Nodal Importance Concept for Computational Efficiency in Optimal Sensor Placement in Water Distribution Systems

Scott W. Rogers[1], Jiabao Guan[1], Morris L. Maslia[2], and Mustafa M. Aral[1]

[1]Multimedia Environmental Simulations Laboratory, Georgia Institute of Technology
[2]Agency for Toxic Substances and Disease Registry

“Computational Efficiency Issues in Optimal Sensor Location”
WDSA Symposium
EWRI 2007
May 17, 2007

Georgia Tech
School of Civil & Environmental Engineering

Disclaimer

- *The findings and conclusions in this presentation have not been formally disseminated by CDC or ATSDR and should not be construed to represent any agency determination or policy.*

Georgia Tech
School of
Civil & Environmental Engineering

Background
& Motivation for Study

- Recent interest in protecting drinking water in water distribution systems (WDSs) in the event of terrorist attack by contaminant injection
- Human variables of uncertainty hinder definitive contaminant sensor placement in WDSs
- Methods documented to date
  - computationally-expensive algorithms
  - oversimplifying assumptions
  - inability for implementation on larger systems
- Method needed to increase efficiency of search for optimal placement schemes without compromising WDS protection.

Georgia Tech
School of Civil & Environmental Engineering

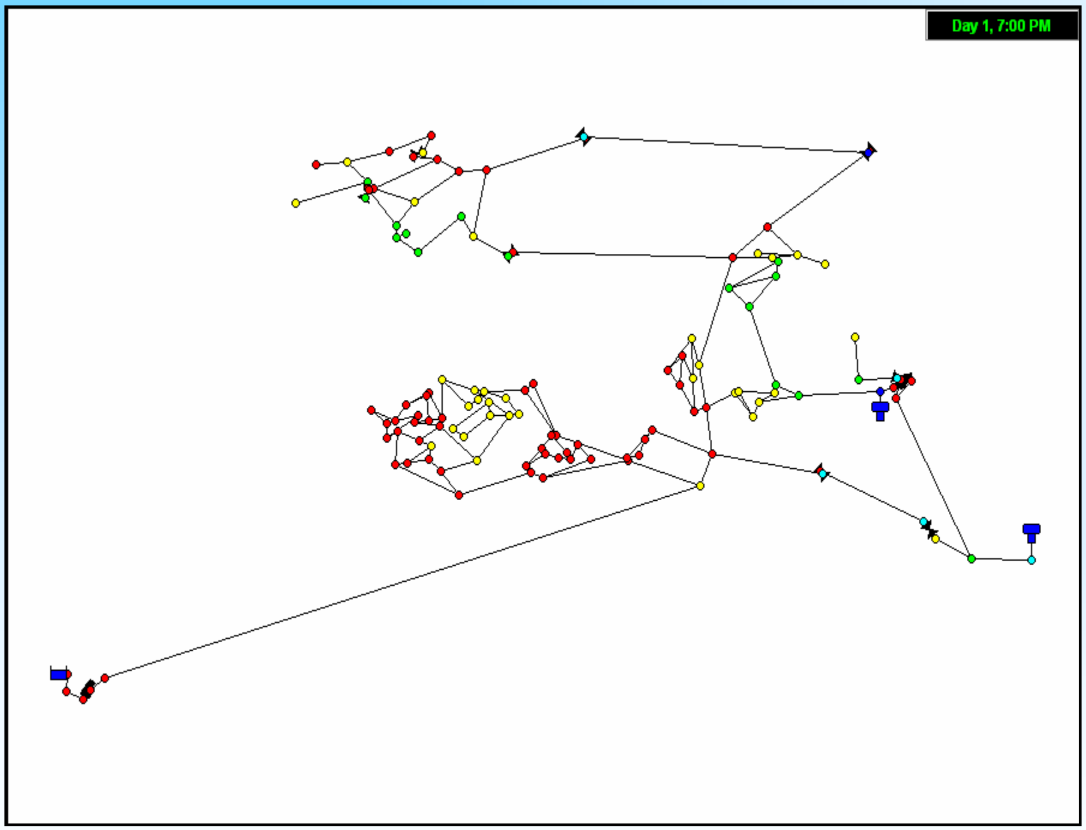
Study Problem

- Objective
  - Allocate contaminant sensors to WDS nodes in a <u>computationally efficient</u> manner to provide maximum WDS protection.
- Performance Measures
  - Z_{time} = expected detection time (minimized)
  - Z_{vol} = expected contaminated water volume (minimized)
  - Z_{lik} = detection likelihood (maximized)
- Limiting Factor
  - M = fixed number of sensors available
- Attack Scenario
  - eligible injection node: any one node in WDS
  - eligible injection time: any 5-minute multiple of first 1/4 of study period
  - injection: constant mass flow of conservative contaminant

Georgia Tech
School of Civil & Environmental Engineering

Study System

- BWSN Network 1 *



- 129 nodes (126 junctions, 2 tanks, 1 reservoir)
- 178 links (168 pipes, 2 pumps, 8 valves)
- localized flow behavior
- large variance in hydraulic demand (63% of junctions with demand)
- 96-hour study period

Georgia Tech School of Civil & Environmental Engineering

* "Battle of the Water Sensor Networks," *8th Annual WDSA Symposium*

Nodal Importance Concept (1)

- Nodal Importance Defined
  - degree to which an individual WDS node should be considered as a candidate for sensor placement
  - related to potential on average for adverse effects to be experienced at an individual node under an unknown attack scenario
- Use of Concept
  - isolate a subset of “more important” nodes to confine search domain
  - test different combinations of subset nodes with optimization algorithm to find sensor placement scheme providing maximum protection
- Expected Advantages
  - better-performing sensor placement schemes
  - reduced computational runtimes

Nodal Importance Concept (2)

➢Nodal Importance Function

$$f_{is} = \alpha_1 \frac{V_{is}^{cont}}{\max_i (V_{is}^{cont})} + \alpha_2 \left(\frac{T_s - t_{is}^d}{T_s} \right) D_{is}$$

V_{is}^{cont} = total contaminated volume associated with node ***i*** under scenario ***s*** assuming no contaminant detection at node ***i***

t_{is}^{d} = time after injection during scenario ***s*** when contaminant is first present at node ***i*** (if not present at any time, t_{is}^{d} = 2 x study period duration)

T_s = time after injection at the end of study period for scenario ***s***

D_{is} = 1 or 0, indicating contaminant presence or absence, respectively, of contaminant at node ***i*** at any time during scenario ***s***

α_1, α_2: scalars in domain [0, 1] ($\alpha_1 + \alpha_2 = 1$)

Georgia Tech
School of Civil & Environmental Engineering

Nodal Importance Concept (3)

- Weighting of importance function terms
  - Use array of ($\boldsymbol{\alpha_1}$, $\boldsymbol{\alpha_2}$) schemes to capture nodes according to different protection preferences
  - ($\boldsymbol{\alpha_1}$, $\boldsymbol{\alpha_2}$) = **(1, 0)**, **(0, 1)**, & **(1/2, 1/2)** for this study
- Relative Importance

$$f_{is}^{rel} = \frac{f_{is}}{\max_i (f_{is})}$$

- Expected Relative Importance

$$F_i = \frac{1}{S} \sum_s f_{is}^{rel}$$

S = number of Monte Carlo scenarios run

  - Used to rank nodes for a particular ($\boldsymbol{\alpha_1}$, $\boldsymbol{\alpha_2}$) scheme

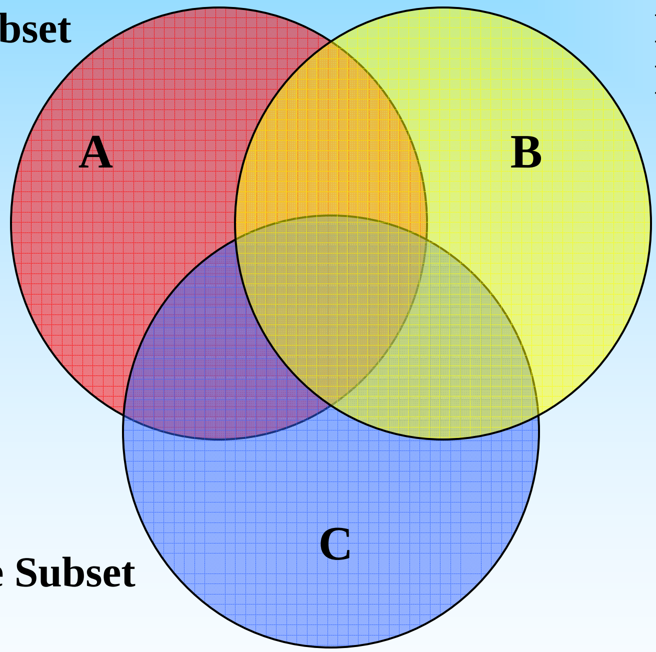
Georgia Tech
School of Civil & Environmental Engineering

Subset Creation (1)

- Base Subsets to Total Subset
  - each "base subset" of more-important nodes corresponds to a particular (α_1, α_2) scheme
  - "total subset" is the union of all base subsets generated

Base Subset A

Base Subset B



Base Subset C

$$(\alpha_1, \alpha_2)_A = (1, 0)$$

$$(\alpha_1, \alpha_2)_B = (0, 1)$$

$$(\alpha_1, \alpha_2)_C = (1/2, 1/2)$$

Base Subset Size = size(A) = size(B) = size(C)

Total Subset = A U B U C

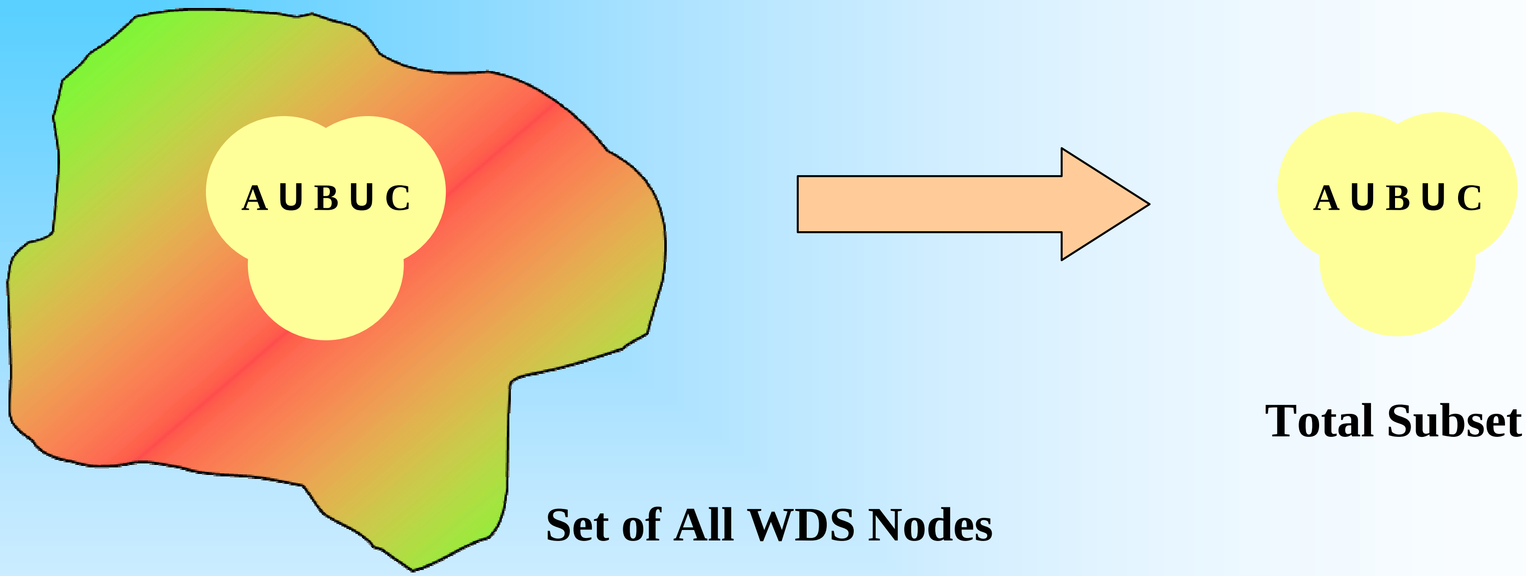
Total Subset Size = size(A U B U C)

Base Subset Size ≤ Total Subset Size ≤ (3 * Base Subset Size)

Georgia Tech

School of Civil & Environmental Engineering

Subset Creation (2)



- Only nodes in total subset tested for sensor placement
- The optimal subset: smallest subset that includes optimal nodes for sensor placement

Georgia Tech
School of Civil & Environmental Engineering

Optimization (1)

- Program

$$\max Z_{comb} = \gamma_1 \frac{\max_s(t_s^d) - Z_{time}}{\max_s(t_s^d)} + \gamma_2 \frac{\max_s(V_s^{cont,d}) - Z_{vol}}{\max_s(V_s^{cont,d})} + \gamma_3 Z_{lik}$$

$$\text{s.t.} \quad \sum_{a=1}^{A} m_a = M \quad \forall\, all\ a$$

$$m_a \in \{0,1\} \quad \forall\, all\ a$$

t_s^d = time after attack when contaminant is first present under scenario ***s*** at any node ***i*** with an assigned sensor

$V_s^{cont,d}$ = total volume contaminated under ***s*** at all ***i*** until t_s^d

m_a = 0 or 1, indicating the absence or presence, respectively, of a sensor at total subset node ***a***

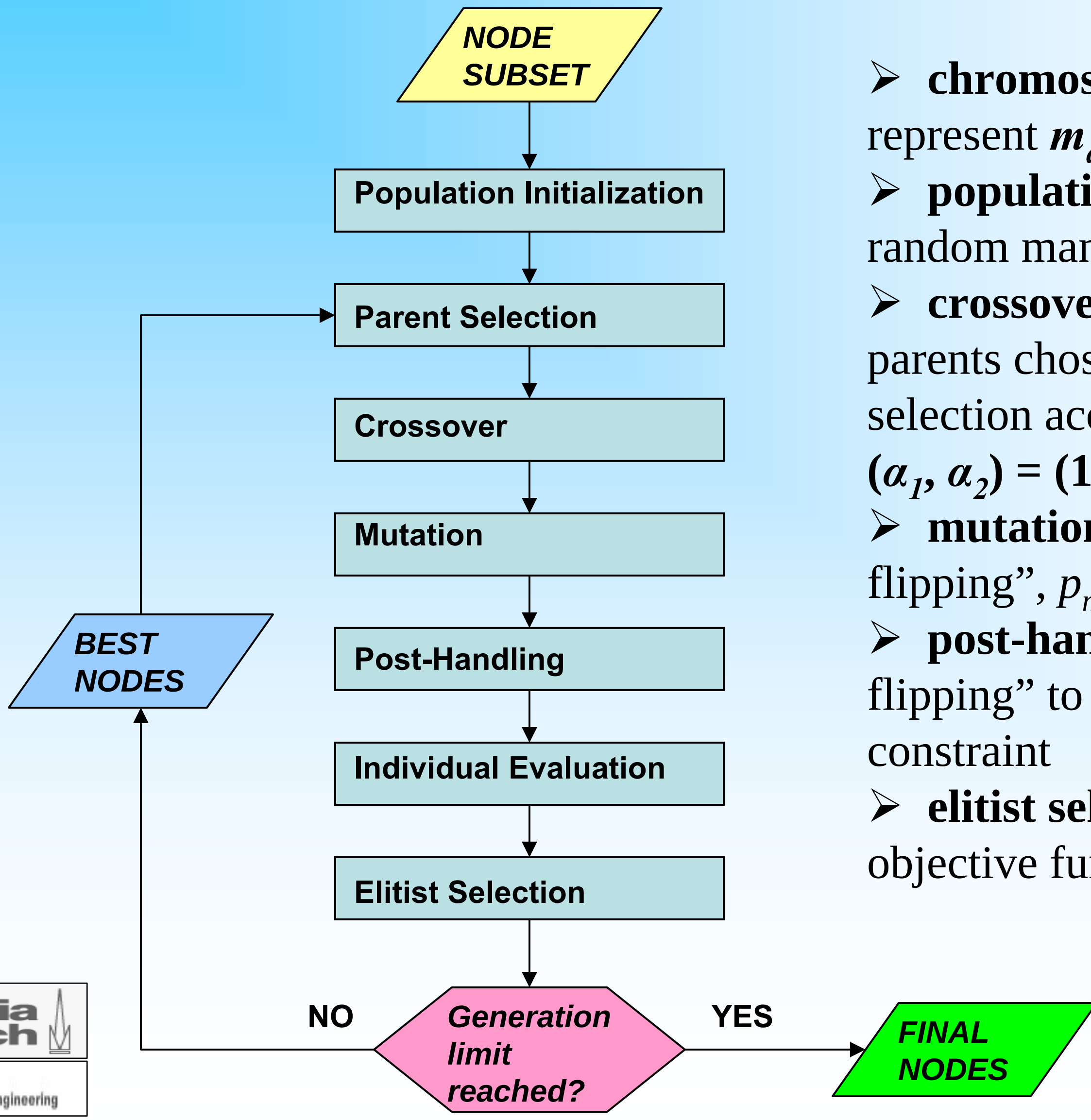
A = number of nodes in total subset

$\gamma_1, \gamma_2, \gamma_3$: scalars in the domain [0, 1] ($\gamma_1 + \gamma_2 + \gamma_3 = 1$)

Georgia Tech
School of Civil & Environmental Engineering

Optimization (2)

- "Simple" Genetic Algorithm



- **chromosome**: binary string; bits represent $\boldsymbol{m_a}$ values
- **population**: initialized in uniform, random manner
- **crossover**: one-point, $p_{cross} = 0.95$; parents chosen through roulette-wheel selection according to $\boldsymbol{F_i}$ values under $\boldsymbol{(\alpha_1, \alpha_2) = (1/2, 1/2)}$
- **mutation**: uniform, random "bit-flipping", $p_{mut} = 0.05$
- **post-handling**: uniform, random "bit-flipping" to satisfy sensor availability constraint
- **elitist selection**: according to objective function

Georgia Tech
School of Civil & Environmental Engineering

Performance Testing (1)

- Decision Variables
  - variables kept constant
    - $\boldsymbol{M} = 5$
    - $\boldsymbol{S} = 3{,}000$
    - GA population size = 500
    - number of GA generations = 500
  - designated weighting schemes
    - $(\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}) = (1, 0), (0, 1), (1/2, 1/2)$
    - $(\gamma_1, \gamma_2, \gamma_3) = (1/4, 1/2, 1/4)$
  - subset size
    - critical variable for testing use of importance concept – allowed to vary
    - base subset size candidates: 5, 10, 20, 30, 40, 50
- Computational Runtimes
  - less than 1 hour for all tests employing nodal importance concept

Georgia Tech
School of Civil & Environmental Engineering

Performance Testing (2)

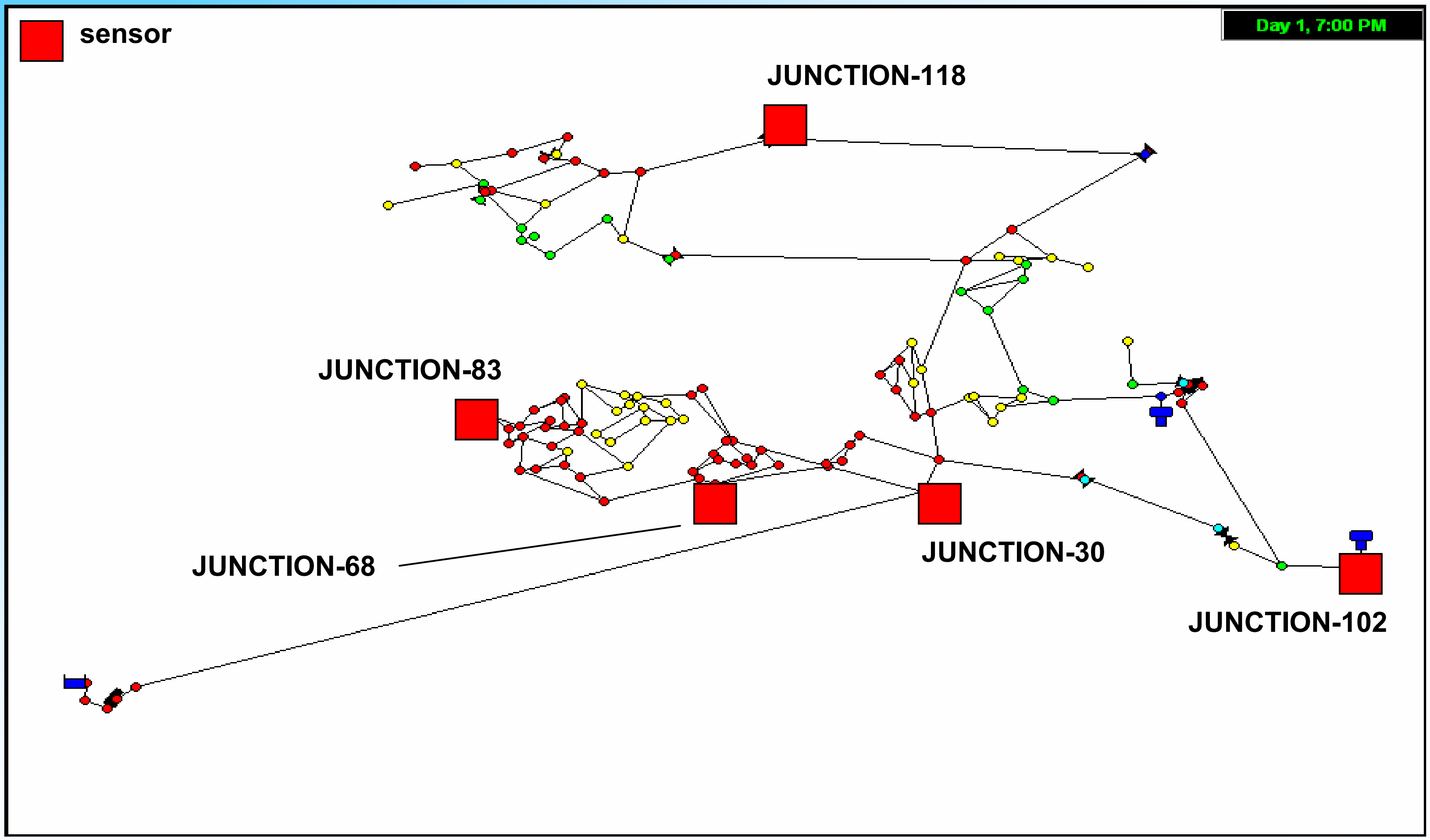
- Method Results

Method	Base Subset Size	Total Subset Size	Sensor Nodes ("JUNCTION-*x*")	Z_{time} (min)	Z_{vol} (gal)	Z_{lik} (%)	Z_{comb} x 10^2
random placement				7382	10994	37.6	20.1
GA-only		*129*	17, 49, 84, 100, 122	3783	2269	70.5	73.7
ranking-GA (a)	5	9	17, 30, 68, 83, 126	4162	3949	66.9	64.9
ranking-GA (b)	**10**	**18**	**30, 68, 83, 102, 118**	3686	1985	71.1	**75.3**
ranking-GA (c)	20	37	17, 68, 83, 100, 103	3272	2635	75.1	74.6
ranking-GA (d)	30	47	68, 83, 100, 102, 118	3034	3038	77.2	74.1
ranking-GA (e)	40	59	45, 68, 83, 103, 118	3236	2665	75.2	74.6
ranking-GA (f)	50	73	68, 83, 100, 103, 117	3222	2954	75.6	73.5

Georgia Tech
School of Civil & Environmental Engineering

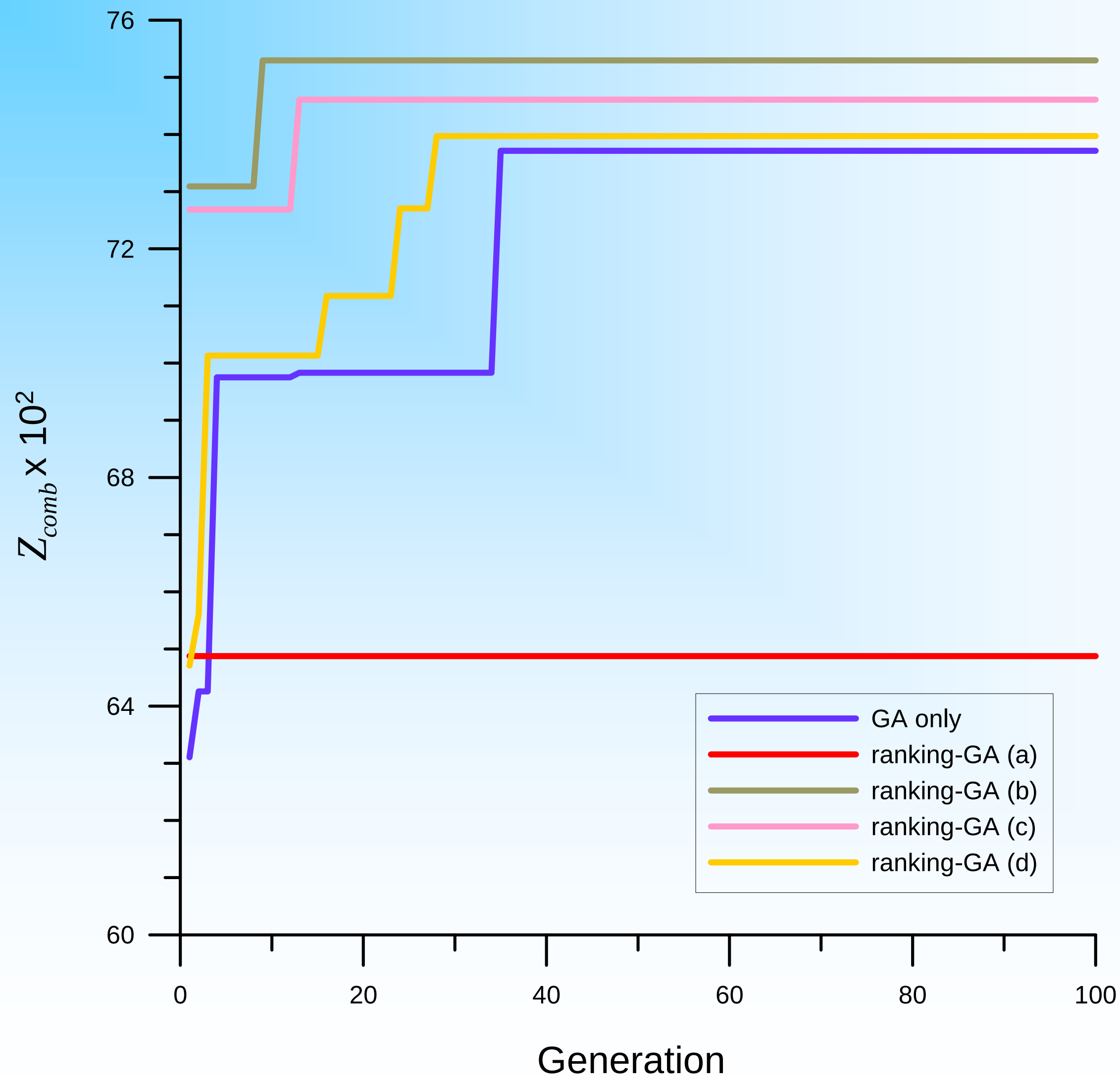
Performance Testing (3)

- Sensor Placement: Method “ranking-GA (b)”



Performance Testing (4)

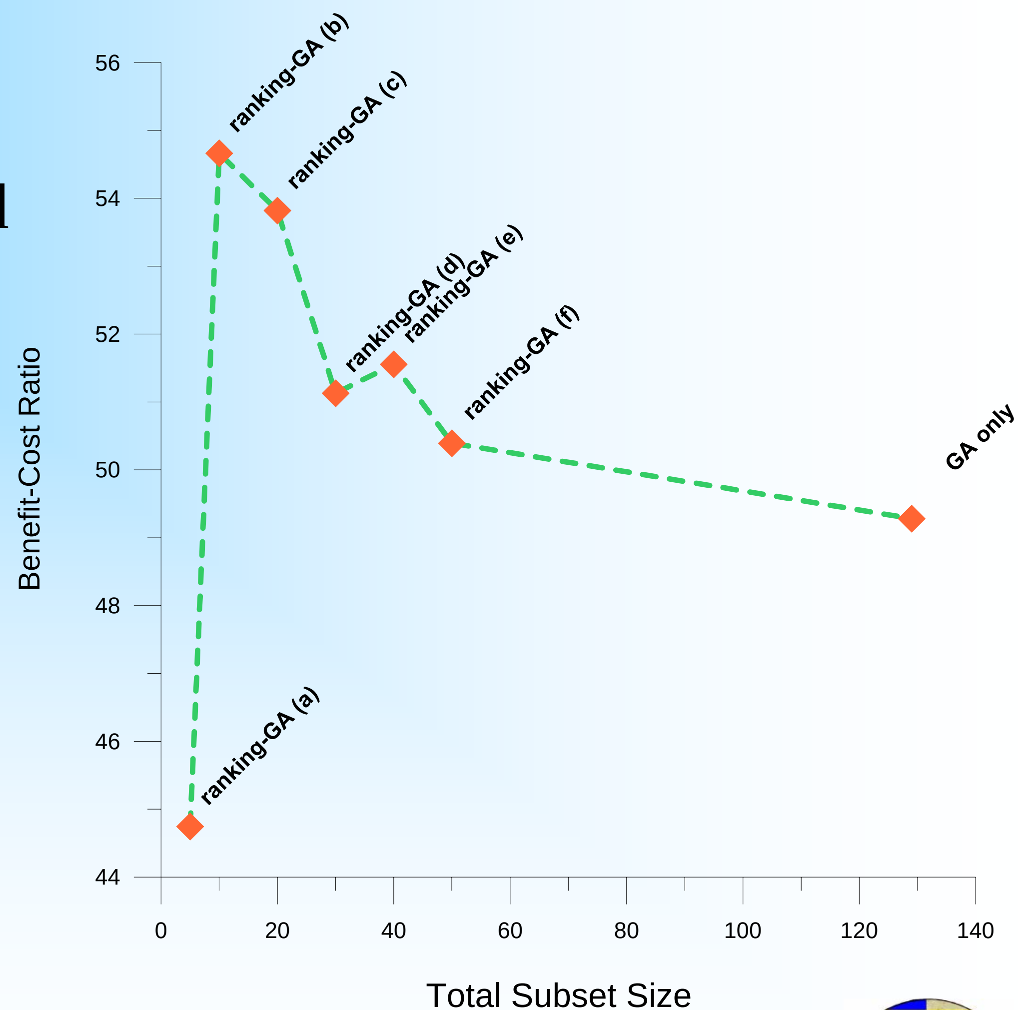
- Performance vs. GA Generation



Georgia Tech
School of Civil & Environmental Engineering

Performance Testing (5)

- Benefit-Cost Ratios
  - Benefit: gained performance for method beyond baseline from random placement
  - B/C Ratio: benefit averaged over all scenarios up to GA generation of convergence <u>for all methods</u>



Georgia Tech
School of Civil & Environmental Engineering

Observations & Conclusions

- Using nodal importance concept can lead to heightened efficiency in the optimization of contaminant sensor placement without compromising WDS protection goals.
- A subset of more-important nodes too small in size may not provide enough diversity for finding a sensor placement scheme of acceptably high performance.
- As the size of a subset increases toward the total number of WDS nodes, performance reaches a peak value then converges to a value resulting from optimization without using the importance concept.

Georgia Tech
School of Civil & Environmental Engineering

Future Work

- Applying nodal importance concept to optimization of sensor placement in larger systems
- Resolving ambiguities
  - definitiveness of importance functions & corresponding variables
  - number of Monte Carlo scenarios to run
  - subset sizes
- Developing means of faster WDS simulation

Georgia Tech

School of Civil & Environmental Engineering