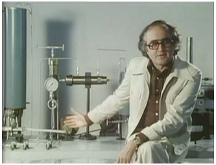
Connections: Data, Decisions, Optimization, Cows and Fish

Michael C. Ferris

University of Wisconsin, Madison

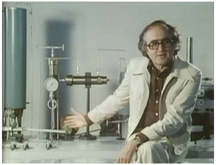
Mathematical Institute, Oxford University, September 5, 2017

Connections, Data and Innovation

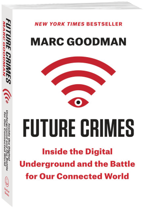


- James Burke, TV historian/science writer, and BBC series: Connections
- modern world...result of a web of interconnected events Turing,weak sharp minima,9/11,video-games,sparse approximation,Chebyshev,butterfly markings,iteration, each one consisting of a person or group acting for reasons of their own motivations (e.g., profit, curiosity, religious) with no concept of the final result

Connections, Data and Innovation

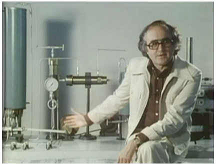


- James Burke, TV historian/science writer, and BBC series: Connections
- modern world...result of a web of interconnected events Turing,weak sharp minima,9/11,video-games,sparse approximation,Chebyshev,butterfly markings,iteration, each one consisting of a person or group acting for reasons of their own motivations (e.g., profit, curiosity, religious) with no concept of the final result

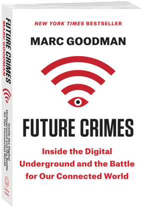


- Target sends adverts for baby clothes and cribs to 15 year old
- Father complains, then later retracts when learns truth
- Target inferred connection between customer purchases and outcomes *from data*

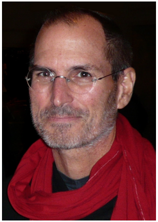
Connections, Data and Innovation



- modern world...result of a web of interconnected events , each one consisting of a person or group acting for reasons of their own motivations with no concept of the final result



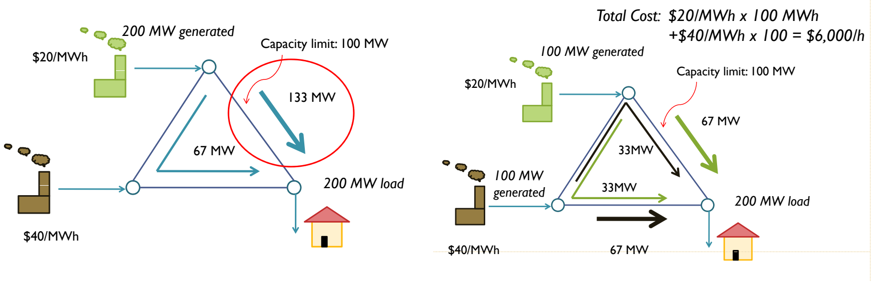
- Target inferred connection between customer purchases and outcomes *from data*



Steve Jobs: "If you're gonna make connections which are innovative ... you have to not have the same bag of experiences as everyone else does, or else you're going to make the same connections [as everybody else], and then you won't be innovative

Smart Grid: Transmission switching

Opening (removing) lines in a transmission network can reduce cost



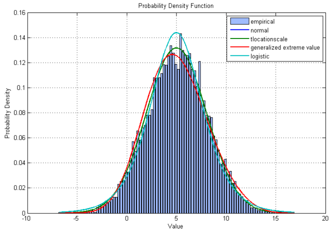
(a) Infeasible due to line capacity (b) Feasible dispatch

Must use expensive generator due to power flow characteristics

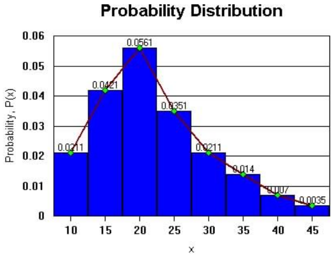
Domain expertise, integrated economic and power models, substantial (4-15%) savings estimated in $300 billion industry

Data, Computation (and Optimization)

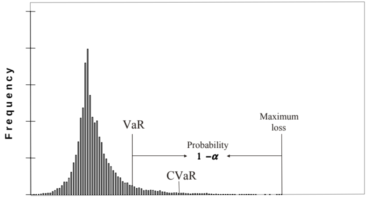
- Uncertainty: Input



- Quantification: Output

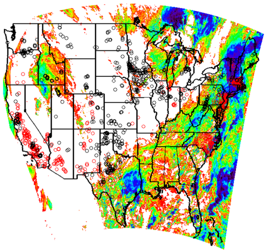


- Data can help manage or control output distribution



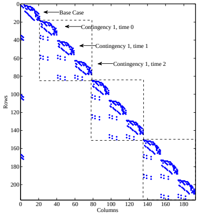
Satellite data, FERC and Reserves

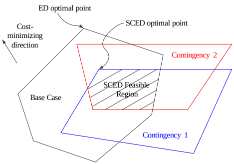
Solar transmittance and power



- Generators set aside capacity for "contingencies" (reserves)
- Separate energy π_d and reserve π_r prices
- Use 12 hour cloud cover forecasts to reduce reserves

- Federal Energy Regulatory Commission (FERC) contract to build models and data
- Provided on NEOS (Network enabled optimization system)





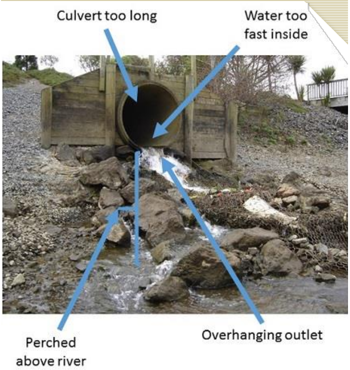
- Integrate satellite forecast data with power system data and smoke models to provide reliability and savings outcomes

Integrating systems, ideas and data to enhance (real) outcomes

- Philosophy
  - Focus on problems of national importance, relevant and able to leverage local impact and resources
  - Collaboratory, creating ownership, engaging at all levels
  - Drive outcomes with data, mitigate uncertainties
  - Enhance understanding of decision space, facilitate policy design and operational improvement
- Examples: Fishwerks and ANMODS
  - Key impact area: decision making in (environmentally) resource constrained problems
  - Feature: shared resource that interacts with complex multi-user systems
  - Visualize input data and results
  - Build appropriate models
  - Ensure fast enough solution for expert interaction

Fish and barriers

- 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 Huron: $>$ 86% of tributaries inaccessible
- Lake Michigan: $>$ 83% of tributaries inaccessible
- Lake Erie: $>$ 50% reduction of population size

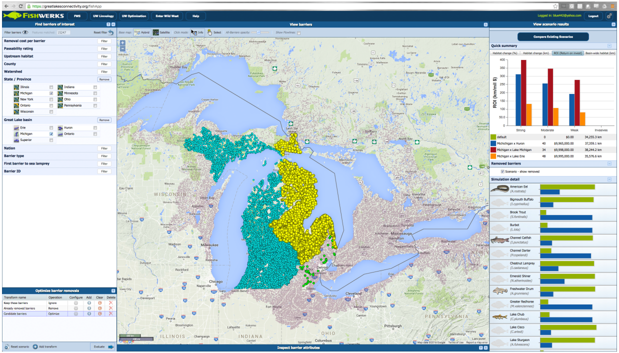
- Barriers can be mitigated to allow for fish passage: Removal of dams, improved road crossings, fish passageways
- However, they are very expensive - Average costs for fixes:
  - Dams: $100,000-650,000$ each
  - Others: \$30,000 - \$150,000 per project

- 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)

Objective

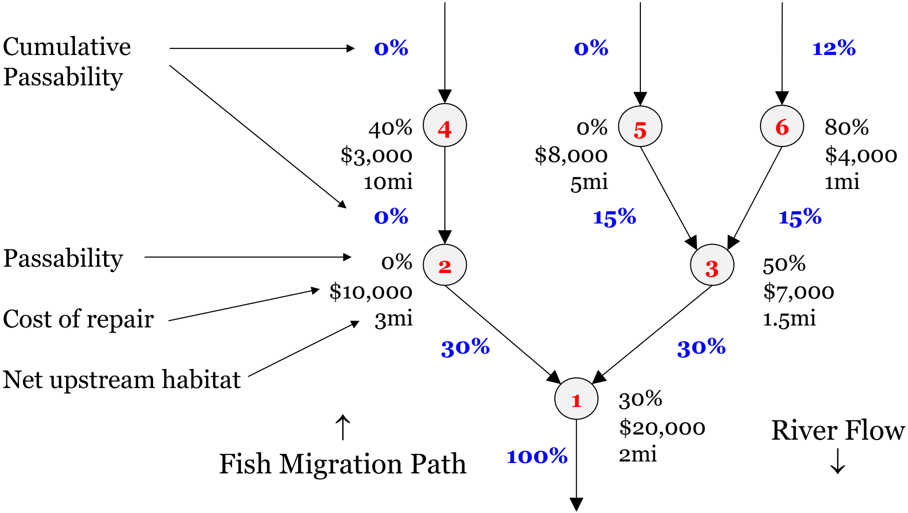
- Provide an interactive tool to consolidate big-data sets across multiple departments (DOT, DNR, FWS, NFPP, etc) and visually display in a meaningful way.
- Utilize optimization to maximize efficiency in policy decisions and funds appropriations.
- Allow any user to dynamically solve a large range of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

Fishwerks: A decision support tool



- Great Lakes basin scale data visualization
- Complex optimization for budget constraints, specific fish guilds, invasives
- Crowd sourcing data
- 230,000+ interdependent barriers
- Adopted by Fish and Wildlife Service `www.greatlakesconnectivity.org`
- Data integration, connecting stream and road systems, enhances outcome

Probability Chain



The Model

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

$$\text{s.t.} \sum_{j \in J} c_j x_j \leq B$$

$$z_{js} = (\bar{p}_{js} + \pi_{js} x_j) \times z_{d(j)s}, \qquad (j, s) \in J \times S$$

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

- 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
- $d(j)$ unique node downstream of j (none at "root" nodes $j \in R$)
- x_j remove barrier j (yes/no)

Linearization of products with binaries

$$z_{js} = (\bar{p}_{js} + \pi_{js} x_j) \times z_{d(j)s}, \quad (j,s) \in J \times S$$

Root nodes have no "downstream" nodes

$$z_{rs} = \bar{p}_{rs} + \pi_{rs} x_r$$

For others, introduce new variable: $y_{js} = x_j \times z_{d(j)s}$, $(j,s) \in (J \setminus R) \times S$

$$z_{js} = \bar{p}_{js} z_{d(j)s} + \pi_{js} y_{js}$$

Add additional constraints:

$$0 \leq y_{js} \leq x_j$$
$$y_{js} \leq z_{d(j)s}$$

Thus $y_{js} \leq x_j \times z_{d(j)s}$ but equality holds due to objective maximization

Basic Mixed Integer Programming (MILP) Model

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

$$\text{s.t.} \sum_{j \in J} c_j x_j \leq B$$

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

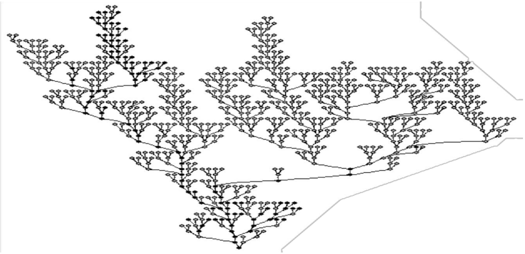
$$z_{js} = \bar{p}_{js} z_{d(j)s} + \pi_{js} y_{js}, \qquad (j, s) \in (J \setminus R) \times S$$

$$0 \leq y_{js} \leq x_j, \qquad (j, s) \in (J \setminus R) \times S$$

$$y_{js} \leq z_{d(j)s}, \qquad (j, s) \in (J \setminus R) \times S$$

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

The underlying issue

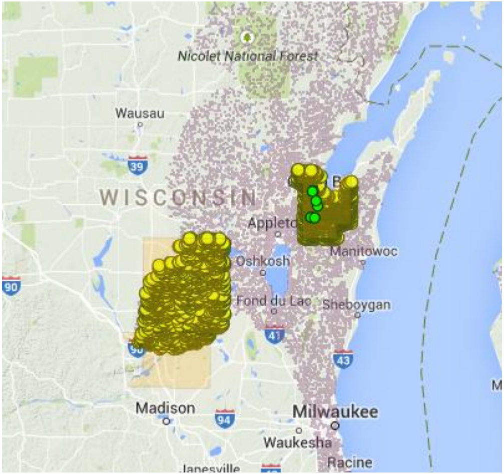


- 236,264 barriers [J], with network, geographical, type and cost data
- 36 Fish Species [S], with classification
- For every [$J \times S$] : 8,505,504
  - Passability Rating - % Chance species can pass this barrier
  - Upstream Habitat

- MILP is NP-Hard
  - Solution time quickly becomes impractical as problem size grows!
  - Application desires ROI curve over entire range of budgets and different scenarios!
  - Web tool requires fast processing to inform user
- Need to find methods to speed up solution time!
- Can we take advantage of the unique structure of our data?

Pre-Processing: data compression

- Disjoint areas: e.g. may desire collaboration between counties
- Downstream barriers affected by upstream decisions
- Barriers in-between are irrelevant
- Can be removed by compressing their data into other nodes!



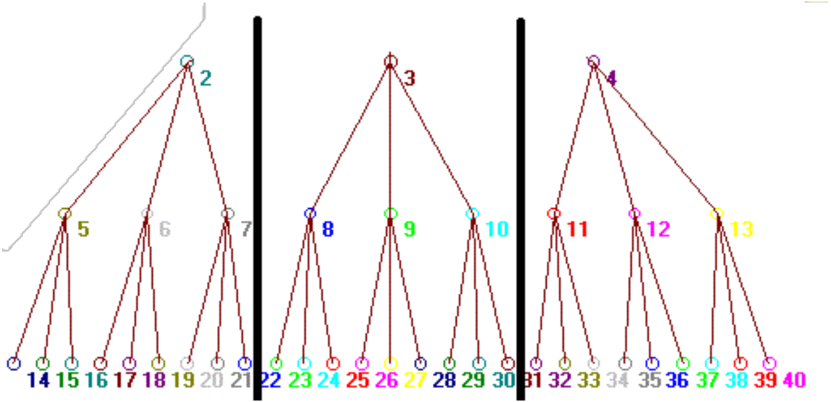
Fish guilds: variable reduction

- 36 species can be divided into 3 different guilds (fast, medium, slow swimmers)
- passability data (but not habitat data) only given by guild

$$
\begin{aligned}
\max \; & \sum_{j \in J} \sum_{s \in S \setminus \text{Inv}} v_{js} z_{jg(s)} \\
\text{s.t.} \; & \sum_{j \in J} c_j x_j \le B \\
& z_{rg} = \bar{p}_{rg} + \pi_{rg} x_r, && (r, g) \in R \times G \\
& z_{jg} = \bar{p}_{jg} z_{d(j)g} + \pi_{jg} y_{jg}, && (j, g) \in (J \setminus R) \times G \\
& 0 \le y_{jg} \le x_j, && (j, g) \in (J \setminus R) \times G \\
& y_{jg} \le z_{d(j)g}, && (j, g) \in (J \setminus R) \times G \\
& x_j \in \{0, 1\}
\end{aligned}
$$

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?



Solution approaches

- Pre-processing data due to spatially separated decision makers, or presence/absence data
- Variable reduction (on cumulative passability) due to fish guild commonalities
- Watershed decoupling (via Lagrangian relaxation)
- Use rMIP to determine budget allocation: replace actions by probabilities
- The combined reformulations provide adequate solution times, but...

The user wants more: additional constraints

- Supplement base model with additional constraints
- Ensure that available habitat for all (non-invasive) species increases from v_{0s} by specific amount U_s:

$$\sum_{j \in J} (v_{js} z_{jg(s)}) \geq v_{0s} U_s$$

- Prevent over-proliferation of invasive species:

$$\sum_{j \in J} (v_{ji} z_{jg(i)}) \leq v_{0i} U_i, \quad i \in \text{Inv}$$

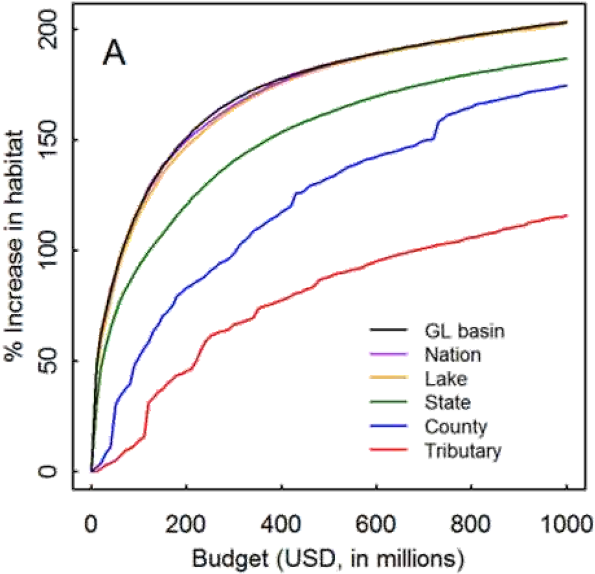
$$y_{ji} \geq z_{d(j)g(i)} + (x_j - 1), \quad (j, i) \in J \times \text{Inv}$$

- Note that last constraint is necessary to enforce equality in definition of y_{ji} - otherwise model will want to set y_{ji} to 0

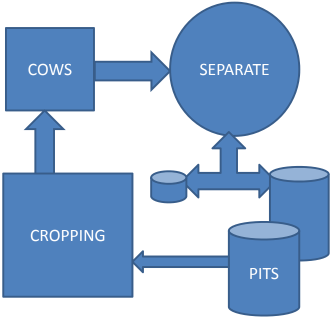
Decomposition approaches

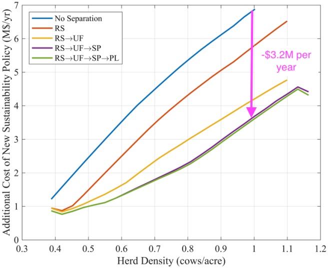
- Imbalanced watersheds (several have most of the barriers) - use better load balancing
- Decompose large watersheds by "precomputing" decisions at nodes near lake
- Minimum increase constraints inactive (remove)
- Force binary decisions on "Left Out" barriers used for load balancing
- Allocate monetary and invasives budgets in each subnetwork based on rMIP solution

Collaboration Matters!



Biomass Research and Development Initiative (BRDI)

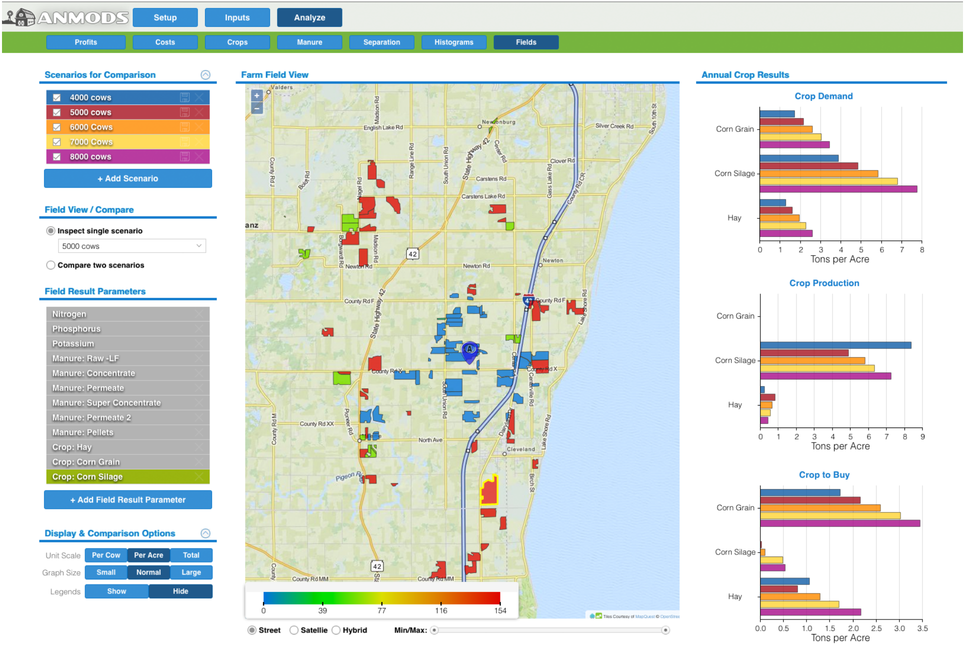




- Whole farm (complex interacting) mathematical model
- Long term sustainable (environment and financial)
- Economic/Logistic Optimization, with phosphorus runoff, other environmental restrictions
- Incorporates data analytics (e.g. WI regulatory tool SNAP+)
- Interconnected complex system linked by domain data with verifiable outcome

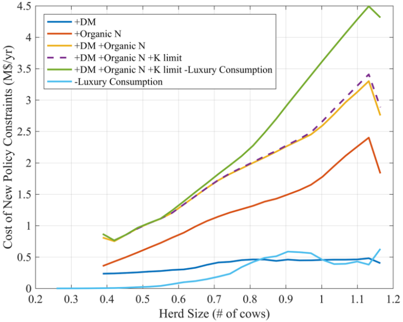
Mathematical details

- rotation = (ofas, hay, hay, hay, corn silage, corn silage, grain, grain)
- sustainable: must be able to repeat indefinitely = same state at end of 8 years as at start
- fields (order 150), rotations (order 30 * 8), separation types (order 8), manure applications
- Key variables: use(rot,year,field), apply(src,year,field), frac(cows,src), amount(src,year,field)
- Large MIP, relaxation bound hard to improve - symmetry in start years
- Model is largely indifferent to start year of rotation, solve for rotations, then balance to enforce contraints


ANMODS
Setup
Inputs
Analyze
Profits
Costs
Crops
Manure
Separation
Histograms
Fields
Scenarios for Comparison
4000 cows
5000 cows
6000 Cows
7000 Cows
8000 cows
+ Add Scenario
Field View / Compare
Inspect single scenario
5000 cows
Compare two scenarios
Field Result Parameters
Nitrogen
Phosphorus
Potassium
Manure: Raw -LF
Manure: Concentrate
Manure: Permeate
Manure: Super Concentrate
Manure: Permeate 2
Manure: Pellets
Crop: Hay
Crop: Corn Grain
Crop: Corn Silage
+ Add Field Result Parameter
Display & Comparison Options
Unit Scale
Per Cow
Per Acre
Total
Graph Size
Small
Normal
Large
Legends
Show
Hide
Farm Field View
0
39
77
116
154
Street
Satellite
Hybrid
Min/Max:
Annual Crop Results
Crop Demand
Corn Grain
Corn Silage
Hay
0 1 2 3 4 5 6 7 8
Tons per Acre
Crop Production
Corn Grain
Corn Silage
Hay
0 1 2 3 4 5 6 7 8 9
Tons per Acre
Crop to Buy
Corn Grain
Corn Silage
Hay
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5
Tons per Acre

Nutrient management = water quality problem

- Nitrogen, as nitrate (NH3), leaches through soil into underground wells/aquifers
- Nitrate poisoning is the leading cause of blue baby syndrome (Methaemoglobinaemia – decreased ability of blood to carry vital oxygen)
- Phosphorus tends to pollute surface waters (rivers, lakes, streams, etc) through runoff
- Nitrogen and phosphorus from agricultural sources contribute to eutrophication, causing harmful algal blooms, fish kills



Risky problem, verifiable outcome, discover value and tradeoffs of new regulations

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 optimization
- 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

Final thought

D ata

I ntegration

S ystems

C omputation

Final thought

D ata

I nnovation

S ystems

C onnections

V isualization

Final thought

D ata

I nnovation

S ystems

C onnections

O utcomes

V erifiable

Final thought

D ata

I nnovation

S ystems

C onnections

O utcomes

V erifiable

E nhancement

R isk taking

Final thought

D ata

I nnovation

S ystems

C onnections

O utcomes

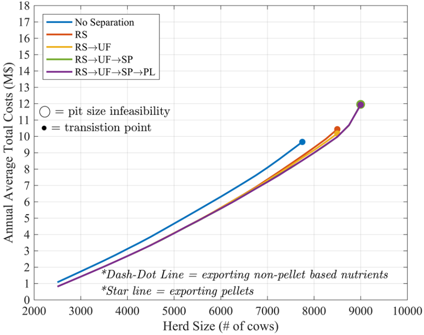
V erifiable

E nhancement

R isk management

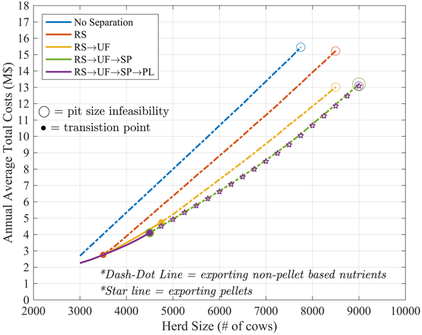
Y do this?

No organic N constraint



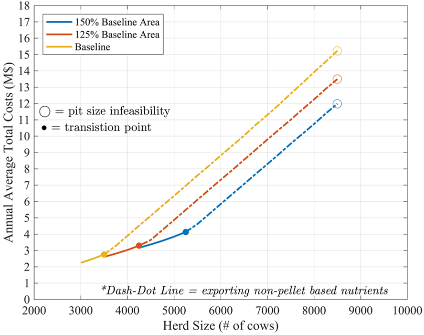
- May compromise sustainability due to nitrogen leaching
- No current regulation on application of organic N – perverse incentive to "dispose of N through over application"

Manure Separation with N constraint



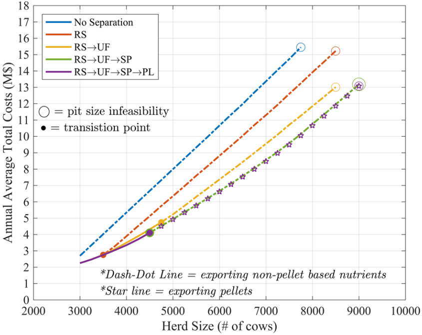
- Separation can enable sustainable nutrient management
- Pellets are valuable in other markets

Rent Land



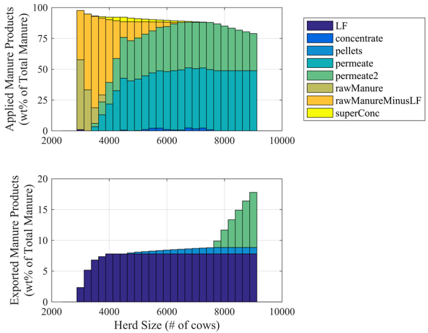
- Renting land would require more organic N to be applied, which may not be available on farm from manure

Manure Separation

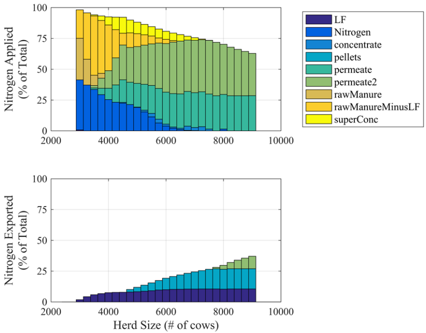


- Advanced separation can enable larger herd sizes per acre of land
- Advanced separation can lead to large economic savings vs. no/single/double stage separation

Separation Products (% weight of total manure)



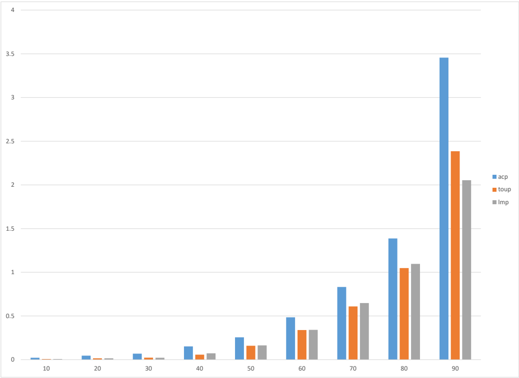
Nitrogen application (fields/export) (% weight of total N)



- Model captures significant detail about products created and applied or exported off farm
- External nitrogen fertilizer purchases essentially go to zero at large herd sizes

Smart Metering Lowers the Cost of Congestion

- average cost pricing (same price wherever and whenever)
- time of use pricing (time variation)
- location marginal pricing (space and time variation)
- connected physical and economic system, integrated data, verifiable outcome



increase in transmission quantities

Health and logistics

- Personalized medicine (use data mining and electronic medical record to provide value of patient centric "access but not own" data)
- Experimental logistics: combinatorial factor screening (software control of robot screening technology. Application to biology and chemistry in both academia and pharmaceutical industries)

Factor Tube Array

GF

1	9
2	10
3	11
4	12
5	13
6	14
7	15
8	16

24 Well Plate Array

24 Well Cell Plate

	1	2	3	4	5	6
A	1	5	9	13	17	21
B	2	6	10	14	18	22
C	3	7	11	15	19	23
D	4	8	12	16	20	24