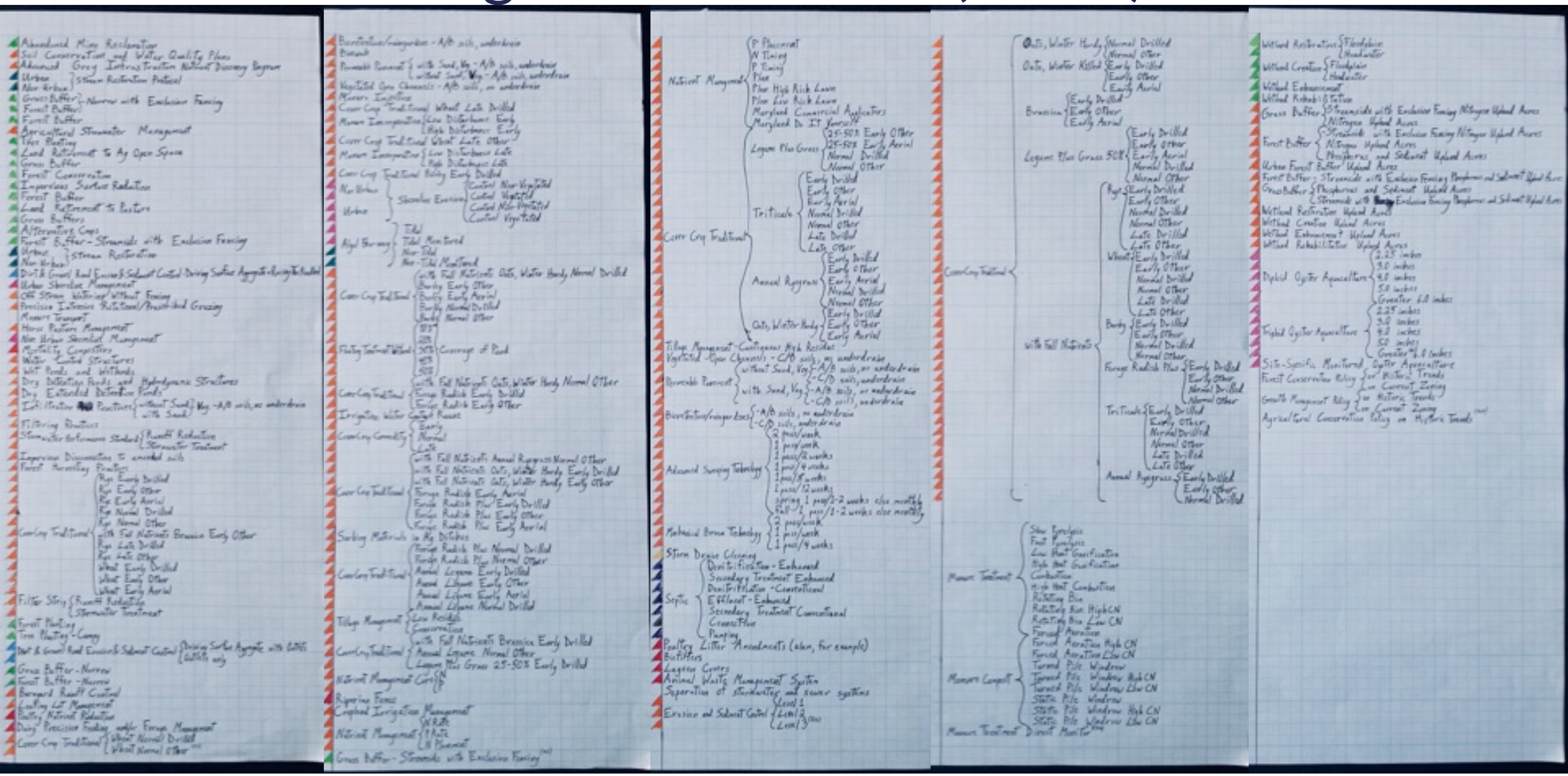
Scenario Optimization Tool for CAST
(the time-averaged Phase 6 watershed model)

11 September 2018 – Modeling Workgroup Meeting

Danny Kaufman

Project Goal: Investigate, develop, test, and implement an optimization system for the Chesapeake Assessment Scenario Tool (CAST) that will facilitate identification of more cost-effective and otherwise optimal approaches to pollutant load reduction for CBP partners.

Best Management Practices (BMPs) in CAST



Analyze potential BMP options and identify low-cost strategies

To help the Chesapeake Bay Program and its Partners restore the Bay and its watershed

Why this update? (outline)

Takeaways from recent Advisory and Support Committee (ASC) meeting

Current working prototype and progress

Next steps

Main Takeaways from optimization ASC meeting

(Monday, 20 Aug 2018)

Working prototype, using subset of BMPs, is well formulated:

- convexity and starting point analyses were useful
- and there are not any fatal flaws

Key elements to build on:

- larger geographic scales
- compare prototype results to 'optimal solutions' obtained by both CAST experts and other users
- include additional BMPs
- other operational issues

Post-ASC meeting discussions

Focus the CBPO “operations research department” on merging ASC ideas with the need to develop a working prototype in 2019

Utilize working prototype to gather feedback and collaboratively build in features desired by users

- Actively engage users that have not been previously engaged
- Work with folks (could be you!) to construct and explore case studies

Concurrent long-term strategizing:

- Algebraically formulating additional BMPs
- Testing heuristic and/or model-based optimization strategies

Working Prototype

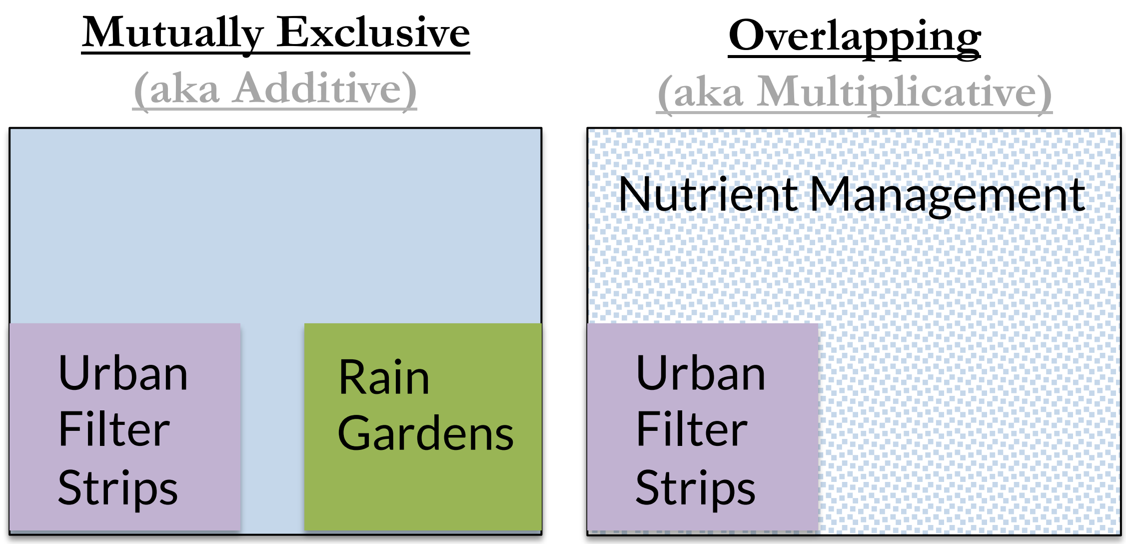
Problem description

Optimization model

- two forms
- example results – land river segment
- example results – county

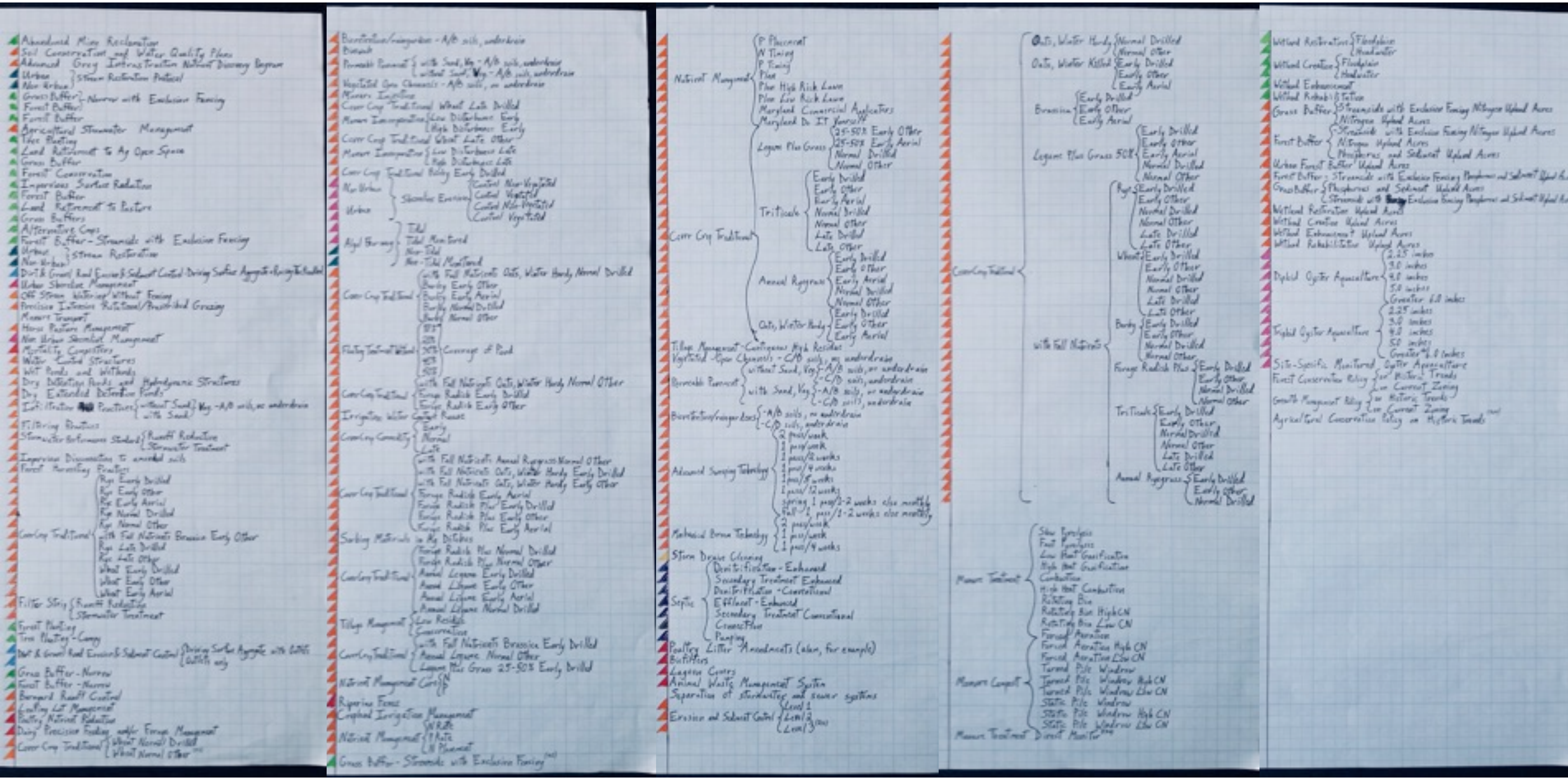
Challenges and long-term tests

Efficiency BMPs



What's the optimal (lowest-cost) implementation of these BMPs for a given geographical area?

BMPs and their simulation rules



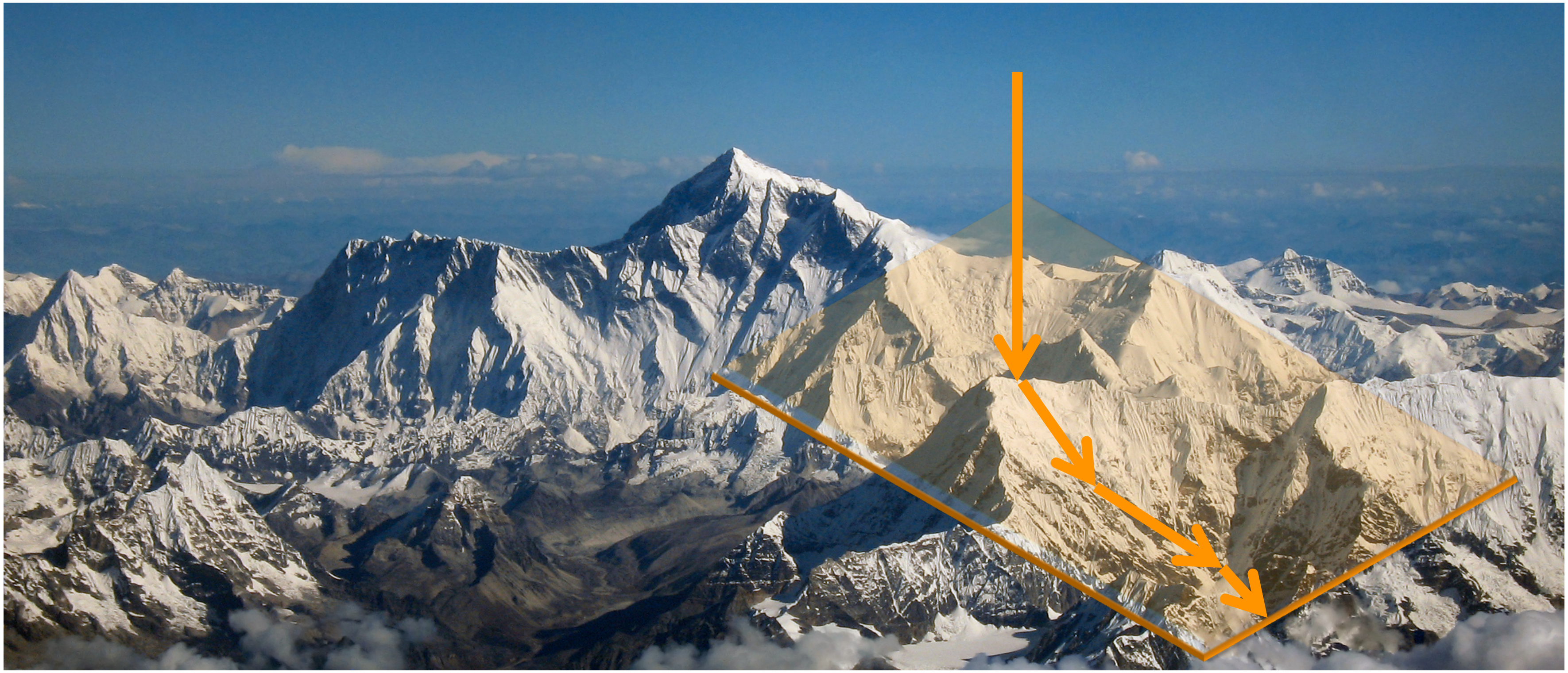
Orange - Efficiency
Purple – Septic efficiency
Yellow – Load reduction
Light blue – dirt and gravel
Green – Land use change
Brown – Septic connections
Pink – shore
Teal – stream
Red - animal

Optimization as search

How would you go about finding the lowest point? Without GPS :(

Optimization as search

Move in the direction of the steepest slope, towards a minimum



Efficiency BMP prototype

total cost

$$\sum_{\substack{Segments\\BMPs\\LoadSources}} (cost * BMPacres)$$

load reduction

$$\% \, LoadReduction_{\{segment, pollutant\}}$$

available acreage

$$\sum_{\substack{BMPs\\in\\Groups}} BMPacres \leq AvailableAcres$$

The same calculations as in CAST

Efficiency BMP prototype

Two Versions

Minimize
(total cost)

$$\sum_{\substack{Segments \\ BMPs \\ LoadSources}} (cost * BMPacres)$$

Constrained by:
(Target load)

Efficiency BMP prototype

Two Versions

Minimize
(total cost)

$$\sum_{\substack{Segments \\ BMPs \\ LoadSources}} (cost * BMPacres)$$

Constrained by:
(Target load)

Maximize
(load reduction)

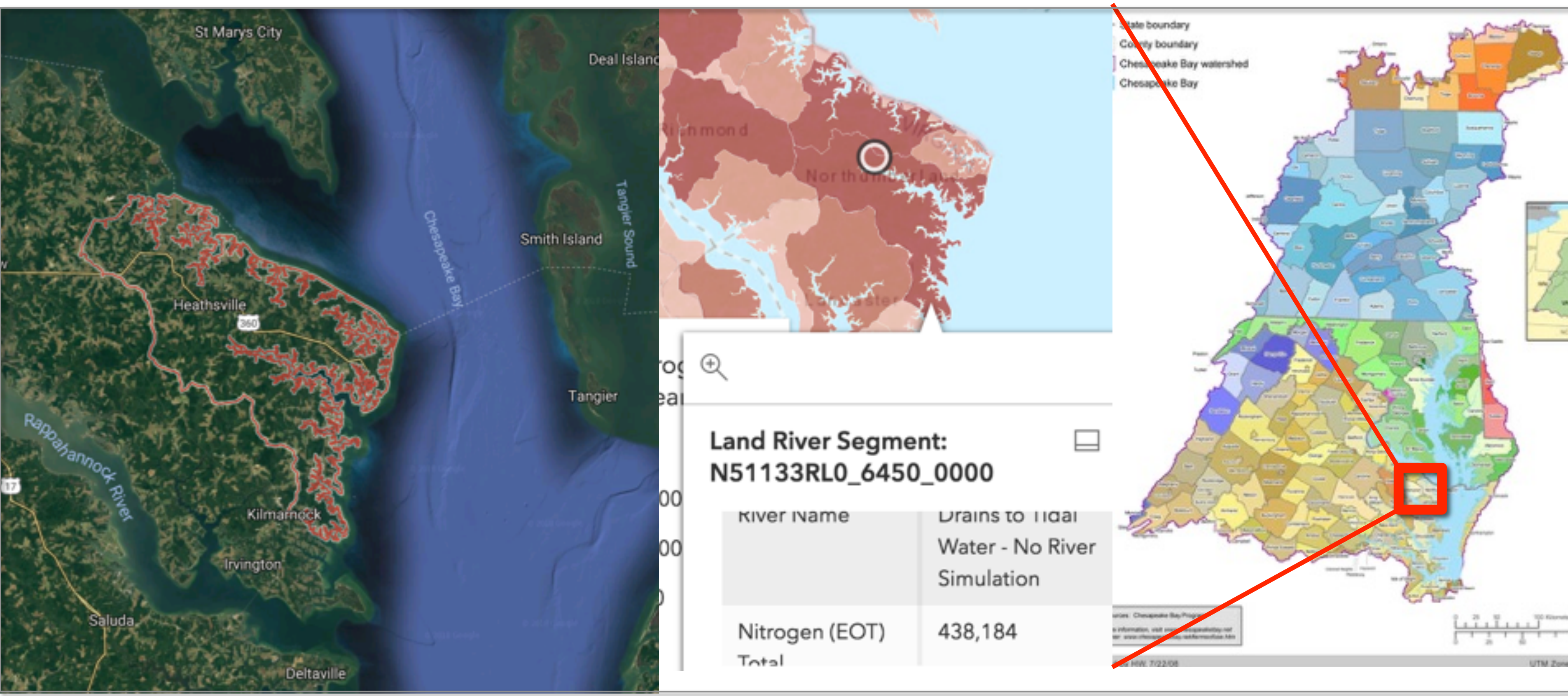
$$\% LoadReduction_{\{segment, pollutant\}}$$

Constrained by:
(Cost bound)

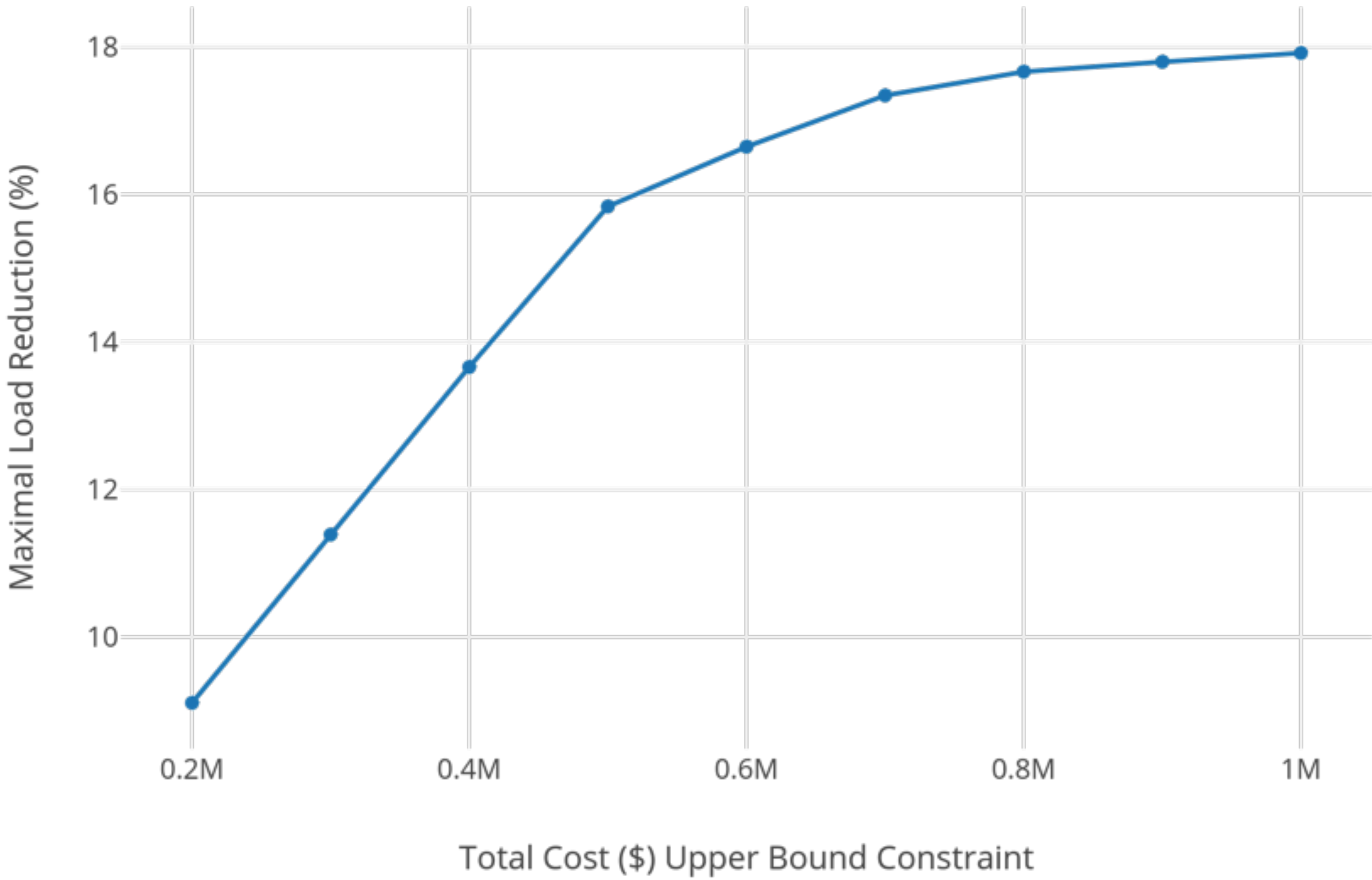
Results

Northumberland county, VA - Land River Segment N51133RL0_6450_0000

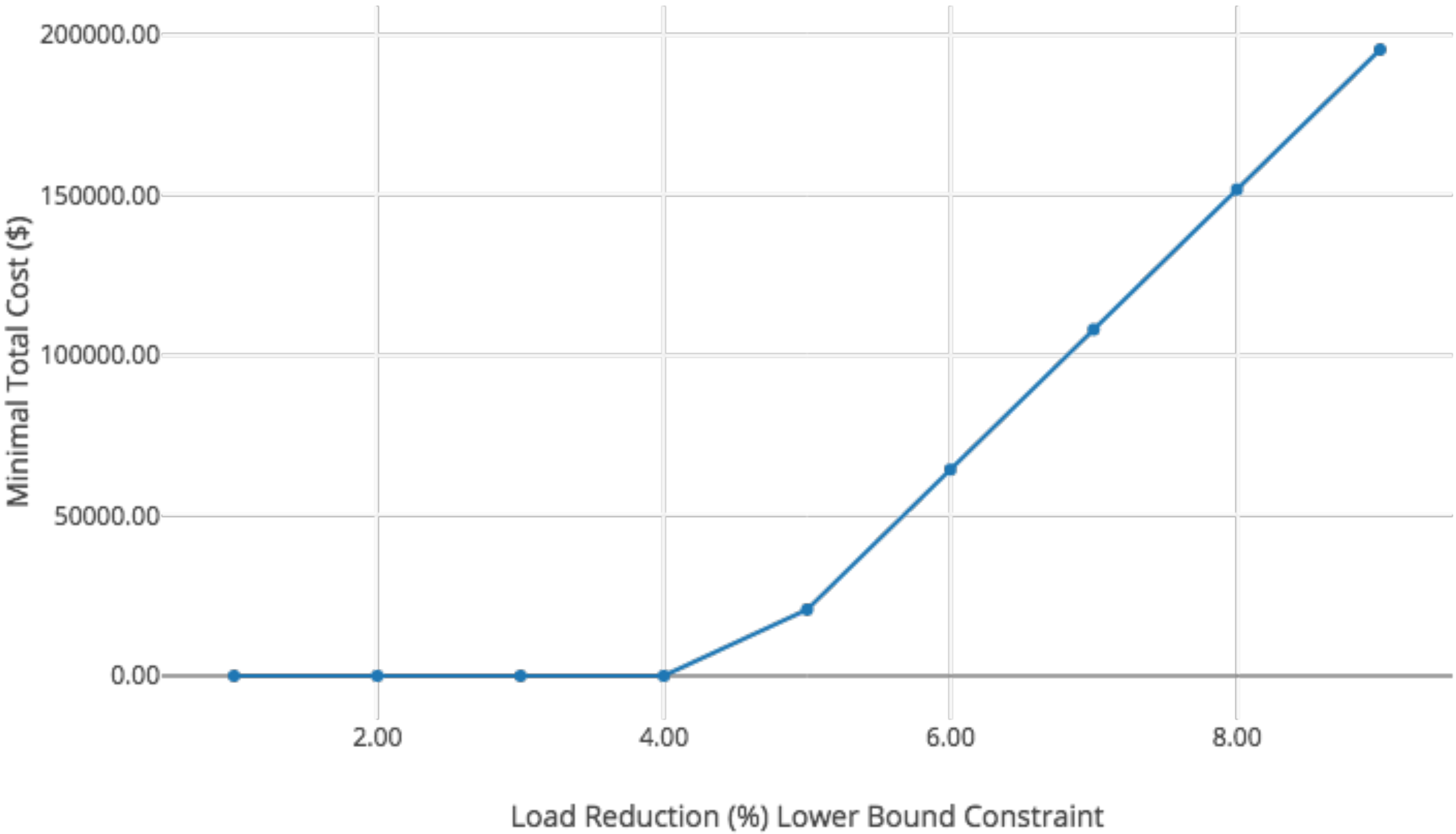
total acres = 58,040.90



Objective: **Maximize** **Load Reduction** **(%)**



Objective: **Minimize** **Total Cost** **($)**



County-level

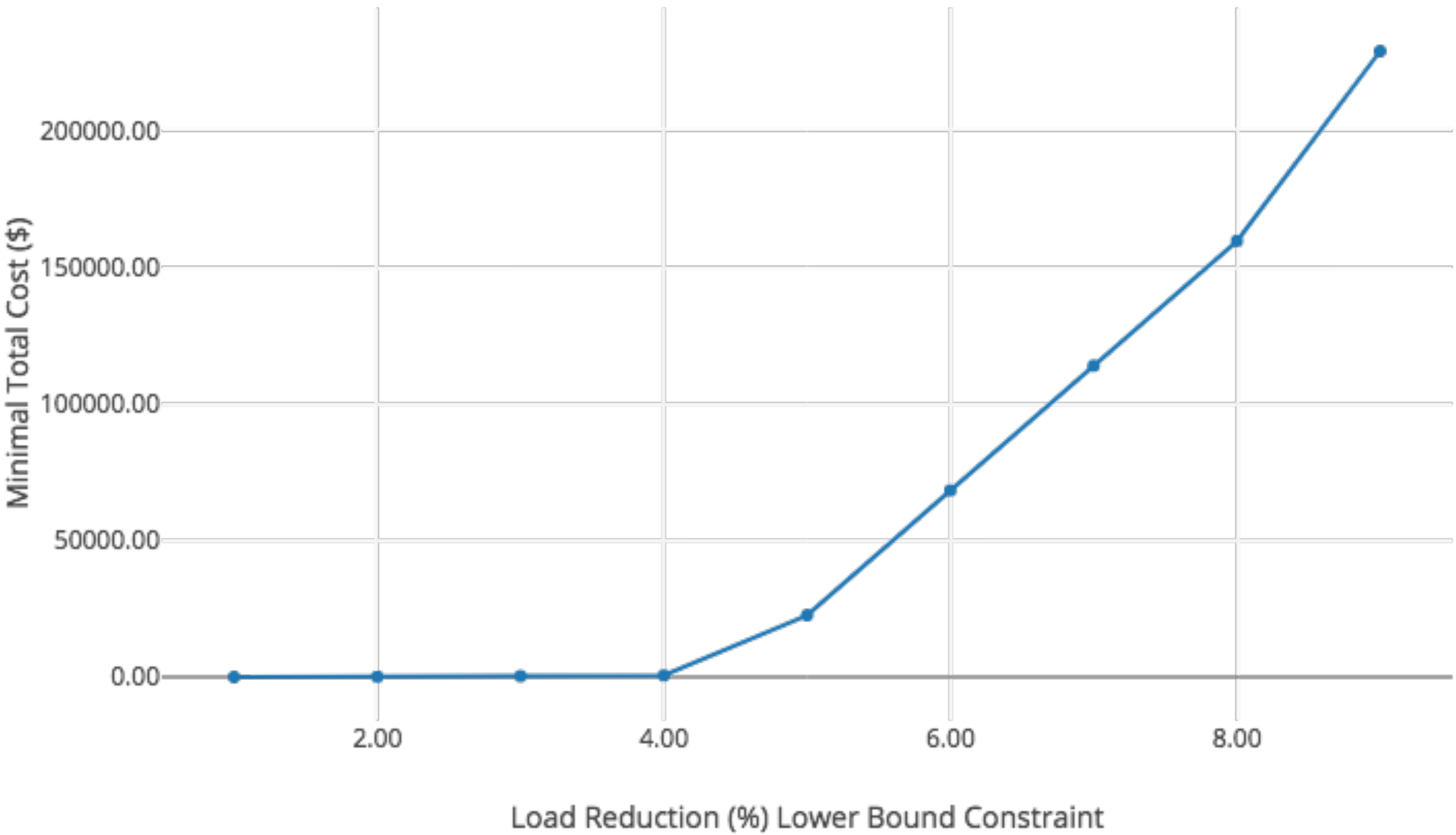
- Total cost summed over all land river segments

- Separate target load reduction constraint for each land river segment

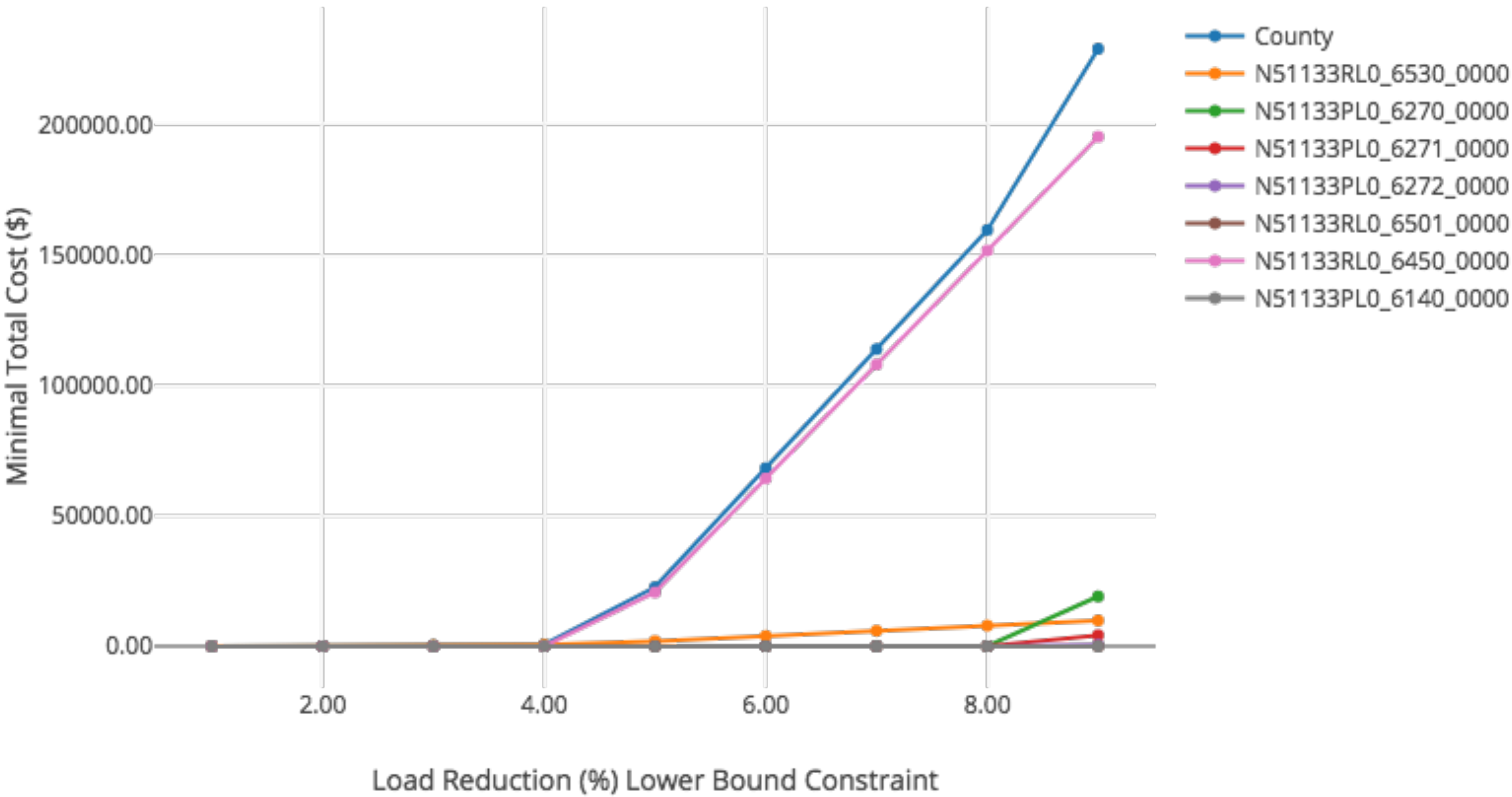
Northumberland county, VA

7 land river segments

Objective: **Minimize** **Total Cost** **($)**



Objective: **Minimize Total Cost ($)**



Starting point analysis

Why?

There may be starting point dependence, because load reduction function is non-convex.

(a local optimum may not be a global optimum)

Starting point analysis

Completed 10 runs for each formulation, for which the variable values were all drawn from a uniform distribution spanning 0 - 6000 (acres). *Note: this means that the initial points were not feasible.*

- Same solution was found each time.
- Numbers of iterations taken to find the solution, and the trajectories, were different for the trials.
- Upper bound (*6000 acres*) used for randomizing start points was never a binding constraint.

Solution was robust to changing the solver starting point

Next steps

Efficiency BMP optimization model:

- Using oxygen damage units to consolidate N & P
- Ensuring robust solutions for more geographic regions
- Accounting for existing constraints, structural BMPs
- Testing county-level targets / unequal target loads between land river segments within a county

Feedback!

Concurrent discovery for incorporating other BMPs

What are others thinking?
Feedback from Presentations

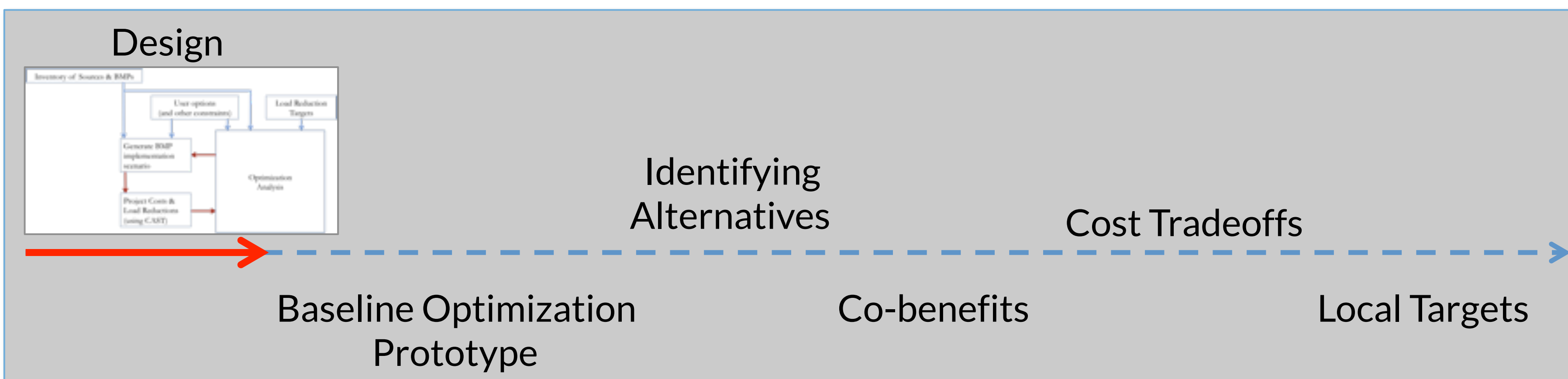
- Cost profiles aren't representative (zeros are strange)
- This is going to be a very useful tool
- Going to make a big difference in conjunction with co-benefit quantification (current GIT proposed funding)

Will be shaped by feedback

Actively searching for ways to engage local decision makers at county and municipal scales for their guidance and feedback on prototype design.

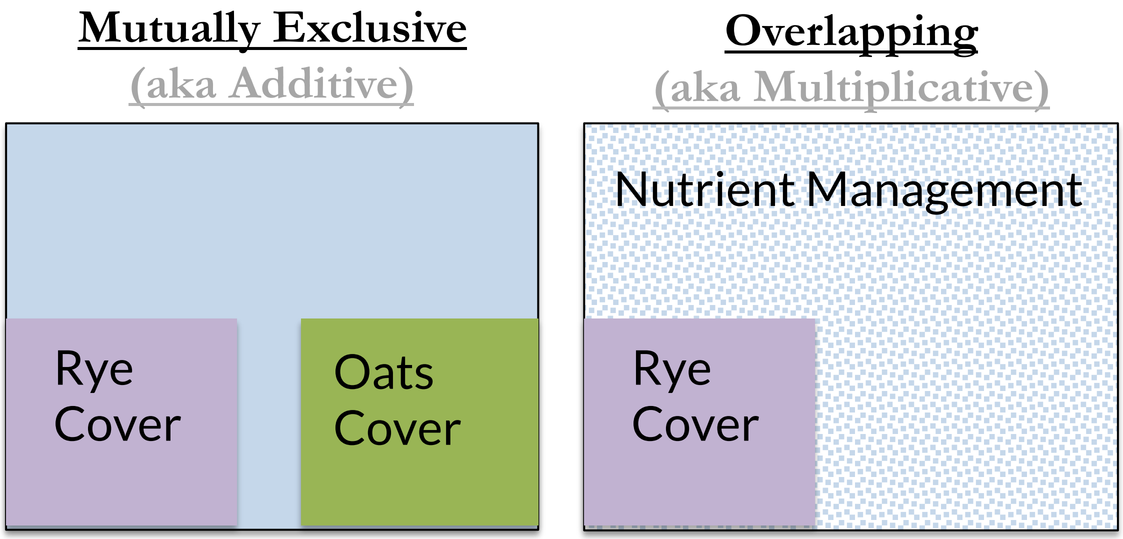
Your area (county, sub-watershed) can be an early case study!

Email me (Danny) at: dkaufman@chesapeakebay.net



EXTRA SLIDES FOLLOW

Efficiency BMPs



What's the optimal (lowest-cost) implementation of these BMPs for a given geographical area?

Prototyping Experiments

Search Space Investigations

Fine-tuned sub-problem

Next steps

Conceptual

- Investigate solutions in more depth
  - Look at z_L, dual outputs at the solution
  - Look at Pynumero gradient results
- Sensitivity analyses
  - Test other randomized start pt. ranges (0-100,000; 0-100; etc.)

Practical

- Solving NLP for multiple land-river segments simultaneously
- Solve cost min. objective formulation with only N constraint to parallel load objective
- Use oxygen damage units to consolidate N & P
- Many more geographic regions
- Feedback!

Heatmap

- for county BMPs (aggregated by land river segment?)
- and then on the next slide (aggregated by load source?)

Efficiency BMP prototype

Two Versions

Minimize
(total cost)

$$\sum_{\substack{Segments \\ BMPs \\ LoadSources}} (cost * BMPacres)$$

Maximize
(load reduction)

$$\% \, LoadReduction_{\{segment,\, pollutant\}}$$

Cost objective

Minimize:
(total cost)

$$\sum_{\substack{Segments \\ BMPs \\ LoadSources}} (cost * BMPacres)$$

Subject to:
(Target load)

$$\%\, LoadReduction_{\{segment \atop pollutant\}} \geq TargetLoadReduction \quad \text{(e.g. 10\%)}$$

(Availability)

$$\sum_{\substack{BMPs \\ in \\ Groups}} BMPacres \leq AvailableAcres$$

Load objective

Maximize:
(load reduction)

$$\%\, LoadReduction_{\left\{\substack{segment\\pollutant}\right\}}$$

Subject to:
(Cost bound)

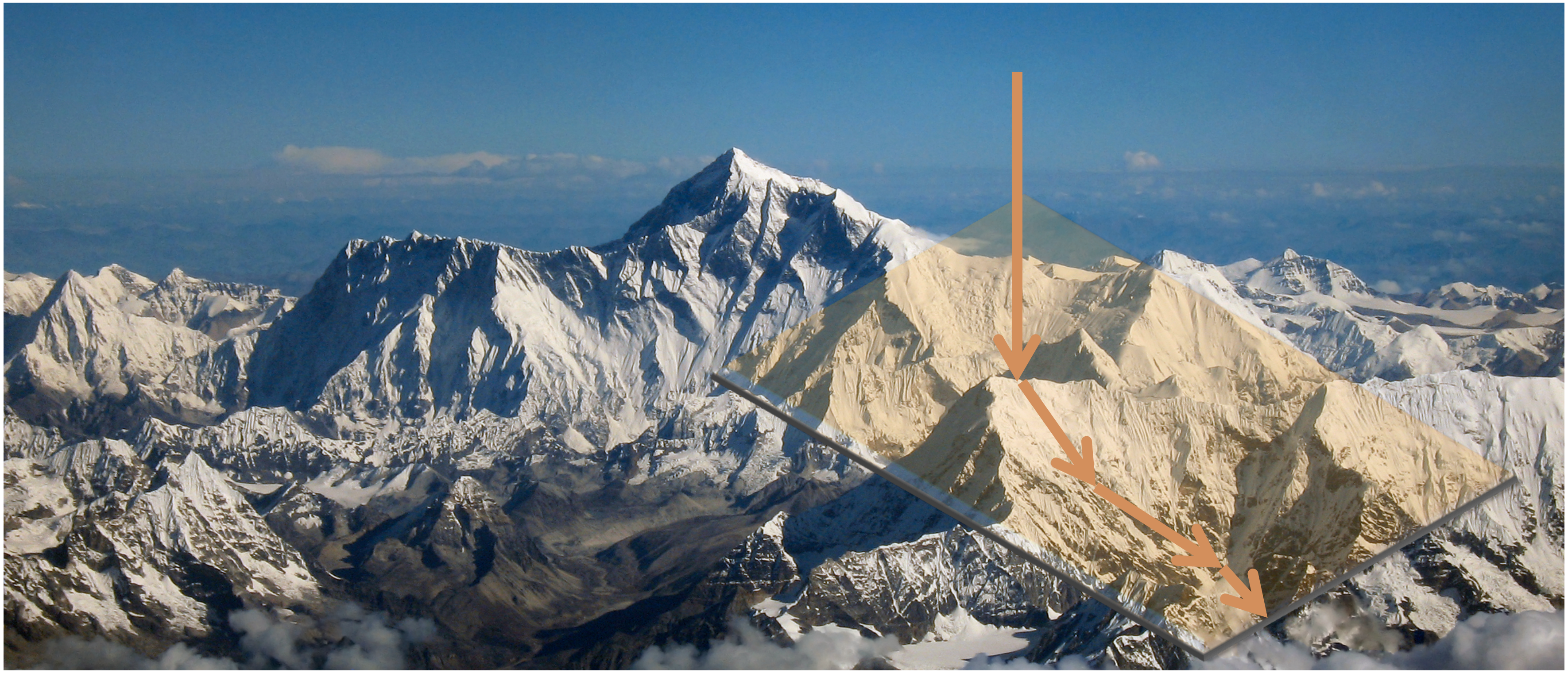
$$\sum_{l\in L}\sum_{b\in B}\sum_{\psi_b^\lambda\in\Psi_b^*}(c_b * X_{b,l,\psi_b^\lambda}) \le C$$

(e.g. $500,000)

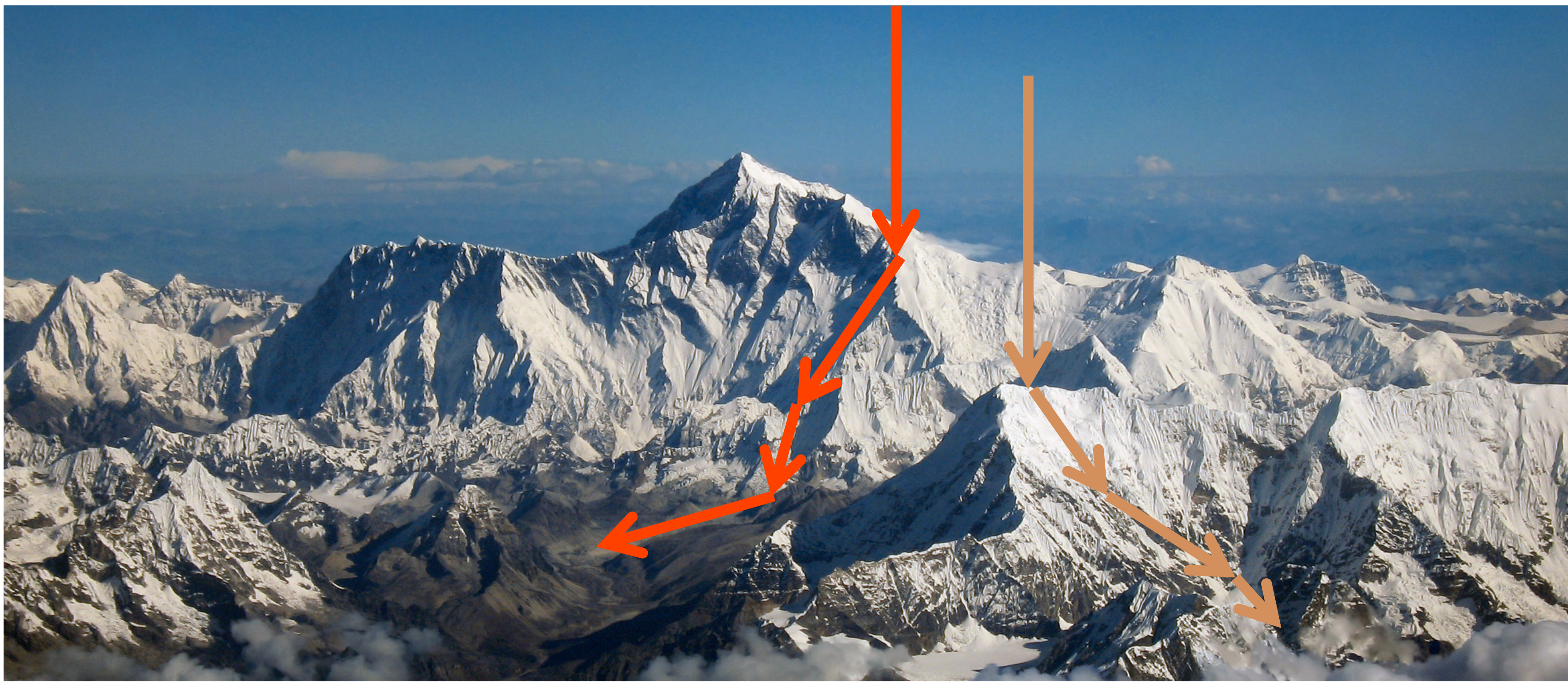
(Availability)

$$\sum_{\substack{BMPs\\in\\Groups}} BMPacres \le AvailableAcres$$

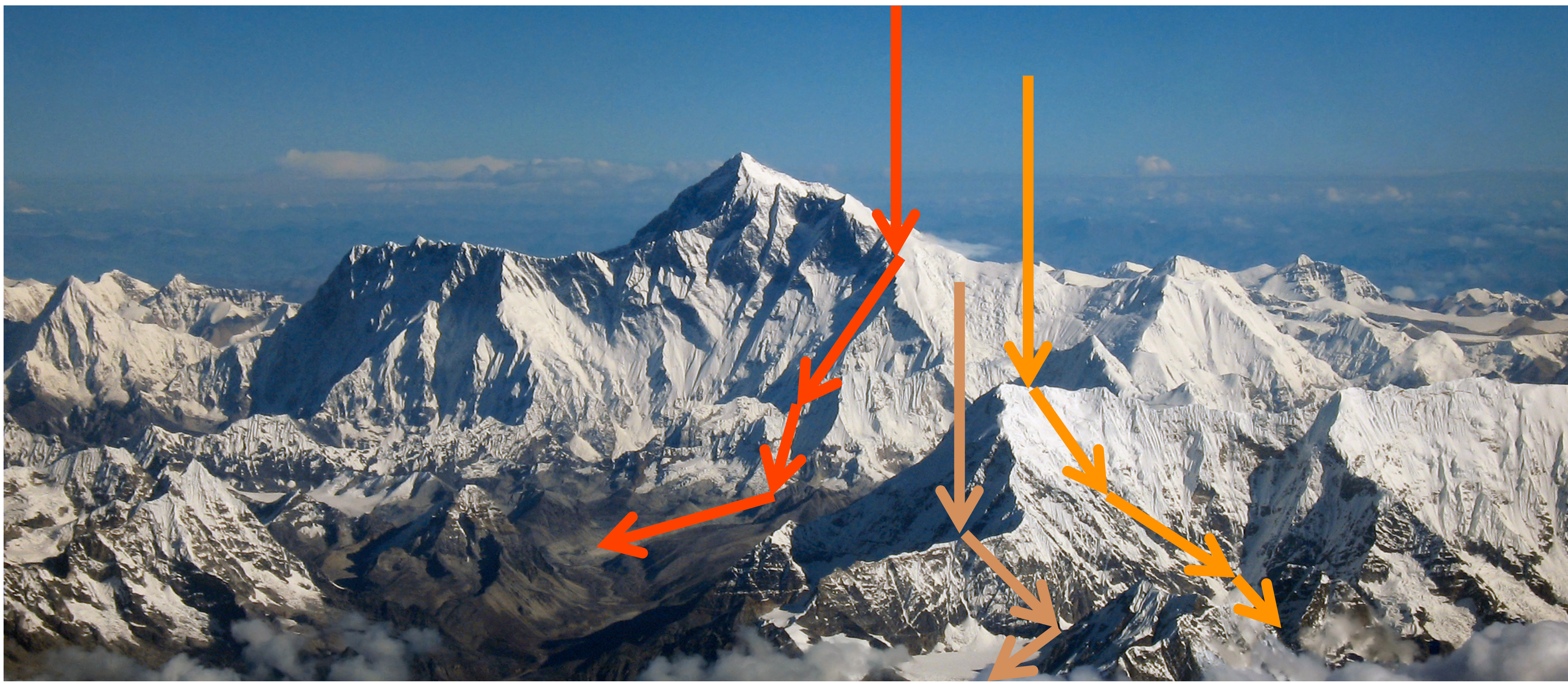
Efficiency BMP prototype



Efficiency BMP prototype



Efficiency BMP prototype



Summary of starting point analysis

- Results for 10 runs for each formulation are shown, for which the variable values were all drawn from a uniform distribution spanning 0 - 6000 (acres). *Note: this means that the initial points are almost assuredly not feasible.*

- Overall, these results are showing that the same solution is being found each time (so far).

- The number of iterations it takes to find the solution, and the trajectory of points through which it travels, is different for the various trials.

- The upper bound (*6000 acres*) used for randomizing the start points never appears to be a binding constraint

Details

Setting up parameters & variables	Sets / Parameters / Variables	Solved using
Query CAST source data	**Sets:** - Land River Segments, Pollutants - BMPs (type = efficiency) - BMP groups - Load Sources - (BMP, BMP group) - (BMP, Load Source) - (BMP group, Load Source)	Pyomo (algebraic modeling language library for python) developed by Sandia National Laboratories
Build variables [lrseg, agency, load source, bmp, unit]	**Parameters** - Costs - Load Reduction Efficiencies - Base loading - Pre-bmp Acres for load sources	IPOPT (interior point / barrier method solver)
Get parameters (costs, efficiency values, base loading, acres)	**Variables** - Acres per (LRseg, Loadsource, BMP)	

Why analyze different starting points?

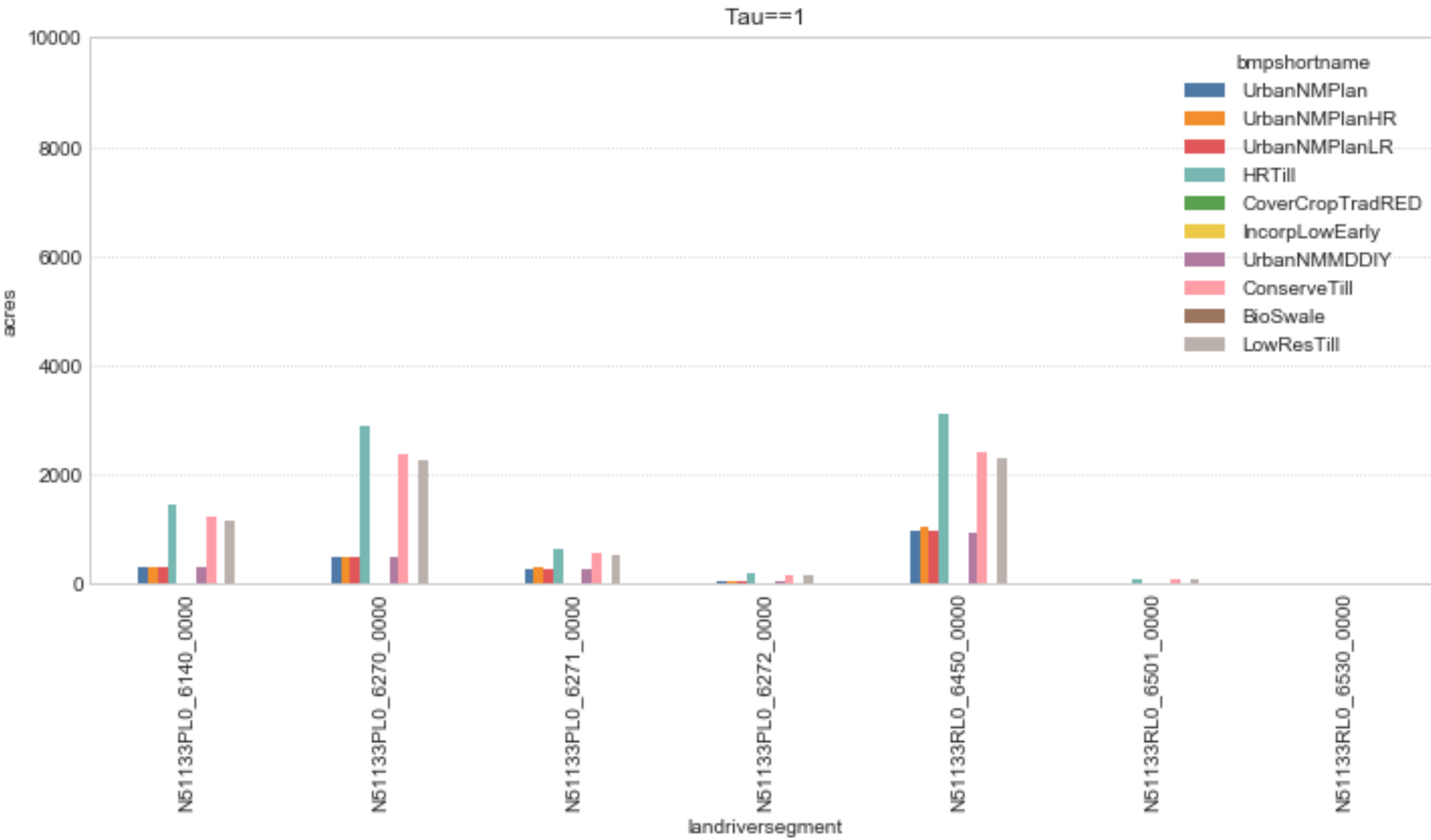
Non-convex load reduction function

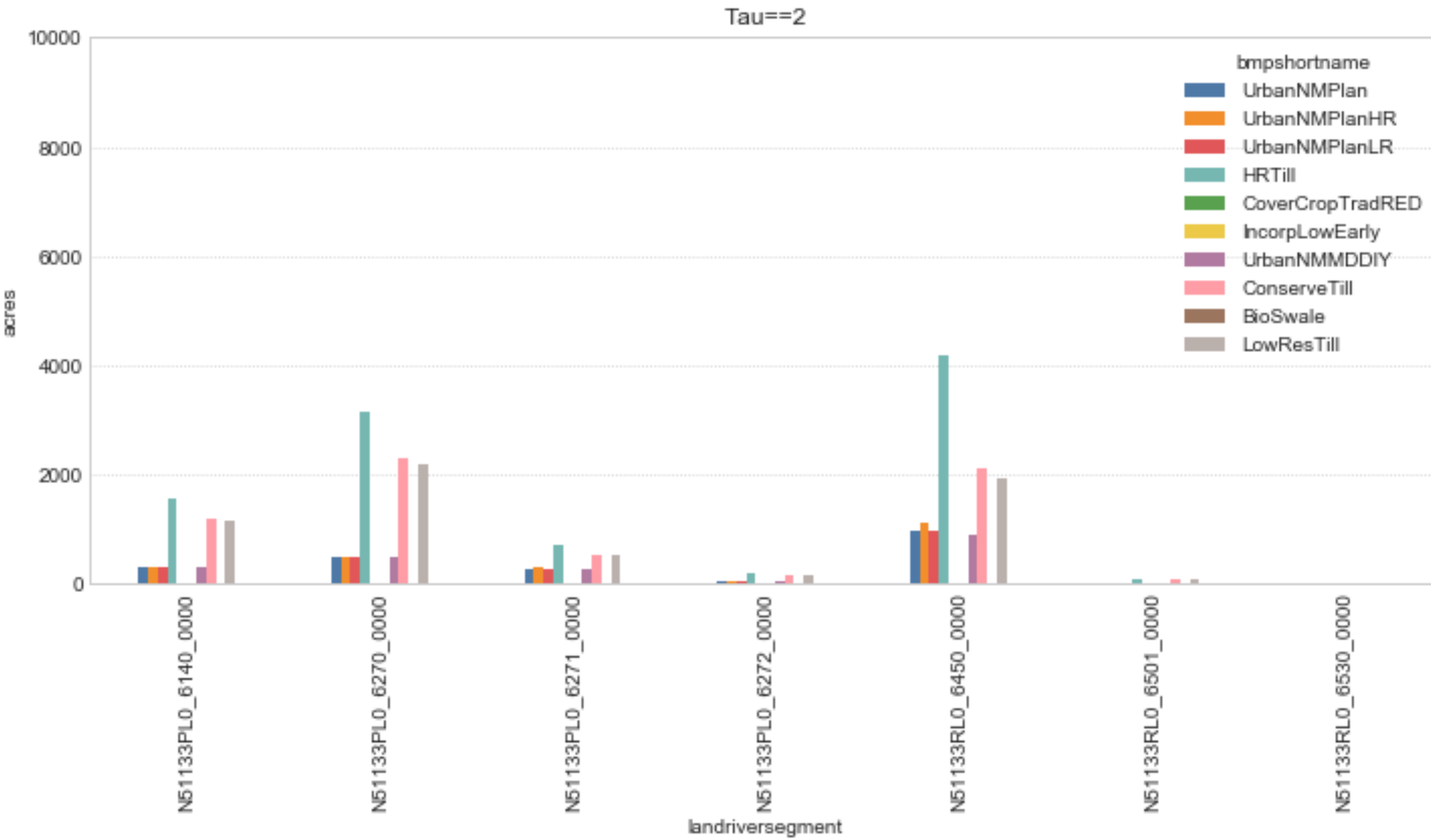
Local optimum may not be global optimum

Means that there may be starting point dependence

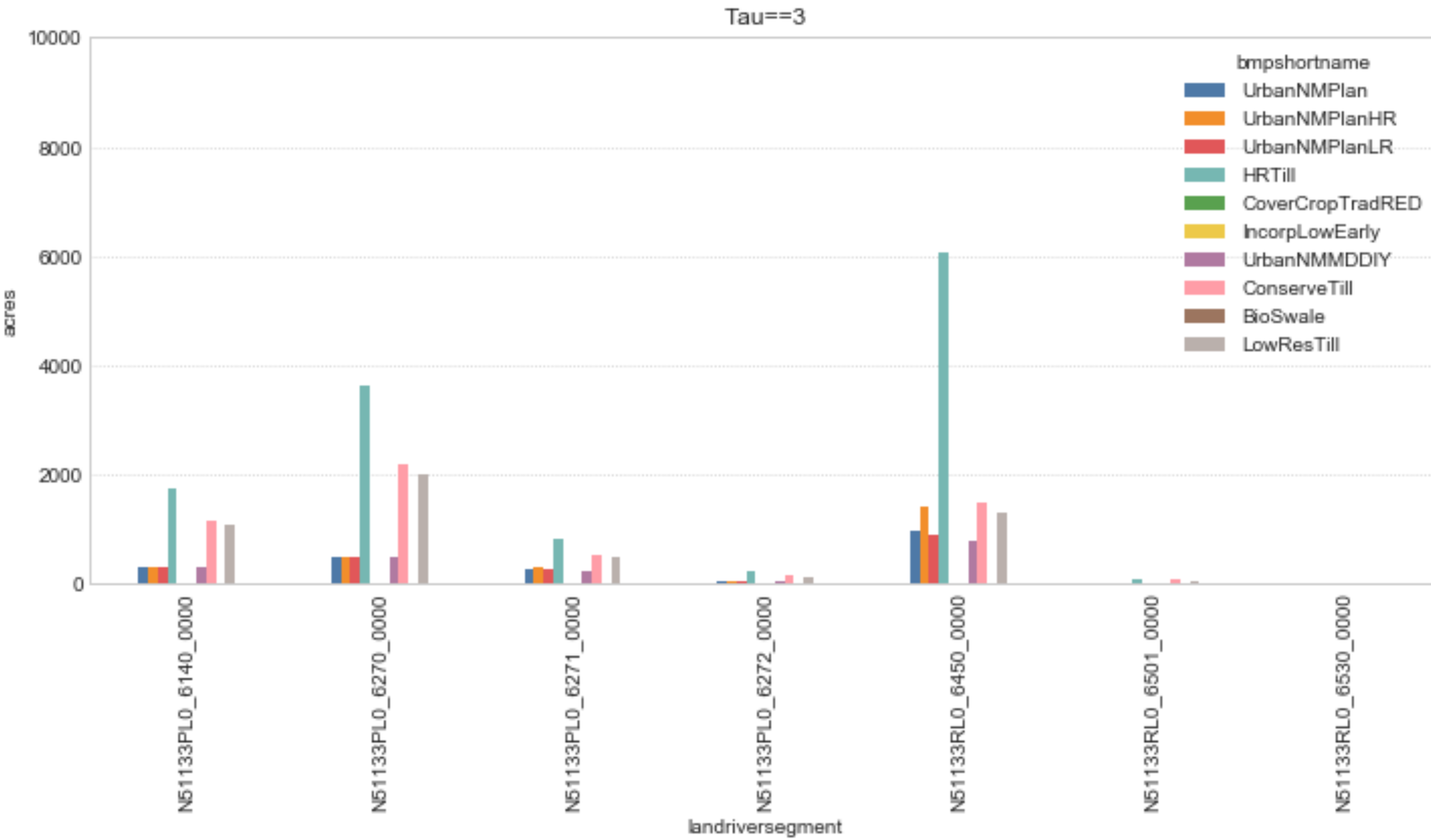
$$\mathrm{eig}(H) = \begin{bmatrix} \phi \\ \phi \\ \alpha_1 \sqrt{(\beta_1^2 + \beta_2^2)(\beta_3^2 + \beta_4^2)} \\ -\alpha_1 \sqrt{(\beta_1^2 + \beta_2^2)(\beta_3^2 + \beta_4^2)} \end{bmatrix}$$

Hessian is not positive definite, therefore, f(x) is not convex

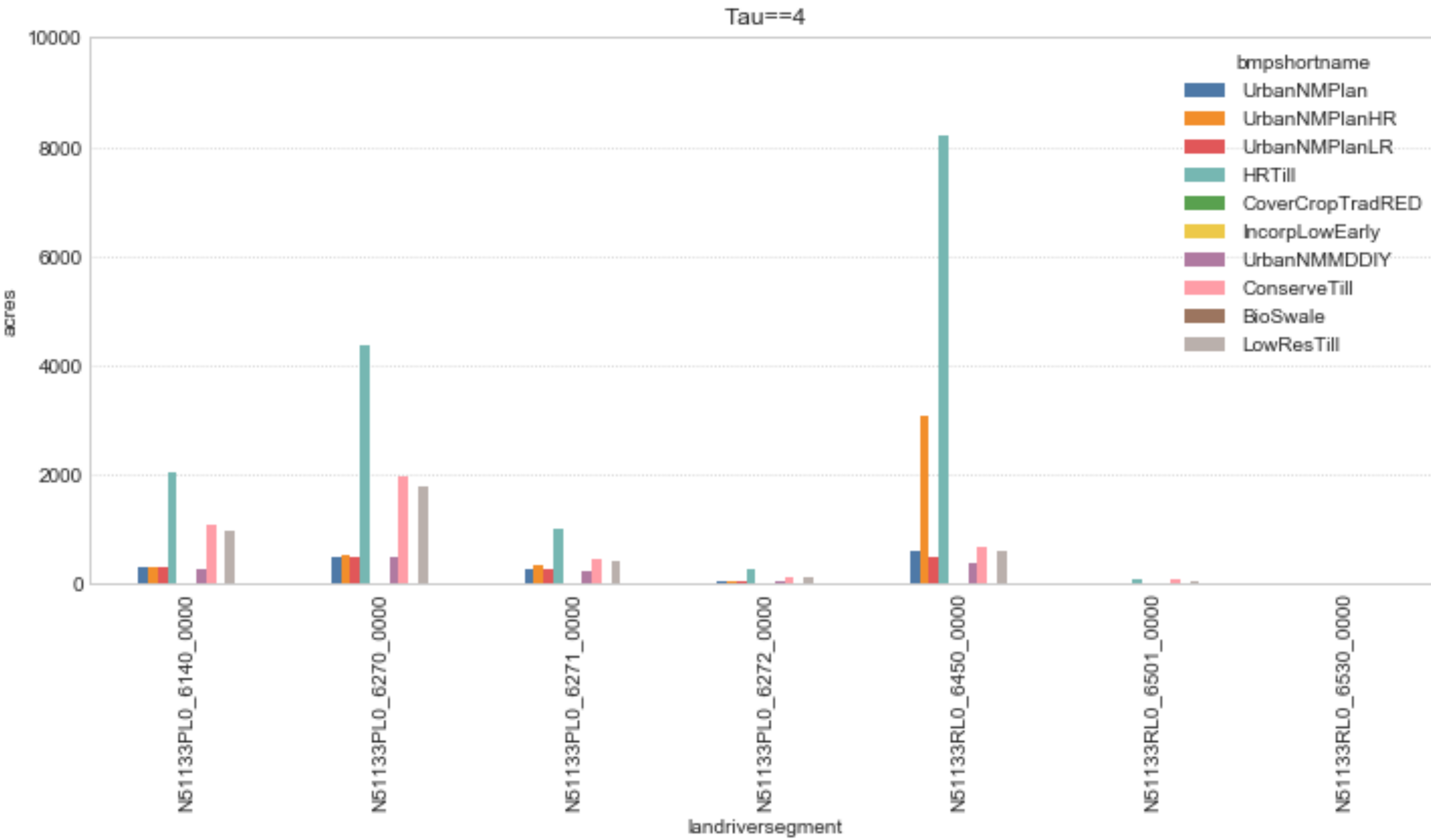

Tau==1
10000
8000
6000
4000
2000
0
acres
bmpshortname
UrbanNMPlan
UrbanNMPlanHR
UrbanNMPlanLR
HRTill
CoverCropTradRED
IncorpLowEarly
UrbanNMMDDIY
ConserveTill
BioSwale
LowResTill
N51133PL0_6140_0000
N51133PL0_6270_0000
N51133PL0_6271_0000
N51133PL0_6272_0000
N51133RL0_6450_0000
N51133RL0_6501_0000
N51133RL0_6530_0000
landriversegment



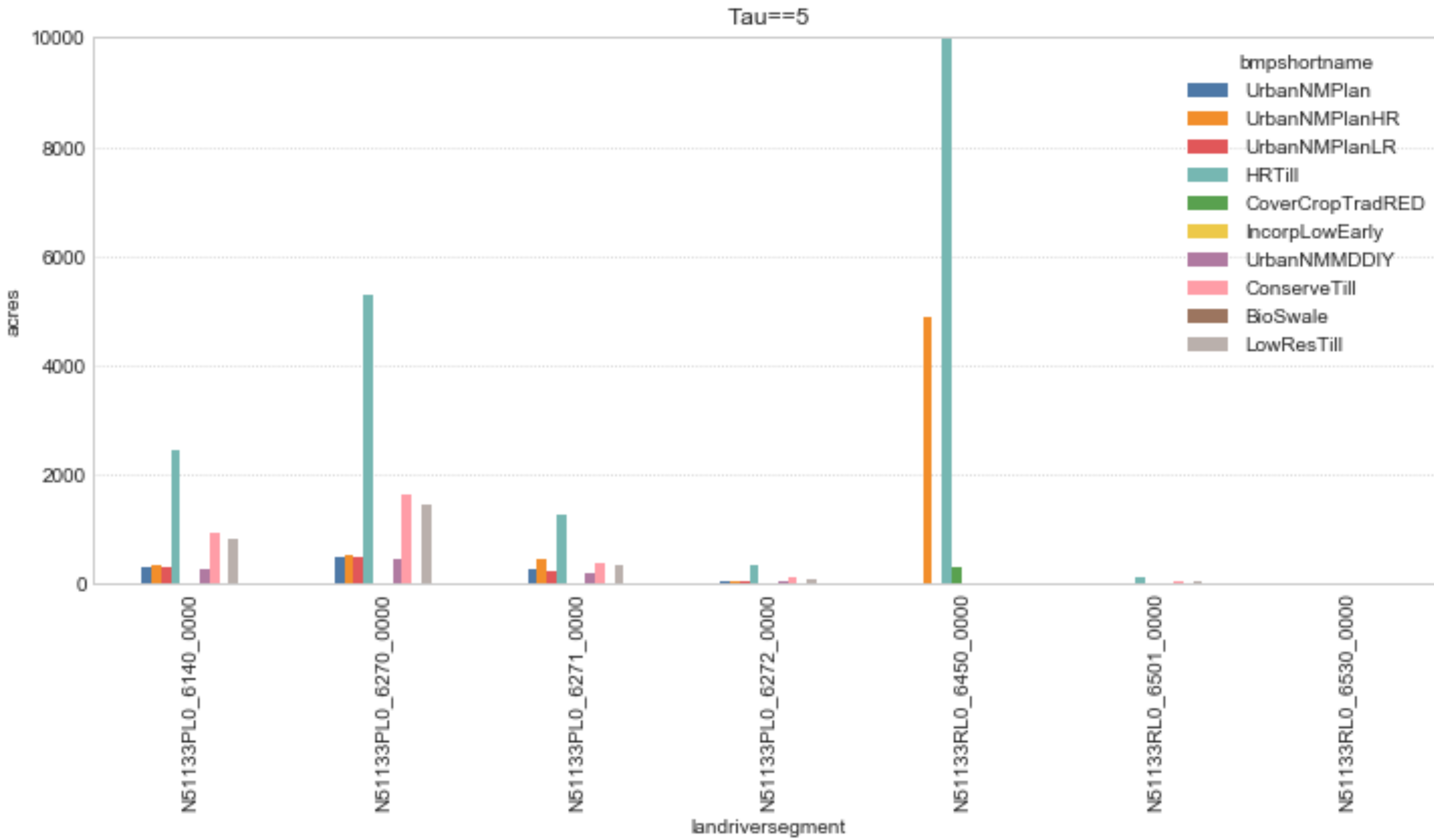
Tau==2
10000
8000
6000
4000
2000
0
acres
bmpshortname
UrbanNMPlan
UrbanNMPlanHR
UrbanNMPlanLR
HRTill
CoverCropTradRED
IncorpLowEarly
UrbanNMMDDIY
ConserveTill
BioSwale
LowResTill
N51133PL0_6140_0000
N51133PL0_6270_0000
N51133PL0_6271_0000
N51133PL0_6272_0000
N51133RL0_6450_0000
N51133RL0_6501_0000
N51133RL0_6530_0000
landriversegment

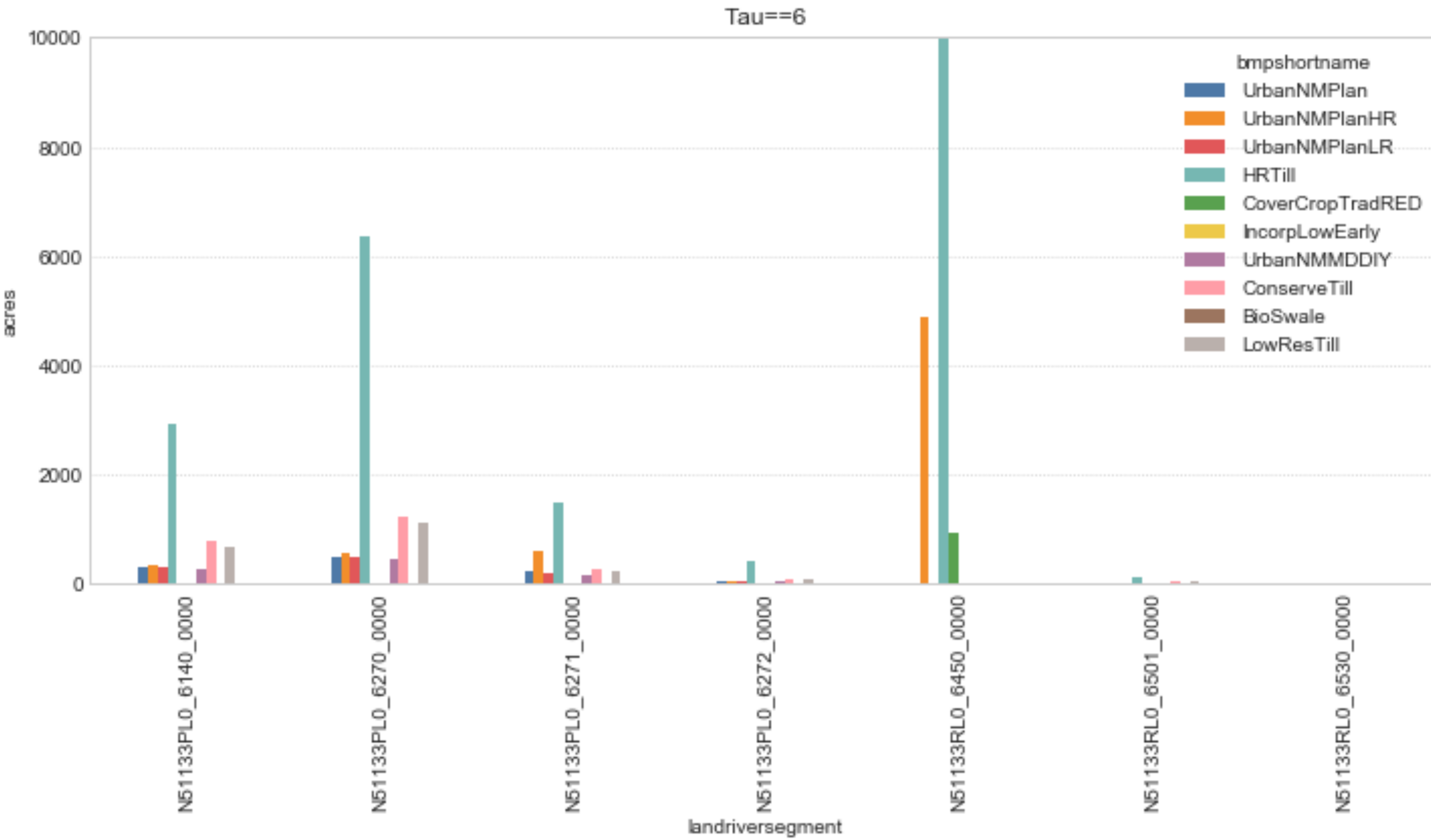


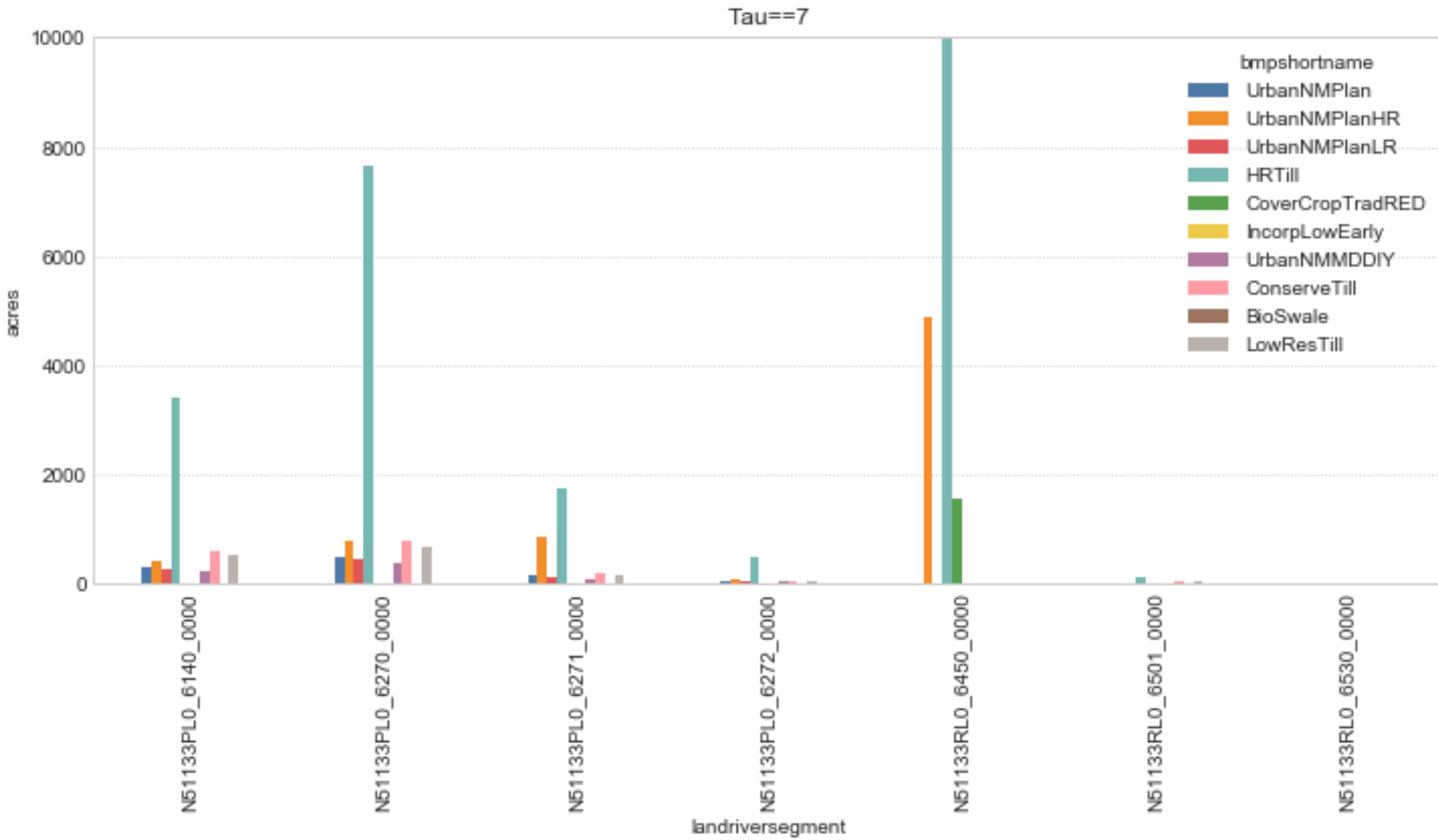
Tau==3
10000
8000
6000
4000
2000
0
acres
bmpshortname
UrbanNMPlan
UrbanNMPlanHR
UrbanNMPlanLR
HRTill
CoverCropTradRED
IncorpLowEarly
UrbanNMMDDIY
ConserveTill
BioSwale
LowResTill
N51133PL0_6140_0000
N51133PL0_6270_0000
N51133PL0_6271_0000
N51133PL0_6272_0000
N51133RL0_6450_0000
N51133RL0_6501_0000
N51133RL0_6530_0000
landriversegment

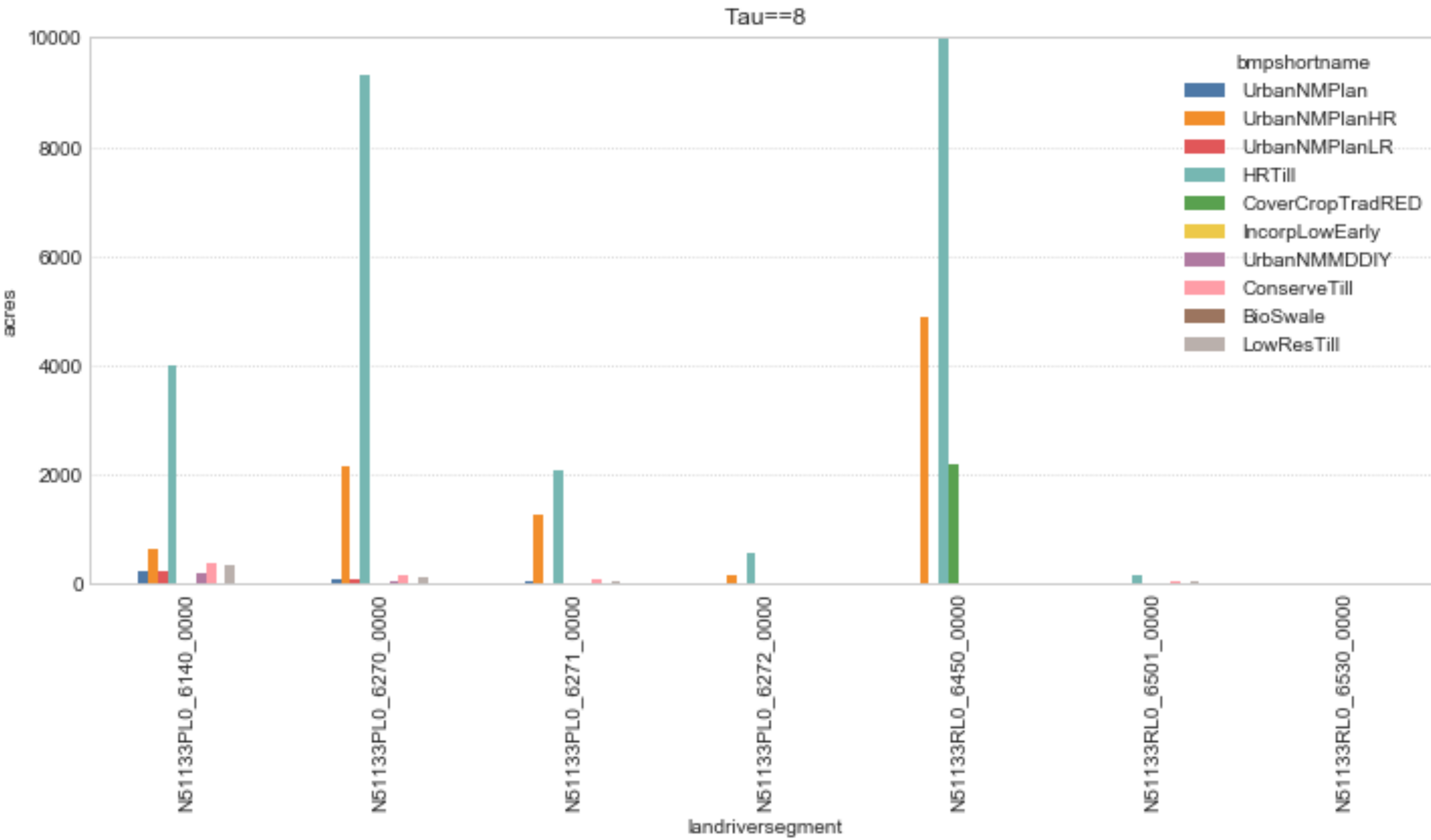


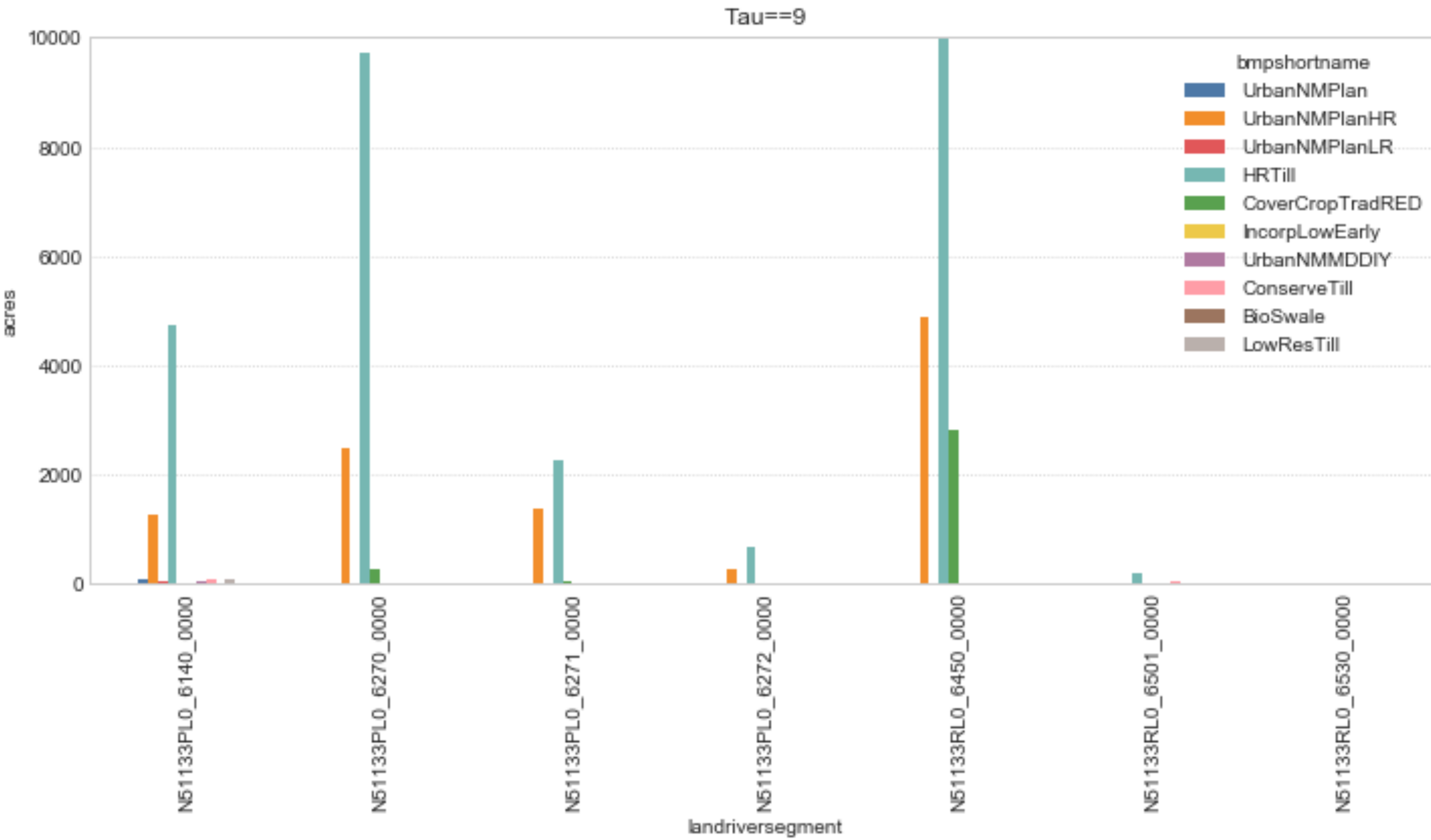
Tau==4
10000
8000
6000
4000
2000
0
acres
bmpshortname
UrbanNMPlan
UrbanNMPlanHR
UrbanNMPlanLR
HRTill
CoverCropTradRED
IncorpLowEarly
UrbanNMMDDIY
ConserveTill
BioSwale
LowResTill
N51133PL0_6140_0000
N51133PL0_6270_0000
N51133PL0_6271_0000
N51133PL0_6272_0000
N51133RL0_6450_0000
N51133RL0_6501_0000
N51133RL0_6530_0000
landriversegment

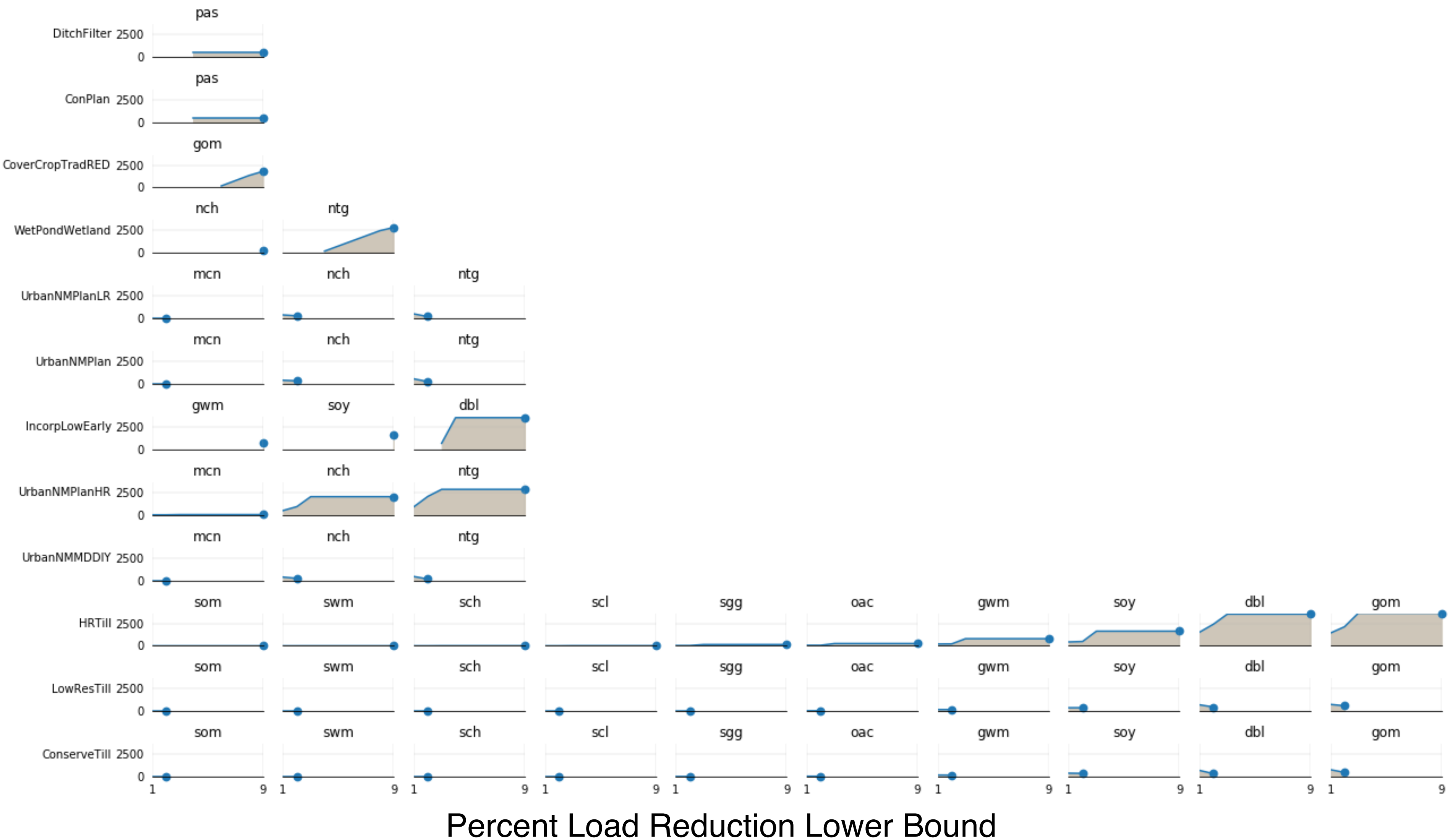

Tau==5
acres
10000
8000
6000
4000
2000
0
bmpshortname
UrbanNMPlan
UrbanNMPlanHR
UrbanNMPlanLR
HRTill
CoverCropTradRED
IncorpLowEarly
UrbanNMMDDIY
ConserveTill
BioSwale
LowResTill
N51133PL0_6140_0000
N51133PL0_6270_0000
N51133PL0_6271_0000
N51133PL0_6272_0000
N51133RL0_6450_0000
N51133RL0_6501_0000
N51133RL0_6530_0000
landriversegment


Tau==6
10000
8000
6000
4000
2000
0
acres
bmpshortname
UrbanNMPlan
UrbanNMPlanHR
UrbanNMPlanLR
HRTill
CoverCropTradRED
IncorpLowEarly
UrbanNMMDDIY
ConserveTill
BioSwale
LowResTill
N51133PL0_6140_0000
N51133PL0_6270_0000
N51133PL0_6271_0000
N51133PL0_6272_0000
N51133RL0_6450_0000
N51133RL0_6501_0000
N51133RL0_6530_0000
landriversegment


Tau==7
10000
8000
6000
4000
2000
0
acres
bmpshortname
UrbanNMPlan
UrbanNMPlanHR
UrbanNMPlanLR
HRTill
CoverCropTradRED
IncorpLowEarly
UrbanNMMDDIY
ConserveTill
BioSwale
LowResTill
N51133PL0_6140_0000
N51133PL0_6270_0000
N51133PL0_6271_0000
N51133PL0_6272_0000
N51133RL0_6450_0000
N51133RL0_6501_0000
N51133RL0_6530_0000
landriversegment

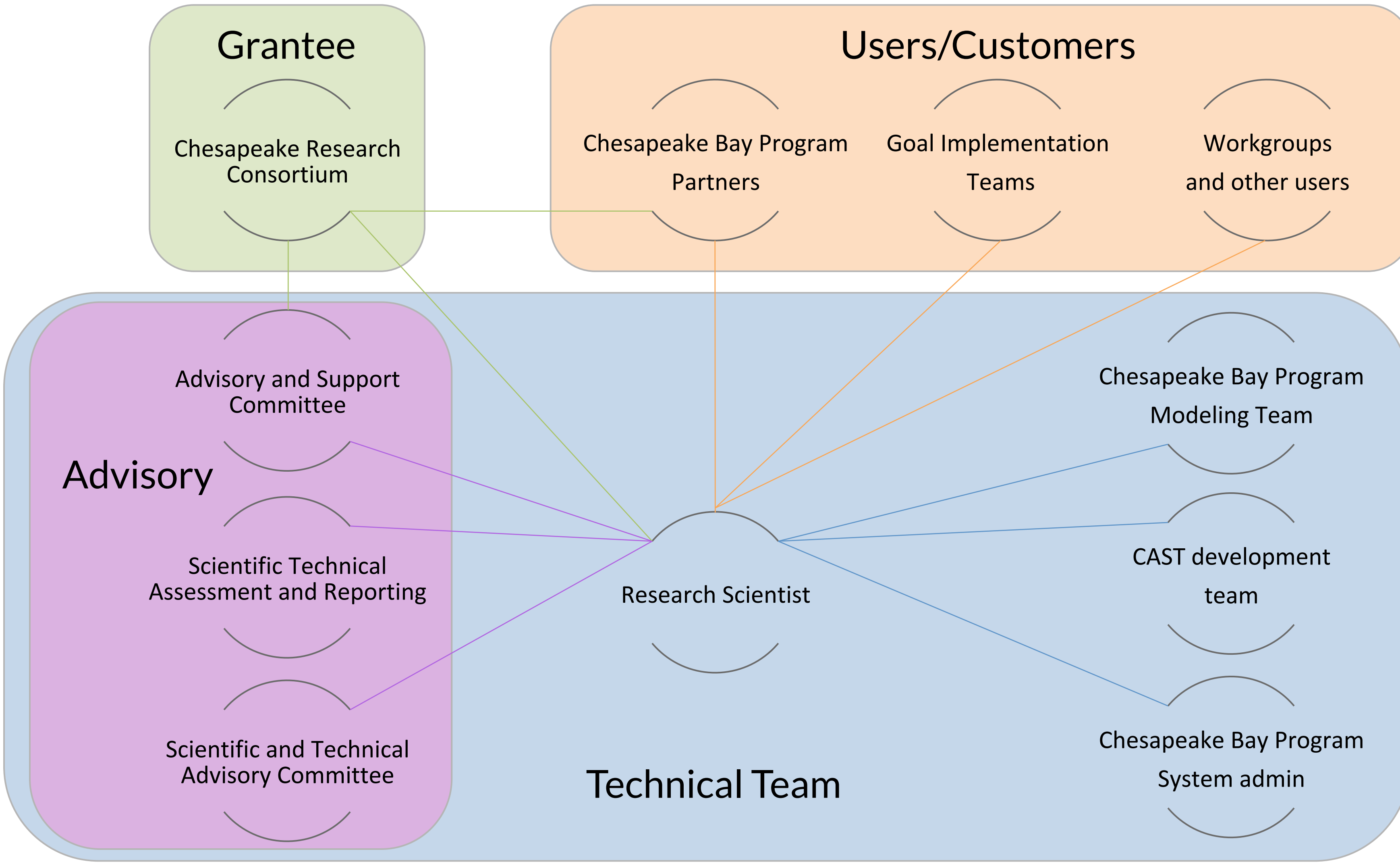

Tau==8
acres
10000
8000
6000
4000
2000
0
bmpshortname
UrbanNMPlan
UrbanNMPlanHR
UrbanNMPlanLR
HRTill
CoverCropTradRED
IncorpLowEarly
UrbanNMMDDIY
ConserveTill
BioSwale
LowResTill
N51133PL0_6140_0000
N51133PL0_6270_0000
N51133PL0_6271_0000
N51133PL0_6272_0000
N51133RL0_6450_0000
N51133RL0_6501_0000
N51133RL0_6530_0000
landriversegment


Tau==9
10000
8000
6000
4000
2000
0
acres
bmpshortname
UrbanNMPlan
UrbanNMPlanHR
UrbanNMPlanLR
HRTill
CoverCropTradRED
IncorpLowEarly
UrbanNMMDDIY
ConserveTill
BioSwale
LowResTill
N51133PL0_6140_0000
N51133PL0_6270_0000
N51133PL0_6271_0000
N51133PL0_6272_0000
N51133RL0_6450_0000
N51133RL0_6501_0000
N51133RL0_6530_0000
landriversegment



Objective:
Minimize Total Cost ($)

Project Organization



Current system

Best Management Practices (BMPs)

Chesapeake Assessment Scenario Tool (CAST)

Loads

Cost

Current system

Best Management Practices (BMPs)

Chesapeake Assessment Scenario Tool (CAST)

Not feasible to exhaustively try potential strategies

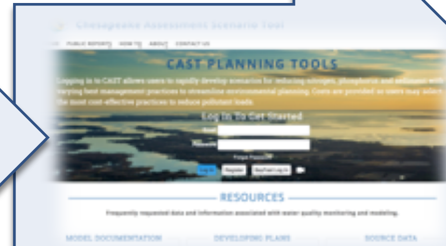
Loads

Cost

Scenario Optimization System

Best Management Practices (BMPs)

Chesapeake Assessment Scenario Tool (CAST)



Loads

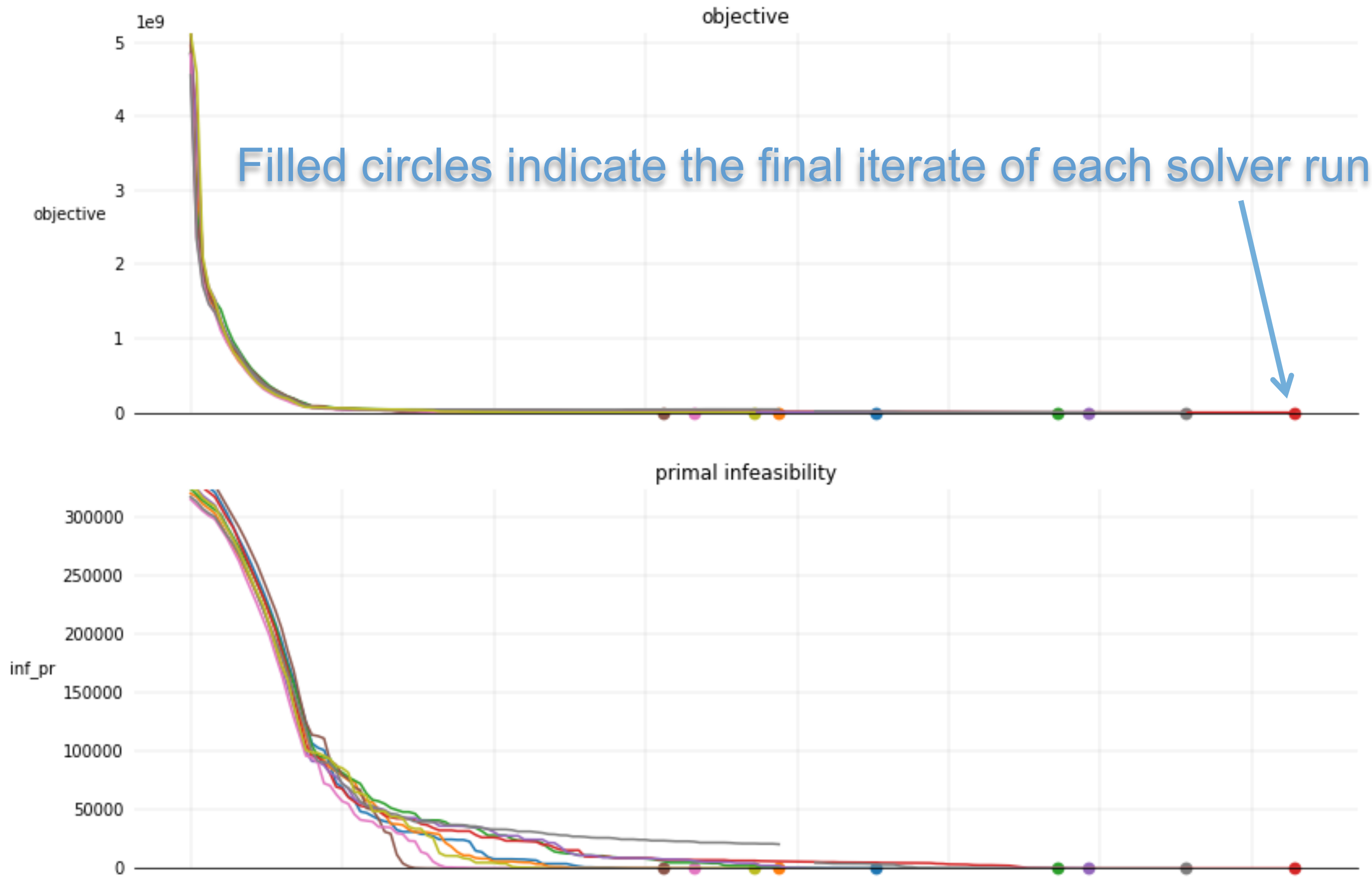
Cost

Analyze potential BMP options and identify low-cost strategies

To help the Chesapeake Bay Program and its Partners restore the Bay and its watershed

Multiple Runs

The objective and primal infeasibility follow similar trends for each of the randomized starting point trials.



Objective:
Minimize
Total Cost ($)

Why analyze different starting points?

$$f(x) = \alpha_1 \left(1 - \beta_3 x_3 + \beta_4 x_4 - \beta_1 x_1 + \beta_2 x_2 + \beta_1 \beta_3 x_1 x_3 + \beta_1 \beta_4 x_1 x_4 + \beta_2 \beta_3 x_2 x_3 + \beta_2 \beta_4 x_2 x_4\right)$$

(load reduction)

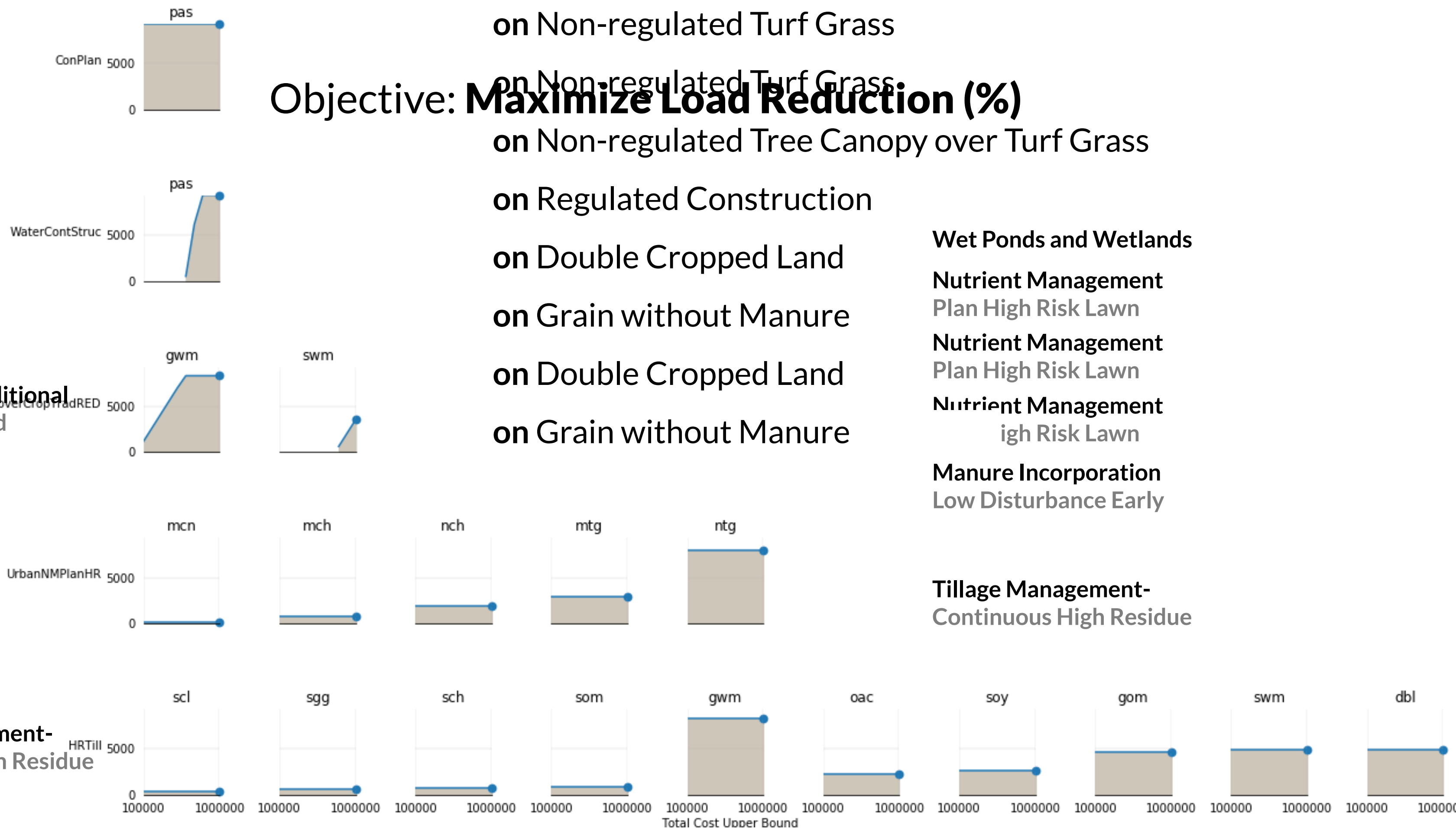
$$H = \begin{bmatrix} 0 & 0 & \alpha_1 \beta_1 \beta_3 & \alpha_1 \beta_1 \beta_4 \\ 0 & 0 & \alpha_1 \beta_2 \beta_3 & \alpha_1 \beta_2 \beta_4 \\ \alpha_1 \beta_1 \beta_3 & \alpha_1 \beta_2 \beta_3 & 0 & 0 \\ \alpha_1 \beta_1 \beta_4 & \alpha_1 \beta_2 \beta_4 & 0 & 0 \end{bmatrix}$$

$$\mathrm{eig}(H) = \begin{bmatrix} 0 \\ 0 \\ \alpha_1 \sqrt{(\beta_1^2 + \beta_2^2)(\beta_3^2 + \beta_4^2)} \\ -\alpha_1 \sqrt{(\beta_1^2 + \beta_2^2)(\beta_3^2 + \beta_4^2)} \end{bmatrix}$$

Cross products of the form x*y

Hessian is not positive definite, therefore, f(x) is not convex

Objective: **Maximize Load Reduction (%)**



on Non-regulated Turf Grass

on Non-regulated Turf Grass

on Non-regulated Tree Canopy over Turf Grass

on Regulated Construction

on Double Cropped Land

on Grain without Manure

on Double Cropped Land

on Grain without Manure

Wet Ponds and Wetlands

Nutrient Management
Plan High Risk Lawn

Nutrient Management
Plan High Risk Lawn

Nutrient Management
igh Risk Lawn

Manure Incorporation
Low Disturbance Early

Tillage Management-
Continuous High Residue

Cover Crop Traditional
Rye Early Drilled

Tillage Management-
Continuous High Residue

Total Cost Upper Bound
($ 200,000 to $ 1,000,000)

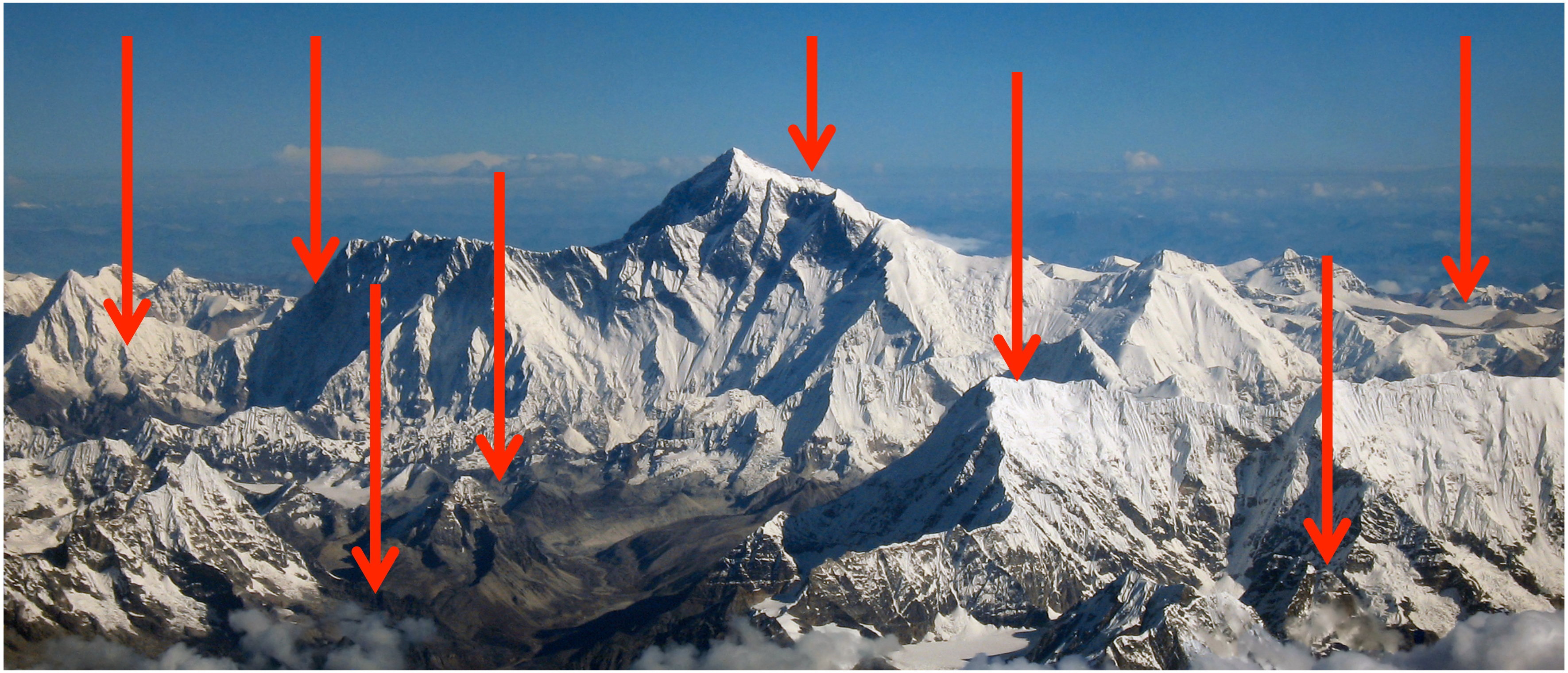
Prototyping Experiments

Search Space Investigations **Fine-tuned sub-problem**

Prototyping Experiments

Search Space Investigations

Fine-tuned sub-problem



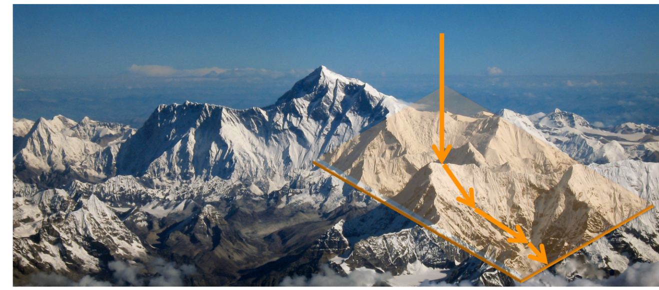
What the future may hold...

Two land-river segments

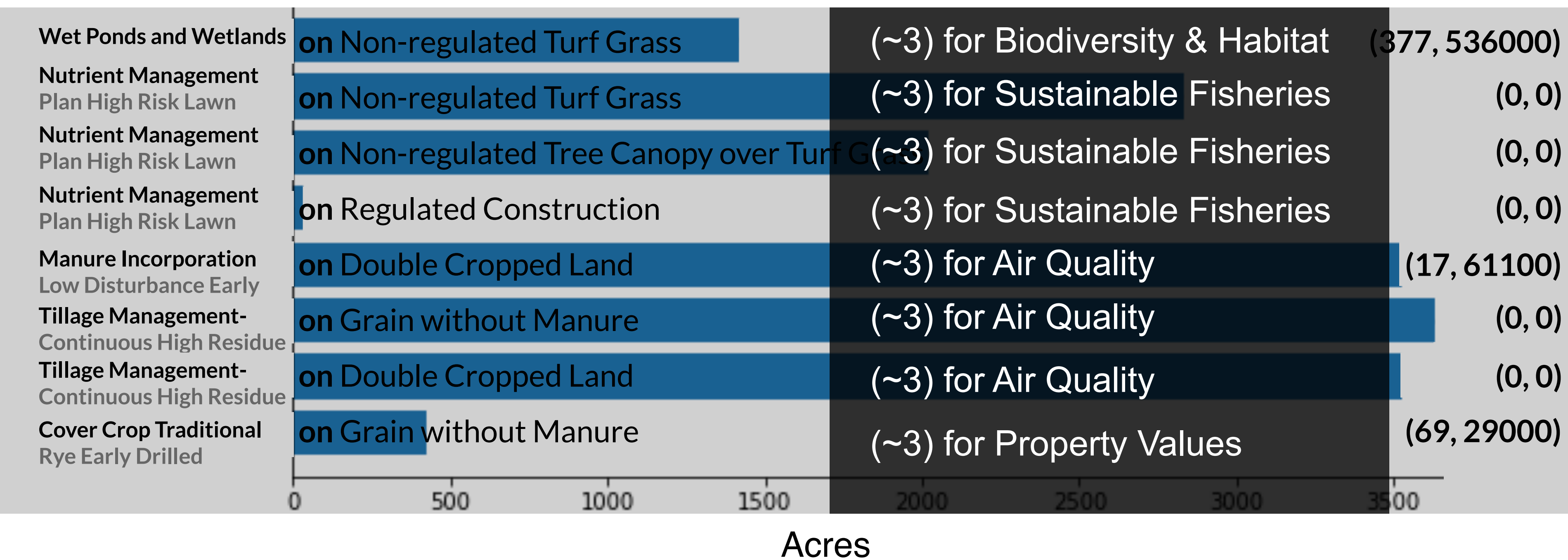
	Northumberland County, VA		Lancaster County, PA	
Objective:	Min. Cost (τ=12)	Max Load Reduction (C=100,000)	Min. Cost (τ=12)	Max Load Reduction (C=100,000)
# of variables	1339	1339	1339	1339
# of inequality constraints	1110	1108	1110	1108
# of nonzeros in inequality constraint Jacobian	3609	2628	3862	2628
# of nonzeros in Lagrangian Hessian	5575	5278	6058	5746
Iterations to solve	57	35	47	29

Experiments –
(2) Fine-tuned sub-problem

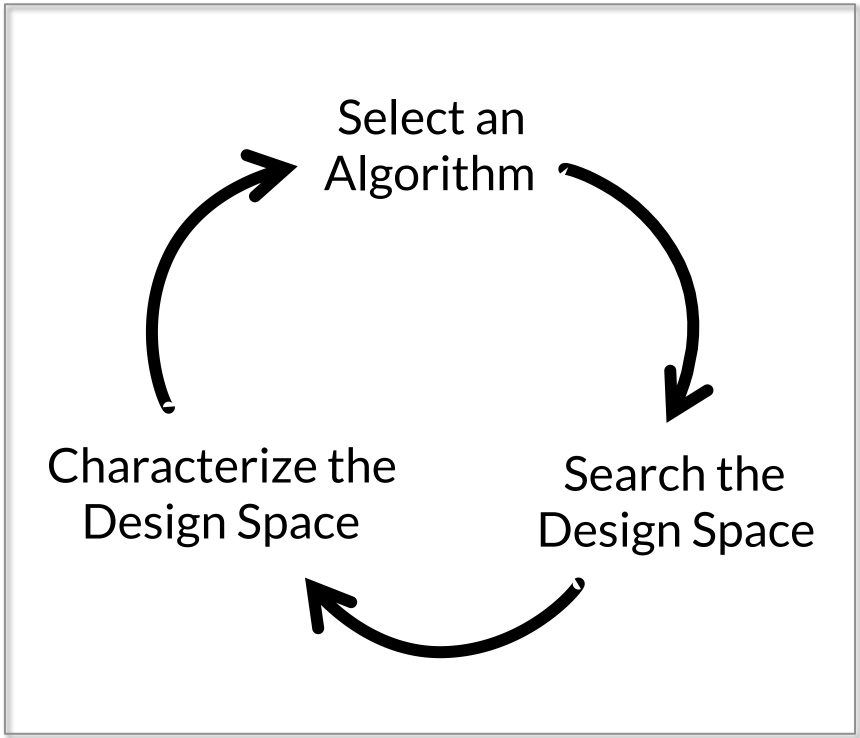
- Guide explorations of increasing complexity
- Solutions for select BMPs



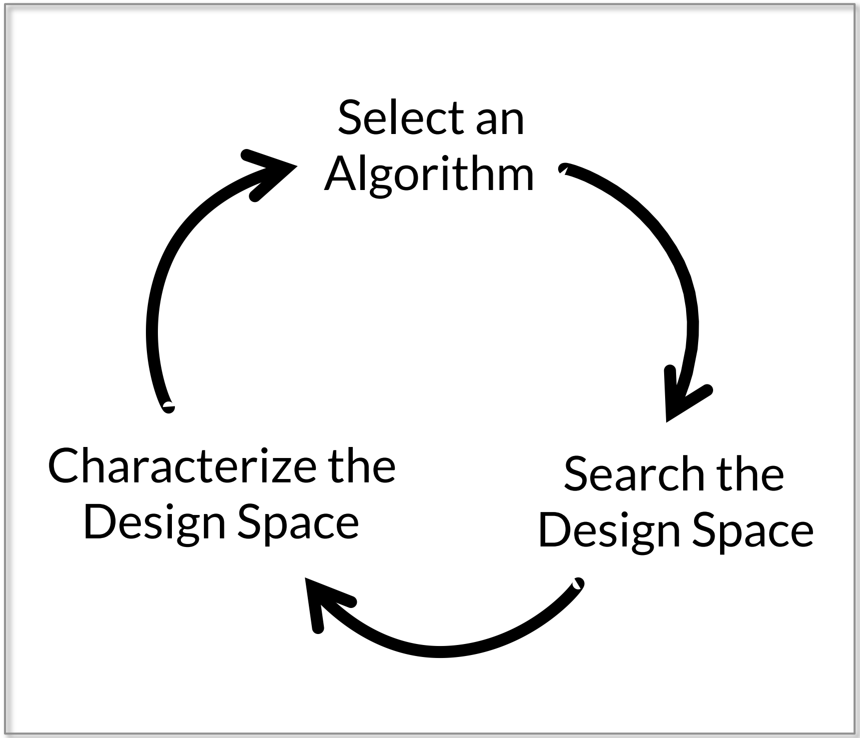
Example Results (6% Reduction Constraints)



Medium-term goals: Post Core-CAST approaches



After prototyping experiments…



Various Possible Approaches

- **Population-based stochastic search**

  (e.g. Genetic algorithm)
- **Decomposing into sub-problems, with multiple algorithms**
  - *Population-based for land use change and/or manure transport*
  - *Greedy algorithm or nonlinear programming for efficiency BMPs*
- **Model training**

Figure modified from Red Cedar Tech., WP-1022

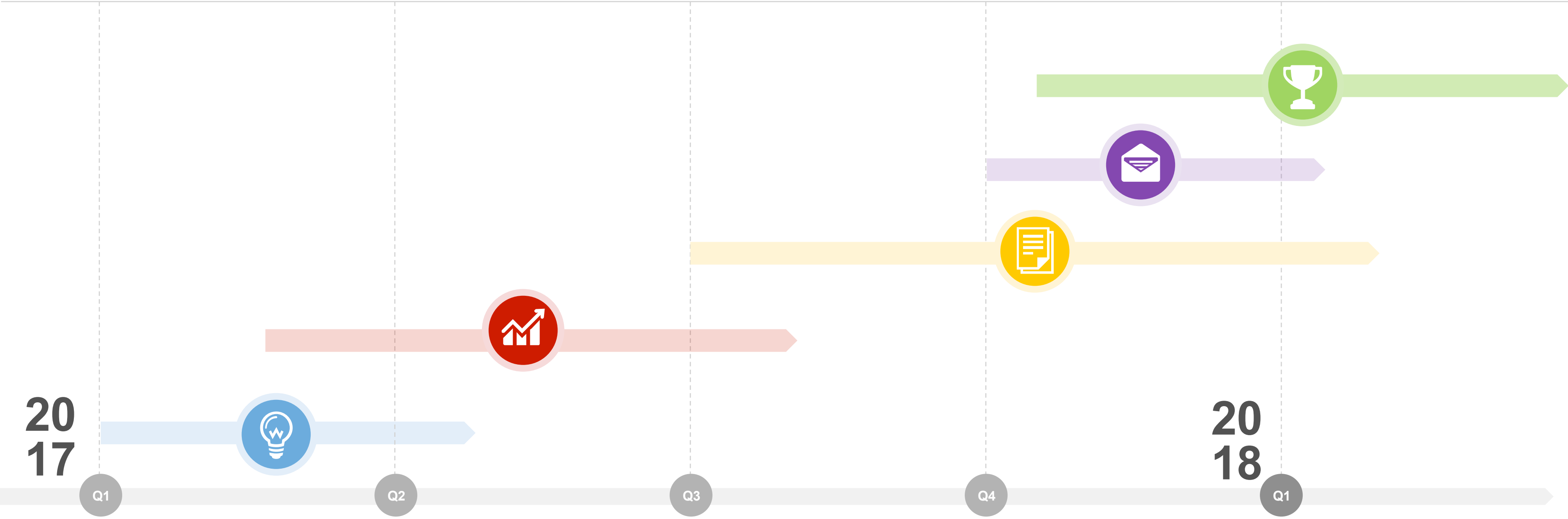
Heuristic + NLP?

Pseudocode for Hybrid algorithm:

P <- Build population of non-efficiency-BMP scenarios
Best <- Ø
repeat
    for each individual P_i in *P* **do**
        Fitness(P_i) <- NLP solution (for efficiency-BMPs)   # using P_i variables, costs, & loads as starting point

        if *Best* = Ø or Fitness(P_i) > Fitness(*Best*) **then**
            Best <- P_i
    P <- modified *P* using a heuristic (select, mutate, recombine, etc.) according to the fitnesses
until *Best* is the ideal solution or we have run out of time
return *Best*

Project Timeline



Dummy Title 1

Mauris rhoncus tortor sed lacus finibus, eget facilisis sapien vulputate. Maecenas rutrum sollicitudin suscipit.

Dummy Title 2

Mauris rhoncus tortor sed lacus finibus, eget facilisis sapien vulputate. Maecenas rutrum sollicitudin suscipit.

Dummy Title 3

Mauris rhoncus tortor sed lacus finibus, eget facilisis sapien vulputate. Maecenas rutrum sollicitudin suscipit.

Dummy Title 4

Mauris rhoncus tortor sed lacus finibus, eget facilisis sapien vulputate. Maecenas rutrum sollicitudin suscipit.

Dummy Title 5

Mauris rhoncus tortor sed lacus finibus, eget facilisis sapien vulputate. Maecenas rutrum sollicitudin suscipit.

Revised timeline

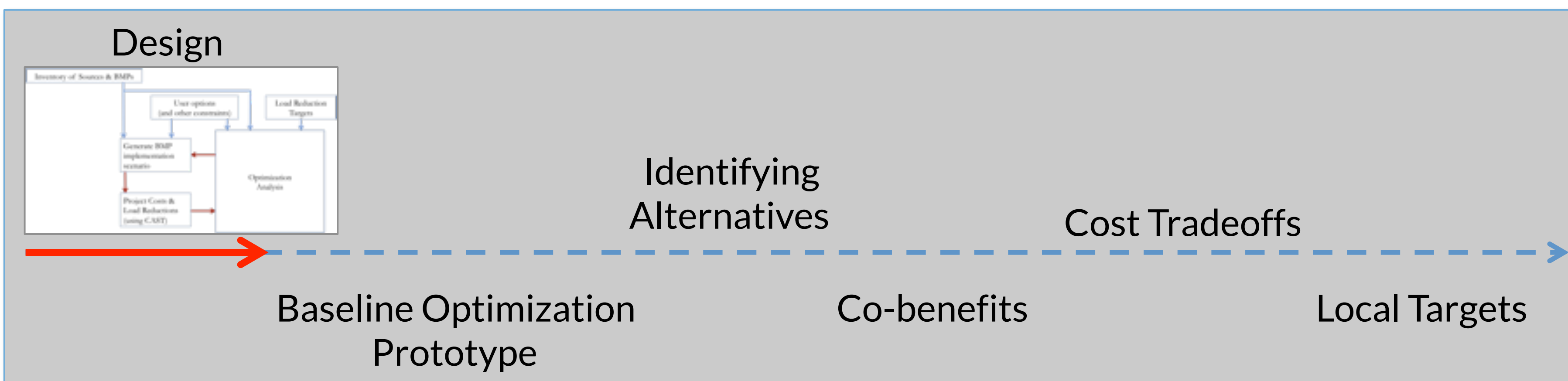
Date	CoreCAST Task	Optimization Task
May 2018	Detailed Design & Review with Jess and Dev Team	Formulate and code NLP sub-problem model
June - July 2018	CoreCAST Development	Conduct NLP sub-problem tests for efficiency BMPs
August 2018	CAST UI Integration & Performance Testing	Analysis of NLP sub-problem results
September 2018	CAST Master Processor Development	Analysis & visualization of batch scenario experiments
October 2018	Beta Testing / Results Verification; Integration with CAST UI	Scenario generator interfacing with CoreCAST Beta & Algorithm/package evaluation
November 2018	Performance Tuning	Design options for constraints and user interface
December 2018	CAST End User Testing	Beta Testing of version 0.1
January 1, 2019	Production Release	Beta Testing of version 0.1

Near-term Milestones

Date	Optimization Task
Summer 2018	Analyses of sampling experiments & sub-problem formulation
End of Summer	Scenario generator interfacing with CAST architecture update
Fall 2018	Algorithm/package evaluation
Winter 2018	Beta testing of version 0.1, constraints & user interface

Near-term Milestones & Looking Ahead

Date	Optimization Task
Summer 2018	Analyses of sampling experiments & sub-problem formulation
End of Summer	Scenario generator interfacing with CAST architecture update
Fall 2018	Algorithm/package evaluation
Winter 2018	Beta testing of version 0.1, constraints & user interface

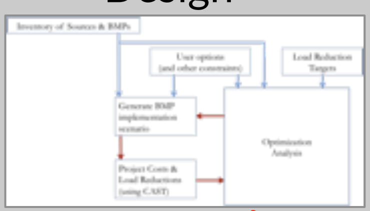


Will be shaped by feedback!

Actively searching for ways to engage local decision makers at county and municipal scales for their guidance and feedback on optimization design.

Email me (Danny) at: dkaufman@chesapeakebay.net

Design



Identifying Alternatives

Cost Tradeoffs

Baseline Optimization Prototype

Co-benefits

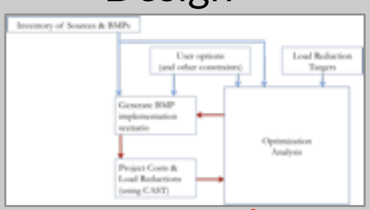
Local Targets

Will be shaped by feedback!

Actively searching for ways to engage local decision makers at county and municipal scales for their guidance and feedback on optimization design.

Email me (Danny) at: dkaufman@chesapeakebay.net

Design



Identifying Alternatives

Cost Tradeoffs

Baseline Optimization Prototype

Co-benefits

Local Targets

Extra Slides Follow

2016 to 2017

2016, February: STAC Workshop held

2016, December

- Grant for optimization tool development awarded to Chesapeake Research Consortium (CRC)
- Advisory and Support Committee (ASC) formed

2017, June

- STAC Workshop Report published
- Dr. Stuart Schwarz of UMBC worked on Phase 1 activities in June and July of 2017, and proposed a recursive greedy algorithm for adding improved feasible choices to a optimization solution set.

2017, December: Research Scientist hired

Optimization – from STAC to Development

2016 — **Advisory Committee**
Convene and financially retain the services of a team of recognized optimization experts to continue to provide their expert input during the development of the CAST-based optimization tool.

2017 — **Research Scientist**
Selection of an appropriate candidate by the Advisory and Support Committee in consultation with an *ad hoc* advisory board comprising key non-federal and federal CBP partners

early 2018 — **Initial steps**
Work with CAST developers, non-federal stakeholders and other users to look for model simplifications, to design an efficient interface between CAST and the proposed optimization software, and to ensure procedures are in place to collect the necessary information

late 2018 — **Development**
Develop a prototype optimization model for selecting management actions to achieve nonpoint source reduction, implemented for a well-characterized watershed

2019 — **Beta Version**
Test the optimization engine within multiple regions in all jurisdictions of the watershed; Update and begin to maintain the engine in response to early user feedback

Programmatic highlights since Dec, 2017

Meetings with:

- Advisory and Support Committee (~Quarterly)
- Dr. Hugh Ellis, at Johns Hopkins (~monthly)
- Modeling team (regularly)
- CBPO optimization technical team (modeling team, system administrators, CAST developers; monthly)
- CBPO cloud-computing group (~bi-weekly)
- CBPO User Experience team

Presentations and feedback gathered from:

- Water Quality Goal Implementation Team (WQGIT)
- Workgroups
  - Watershed Technical
  - Modeling
  - Urban Stormwater
  - Wastewater Treatment
- Scientific, Technical Assessment, and Reporting (STAR) team
- Chesapeake Research & Modeling Symposium

- Response to STAC workshop drafted for CBP Management Board
- Project workplan updated, and no-cost extension granted (to March 31, 2020)

Technical highlights since Dec, 2017

- Pre-prototype software for on-the-fly generation of Best Management Practice (BMP) scenario files for use with CAST.

- Two prototype optimization models for efficiency BMPs (a sub-population of all BMPs) have been developed for
  1. minimizing total scenario cost while achieving nonpoint source reduction, and
  2. maximizing total load reductions subject to a cost constraint

- Efficiency BMP optimization problem analyses established:
  - non-convexity of the load reduction function
  - negligible starting point sensitivity (for single land river segments)
  - results for constraint variations