High performance computing application to address non-point source pollution at a watershed level

Chetan Maringanti, Indrajeet Chaubey

Agricultural and Biological Engineering

Purdue University

PURDUE UNIVERSITY
AGRICULTURAL AND BIOLOGICAL ENGINEERING

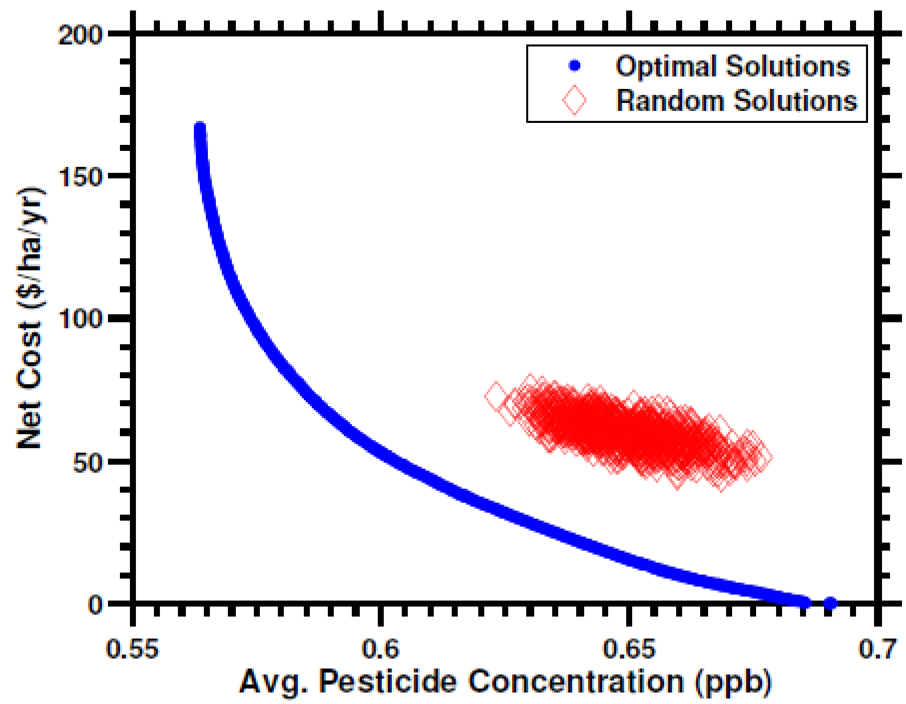
Introduction

- Best Management Practices (BMP) selection and placement problem
  - Provide optimal solutions for maximum pollution reduction and minimum increase in net costs
  - Genetic algorithm in combination with watershed model
  - BMP tool for estimation of Nonpoint Source (NPS) pollutant efficiency of the BMPs

Maringanti et al., (2009)-In press (WRR)

Evaluation of solutions from BMP optimization

- Solutions from the Pareto-optimal front are selected and evaluated using the watershed model
  - Check if the reductions are achieved at watershed level
  - 500 model simulations ~ 83 hours on single Windows machine
  - Requires using High Performance Computing (HPC)

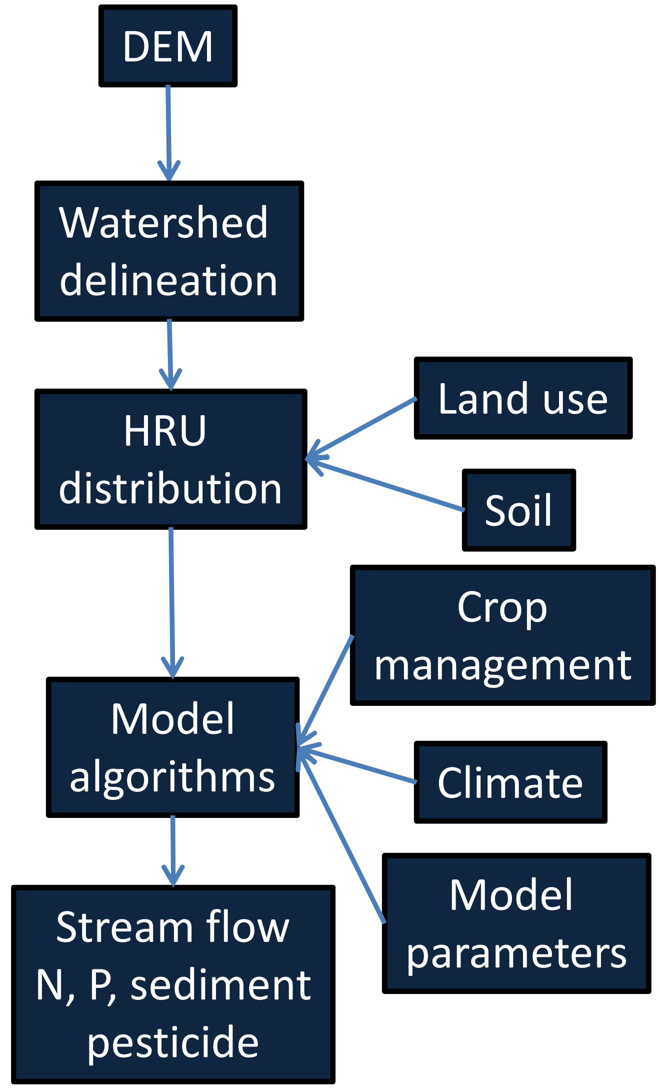


Objective and tasks

- Evaluate the performance of HPC based on a UNIX based cluster when compared to the performance of a single processor in windows
  - The BMP optimization tool based on a genetic algorithm is run for nitrogen, phosphorus, and sediment reduction in Lincoln Lake Watershed, Arkansas
  - The validation of the solutions obtained from the optimization of the BMP optimization tool are used to evaluate the pollutant loads using a watershed model to estimate nitrogen, phosphorus, and sediment in L'Anguille River Watershed; and pesticide in Wildcat Creek Watershed

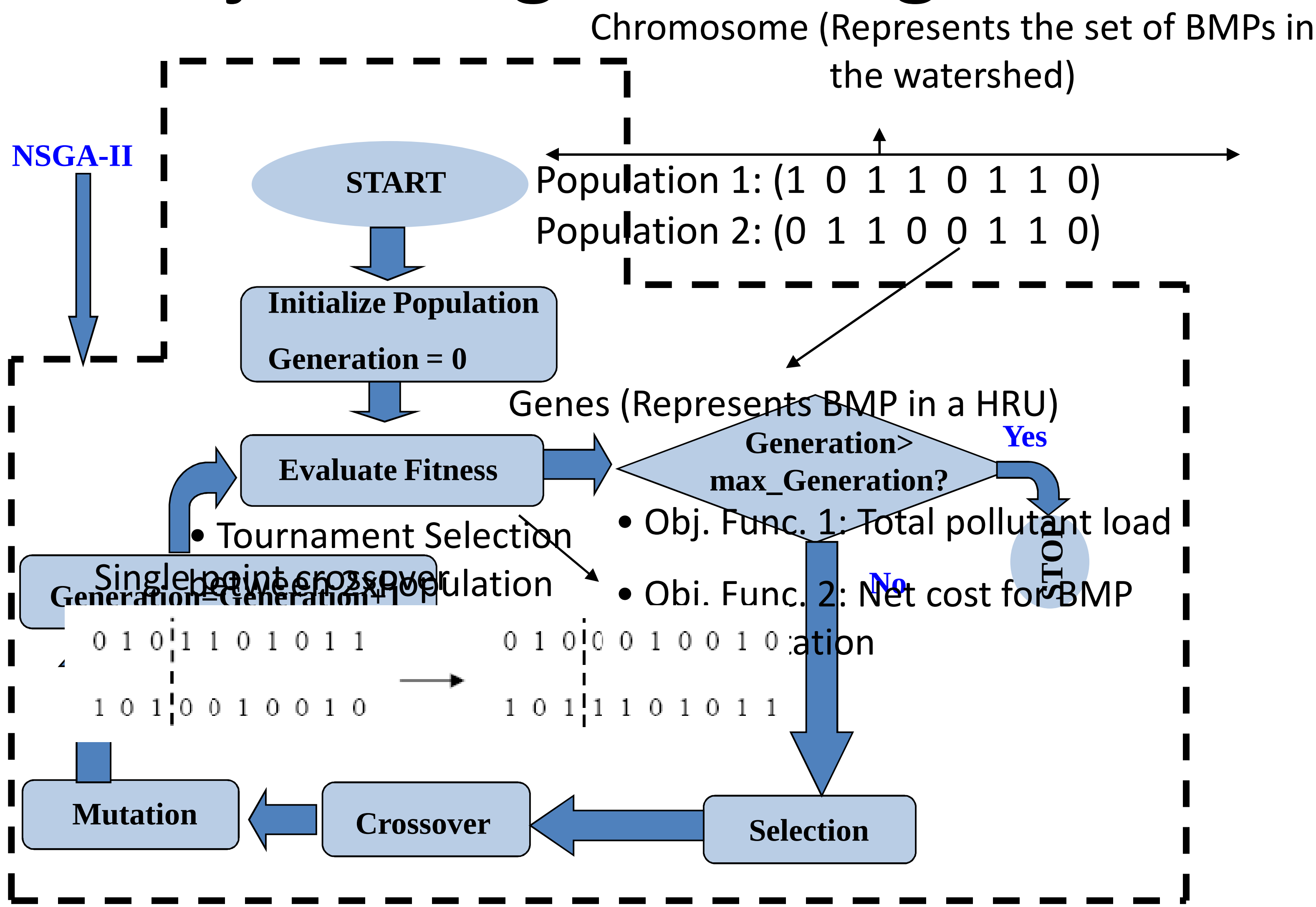
Soil and Water Assessment Tool (SWAT)

- Soil and Water Assessment Tool (SWAT)
  - Hydrologic and water quality modeling
  - Simulate stream flow, nitrogen (N), phosphorus (P), sediment, and pesticide
  - Agricultural management practices
  - Dynamic land use
  - Climate change



PURDUE UNIVERSITY
AGRICULTURAL AND BIOLOGICAL ENGINEERING

Multi-objective genetic algorithm

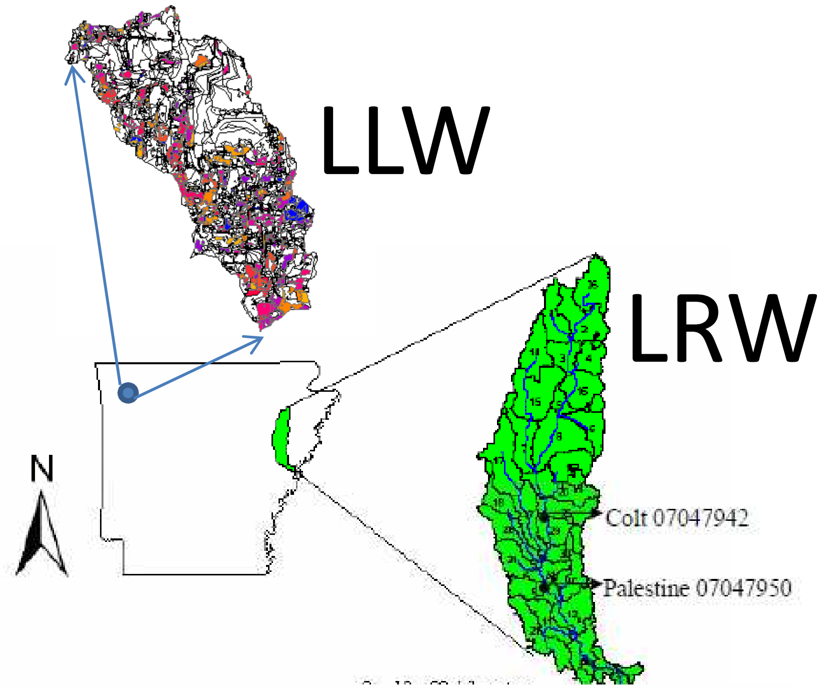


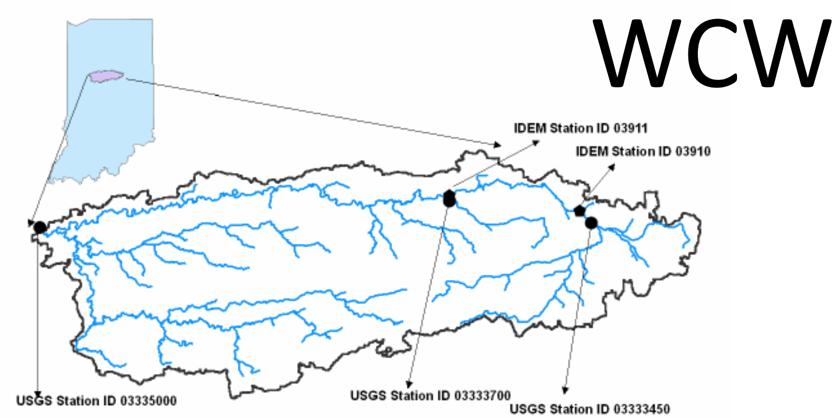
High performance computing at Purdue University

- Steele super computer
  - Red Hat Linux operating system
  - Dell dual quad-core computer nodes
  - Estimated peak performance of 60 teraflops
  - Largest super computer in the Midwest univesities
  - 104 in the TOP500 list of global supercomputers

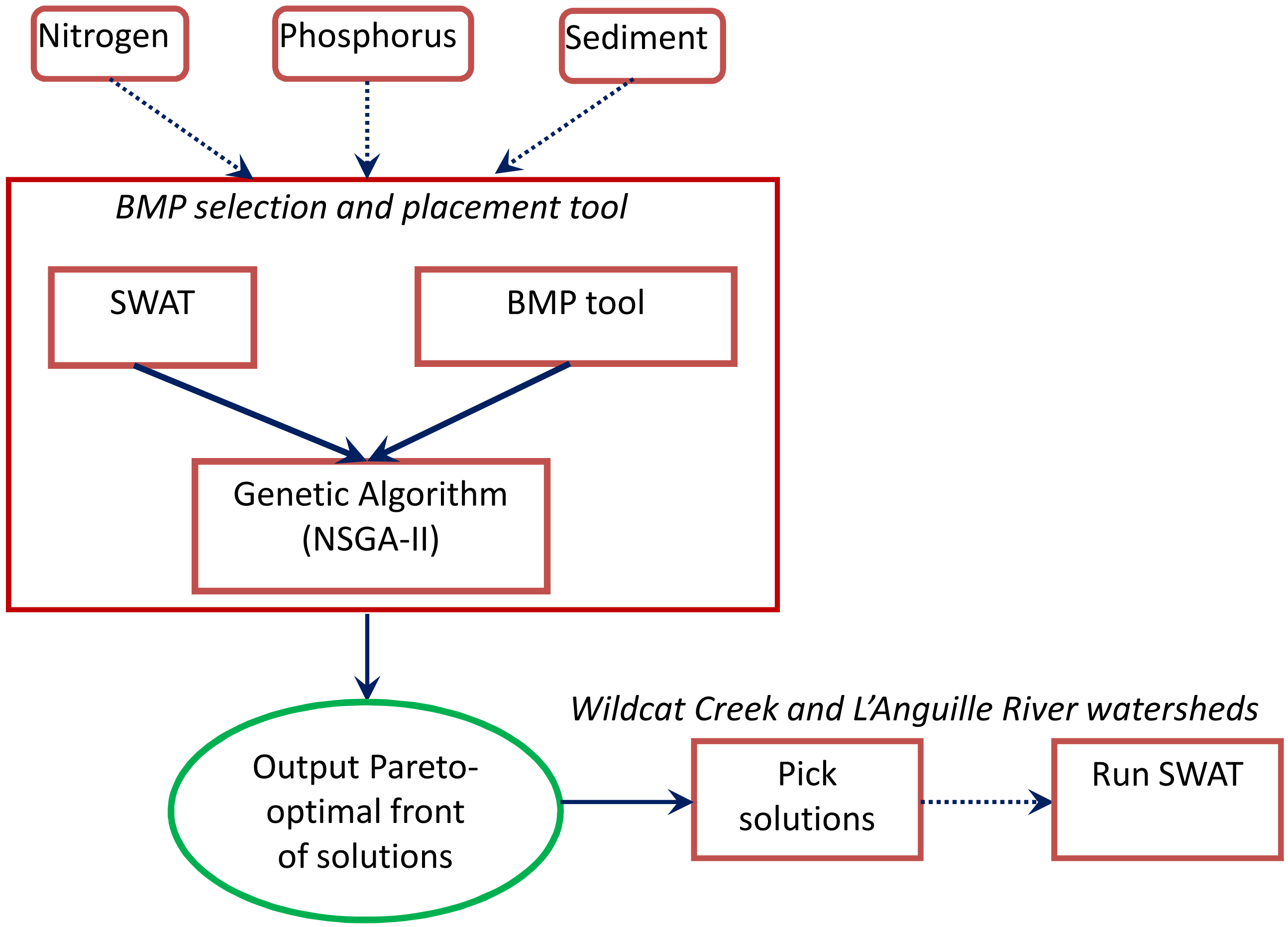
Study area

- Lincoln Lake watershed (LLW)
  - Located in Northwest Arkansas
  - Area: 32 km^2
  - Pasture dominated
- L'Anguille River watershed (LRW)
  - Located in Eastern Arkansas
  - Area: 2520 km^2
  - Rice and soybean dominated
- Wildcat Creek watershed (WCW)
  - Located in Northcentral Indiana
  - Area: 2300 km^2
  - Corn and soybean dominated





Methodology

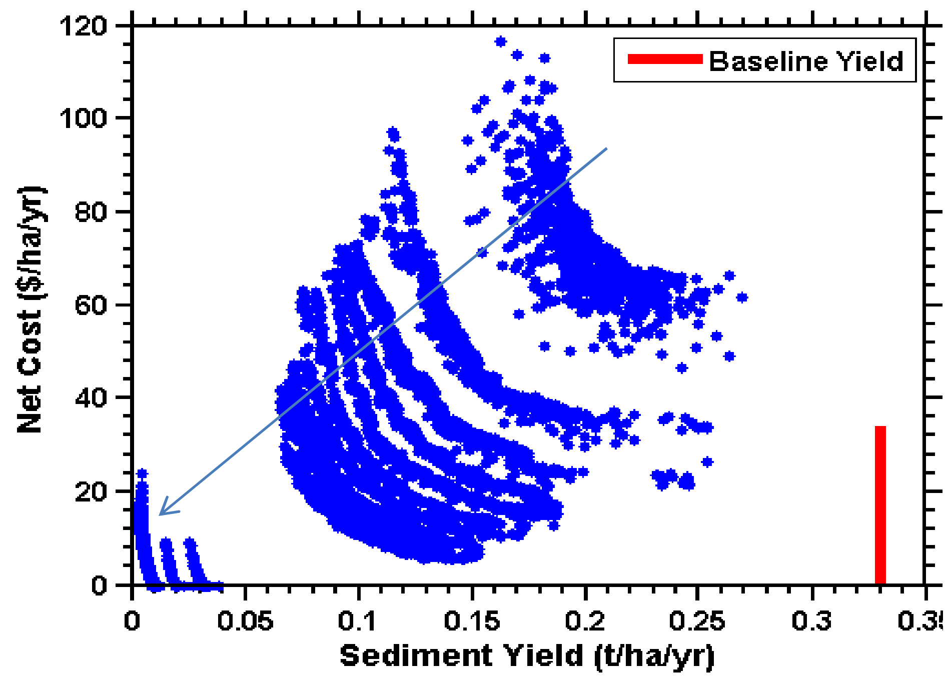


Results and discussion

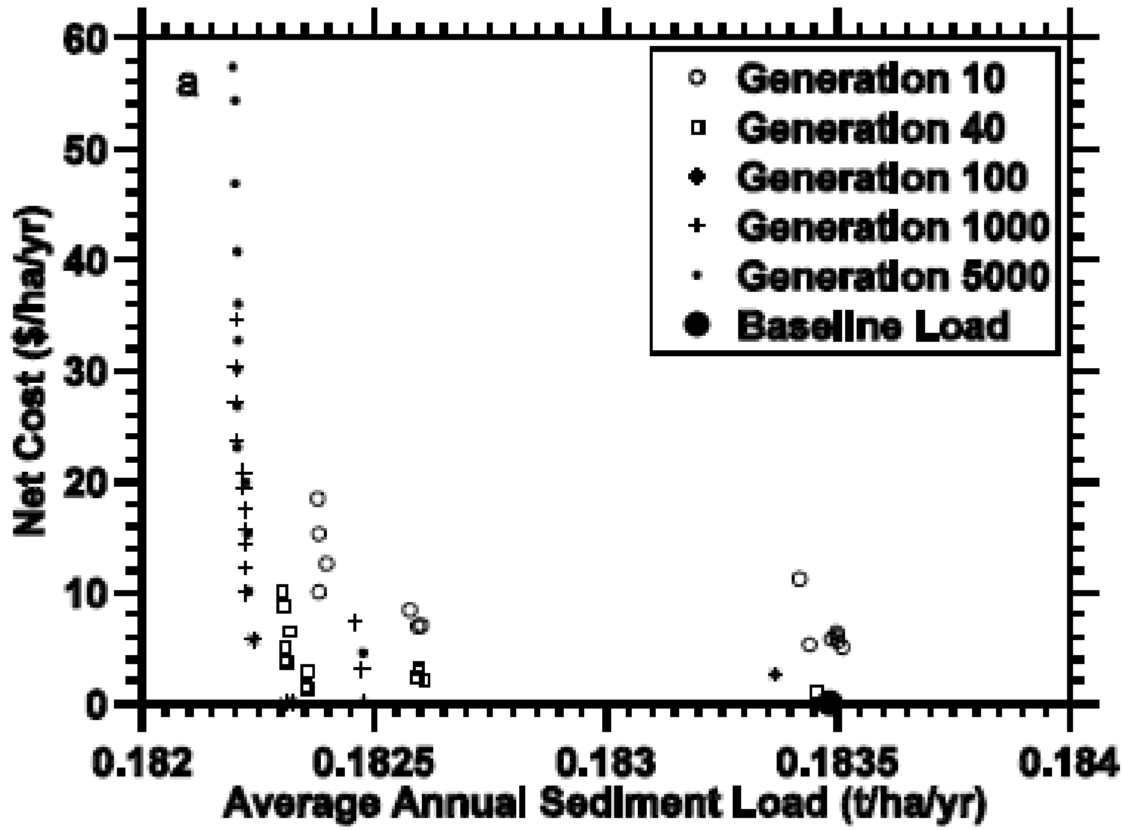
Watershed	Years of simulation	No. of HRUs	Computation time	
			Windows	UNIX
L'Anguille River Watershed	18	433	10 mins	30 secs
Wildcat Creek Watershed	7	403	4 mins	15 secs
Lincoln Lake Watershed	28	1461	30 mins	3 mins

Progress of solutions in LLW

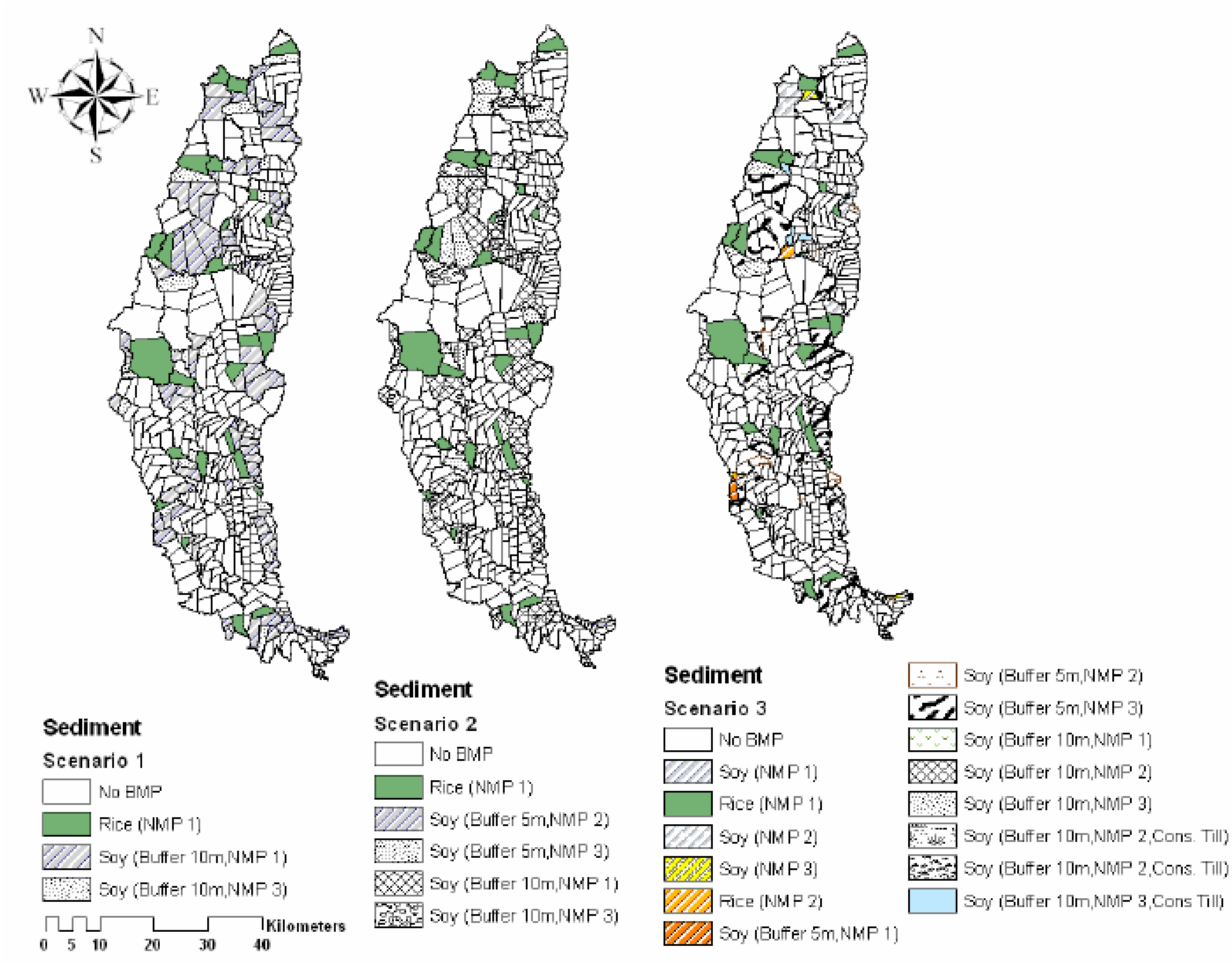
- The solutions get closer to origin (reduction in both the objective functions) as the number of generations increase
- Generations 10, 40, 100, 1000, 5000 selected for simulation with the watershed model



Simulation of the solutions obtained during optimization using the SWAT model for estimating the sediment

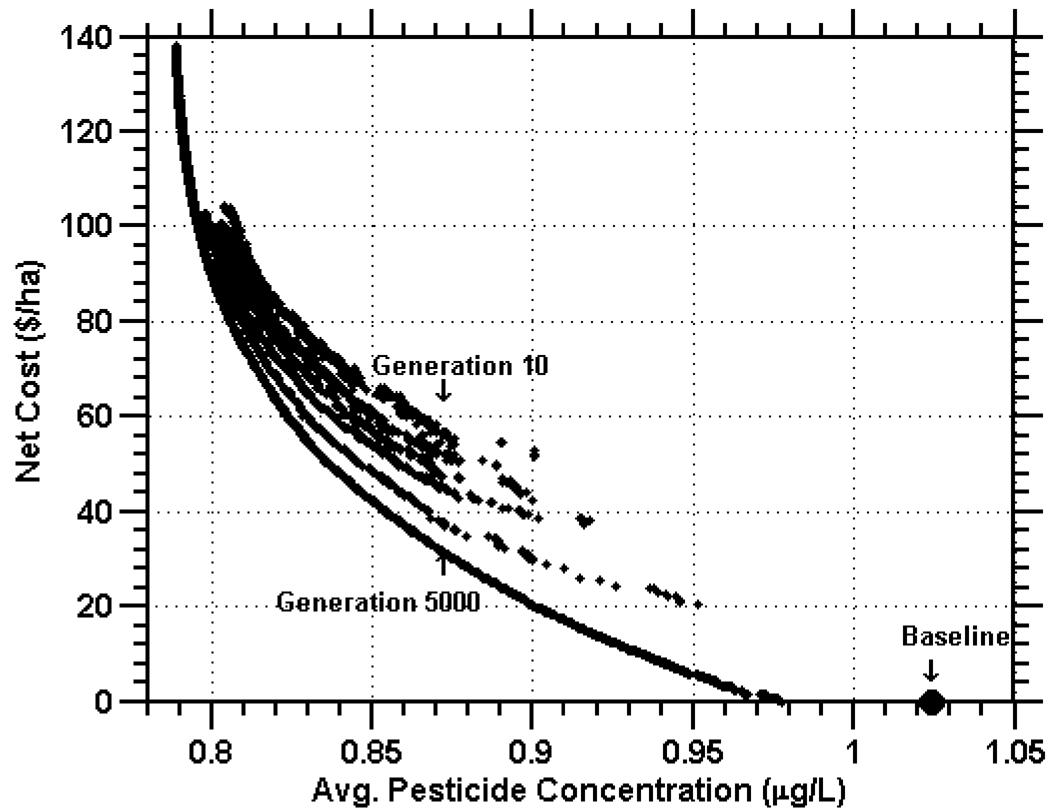


Optimal solution representation in the watershed



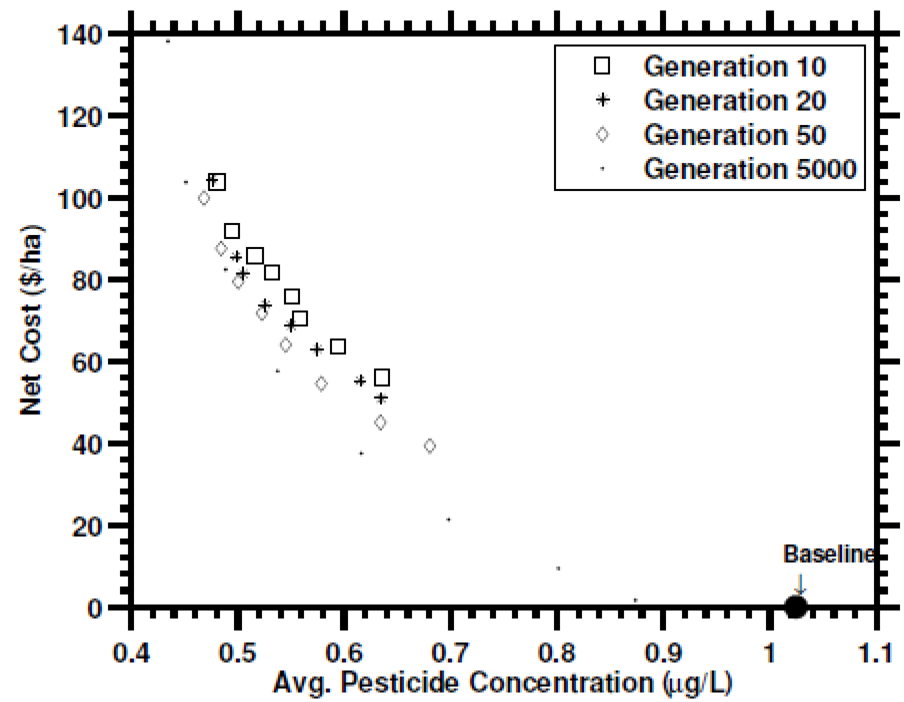
Progress of the Pareto-optimal front in WCW

- 5000 generations
- 800 population
- 8 solutions from the front at end of 10, 20, 50, and 5000 generations selected



Performance of the solutions when implemented at the watershed level using SWAT

- Solutions improve with the generation at the watershed scale



Conclusions

- HPC reduces the computation time by distributing the different processes to different processes
  - More than 90% reduction in computation
- UNIX based machines are computationally more efficient when compared to Windows machines
  - More than 10 times faster

Future work

- Use HPC to perform the BMP selection and placement where the watershed model is dynamically linked with the optimization algorithm
  - Distribute the various population in each generation to different slave computer (processors)
  - Host computer will collect the simulations from slave computers and proceed to next generation

THANK YOU

Questions??