Modeling for design and impact

Michael C. Ferris

Computer Sciences Department and Wisconsin Institute for Discovery
University of Wisconsin-Madison

University of Auckland, New Zealand
March 27, 2024

- Data science: motivation
- Optimization: basic components and tradeoffs
- Modeling: using constraints to add domain knowledge/structure
- Uncertainty: how to deal with randomness

Descriptive, Predictive, Prescriptive

- Use collected information for reporting

$$y = \phi(a)$$

- Apply models to forecast future events

$$y_j \approx \phi(a_j; x)$$

- Increase sophistication of analysis to evaluate which decisions lead to desired outcomes

$$\min_{(a,y)\in\Omega} v(a, y) \text{ s.t. } y = \phi(a; x)$$

And then add uncertainty...

Successful data analytics: some features

- Financial:
  - standards for interconnectivity (transfers)
- Weather:
  - large scale, real time
  - open source/access
  - no private information (but apps that present information differently)
  - data provider is not the same as user
- Medical
  - shared/private information
  - multiple data types
- Travel:
  - links different types of agents (drivers, riders, administrators)
  - real time, large scale
  - congestion pricing (public/summary information)
  - trips (private information)
  - required (user) inputs to generate specific user outputs

All have reliable acquisition. Need to name things consistently.

1: Data Science

- Extract meaning from data: learning
- Use this knowledge to make predictions: inference
- Optimization provides tools for modeling / formulation / algorithms
- Modeling and domain-specific knowledge is vital in practice: "80% of data analysis is spent on the process of cleaning and preparing the data."

2: Optimization

- There is an objective (function) which we are seeking to maximize or minimize described by:

$$f : S \subseteq \mathbb{R}^n \to \mathbb{R} \cup \{+\infty / -\infty\} =: \bar{\mathbb{R}}$$

- Objective function of variables (or unknowns) $f(x)$ where $x \in \mathbb{R}^n$.
- Variables could be subject to constraints such as $h(x) = 0$.
- The feasible set is described by

$$\Omega = \{x | h(x) = 0,\ g(x) \leq 0\}$$

- This generates a program of the form

$$\min_{x \in \Omega} f(x)$$

- Unconstrained problems have $\Omega = \mathbb{R}^n$ which is the whole space
- What about $\Omega \neq \mathbb{R}^n$?
- Constrained problems can be treated in various ways, including nonlinear, nonconvex problems and convex cones for example.

Four components to optimization

1. Calculus (analysis, probability)
2. Geometry or structure (convexity, polyhedral, discrete)
3. Computation (using linear algebra and sparse tools)
4. Data

- Iterative algorithms generate a series of points which hopefully converge to the solution: issues about well defined (computable), how fast, what they converge to, and how to check properties of the end point.
- Will need all four components; understanding how they link together is important for full command of optimization

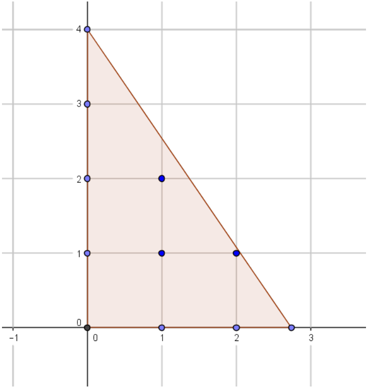
Linear vs. Nonlinear / Stochastic vs Deterministic

- If $f,\ g_i,\ h_i$ are affine then we have a linear program.

$$\begin{aligned} \min \quad & f(x) & \\ \text{s.t.} \quad & g_i(x) & \leq 0 \\ & h_i(x) & = 0 \end{aligned}$$

- Linear problems tend to come from the decision sciences whereas nonlinear problems often arise from physical systems.
- A problem is stochastic if data is not known beforehand. It may arise from some known distribution or assumed via statistical measurements.
- Note the difference between stochastic data and stochastic programs and stochastic algorithms.

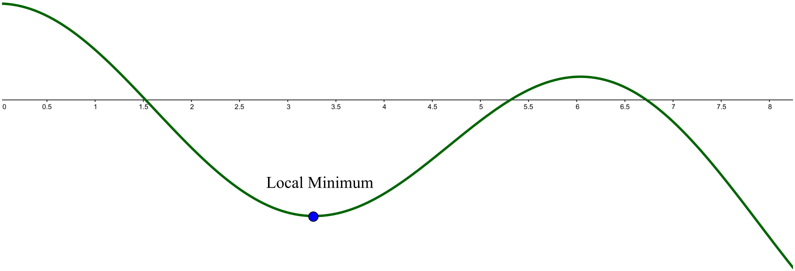
Continuous Vs. Discrete



- In a discrete problem, only the points would be feasible. In a continuous problem, the whole shaded region is feasible.
- Use case: discrete entities, logic

Global vs Local

We use the notion of local and global minimizers



The local minimum is clearly a minimum only within its neighborhood.

Convex functions are ones for which local minimizers are global minimizers.

3: What is meant by a model?

- Many of us build (computer/mathematical) models that capture physics, dynamics, stochastics, discrete choices, and to some extent behavior: collaboration, competition
- Model of system $\phi(a; x)$
- $a = (s, d)$: Actions or designs d affect state s, parameters x
  energy example: state s = electricity flow, actions d = investment/operations, parameters x = loss rate/fuel cost
- Optimization determines model parameters x (based on data - machine learning) (training)
- Can use $\phi(s, d; x)$ to predict state evolution or specfic outcomes
- Validation ensures predictions are good (testing)

Typical Setup

After cleaning and formatting, obtain a data set of m objects:

- Vectors of features: a_j, $j = 1, 2, \ldots, m$
- Outcome / observation / label y_j for each feature vector

The outcomes y_j could be:

- a real number: regression
- a label indicating the a_j lies in one of M classes (for $M \geq 2$): classification. (M can be very large)
- no labels (y_j is null):
  - subspace identification: locate low-dimensional subspaces that approximately contain the (high-dimensions) vectors a_j
  - clustering: partition the a_j into clusters; each cluster groups objects with similar features.

Fundamental Data Analysis

Seek a function ϕ that:

- approximately maps a_j, to y_j for each j; $\phi(a_j) \approx y_j$, $j = 1, 2, \ldots, m$
- satisfies additional properties to make it "plausible" for the application, robust to perturbations in the data, generalizable to other data samples from the same distribution.

Can usually define ϕ in terms of some parameter vector x - thus identification of ϕ becomes a data-fitting problem:

- Find a nice x such that $\phi(a_j; x) \approx y_j$ for $j = 1, 2, \ldots, m$
- Objective function in this problem often built up of m terms that capture mismatch between predictions and observations for data item (a_j, y_j)
- The process of finding ϕ is called learning or training.

What's the use of the mapping ϕ?

- Prediction: Given new data vector a_k, predict outputs $y_k \leftarrow \phi(a_k; x)$.
- Analysis: ϕ (more particularly the parameter x) reveals structure in the data

Many possible complications:

- Noise or errors in a_j and y_j
- Missing data:
- Overfitting: ϕ exactly fits the set of training data (a_j, y_j) but predicts poorly on "out-of-sample" data (a_k, y_k)

ML models in practice

- Regression: $\phi(a_j; x) = a_j^T x$.

$$\min_x f(x) := \frac{1}{2} \sum_{j=1}^{m} (a_j^T x - y_j)^2$$

- Add $\ell_2 = \|x\|^2$ reduces sensitivity to noise in y
- Add $\ell_1 = \|x\|_1$ yields solutions x with few non-zeros (Feature selection)
- loss function $+\lambda * R(x)$
- Sparse PCA.
- Linear Support Vector Machines (kernel SVM)
- Logistic Regression
- Deep learning

All of these models can be augmented by domain specific knowledge, leading to nonlinear and/or constrained optimization

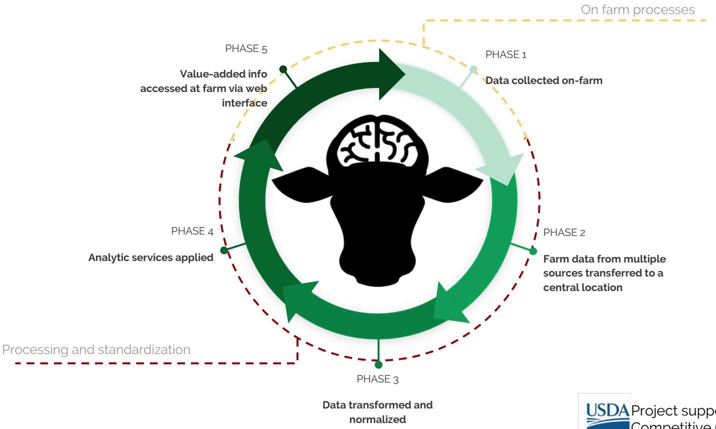
What does optimization add?

- Value of outcome $v(s, d)$ (e.g s electricity flows in network, d capacity expansion, v is operation profit)
- How to use model to suggest good actions/designs (s, d)?
- Constrained optimization chooses (feasible) actions to maximize value

$$\max_{(s,d)\in\Omega} v(s, d) \text{ s.t. } \phi(s, d; x)$$

- Optimization can be hard to solve (non-convex)
- Models can be complex and difficult to explain, often ignored by decision makers, yet their solution can lead to fundamentally new insights
- Simple rules (policies) $d = \pi(s)$, reduce complexity of optimization, enhance explainability

Dairy Brain - a continuous decision aiding engine



DEPARTMENT OF DAIRY SCIENCE University of Wisconsin-Madison

MORGRIDGE INSTITUTE FOR RESEARCH

WISCONSIN INSTITUTE FOR DISCOVERY AT THE UNIVERSITY OF WISCONSIN-MADISON

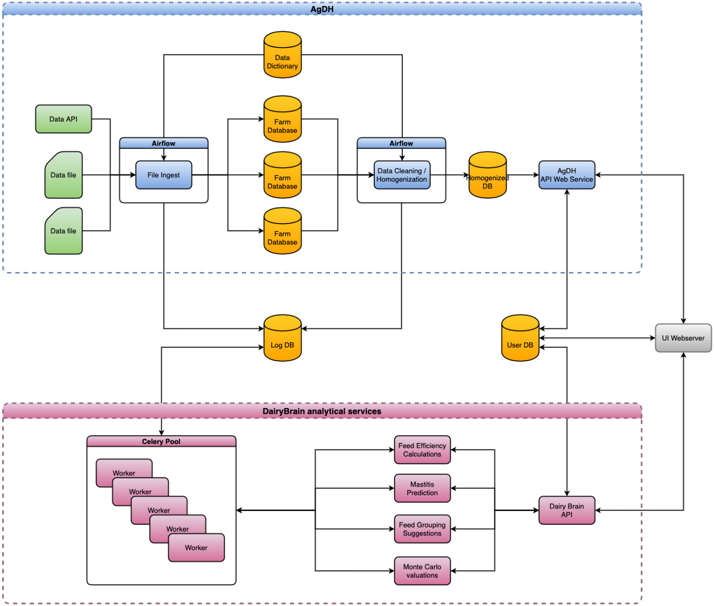
COMPUTER SCIENCES

WISCONSIN UNIVERSITY OF WISCONSIN-MADISON

USDA Project supported by NIFA-AFRI-FACT Competitive **Grant No. 2018-09256**

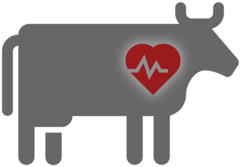
- Translate research outcomes to practical applications
- Provide access to analytical services to enhance operations

Application Programming Interface (API) design



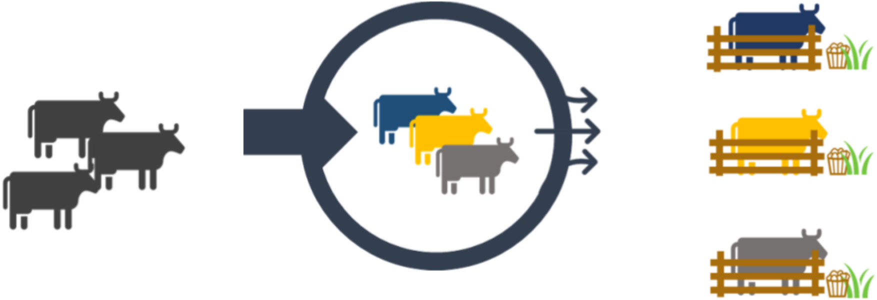
Cow health

- Early ketosis identification
- Early Prediction of Clinical Mastitis
- Monitoring the Risk of CM for 1st Lactation Heifers

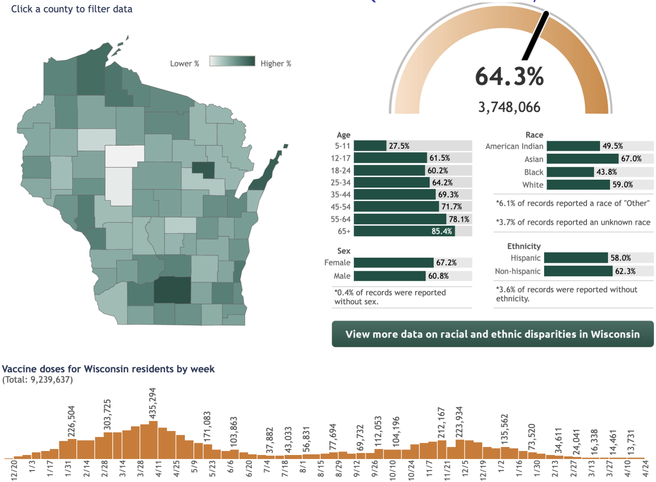


Nutritional grouping

Group of cows Cluster cows Differentiated diet



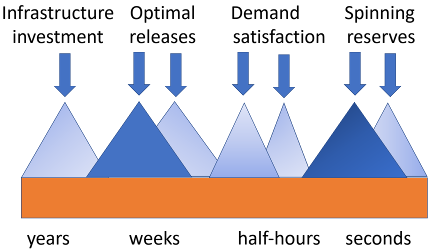
Covid-19 vaccine allocation (with DHS/National Guard)



- Two phase optimization(first phase: fair allocation, second phase: logistics)
- `https://www.dhs.wisconsin.gov/covid-19/vaccine-data.htm`

4: Planning models have uncertainty at various time scales

- Demand growth, technology change, capital costs are long-term uncertainties (years)
- Seasonal inflows to hydroelectric reservoirs are medium-term uncertainties (weeks)
- Levels of wind and solar generation are short-term uncertainties (half hours)
- Very short term effects from random variation in renewables and plant failures (seconds)



- Tradeoff: Uncertainty, cost and operability, regulations, security/robustness/resilience
- Needs modelling at finer time scales

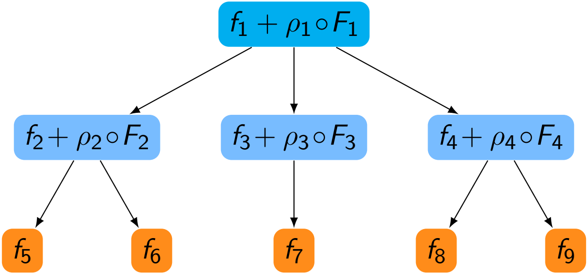
Simplified two-stage stochastic optimization model

- Investment decisions are z at cost $K(z)$
- Operating decisions are: generation y at cost $C(y)$, loadshedding q at cost Vq.
- Random demand is $d(\omega)$.
- Minimize capital cost plus expected operating cost:

$$
\begin{aligned}
\text{P:} \quad \min_{z,y,q \in X} \quad & K(z) \; + \; \mathbb{E}_\omega[C(y(\omega)) + Vq(\omega)] \\
\text{s.t.} \quad & y(\omega) \le z, \\
& y(\omega) \ge d(\omega) - q(\omega), \\
& z_{\mathcal{N}} \le (1-\theta) z_{\mathcal{N}}(2017)
\end{aligned}
$$

Who do you have on your bench, what reserves are in your plan?

Scenario tree with nodes $\mathcal{N} = \{1, 2, \ldots, 9\}$, and $T = 3$



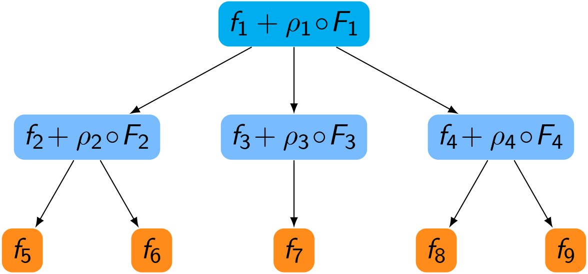
At leaf nodes:

$$\min_{x_{a\ell} \in \mathcal{X}_{a\ell}} \leftarrow f_{a\ell}(x_{a\ell}; x_{\text{-}a\ell}, \pi_\ell) \quad \forall a \in \mathcal{A},$$

$$0 \in H_\ell(\pi_\ell; x_{\cdot\ell}) + N_{P_\ell}(\pi_\ell)$$

";" separates variables from parameters in function definition

Stochastic equilibrium (MOPEC)



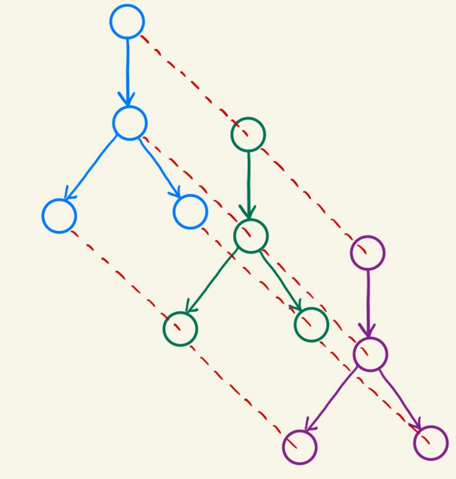
Agents solve problem at root node, linking at all nodes:

$$\min_{x_{a\cdot} \in \mathcal{X}_{a0}} f_{a1}(x_{a1}; x_{-a1}, \pi_1)$$

$$+ \rho_{a1}\left(\left[f_{aj}(x_{aj}; x_{-aj}, \pi_j) + \rho_{aj}\left(\left[f_{a\ell}(x_{a\ell}; x_{-a\ell}, \pi_\ell)\right]_{\ell \in j_+}\right)\right]_{j \in 1_+}\right) \quad \forall a \in \mathcal{A},$$

$$0 \in H_j(\pi_j; x_{\cdot j}) + N_{P_j}(\pi_j), \qquad \forall j \in \mathcal{T}.$$

Scenario trees linked across agents



- Dynamics link over time
- Complementarity links nodes of scenario tree across agents

Three sources of difficulty:

1. Size: number of scenarios, agents, details
2. Non-convexity: Nash behavior
3. Risk aversion: Nonsmooth or Nonlinear (product of probabilities)

Risk Measures: example of structure

Problem type

Objective function or Constraint

$$\min_{x \in X} \theta(x) + \rho(F(x)) \qquad \min_{x \in X} \theta(x) \text{ s.t. } \rho(F(x)) \leq \alpha$$

- Dual representation (of coherent r.m.) in terms of risk sets

$$\rho(Z) = \sup_{y \in \mathcal{D}} \mathbb{E}_y[Z]$$

- If $\mathcal{D} = \{p\}$ then $\rho(Z) = \mathbb{E}[Z]$
- If $\mathcal{D}_{\alpha,p} = \{y \in [0, p/(1-\alpha)] : \langle \mathbf{1}, y \rangle = 1\}$, then $\rho(Z) = \overline{CVaR}_\alpha(Z)$
- Combinations - increasing risk aversion as λ increases

$$\rho(Z) = (1-\lambda)\mathbb{E}[Z] + \lambda \overline{CVaR}_\alpha(Z)$$

The transformation to complementarity

$$\min_{x \in X} \theta(x) + \rho(F(x))$$

$$\text{where } \rho(u) = \sup_{y \in \mathcal{D}} \left\{ \langle y, u \rangle - \frac{1}{2} \langle y, My \rangle \right\}$$

optimality condition:

$$0 \in \partial\theta(x) + \nabla F(x)^T \partial\rho(F(x)) + N_X(x)$$

calculus:

$$0 \in \partial\theta(x) + \nabla F(x)^T y + N_X(x)$$

$$0 \in -y + \partial\rho(F(x)) \iff 0 \in -F(x) + My + N_{\mathcal{D}}(y)$$

- This is a complementarity problem: opt conds in x coupled with opt conds in y - separated

Stochastic Equilibrium as (extended) MOPEC

$$\min_{x_{a\cdot} \in \mathcal{X}_{a0}} f_{a1}(x_{a1}; x_{-a1}, \pi_1) + \sum_{j \in 1_+} y_{aj} \left(f_{aj}(x_{aj}; x_{-aj}, \pi_j) + \sum_{\ell \in j_+} y_{a\ell} f_{a\ell}(x_{a\ell}; x_{-a\ell}, \pi_\ell) \right), \quad \forall a \in \mathcal{A} \tag{1}$$

$$0 \in H_j(\pi_j; x_{\cdot j}) + N_{P_j}(\pi_j), \quad \forall j \in \mathcal{T} \tag{2}$$

$$\begin{aligned} r_{a1}(x,\pi) &= \max_{y_{a1_+} \in \mathcal{D}_{a1}} \sum_{j \in 1_+} y_{aj}(f_{aj}(x_{aj}; x_{-aj}, \pi_j) + r_{aj}(x,\pi)) \\ r_{a2}(x,\pi) &= \max_{y_{a2_+} \in \mathcal{D}_{a2}} \sum_{\ell \in 2_+} y_{a\ell} f_{a\ell}(x_{a\ell}; x_{-a\ell}, \pi_\ell) \\ r_{a3}(x,\pi) &= \max_{y_{a3_+} \in \mathcal{D}_{a3}} \sum_{\ell \in 3_+} y_{a\ell} f_{a\ell}(x_{a\ell}; x_{-a\ell}, \pi_\ell) \\ r_{a4}(x,\pi) &= \max_{y_{a4_+} \in \mathcal{D}_{a4}} \sum_{\ell \in 4_+} y_{a\ell} f_{a\ell}(x_{a\ell}; x_{-a\ell}, \pi_\ell) \end{aligned} \tag{3}$$

Algorithms and problems

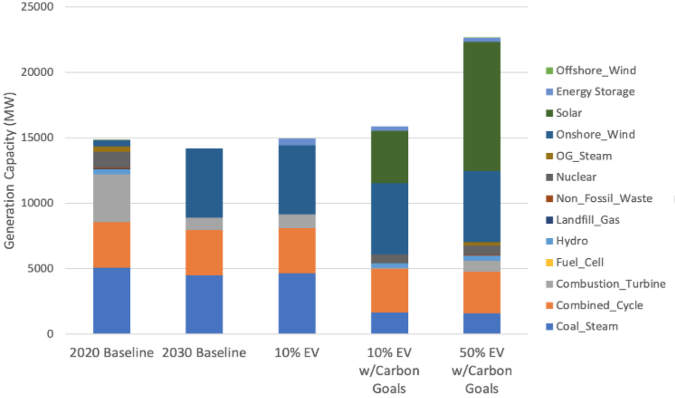
- PATH: nonsmooth Newton method (defaults) (blue+red+black)
- PD (Primal-dual): iteratively blue+red then black
- PD-PTH (Primal-dual + PATH)
- PD-CC-PTH (Primal-dual + convex-comb(black) + PATH)
- Homot(λ) + Primal-dual + convex-comb(black) + PATH
- Multistage economic dispatch, capacity expansion, hydroelectric system
- 3 types of demand formulation (I,II and III)
- Two scenario trees (4 stages, 40 nodes) and (4 stages, 156 nodes)
- 32 data instances for each formulation
- Several modulus of convexity and risk aversion parameters

$$\rho(Z) = (1-\lambda)\mathbb{E}[Z] + \lambda\overline{CVaR}_\alpha(Z)$$

Dispatch example, large tree, type I

quad	λ	PATH	PD	PD-PTH	PD-CC-PTH	Homotopy
0	0.1	0.0	0.0	59.4	100.0	100.0
0	0.3	0.0	0.0	12.5	96.9	100.0
0	0.5	0.0	0.0	9.4	71.9	87.5
0	0.7	0.0	0.0	3.1	18.8	53.125
0	0.9	0.0	0.0	0.0	9.4	21.875
1e-2	0.1	28.1	15.6	100.0	100.0	100.0
1e-2	0.3	0.0	0.0	90.6	100.0	100.0
1e-2	0.5	0.0	0.0	40.6	100.0	100.0
1e-2	0.7	0.0	0.0	21.9	84.4	93.8
1e-2	0.9	0.0	0.0	6.2	53.1	68.75
1e-1	0.1	0.0	59.4	100.0	100.0	100.0
1e-1	0.3	0.0	43.8	100.0	100.0	100.0
1e-1	0.5	0.0	18.8	96.9	100.0	100.0
1e-1	0.7	0.0	12.5	100.0	100.0	100.0
1e-1	0.9	0.0	15.6	93.8	100.0	100.0

Impact of Electric Vehicles on Generator Investments



- Carbon Goals: 60% reduction on in-state carbon emissions
- Nuclear (low-carbon) used
- Coal steam generators shut down, supplanted by renewables
- Additional 180,000 MWh demand for EVs
- Storage investment needed
- Additional demand or carbon goals give more dramatic effects

Conclusions

Value of (constrained) optimization

- Constraints can capture domain knowledge much more than a single objective
- Machine learning can be used to inform models
- Informed strategic decisions and tradeoffs
- Horses for courses: simple policies are effective
- Facility location: where to locate reserves, agents, sizing
- Disaster recovery: hedging risk, promoting flexibility, dynamics, windows and staging
- Risk models: not all outcomes are equally bad, trade risk

Truth is in the details

- Thanks to collaborators: Andy Philpott, Olivier Huber, Jiajie Shen, Steve Wangen, Kristine Palmer, Adam Christensen, Victor Cabrera

Issues regarding what to do for who?

- Policy or individual farm?
- Operational (precision) or strategic?
- When are decisions made: yearly, seasonal, daily, hourly?
- Inform human-in-the-loop decision making
- Ownership: whose data is it, after change/cleaning
- Privacy: who can see what and when
- Scale: the big data issue
- Missing data

A mathematical modelling approach to planning

- Build and solve a social plannning model that optimizes electricity capacity investment with constraints on CO2 emissions.
- Social planning solution should be stochastic: i.e. account for future uncertainty
- Social planning solution should be risk-averse: because the industry is.
- Approximate the outcomes of the social plan by a competitive equilibrium with risk-averse investors.
- Compensate for market failures from imperfect competition or incomplete markets.

Implementation details

- Use optimization modeling system, api's to sophisticated solvers (Cplex, Gurobi, Mosek, Conopt, Knitro)
  1. Aggregate: build a (rigorous) approximation of underlying physics and stochastics to generate a "system model"
  2. Solve: Use simple approximation to detemine key design, incorporate (some level of) operation
  3. Validate/Visualize: Use detailed model evaluations to verify operations are effective
  4. Rinse and repeat
- Key use of constraints to modify solutions, capture appropriate detail
- Address issues of risk and uncertainty
- Data driven hybrid approach - model based learning by interaction
- Extensions: sequential decision making (multiple time steps with dynamic model updates).