Mining Environmental - Ecological Data

Nikos Giatrakos

Information Systems Laboratory,

Department of Informatics, University of Piraeus

http://infolab.cs.unipi.gr

Outline

- Introduction
- Mining Large Scale Environmental Data
  - Preprocessing
  - Clustering
  - Association Rules
- Mining Medium and Low Scale Ecological Data
  - Co-location Mining
  - STAMM
  - PLUMS
  - Regression Trees, Neural Networks, Clustering
- Sensor Networks for Environmental Applications

Outline

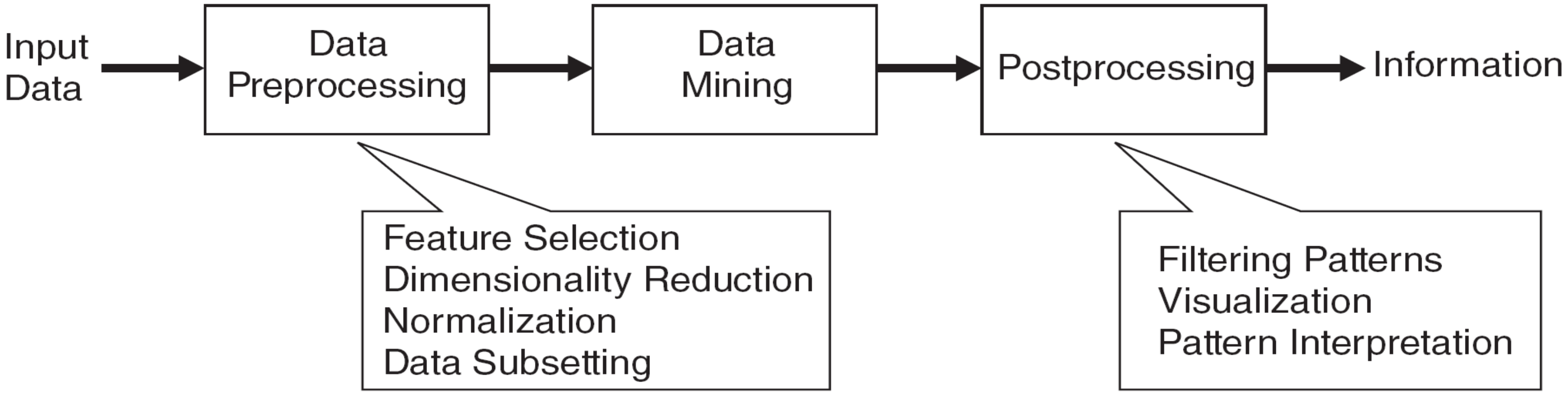
- Introduction
- Mining Large Scale Environmental Data
  - Preprocessing
  - Clustering
  - Association Rules
- Mining Medium and Low Scale Ecological Data
  - Co-location Mining
  - STAMM
  - PLUMS
  - Regression Trees, Neural Networks, Clustering
- Sensor Networks for Environmental Applications

Why Statistics Is Not Sufficient

- Hypothesize-and-test paradigm is extremely labor-intensive.
  - Extremely large and growing families of interesting spatio-temporal hypotheses and patterns in ecological datasets.
- Classical statistics deals primarily with numeric data whereas ecological data contains many categorical attributes.
  - Types of vegetation, ecological events and geographical landmarks.
- Ecological datasets have selection bias in terms of being convenience or opportunity samples.
  - Not traditional statistical idealized random samples from independent, identical distributions.

Benefits of Data Mining

- Data mining provides earth scientist with tools that allow them to spend more time choosing and exploring interesting families of hypotheses.
- By applying the proposed data mining techniques, some of the steps of hypothesis generation and evaluation will be automated, facilitated and improved.
- Association rules provide a “new” framework for detecting relationships between events.



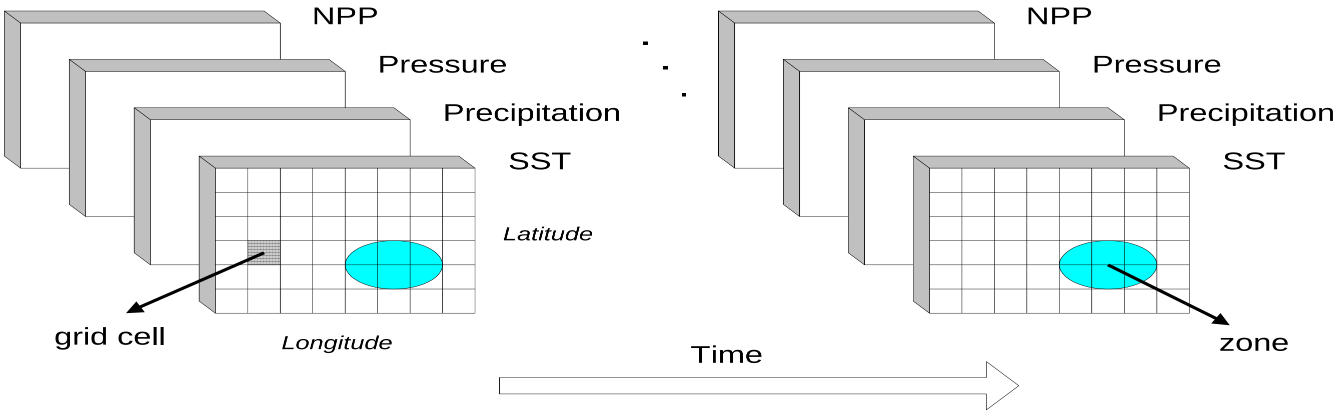
V. Kumar: Mining Earth Science Data (http://www-users.cs.umn.edu/~kumar/nasa-umn)

Outline

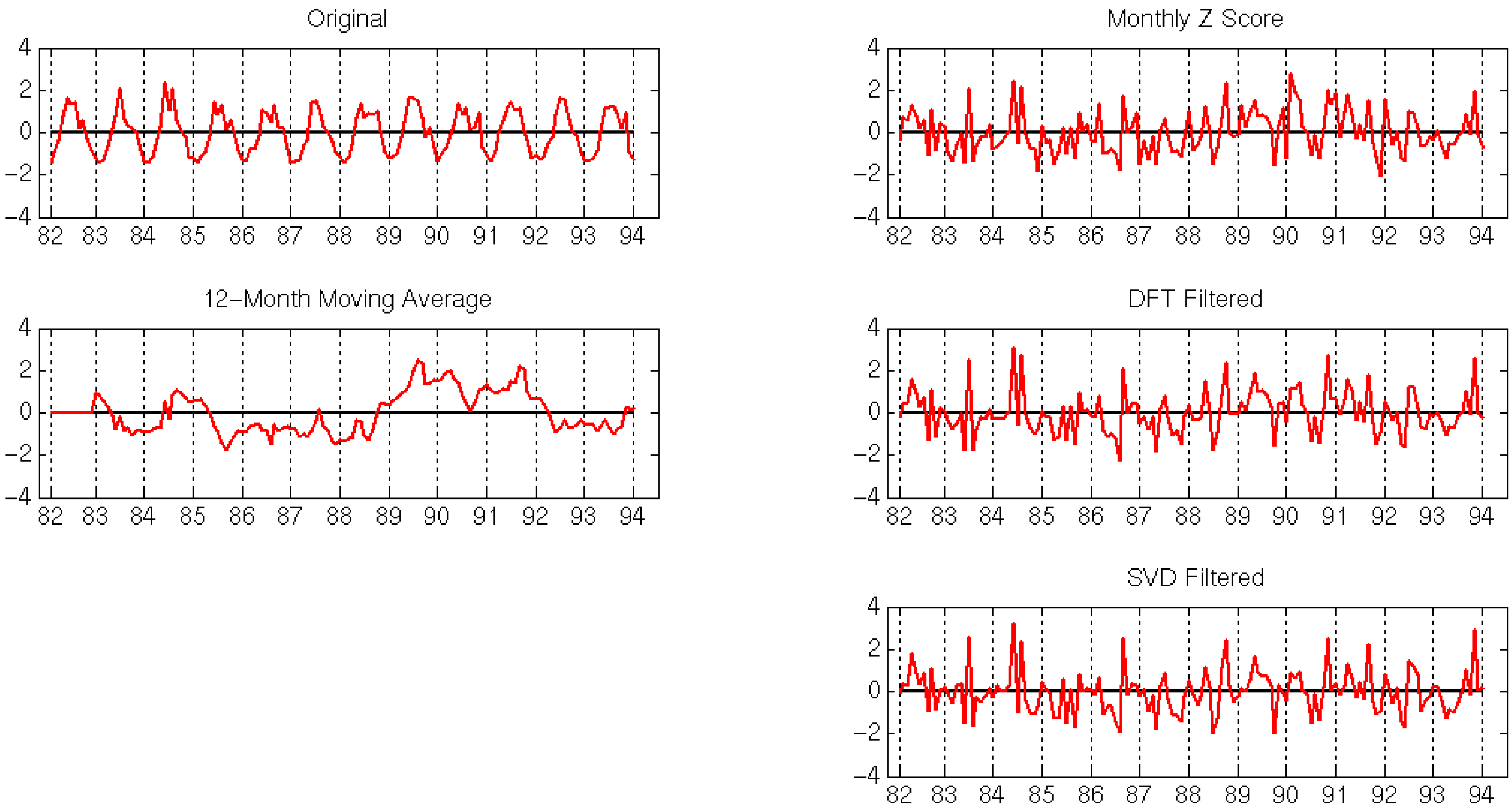
- Introduction
- Mining Large Scale Environmental Data
  - Preprocessing
  - Clustering
  - Association Rules
- Mining Medium and Low Scale Ecological Data
  - Co-location Mining
  - STAMM
  - PLUMS
  - Regression Trees, Neural Networks, Clustering
- Sensor Networks for Environmental Applications

Preprocessing

- Must remove seasonality to obtain the “interesting” data.
  - 12 month moving average
    - Smoothes as well as removes seasonality
  - Discrete Fourier Transform
  - Monthly Z Score
    - Subtract of monthly mean and divide by monthly standard deviation
  - Singular Value Decomposition



Removing Seasonality from Atlanta TS



Original

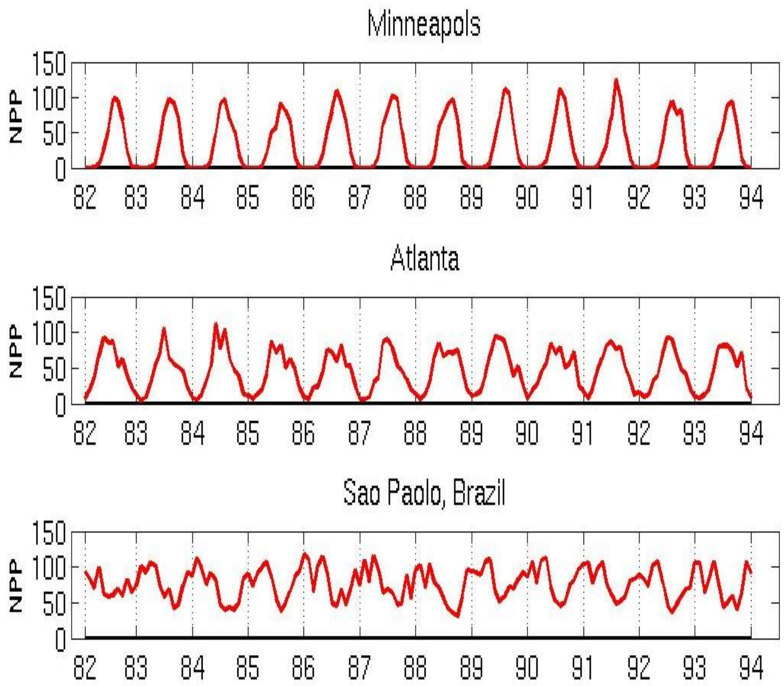
12–Month Moving Average

Monthly Z Score

DFT Filtered

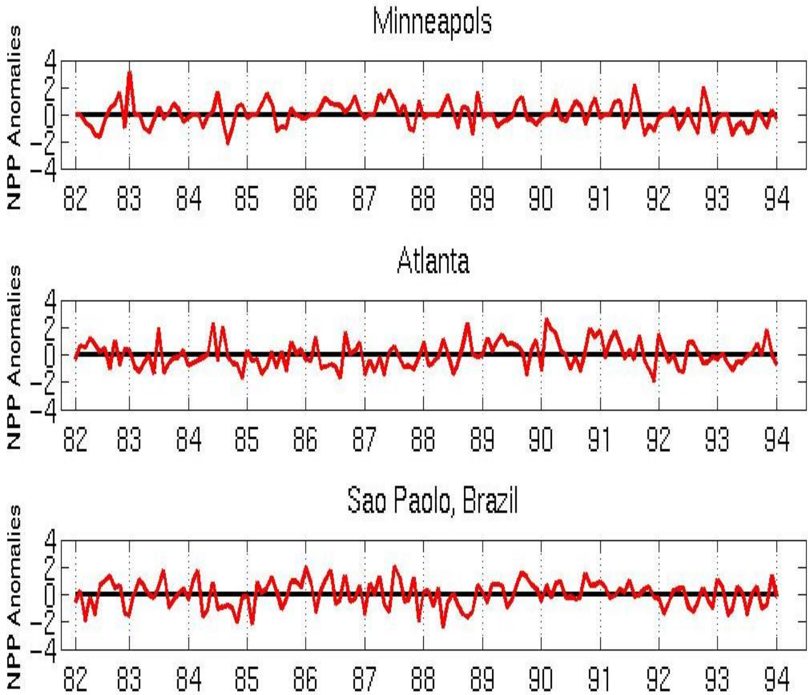
SVD Filtered

Seasonality Accounts for Much Correlation



Correlations between time series

	Minneapolis	Atlanta	Sao Paolo
Minneapolis	1.0000	0.7591	-0.7581
Atlanta	0.7591	1.0000	-0.5739
Sao Paolo	-0.7581	-0.5739	1.0000



Correlations between time series

	Minneapolis	Atlanta	Sao Paolo
Minneapolis	1.0000	0.0492	0.0906
Atlanta	0.0492	1.0000	-0.0154
Sao Paolo	0.0906	-0.0154	1.0000

V. Kumar: Mining Earth Science Data (http://www-users.cs.umn.edu/~kumar/nasa-umn)

Outline

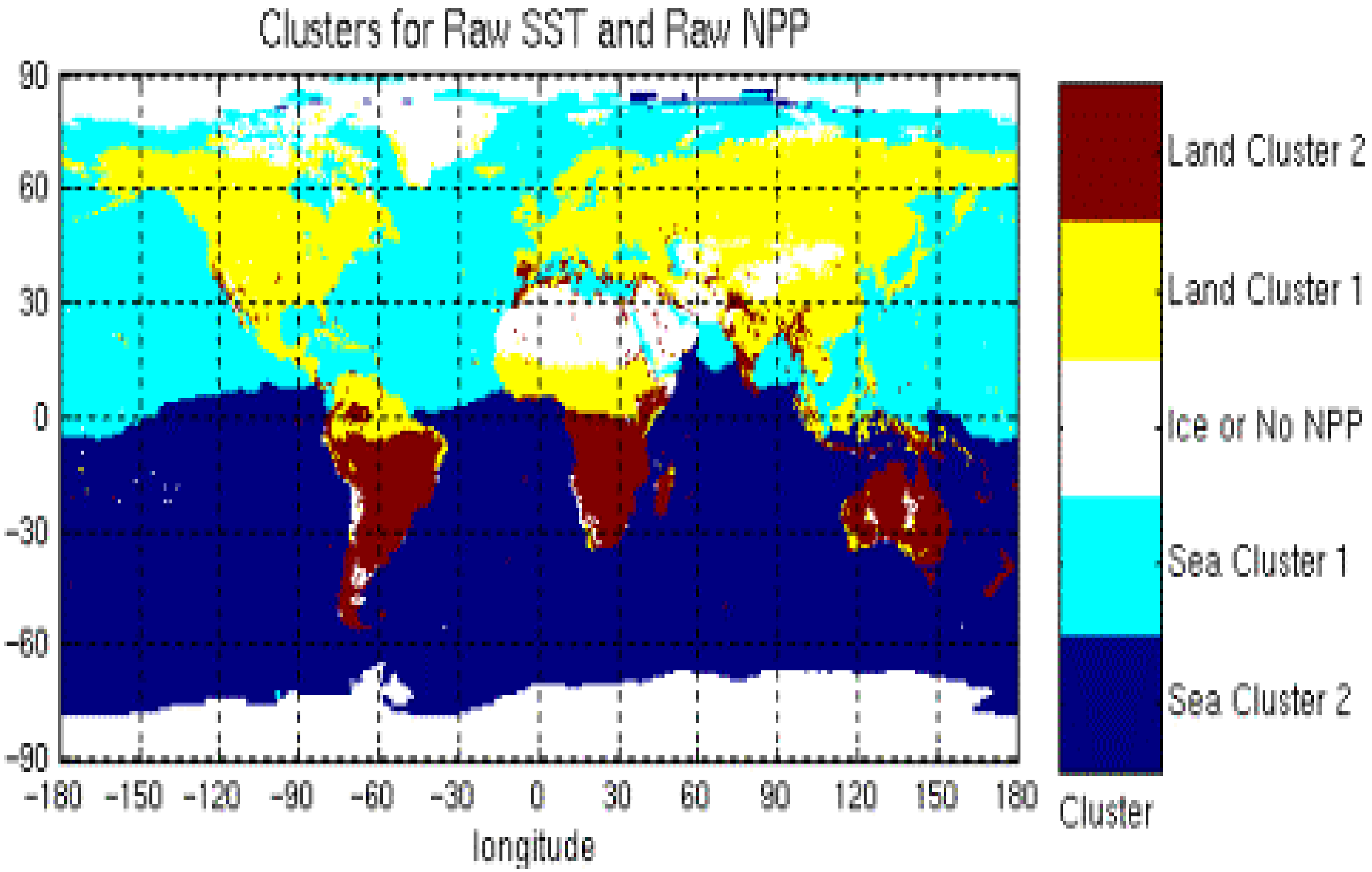
- Introduction
- Mining Large Scale Environmental Data
  - Preprocessing
  - Clustering
  - Association Rules
- Mining Medium and Low Scale Ecological Data
  - Co-location Mining
  - STAMM
  - PLUMS
  - Regression Trees, Neural Networks, Clustering
- Sensor Networks for Environmental Applications

Clustering for Zone Formation

- Interested in relationships between regions, not “points.”
- For land, clustering based on NPP or other variables, e.g., precipitation, temperature.
- For ocean, clustering based on SST (Sea Surface Temperature).
- When “raw” NPP and SST are used, clustering can find seasonal patterns.
  - Anomalous regions have plant growth patterns which reversed from those typically observed in the hemisphere in which they reside, and are easy to spot.

K-Means Clustering of Raw NPP and Raw SST

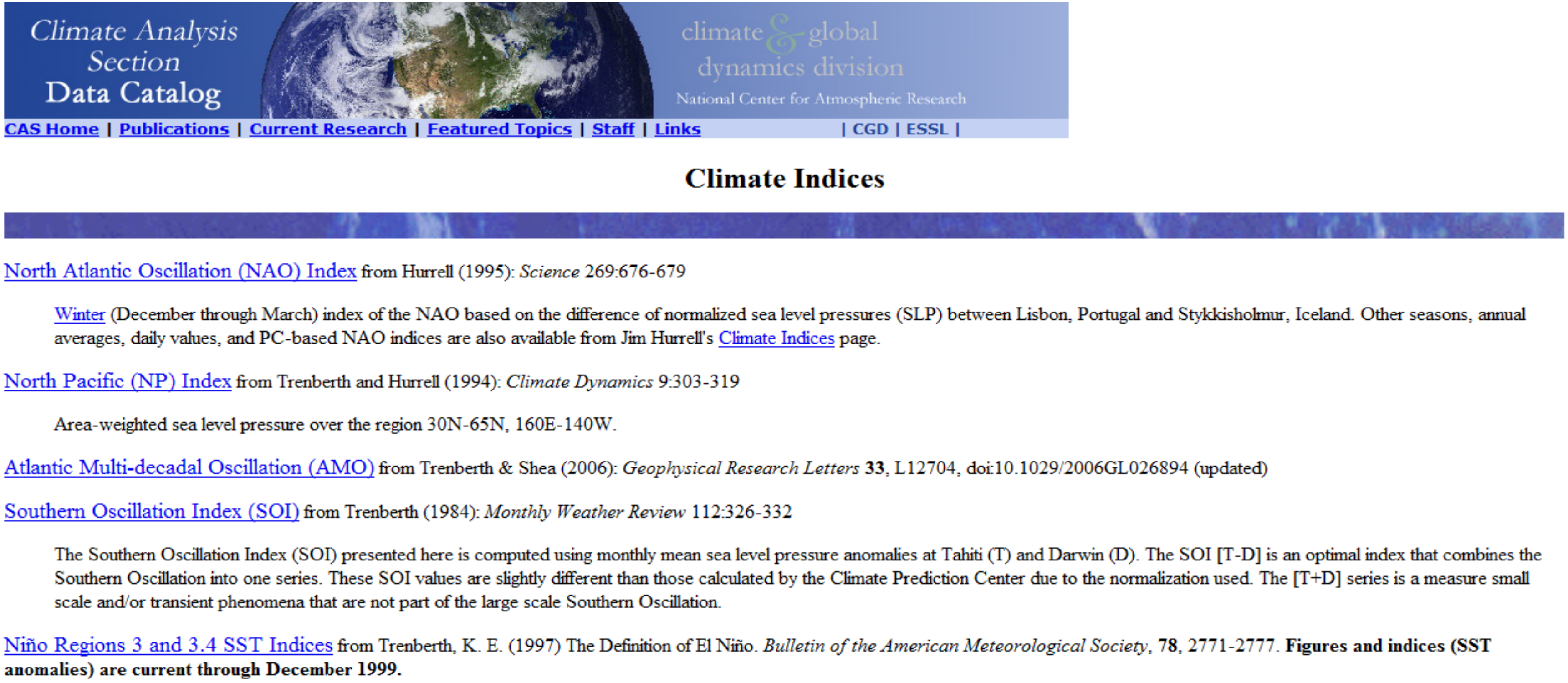
- K-Means, Number of Clusters=2



V. Kumar: Mining Earth Science Data (http://www-users.cs.umn.edu/~kumar/nasa-umn)

Climate Indices

- A Climate Index is a time series of SST, SLP etc
- Climate Indices capture teleconnections
  - The simultaneous variation in climate related processes over widely separated points on Earth


Climate Analysis Section Data Catalog
climate & global dynamics division
National Center for Atmospheric Research
CAS Home | Publications | Current Research | Featured Topics | Staff | Links
| CGD | ESSL |

Climate Indices

North Atlantic Oscillation (NAO) Index from Hurrell (1995): *Science* 269:676-679

Winter (December through March) index of the NAO based on the difference of normalized sea level pressures (SLP) between Lisbon, Portugal and Stykkisholmur, Iceland. Other seasons, annual averages, daily values, and PC-based NAO indices are also available from Jim Hurrell's Climate Indices page.

North Pacific (NP) Index from Trenberth and Hurrell (1994): *Climate Dynamics* 9:303-319

Area-weighted sea level pressure over the region 30N-65N, 160E-140W.

Atlantic Multi-decadal Oscillation (AMO) from Trenberth & Shea (2006): *Geophysical Research Letters* **33**, L12704, doi:10.1029/2006GL026894 (updated)

Southern Oscillation Index (SOI) from Trenberth (1984): *Monthly Weather Review* 112:326-332

The Southern Oscillation Index (SOI) presented here is computed using monthly mean sea level pressure anomalies at Tahiti (T) and Darwin (D). The SOI [T-D] is an optimal index that combines the Southern Oscillation into one series. These SOI values are slightly different than those calculated by the Climate Prediction Center due to the normalization used. The [T+D] series is a measure small scale and/or transient phenomena that are not part of the large scale Southern Oscillation.

Niño Regions 3 and 3.4 SST Indices from Trenberth, K. E. (1997) The Definition of El Niño. *Bulletin of the American Meteorological Society*, 78, 2771-2777. **Figures and indices (SST anomalies) are current through December 1999.**

- http://www.cgd.ucar.edu/cas/catalog/climind/

Shared Nearest Neighbor (SNN) Clustering

- Find the nearest neighbors of each data point.
  - In this case data points are time series.
  - Redefine the similarity between pairs of points in terms of how many nearest neighbors the two points share.
- Calculate the density at each point by summing the similarities of its nearest neighbors.
- Identify and eliminate noise and outliers, which are points with low density.
- Identify core points, which are points with high density.
- Build clusters around the core points.

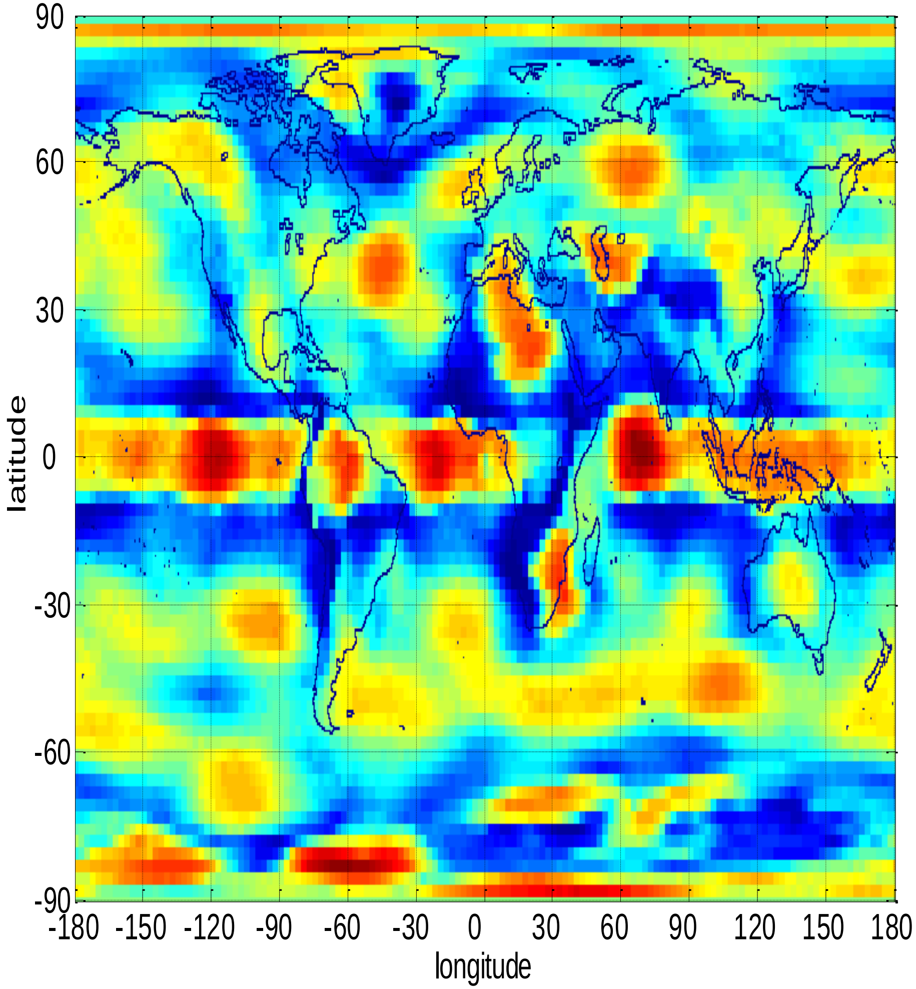
SNN Clustering - Advantages

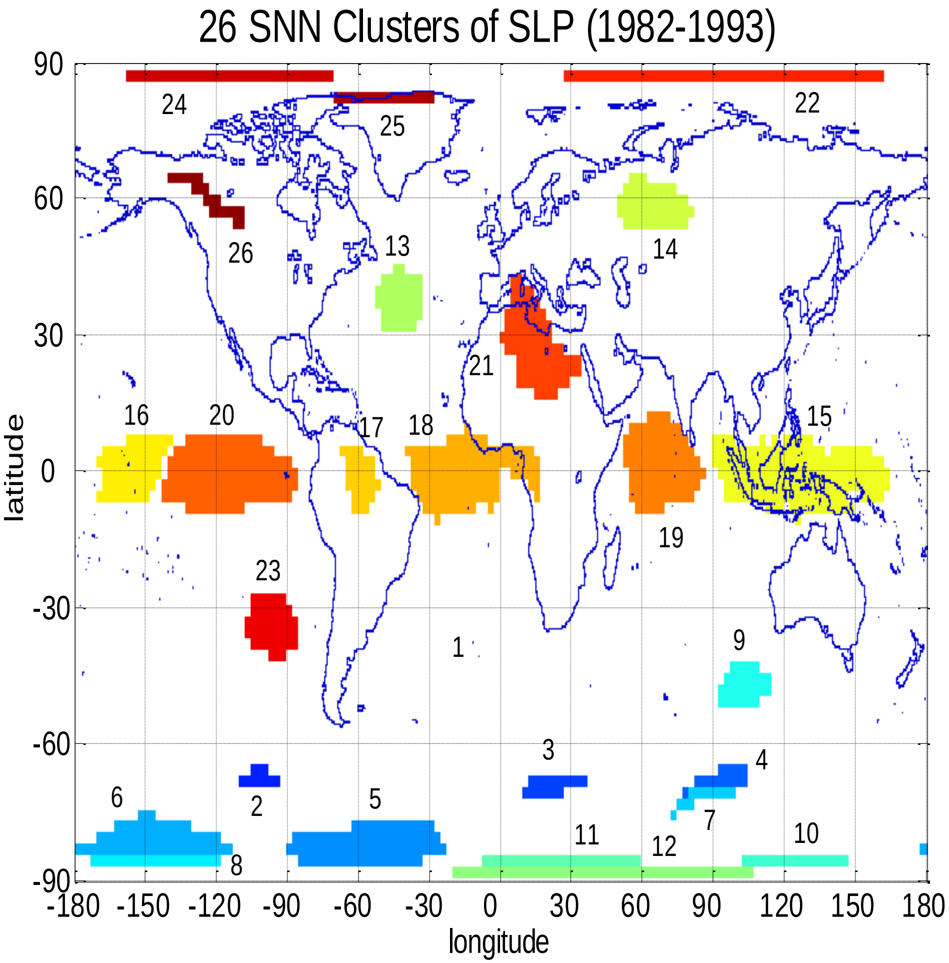
- The use of a shared nearest neighbor definition of similarity removes problems with varying density, while the use of core points handles problems with shape and size.

- Finding clusters of different shapes and sizes, especially in the presence of noise is a difficult clustering problem.
  - Earth Science data is noisy

- Find the number of clusters automatically.

SLP Clusters

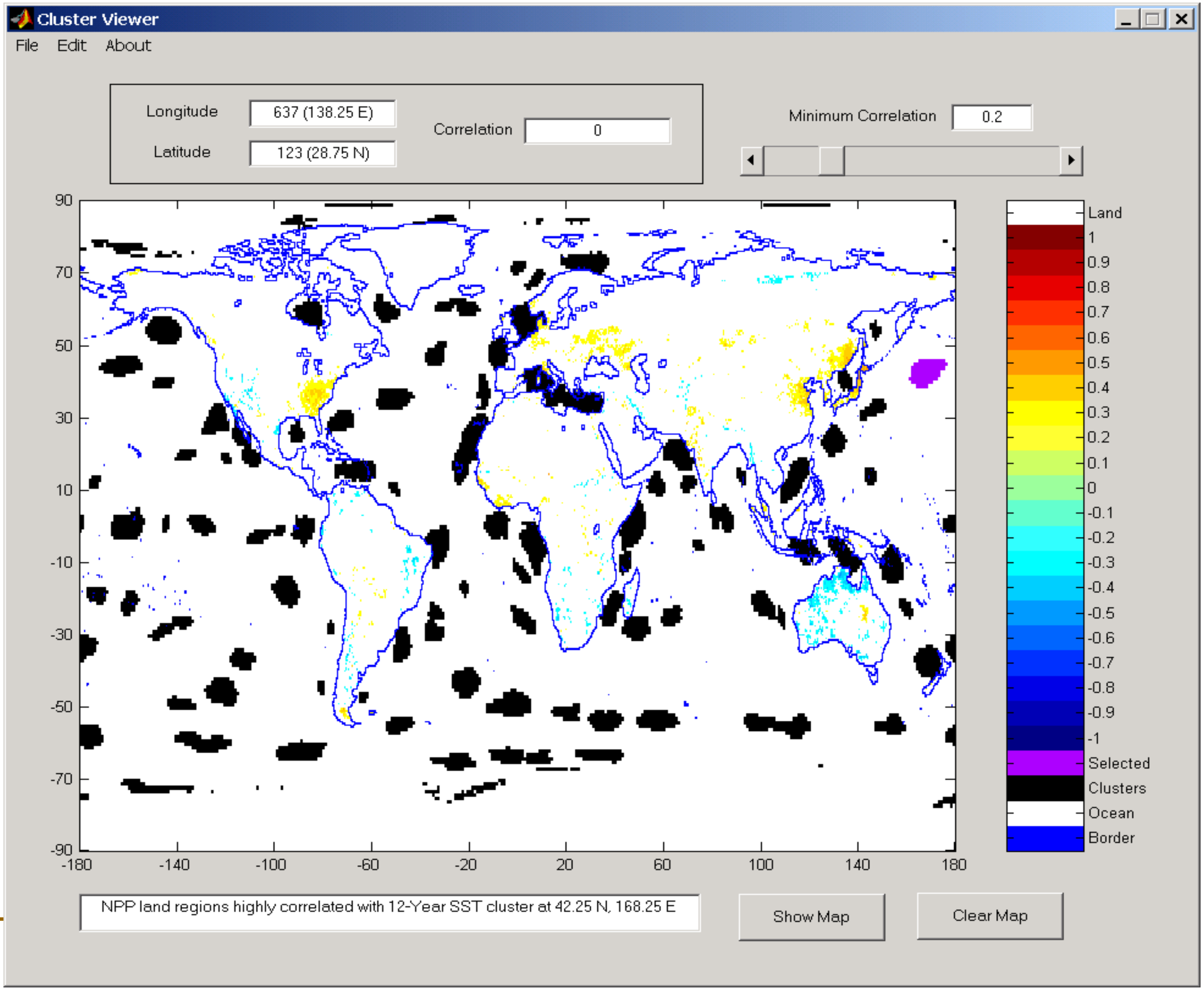




V. Kumar: Mining Earth Science Data (http://www-users.cs.umn.edu/~kumar/nasa-umn)

Teleconnections (1/3)

Cluster viewer showing land regions with positive or negative correlation > 0.2 with highlighted ocean cluster.



Cluster Viewer

Longitude: 637 (138.25 E)

Latitude: 123 (28.75 N)

Correlation: 0

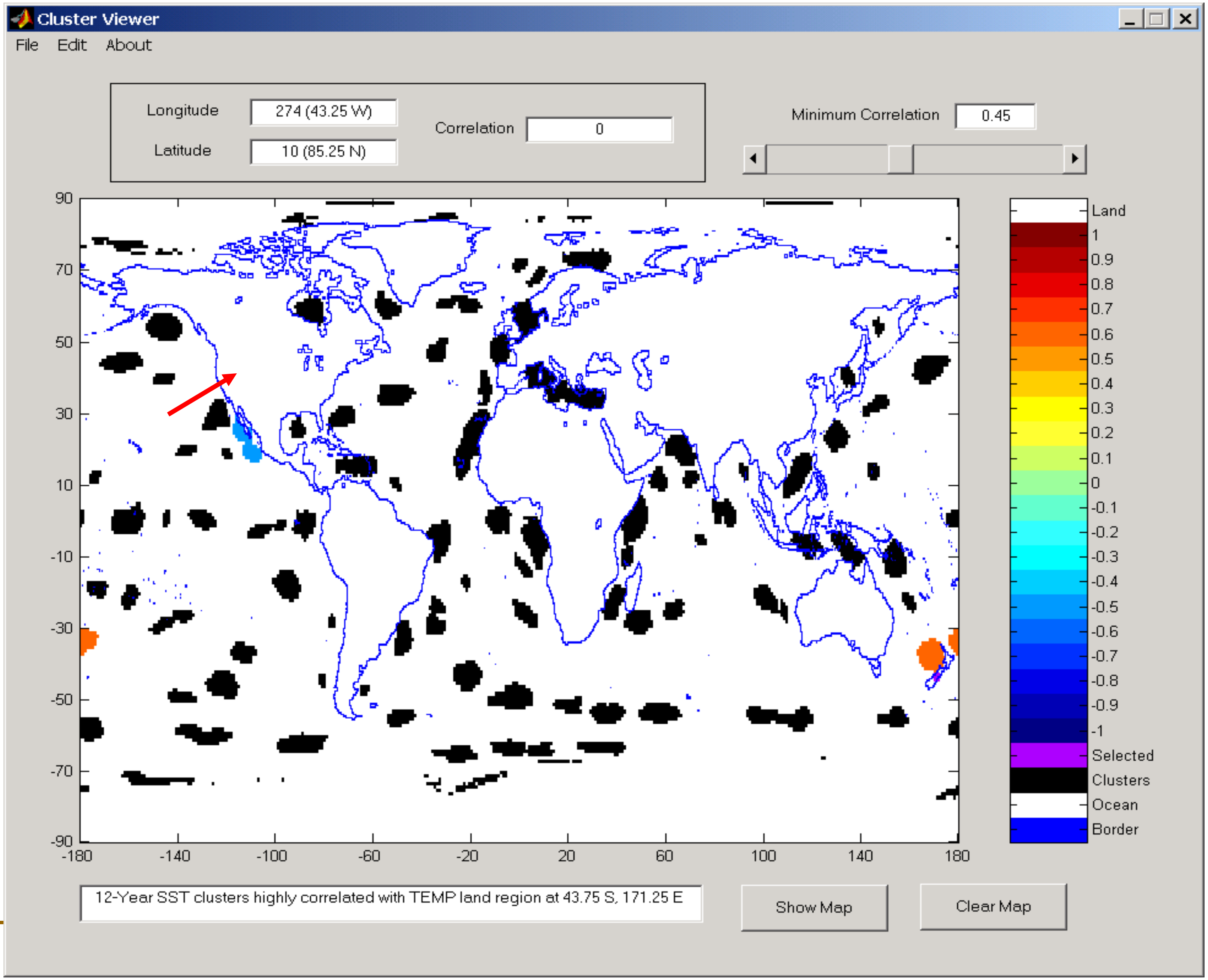
Minimum Correlation: 0.2

NPP land regions highly correlated with 12-Year SST cluster at 42.25 N, 168.25 E

Teleconnections (2/3)

Cluster viewer showing clusters correlated (> 0.45) to a New Zealand land point)

Notice cluster off the coast of western Mexico, which is negatively correlated.



Longitude: 274 (43.25 W)

Latitude: 10 (85.25 N)

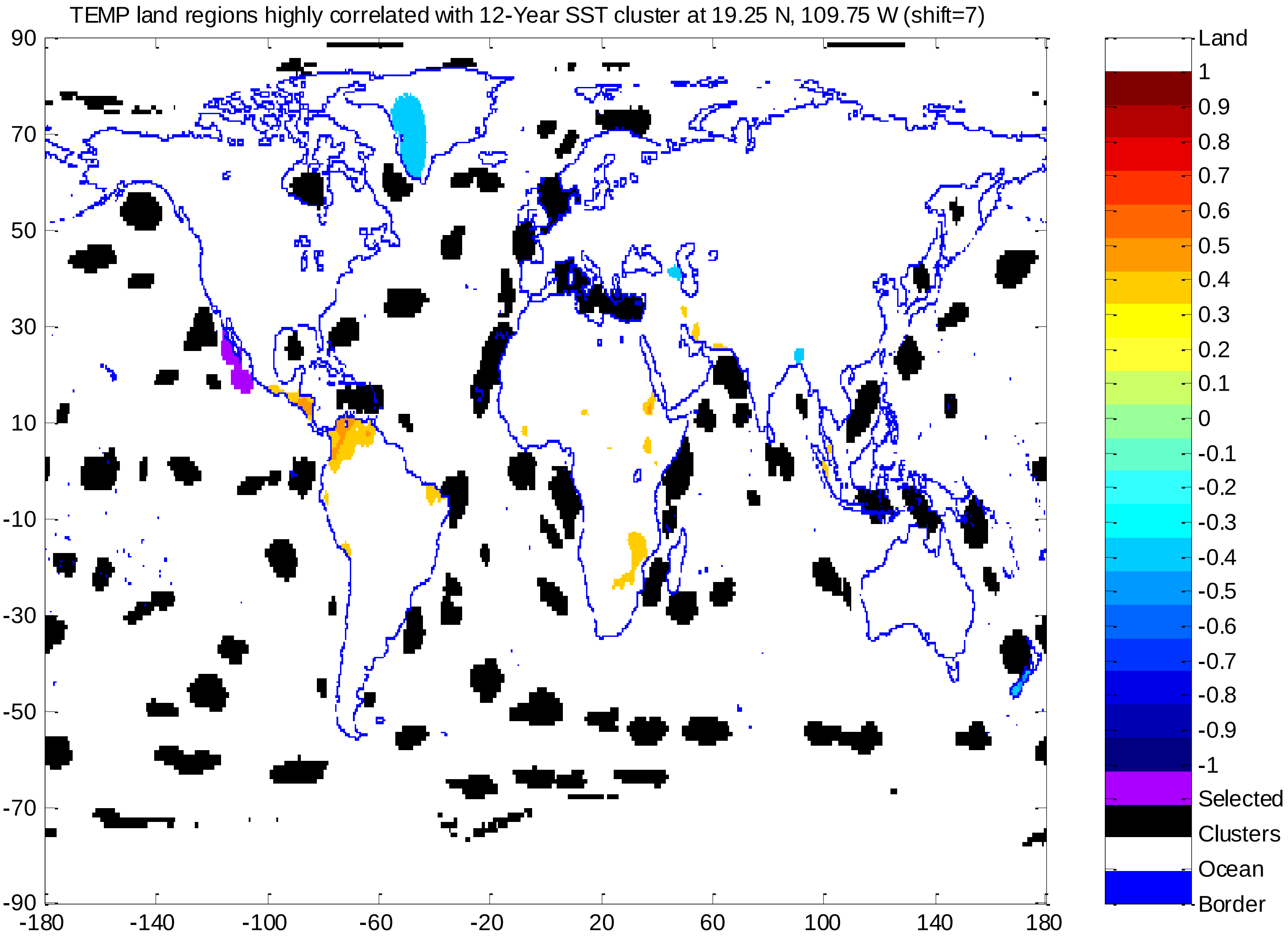
Correlation: 0

Minimum Correlation: 0.45

12-Year SST clusters highly correlated with TEMP land region at 43.75 S, 171.25 E

Teleconnections (3/3)

Cluster viewer showing land points (Temp) correlated (> 0.34) to a cluster off the coast of western Mexico.



Outline

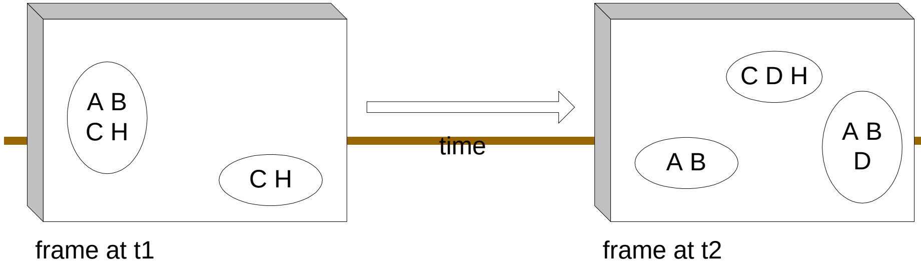
- Introduction
- Mining Large Scale Environmental Data
  - Preprocessing
  - Clustering
  - Association Rules ←
- Mining Medium and Low Scale Ecological Data
  - Co-location Mining
  - STAMM
  - PLUMS
  - Regression Trees, Neural Networks, Clustering
- Sensor Networks for Environmental Applications

Types of Spatio-Temporal Association Patterns

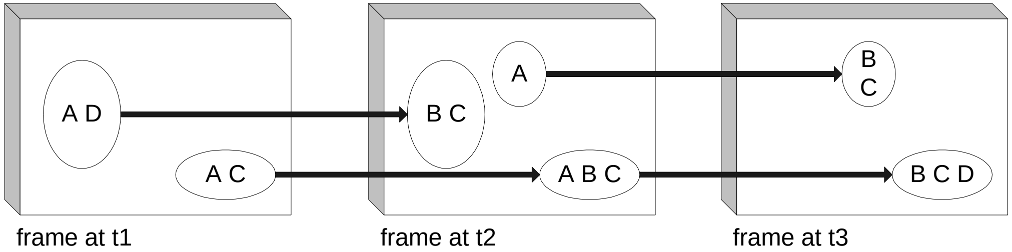
Type of Pattern	Description
Intra-zone non-sequential	relationships among events in the same grid cell or zone, ignoring the temporal aspects of the data.
Intra-zone sequential	temporal relationships among events occurring within the same grid cell or zone.
Inter-zone non-sequential	relationships among events happening in different grid cells or zones, ignoring temporal aspects of the data.
Inter-zone sequential	temporal relationships among events occurring at different spatial locations.

V. Kumar: Mining Earth Science Data (http://www-users.cs.umn.edu/~kumar/nasa-umn)

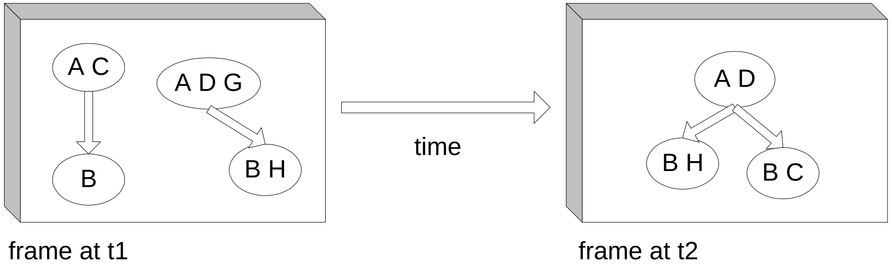
Types of Spatio-temporal Patterns



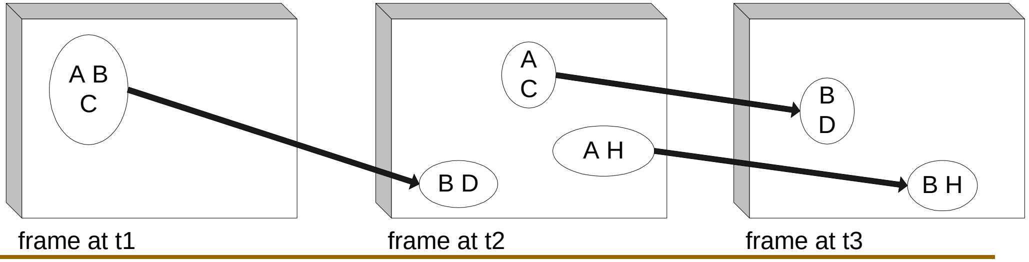
(a) Intra-zone non-sequential (e.g. {A,B}, {C,H})



(b) Intra-zone sequential (e.g. A ==> C)



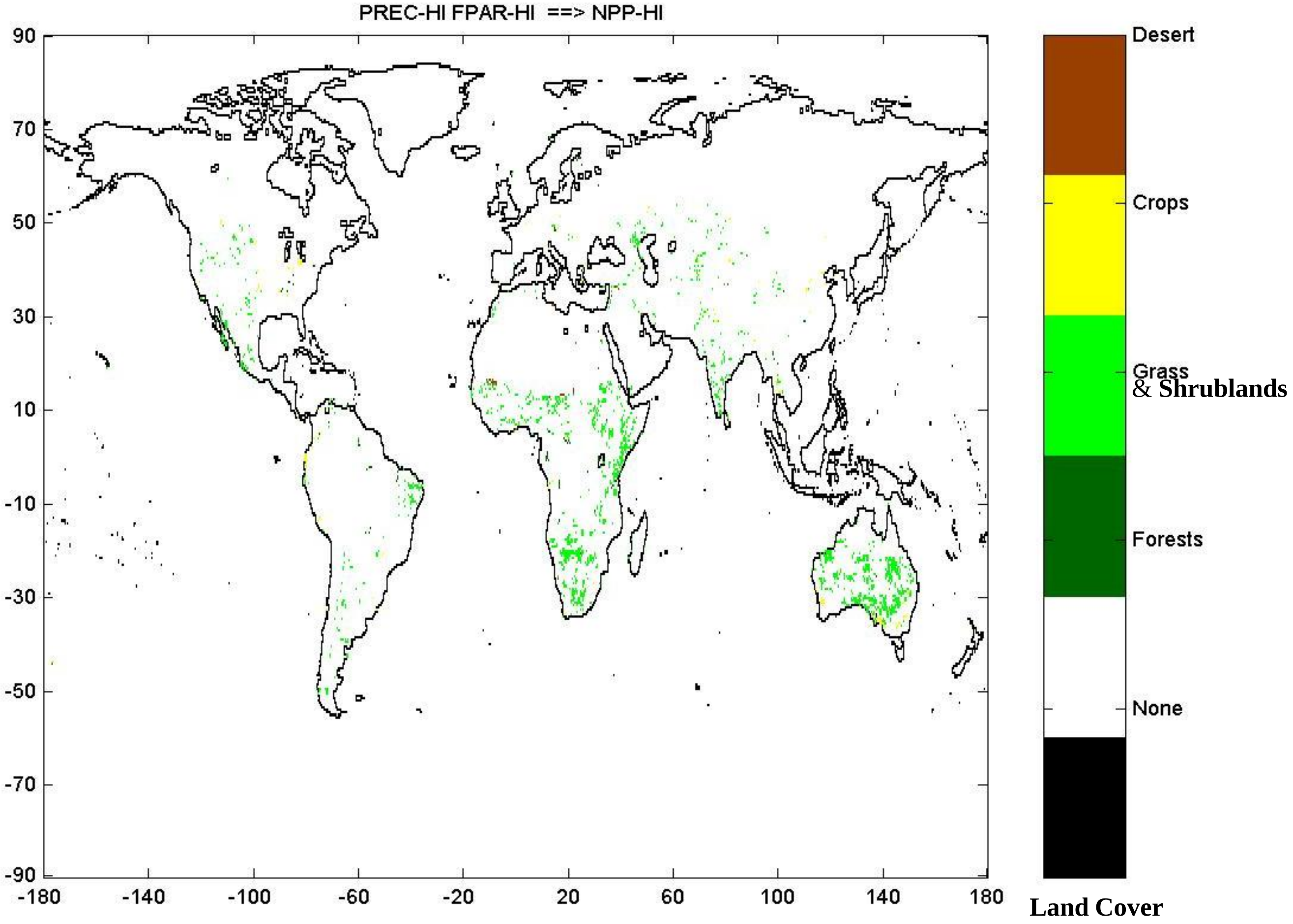
(c) Inter-zone non-sequential (e.g. B to the south of A)



(d) Inter-zone sequential (e.g. B to the south of A in the future)

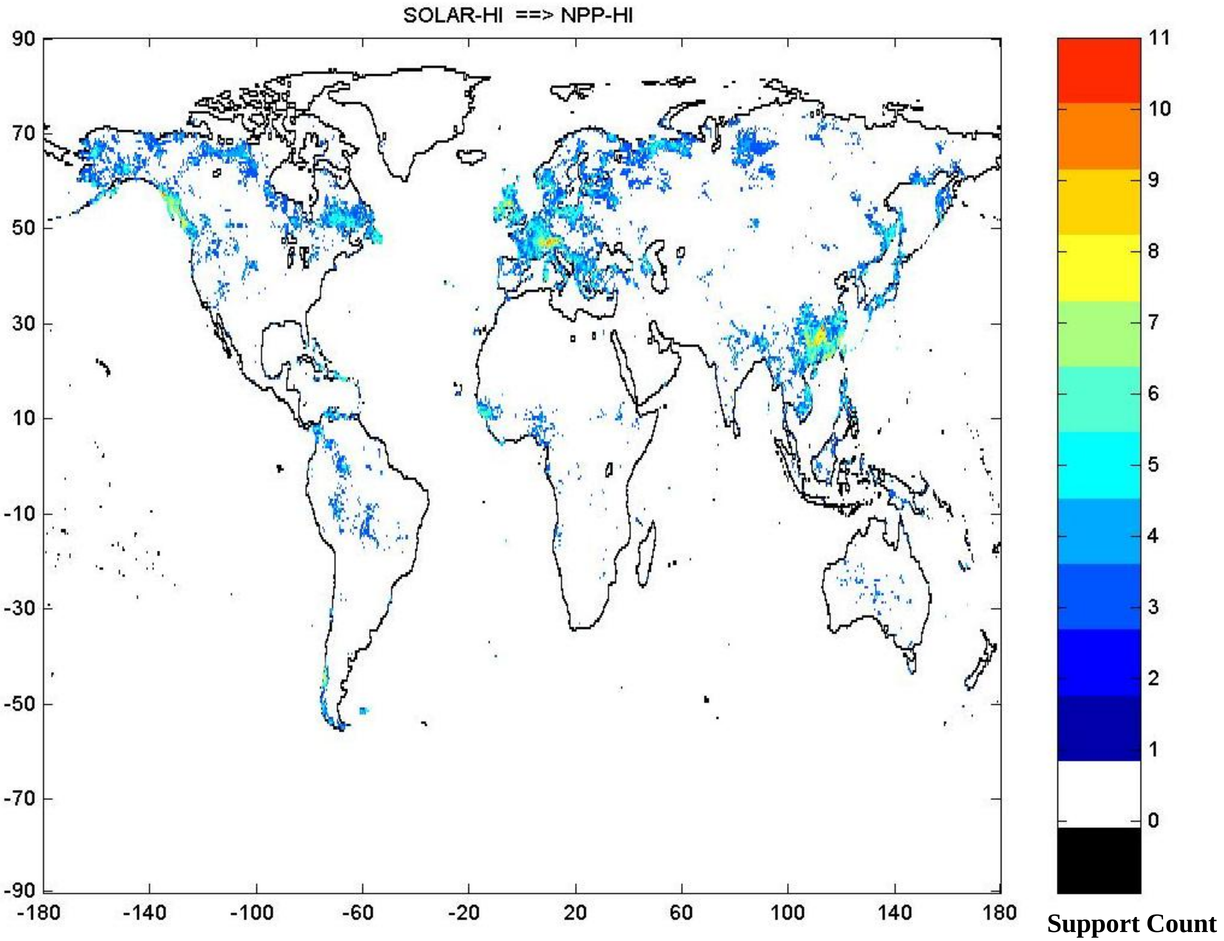
V. Kumar: Mining Earth Science Data (http://www-users.cs.umn.edu/~kumar/nasa-umn)

Intra-zone non-sequential Patterns (1/4)



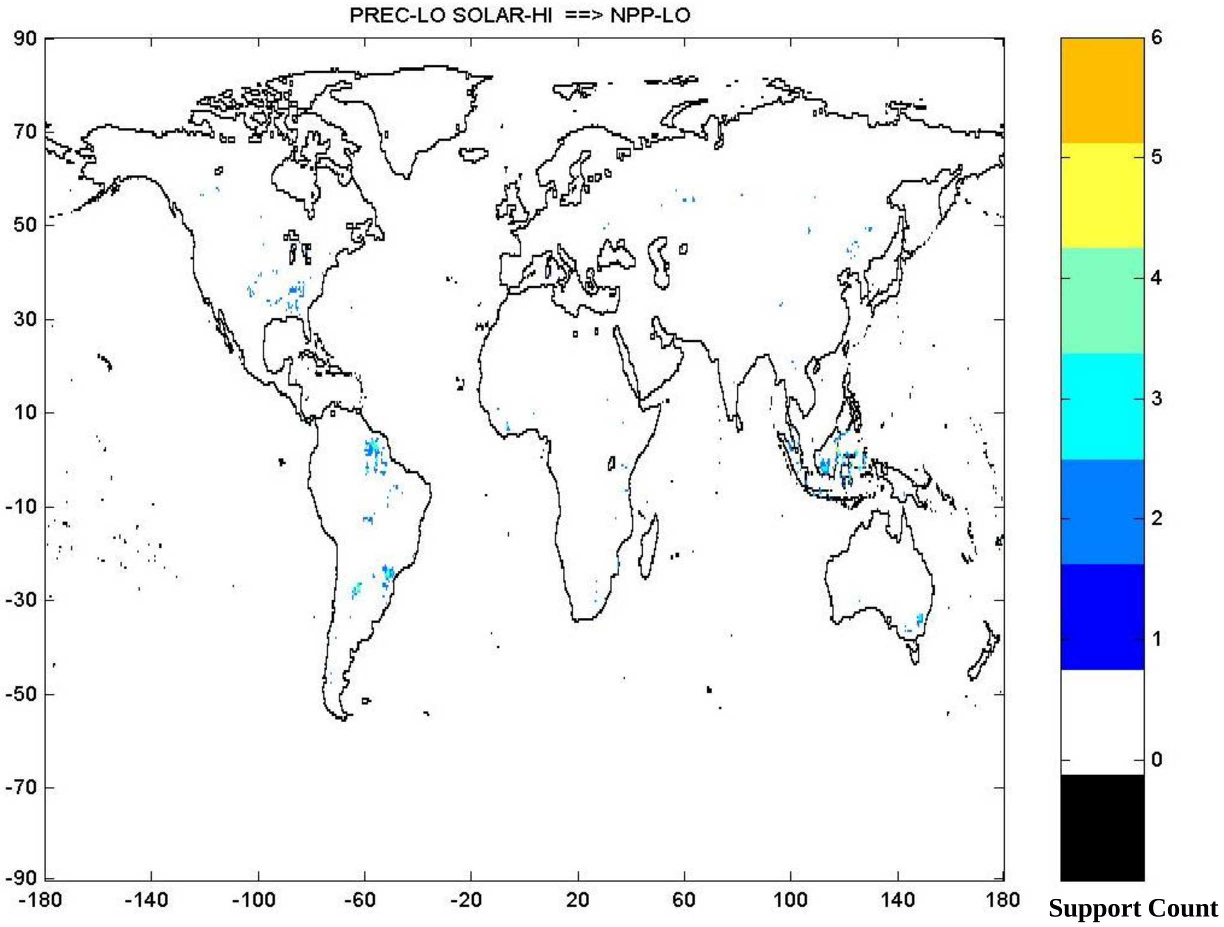
- Map agrees with hypothesis that Prec-Hi Fpar-Hi → NPP-Hi occurs mostly in shrubland and other type of grassland regions (support ≥ 3).

Intra-zone non-sequential Patterns (2/4)



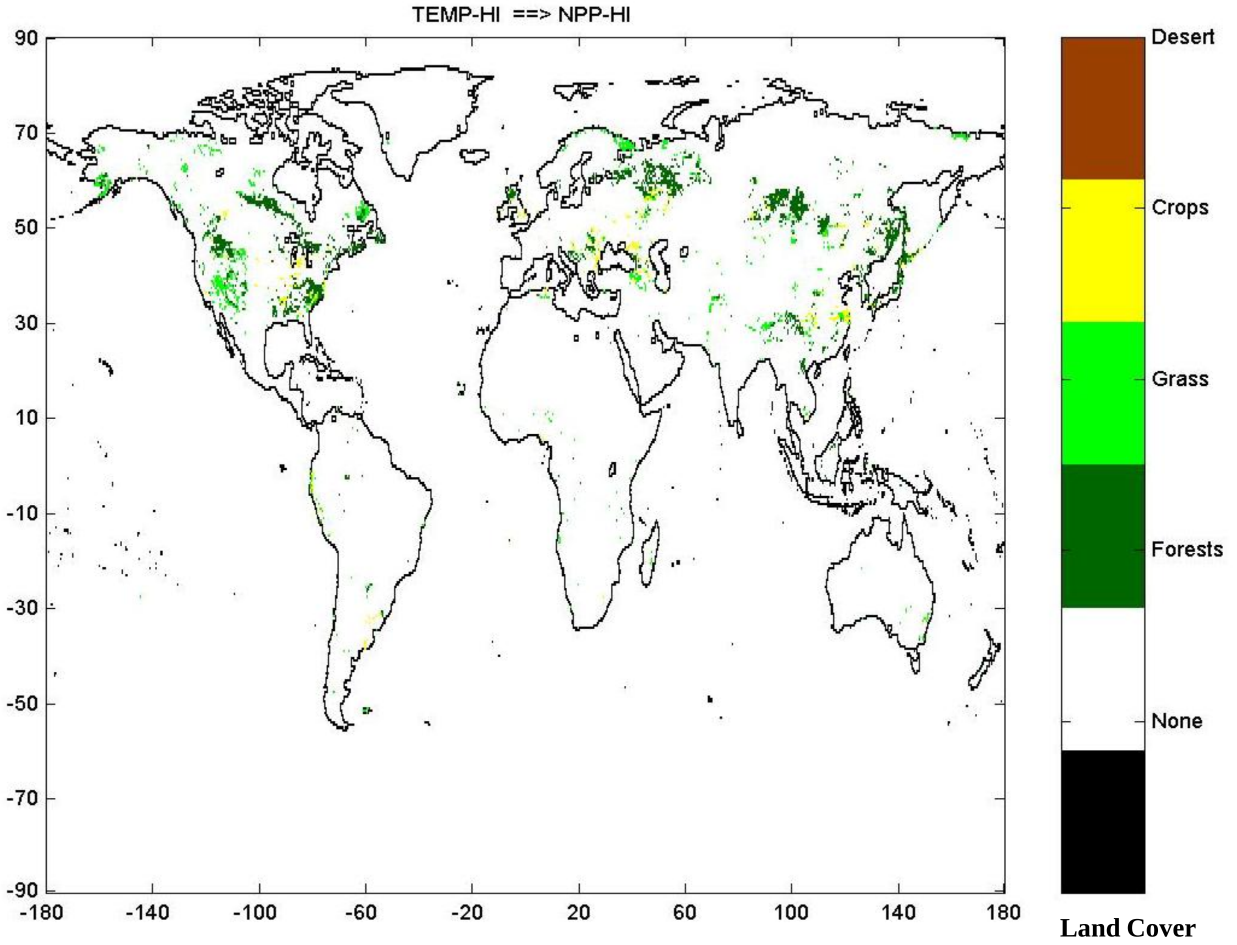
- Solar-Hi → NPP-Hi tends to occur in very cloudy (light limited) areas, like the Pacific NW and Canada/Alaska (support ≥ 3).

Intra-zone non-sequential Patterns (3/4)



- Prec-Lo Solar-Hi → NPP-Lo tends to occur in drought-prone areas of tropical and sub-tropical zones, and areas of major forest fires (support ≥ 2).

Intra-zone non-sequential Patterns (4/4)



- Temp-Hi → NPP-Hi tends to occur in the forest regions of the northern hemisphere (support ≥ 4).

V. Kumar: Mining Earth Science Data (http://www-users.cs.umn.edu/~kumar/nasa-umn)

Outline

- Introduction
- Mining Large Scale Environmental Data
  - Preprocessing
  - Clustering
  - Association Rules
- Mining Medium and Low Scale Ecological Data
  - Co-location Mining
  - STAMM
  - PLUMS
  - Regression Trees, Neural Networks, Clustering
- Sensor Networks for Environmental Applications

Co-Location Mining (1/5)

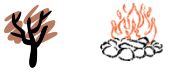
- Introduced in S. Shekar et al, SSTD 2001
- Problem: Given a set of boolean spatial features
  - find subsets of co-located features, e.g. (fire, drought, vegetation)
  - Data - continuous space, partition not natural, no reference feature
- Classical data mining approach: association rules
  - But, Look Ma! No Transactions!!! No support measure!
- Approach: Work with continuous data without transactionizing it!
- Co-location patterns may reveal
  - Hunter – Chase relationships between species
  - Set of required conditions for certain kinds of species to breed
  - Correlation between the presence of certain pollutants and human, animal deceases

S. Shekar: www.cs.umn.edu/~shekhar/talk/**ucgis.ppt**

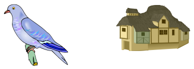
Co-Location Mining (2/5)

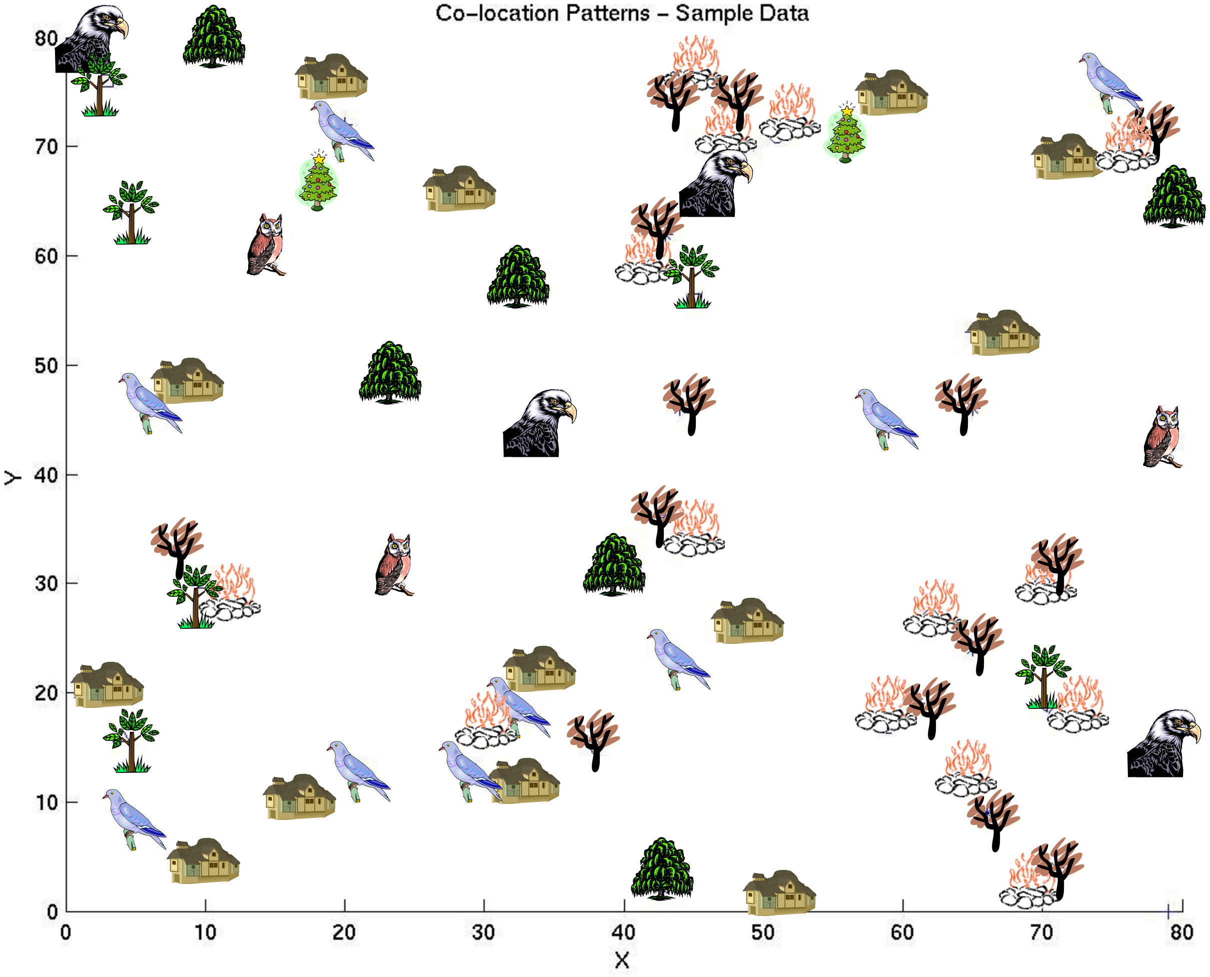
- ❑ Can you notice co-location patterns from the following sample dataset?

Answers:



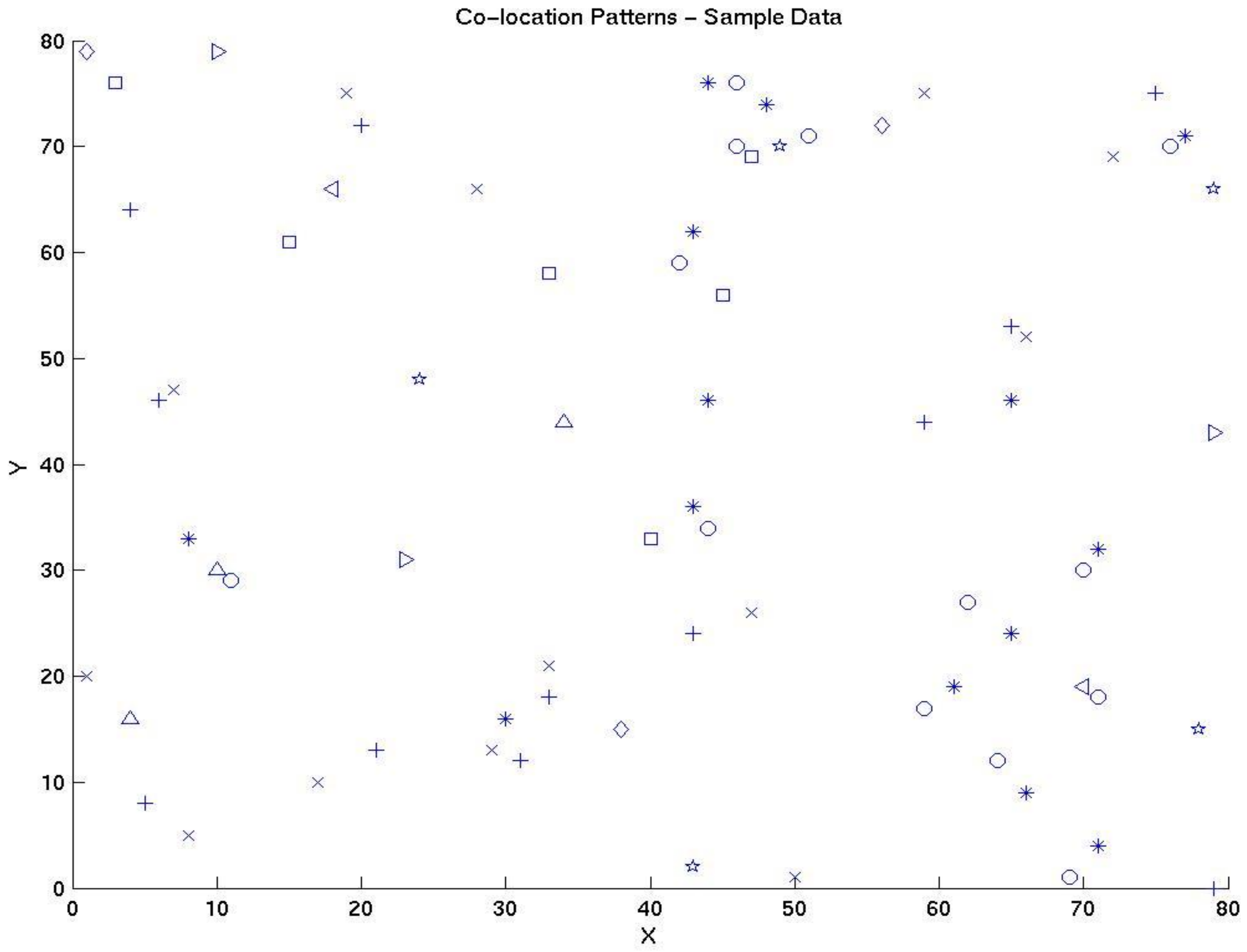
and





Co-Location Mining (3/5)

- Can you find co-location patterns from the following sample dataset?



Co-Location Mining (4/5)

Spatial Co-location

A set of features frequently co-located

Given

A set T of K boolean spatial feature types $T=\{f_1,f_2, \ldots , f_k\}$

A set P of N locations $P=\{p_1, \ldots, p_N\}$ in a spatial frame work S, $p_i \in P$ is a vector <instance-id, feature type,loc>

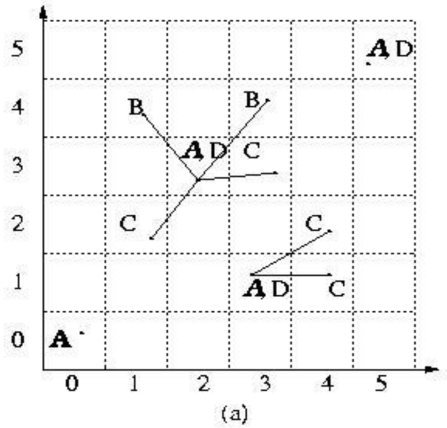
A neighbor relation R over locations in S

A min-prevalence threshold and a

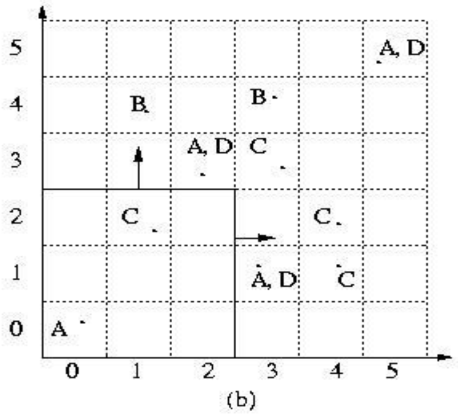
min conditional probability threshold

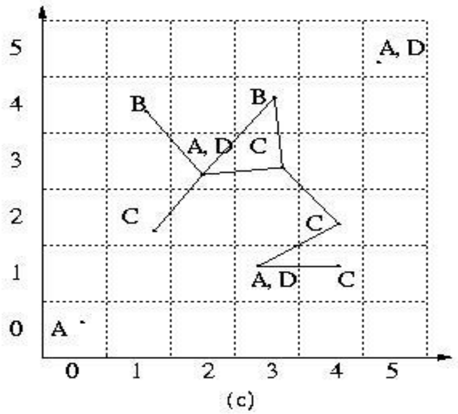
Find

$T_c = \cup$subsets of T frequently co-located



Reference Feature Centric



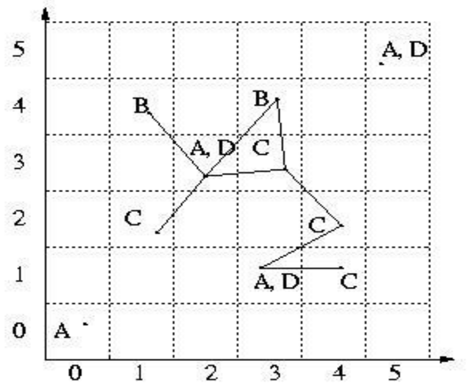


Window Centric

Event Centric

S. Shekar: www.cs.umn.edu/~shekhar/talk/**ucgis.ppt**

Co-Location Mining (5/5)



(c)

Initial Records

Instance ID	Location	Feature Type
1	(0,0)	A
2	(1,2)	C
…	…	…

Table instance of co-location {A,C}

A	C
(3,1)	(4,1)
(3,1)	(4,2)
(2,3)	(1,2)
(2,3)	(3,3)

Participation index

Participation ratio pr(f_i, c) of feature f_i in co-location c = {f_1, f_2, …, f_k}: fraction of instances of f_i with feature {f_1, …, f_{i-1}, f_{i+1}, …, f_k} nearby

Participation index (c)= Πpr(f_i, c)}

Conditional Probability

Pr.[{c2} in N(L) | c1 at L]

Algorithm

Hybrid Co-location Miner

Participation ratio pr(C, {A,C})=4/4=1 (i.e 4 (distinct) out of 4 instances of C participate in {A,C})

Participation ratio pr(A, {A,C})=2/4

Outline

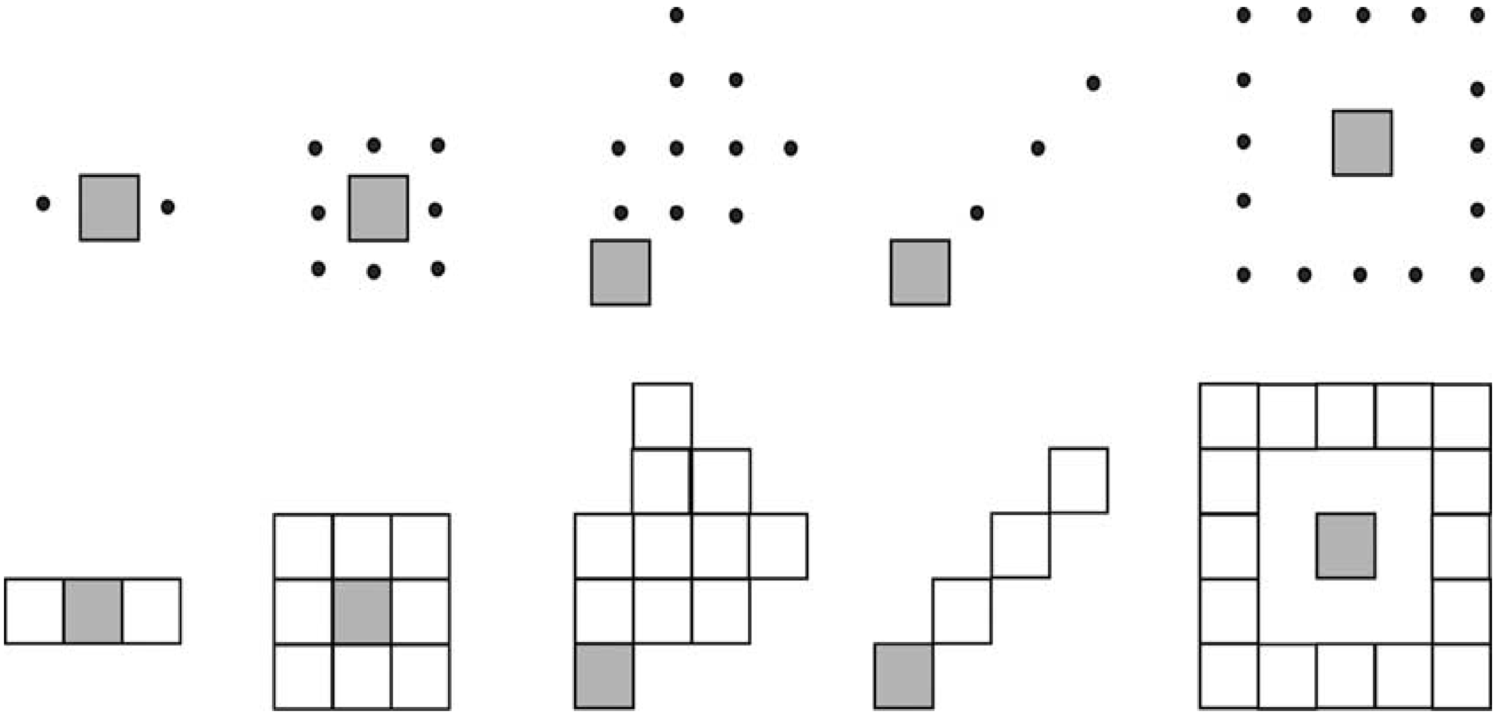
- Introduction
- Mining Large Scale Environmental Data
  - Preprocessing
  - Clustering
  - Association Rules
- Mining Medium and Low Scale Ecological Data
  - Co-location Mining
  - STAMM
  - PLUMS
  - Regression Trees, Neural Networks, Clustering
- Sensor Networks for Environmental Applications

STAMM (1/4)

- Citation: Su et al, Ecological Modeling 174 (2004) 421-431
- Problem: Identify the effect of environmental factors on the behavior of living organisms
  - Select a set of environmental factors under study
  - Place values to corresponding cell of a grid
  - The behavior of a living organism is considered as an ecological event (EE)
- Approach: Built a Spatiotemporal Assignment Mining Model
  - Extracts Ecological Association Rules (EARs)
  - EE always appears at the left side of the rule
  - Use of Apriori algorithm to extract EARs

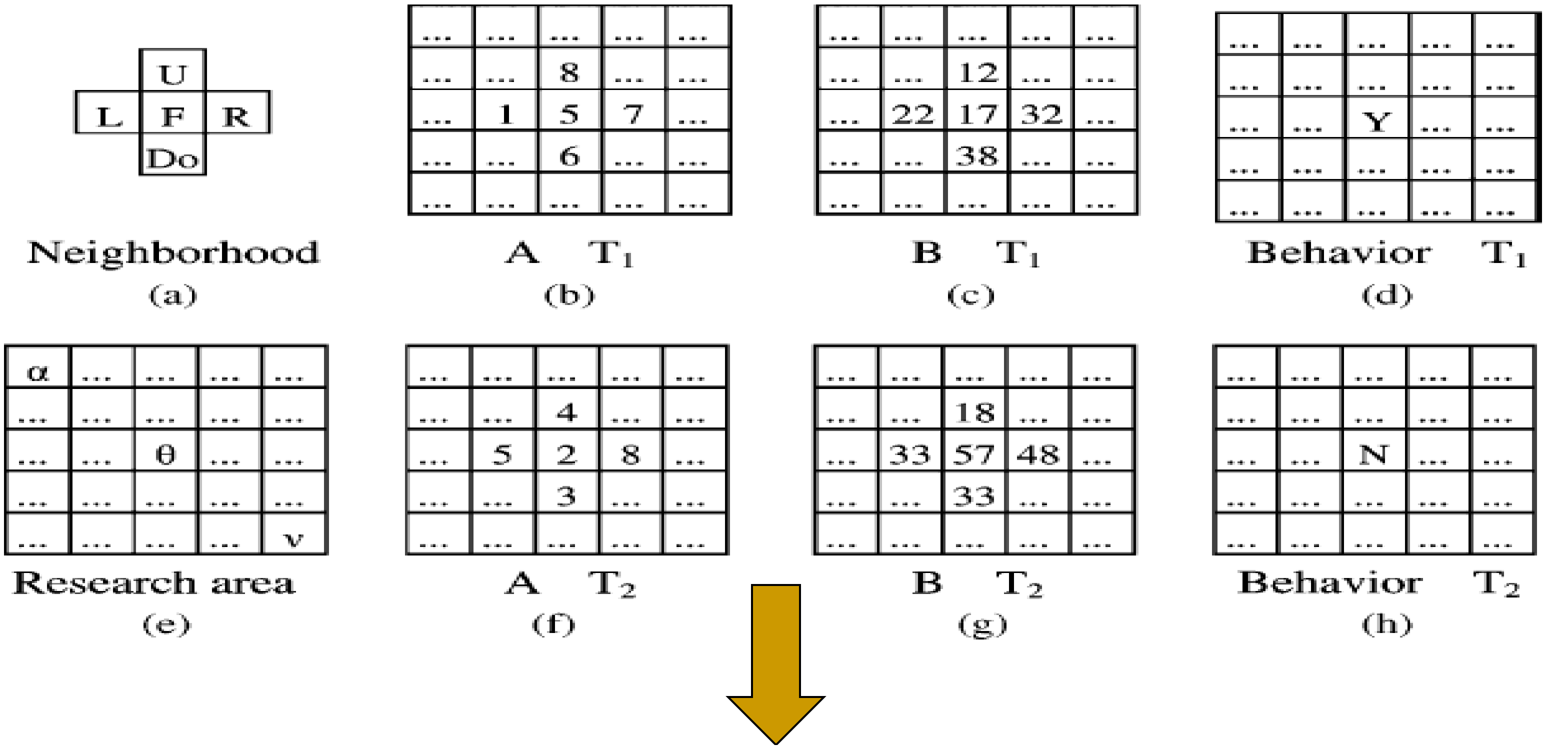
STAMM (2/4)

- Step 1: Place a grid on the area of study
- Step 2: For each focus cell identify its neighborhood based on prior knowledge



STAMM (3/4)

- Step 3: Construct Ecological Decision Table (EDT)



	U	
L	F	R
	Do	

Neighborhood
(a)

...	...	...	...	...
...	...	8	...	...
...	1	5	7	...
...	...	6	...	...
...	...	...	...	...

A T$_1$
(b)

...	...	...	...	...
...	...	12	...	...
...	22	17	32	...
...	...	38	...	...
...	...	...	...	...

B T$_1$
(c)

...	...	...	...	...
...	...	...	...	...
...	...	Y	...	...
...	...	...	...	...
...	...	...	...	...

Behavior T$_1$
(d)

α	...	...	...	...
...	...	...	...	...
...	...	θ	...	...
...	...	...	...	...
...	...	...	...	ν

Research area
(e)

...	...	...	...	...
...	...	4	...	...
...	5	2	8	...
...	...	3	...	...
...	...	...	...	...

A T$_2$
(f)

...	...	...	...	...
...	...	18	...	...
...	33	57	48	...
...	...	33	...	...
...	...	...	...	...

B T$_2$
(g)

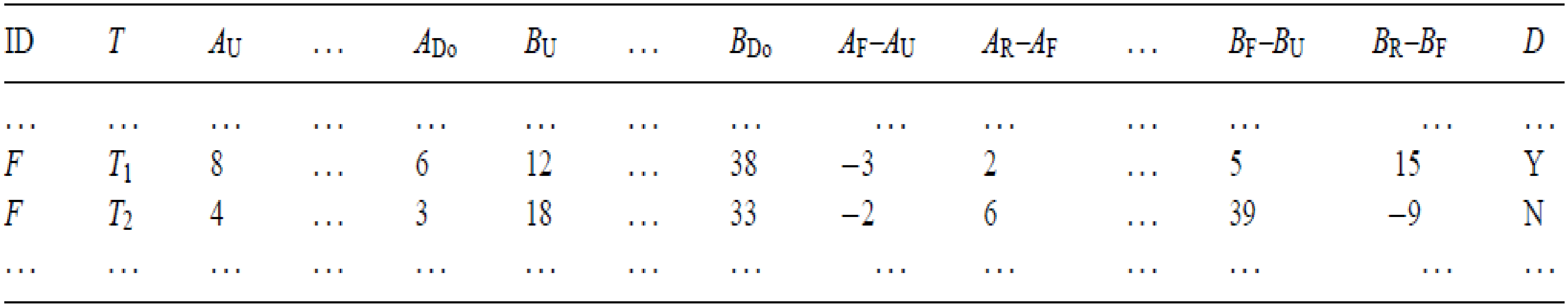
...	...	...	...	...
...	...	...	...	...
...	...	N	...	...
...	...	...	...	...
...	...	...	...	...

Behavior T$_2$
(h)

ID	T	A_U	A_L	A_F	A_R	A_{Do}	B_U	B_L	B_F	B_R	B_{Do}	D
...	...	...	...	...	...	...	...	...	...	...	...	...
θ	T_1	8	1	5	7	6	12	22	17	32	38	Y
θ	T_2	4	5	2	8	3	18	33	57	48	33	N
...	...	...	...	...	...	...	...	...	...	...	...	...

STAMM (4/4)

- We may include relative spatial relations of environmental factors



ID	T	A_U	…	A_{Do}	B_U	…	B_{Do}	$A_F - A_U$	$A_R - A_F$	…	$B_F - B_U$	$B_R - B_F$	D
…	…	…	…	…	…	…	…	…	…	…	…	…	…
F	T_1	8	…	6	12	…	38	−3	2	…	5	15	Y
F	T_2	4	…	3	18	…	33	−2	6	…	39	−9	N
…	…	…	…	…	…	…	…	…	…	…	…	…	…

- Step 4: Run Apriori on EDT
- Obtain EARs of the form
  - $(ID, \theta) \wedge (T, t_1) \wedge (A_U, 8) \wedge (A_F, 5) \wedge (B_U, 12) \wedge (B_F, 17) \wedge (B_R - B_F, 15) \rightarrow (D, Y)$
  - *(chlorophyll,R) ∧ (Position,A) → (fish, assembling)*

Outline

- Introduction
- Mining Large Scale Environmental Data
  - Preprocessing
  - Clustering
  - Association Rules
- Mining Medium and Low Scale Ecological Data
  - Co-location Mining
  - STAMM
  - PLUMS
  - Regression Trees, Neural Networks, Clustering
- Sensor Networks for Environmental Applications

Location Prediction – PLUMS (1/3)

- Problem: predict nesting site in marshes
  - given vegetation, water depth, distance to edge, etc.
- Data - maps of nests and attributes
  - spatially clustered nests, spatially smooth attributes
- Classical method: logistic regression, decision trees, bayesian classifier
  - but, independence assumption is violated ! Misses auto-correlation !
  - Spatial auto-regression (SAR), Markov random field bayesian classifier
  - Open issues: spatial accuracy vs. classification accurary
  - Open issue: performance - SAR learning is slow!

Location Prediction – PLUMS (2/3)

Given:

1. Spatial Framework $S = \{s_1, \ldots s_n\}$

2. Explanatory functions: $f_{X_k} : S \to R$

3. A dependent function $f_Y : S \to \{0,1\}$

4. A family $\mathfrak{I}$ **of function mappings:**

$$R \times \ldots \times R \to \{0,1\}$$

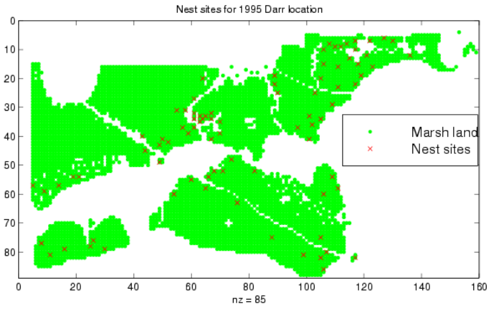
Find: A function $\hat{f}_y \in \mathfrak{I}$

Objective:maximize

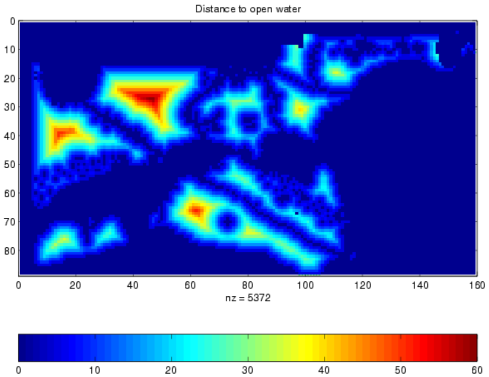
classification_accuracy $(\hat{f}_y, f_y)$

Constraints:

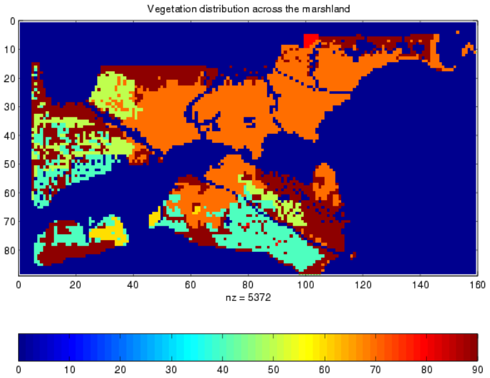
Spatial Autocorrelation exists



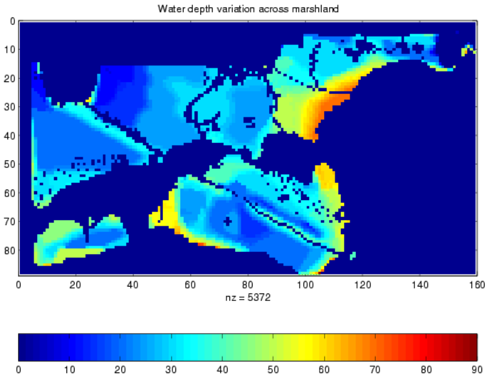
Nest locations



Distance to open water



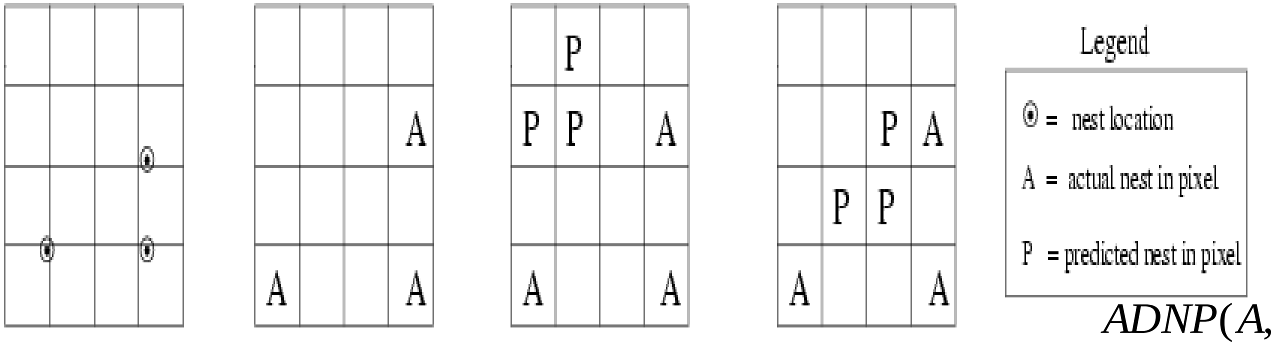
Vegetation durability



Water depth

S. Shekar: www.cs.umn.edu/~shekhar/talk/**ucgis.ppt**

Location Prediction – PLUMS (3/3)



Legend

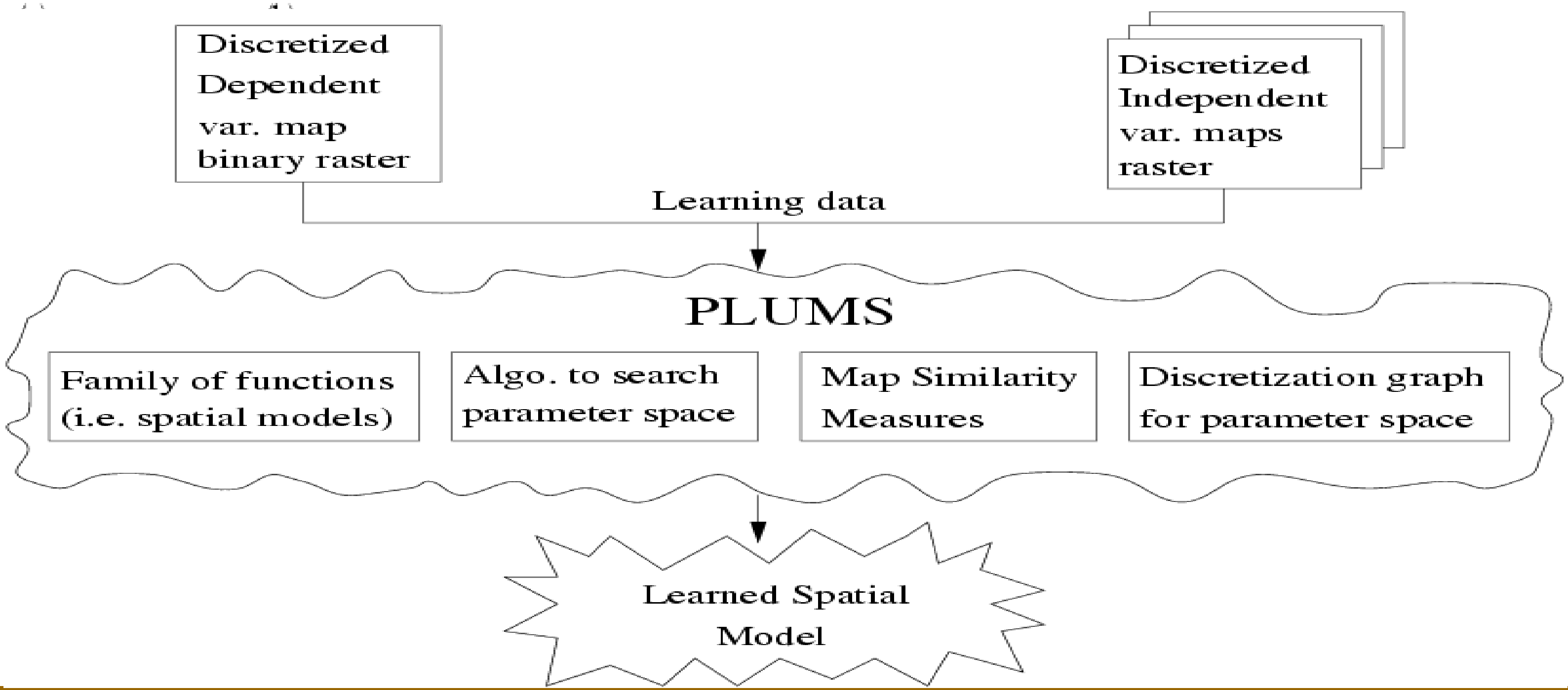
⊙ = nest location

A = actual nest in pixel

P = predicted nest in pixel

New measure:

$$ADNP(A,P) = \sum_k dist(A_k, A_k.nearest(P))$$



Discretized Dependent var. map binary raster

Discretized Independent var. maps raster

Learning data

PLUMS

Family of functions (i.e. spatial models)

Algo. to search parameter space

Map Similarity Measures

Discretization graph for parameter space

Learned Spatial Model

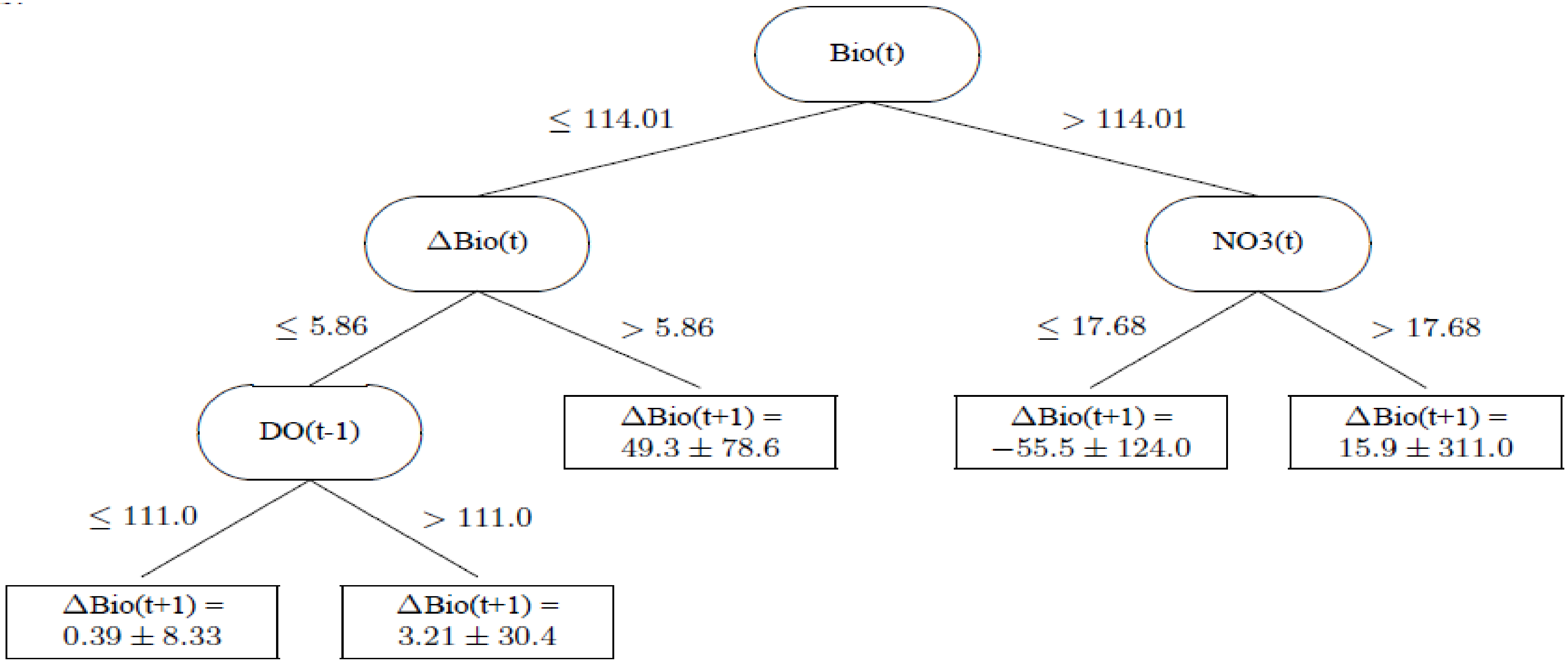
S. Shekar: www.cs.umn.edu/~shekhar/talk/**ucgis.ppt**

Outline

- Introduction
- Mining Large Scale Environmental Data
  - Preprocessing
  - Clustering
  - Association Rules
- Mining Medium and Low Scale Ecological Data
  - Co-location Mining
  - STAMM
  - PLUMS
  - Regression Trees, Neural Networks, Clustering ⬅
- Sensor Networks for Environmental Applications

Regression Trees,Neural Nets,Clustering(1/3)

- Case Study 1: Lagoon of Venice
  - Use of Regression Trees to study the effect of agricaltural activities to excessive algal growth



Bio(t)

≤ 114.01

> 114.01

ΔBio(t)

NO3(t)

≤ 5.86

> 5.86

≤ 17.68

> 17.68

DO(t-1)

$\Delta\text{Bio}(t+1) = 49.3 \pm 78.6$

$\Delta\text{Bio}(t+1) = -55.5 \pm 124.0$

$\Delta\text{Bio}(t+1) = 15.9 \pm 311.0$

≤ 111.0

> 111.0

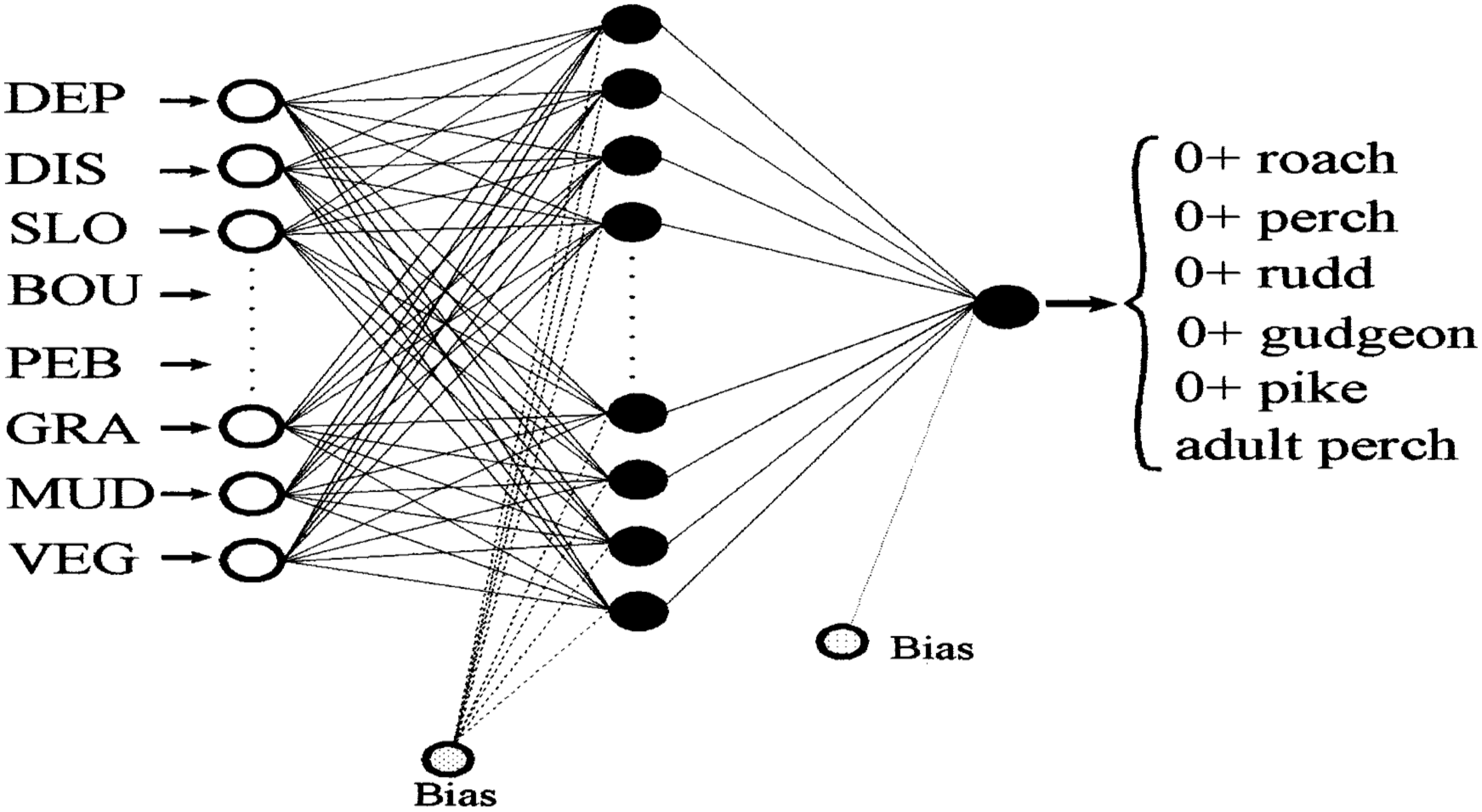
$\Delta\text{Bio}(t+1) = 0.39 \pm 8.33$

$\Delta\text{Bio}(t+1) = 3.21 \pm 30.4$

- Similar Story: Danish Lake Glumose

Regression Trees,Neural Nets,Clustering(2/3)

- Case Study 2: French Lake Preloup
  - Use of feed-forward neural networks to predict the distribution of fish based on soil characteristics



Regression Trees,Neural Nets,Clustering(3/3)

- Other…
  - Use of K-means to cluster Australian rivers based on 286 catchments [www.maths.anu.edu.au/research.reports/mrr/02/003/MRR02-003.pdf]
  - Various clustering algorithms have been tested to extract groups of rivers sharing common flow characteristics in Colorando, Oregon, Washington

    [Stephen et al, Predicting streamflow regime metrics for ungauged streams in Colorado, Washington, and Oregon, Journal of Hydrology, Volume 325, Issues 1-4, 30 June 2006, Pages 241-261]

Issues…

- Study the evolution of environmental, ecological patterns
  - SNN clusters may survive, split, merge etc
  - How do teleconnections evolve as SNN clusters change?
  - Co-location occurrences may survive or be replaced in different time windows, but also they may move one by one or in groups as time passes
  - How EAR rules change over time? How to detect sequential patterns or seasonalities in the behavior of living organisms?
- Extraction of the "interesting" co-location or EAR rules?
- Co-location and EAR rules in Neighborhood Hierarchies, automatically - dynamically identified neighborhoods, fuzzy neighborhoods?
- Semi-supervised methods tailored for Ecological Data ?

Outline

- Introduction
- Mining Large Scale Environmental Data
  - Preprocessing
  - Clustering
  - Association Rules
- Mining Medium and Low Scale Ecological Data
  - Co-location Mining
  - STAMM
  - PLUMS
  - Regression Trees, Neural Networks, Clustering
- Sensor Networks for Environmental Applications

Why?

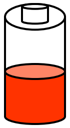
- Certain areas of interest may not be easily approachable or even hostile for humans to install infrastructures
- ….the installation of mazes of cables and devices in such scenarios require excessive infrastructure and administrative costs
- Even when this is achieved, subjects of study tend to alter their natural behavior due to the presence of observation devices

Introduction

- Motes are capable of monitoring a wide variety of ambient conditions including:
  - Temperature, humidity, pressure, lighting
  - Soil makeup
  - Noise levels
  - Presence of certain kinds of objects as well as location, size, direction of movement, speed etc
- Wireless Sensor Networks utility:
  - Scatter cheap, tiny motes in an area of interest
  - Perform querying operations
  - Obtain reports of physical quantities and species under study
  - Support sampling procedures, decision making processes etc

Mote & Network Features

- Mote Features
  - Low Power Supply, Low Power, Low Power...
  - Low processing capabilities
  - Constrained memory capacity



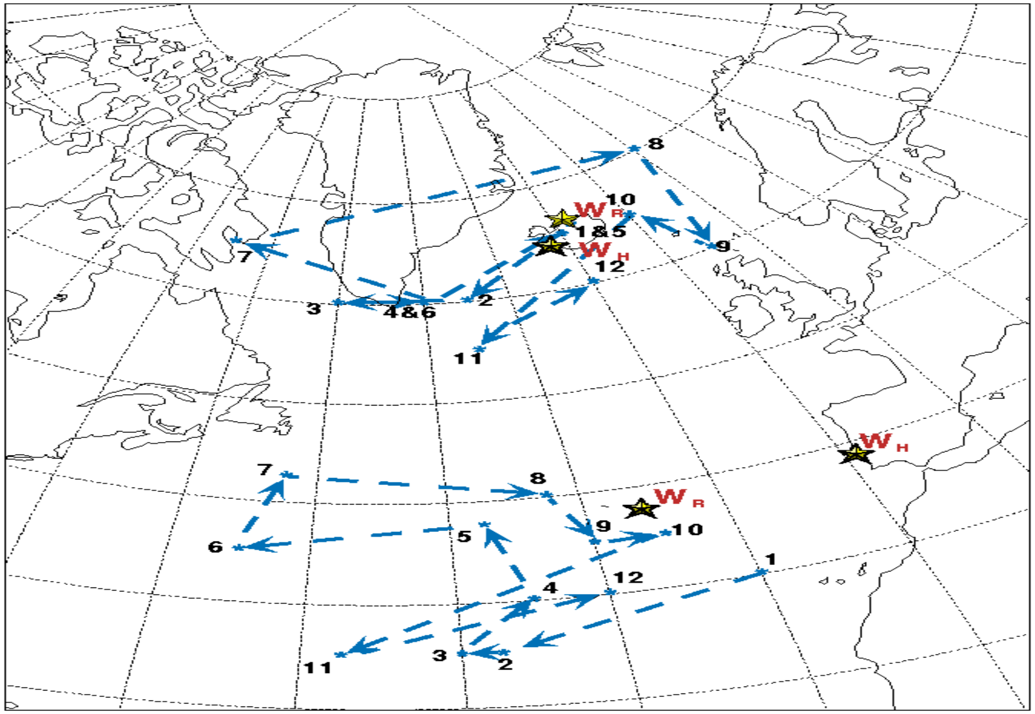
- Network Features
  - Wireless, multi-hop communication using ISM radio zones (433MHz – 2,4GHz)
  - Ad-hoc network topologies

Sensor Net Sample Apps

Habitat Monitoring Case Studies:
Storm petrels on great duck island, microclimates on James Reserve.

Research Projects:

- Cougar
- Dataspace
- Ocean Drifters:
  - ARGO
  - NEPTUNE



Source: Portis et al, Seasonality of the NAO, AGU Chapman Conference, 2000.

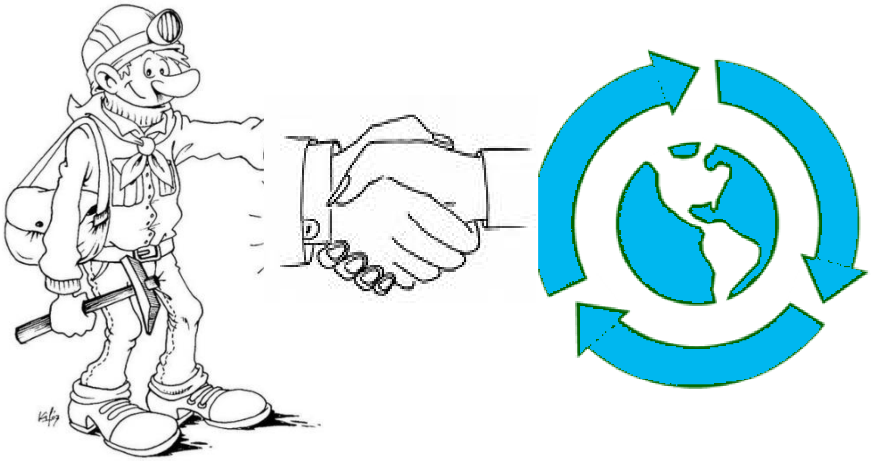
Challenges

- Given sensor readings at different points in the ocean, how to
  - Perform typical aggregate queries?

    Group by…Having…?
  - Identify clusters moving in space and time?
  - Extract co-location patterns and study their evolution?
  - Continuously report the boundary of undergoing phenomena?
  - Continuously detect and report topological relations of undergoing phenomena?

Thank you!

Q&A



© www.schoolplaten.com

Reading List

- Pusheng Zhang, Michael Steinbach, Vipin Kumar, Shashi Shekhar, Pang-Ning Tan, Steve Klooster, and Chris Potter, Discovery of Patterns of Earth Science Data Using Data Mining, as a Chapter in Next Generation of Data Mining Applications, Jozef Zurada and Medo Kantardzic(eds), Wiley-IEEE Press, March 2005.
- Zhang, X., Mamoulis, N., Cheung, D. W., and Shou, Y. Fast mining of spatial collocations. In KDD '04.
- Saso Dzeroski, Applications of symbolic machine learning to ecological modelling, Ecological Modelling, Volume 146, Issues 1-3, 1 December 2001, Pages 263-273
- Su, F., C. Zhou, V. Lyne, Y. Du, and W. Shi. A data-mining approach to determine the spatiotemporal relationship between environmental factors and fish distribution. Ecological Modelling, 174(4):421–431, June 2004
- Cerpa, J. Elson, M. Hamilton, J. Zhao, Habitat monitoring: application driver for wireless communications technology, ACM SIGCOMM’2000, Costa Rica, April 2001.
- Gould, J.; D. Roemmich; et.al. May 2004. Argo Profiling Floats bring New Era of in situ Ocean Observations. EOS Transactions 85(19):185-190