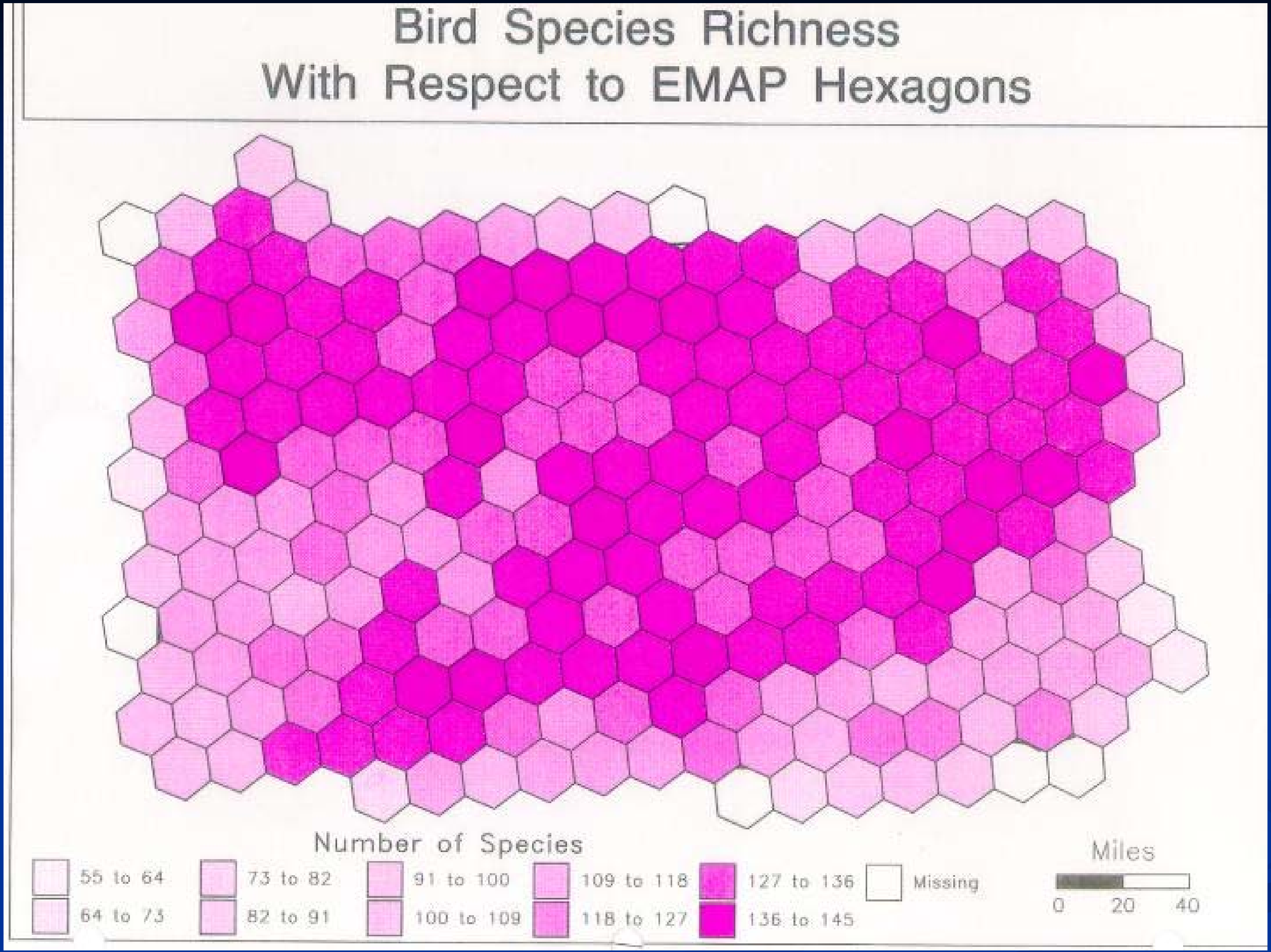
Geographical Surveillance

- Discrete response
- Hotspot detection and tree-structured SatScan
- Hotspot delineation and hot-spot rating
- Multiple hotspot detection and delineation
- Hotspot prioritization and poset ranking
- Space-time detection and early warning
- Continuous response
- User friendly software and downloadable website

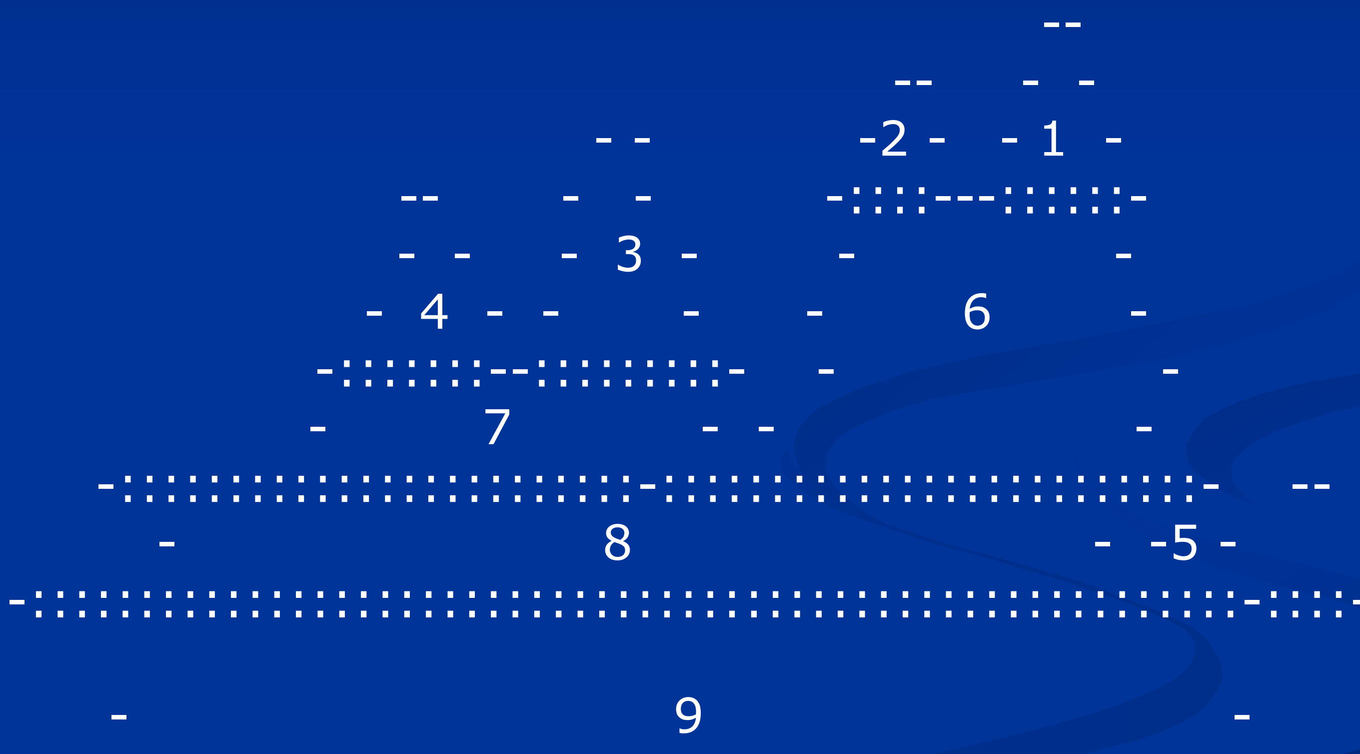

Breast Cancer by ZIP Code
New York State, 1993-1997
Simple SIRs as observed/expected
SIR (maximum likelihood estimate)
more than 100% above expected (28)
50% to 100% above expected (93)
15% to 49% above expected (279)
within 15% of expected (471)
15% to 50% below expected (338)
more than 50% below expected (100)
very sparse data (104)

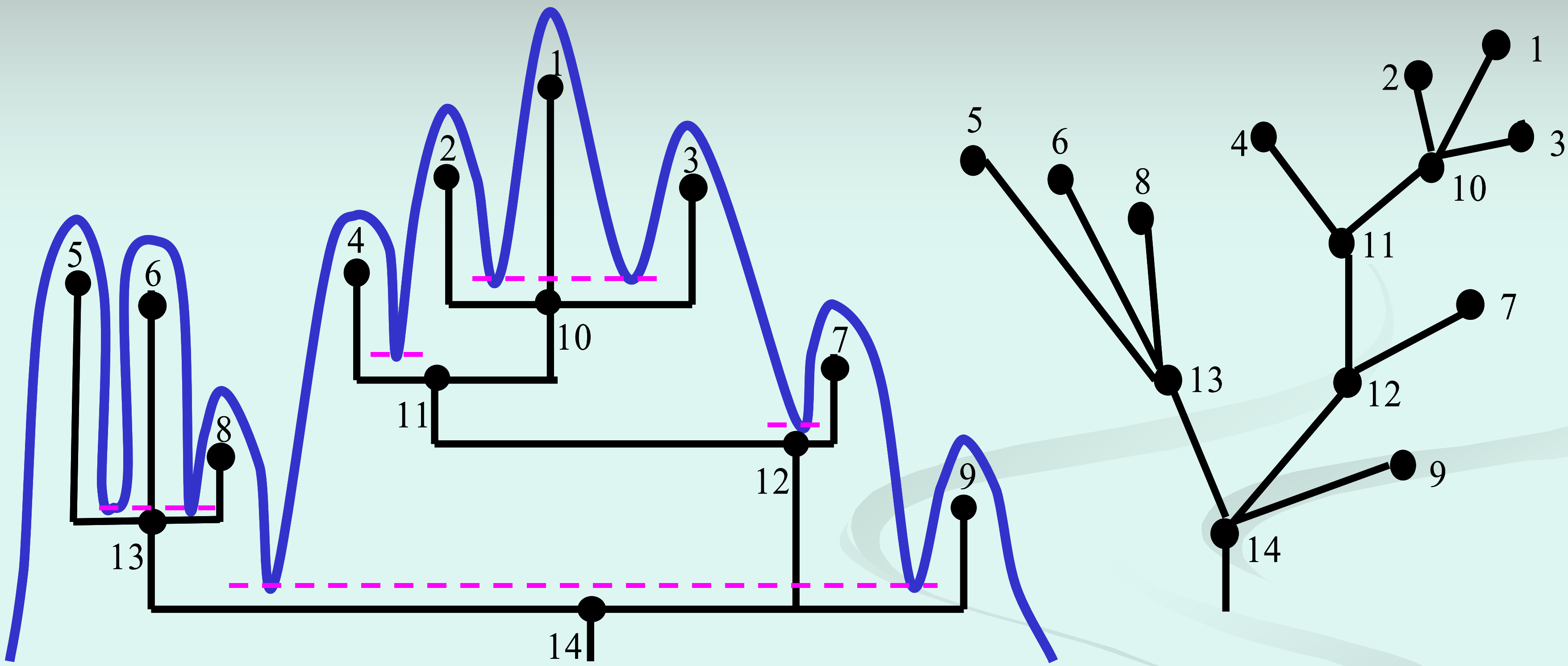


Age-adjusted death rates by HSA, 1988-92
Heart disease
White male
(a)
Age-adjusted

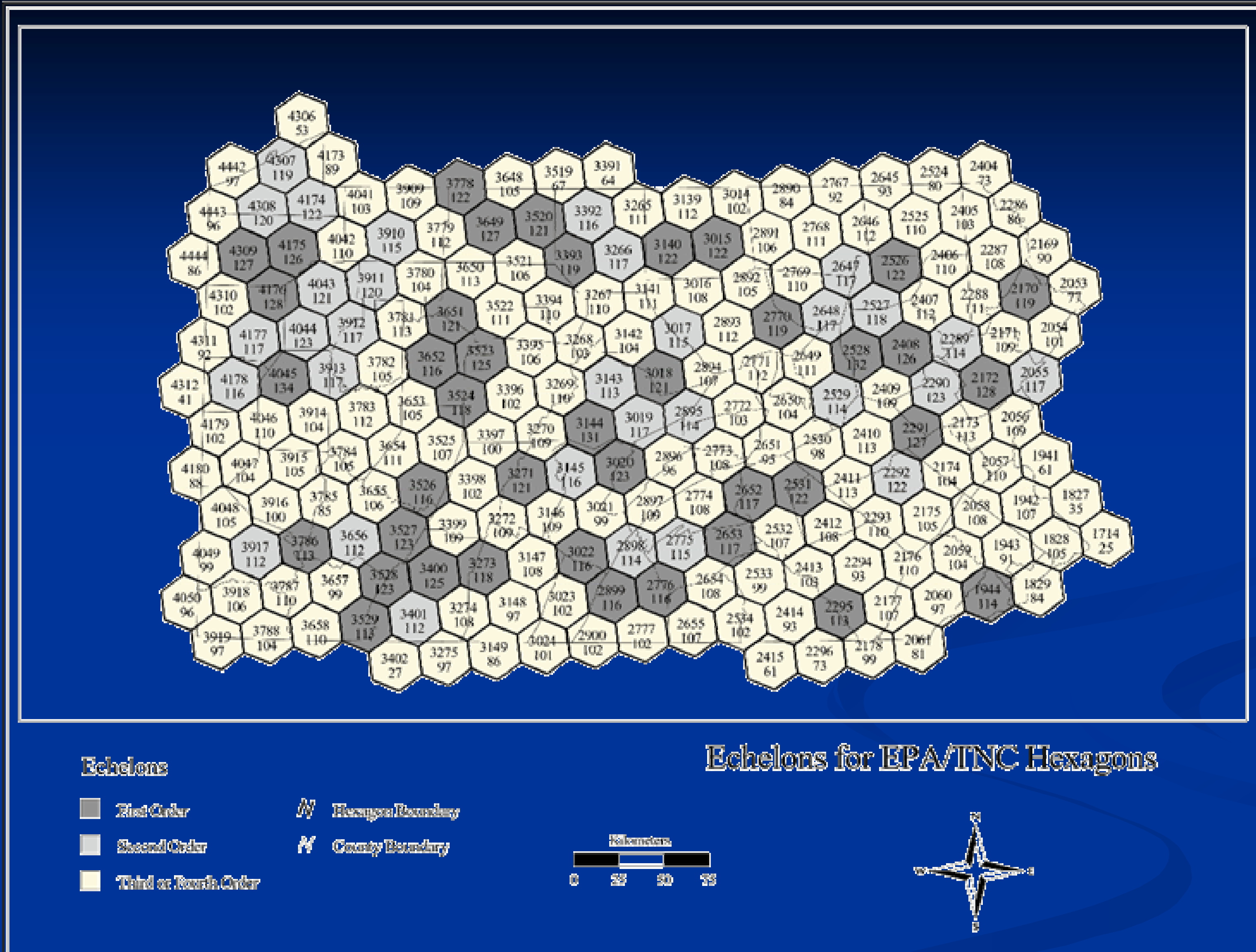


Bird richness in the hexagons.

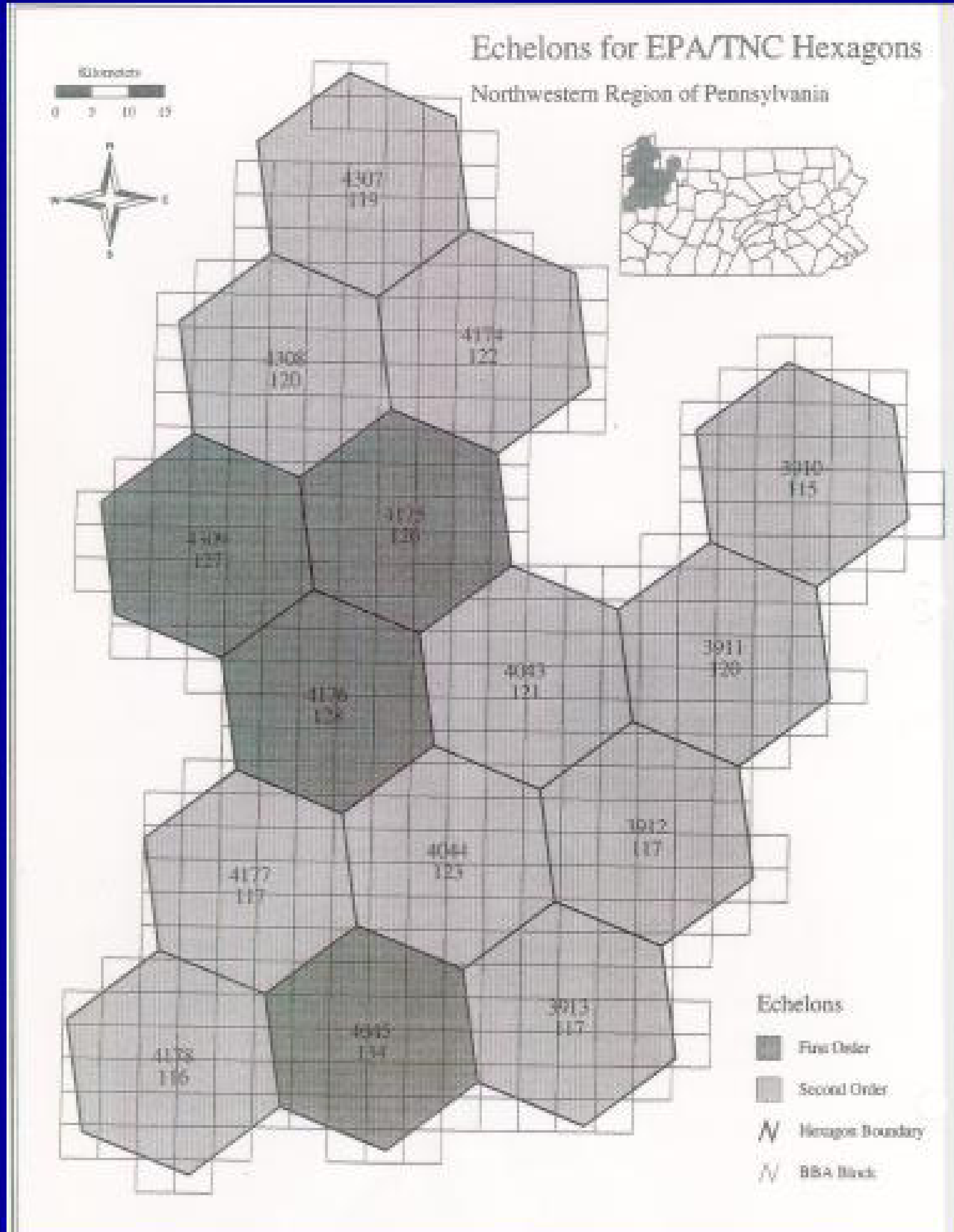

2
1
3
4
6
7
8
5
9



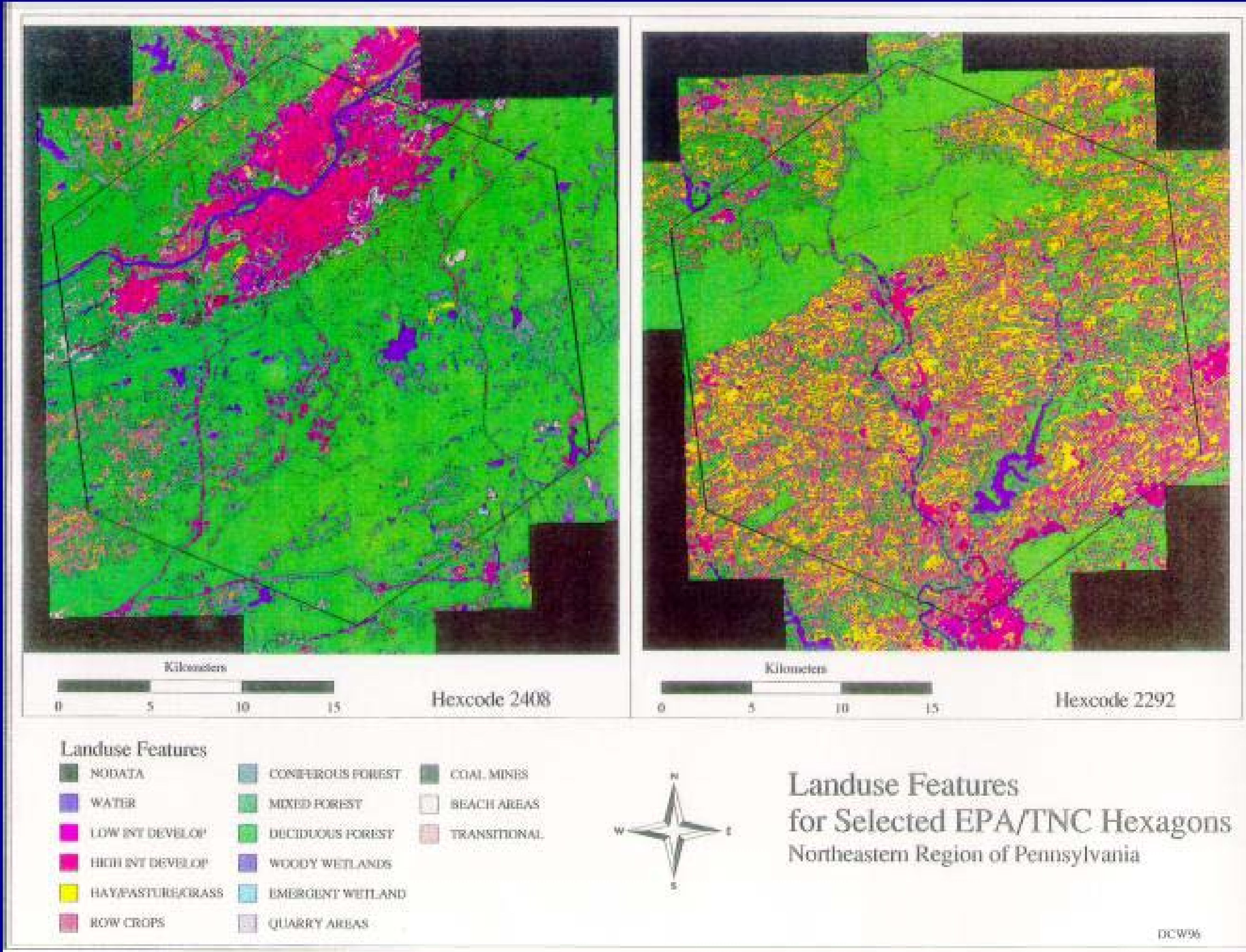
Echelon decomposition of a surface (left) and associated echelon tree (right).



Statewide echelon map based on EMAP hexagons. The 4-digit number in each hexagon is the EPA-EMAP identifier, while the number below is species richness.

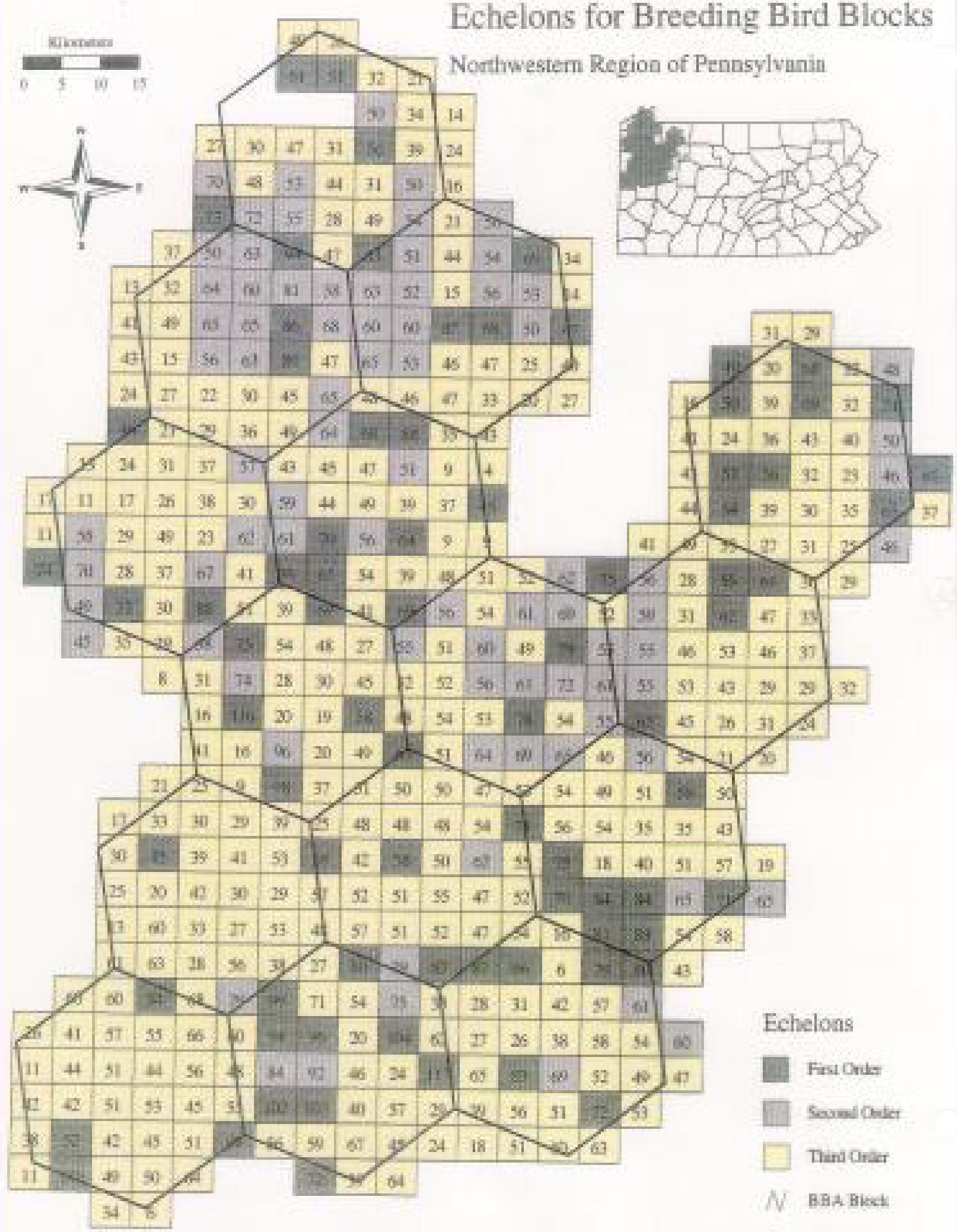


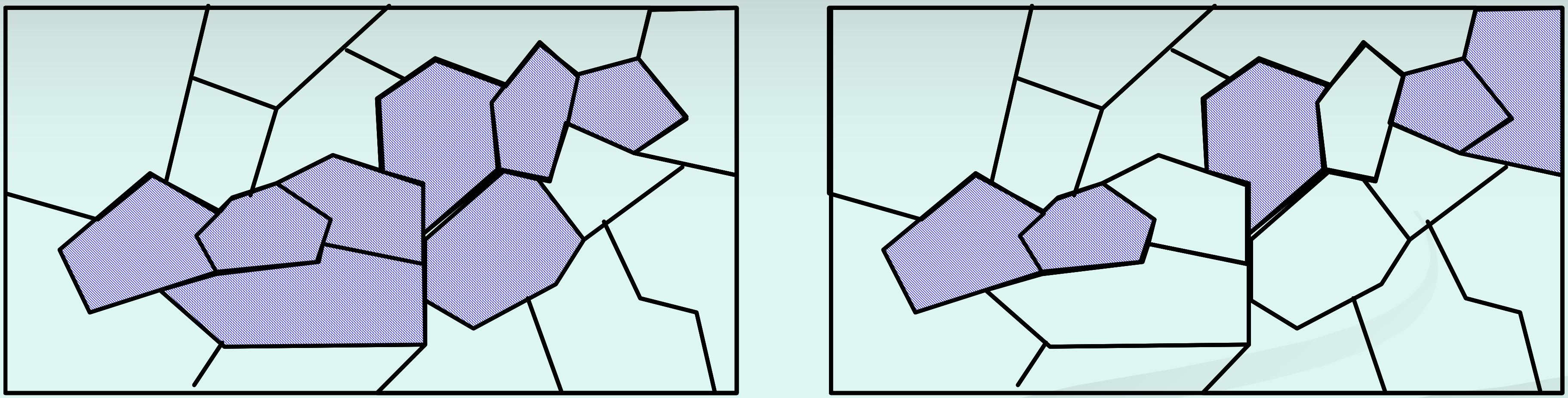
Echelons for EPA/TNC Hexagons
Northwestern Region of Pennsylvania
Kilometers
0
5
10
15
4307
4308
120
122
127
4083
121
3911
120
3910
115
4044
125
3912
117
4177
117
3913
117
134
116
Echelons
First Order
Second Order
Hexagon Boundary
BBA Block


Kilometers
0 5 10 15
Hexcode 2408
Kilometers
0 5 10 15
Hexcode 2292
Landuse Features
NODATA
WATER
LOW INT DEVELOP
HIGH INT DEVELOP
HAY/PASTURE/GRASS
ROW CROPS
CONIFEROUS FOREST
MIXED FOREST
DECIDUOUS FOREST
WOODY WETLANDS
EMERGENT WETLAND
QUARRY AREAS
COAL MINES
BEACH AREAS
TRANSITIONAL
N
W
E
S
Landuse Features
for Selected EPA/TNC Hexagons
Northeastern Region of Pennsylvania
DCW96

Echelons for Breeding Bird Blocks

Northwestern Region of Pennsylvania





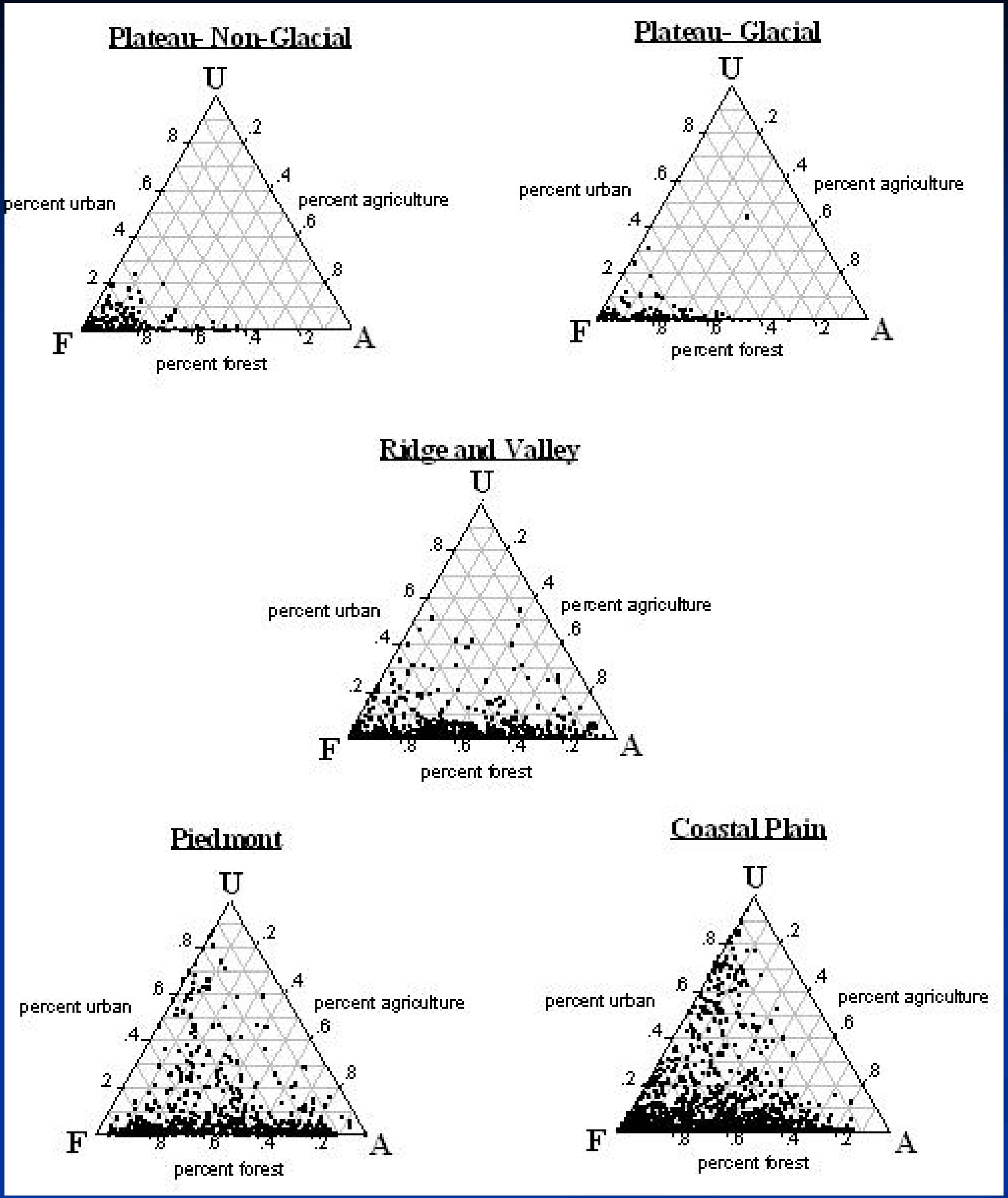
A tessellated region. The collection of shaded cells in the left-hand diagram is connected and, therefore, constitutes a zone in . The collection on the right is not connected.

SATScan Setup – 6

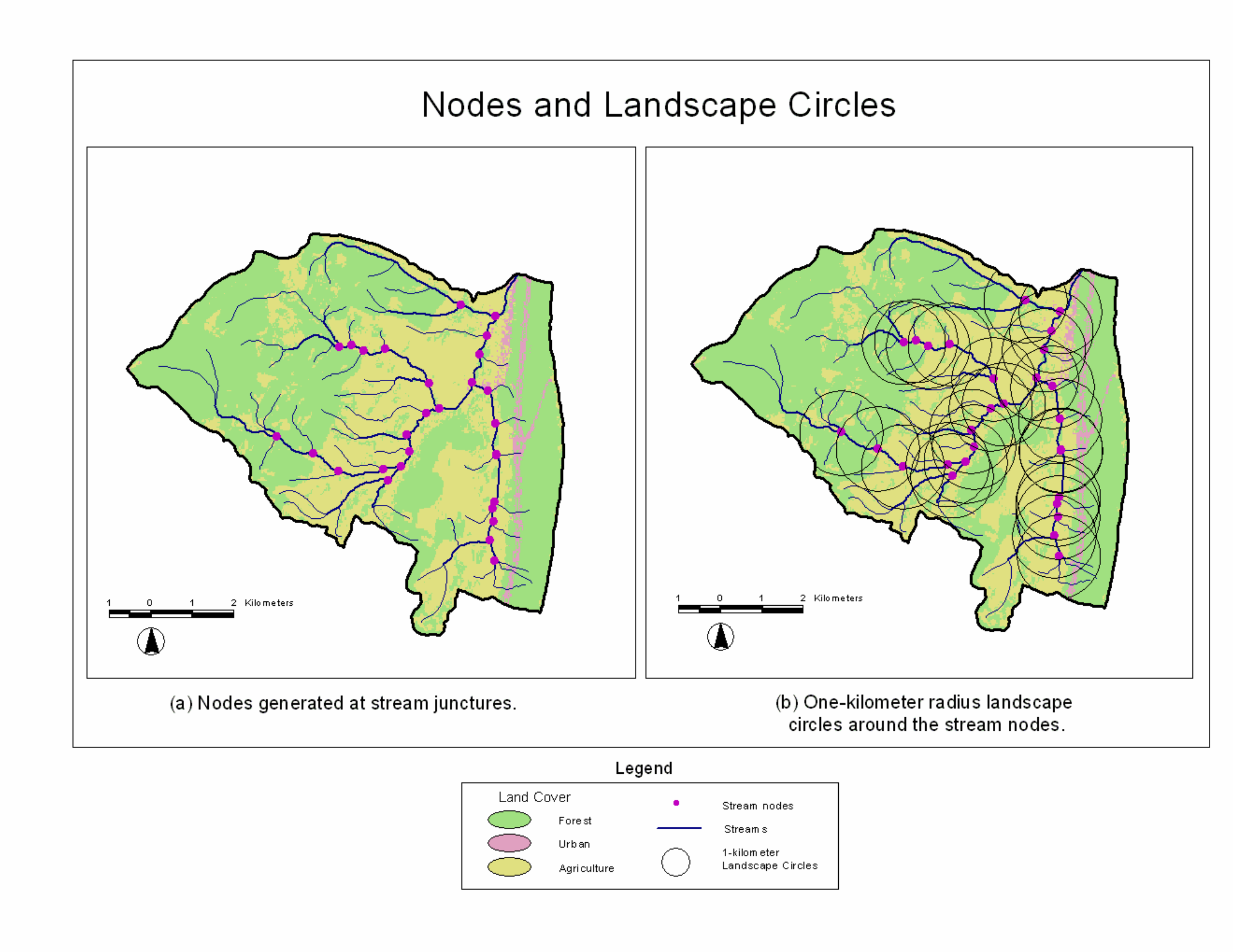
- Goal: Identify geographic zone(s) in which a response is significantly elevated relative to the rest of a region
- A list of candidate zones *Z* is specified *a priori*.
  - This list becomes part of the parameter space and the zone must be estimated from within this list.
  - Each candidate zone should generally be spatially connected, e.g., a union of contiguous spatial units or cells.
  - Longer lists of candidate zones are usually preferable
  - Expanding circles about specified centers are a common method of generating the list

Atlantic Slope	NLCD – Anderson 1	NLCD – Anderson 2
Water	Water	11 Open Water
Not Applicable		12 Perennial Ice/Snow
Suburban	Developed	21 Low Intensity Residential
Urban		22 High Intensity Residential
Urban		23 Commercial/Industrial/Transportation
Rock	Barren	31 Bare Rock/Sand/Clay
Rock		32 Quarries/Strip Mines/Gravel Pits
Transitional		33 Transitional
Forest	Forest Upland	41 Deciduous Forest
Forest		42 Evergreen Forest
Forest		43 Mixed Forest
Forest	Shrubland	51 Shrubland
Row Crops	Non-natural Woody	61 Orchards/Vineyards/Other
N-A	Herbaceous Upland	71 Grasslands/Herbaceous
Pasture	Herbaceous Planted/Cultivated	81 Pasture/Hay
Row Crops		82 Row Crops
N-A		83 Small Grains
N-A		84 Fallow
Suburban		85 Urban/Recreational Grasses
Forest	Wetlands	91 Woody Wetlands
Emergent Wetlands		92 Emergent Herbaceous Wetlands

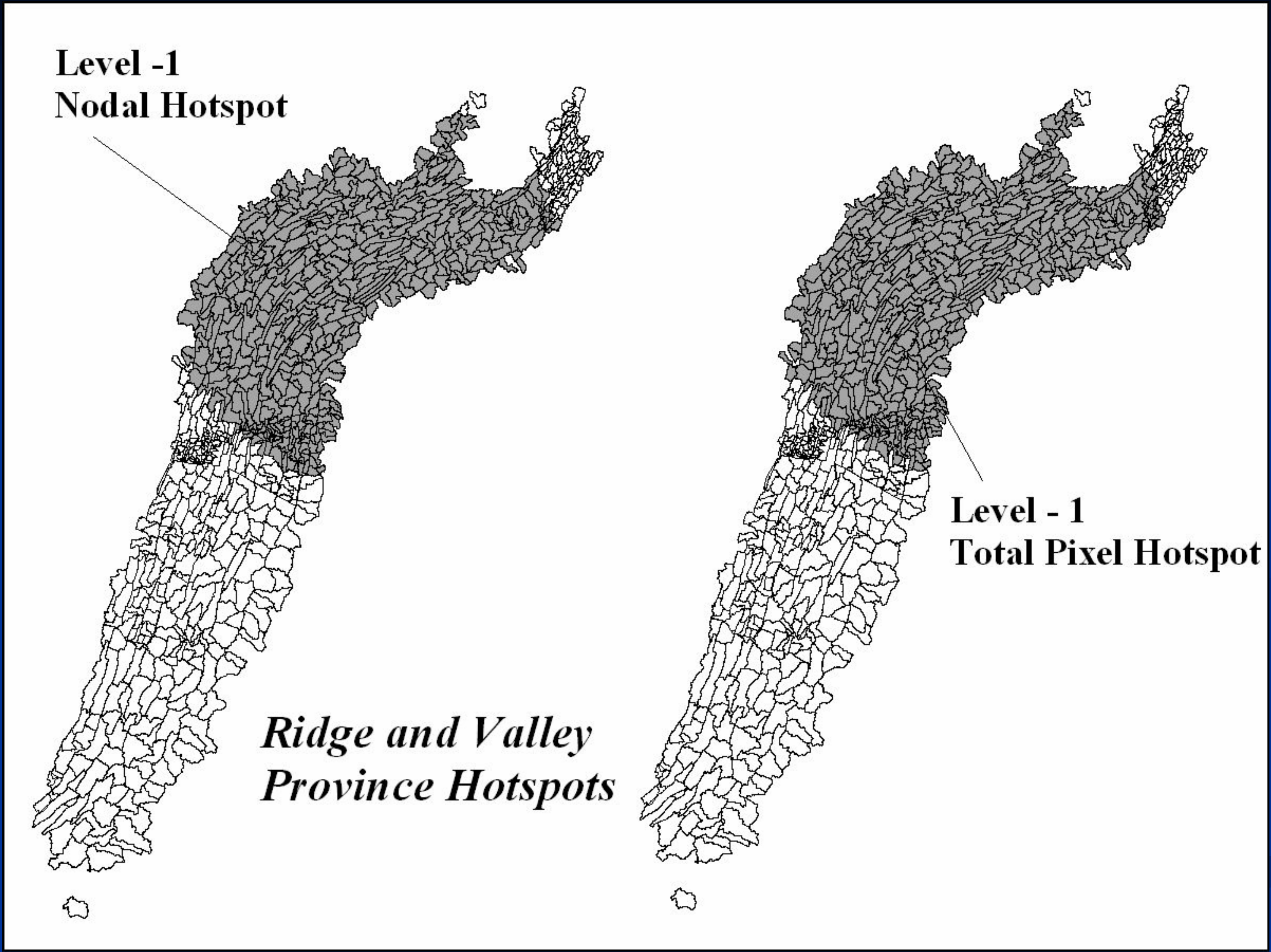
Land cover/land use classes utilized in the Atlantic Slope project, and their corresponding NLCD classes.



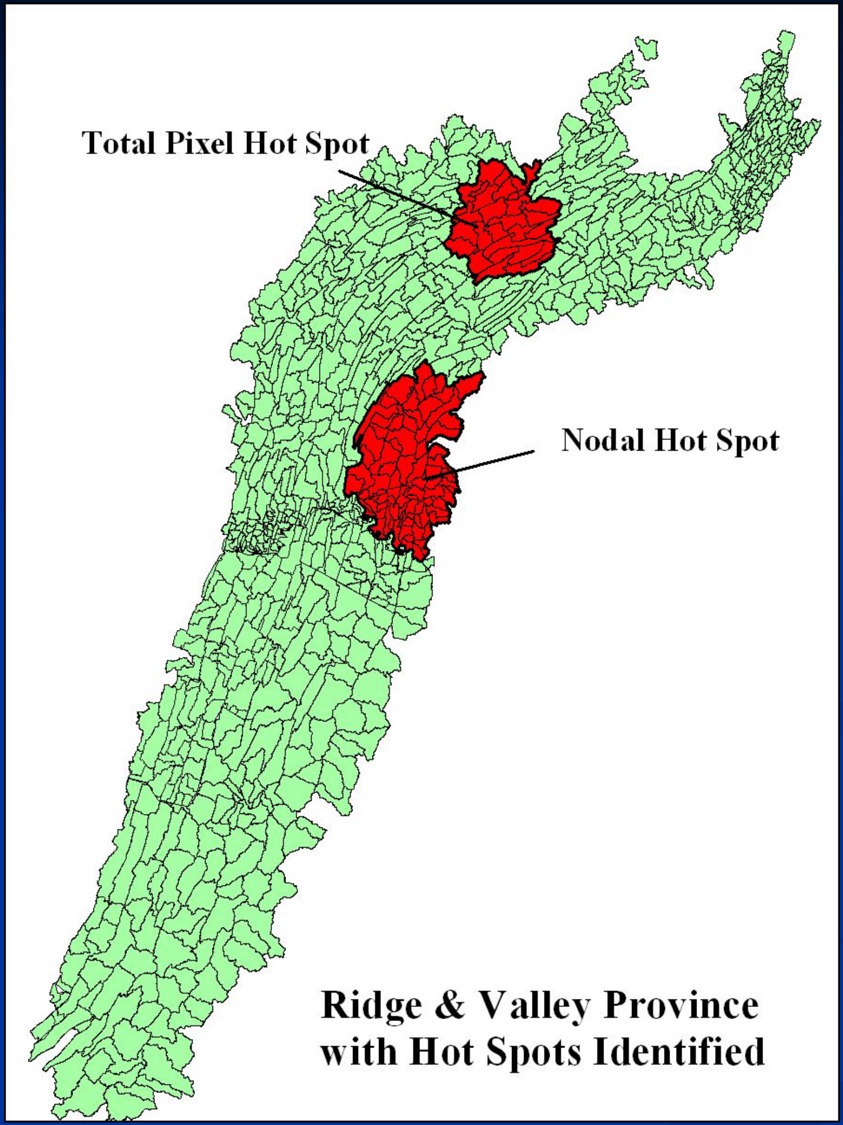
Ternary plots of watersheds of Atlantic Slope by physiographic province.



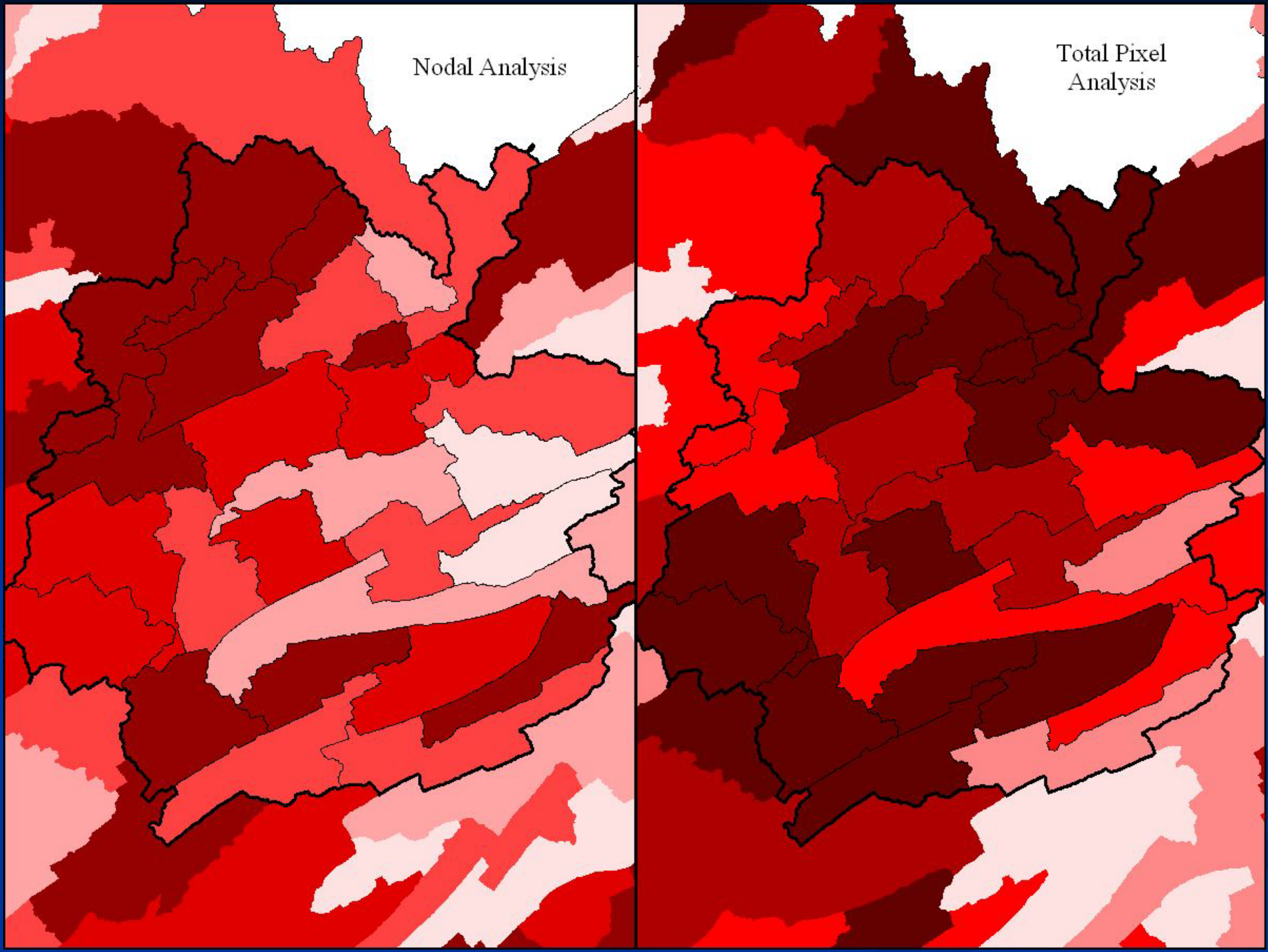
Nodes and landscape circles.



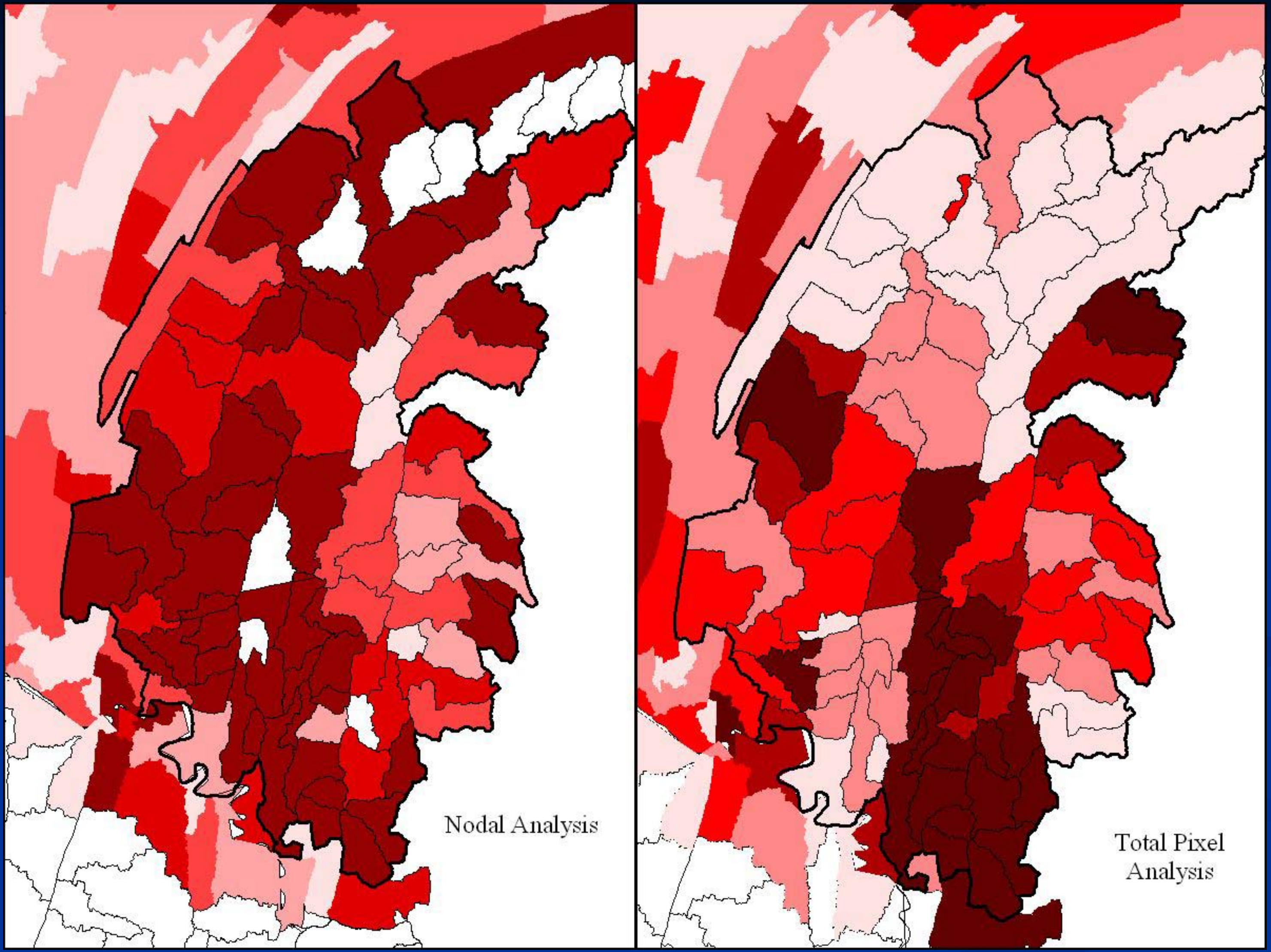
Ridge and Valley Hotspots



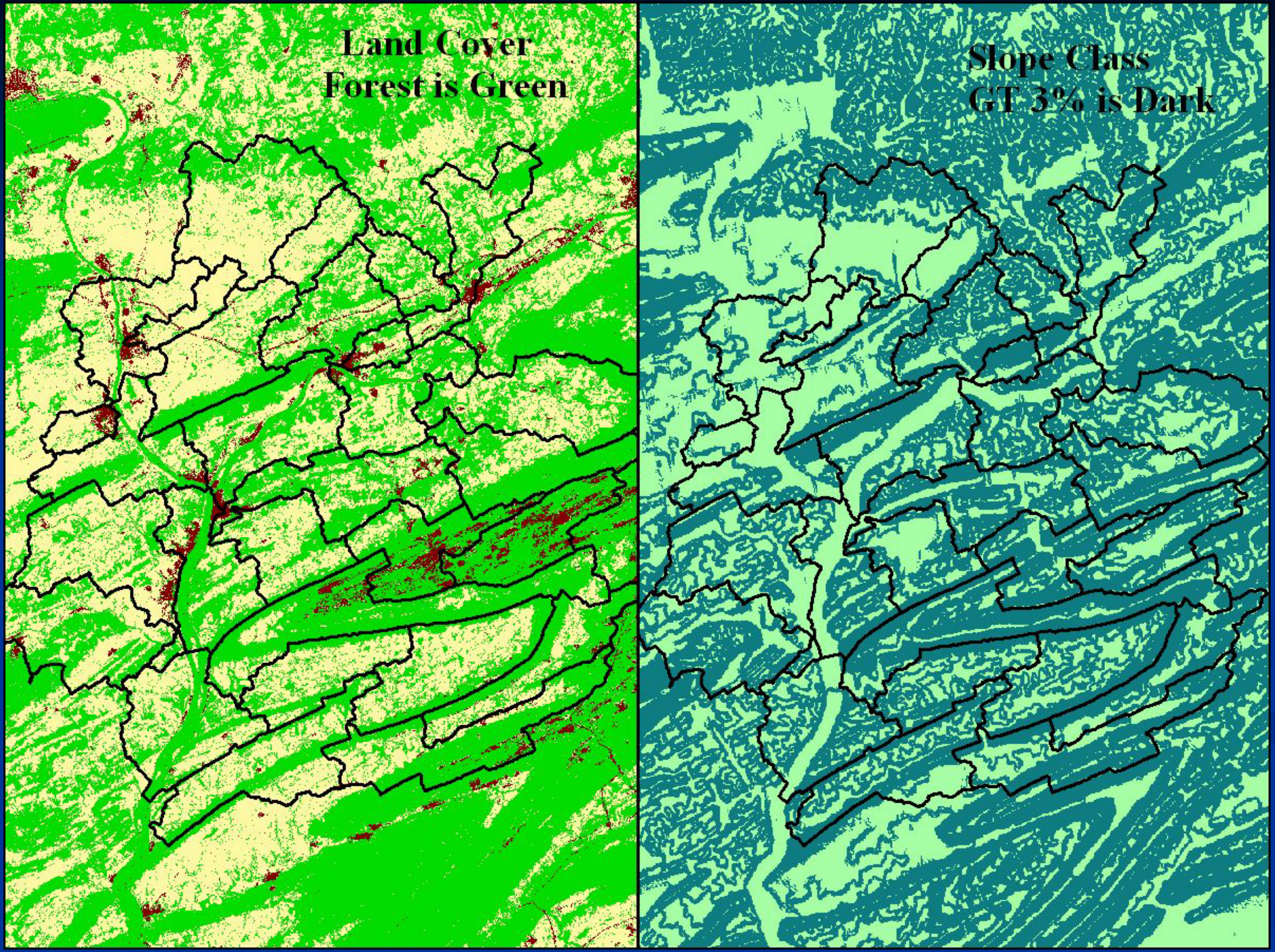
Ridge and Valley hotspots within hotspots.



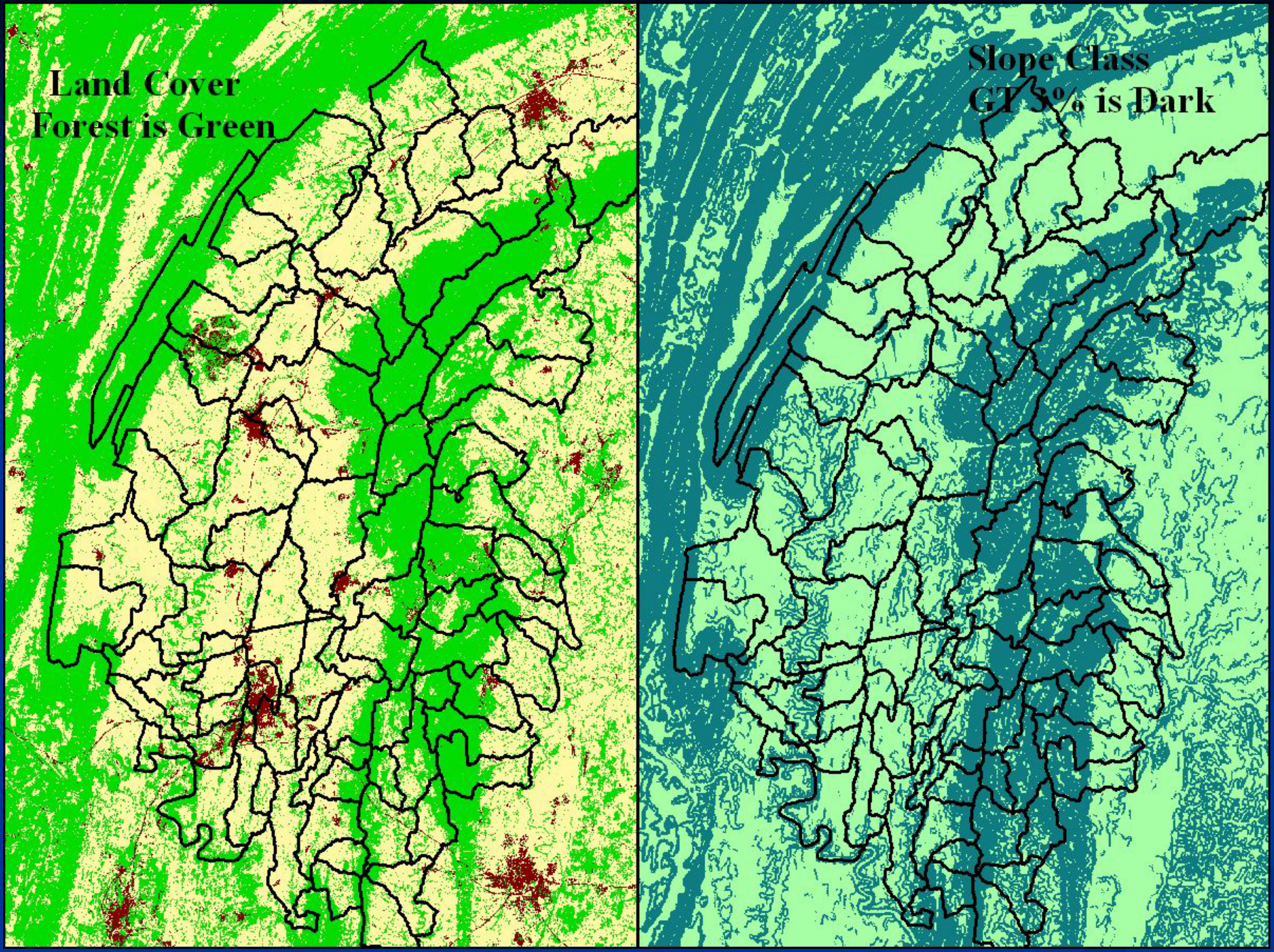
Shaded ordinal view of total pixel hotspot.



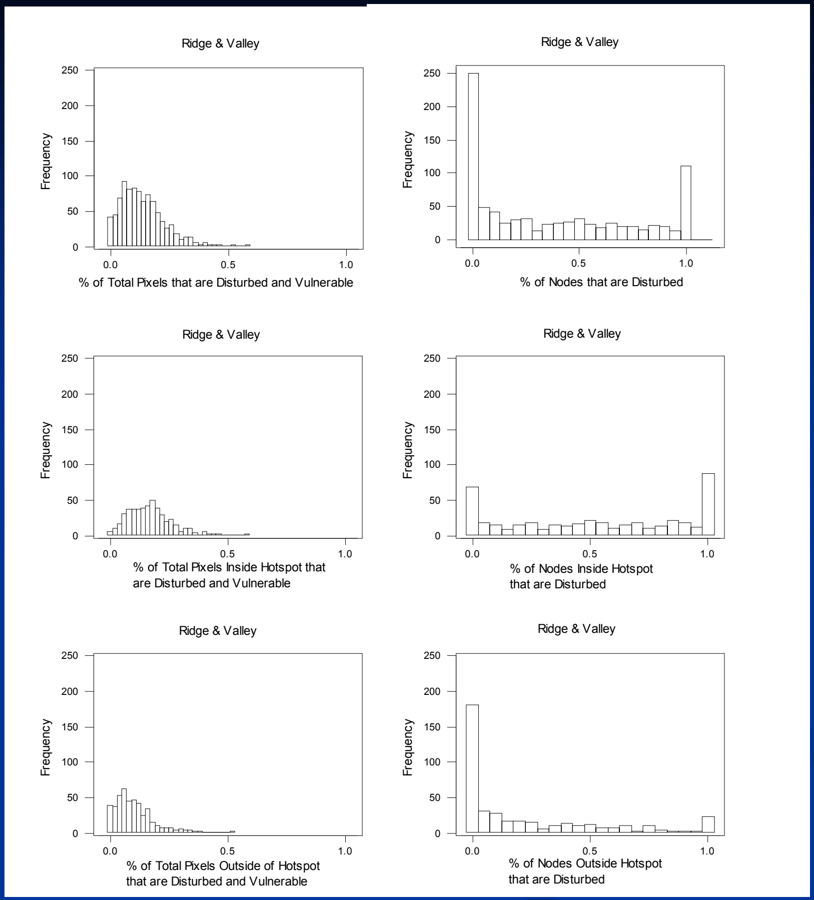
Shaded ordinal view of nodal hotspot.



Land cover and slope class for the nodal hotspot.

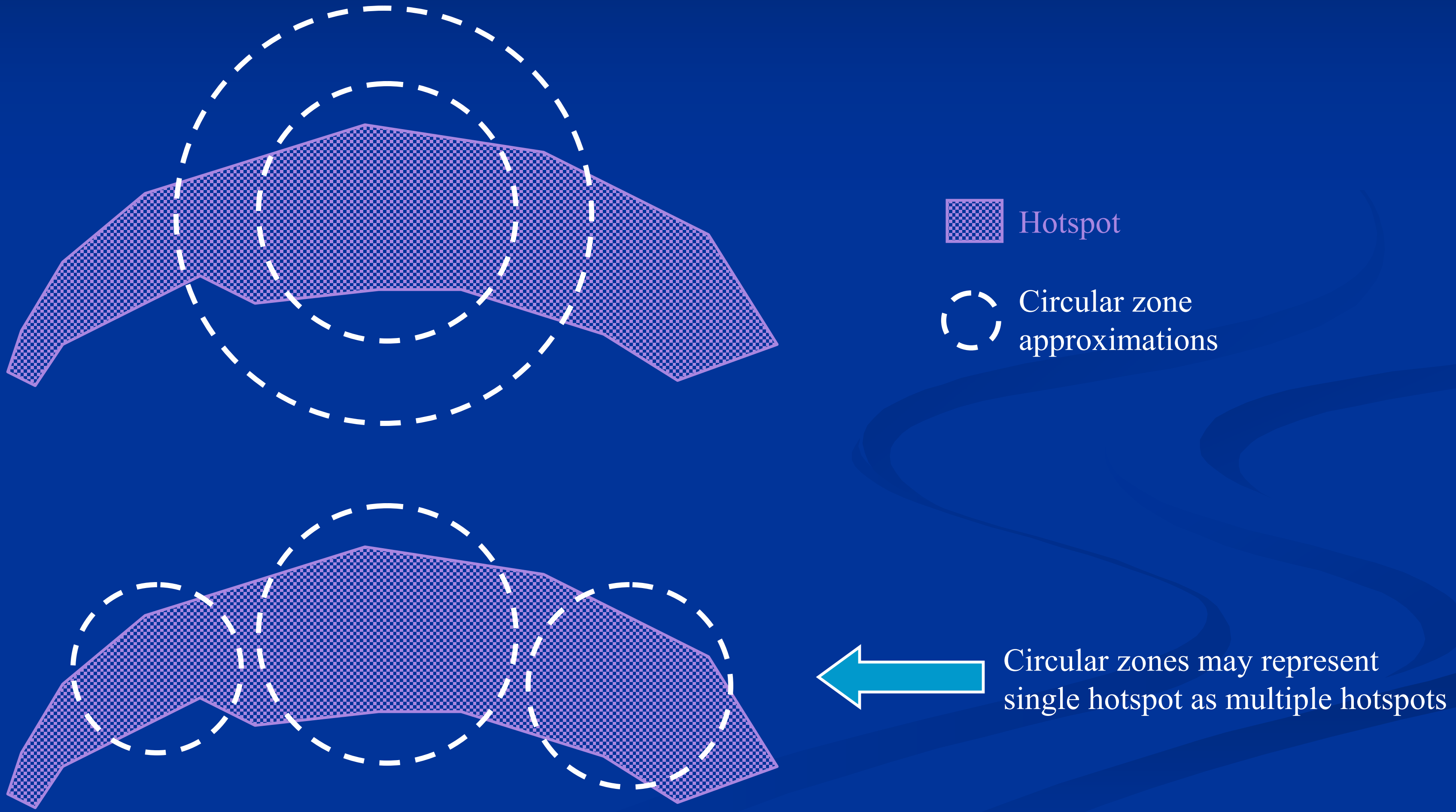


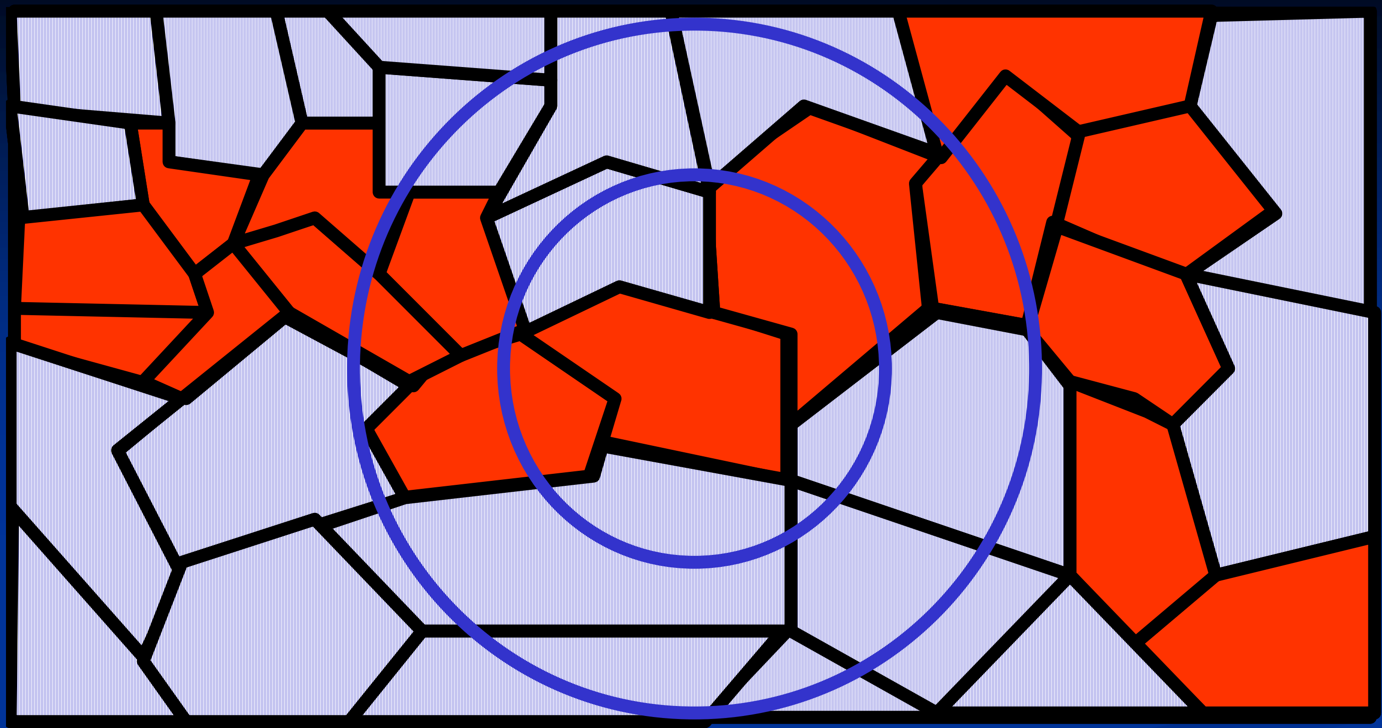
Land cover and slope class for the total pixel hotspot.



Frequency histograms showing how the proportions of affected pixels/nodes vary across the watersheds. The diagrams on the left refer to the total pixels analysis, where affected pixels are those that are both disturbed and vulnerable. The diagrams on the right refer to the nodal analysis, where affected nodes are those that are disturbed. The frequency distributions are for all watersheds (top), watersheds in the hotspot (middle) and watersheds outside the hotspot (bottom).

Poor Hotspot Delineation by Circular Zones





Cholera outbreak along a river flood-plain

- Small circles miss much of the outbreak
- Large circles include many unwanted cells

Circular spatial scan statistic zonation.

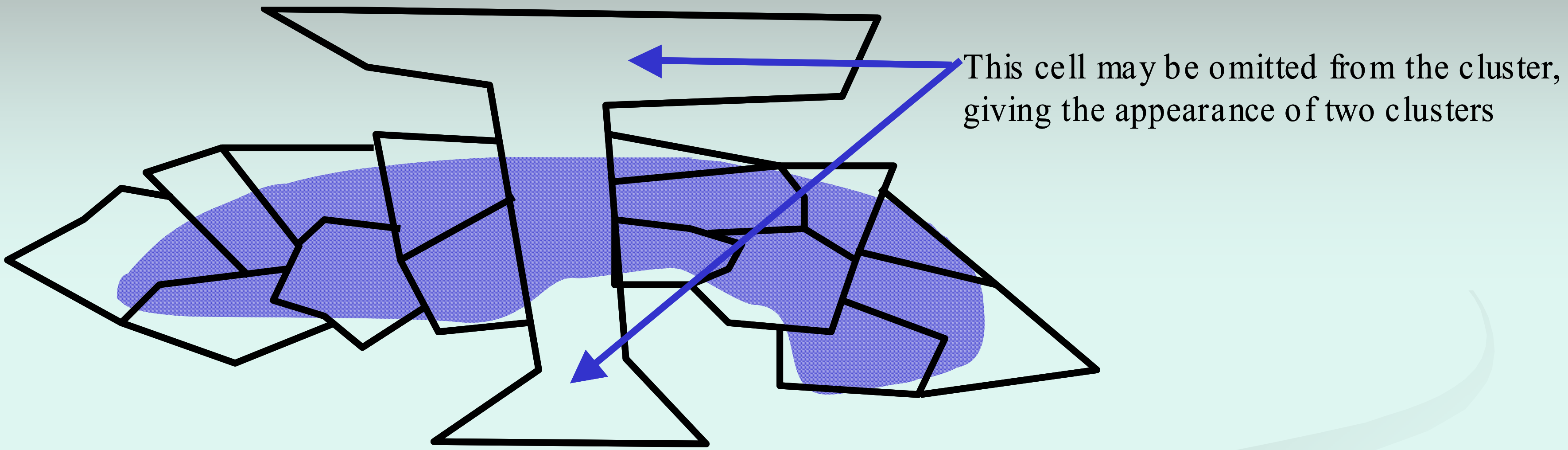
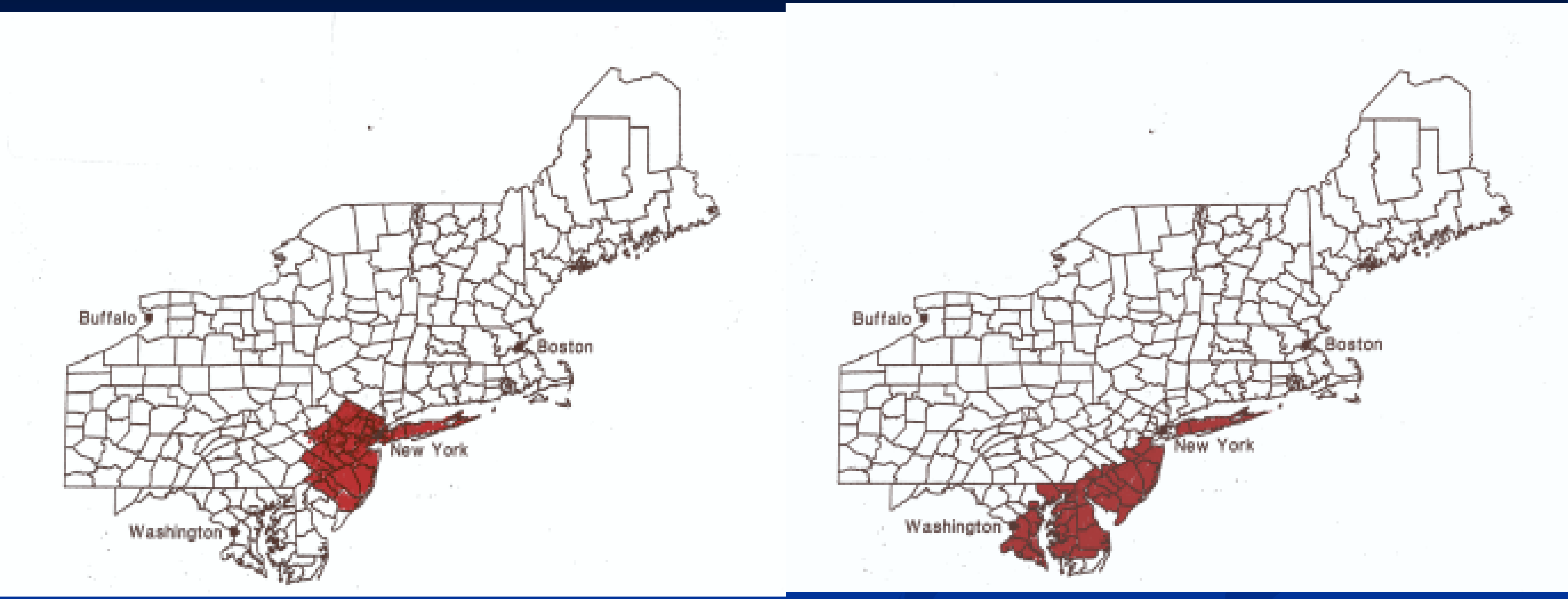
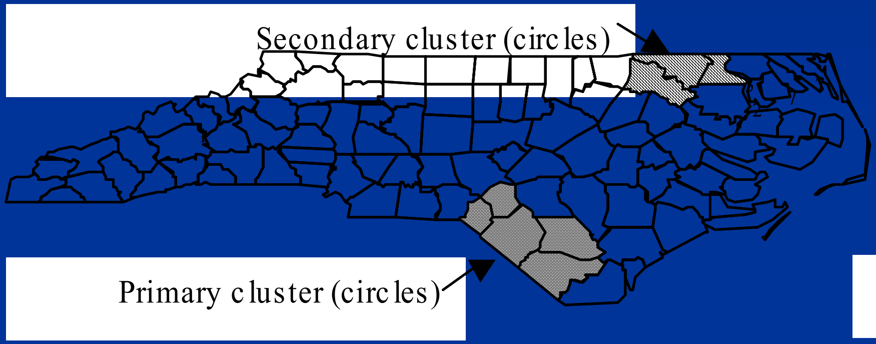
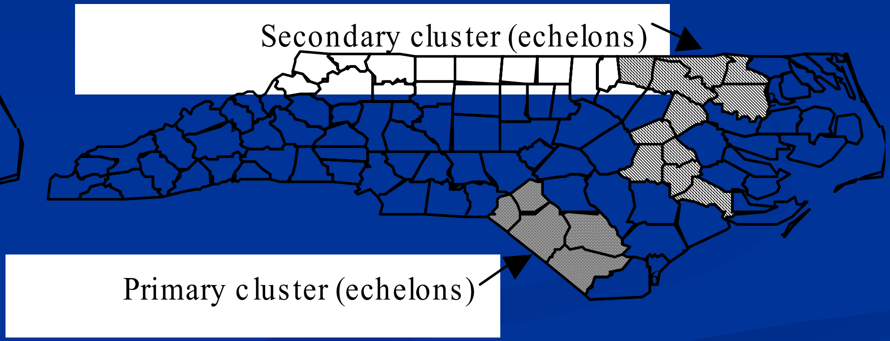


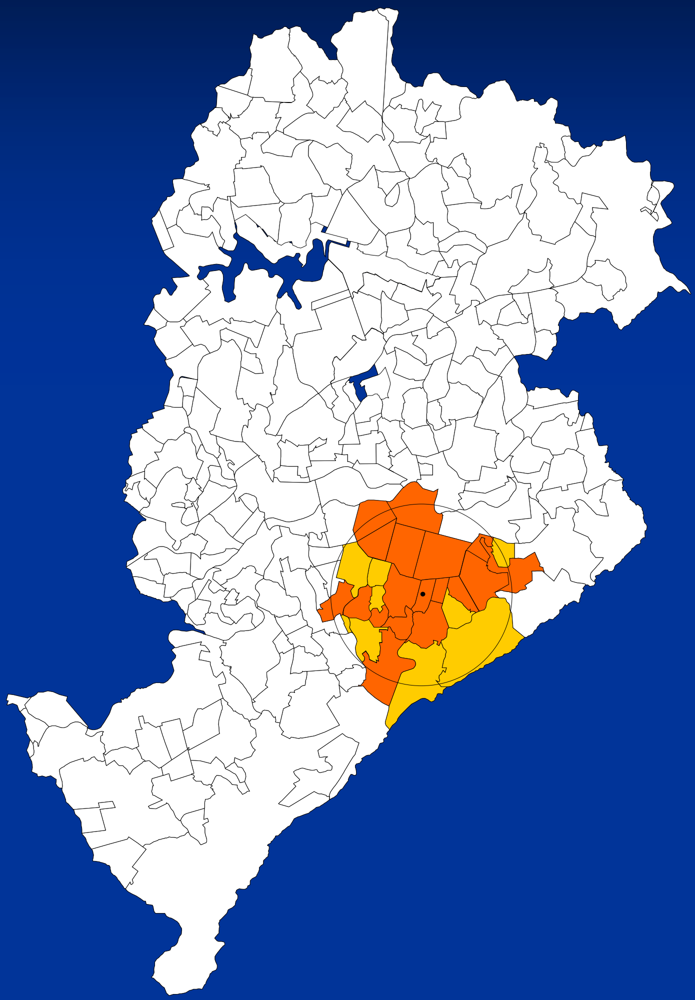
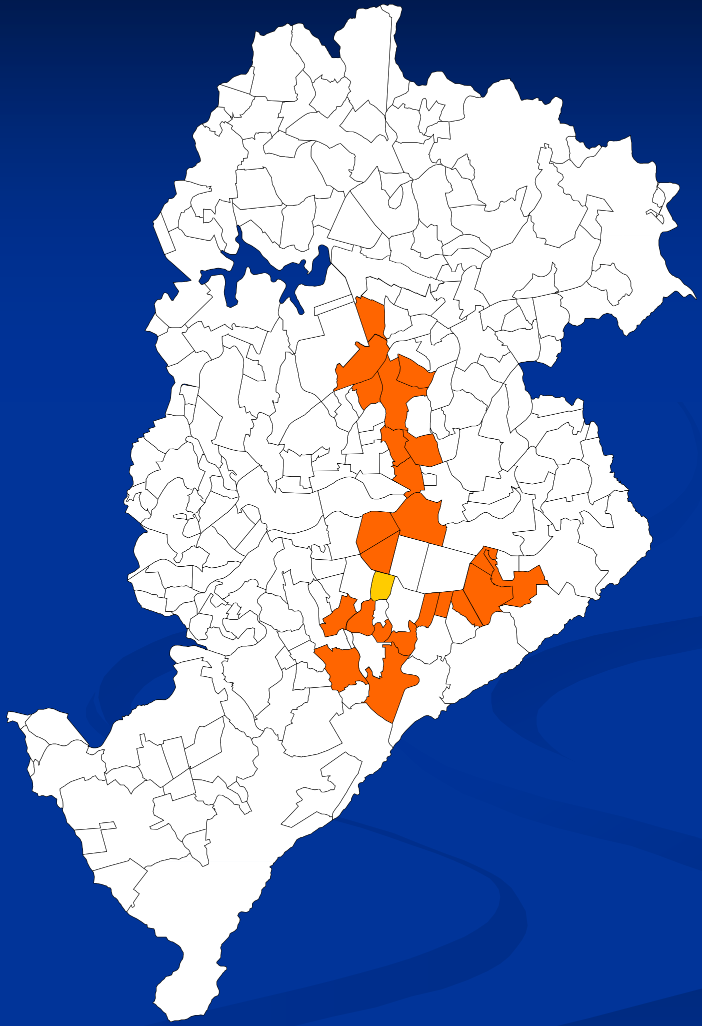
Figure 11. *Actual cluster (shaded) appears as two clusters because a large cell of comparatively low overall incidence rate bisects it. Only pertinent cells are shown.*



The most likely cluster as found by the circular spatial scan statistic (left) and the elliptic spatial scan statistic (right) for an analysis of county-based breast cancer mortality in Northeastern United States, 1988–1992.
For the circular-detected cluster, the relative risk is 1.07 and p=0.0001 (Kulldorff et al, 1997), while the elliptic-detected cluster has a relative risk of 1.08 and p=0.0001. Note that the elliptic-detected cluster is not connected, since the New York City area is not part of the cluster.


Secondary cluster (circles)
Primary cluster (circles)

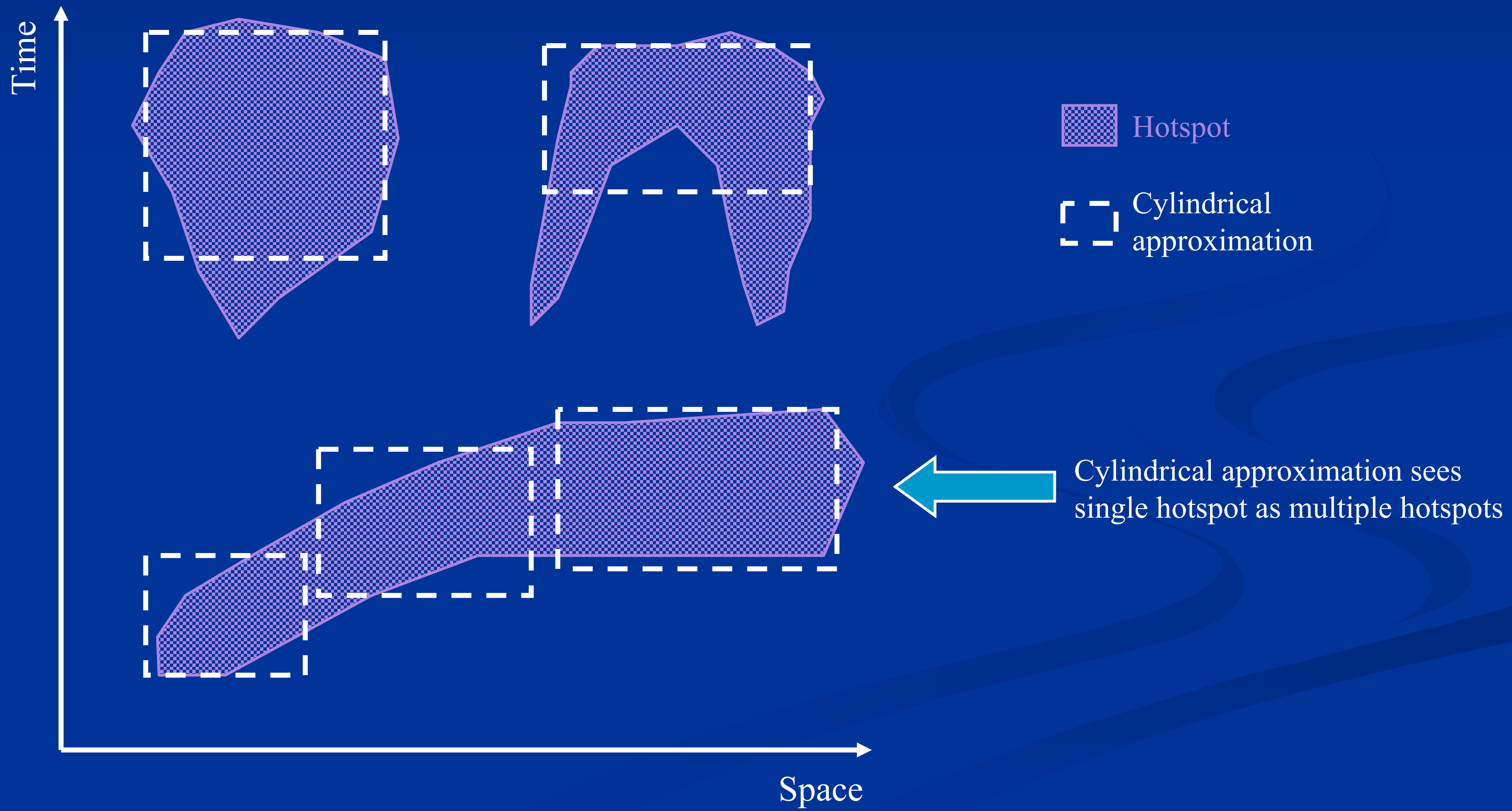

Secondary cluster (echelons)
Primary cluster (echelons)

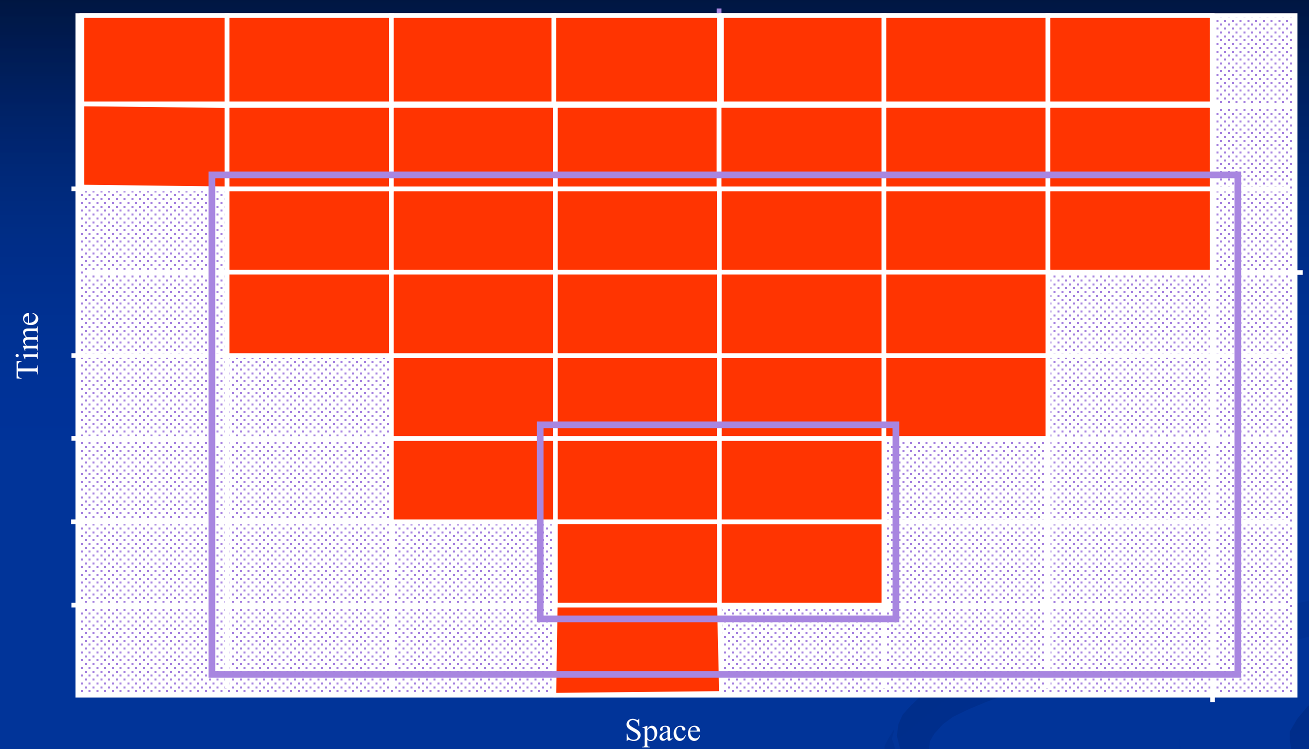



Space-Time Detection and Early Warning-1 Tree-Structured SatScan

- The traditional space-time scan statistic employs cylinders as the candidate zones in the reduced parameter space . In many instances, the cylindrical shape may be a poor approximation to actual space-time hotspots, whereas the ULS approach is able to adapt its shape to the actual hotspot.
- Since the ULS tree is derived from the adjacency matrix, the same software will work once the notion of adjacency has been specified for space-time cells.

Some Space-Time Hotspots and Their Cylindrical Approximations



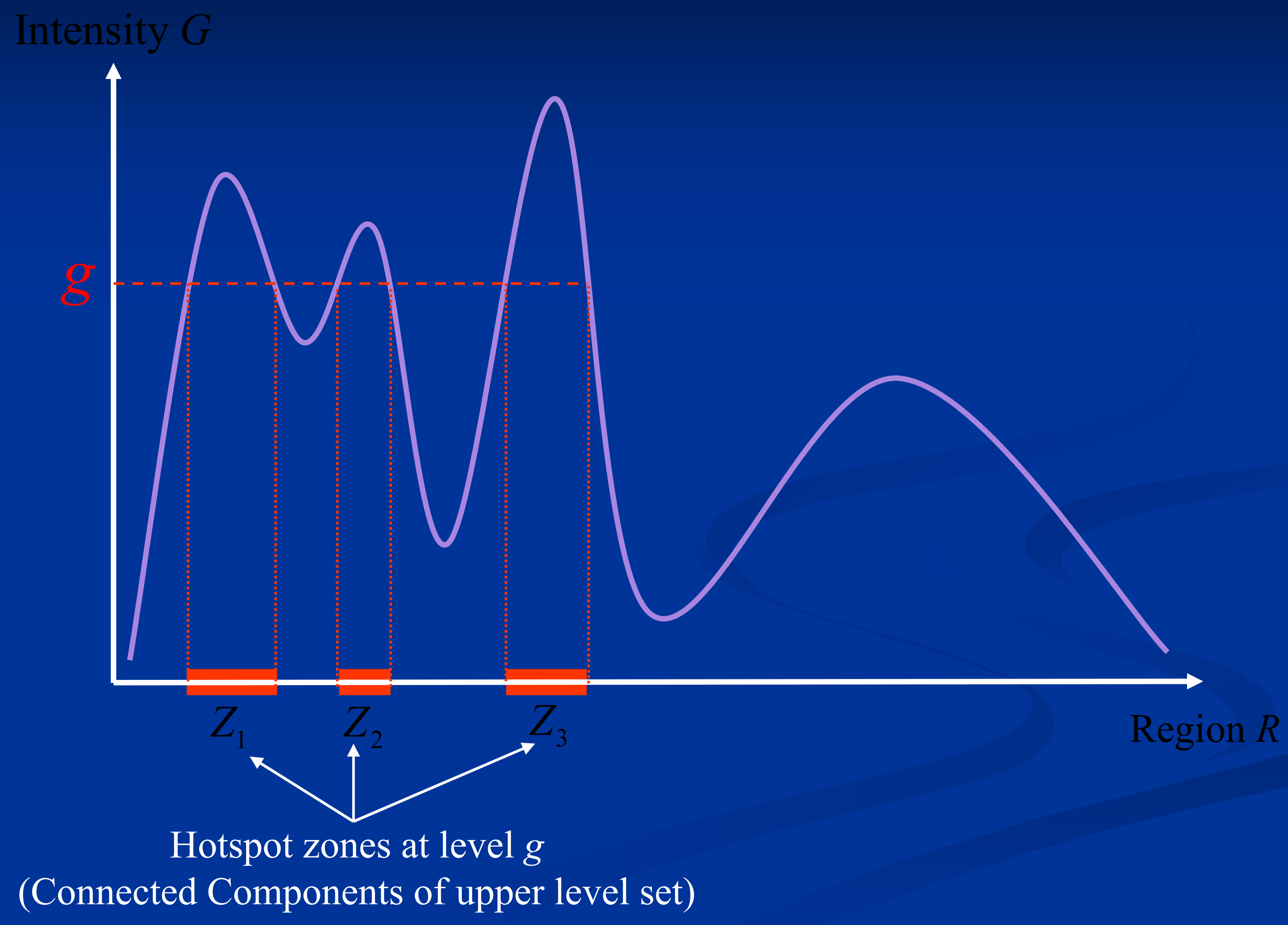


Outbreak expanding in time

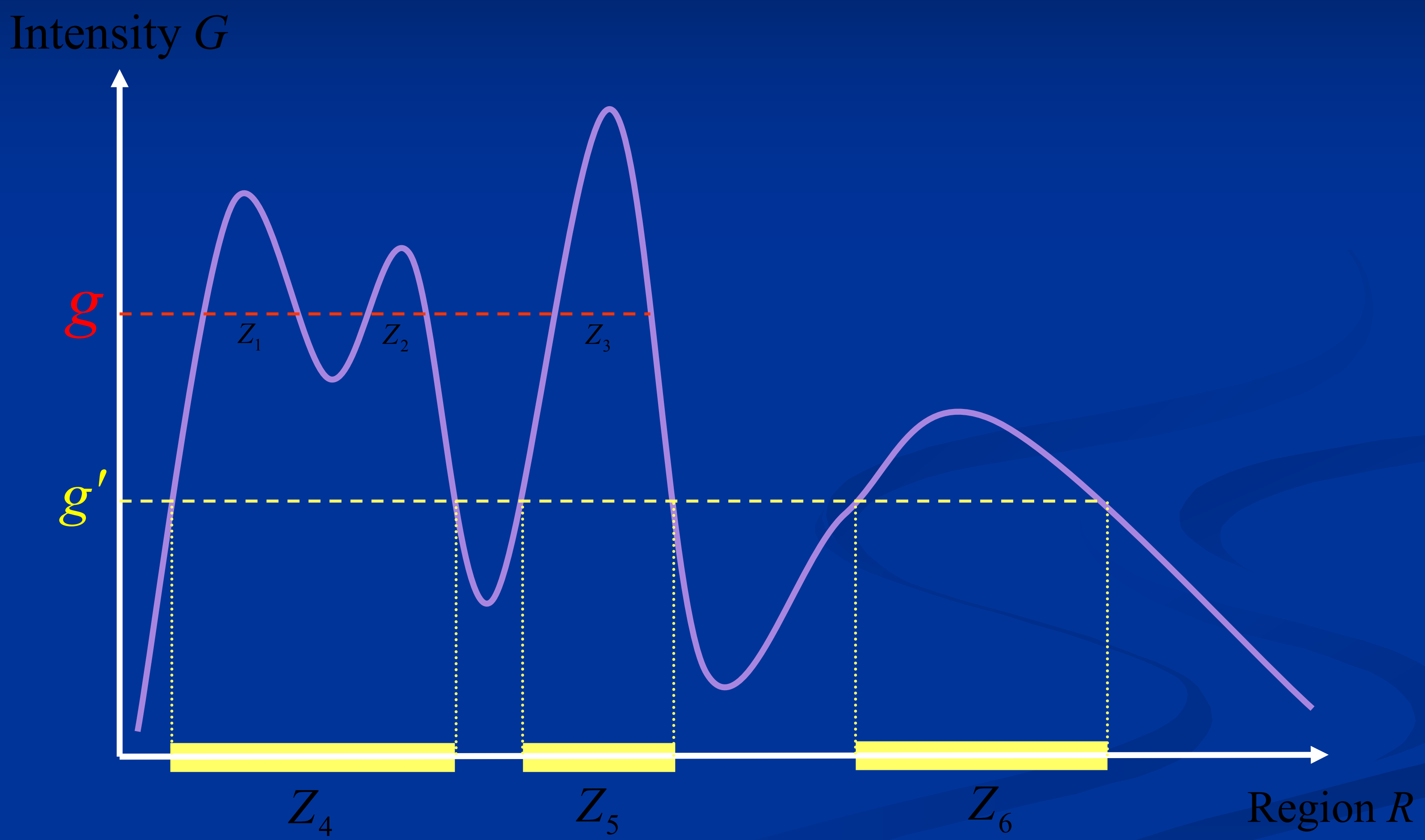
•Small cylinders miss much of the outbreak
•Large cylinders include many unwanted cells

Cylindrical space-time scan statistic zonation.

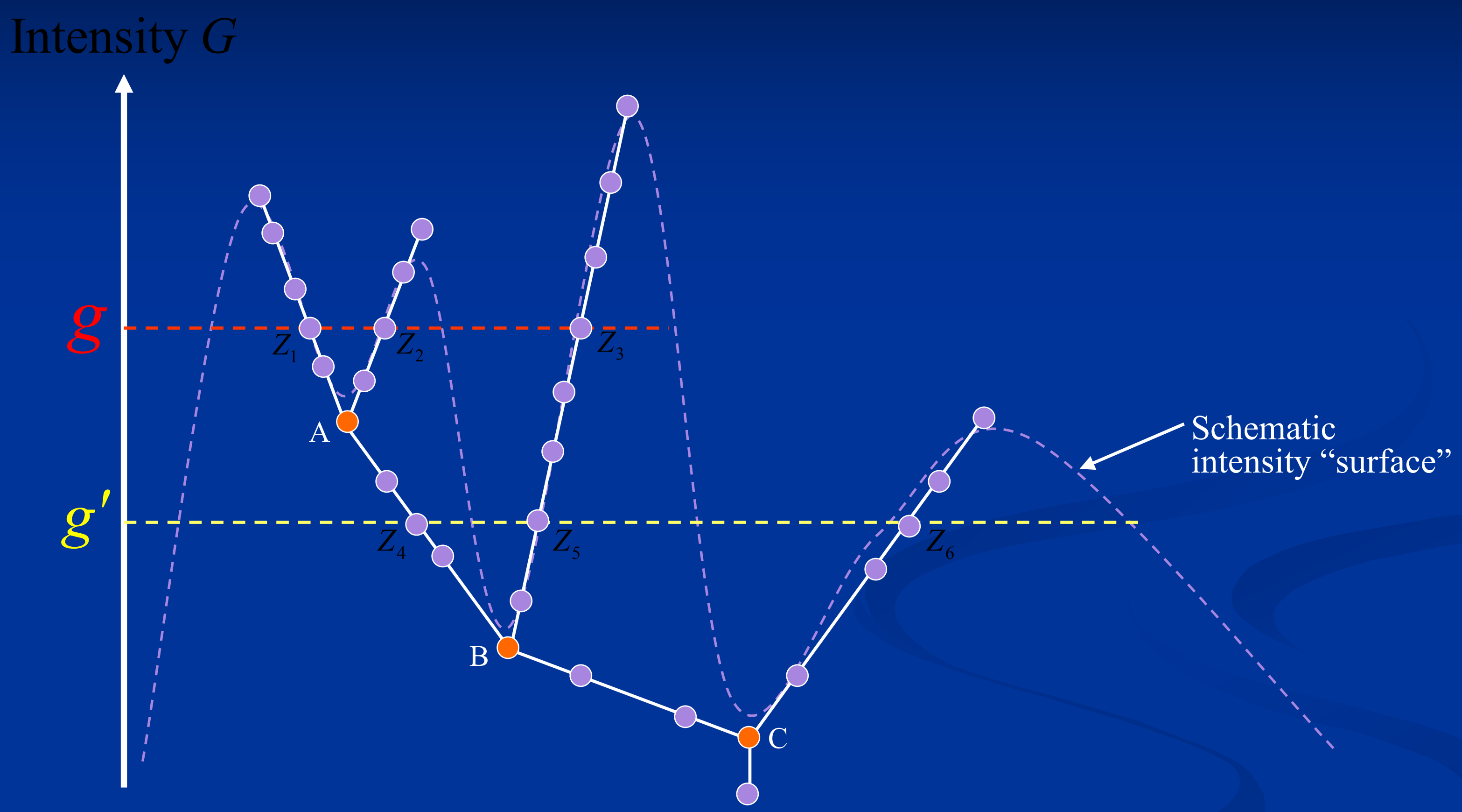
Upper Level Set (ULS) of Intensity Surface



Changing Connectivity of ULS as Level Drops

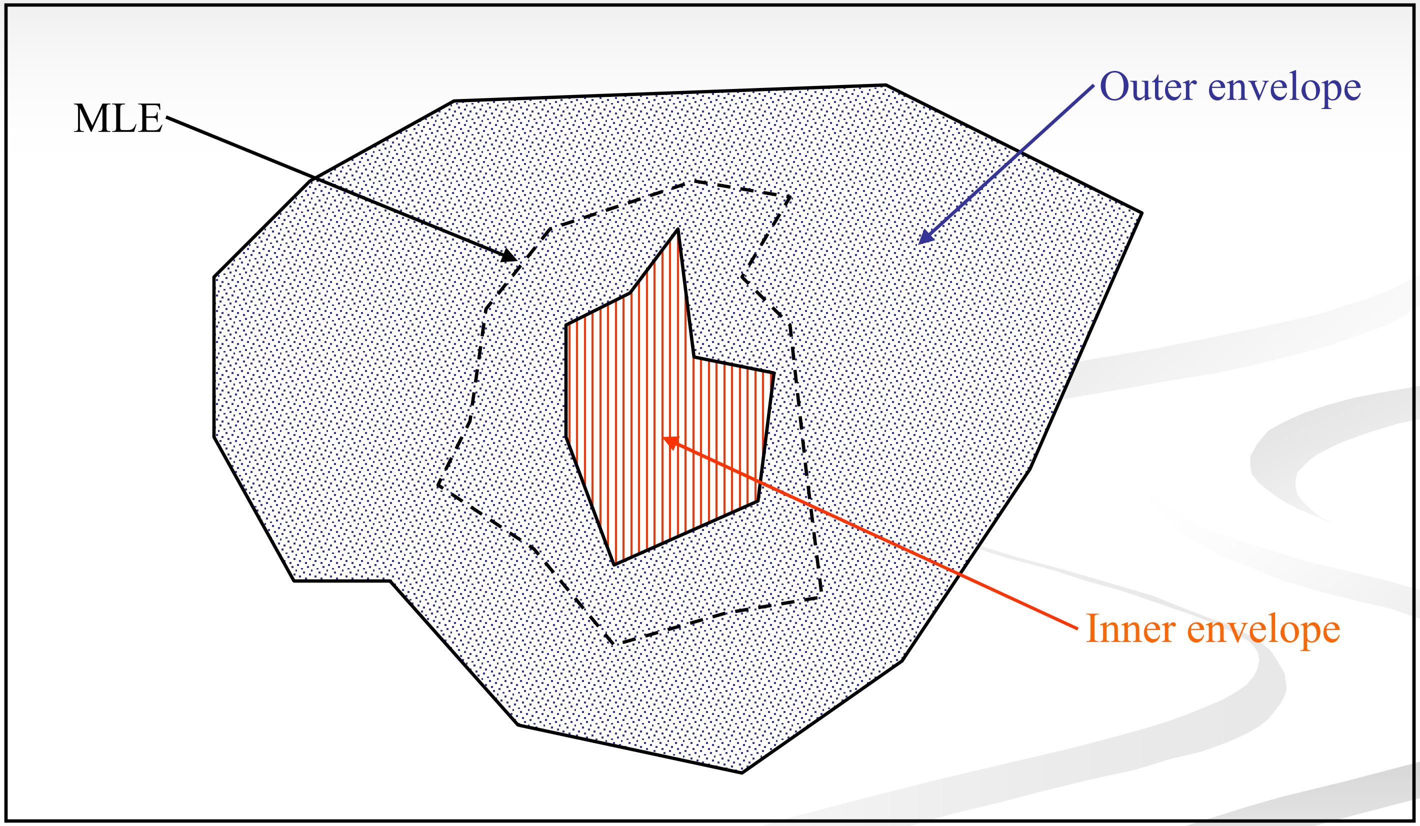


ULS Connectivity Tree

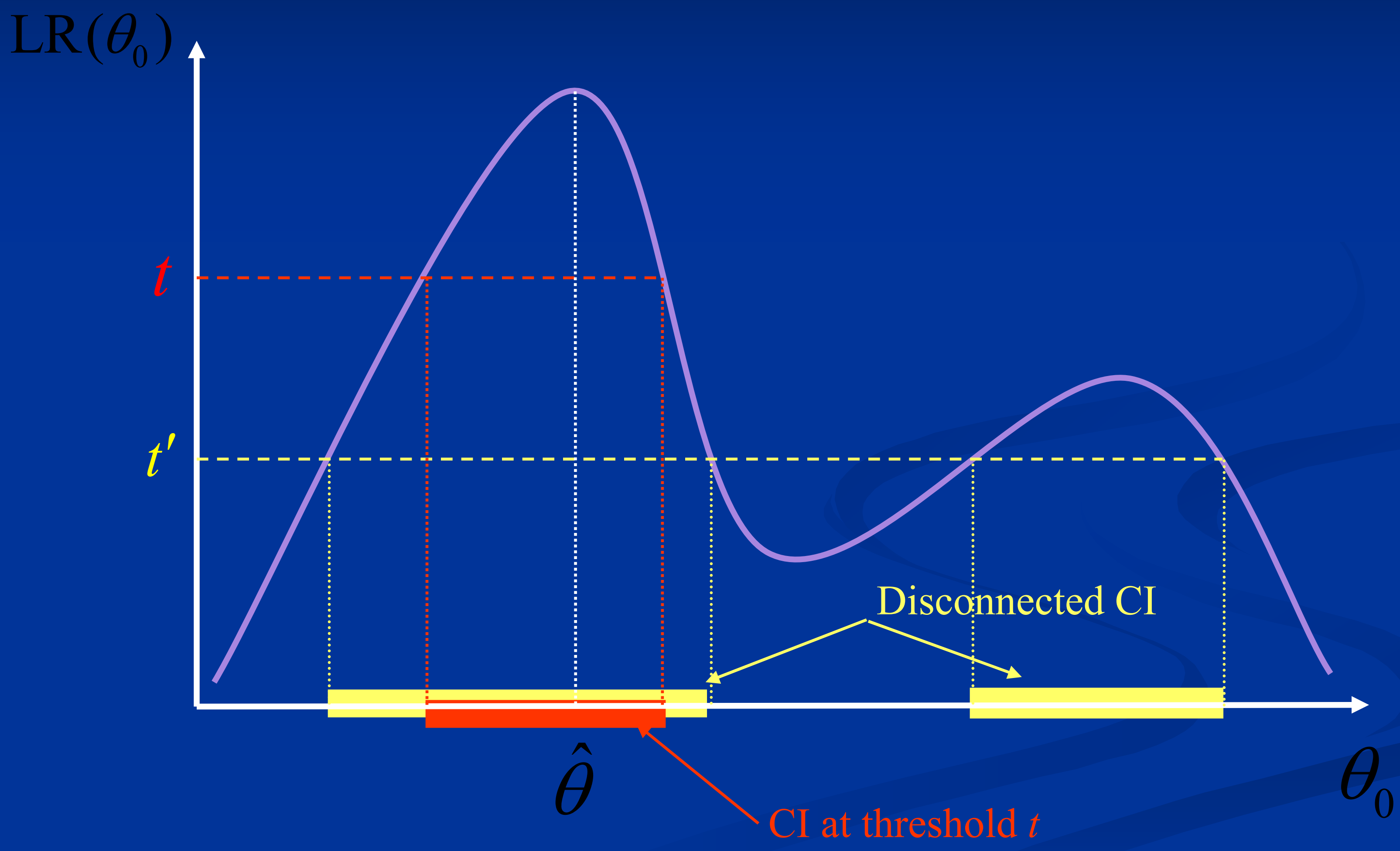


N.B. Intensity surface is cellular (piece-wise constant), with only finitely many levels
A, B, C are junction nodes where multiple zones coalesce into a single zone

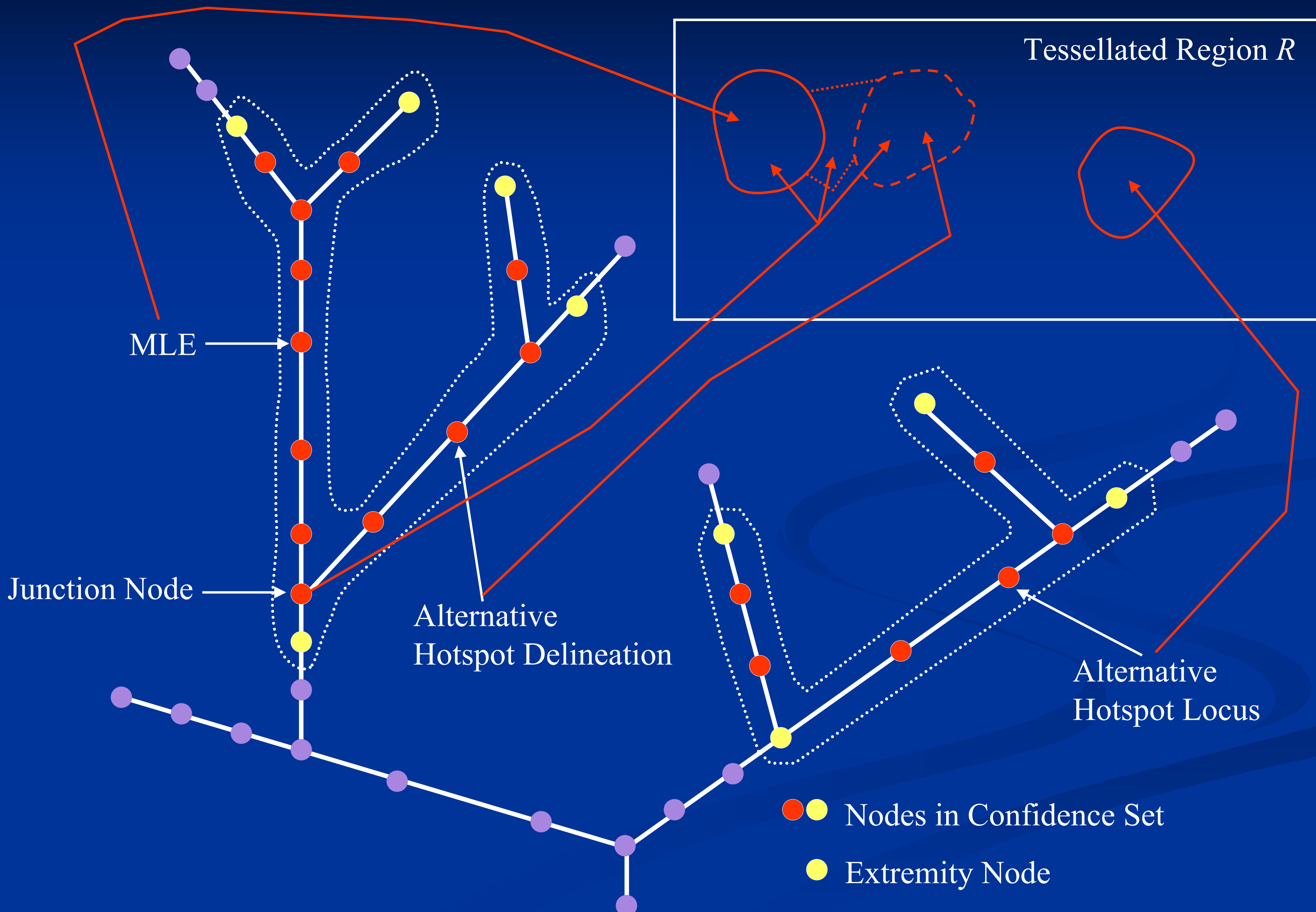
Estimation Uncertainty in Hotspot Delineation



LR Confidence Intervals



Confidence Region on ULS Tree



Hotspot Detection, Delineation, Prioritization

Continuous Responses

- Both human health and environmental contexts
- Simplest distributional model:

$$Y_a \sim \text{Gamma}(k, \beta)$$

- Additivity with respect to the index parameter k suggests that we model k as proportional to size:

$$k_a = A_a / c.$$

- Other distribution models (e.g., lognormal) are possible but are computationally complex and applicable to only a single spatial scale

Hotspot Detection, Delineation, Prioritization

Examples of Continuous Responses

- Human Health Context:
  - Blood pressure levels for spatial variation in hypertension
  - Estrodial levels in women for breast cancer and osteoporosis
  - Cancer survival (censoring issues)
- Environmental Context:
  - Landscape metrics such as forest cover, fragmentation, etc.
  - Pollutant loadings
  - Animal abundance

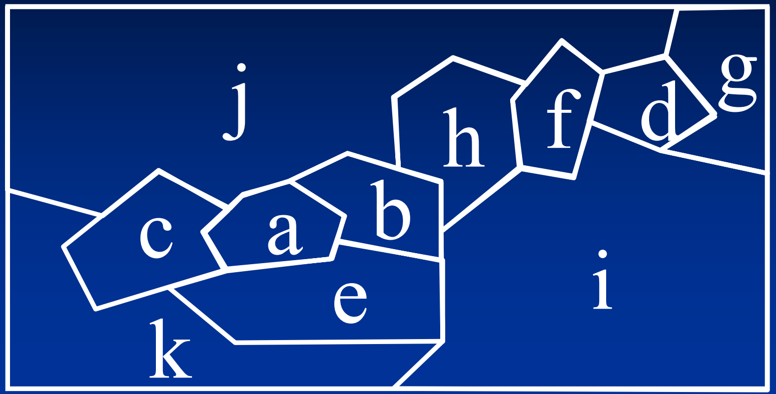
Multiple Criteria Analysis
Multiple Indicators and Choices
Health Statistics
Disease Etiology, Health Policy, Resource Allocation

- First stage screening
  - Significant clusters by SaTScan and/or upper level sets
- Second stage screening
  - Multicriteria noteworthy clusters by partially ordered sets and Hass diagrams
- Final stage screening
  - Follow up clusters for etiology, intervention based on multiple criteria using Hass diagrams

Logo for Statistics, Ecology, Environment, and Society

SATScan Setup -- 1

- Tessellation of a geographic region



a, b, c, … are cell labels

- Region R, Tessellation $T = \{a\}$ of R
- Cell a, Response Y_a, Cell "Size" A_a
- Two distributional settings:
  - Y_a is Binomial (N_a , p_a), $A_a = N_a$, p_a = cell rate/intensity
  - Y_a is Poisson $(\lambda_a A_a)$, λ_a = cell rate/intensity
- Cell sizes A_a are known and fixed
- Cell responses Y_a, $a \in A$, are independent

SATScan Setup -- 2

- $G_a = Y_a / A_a$ empirical cell intensity
  determines a cellular (piece-wise constant) surface
  defined over the tessellated region
- Zones Z are connected unions of cells from tessellation
  Ω = collection of all possible zones
- SatScan hotspot model:
  - Zone Z such that
  - $p_a = p_1$ for all $a \in Z$
  - $p_a = p_0$ for all $a \in R - Z$
  - $p_1 > p_0$

SATScan Setup – 3
Hotspot Detection

- Hypothesis testing approach:

  H_0: There is no hotspot; p_a constant for all cells a

  H_1: There is a hotspot $Z \in \Omega$; Z unknown

- Parameter space for full model:

  $$\{ (Z, p_1, p_0) : \; Z \in \Omega, 0 < p_0 \leq p_1 < 1 \}$$

- For fixed Z, expression for likelihood and MLE for p_1, p_0 are straightforward

- Profile likelihood for Z:

  $$L(Z) = \mathrm{Max} \{L(Z, p_1, p_0) : 0 < p_0 \leq p_1 < 1 \}, \qquad Z \in \Omega$$

SATScan Setup – 4
Hotspot Estimation

- Ω is finite but large
- Maximizing $L(Z)$, $Z \in \Omega$, by exhaustive search impractical
- Possible optimization methods:
  - Stochastic optimization (annealing, GA, etc.)
  - Reduction of Ω to Ω_0 with $\Omega_0 \subset \Omega$ and Ω_0 small enough for exhaustive search
- Possible reductions of Ω to Ω_0 :
  - Expanding circles
    An *a priori* reduction which depends only on the tessellation and not on the data
  - Upper level sets of empirical intensity surface **(Tree-Structured SATScan)**
    An adaptive, data-dependent, reduction allowing flexible shapes for zones in Ω_0 . Data dependence must be incorporated into simulations.

SATScan Setup – 5
Hotspot Testing

- Test H_0 versus H_1 for significance of hotspot
- Use reduced parameter space Ω_0
- $\hat{Z} =$ hotspot MLE, maximizes $L(Z)$, $Z \in \Omega_0$
- Likelihood ratio test

  Test statistic: $\mathrm{LR} = L(\mathrm{H}_0) / L(\hat{Z})$

  Reject H_0 when LR is small
- Nonstandard likelihood situation
  (Ω_0 is finite discrete; parameter Z is non-identifiable under H_0)
- Asymptotic chi-squared not applicable
- Null distribution of LR to be determined by simulation
  - Eliminate nuisance parameters p_1, p_0 under H_0 by conditioning on the total response $\sum_a Y_a$

SATScan Setup – 7

- Question: Are there data-driven (rather than *a priori*) ways of selecting the list of candidate zones ?
- Motivation for the question: A human being can look at a map and quickly determine a reasonable set of candidate zones and eliminate many other zones as obviously uninteresting. Can the computer do the same thing?
- A data-driven proposal: Candidate zones are the connected components of the upper level sets of the response surface. The candidate zones have a tree structure, which may assist in automated detection of multiple, but geographically separate, elevated zones.
- Null distribution: If the list is data-driven (i.e., random), its variability must be accounted for in the null distribution. A new list must be developed for each simulated data set.

Tree-Structured SATScan

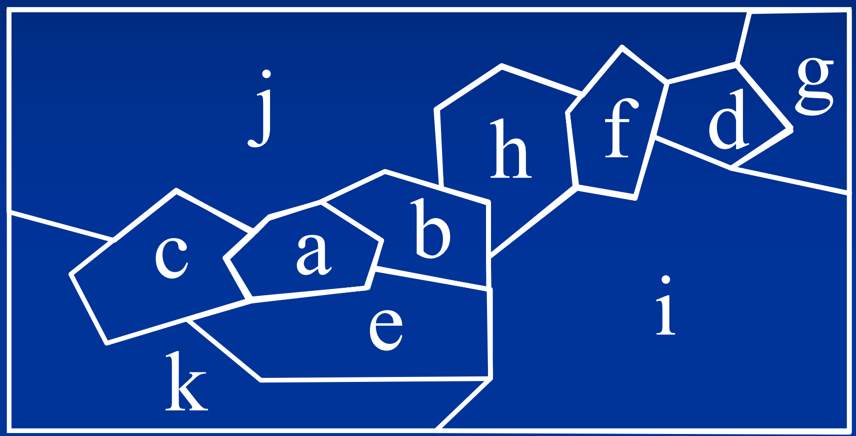
- Data-adaptive approach to reduced parameter space Ω_0
- Zones in Ω_0 are connected components of upper level sets of the empirical intensity function $G_a = Y_a / A_a$
- Upper level set (ULS) at level g consists of all cells a where

$$G_a \geq g$$

- Upper level sets may be disconnected. Connected components are the candidate zones in Ω_0
- These connected components form a rooted tree under set inclusion.
  - Root node = entire region R
  - Leaf nodes = local maxima of empirical intensity surface
  - Junction nodes occur when connectivity of ULS changes with falling intensity level

ULS Connectivity Tree -- 1

- Ingredients:
  - Tessellation of a geographic region:



a, b, c, … are cell labels

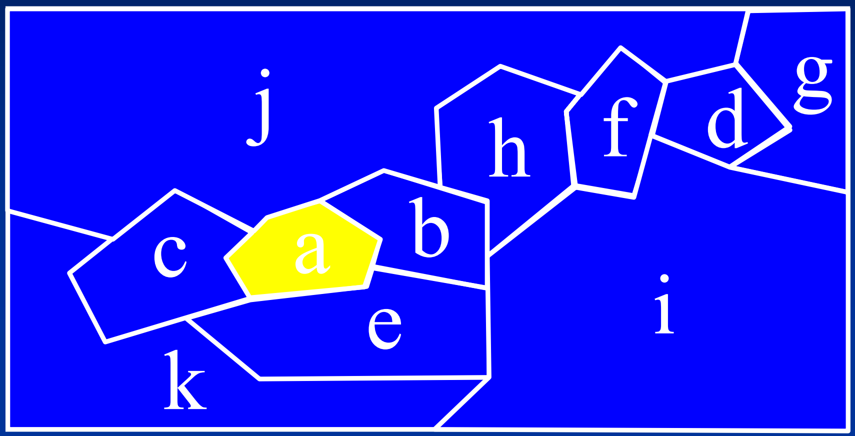
  - Intensity value *G* on each cell. Determines a cellular (piece-wise constant) surface with *G* as elevation.
- Imagine surface initially inundated with water
- Water evaporates gradually exposing the surface which appears as islands in the sea
- How does **connectivity** (number of connected components) of the exposed surface change with falling water level?

ULS Connectivity Tree -- 2

- Think of the tessellated surface as a landform
- Initially the entire surface is under water
- As the water level recedes, more and more of the landform is exposed
- At each water level, cells are colored as follows:
  - **Green** for previously exposed cells (green = vegetated)
  - **Yellow** for newly exposed cells (yellow = sandy beach)
  - **Blue** for unexposed cells (blue = under water)
- For each newly exposed cell, one of three things happens:
  - **New island emerges**.
    Cell is a local maximum. Morse index=2. Connectivity increases.
  - **Existing island increases in size**.
    Cell is not a critical point. Connectivity unchanged.
  - **Two (or more) islands are joined**.
    Cell is a saddle point Morse index=1. Connectivity decreases.

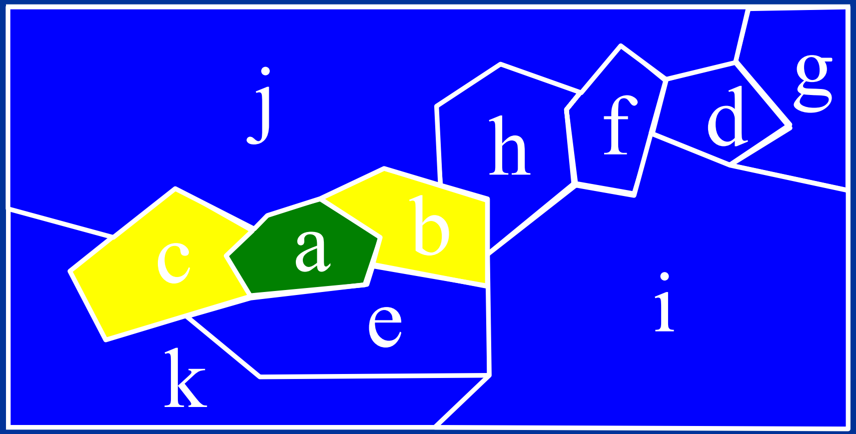
ULS Connectivity Tree -- 3

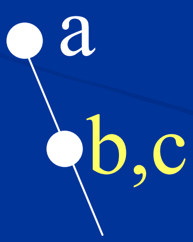
Newly exposed island



ULS Tree

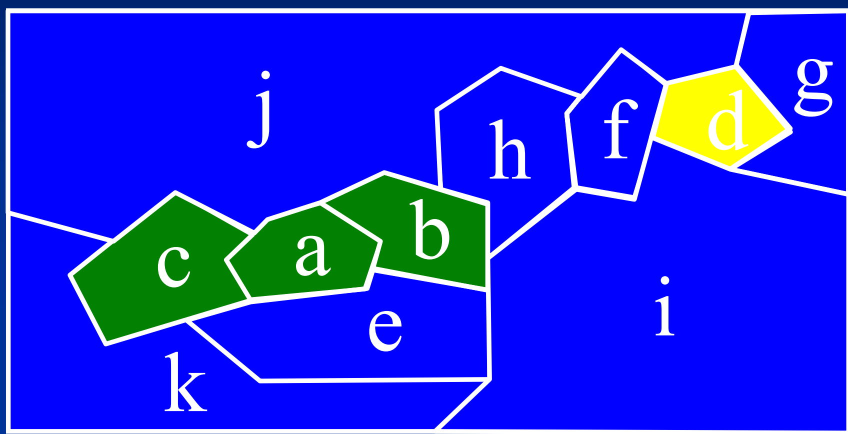
Island grows



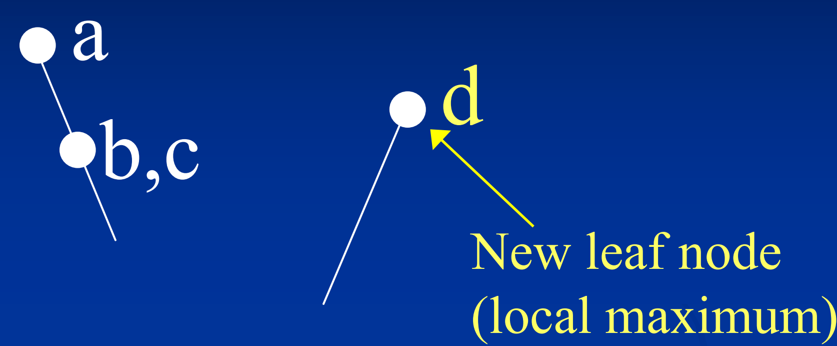


ULS Connectivity Tree -- 4

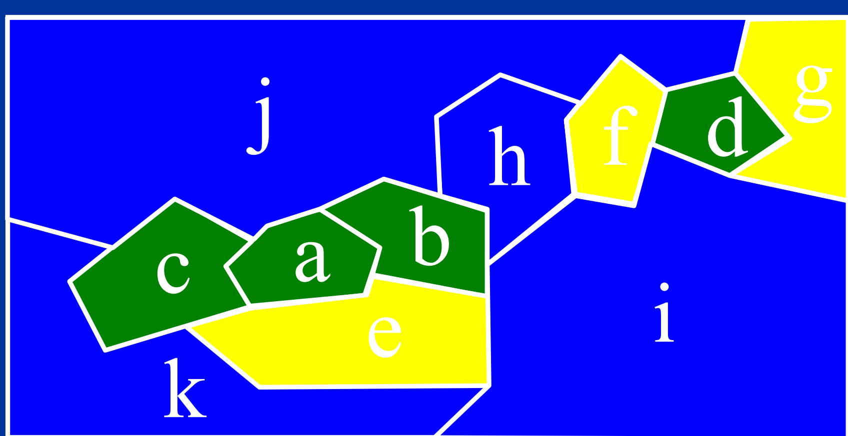
Second island appears

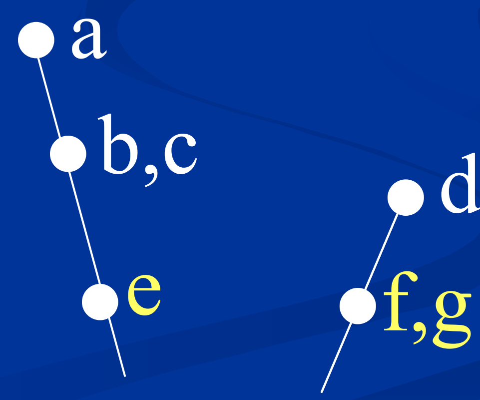


ULS Tree



Both islands grow

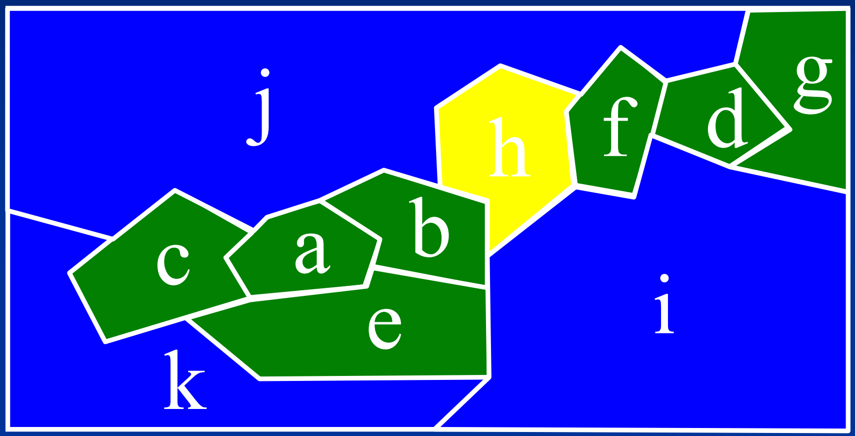


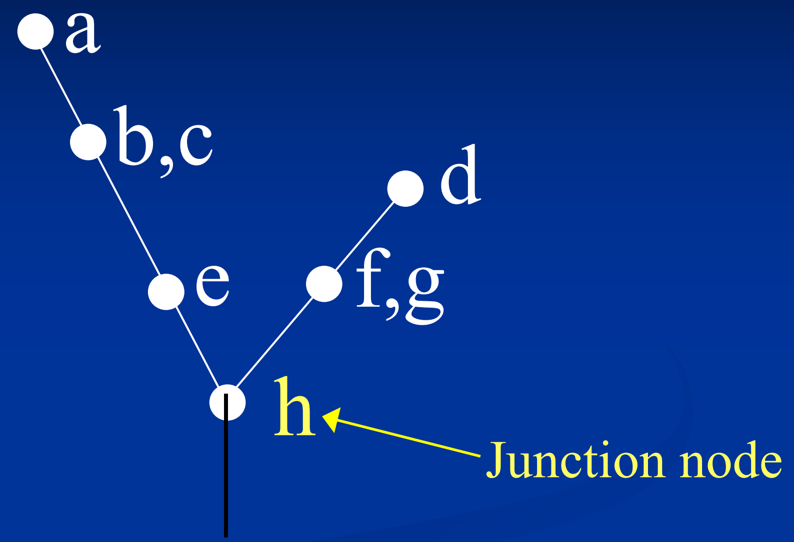


ULS Connectivity Tree -- 5

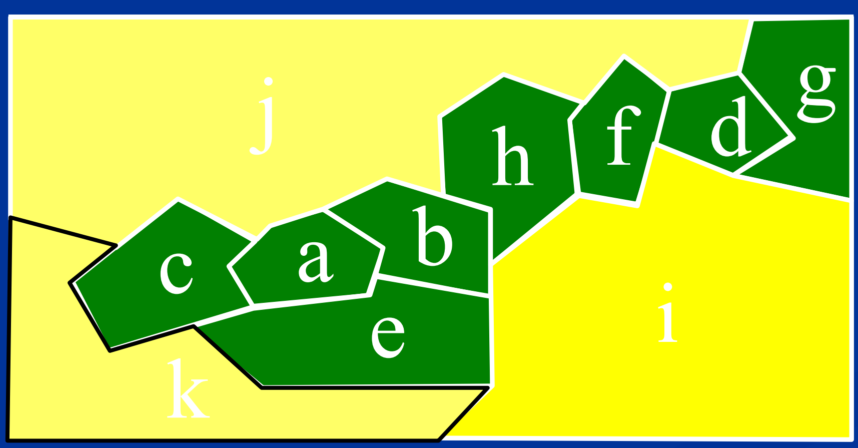
ULS Tree

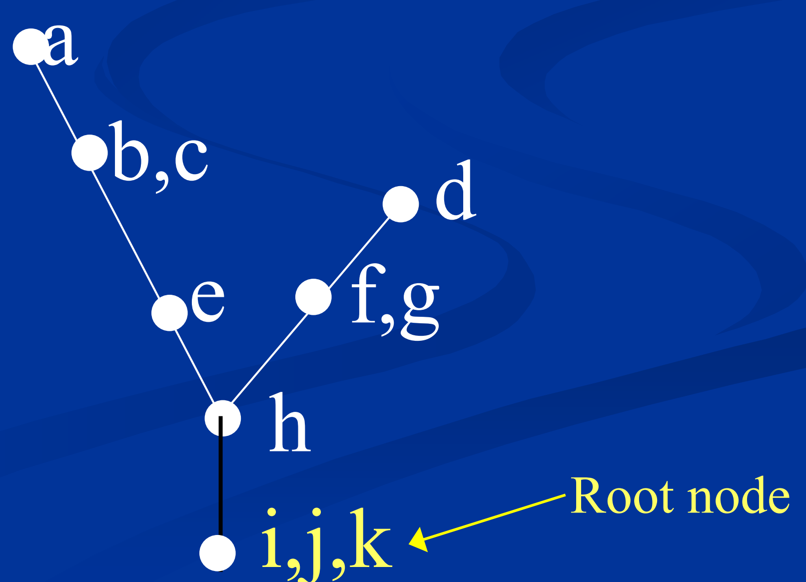
Islands join – saddle point





Exposed land grows





Comparison of Tree-Structured and Circle-Based SATScan

- Agreement/Disagreement regarding hotspot locus

$$\Pr[\ \hat{Z}_{\text{ULS}} \cap \hat{Z}_{\text{circles}} \neq \varnothing\ ]$$

- Comparative plausibility and accuracy of hotspot delineation

$$\Pr[\ L(\hat{Z}_{\text{ULS}}) \geq L(\hat{Z}_{\text{circles}})\ ]$$

- Execution time and computer efficiency

Hotspot Delineation and Hotspot Rating -- 1

- Determine a confidence set for the hotspot
- Each member of the confidence set is a zone which is a statistically plausible delineation of the hotspot at specified confidence
- Confidence set lets us rate individual cells *a* for hotspot membership
- Rating for cell *a* is percentage of zones in confidence set that contain *a*. (More generally, use weighted proportion.)
- Map of cell ratings:
  - Inner envelope = cells with 100% rating
  - Outer envelope = cells with positive rating

Hotspot Delineation and Hotspot Rating -- 2
Confidence Set Determination

- Confidence set is all null hypotheses that cannot be rejected
- As hypotheses, use

$$\tilde{\mathrm{H}}_0: \text{ hotspot } Z = Z_0$$

$$\tilde{\mathrm{H}}_1: \text{ hotspot } Z \neq Z_0$$

  where $Z_0 \in \Omega_0$ is a given zone.
- Confidence set is all $Z_0 \in \Omega_0$ for which $\tilde{\mathrm{H}}_0$ cannot be rejected.
- Likelihood ratio test:

  Test statistic: $\mathrm{LR} = L(Z_0) / L(\hat{Z})$ where $\hat{Z} = \text{MLE under } \tilde{\mathrm{H}}_0 \cup \tilde{\mathrm{H}}_1$

  Reject H_0 when LR is small
- Null distributions have to be determined by simulation

- 3
Review of LR Confidence Set Determination

- $\mathrm{H}_0 : \theta = \theta_0$

  $\mathrm{H}_1 : \theta \neq \theta_0$

  $\alpha =$ significance level, $c =$ confidence level

  Test statistic: $\mathrm{LR} = L(\theta_0) / L(\hat{\theta})$

  $\Pr[\,\mathrm{LR} \geq t(\theta_0) \mid \mathrm{H}_0 : \theta = \theta_0\,] = 1 - \alpha = c$

  $t(\theta_0) = t,$ critical point approximately free of θ_0
- Null distribution free of parameter (approximately)
- Simulation of null distribution at endpoints of confidence interval

Hotspot Delineation and Hotspot Rating -- 6 Tree-Structured SATScan

- How is the null distribution to be simulated for given $Z_0 \in \Omega_0$?
- What is the analogue of extemity or boundary of the confidence set, when the parameter set Ω_0 is finite?
- How do we handle and interpret multimodality of LR giving rise to disconnected confidence set for the hotspot?

Hotspot Delineation and Hotspot Rating -- 7 Tree-Structured SATScan

- Is the null distribution fairly constant across much of the tree?
- Assignment of *p*-value to every LR value and hence to every node in the ULS tree
- Secondary hotspots, *p*-values *versus* pseudo *p*-values

Multiple Hotspot Detection and Delineation-1 Tree-Structured SatScan

$\mathrm{H}_0 : p_a$ constant, $a \in T$. No hotspot

$\mathrm{H}_1' : \exists M > 0$ and $\{Z_i : i = 1,2,\cdots, M;\text{separated}\}$, such that

$$p_a = p_i, \quad a \in Z_i$$

$$p_a = p_0, \quad a \in R - \bigcup Z_i = Z_0$$

$$p_i > p_0, \quad i = 1,2,\cdots, M$$

Under H_1', the following parameters :

$M; Z_1, Z_2, \cdots, Z_M; p_0, p_1, p_2, \cdots, p_M$.

Multiple Hotspot Detection and Delineation-2 Tree-Structured SatScan Maximized Profile Likelihood

$$L(M, Z_1, Z_2, \cdots, Z_M) = \underset{p_0, \ldots, p_M}{\text{Max}} \, L\left(M; Z_1, \cdots; Z_M; p_0, p_1, \cdots; p_M\right)$$

$$= L\left(M; Z_1, \cdots, Z_M; \hat{p}_0, \hat{p}_1, \cdots, \hat{p}_M\right)$$

where $\hat{p}_i = \dfrac{\sum Y_a}{\sum N_a}, a \in Z_i$.

Exhaustive search to maximize L over $\Omega = \left\{M; Z_1, \cdots, Z_M\right\}$

Ω enormously large.

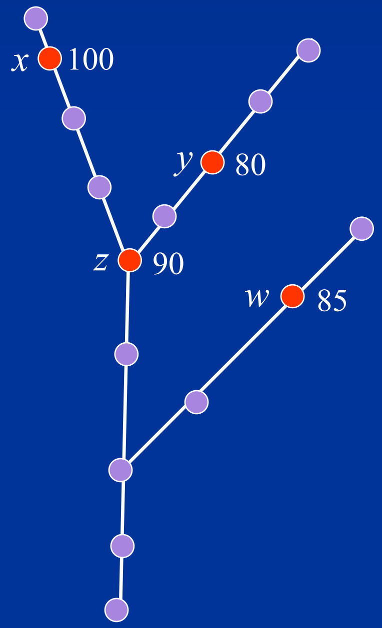
Apply stochastic optimation or

choose manageable subset $\Omega' \subset \Omega$ and exhaustive search over Ω'

Multiple Hotspot Detection and Delineation-3 Tree-Structured SATScan

- Three parameter space reduction schemes:
  - Falling waterline model (modes of intensity function on ULS tree)
  - Modes of LR on ULS connectivity tree
  - Sequential determination of secondary hotspots

Likelihood Function Defined on ULS Tree
Multiple Hotspots



Numerical labels are values of the likelihood function

Highlighted nodes are local maxima (modes) of the likelihood function over the tree. Each is a candidate for a hotspot locus

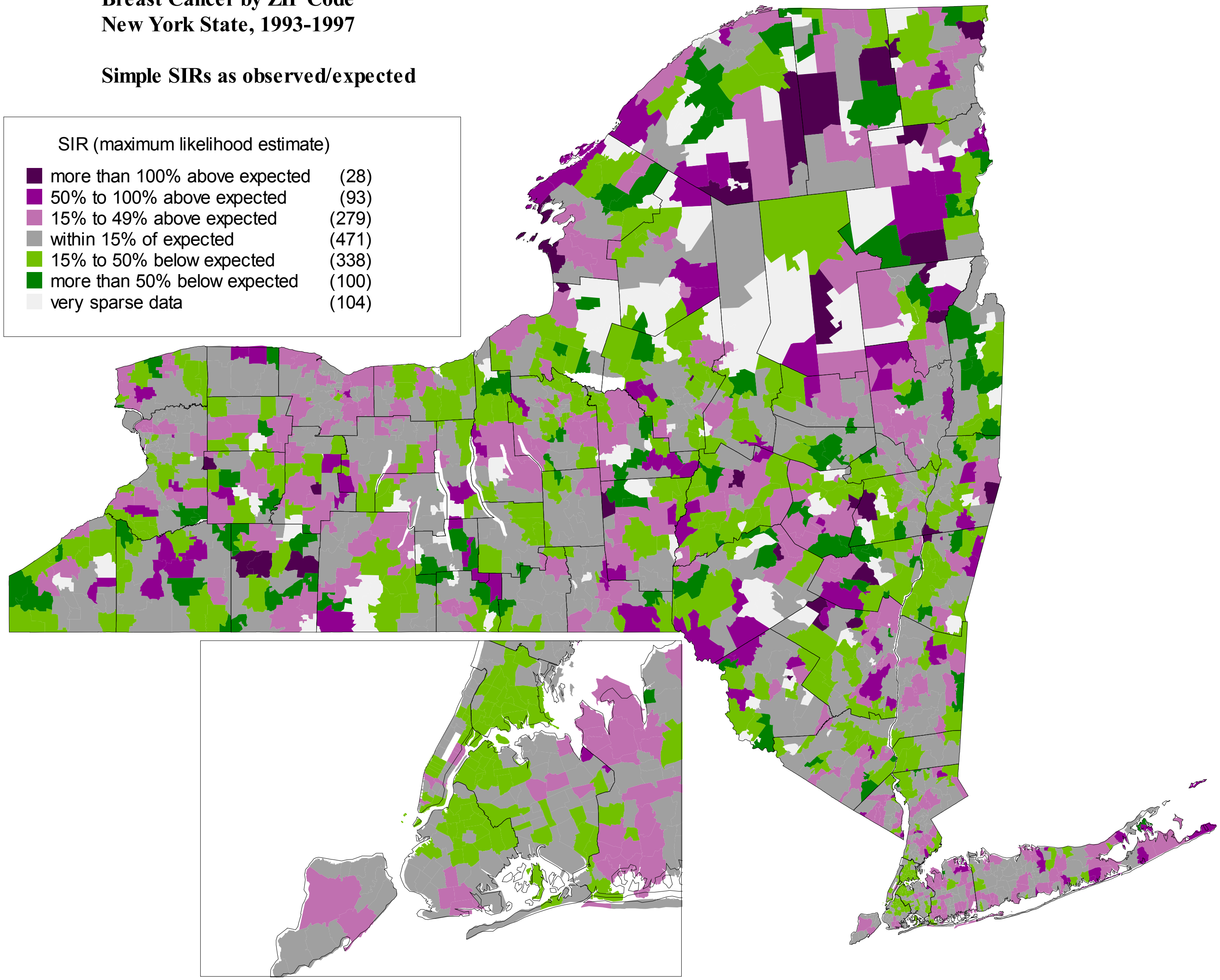
Node *x* is the MLE (global maximum)

Hotspot Prioritization and Poset Ranking

- Multiple hotspots with intensities significantly elevated relative to the rest of the region
- Ranking based on likelihood values, and additional attributes: raw intensity values, socio-economic and demographic factors, feasibility scores, excess cases, seasonal residence, atypical demographics, etc.
- Multiple attributes, multiple indicators
- Ranking without having to integrate the multiple indicators into a composite index

Breast Cancer by ZIP Code
New York State, 1993-1997

Simple SIRs as observed/expected



Ranking Possible Disease Clusters in the State of New York

Data Matrix

cluster *	SIR	LL	Young Cases	Multiple Cancers	Atypical Demographics	Late Stage of Diagnosis
LF2	2.09	10.36	2	1	1	2
LM14	1.5	36	2	0	0	2
LM4	2.04	19.21	2	0	0	2
LF7	1.51	15.43	1	1	1	1
B2	1.21	31.3	2	1	0	2
B4	1.25	28.4	1	0	0	0
LM1	2.32	21.91	0	1	0	2
LM3	2.13	21.26	1	1	0	1
LM7	2.12	13.33	1	0	0	2

* LF = lung, female; LM = lung, male; B = breast

Multiple Criteria Analysis
Multiple Indicators
Partial Ordering Procedures

- Cells are objects of primary interest, such as countries, states, watersheds, counties, etc.
- Cell comparisons and rankings are the goals
- Suite of indicators are available on each cell
- Different indicators have different comparative messages, i.e., partial instead of linear ordering
- Hasse diagrams for visualization of partial orders. Multi-level diagram whose top level of nodes consists of all maximal elements in the partially ordered set of objects. Next level consists of all maximal elements when top level is removed from the partially ordered set, etc. Nodes are joined by segments when they are immediately comparable.

HUMAN ENVIRONMENT INDEX
LAND, AIR, WATER INDICATORS

for land - % of undomesticated land, i.e. total land area - domesticated (permanent corps and pastures, built up areas, roads, etc.)
for air - % of renewable energy resources, i.e. hydro, solar, wind, geothermal
for water - % of population with access to safe drinking water

RANK	COUNTRY	LAND	AIR	WATER	HEI
1	*Sweden*	69.01	35.24	100	0.68
2	*Finland*	76.46	19.05	98	0.65
3	*Norway*	27.38	63.98	100	0.64
5	*Iceland*	1.79	80.25	100	0.61
13	*Austria*	40.57	29.85	100	0.57
22	*Switzerland*	30.17	28.10	100	0.53
39	*Spain*	32.63	7.74	100	0.47
45	*France*	28.34	6.50	100	0.45
47	*Germany*	32.56	2.10	100	0.45
51	*Portugal*	34.62	14.29	82	0.44
52	*Italy*	23.35	6.89	100	0.43
59	*Greece*	21.59	3.20	98	0.41
61	*Belgium*	21.84	0.00	100	0.41
64	*Netherlands*	19.43	1.07	100	0.40
77	*Denmark*	9.83	5.04	100	0.38
78	*United Kingdom*	12.64	1.13	100	0.38
81	*Ireland*	9.25	1.99	100	0.37

Hasse Diagram (all countries)

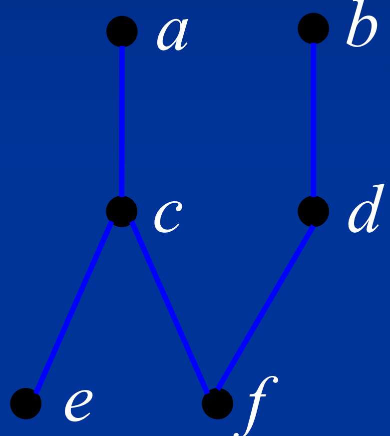
Hasse Diagram (western Europe)

Ranking Partially Ordered Sets – 2

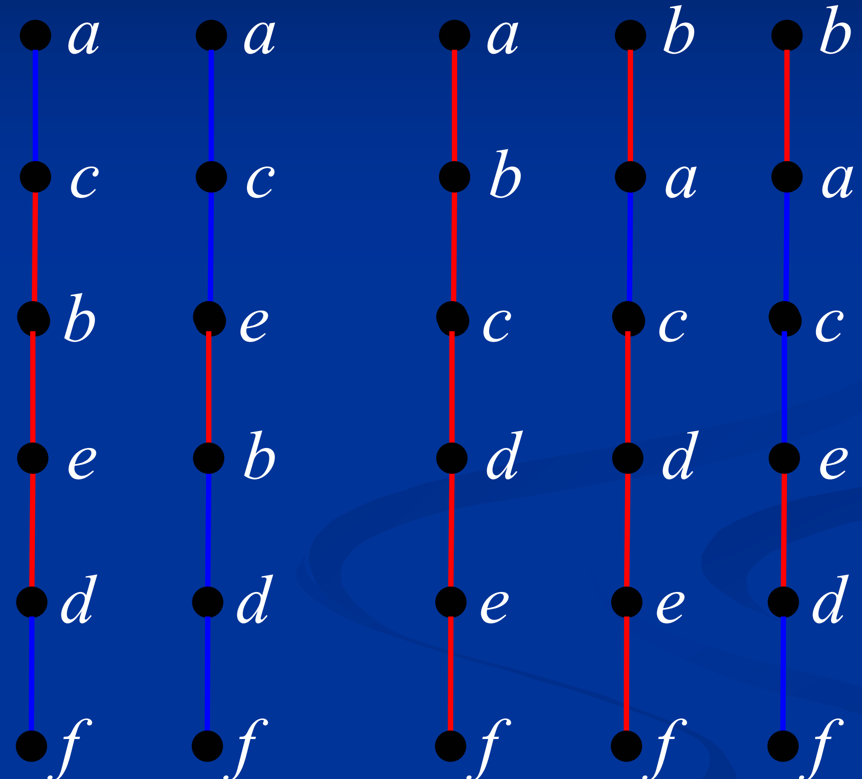
An Example

Poset

(Hasse Diagram)



Some linear extensions



Jump Size: 3 1 5 4 2

Jump or **Imputed Link** (-------) is a link in the ranking that is not implied by the partial order

Ranking Partially Ordered Sets – 3

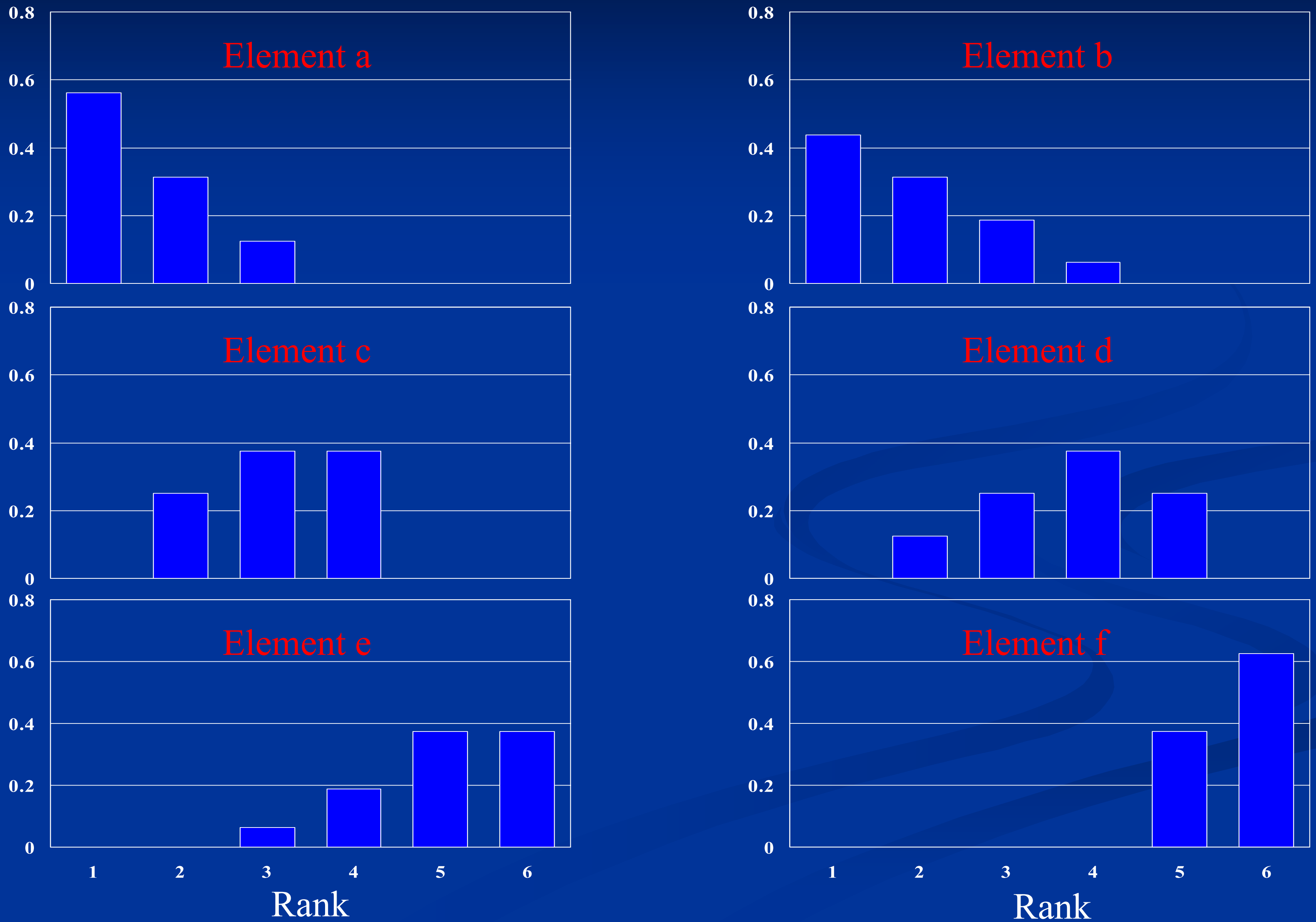
In the example from the preceding slide, there are a total of 16 linear extensions, giving the following frequency table.

	Rank						
Element	1	2	3	4	5	6	Totals
a	9	5	2	0	0	0	16
b	7	5	3	1	0	0	16
c	0	4	6	6	0	0	16
d	0	2	4	6	4	0	16
e	0	0	1	3	6	6	16
f	0	0	0	0	6	10	16
Totals	16	16	16	16	16	16	

- Each (normalized) row gives the rank-frequency distribution for that element
- Each (normalized) column gives a rank-assignment distribution across the poset

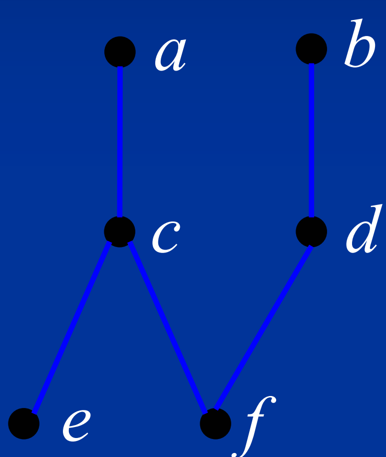
Ranking Partially Ordered Sets – 3a

Rank-Frequency Distributions

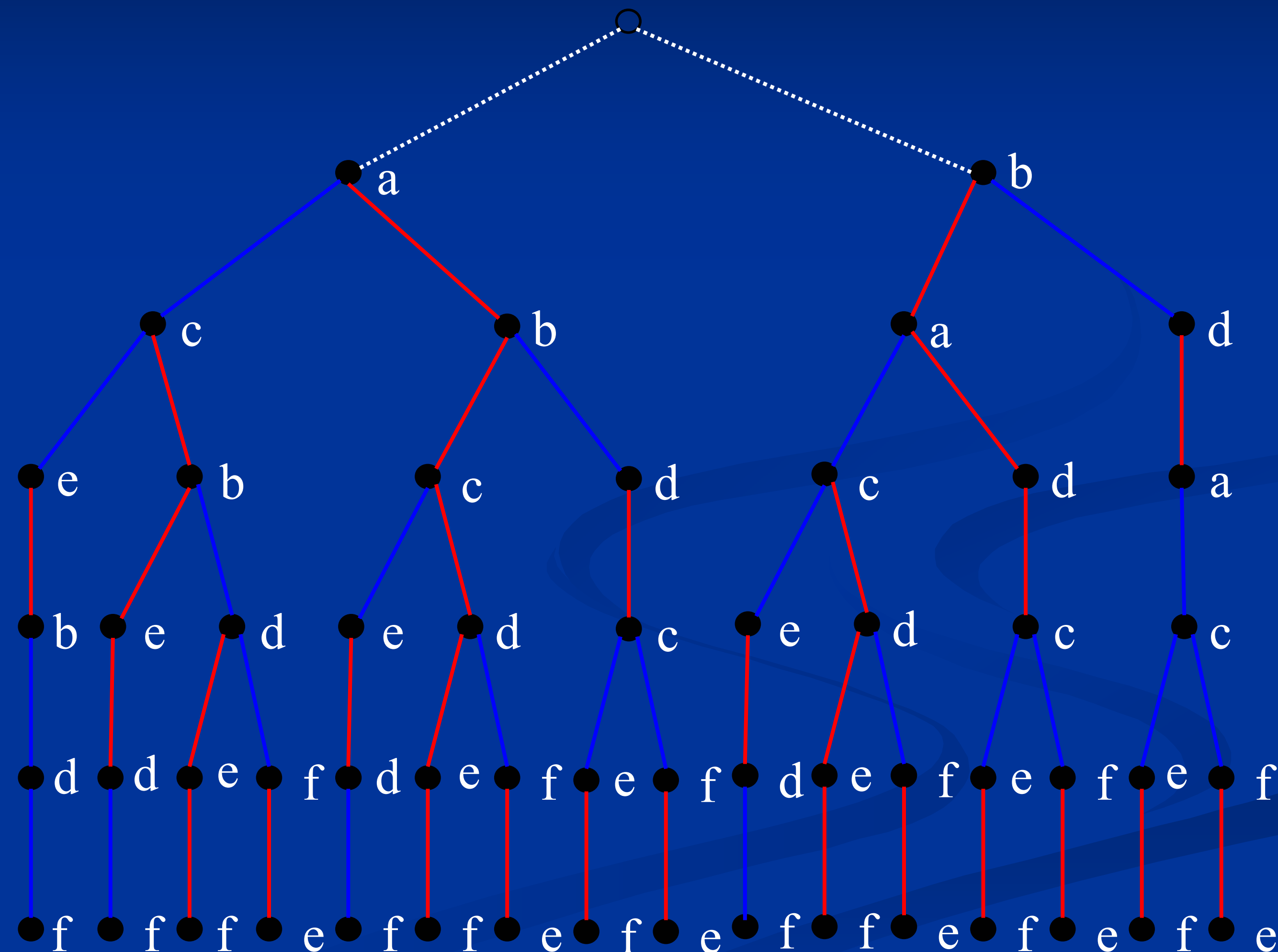


Ranking Partially Ordered Sets – 5

Poset
(Hasse Diagram)



Linear extension decision tree



Jump Size: 1 3 3 2 3 5 4 3 3 2 4 3 4 4 2 2

Cumulative Rank Frequency Operator – 5

An Example of the Procedure

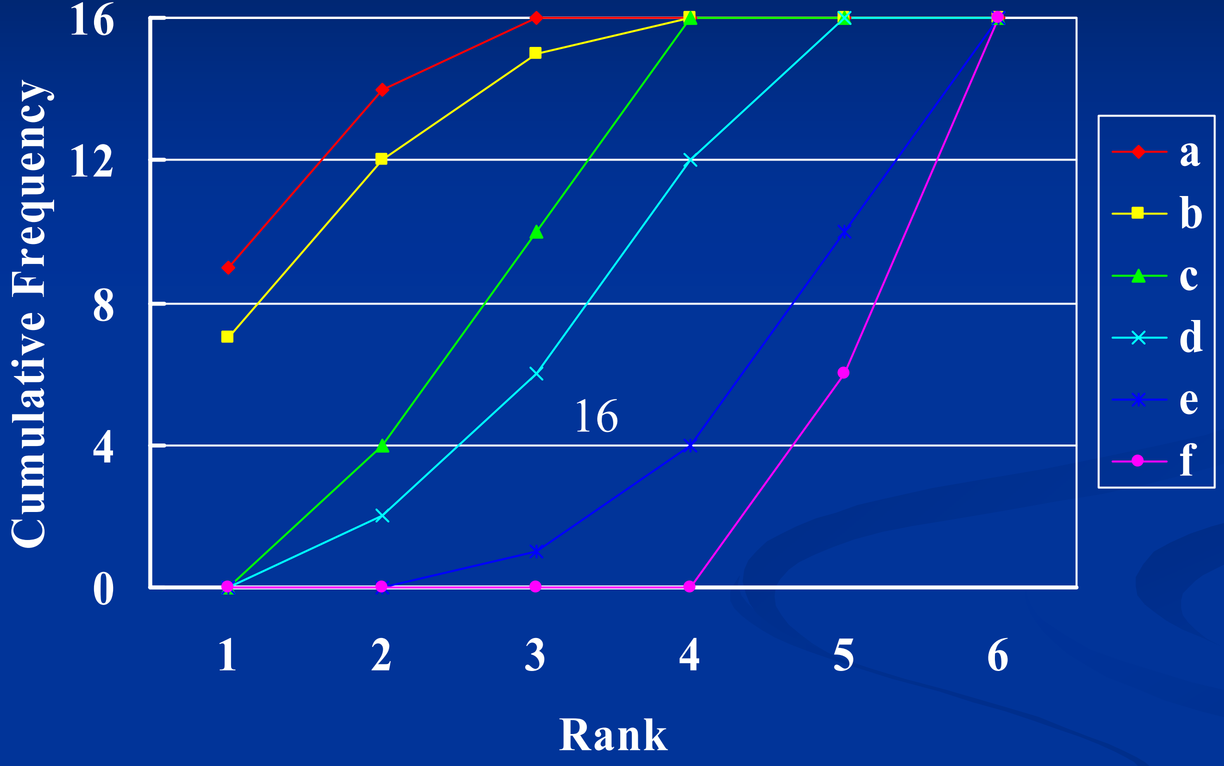
In the example from the preceding slide, there are a total of 16 linear extensions, giving the following cumulative frequency table.

	Rank					
Element	1	2	3	4	5	6
a	9	14	16	16	16	16
b	7	12	15	16	16	16
c	0	4	10	16	16	16
d	0	2	6	12	16	16
e	0	0	1	4	10	16
f	0	0	0	0	6	16

Each entry gives the number of linear extensions in which the element (row label) receives a rank equal to or better that the column heading

Cumulative Rank Frequency Operator – 6

An Example of the Procedure

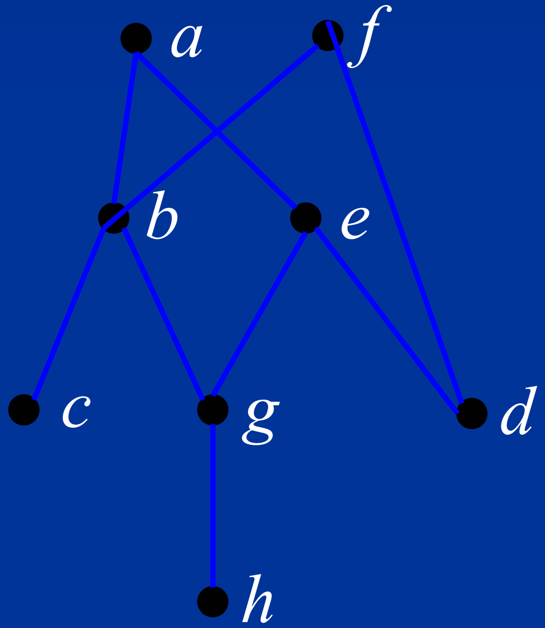


The curves are stacked one above the other and the result is a linear ordering of the elements: $a > b > c > d > e > f$

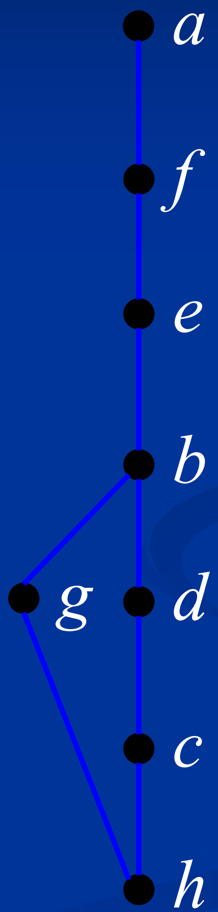
Cumulative Rank Frequency Operator – 7

An example where $\mathscr{F}$ must be iterated

Original Poset
(Hasse Diagram)



$\mathscr{F}$



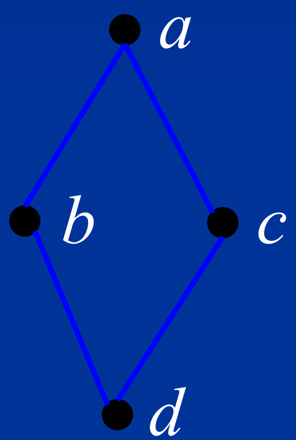
$\mathscr{F}^2$

Cumulative Rank Frequency Operator – 8

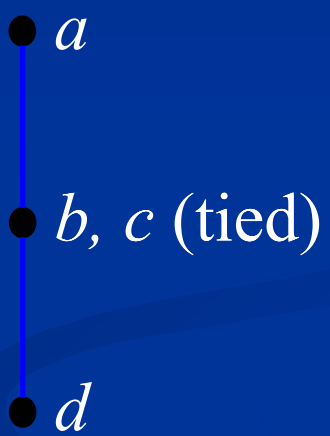
An example where $\mathscr{F}$ results in ties

Original Poset

(Hasse Diagram)



$\mathscr{F}$



- Ties reflect symmetries among incomparable elements in the original Hasse diagram
- Elements that are comparable in the original Hasse diagram will not become tied after applying $\mathscr{F}$ operator