Interdisciplinary Research: Two War Stories

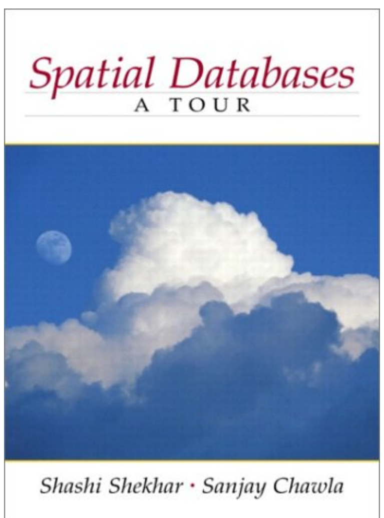
Shashi Shekhar

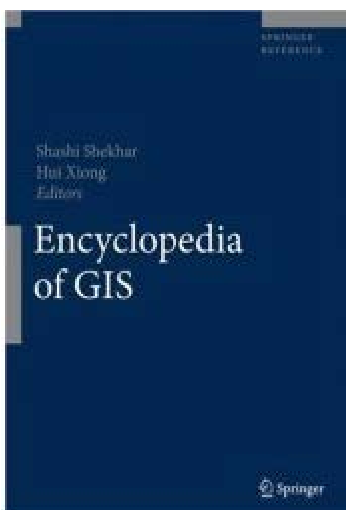
McKnight Distinguished University Professor

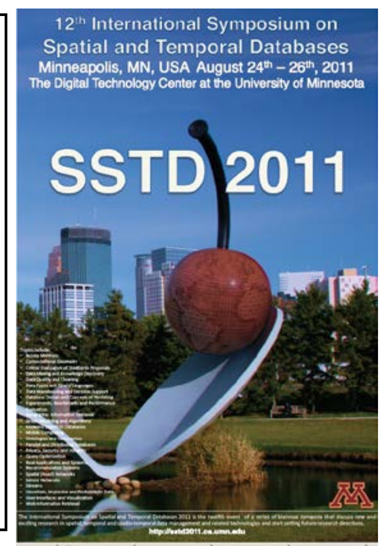
Department of Computer Science and Eng., University of Minnesota

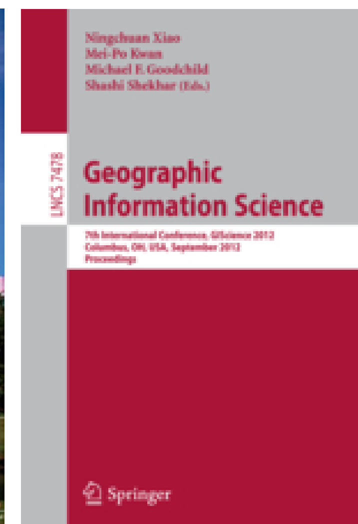
www.cs.umn.edu/~shekhar

Presented in Jan. 7th, 2013 Workshop to Kickoff
Technology Innovation Center in Geographic Information Systems,
Umm Al-Qura University, Abdia, Makkah-21955, Saudi Arabia







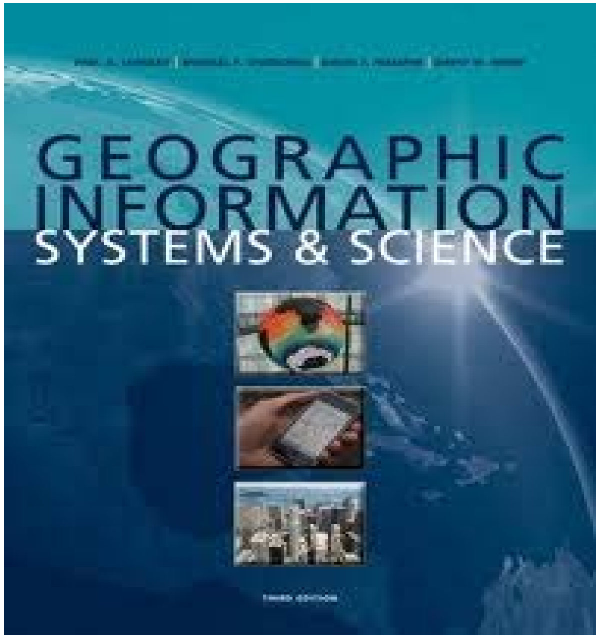


Collaborations With Geographers

- All I know about GIS, I learned from Geographers!
  - And I used that to discover limitations of Computer Science
  - It helped advance Computer Sc. While serving societal application domains
- University of Minnesota
  - R. McMaster: co-taught seminar (1994), co-advised students
  - T. Burk: UMN Map Server (1995-2005) scale up searches
  - F. Harvey: co-PI on U-Spatial (2012-15), MN Future Workshop (2009)
  - S. Ruggles, Minnesota Population Center, NSF Datanet Terrapop
- USA
  - M. Goodchild (UCSB), J. Dangermond (ESRI) D. Mark (Buffalo), M. Egenhofer (Maine)
  - National Academy of Science: Mapping Sciences Committee (2003-9)
  - UCGIS: Board of Director (2003), Congressional breakfast (2004), …
  - Intl. Conf. on Geographic Info. Science (2012 Co-chair, 2004 keynote)
  - American Association of Geographers (AAG) workshops with NIH, …
- International
  - Canada (GEOIDE), Ireland (Geo-computation Center, NUI), India (GIS Center @ IIT-Mumbai, UNDP Project), China (Wuhan U, Beijing U), Ukraine (CRDF), …
  - Keynote @ Geo-Computation conference (2011, UCL, UK), ISPRS Spatial Data Mining (2006, Turkey), Finland (2013), …

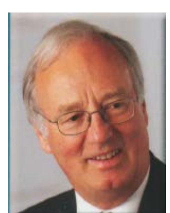
GIS contributions highlighted in a popular Textbook

GIS is not just about machines, but also about people. We present vignettes of committed, moti- vated **individuals whose contributions have also made a difference to GIS.**



WILEY
www.wiley.com/college/longley



PAUL A. LONGLEY | MICHAEL F. GOODCHILD | DAVID J. MAGUIRE | DAVID W. RHIND

Figure 10.21 Shashi Shekhar.

Biographical Box 10.3

Shashi Shekhar, Computer Scientist

Shashi Shekhar (Figure 10.21) received a computer science and engineering (CSE) education from the Indian Institute of Technology, Kanpur (1981–1985) and the University of California, Berkeley (1985–1989). He is presently a McKnight Distinguished University Professor of CSE at the University of Minnesota. His knowledge of GIS started in the early 1990s from sponsored research projects on computational aspects of in-vehicle and Web-based navigation systems (USDOT), high-performance GIS for vehicle simulators (USDOD), Minnesota Mapserver (NASA), and the like. Soon, he realized there was a strong and growing demand for CSE advances for GIS, yet very few CSE scholars were dedicated to this promising area. Thus, in the mid-1990s, he decided to focus his research on understanding the structure of very large spatial computations (e.g., data analysis via spatial querying and spatial data mining). Illustrative contributions include: the Capacity Constrained Route Planner, an evacuation route planning algorithm, which is orders of magnitude faster than traditional linear programming-based methods; the Connectivity-Clustered Access Method, a min-cut graph-partitioning-based storage method, which outperforms geometry-based indices (e.g., R-tree family) in carrying out network computations; and the notion of "co-location" patterns in spatial datasets to provide a trade-off between computational scalability and spatial statistical rigor.

CCC Visioning Workshop: Making a Case for Spatial Computing 2020
http://cra.org/ccc/spatial_computing.php

Computing Community Consortium
We support the computing research community in creating compelling research visions and the mechanisms to realize these visions.

HOME | ABOUT | YOUR VISION | ACTIVITIES | RESOURCES | CONTACT | GO

Funded Visioning Activities

Disaster Management · SEES IT · HealthIT · Interactive Tech · Architecture · XLayer · Robotics · Learning Tech · Open Source · Cyber Physical Systems · Global Development · Theoretical CS · Big Data Computing · NetSE · Spatial Computing

From GPS and Virtual Globes to Spatial Computing-2020

About the workshop

This workshop outlines an effort to develop and promote a unified agenda for Spatial Computing research and development across US agencies, industries, and universities. See the original workshop proposal **here**.

Spatial Computing

Spatial Computing is a set of ideas and technologies that will transform our lives by understanding the physical world, knowing and communicating our relation to places in that world, and navigating through those places.

The transformational potential of Spatial Computing is already evident. From Virtual Globes such as Google Maps and Microsoft Bing Maps to consumer GPS devices, our society has benefitted immensely from spatial technology. We've reached the point where a hiker in Yellowstone, a schoolgirl in DC, a biker in Minneapolis, and a taxi driver in Manhattan know precisely where they are, nearby points of interest, and how to reach their destinations. Large

Logistics

Date: Sept. 10th-11th, 2012
Location: Keck Center
Hotel: Liaison Hotel

Steering Committee

Erwin Gianchandani

Hank Korth

Organizing Committee

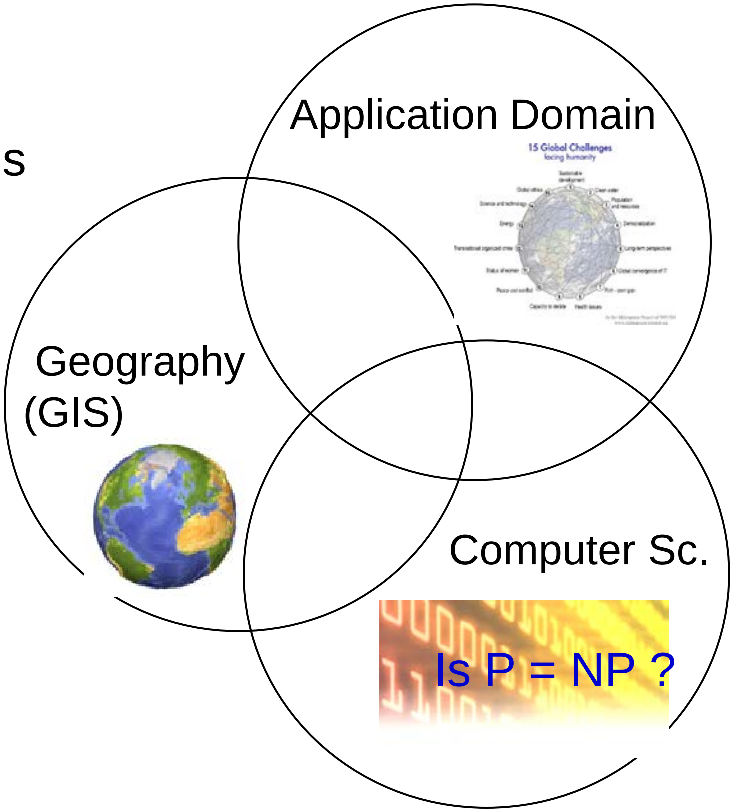
Peggy Agouris, George Mason University

Walid Aref, Purdue University

Michael F. Goodchild, University of California - Santa Barbara

Outline

- Binary Interdisciplinary War Story
  - Computer Science (CS) advancing a Domain Science
  - Advance both Computer SC. and Domain Science
- Ternary Interdisciplinary War Story
  - GIS and CS together address societal needs



Story 1 – Clean Water

March 11, 2008, 10:54 am

We Are What We Drink Is What We Are

By ANDREW C. REVKIN

(Tony Cenicola/The New York Times)

That headline isn't a typo.

By 2025,1.8 billion people could be living in water scarce areas

- **Today**, 750 million people live below the water-stress threshold of 1.7 K

Two Disciplines

Civil Engineering

- Professors
  - William Arnold
  - Ray Holzalski
  - Miki Hondzo
  - Paige Novak
- Students
  - Mike Henjum
  - Christine Wennen

Computer Science

- Professors
  - Shashi Shekhar
- Students
  - Jim Kang

Hydrology

Professors

- David Maidment

What is Interdisciplinary Research?

- Is it multiple Disciplines working on a single project?
- Is it one discipline helping another?
- My Thoughts:
  - Ideal: Perform research that enhances all disciplines involved.
    - Not just a subset!
  - Very Hard To Do!!!
  - A lot of asking questions back and forth

Why Interdisciplinary Research?

Interdisciplinary research

- integrates information, data, techniques, tools, perspectives, concepts, and/or theories from two or more disciplines or bodies of specialized knowledge

Ex. What is computational complexity of Line-simplification? (e.g., Douglas-Pucker algorithm). How may one geo-locate an IP-address on Internet?

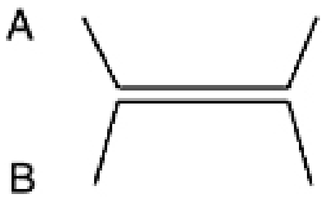
- to advance fundamental understanding or to solve problems whose solutions are beyond the scope of a single discipline or area of research practice.

- Ex.: Grand Societal Challenges, climate change, geo-privacy, evacuation, …

Evacuation: How may we estimate (day-time) population in evacuation zone? What are best evacuation routes (in system-optimal or game-theoretic sense)? Will evacuees follow those?

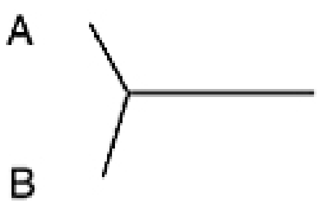
Source: Facilitating Interdisciplinary Research, National Academy Press, 2004.

Different Possibilities



A
B

Multidisciplinary:
Join together to
work on common problem,
split apart unchanged
when work is done.

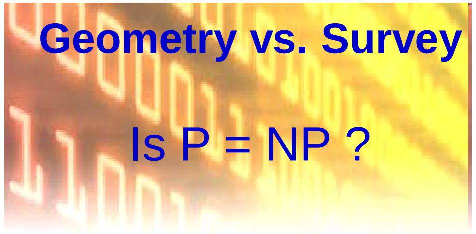


C

Interdisciplinary:
Join together to
work on common question or problem.
Interaction may forge a new
research field or discipline.

Challenges: Communication, Trust, Finding Win-Win

- Each Discipline has its own language!
  - We need to learn language of other discipline
  - Almost like learning a foreign language
  - Homonym (Civil Eng., C. S.): infrastructure, network, bridge,
  - Homonym (GIS, C. S.).: scale, map, generalization (map vs. across problems), …
  - Synonym (GIS & C. S.): conflation vs. data integration, map overlay vs. spatial join …
- Disciplines have different values, cultures and goals
  - Different norms for ordering of authors in a joint paper
  - Different emphasis on grants, citations, service, …
- Mis-understanding of what each discipline really is
  - e.g., “I thought Civil Eng. was all about building bridges!”
  - e.g., “I thought Computer Sc. was all about programming!”

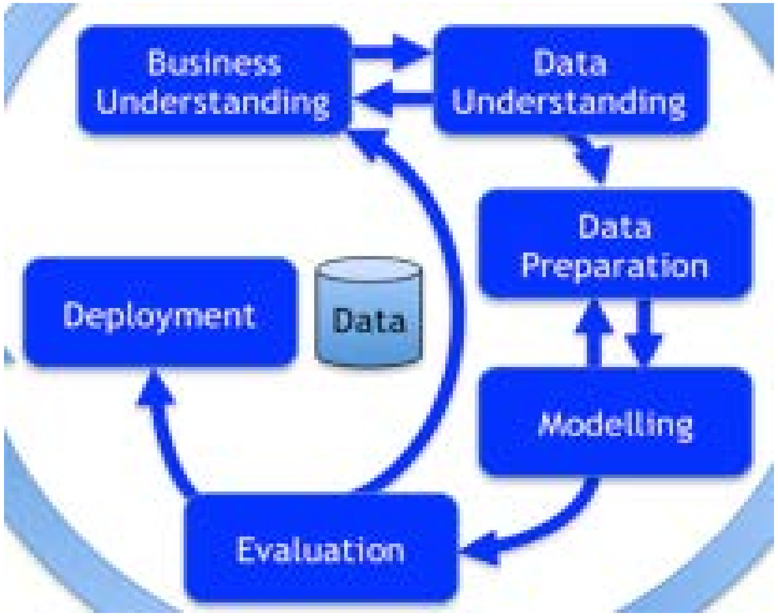


- Break down barriers
  - Build mutual trust and respect via social activities, sharing core competencies, …
  - Keep talking to each other to find a win-win
  - Have an open mind when discussing each others’ interests

Brainstorming: In the Beginning…

Civil Engineering Questions

- How is Computer Science involved in this work?
- CS: I don't know!
- Need to understand domain questions and dataset first



CRISP-DM Methodology

Computer Science Questions

- How many sensors will there be? Like 1000 or 10,000 or more?
- CE Ans: No Way! A sensor cost $30,000
- CS: Large data or computations ?
- CE: Time-series length is in 50,000
- CS: time-series similarity query?
- CE: What is similarity?
- CS: time-series correlation
- CE: Not yet!

Brainstorming: A little later…

Civil Eng. Questions

- Can you help publish field-sensor data on national website (HIS) ? dealing with wireless communication, data format conversion, XML, etc.?
- Can you remove errors from the dataset?
- CS Ans:
  - Yes, we will help with these.
  - But these do not advance CS
  - Techniques already exist

Computer Science Questions

- Do you want to know how fast the river is flowing?
- CE Ans:
  - Not really,
  - We can already determine that by the discharge, water depth, and physical characteristics of the river

Brainstorming: Light at the end of the tunnel

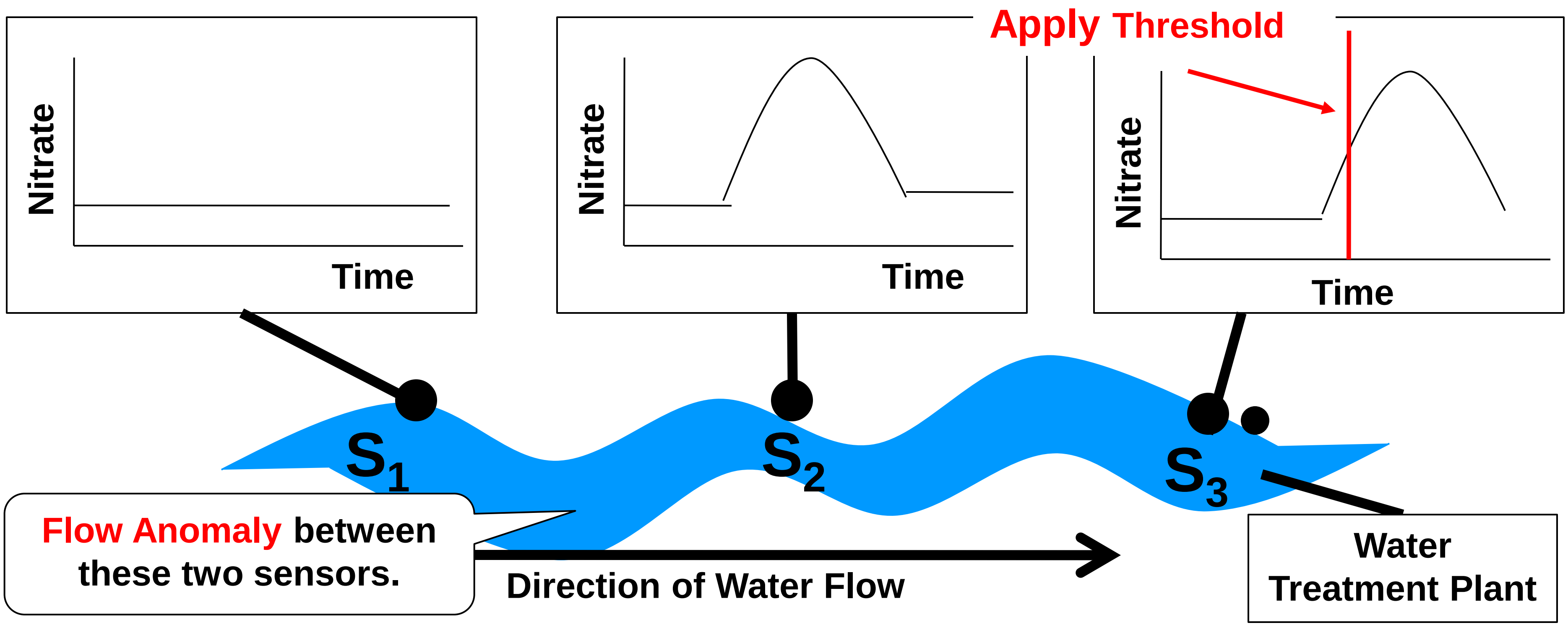
Civil Engineering

- Can you find when and where a contaminant enter a river?
- CS Ans.
  - What is signature in data from sensors ?
- CE Ans.
  - Signature across a pair of sensors.
  - Upstream sensor sees clean water.
  - Downstream sensor sees contamination.
- CS.:
  - How hard is it manually? computationally?
- CE Ans.
  - Its too hard to find this manually
  - e.g., hours to sift through the data
  - 50k data points per measured variable

Computer Science

- Let me see. Spatial outlier detection won't not catch it!
- A new pattern family: Flow Anomaly: time periods of flow mismatch between up/downstream sensors?
- It may advance Computer Sc.
- Which Flow Anomalies (F.A.) ?
  - Transient
  - Persistent
  - Dominant Persistent
- CE Ans:
- Yes!
- Dominant Persistent F.A.s please.

Problem Formulation: Flow Anomaly



Two Use Cases:

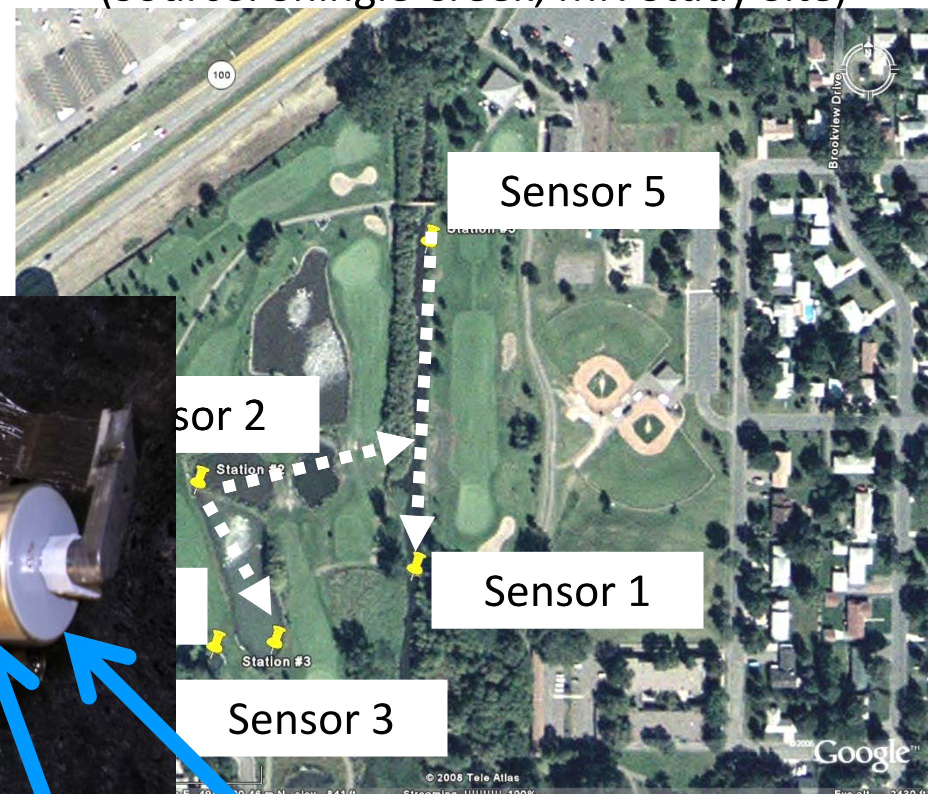
- At the water treatment plant, when should it turn off the water supply from the river?

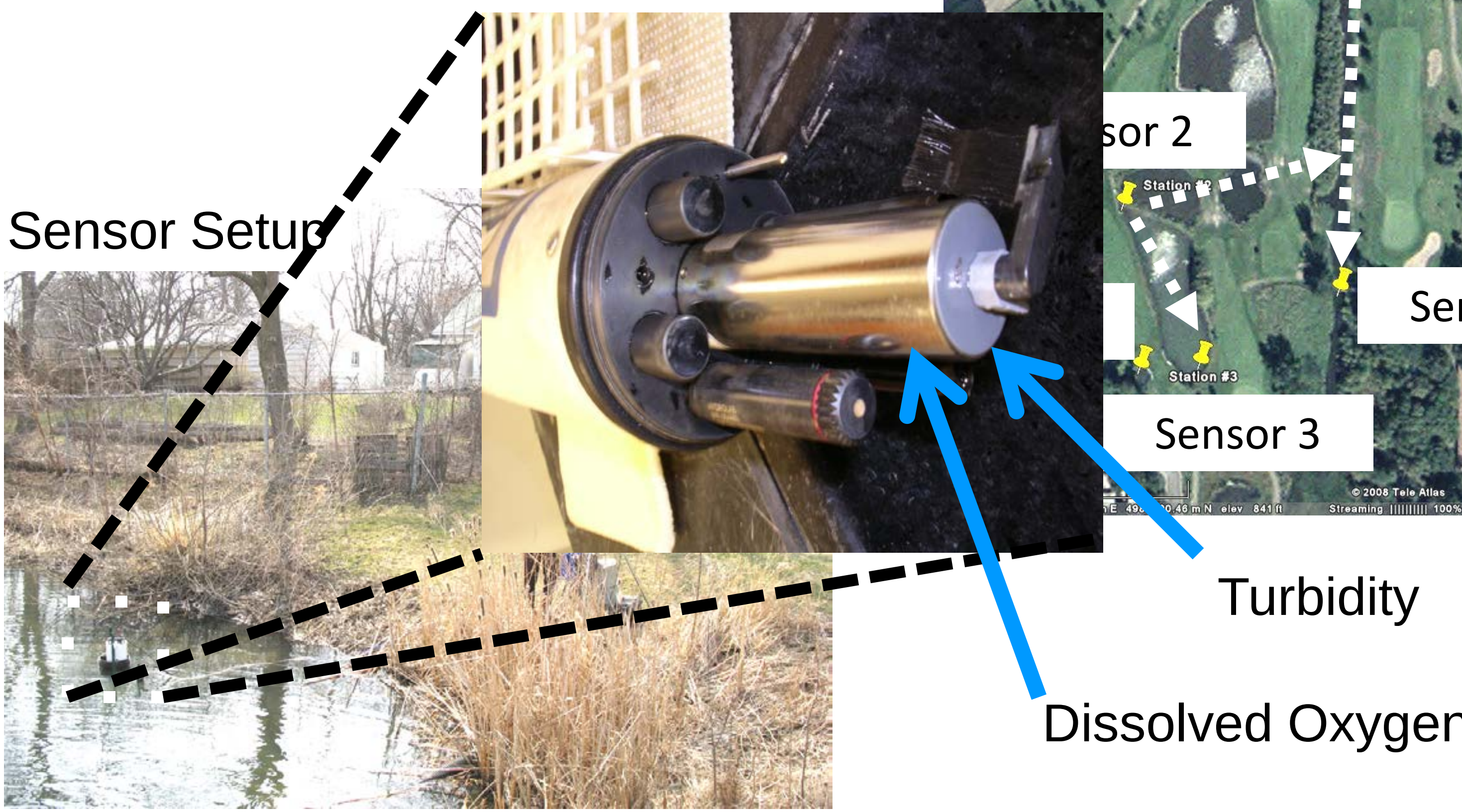
Where is the source of the contaminant?

Evaluation with Field Data Sets

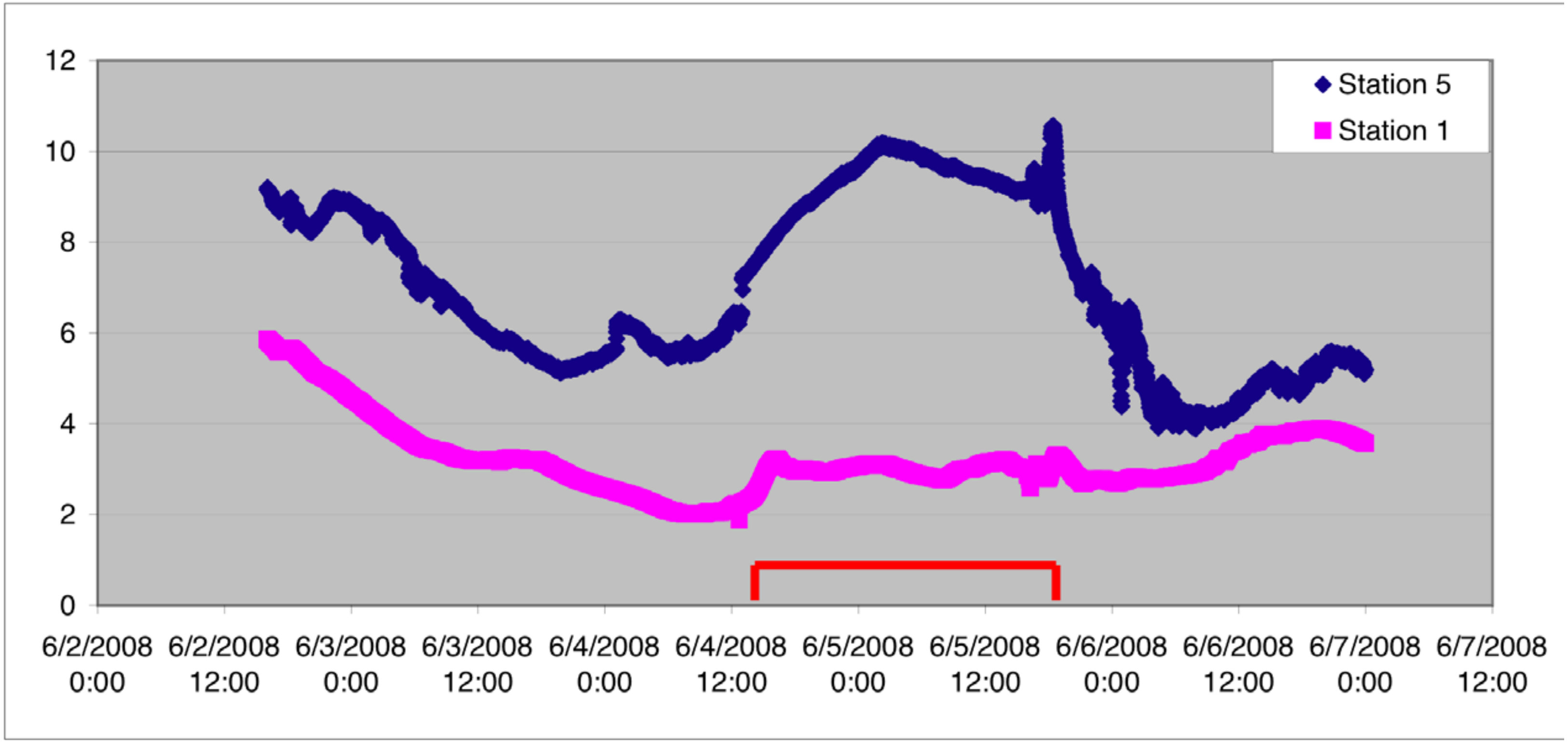
- Shingle Creek
  - Stations 5 to 1
  - Dataset 1: Turbidity: 3K-15K time intervals
  - Dataset 2: DO: 5K time intervals

(Source: Shingle Creek, MN Study Site)





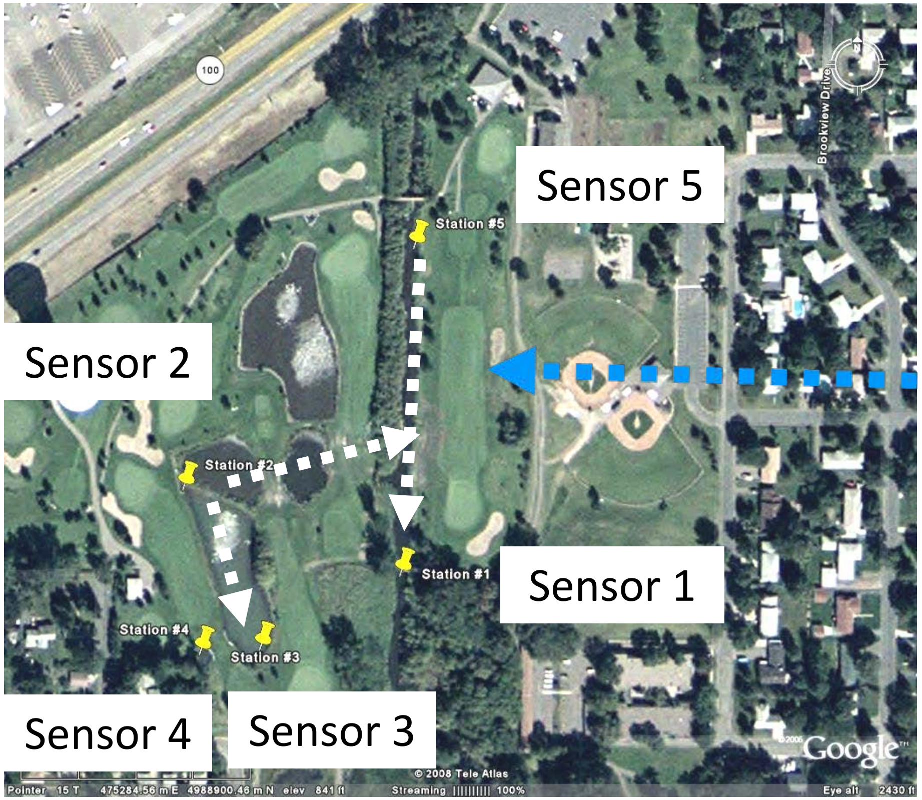
Domain-based Validation: Dissolved Oxygen



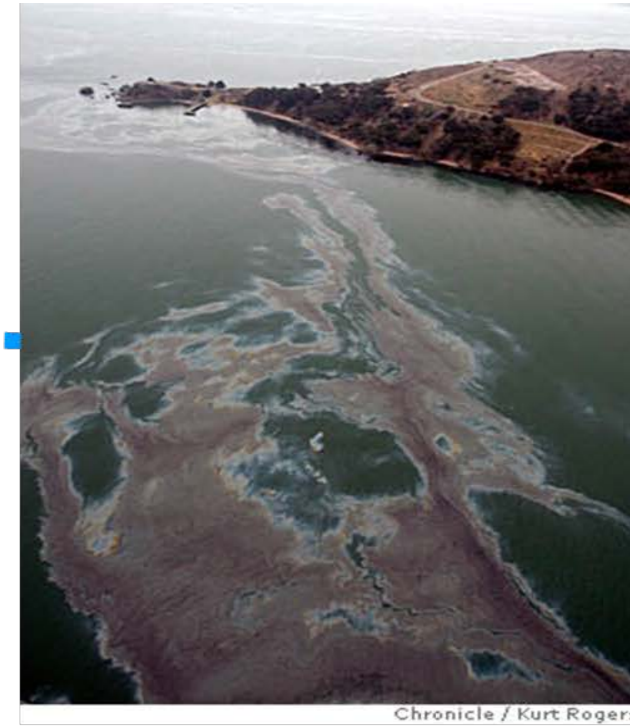
Longest Flow Anomaly Result (Error: +/- 5, Persistent: 80%)

Start: 6/4/2008 13:06 End: 6/5/2008 19:34

Domain-based Validation



(Source: Shingle Creek, MN Study Site)



(Source: http://www.sfgate.com/cgi-bin/news/oilspill/busan)

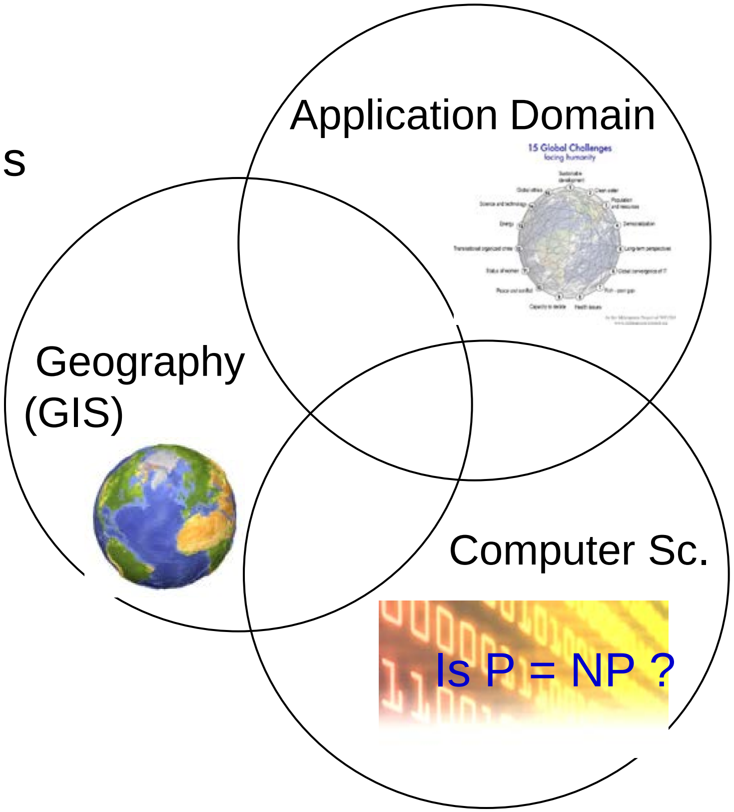
Contributions:

- **Computer Science:** New algorithm for problems where dynamic programming fails
- **Environmental Science:** Localize pollution sources
  - Explanation: the retention pond near sensor 4 has very low DO
  - After rain event, retention pond water flushes into stream between sensors 5 and 1
  - Resulting in a Flow Anomaly for DO

Outline

- Binary Interdisciplinary War Story
  - Computer Science (CS) advancing a Domain Science
  - Advance both Computer SC. and Domain Science

- Ternary Interdisciplinary War Story
  - GIS and CS together address societal needs
  - Problem Formulation Constraints
    - Deliver value to application domain
    - Computational Feasibility
    - (Geographic) Data Feasibility



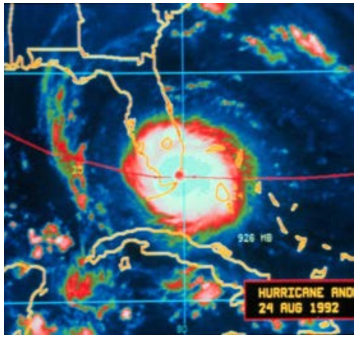
Large Scale Evacuation

Hurricane: Andrews, Rita

- Traffic congestions on all highways
    - E.g. 100-mile congestion (TX)
- Great confusions and chaos

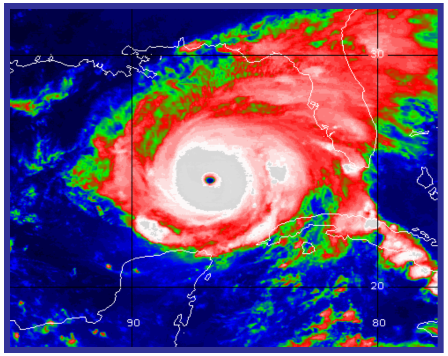
"We packed up Morgan City residents to evacuate in the a.m. on the day that Andrew hit coastal Louisiana, but in early afternoon the majority came back home. **The traffic was so bad that they couldn't get through Lafayette**."
Mayor Tim Mott, Morgan City, Louisiana
(http://i49south.com/hurricane.htm)

Florida, Lousiana (Andrew, 1992)



(National Weather Services)

Houston (Rita, 2005)



(National Weather Services)

(www.washingtonpost.com)

I-45 out of Houston

(FEMA.gov)

Problem Statement: Evacuation Route Planning

Given

- Number of evacuees and their initial locations
- Evacuation destinations
- A transportation network, a directed graph ***G* = (*Nodes, Edges*)** with
  - Travel time for each edge
  - Capacity constraint for each edge and node

Output

- Evacuation plan = a set of origin-destination routes & schedule

Objective

- Minimize evacuation time

Constraints

- (Spatio-temporal) Data Feasibility
- Computational Scalability to large population and geographies
- (Societal) Application Domain Interpretability and value
  - Transportation Network Capacity << Number of evacuees

Spatio-temporal Data Feasibility

- **(Spatio-temporal) Data Feasibility**
  - Is it doable with available geographic dataset ?
  - Does it require new geo-datasets, surveys, mapping efforts?
- **Digital Road Maps**
  - Are Navteq, OpenStreetMap digital roadmaps adequate?
  - Capacity constraints: TP+ (Tranplan) has major roads
- **Evacuee Population**
  - Is Census Data adequate? Timely?
  - Day-time population – employment, school, tourist, events, …
  - Location-aware Smart-phones ?estimate real-time population

Computational Feasibility

Computational Scalability to large population and geographies

Is it computable in reasonable time with current hardware, software ?

Does it require computer science advances?

A. Related Work in Transportation Sc.: Dynamic Traffic Assignment

- Game Theory: Wardrop Equilibrium, e.g. DYNASMART (FHWA), DYNAMIT(MIT)

Limitation: Nice Game Theory,but Extremely high computation time

- Does not scale to medium size networks and populations!

B. Related Work Operations Research: Time-Expanded Graph + Linear Prog.

- Optimal solution, e.g. EVACNET (U. FL), Hoppe and Tardos (Cornell U).

Limitation: - High computational complexity => Does not scale to large problems

- Users need to guess an upper bound on evacuation time

Inaccurate guess => either no solution or increased computation cost!

Number of Nodes	50	500	5,000	50,000
EVACNET Running Time	0.1 min	2.5 min	108 min	> 5 days

C. Computer Science: Capacity Constrained Route Planner

- Extends shortest-path algorithms to honor capacity constraints
- Scales up to to Millions of evacuees over hundreds of square miles

Win-Win

Contributions:

- Computer Science: New algorithm to scale up to metropolitan scenarios
- Transportation Science: Walking first mile speeds up evacuation by factor of 3

[Societal Impact highlighted in Fox TV (KMSP) evening news]

A 4-miute video-clip is at http://www.cs.umn.edu/~shekhar/talk/video/fox9_aired_mpg.avi
It highlighted the evacuation route planning work along with its social impact including those on evacuation plan for Minneapolis-St. Paul Twincities.


TWIN CITIES METRO
EVACUATION PLAN

TECHNICAL
MEMORANDUM #1

UNCLASSIFIED/FOR OFFICIAL USE ONLY (FOUO)
MAY BE DISSEMINATED ON A "NEED TO KNOW" BASIS
NOT FOR MEDIA OR PUBLIC DISTRIBUTION
PROPERTY OF THE MINNESOTA DEPARTMENT OF
TRANSPORTATION

A summary paper titled "Evacuation Planning: A Spatial Network Database Approach" from IEEE Data Eng. Bulletin, 33(2): 26-31 (2010) is at http://sites.computer.org/debull/A10june/Shashi.pdf

Dimensions Beyond Data & Computing

- (Spatio-temporal) Data Availability
  - Estimating evacuee population, available transport capacity
  - Pedestrian data: walkway maps, link capacities based on width
- Traffic Eng.
  - Link capacity depends on traffic density
  - Modeling traffic control signals, ramp meters, contra-flow, …
- Evacuee Behavior (Social Science)
  - Unit of evacuation: Individual or Household
  - Heterogeneity: by physical ability, age, vehicle ownership, language, …
- Policy Decisions
  - How to gain public's trust in plans? Will they comply?
  - When to evacuate? Which routes? Modes? Shelters? Phased evacuation?
  - Common good with awareness of winners and losers due to a decision
- Science
  - How does one evaluate an evacuation planning system ?

Lessons Learned

- Interdisciplinary Research is rewarding but HARD

- Reward: Solve complex societal problems
  - Which may not be amenable to a single discipline

- Hardest part is trying to understand the other domain
  - Quote: “If you think Mathematics is hard, try learning Chinese.” - Peter Bol
  - Analogy: “If you think learning Chinese is hard, try
    - Learning language, culture and strengths of another discipline
    - Developing trust across disciplines to find win-win
    - Introspecting to understand when one needs help from another discipline

- Crucial that both sides understand each other before research can begin
- A lot of trial and error between both sides
- Once an “Ah-ha” moment occurs
  - The number of opportunities can be unlimited!