Optimization and Modeling at UW: Fish, Cows, Sanctions and Energy

Michael C. Ferris

University of Wisconsin, Madison

Optimization Research Consortium
Wisconsin Institutes for Discovery, University of Wisconsin
November 20, 2015

How to enhance the impact of optimization in applications?

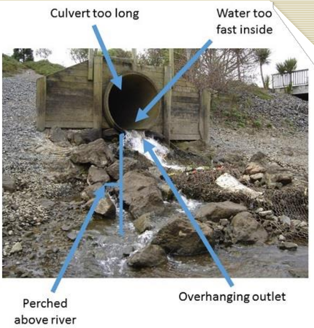
- Hire (and/or engage) people with breadth of, and complementary expertise - theory, algorithms, computation, applications

How to enhance the impact of optimization in applications?

- Hire (and/or engage) people with breadth of, and complementary expertise - theory, algorithms, computation, applications
- Key impact area: decision making in (environmentally) resource constrained problems
- Feature: shared resource that interacts with complex multi-user systems
- Enhance understanding of decision space, facilitate policy design and operational improvement
- Build appropriate models, fast enough solution for expert interaction, visualize results

Overview

- Anadromous fish migrate from the sea upstream into freshwater to spawn.
- Natural & man-made barriers break stream connectivity and prevent fish from penetrating deep into inland lakes and rivers



- There are over 235,000 identified barriers to migration in the Great Lakes Basin
  - Lake Michigan: >83% of tributaries inaccessible
  - Lake Huron: >86% of tributaries inaccessible
  - Lake Erie: >50% reduction of population size

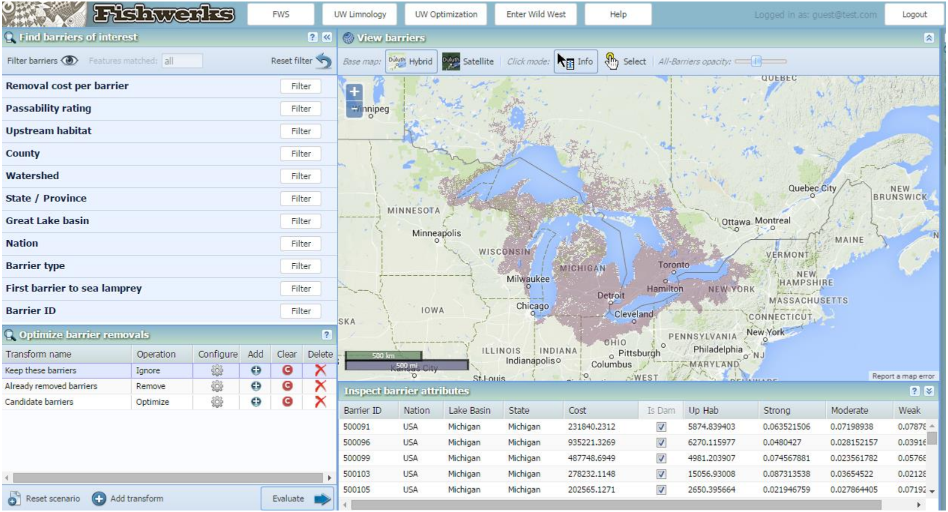
Cont'd

- Barriers can be mitigated to allow for fish passage:
  - Removal of dams, improved road crossings, fish passageways
- However, they are very expensive – Average costs for fixes:
  - Dams: $100,000 - $650,000 each
  - Others: $30,000 - $150,000 per project
- Limited funds necessitate ideal selection of projects
  - Difficult to assess where funds should be used
  - Country/State/County lines make appropriation difficult
- Increasing passability increases risk for the spread of invasive aquatic species (e.g. Sea Lamprey)

The Goal (Customer #1)

1. Provide an interactive tool to consolidate big-data sets across multiple departments (DNR, FWS, NFPP, etc) and visually display in a meaningful way.
2. *Utilize optimization to maximize efficiency in policy decisions and funds appropriations.
3. *Allow any user to dynamically solve a large range of models and scenarios without requiring background knowledge of optimization.
4. Provide means for certified users to update/validate data sets.

Data Visualization: http://www.greatlakesconnectivity.org/



The Data

For every Barrier [J]: 236,264

- Barrier ID – A unique string identifier
- Geographical Info – Nation, State, County, Lake Basin, Watershed
- Barrier Type – Dam or Road Passage
- Cost – Estimated cost to mitigate the barrier
- Root – If the barrier is the first in the stream (no downstream barriers)
- Downstream ID – Identifier of the downstream barrier

For every Fish Guild [S]: 36

- Invasive – If it is an invasive species or not

For every [$J \times S$]: 8,505,504

- Passability Rating – % Chance species can pass this barrier
- Upstream Habitat – Amount of usable habitat upstream of barrier

The Model

Objective:

$$\max \sum_{j \in J} \sum_{s \in S \setminus \text{Inv}} v_{js} * z_{js}$$

Subject To:

$$\sum_{j \in J} x_j * c_j \leq B$$

$$z_{js} = \left(\bar{p}_{js} + \pi_{js} \cdot x_j\right) * z_{ds}, \qquad \forall j \in J, d \in D(j), s \in S$$

$$x_j \in \{0,1\}$$

Where:

- v_{js} := Upstream Habitat, $\bar{p}_{js}$:= Passability Rating, π_{js} := Probability Increase (if mitigated)
- c_j := Cost of mitigation, B := Total Available Budget
- z_{js} := Cumulative passability rating, D(j) := Set of nodes downstream of j. Note: $|D(j)| \leq 1$.
- x_j := Decision to Remove barrier 'j'

Smart Modelling - Linearization

$$z_{js} = (\bar{p}_{js} + \pi_{js} \cdot x_j) * z_{ds}, \qquad \forall j \in J, \forall d \in D(j)$$

Use set of roots ($R \subset J$):

$$z_{rs} = \bar{p}_{rs} + \pi_{rs} \cdot x_r, \qquad \forall r \in R, s \in S$$

Introduce new variable $y_{js} = x_{js} * z_{ds}$:

$$z_{js} = \bar{p}_{js} \cdot z_{ds} + \pi_{js} \cdot y_{js}, \qquad \forall j \in J \backslash R, s \in S$$

Add additional constraints:

$$y_{js} \leq x_j, \qquad \forall j \in J \backslash R, s \in S$$

$$y_{js} \leq z_{ds}, \qquad \forall j \in J \backslash R, s \in S$$

Basic {0,1} LP Model:

$$\max \sum_{j \in J} \sum_{s \in S \setminus \mathrm{Inv}} v_{js} * z_{js}$$

Subject To:

$$\sum_{j \in J} x_j * c_j \leq B$$

$$z_{rs} = \bar{p}_{rs} + \pi_{rs} \cdot x_r, \qquad \forall r \in R, s \in S$$

$$z_{js} = \bar{p}_{js} \cdot z_{ds} + \pi_{js} \cdot y_{js}, \qquad \forall j \in J \setminus \mathrm{R}, s \in S$$

$$y_{js} \leq x_j, \qquad \forall j \in J \setminus \mathrm{R}, s \in S$$

$$y_{js} \leq z_{ds}, \qquad \forall j \in J \setminus \mathrm{R}, s \in S$$

$$x_j \in \{0,1\} \qquad \forall j \in J$$

Interactive Modelling

Allow user to:

- Select their range of influence (i.e. State, County, etc)
- Select mitigatable barriers using a broad range of criteria
- Manipulate Constraints
- Visualize Results

Let's check it out! ($B = 10^7$)

- Minnesota : 3,458 – 6s.
- Wisconsin: 19,854 – Timed Out!?

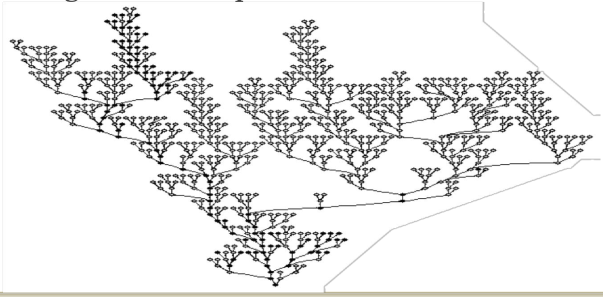
Problem!

{0,1} Linear Programming is $\mathcal{NP} - Complete$!

- Solution time quickly becomes unpractical as problem size grows!
- Web tool requires fast processing to inform user.

Need to find methods to speed up solution time!

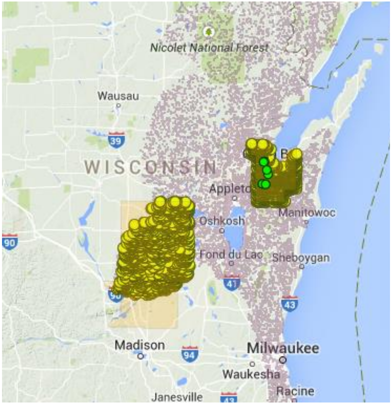
Could we take advantage of the unique structure of our data?



Pre-Processing

Disjoint Counties: Data Compression

- May desire collaboration between counties
- Downstream barriers effected by upstream decisions
  - Barriers in-between are irrelevant
  - Can be removed by smartly incorporating their data into other nodes!



Representative Species

- 36 total fish guilds – Many have very similar parameter data!
- Use QAP to separate guilds into ‘representative groups’
  - Smaller overall data set – improves speed of (relaxed) master solution

$$\min\left\{\sum_{g\in G}\sum_{i\in S}\sum_{j\in S}(d_{ij} * z_{ijg})\right\}$$

Subject to:

$$\sum_{g\in G} x_{sg} = 1, \qquad \forall s \in S$$

$$z_{ijg} \leq x_{ig}, \qquad \forall i,j \in S, g \in G$$

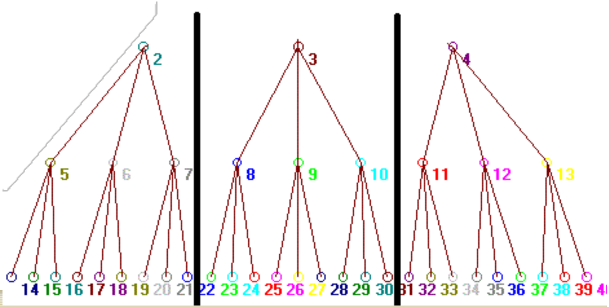
$$z_{ijg} + 1 \geq x_{ig} + x_{jg}, \qquad \forall i,j \in S, g \in G$$

Independent Streams?

Each root node corresponds to a completely independent tree!

Can solve separate, smaller MIP on each tree.

- However, budget constraint is global!
- *How do determine budget in each tree?*



The Goal (Customer #2)

Quickly and accurately create return-on-investment (ROI) curves for a wide-breadth of project scenarios.

- Each curve requires > 20 data points to cover all range of possible budgets!

Supplement base model with additional constraints:

- Ensure that available habitat for ALL species increases by specific amount
  - While still maximizing total habitat
- Prevent invasive species from gaining too much habitat.

Test Data Set: Lake Huron Basin

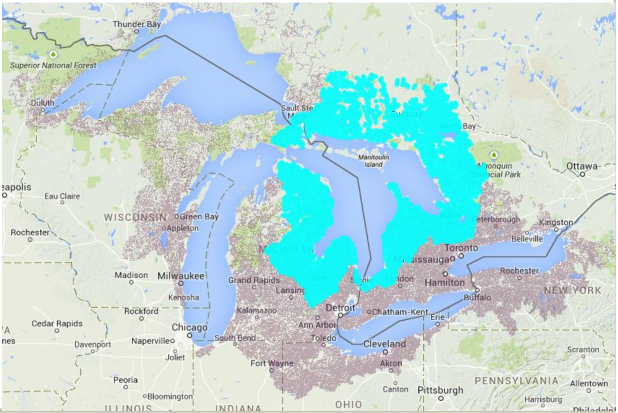
51,149 Barriers

36 Species

- 2 Invasive Species

Model Size:

- 1,934,421 rows
- 1,274,454 columns
  - 753 discrete-columns
- 4,896,386 non-zeroes



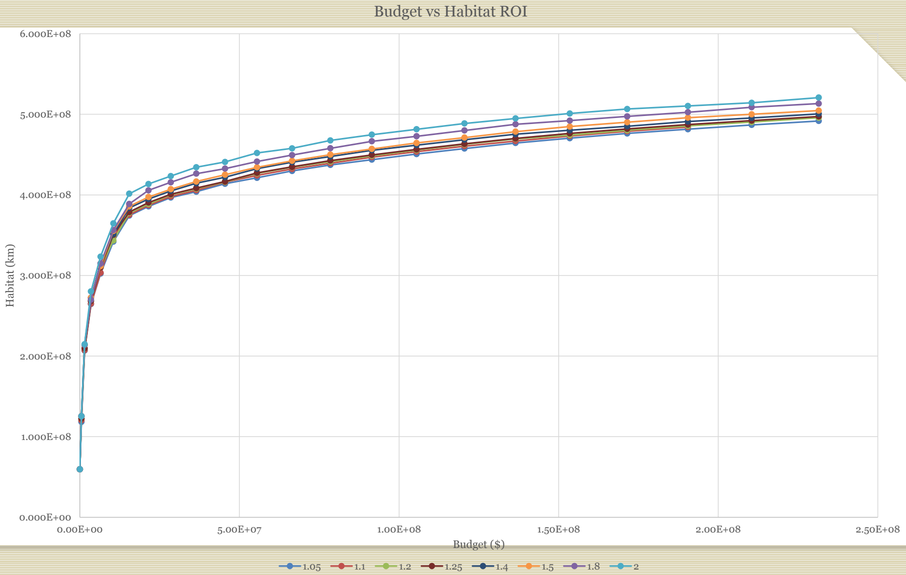
The Problem:

- {0,1} Linear Programming is $\mathcal{NP} - Complete$!
- Our Data Set is extremely large.
- Solution times grow exponentially with budget [CPLEX, WID Clusters]:
  - $B = 10^6$: 8211 s (Gap = 0%)
  - $B = 10^7$: 2132 s (Gap = 0%)
  - $B = 10^8$: >4 days (Gap = 1%)
  - $B = 5 * 10^8$: >4 days (Gap = 10%)
- Customer desires ROI Curve generation, requiring data points over the entire range of budgets and different scenarios!
- Solution time is unpractical for dynamic web-app modelling!

Performance:

Budget ($)	Sol Time (s)	Gap (%)	Sol Time for Best (s)	% Speedup
10^6	573	0.53	8211	1,333 %
10^7	668	0.88	2132	219 %
10^8	2431	1.31	> 4 days	14,116 %

As we can see, we are able to obtain reasonable solutions for most budgets in less than 10 minutes!

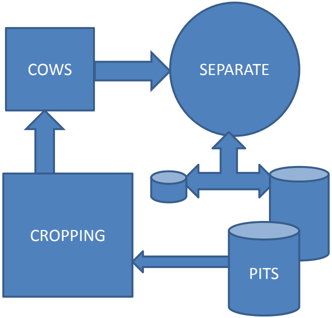


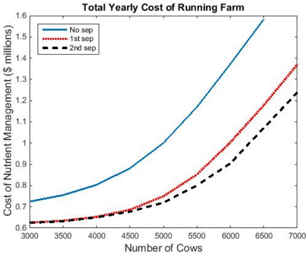
Budget vs Habitat ROI
6.000E+08
5.000E+08
4.000E+08
3.000E+08
2.000E+08
1.000E+08
0.000E+00
Habitat (km)
0.00E+00
5.00E+07
1.00E+08
1.50E+08
2.00E+08
2.50E+08
Budget ($)
1.05
1.1
1.2
1.25
1.4
1.5
1.8
2

Already Impactful!

- Researchers at UW Limnology believe(d) that invasives constraint is vital to amount of attainable habitat.
  - Large amounts of research conducted to identify spread threats
  - Investing research $ into improving mitigation/treatment techniques
    - Pheromones, lampricide, traps, low-head barriers, etc
- ROI Curves show otherwise!
- Either…
  - We've discovered a flaw in current theories on invasive species spreading
  - Or… (More Likely), a flaw in the data set.

Biomass Research and Development Initiative (BRDI)

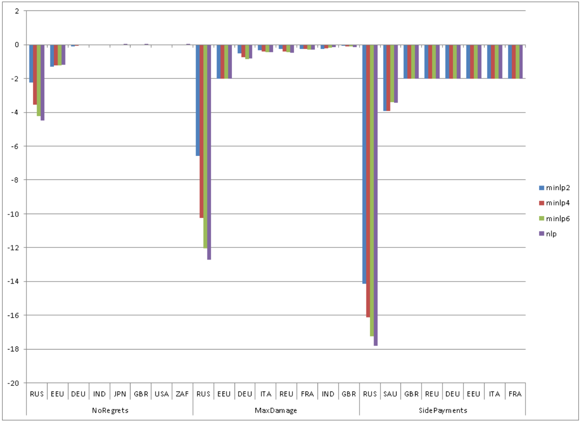




- Whole farm (complex interacting) mathematical model
- Long term sustainable (environment and financial)
- Economic/Logistic Optimization, taking into account phosphorus runoff, other environmental restrictions
- Incorporates data analytics (e.g. SNAP+)
- New insights to operate system efficiently, how to enforce much stricter environmental constraints using blend of rotations, NMP and separations
- Large (mixed integer) optimization

Optimal Sanctions (Boehringer/F./Rutherford)

- GTAP global production/trade database: 113 countries, 57 goods, 5 factors
- Coalition members strategically choose trade taxes to *minimize* Russian welfare
- Russia chooses trade taxes to *maximize* Russian welfare in response
- Nash equilibrium



Resulting equilibrium with no regrets (coalition), maximize damage, side payments

Security-constrained Economic Dispatch (SCED)

$$\min_{u, x_0, \ldots, x_k} \quad c^T u + \rho(u) \qquad \triangleright \text{ Total cost}$$

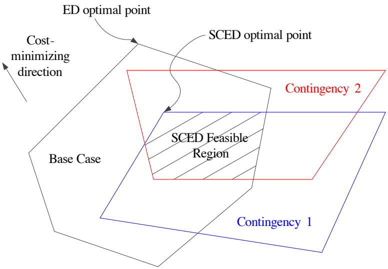
$$\text{s.t.} \quad 0 \le u \le \bar{u} \qquad \triangleright \text{ GEN capacity const.}$$

$$g_0(x_0, u) = 0 \qquad \triangleright \text{Base-case network eqn.}$$

$$-\bar{x} \le x_0 \le \bar{x} \qquad \triangleright \text{Base-case flow limit}$$

$$g_k(x_k, u) = 0, \quad k = 1, \ldots, K \qquad \triangleright \text{Ctgcy network eqn.}$$

$$-\bar{x} \le x_k \le \bar{x}, \quad k = 1, \ldots, K \qquad \triangleright \text{Ctgcy flow limit}$$



- Base-case topology g_0 and line flow x_0
- If k-th line fails, line flow jumps to x_k : $g_k = 0$
- Ensure x_k in bounds $\forall k$

Model structure

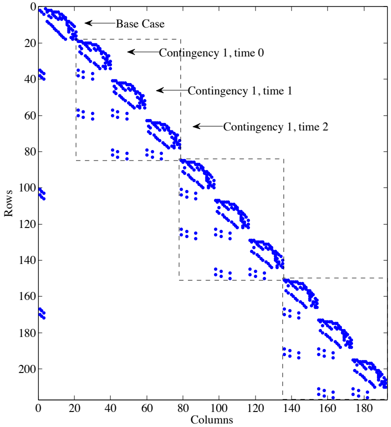


Figure : Sparsity structure of the Jacobian matrix of a 6-bus case, considering 3 contingencies and 3 post-contingency checkpoints.

- Corrective actions are not modeled in ISO's dispatch software (deemed unsolvable!)
- We **model** the *multi-period* corrective rescheduling in SCED; solutions much better quality
- **Enhance** the Benders' **algorithm** to solve the problem faster
- **Achieve** about 50× **speedup** compared to traditional approaches

Conclusions

- Optimization guides the development of complex interaction processes within application domains
- Combination of models provides effective decision tool at multiple scales
- Policy implications addressable using MOPEC
- Problems solved by combination of domain expertise, modeling prowess, good theory/algorithms and efficient implementations all facets needed
- Many new settings available for deployment; need for more theoretic and algorithmic enhancements