Water resources system analysis: tools and challenges

Avi Ostfeld

Professor
Faculty of Civil and Environmental Engineering
Technion - Israel Institute of Technology
Haifa, ISRAEL
Email: ostfeld@tx.technion.ac.il
Web page: www.technion.ac.il/~avi/avi.htm
Outline

- Scientific and practical challenges in modeling water resources systems *(Part 1)*
- Example application: “A Hybrid Model Tree (MT) - Genetic Algorithm (GA) Scheme for Toxic Cyanobacteria Predictions in Lake Kinneret” *(Part 2)*
- Questions
Outline

Part 1

1. Introduction
2. Water resources systems basic components
3. Simulation tools
4. Optimization tools
5. Systems perspective
6. Integration methodologies

Part 2

7. Example application
Outline

1. Introduction
   - Water resources systems
     - Basic components
   - Simulation tools
   - Optimization tools
   - Systems perspective
   - Integration methodologies
   - Example application
Introduction

- **General**: water resources systems analysis - operations research (OR) of “water components” (e.g., reservoirs, rivers, watersheds, groundwater, distribution systems, etc.), as standalone or integrated, for single or multiobjective problems, deterministic or stochastic

- **Specific**: water distribution systems analysis (water security), watershed management, surface water

- **Tools**: “traditional” OR (LP, NLP), data driven modeling (neural networks, model trees), evolutionary computation (Genetic Algorithms, Ant Colony, Cross Entropy) - single/multiobjective
Introduction

- The scientific and practical challenge in dealing quantitatively with water resources management problems is in taking into consideration from a systems perspective, social, economical, environmental, and technical dimensions, and integrating them into a single framework for trading-off in time and in space competing objectives.

- Inherently, such problems involve modeling of water quantity and quality for water resources systems components, such as: surface water, groundwater, water distribution systems, reservoirs, rivers, lakes, and others, as stand alone or linked elements.
Introduction

- **Conceptual issues** which should be overcome for constructing and solving a water resources management problem.

- **Description of available tools** for dealing with water resources management problems: available *simulation programs* (e.g., WMS, GMS, SWAT, AVGWLF, EPANET), *optimization methodologies* (e.g., simulated annealing, genetic algorithms, ant colony, cross entropy, non-linear programming, linear programming).

- **Integration approaches**: "embedding"; "linking"; "hybridizing".

- **Case study** for demonstrating the "hybridizing" approach.
Outline

Introduction

2. Water resources systems basic components

Simulation tools

Optimization tools

Systems perspective

Integration methodologies

Example applications
River basins can be viewed as the integrating units of water resources components from an integrated water resources systems management perspective. Everything interacts within river basins.

Outline

Introduction

Water resources systems
basic components

3. Simulation tools

Optimization tools

Systems perspective

Integration methodologies

Example application
Simulation tools

- Surface water - WMS, BASINS, others

**WMS - Watershed Modeling System**

Merging information obtained from terrain models and GIS with lumped parameter traditional hydrologic analysis models such as HEC-1 and TR-20, having an ability to take advantage of digital terrain for hydrologic data development.

WMS uses three primary data sources for model development:

1. Geographic Information Systems (GIS) Data
2. Digital Elevation Models (DEMs)
3. Triangulated Irregular Networks (TINs)

http://www.ems-i.com/
Simulation tools

- Groundwater - GMS, Visual MODFLOW, others

GMS - Groundwater Modeling System

Providing GIS tools for groundwater simulation including site characterization, model development, calibration, post-processing, and visualization, supporting both finite-difference and finite-element models in 2D and 3D including links to MODFLOW, MODPATH, and others

http://www.ems-i.com/
Simulation tools

- Water networks - EPANET, WaterCad, others

**EPANET**

Performing *extended period simulation of hydraulic and water-quality behavior* within pressurized pipe networks consisting of pipes, consumers, pumps, valves and storage tanks

http://www.epa.gov/nrmrl/wswrd/epanet.html
Simulation tools

- Lakes and reservoirs - CE-QUAL-W2, ELCOM, others

CE-EQUAL-W2

A two-dimensional, longitudinal/vertical, hydrodynamic and water quality model for reservoirs, lakes and estuaries, which models basic eutrophication processes such as temperature-nutrient-algae-dissolved oxygen-organic matter and sediment relationships.

http://www.ce.pdx.edu/w2/
### CE-QUAL-W2 Hydrodynamic equations

#### X- momentum

\[
\frac{\partial U_B}{\partial t} + \frac{\partial U_B U_B}{\partial x} + \frac{\partial W_B U_B}{\partial z} = g_B \sin \alpha_s + g_B \cos \alpha_s \frac{\partial \eta}{\partial x} \frac{\partial \rho}{\partial x} - \frac{g_B \cos \alpha_s}{\rho} \int_{\eta} \frac{\partial \rho}{\partial x} \partial z + \frac{1}{\rho} \frac{\partial B \tau_{xx}}{\partial x} + \frac{1}{\rho} \frac{\partial B \tau_{xz}}{\partial z} + q_B U_x
\]

#### Continuity

\[
\frac{\partial U_B}{\partial x} + \frac{\partial W_B}{\partial z} = q_B
\]

#### Water level

\[
B_\eta \frac{\partial \eta}{\partial t} = \frac{\partial}{\partial x} \int_{\eta} U_B dz - \int_{\eta} q_B dz
\]

#### Equation of state

\[
\rho = f(T, TDS, SS) = \rho_T + (\Delta \rho_{sal} \text{ or } \Delta \rho_{TDS}) + \Delta \rho_{SS}
\]

#### Turbulent

\[
\frac{\partial B \Phi}{\partial t} + \frac{\partial U_B \Phi}{\partial x} + \frac{\partial W_B \Phi}{\partial z} - \frac{\partial}{\partial x} \left( BD_x \frac{\partial \Phi}{\partial x} \right) - \frac{\partial}{\partial z} \left( BD_z \frac{\partial \Phi}{\partial z} \right) = q_\Phi B + S_\Phi B
\]

#### Two observations:

- The physical representation of water resources elements might be quite complex

This directs the possible solution methods and approaches employed for optimizing water resources systems.
Outline

- Introduction
- Water resources systems
  - basic components
- Simulation tools
- 4. Optimization tools
  - Systems perspective
  - Integration methodologies
  - Example application
Optimization tools

Two main categories: (1) Classical (2) Heuristic

- **Classical**: LP (Linear Programming), NLP (Nonlinear Programming, e.g. GRG = General Reduced Gradient), Dynamic Programming, others

  **Advantages**: provide analytical tools to solve optimization problems;
  **Limitations**: restricted (number of constraints, decision variables, model properties)

- **Heuristic**: Simulated Annealing, Genetic Algorithms, Ant Colony, Tabu Search, Cross Entropy, others

  **Advantages**: not restricted;
  **Limitations**: no analytical assurance of an optimal solution, highly computational intensive
Outline

Introduction

Water resources systems
basic components

Simulation tools
Optimization tools

5. Systems perspective

Integration methodologies
Example application
Systems perspective

- The goal is to find a **Pareto optimal** solution set or a non-dominated solution set.

- Each solution in the Pareto optimal set is optimal in the sense that it is not possible to improve one objective without making at least one of the others worse.

Optimize: \( F(x) = (f_1(x), f_2(x), ..., f_M(x))^T \)

Subject to:

\[ g_i(x) > 0, \quad i = 1, 2, ..., k \quad \text{Inequality constraints} \]

\[ e_j(x) = 0, \quad j = 1, 2, ..., l \quad \text{Equality constraints} \]

where \( x = (x_1, x_2, ..., x_n)^T \) is the vector of decision variables.
Outline

- Introduction
- Water resources systems
  - Basic components
- Simulation tools
- Optimization tools
- Systems perspective
- 6. Integration methodologies
- Example applications
- Challenges
Integration methodologies

Three basic approaches:

- **Embedding** - direct; traditional
- **Simulation - Optimization (NLP)** (Late 1980's)
- **Simulation - Optimization (Heuristic)** (Last decade)
- **Hybrid modeling**: Simulation - Data Driven Modeling - Heuristic Optimization (Lately)
Integration methodologies

- Simulation - Optimization (NLP) (Late 1980's)
- Simulation - Optimization (Heuristic) (Last decade)

Optimization [NLP (GRG), Heuristic (GA)]

If the simulation stage is time consuming then the optimization process can become endless (e.g., a one 5 minute simulation duration for 100 GA generations with 100 strings at each generation, will result a computational time effort of about 35 days)
Integration methodologies

- Hybrid modeling: Simulation - Data Driven modeling - Heuristic Optimization (Lately)

**Optimization** [NLP (GRG), Heuristic (GA)]

- Simplified Simulation + Data Driven Modeling

- Decision variables values

- Simulation results
A Hybrid Model Tree (MT) – Genetic Algorithm (GA) Scheme for Toxic Cyanobacteria Predictions in Lake Kinneret

By Avi Ostfeld¹, Ariel Tubaltzev¹, Meir Rom², and Lea Kronaveter²

¹ Civil and Environmental Engineering, Technion – I.I.T, Israel
² Lake Kinneret Watershed Unit, Mekorot, Israel National Water Company, Israel
Introduction

- This work presents a hybrid model tree (MT) - genetic algorithm (GA) scheme for toxic Cyanobacteria (Blue Algae) predictions in Lake Kinneret (Sea of Galilee), which is the most important surface water resource in Israel.
- For more than 30 years, up to 1994, there have been no major problems with respect to the water quality of the lake.
- In 1994 toxic Cyanobacteria (blue algae) blooms appeared, which suggests that the future water quality of the lake might be at risk.
Introduction

- A full physical understanding of the reasons for the toxic Cyanobacteria blooms is lacking.
- This study suggests a data driven modeling approach, relying on the vast existing data base of the lake, to explore the possible major factors causing the toxic Cyanobacteria to bloom, and to predict their possible appearance.
- The suggested model is a hybrid model-trees (MT) genetic algorithm (GA) scheme.
- A few words about Lake Kinneret, model-trees (MT), and genetic algorithms (GA's) ---
Lake Kinneret
32:50N, 35:35E; -209 m above sea level.

A warm, monomictic lake, located at the northern end of the Afro-Syrian Rift Valley in Northern Israel.

**PHYSICAL DIMENSIONS**

- **Surface area**: 170 km²
- **Volume**: 4 km³ (4000 MCM)
- **Maximum depth**: 43 m
- **Mean depth**: 25.6 m
- **Length of shoreline**: 53 km
- **Residence time**: 4.8 yr
- **Catchment area**: 2,730 km²
Lake Kinneret

The largest natural fresh surface water in Israel providing approximately 35% of its annual drinking water, a proportion that is constantly increasing.

[KLL report (2001)]
Monitored Data

LDS - Wind recorder, Short & long-wave radiation, air temp, thermister chain & radio antenna installed at sta. A sends real-time data to KLL.
Methodology

Input (Database)

1. Physical
   - Physical - lake water level, wind velocity, water temperature, etc.

2. Chemical
   - Chemical - average lake values for Nitrogen, Phosphorus, Alkalinity, PH, etc.

3. Biological
   - Biological - species of Zooplankton such as Coppoda, Cladocera, Rotifera, etc.

4. External loading
   - External loading - external pollutant loads entering the lake from the north through the Jordan River (e.g., Nitrate, Nitrogen, Phosphorus, etc.)

- Each of the parameters is measured on a fortnight basis.
- Different imputation methodologies were used to expand the database to a weekly basis.
- The database spans a duration of 25 years.
- What classifiers should be chosen so as to maximize the model prediction ability?
- The model trees are predicting; the GA searches the best classifiers.
Methodology

An iterative scheme for searching the optimal set of classifiers and their corresponding lag times which maximizes the model prediction ability for a given database component

Population of classifiers and lag times

\[(P_1, DT_1), ..., (P_n, DT_n)\]

\*(P_1, DT_1), ..., (P_n, DT_n)\*

Model tree

Fitness

Stop

Genetic algorithm
Methodology

Stage A. INITIALIZATION
* Choose randomly a predefined number of parameters from the database (e.g., one parameter from each database group OR any four parameters).
* Choose randomly a time lag for each of the chosen parameters (e.g., three weeks).
* Repeat until a set of parameters and corresponding time lags are selected (i.e., setting the first GA population).

Stage B. MODEL TREE (MT) CONSTRUCTION
* Build a model tree for each of the selected parameters and time lags (i.e., the GA strings) using Cubist (Quinlan, 1993) for predicting a chosen parameter (e.g., Toxic Nitrogen Fixing Cyanobacteria Algae).
* Assign each string a fitness equal to its correlation coefficient calculated by Cubist using a predefined cross validation dataset.

Stage C. GENETIC ALGORITHM (GA)
* Perform Selection, Crossover, and Mutation on the current population using optiGA (Salomons, 2002).
* Construct a new population of strings (i.e., parameters and time lags).
* Check if STOPPING conditions are met. If stopping conditions are met define the corresponding highest correlation coefficient string as the OPTIMAL solution (i.e., parameters and time lags), otherwise go back to Stage B.
Methodology

The methodology is cast in a program entitled KAPM (Kinneret Algae Prediction Model), it’s main interface menu is:
Application

Predicting (CTXF 1m) Toxic Nitrogen Fixing Cyanobacteria Algae blooms at station A (deepest point in the lake) at a depth of 1 m (validation results):

Training period: 1994 - 2002; Cross validation period: 2003 - 2004; Correlation coefficient: 0.92
Application
Optimal model tree rules:

**Rule 1:**
\[
\text{IF} \quad \begin{align*}
\text{W Temp } & \leq 26.55 \\
\text{CTXF 5m } & \leq 71.23 \\
\end{align*}
\text{THEN} \quad \text{CTXF 1m } = 3.91
\]

**Rule 13:**
\[
\text{IF} \quad \begin{align*}
\text{W Temp } & > 26.81 \\
\text{CTXF 5m } & > 725.00 \\
\text{CTXF 5m } & \leq 1328.42 \\
\end{align*}
\text{THEN} \quad \text{CTXF 1m } = -208.3 + 38 (\text{W Temp}) - 855 (\text{NTOT 5-10}) + 0.38 (\text{CTXF 5m})
\]

**Rule 14:**
\[
\text{IF} \quad \begin{align*}
\text{NTOT (Huri) } & \leq 1.90 \\
\text{NTOT 5-10 } & \leq 0.37 \\
\text{CTXF 5m } & > 725.00 \\
\end{align*}
\text{THEN} \quad \text{CTXF 1m } = 792.54 + 0.38 (\text{CTXF 5m})
\]

**Rule 15:**
\[
\text{IF} \quad \begin{align*}
\text{CTXF 5m } & > 1328.42 \\
\end{align*}
\text{THEN} \quad \text{Output } = 4463.67 - 1.06 (\text{CTXF 5m})
\]
Conclusions

- The methodology and application of a coupled model trees (MT) - genetic algorithm (GA) scheme for predicting blue-algae blooms in Lake Kinneret was presented and demonstrated.

- The model was able to satisfactorily predict blue-algae blooms using a database which incorporates physical, chemical, biological, and external loading data for the lake.

- One of the possible extension options of the model is to explore the tradeoff between the model prediction ability versus the optimal correlation coefficient (i.e., fitness) using multi-objective optimization, assuming that these two objectives compete.
Thanks!
Questions?