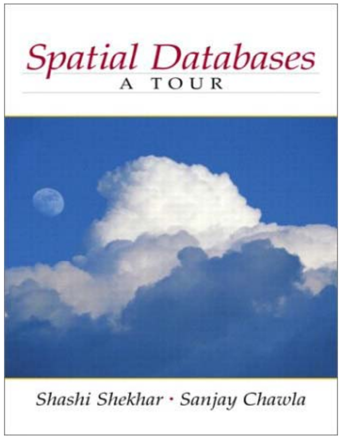
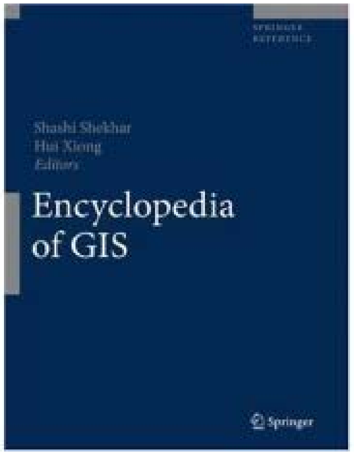
Shashi Shekhar

McKnight Distinguished University Professor
Faculty of Computer Sc. and Eng., Univ. of Minnesota
www.cs.umn.edu/~shekhar

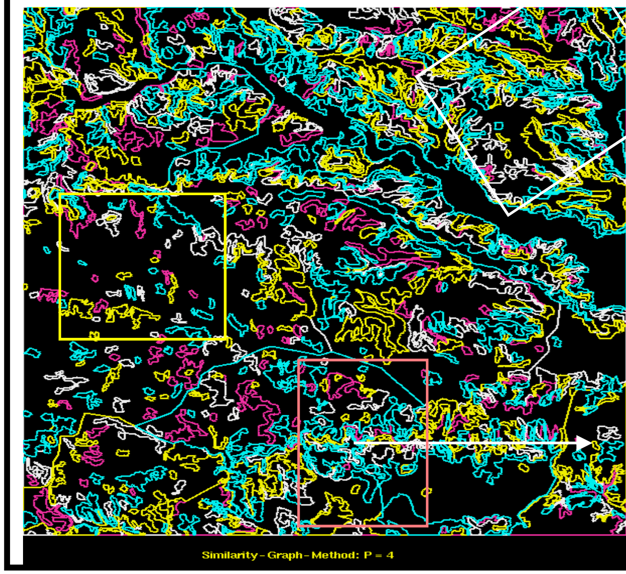
Outline:

- Spatial/Spatio-temporal Database
- Spatial/Spatio-temporal Data Mining



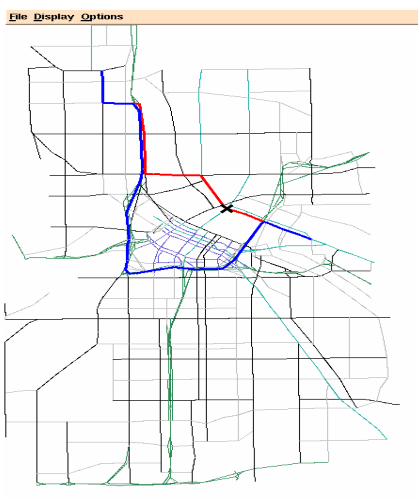


Spatial / Spatio-temporal Databases: Example Projects



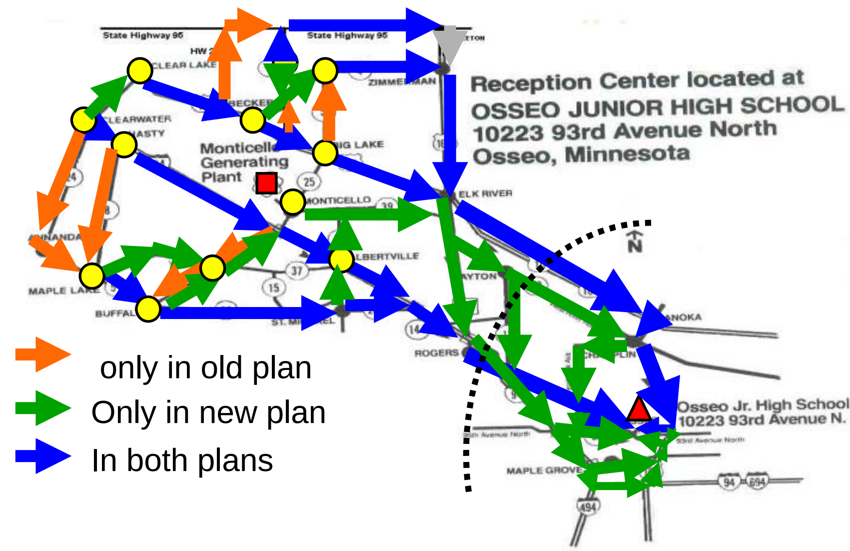
Parallelize Range Queries

Shortest Paths



Storing graphs in disk blocks

Evacutation Route Planning



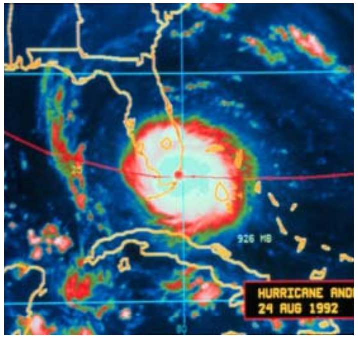
Evacuation Route Planning - Motivation

- No coordination among local plans means
  - Traffic congestions on all highways
  - e.g. 60 mile congestion in Texas (2005)

- Great confusions and chaos

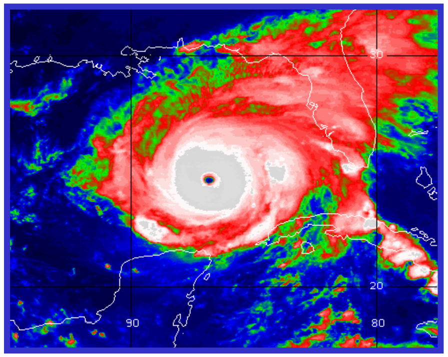
"We packed up Morgan City residents to evacuate in the a.m. on the day that Andrew hit coastal Louisiana, but in early afternoon the majority came back home. **The traffic was so bad that they couldn't get through Lafayette**."
Mayor Tim Mott, Morgan City, Louisiana
(http://i49south.com/hurricane.htm)

Florida, Lousiana
(Andrew, 1992)



(National Weather Services)

Houston
(Rita, 2005)



(National Weather Services)

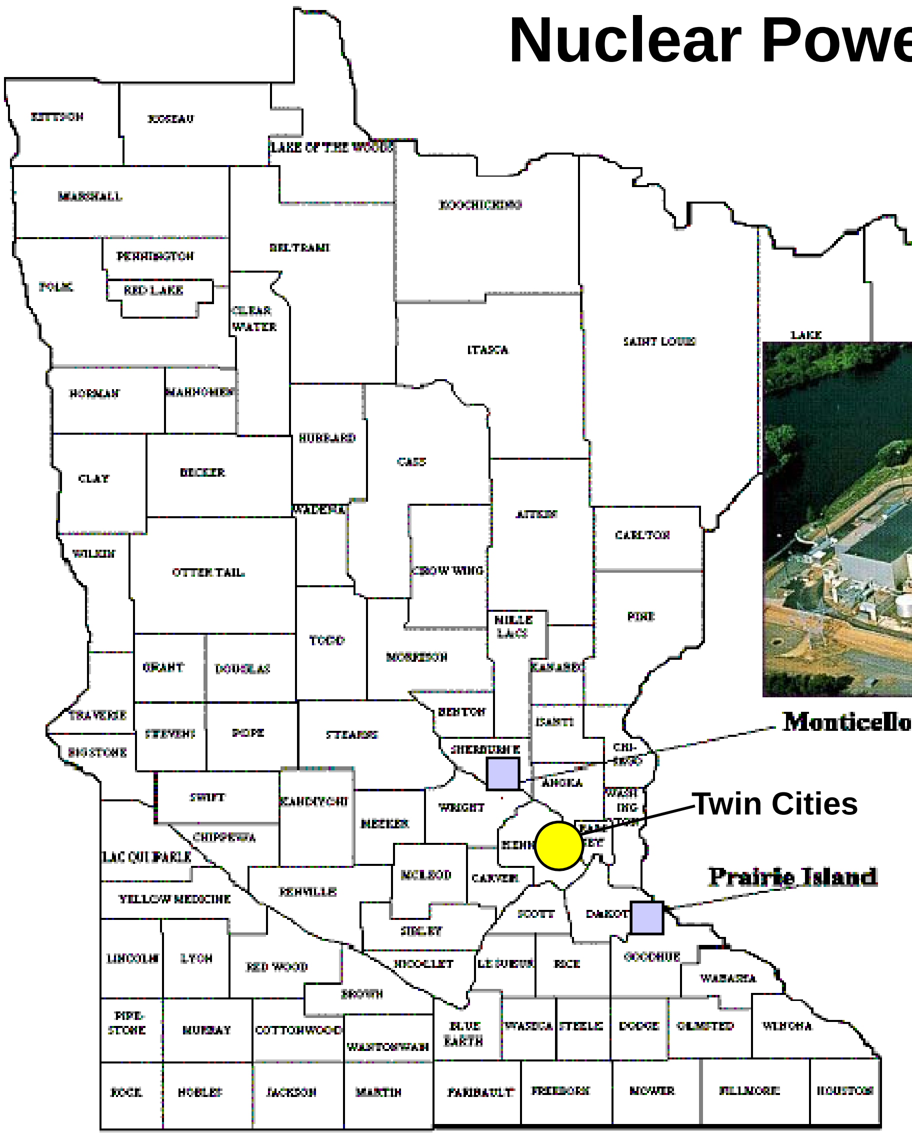
(www.washingtonpost.com)

I-45 out of Houston
(FEMA.gov)

Evacuation Route Planning: A Scenario

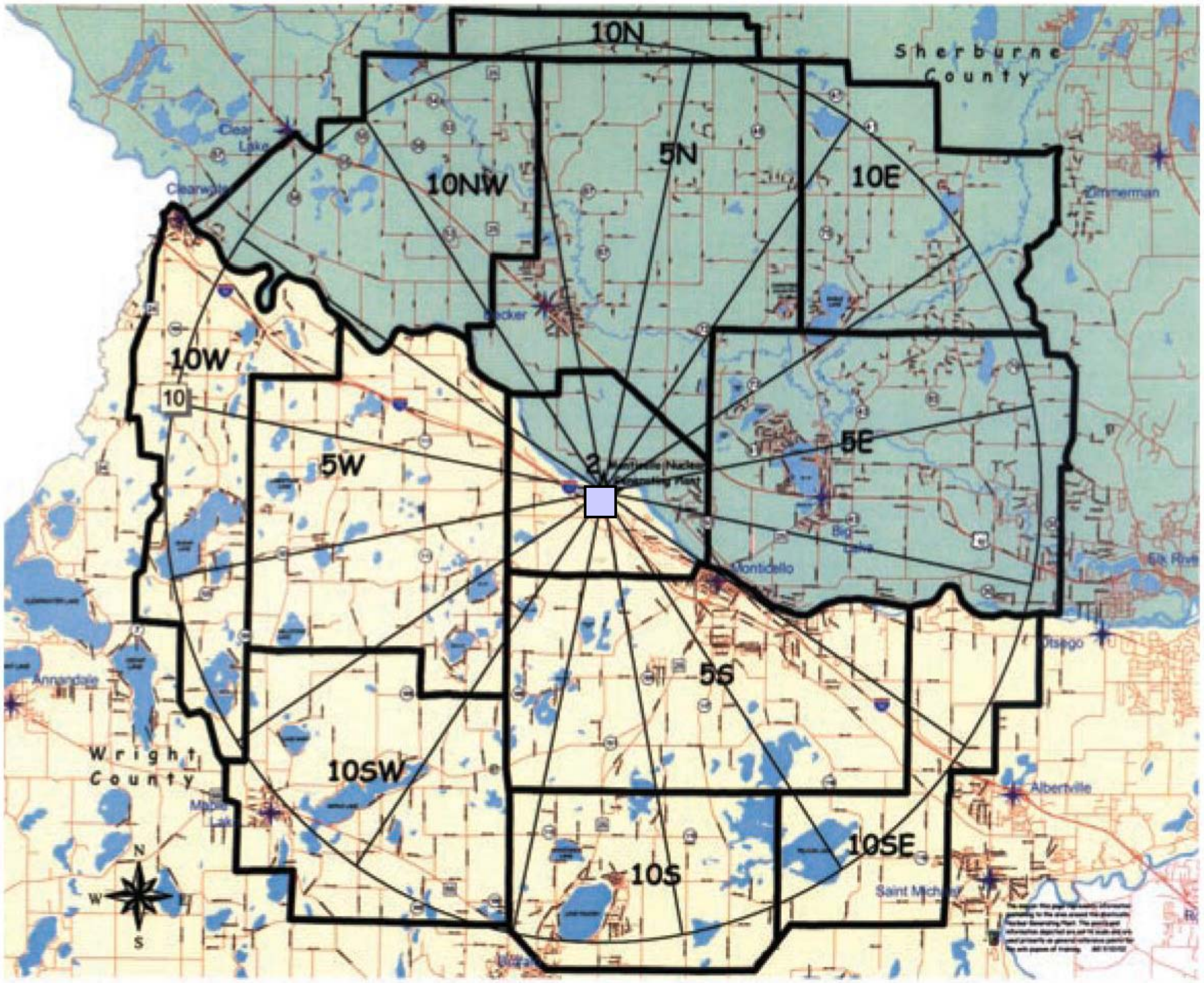
Minnesota Nuclear Power Plants

Nuclear Power Plants in Minnesota



Monticello Emergency Planning Zone

Emergency Planning Zone (EPZ) is a 10-mile radius around the plant divided into sub areas.



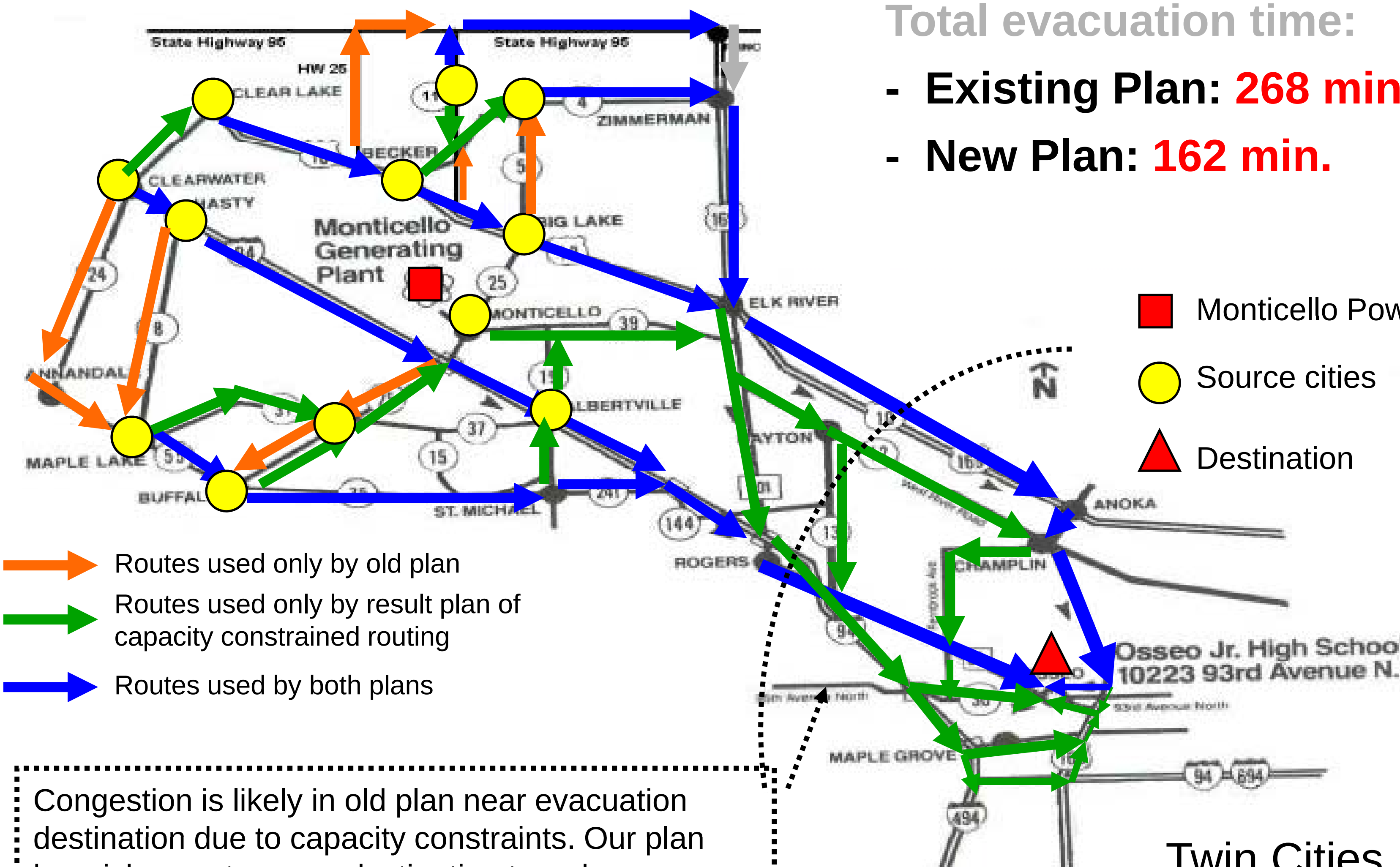
Monticello EPZ

Subarea	Population
2	4,675
5N	3,994
5E	9,645
5S	6,749
5W	2,236
10N	391
10E	1,785
10SE	1,390
10S	4,616
10SW	3,408
10W	2,354
10NW	707
Total	**41,950**

Estimate EPZ evacuation time:
Summer/Winter (good weather):
3 hours, 30 minutes
Winter (adverse weather):
5 hours, 40 minutes

Data source: Minnesota DPS & DHS
Web site: http://www.dps.state.mn.us
http://www.dhs.state.mn.us

A Real World Testcase



Experiment Result

Total evacuation time:

- **Existing Plan: 268 min.**
- **New Plan: 162 min.**

Monticello Power Plant

Source cities

Destination

Routes used only by old plan

Routes used only by result plan of capacity constrained routing

Routes used by both plans

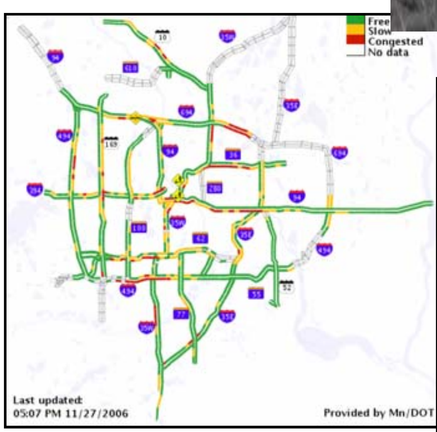
Congestion is likely in old plan near evacuation destination due to capacity constraints. Our plan has richer routes near destination to reduce congestion and total evacuation time.

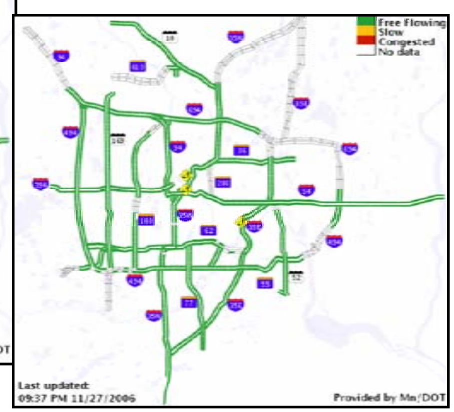
Twin Cities

Computer Sc. Challenge: Time-varying Networks

"U.P.S. Embraces High-Tech Delivery Methods - (*by Claudia Deutsch)*
The research at U.P.S. is paying off. Last year, it cut 28 million miles from truck routes — saving roughly three million gallons of fuel — in good part by *mapping routes that minimize left turns*"
- *New York Times* (July 12, 2007)

Congestion Levels in road networks change with time resulting in travel time changes.





Increasing availability of traffic data through sensors on road networks.

Example queries on a time-varying network

1) Transportation network Routing

- Varying Congestion Levels and turn restrictions ⇒ travel time changes.

Static	Time-Variant
Which is the shortest travel time path from downtown Minneapolis to airport?	Which is the shortest travel time path from downtown Minneapolis to airport at different times of a work day?
What is the capacity of Twin-Cities freeway network to evacuate downtown Minneapolis ?	What is the capacity of Twin-Cities freeway network to evacuate downtown Minneapolis at different times in a work day?

2) Crime Analysis

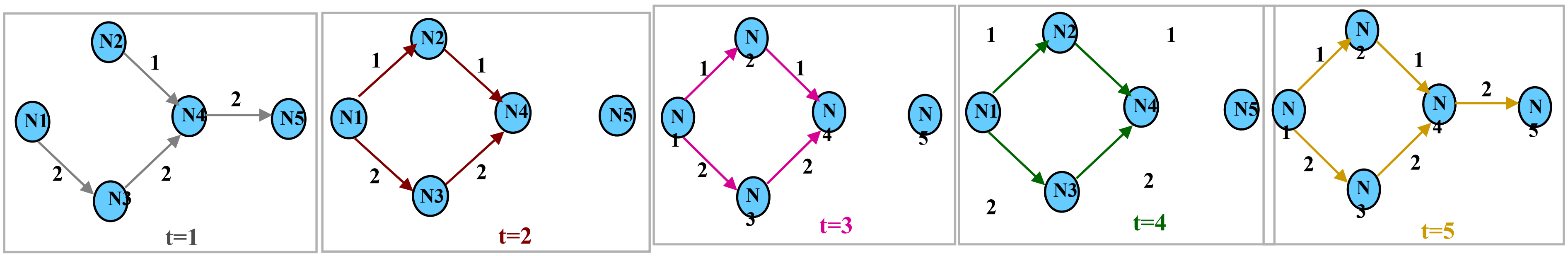
- Identification of frequent routes *(i.e.)* Journey to Crime

3) Knowledge discovery from Sensor data.

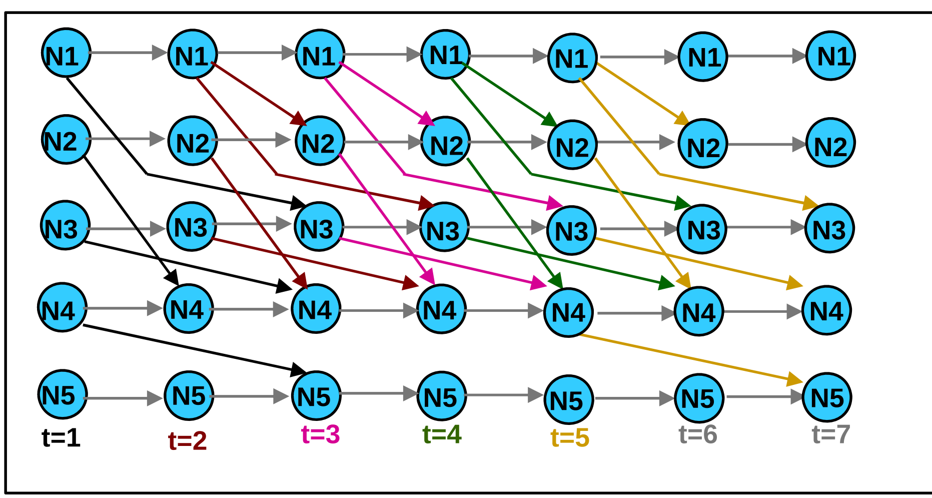
- Spreading Hotspots

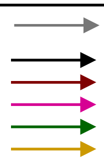
Secret Sauce: Representation of (Spatio-)temporal Networks

(1) Snapshot Model [Guting 04]   Node: N..   Edge: Travel time



(2) Time Expanded Graph (TEG) [Ford 65]



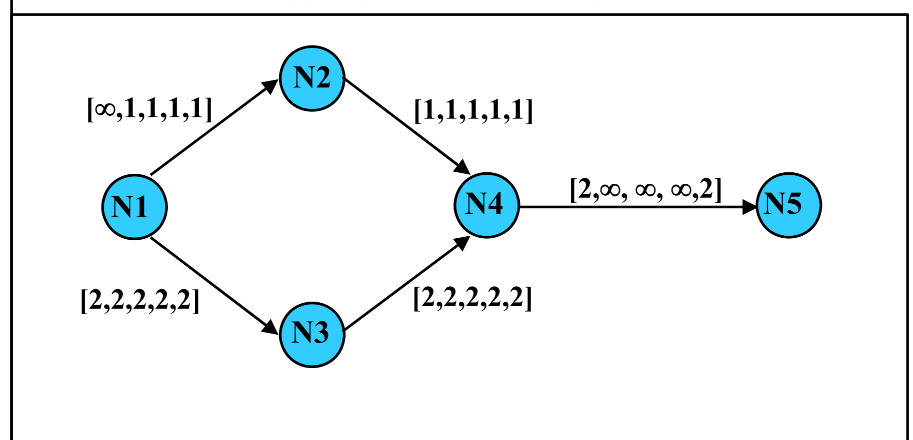


Holdover Edge

Transfer Edges

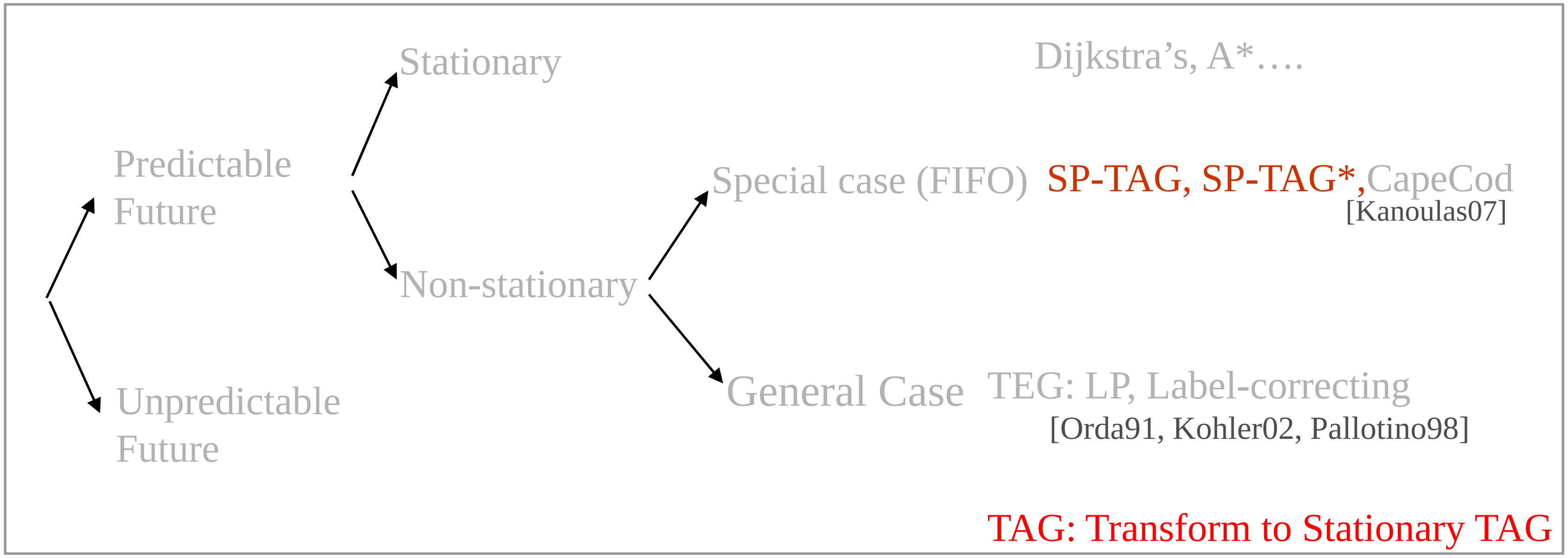
(3) Time Aggregated Graph (TAG) [Our Approach]

¤ Attributes aggregated over edges and nodes.

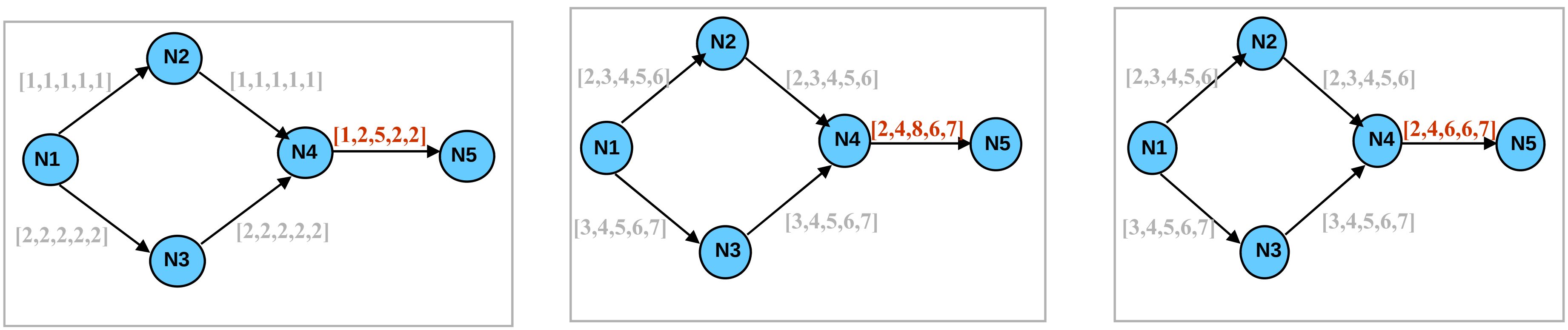


Edge $[m_1,\ldots,(m_T]$ → m_i- travel time at t=i

Power of Representation: Ex. Routing Algorithms



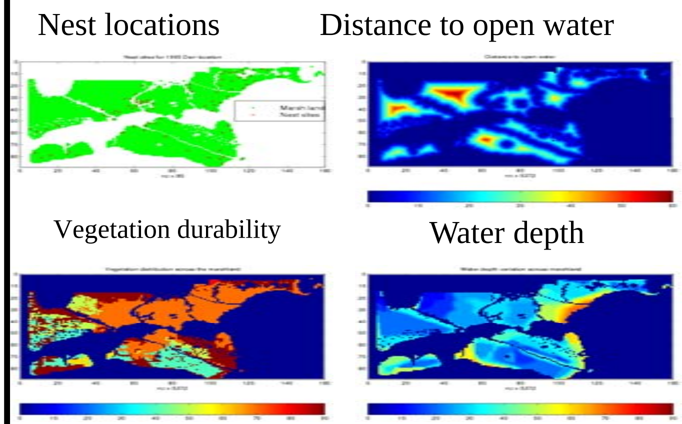
travel times → arrival times at end node → Min. arrival time series



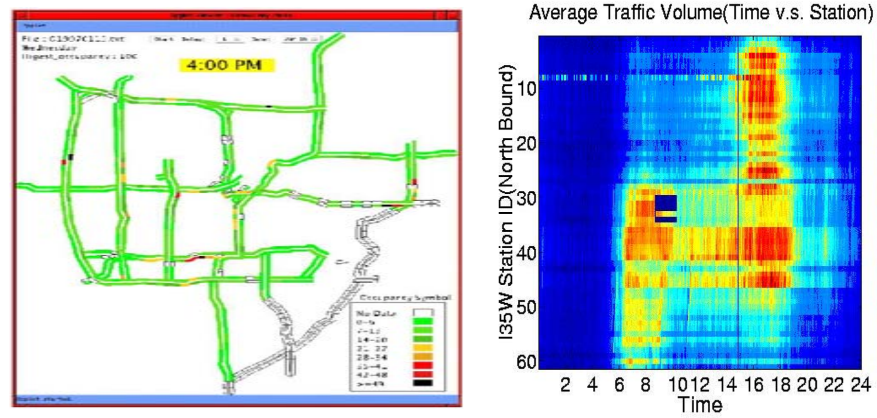
Non-stationary TAG → Stationary TAG

Spatial / Spatio-temporal Data Mining: Example Projects

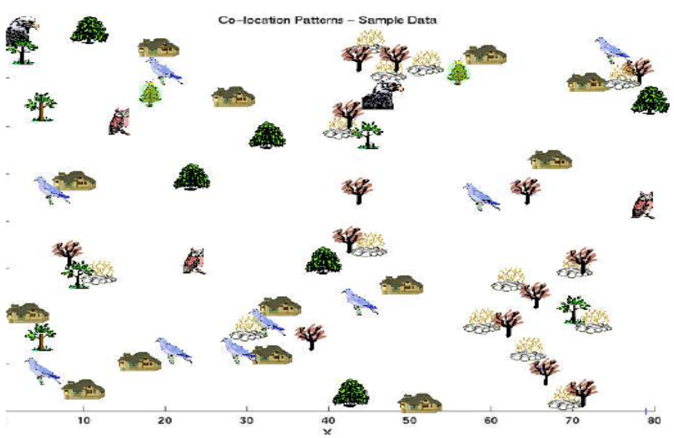
Location prediction: nesting sites



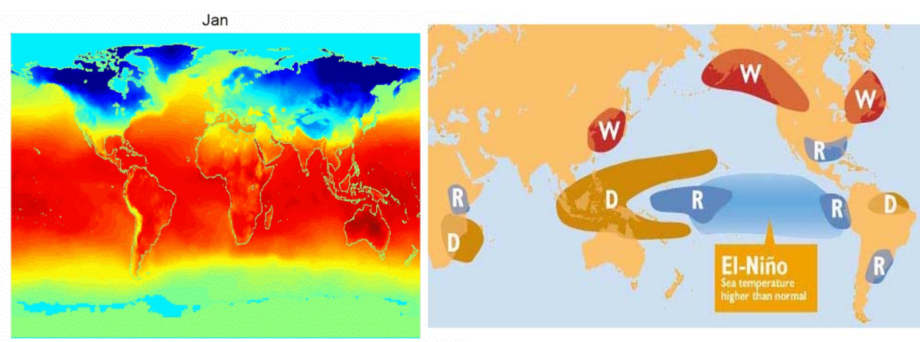
Spatial outliers: sensor (#9) on I-35



Co-location Patterns

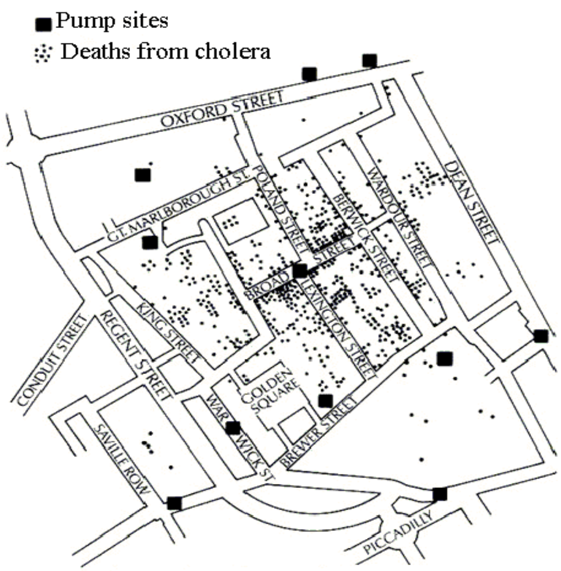


Tele connections



(Ack: In collaboration w/V. Kumar, M. Steinbach, P. Zhang)

Spatial and Spatio-temporal Data Mining

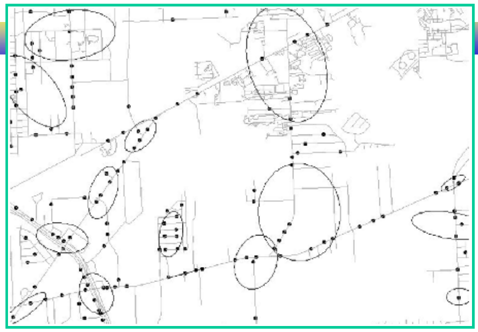


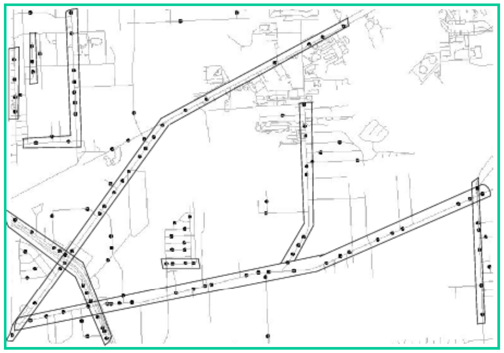
- What is it?
  - Identifying interesting, useful, non-trivial patterns
    - Hot-spots, discontinuities, co-locations, trends, …
  - in large spatial or spatio-temporal datasets
    - Satellite imagery, geo-referenced data, e.g. census
    - gps-tracks, geo-sensor network, …

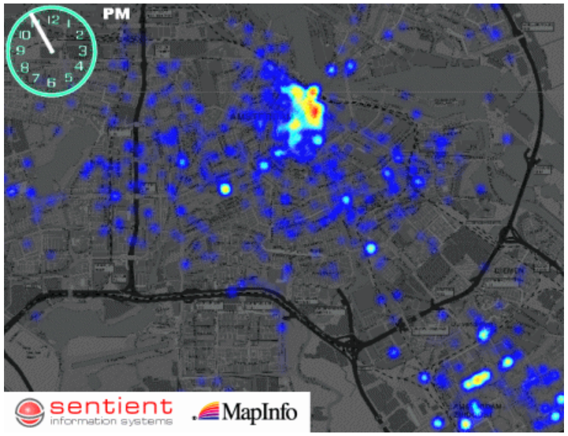
- Why is it important ?
  - Potential of discoveries and insights to improve human lives
    - Environment: How is Earth system changing? Consequences for humans?
    - Public safety: Where are hotspots of crime? Why?
    - Public health: Where are cancer clusters? Environmental reasons?
    - Transportation, National Security, …
  - However, (d/dt) (Spatial Data Volume) >> (d/dt) (Number of Human Analysts)
    - Need automated methods to mine patterns from spatial data
    - Need tools to amplify human capabilities to analyze spatial data

HotSpots

- What is it?
  - Unusally high spatial concentration of a phenomena
    - Cancer clusters, crime hotspots
- Traditional Approach:
  - Spatial statistics based ellipsoids
- Our Recent Focus:
  - Computational Structure
    - Spatial Join-index reduces computational costs
  - Transportation network based hotspots
- Next: Spatio-temporal
  - Ex. Emerging hot-spots

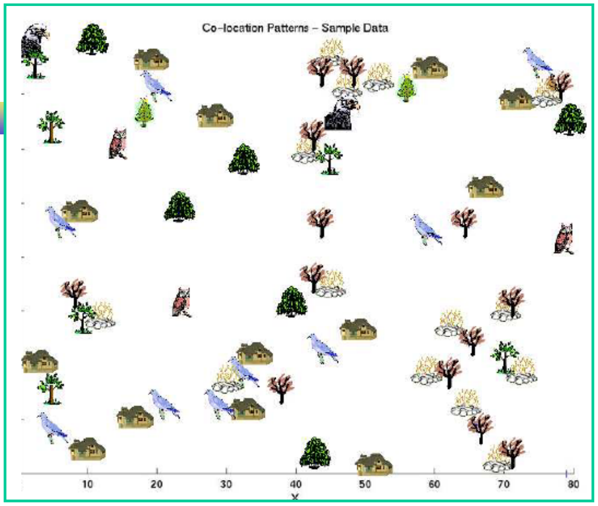


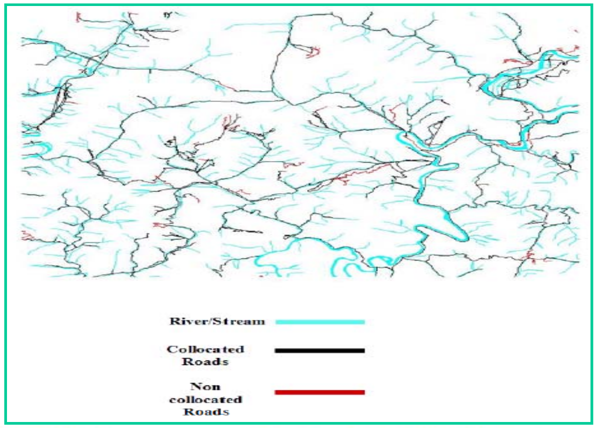


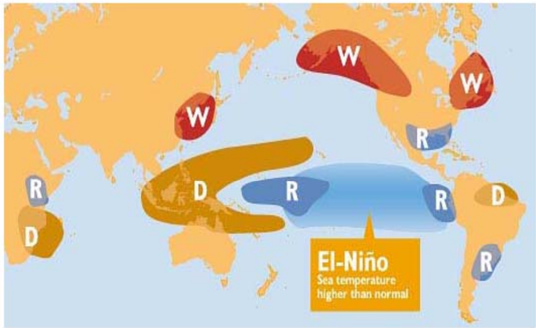


Colocation, Co-occurrence, Interaction

- What is it?
  - Subset of event types, whose instances occur together
  - Ex. Symbiosis, (bar, misdemeanors), …
- Traditional Approach:
  - Neighbor-unaware Transaction based approaches
- Our Approach:
  - Aggregate Functions on Neighbor relationships
  - Balance statistical rigor and computational cost
- Next: Spatio-temporal interactions
  - Item-types that sell well before or after a hurricane
  - Object-types that move together
  - Tele-connections

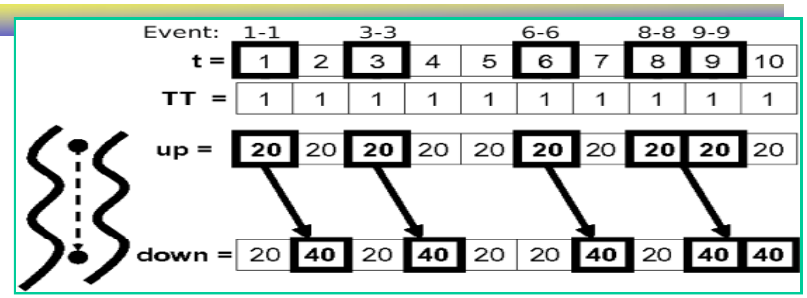


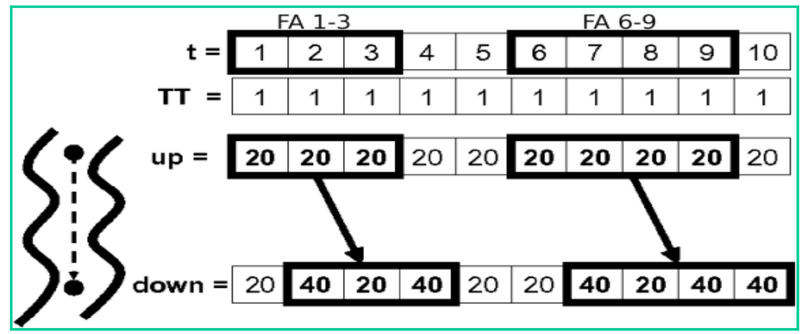


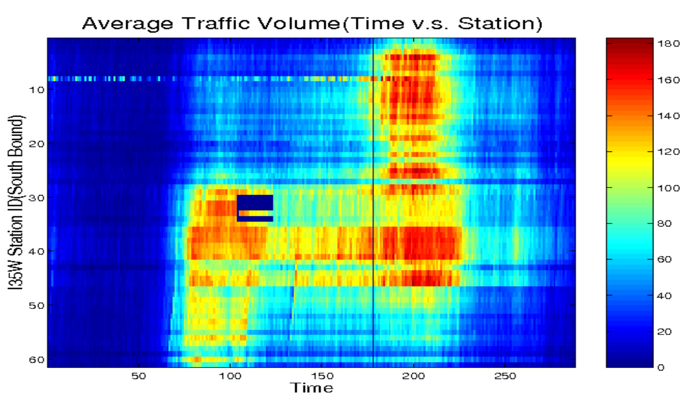


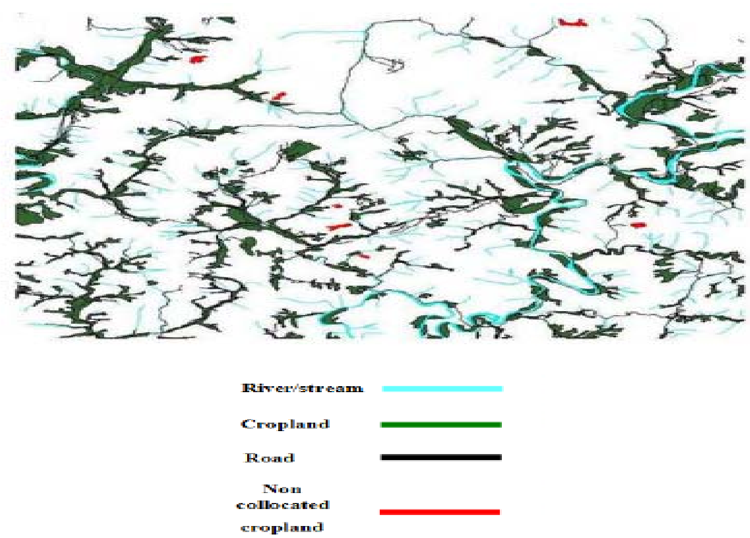
Spatial/Spatio-temporal Outliers, Anamolies

- What is it?
  - Location different from their neighbors
    - Discontinuities, flow anomalies
- Related Work
  - Transient spatial outliers
  - Anomalous trajectories
  - Computational Structure: Spatial Join
    - Very scalable using spatial DBMS
- Next
  - (Dominant) Persistent anomalies
  - Multiple object types, Scale

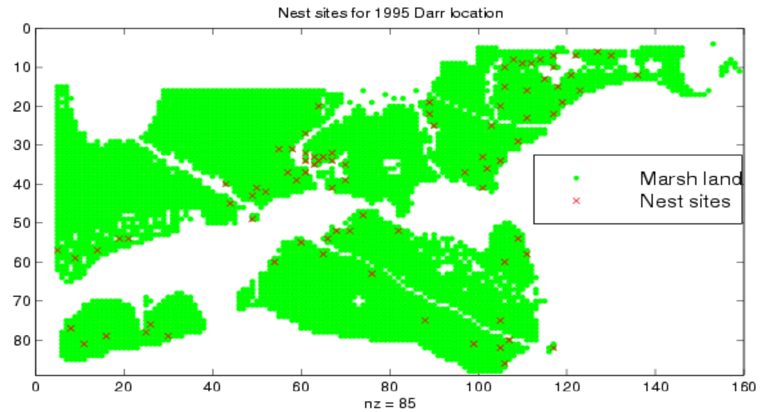




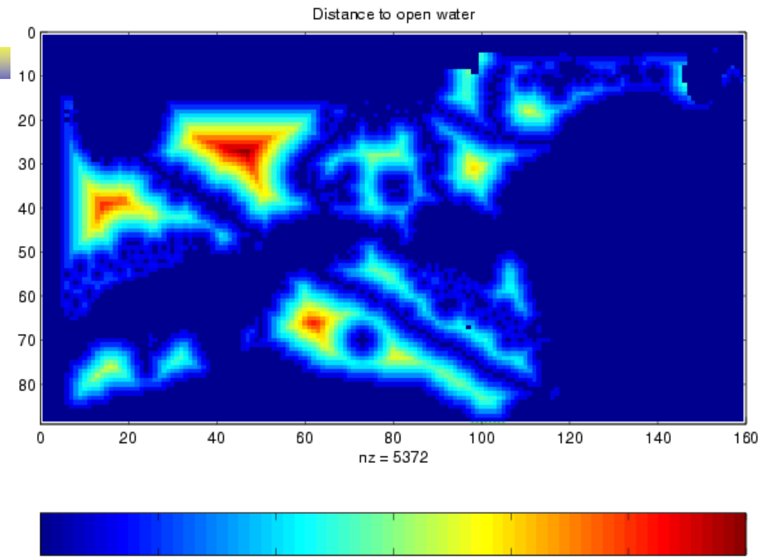




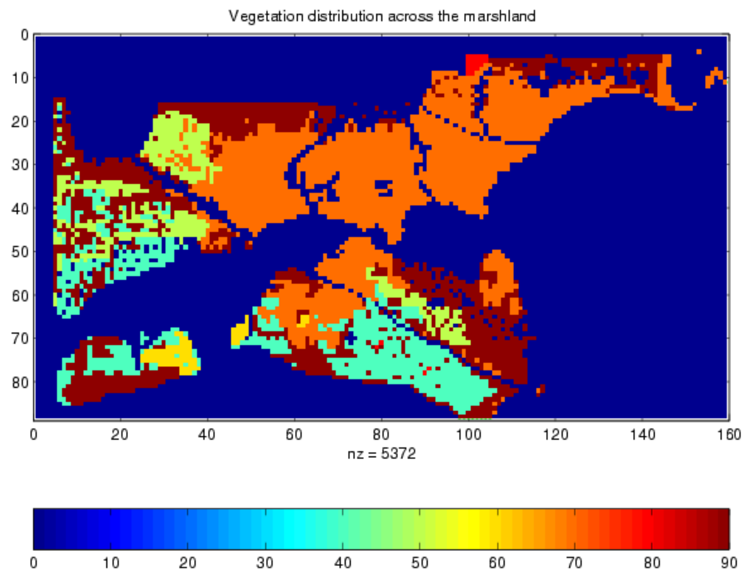
Location Prediction – An Example



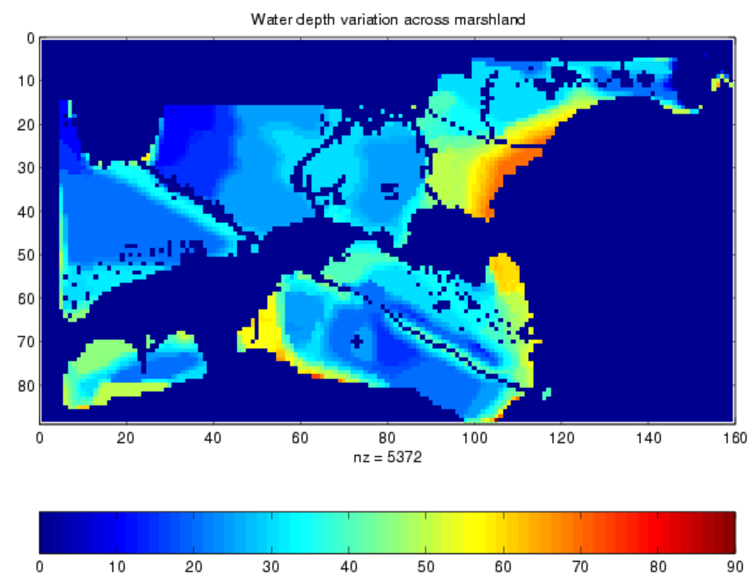
Nest locations



Distance to open water



Vegetation durability



Water depth

Implication of Auto-correlation

Name	***Model***	***Classification Accuracy***
Classical Linear Regression	$\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$	Low
Spatial Auto-Regression	$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$	High

ρ : the spatial auto-regression (auto-correlation) parameter

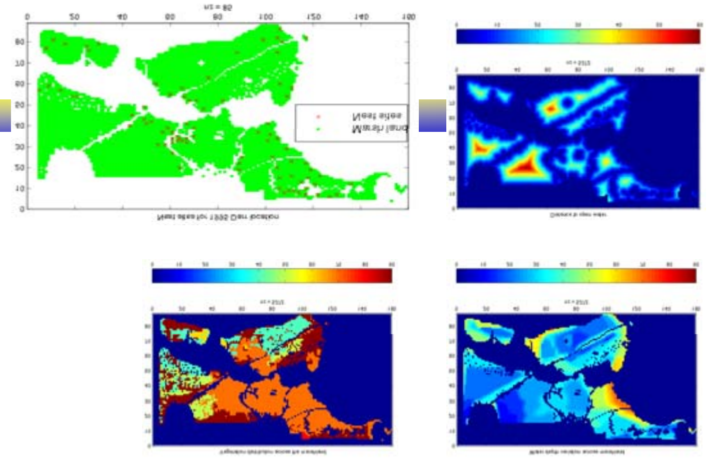
$\mathbf{W}$: n-by-n neighborhood matrix over spatial framework

<u>Computational Challenge</u>:

Computing determinant of a very large matrix
in the Maximum Likelihood Function:

$$\ln(L) = \ln\left|\mathbf{I} - \rho\mathbf{W}\right| - \frac{n\ln(2\pi)}{2} - \frac{n\ln(\sigma^2)}{2} - SSE$$

Space/Time Prediction



- What is it?
  - Models to predict location, time, path, …
    - Nest sites, minerals, earthquakes, tornadoes, …
- Related Work
  - Interpolation, e.g. Krigging
  - Heterogeneity, e.g. geo. weighted regression
  - Auto-correlation, e.g. spatial auto-regression
- Challenge: Independence assumption
  - Models, e.g. Decision trees, linear regression, …
  - Measures, e.g. total square error, precision, recall
- Next
  - Spatio-temporal vector fields (e.g. flows, motion), physics
  - Scalable algorithms for parameter estimation
  - Distance based errors

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\ln(L) = \ln\left|\mathbf{I} - \rho\mathbf{W}\right| - \frac{n\ln(2\pi)}{2} - \frac{n\ln(\sigma^2)}{2} - SSE$$