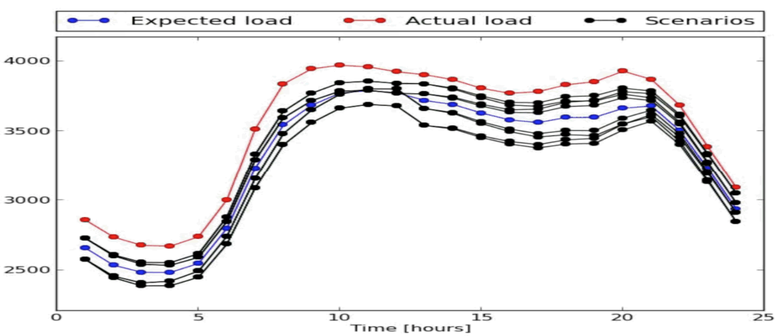
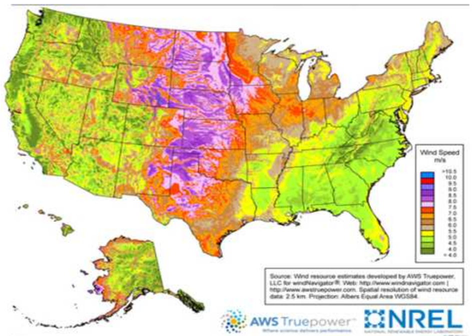
Exceptional service in the national interest

SAND2016-2564C

Sandia National Laboratories





(Some) Directions for Impactful Research at the Intersection of Data Science and the Power Grid

Jean-Paul Watson

Discrete Math and Optimization Department

Sandia National Laboratories

CCR Center for Computing Research

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Thanks To My Collaborators…

Sandia National Laboratories

Francisco Munoz
U. Adolfo-Ibanez

David Woodruff
UC Davis

Ross
Guttromson
Sandia

Andrea Staid
Sandia

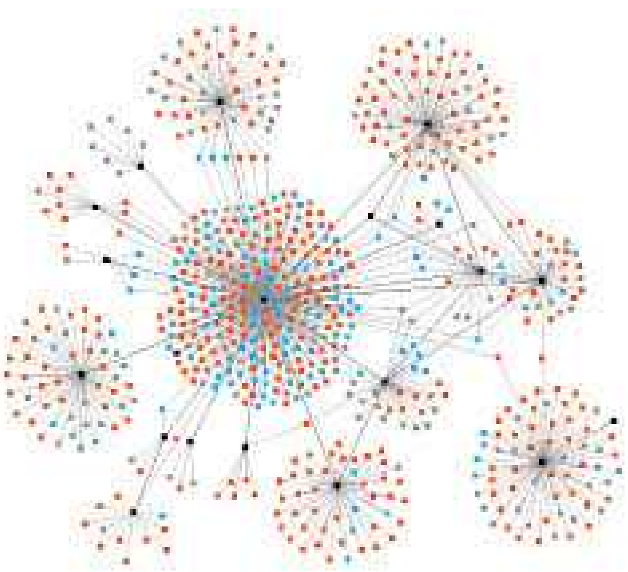
Sarah Ryan
Iowa State

Roger Wets
UC Davis

Cesar Silva Monroy
Sandia

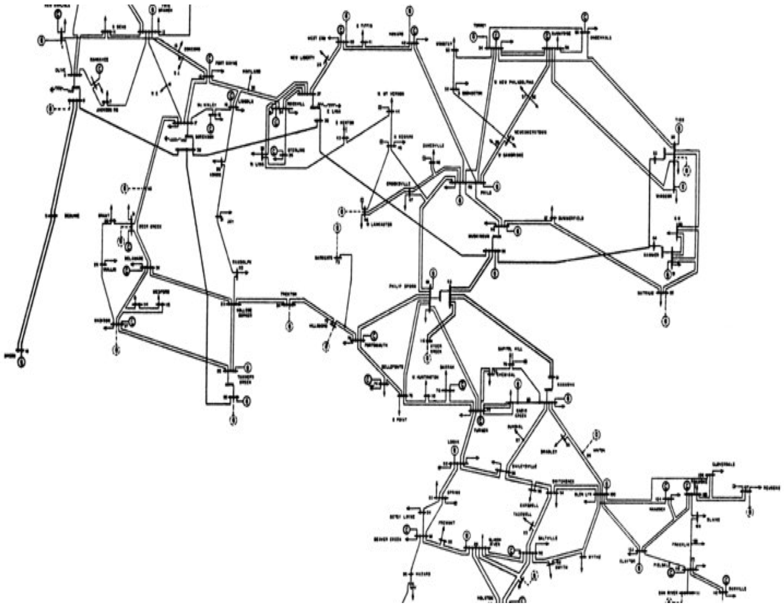
The Power Grid and Big Data (1)

A Typical Small-World Graph



Modified IEEE 118 Bus Test Case System

http://motor.ece.iit.edu/data/ltscuc



Some key differences between typical data science graphs and power grid "graphs"

- One is emergent, the other is engineered – very different structures
- In the power grid
  - Physical laws (e.g., KCL and KVL) dominate behavior…
  - … as do operational constraints on physical equipment (e.g., generators)

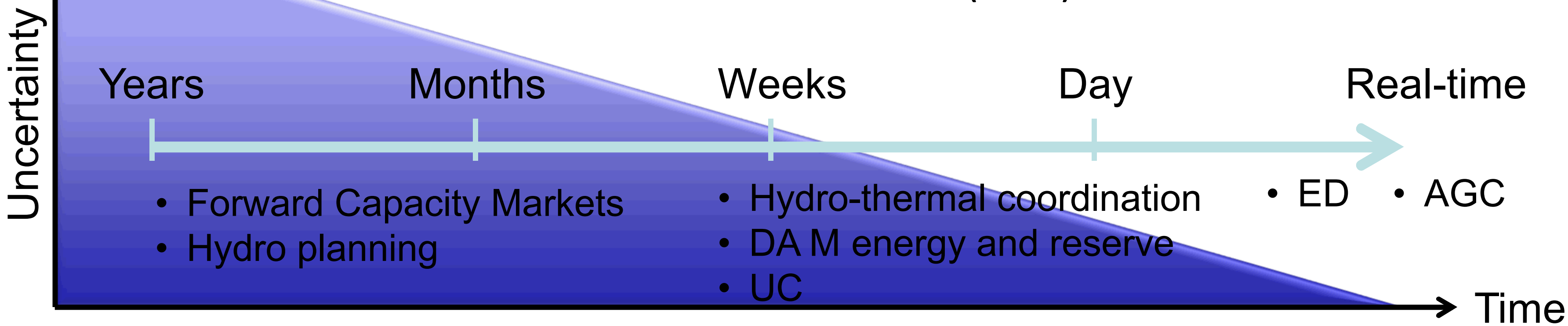
The Power Grid and Big Data (2)

Sandia National Laboratories

- But the biggest differentiator of power grid problems in data science is
  - *They are intimately and almost without exception linked with <u>formal</u>, <u>mature</u>, and <u>existing</u> decision problems that are used to for operations (universally) and planning (less so)*

Power System Planning/Operations

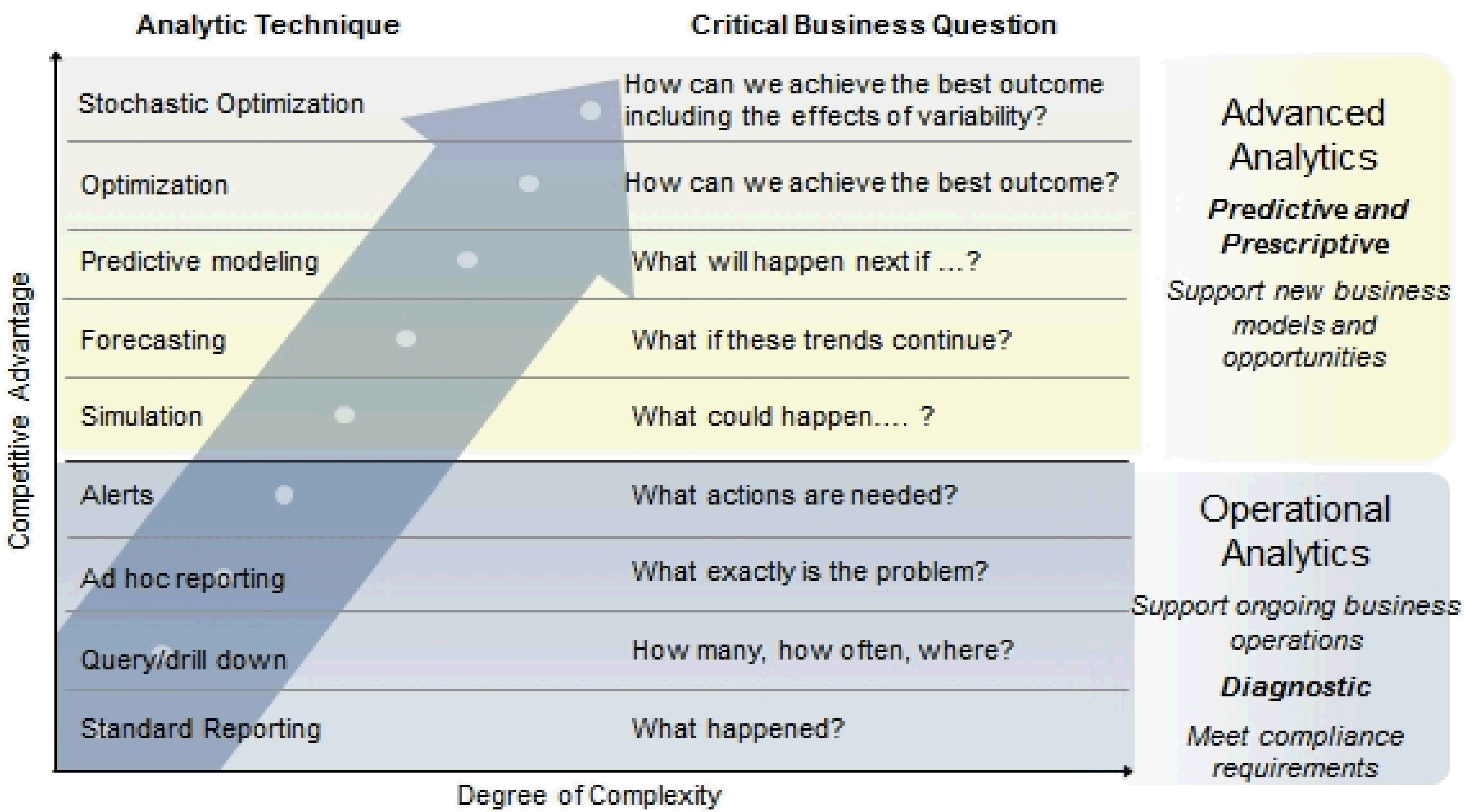
- Decision making in power systems looks at processes ranging from very large time constants to near real-time
  - **Years, Seasons, Months, Weeks**: Resource adequacy, transmission and hydro resource planning.
  - **Days**: Hydro-thermal coordination, day-ahead UC of energy and reserves, intra-day UC.
  - **Hours**: intra-day look-ahead processes, dynamic economic dispatch.
  - **Minutes**: Economic Dispatch (ED).
  - **Seconds**: Automatic Generation Control (AGC).



Key Talk Objectives

- Illustrate the tight linkage between data science and formal power grid operations and planning problems
  - Particularly in the prescience of weather/climate and renewables production uncertainty
- Provide an overview of stochastic optimization, and argue that data science has a major role to play in the solution of real-world stochastic optimization problems for the power grid
- Illustrate these two points through 4 case studies, spanning the range of decision time-scales from minutes-ahead to years-ahead

Data Science, Statistics, and Stochastic Optimization: One View



Taken from Davenport and Harris (2007)

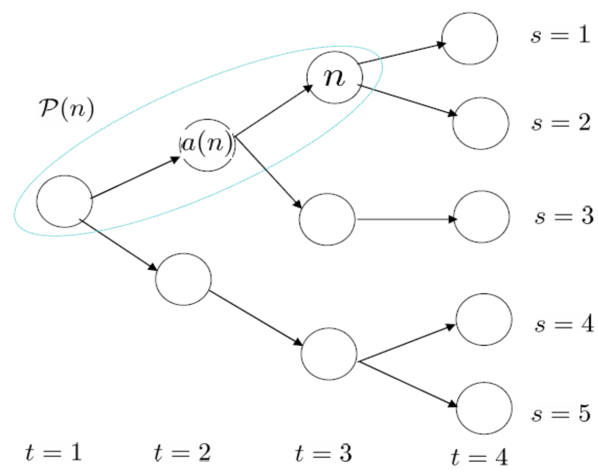
Algebraic Optimization: Deterministic and Stochastic

- Deterministic Mixed-Integer Programming (MIP)
  - The workhorse of (rigorous) Operations Research

$$
\begin{aligned}
\min \quad & c'x + h'y \\
\text{s.t.} \quad & Ax + By \le b \\
& x \in Z^n_+ \,(x \ge 0,\ x \text{ integer}) \\
& y \in R^n_+ \,(y \ge 0)
\end{aligned}
$$

  - Approximable for most real-world problems (NP-Hard)
- Stochastic Mixed-Integer Programming (SMIP)
  - SMIP = MIP + _uncertainty_ + _recourse_

$$
\begin{aligned}
\min \quad & f(\mathbf{x}) = \mathbf{c}^T\mathbf{x} + \mathbb{E}[Q(\mathbf{x},\omega)] \\
\text{s.t.} \quad & A\mathbf{x} \ge \mathbf{b}, \quad \mathbf{x} \in \mathbb{R}^{n_1-p_1}_+ \times \mathbb{Z}^{p_1}_+ \\
Q(\mathbf{x},\omega) = \quad & \min \quad \mathbf{q}(\omega)^T\mathbf{y} \\
& \text{s.t.} \quad W\mathbf{y} \ge \mathbf{h}(\omega) - T(\omega)\mathbf{x} \\
& \mathbf{y} \in \mathbb{R}^{n_2-p_2}_+ \times \mathbb{Z}^{p_2}_+
\end{aligned}
$$



  - Still NP-Hard, but far more difficult than MIP in practice

Why Stochastic (as Opposed to Deterministic) Optimization?

A Motivating Example…

Courtesy Roger J-.B. Wets

Production Planning for a Furniture Warehouse

- A furniture manufacturer must choose how many desks of type to manufacture so as to maximize profit

$$\sum_{j=1}^{4} c_j x_j = 12x_1 + 25x_2 + 21x_3 + 40x_4 = \langle c, x \rangle \qquad x_j \geq 0,$$

- The constraints

$$Tx \leq d$$

$$t_{cj} \sim (t_{fj}) \text{ carpentry (finishing) man-hours: for dresser } j$$

$$d_c \ (d_f) = \text{total time availalble for carpentry (finishing)}$$

$$\begin{aligned} t_{c1}x_1 + t_{c2}x_2 + t_{c3}x_3 + t_{c4}x_4 &\leq d_c \\ t_{f1}x_1 + t_{f2}x_2 + t_{f3}x_3 + t_{f4}x_4 &\leq d_f, \end{aligned} \qquad \begin{pmatrix} 4 & 9 & 7 & 10 \\ 1 & 1 & 3 & 40 \end{pmatrix} x \leq \begin{pmatrix} 6000 \\ 4000 \end{pmatrix}$$

- Optimal:

x^d = (1.333, 0, 0, 67)

Thinking More Realistically (and Stochastically)…

- Induce very small perturbations on key parameters
  - Time required, per desk, for carpentry and finishing
  - Total man-hour availability for carpentry and finishing
  - Consider expected value +- 5% deviation

- 10 random parameters with each 4 possible values = 1,048,576 possible pairs

- => A very large scale (> $2 \cdot 10^6$ variables and constraints) optimization problem

How is the stochastic solution different from deterministic solution?

- Stochastic programming model (a large LP)
- Optimal: $x^* = (257,\ 0,\ 665,\ 34)$ expected profit: $\$18,051.-$
- Solution is robust: It considered all 10^6 possibilities
- Recall: $x^d = (1.333, 0, 0, 67)$ – expected profit = $16,942
- x^d is not close to optimal (-6.5%)
- x^d *isn't pointing in the right direction*

Vignette #1:

Predicting Climate Change Impacts on Hurricane-Related Power Outages

Courtesy: Andrea Staid

Background: Power Outage Prediction

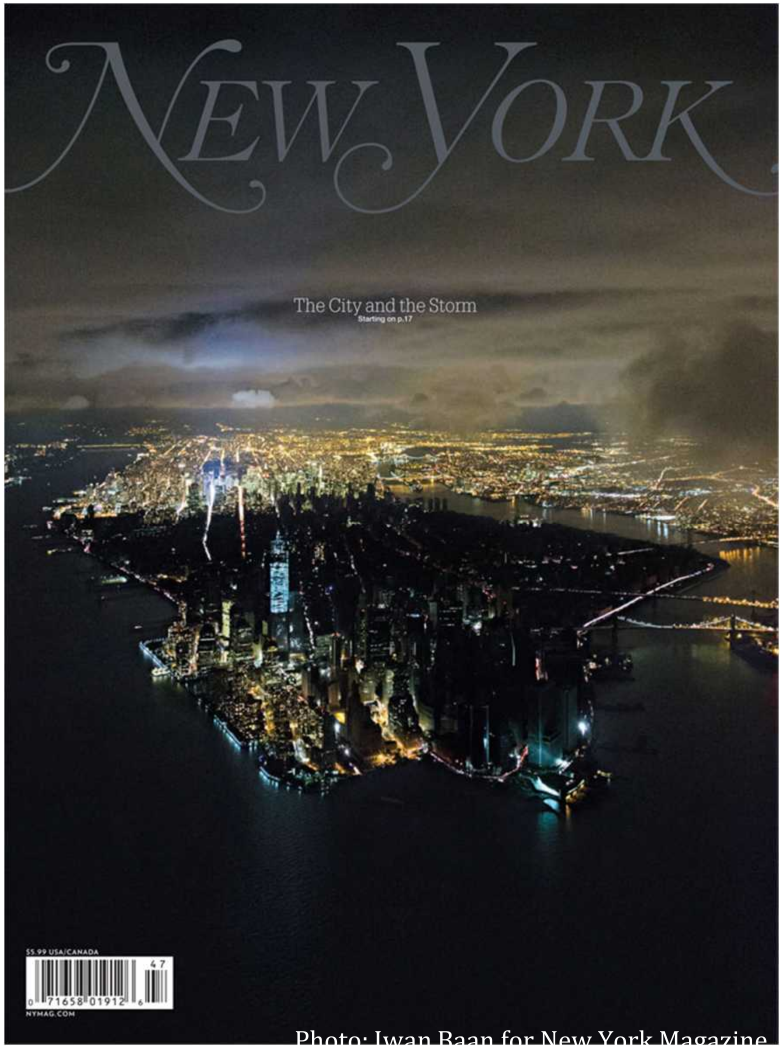


Photo: Iwan Baan for New York Magazine

- Anticipating number and location of outages can reduce duration
- Ideally: Real-time predictions using storm track projections
- Allows utilities and emergency response agencies to plan ahead

Anticipating Power Outages

Typical Utility Response Cycle:

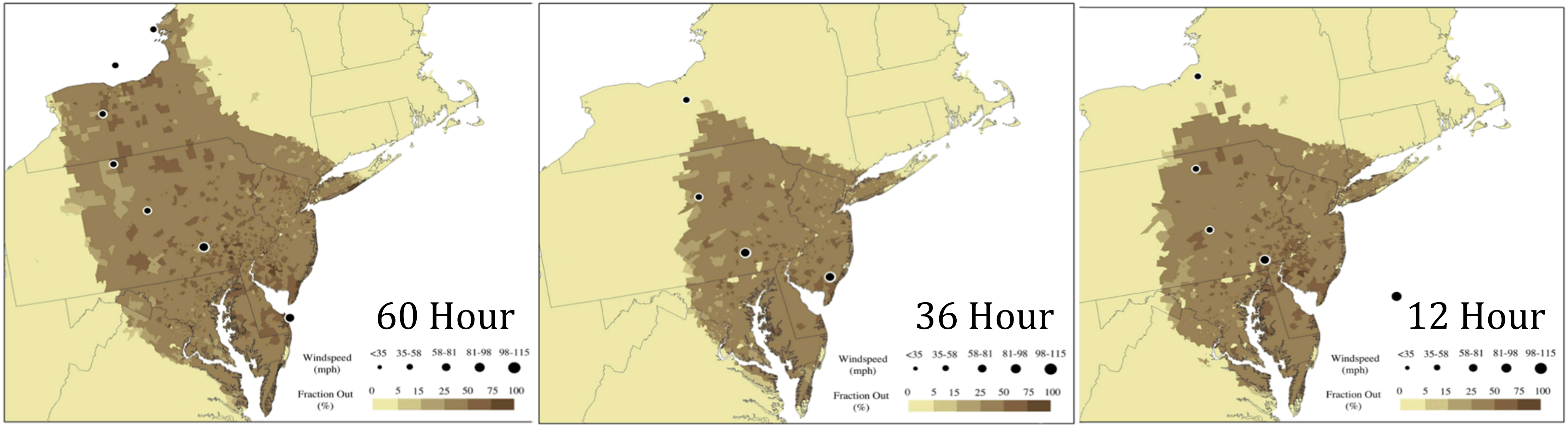
Hurricane Progress	Off-coast, outer bands begin contact	Rain begins for large storms, wind increases	Strongest winds, heaviest rain, highest surge	Wind, rain, surge gradually diminish
Time to Landfall	T - ~5 days	T - ~2-3 days	**T=0** T + 0-1 day	T > 1 day
Utility Activities & Decisions	Monitor track, begin estimating crew needs, begin crew requests	Finalize crew requests, position materials and crews	Begin repairs when safe	Conduct repairs according to prioritization plan

The (Now Prior) State-of-the-Art

Sandia National Laboratories

Outage Prediction Model

- Run in real-time for oncoming storms
- First Version – Trained with utility-specific data
- Second Version – Uses only publicly available data
  - Generalized for other areas using cross-validation to ensure accuracy
  - Can be applied for all areas of Gulf and Atlantic coasts



This Leads to the Question…

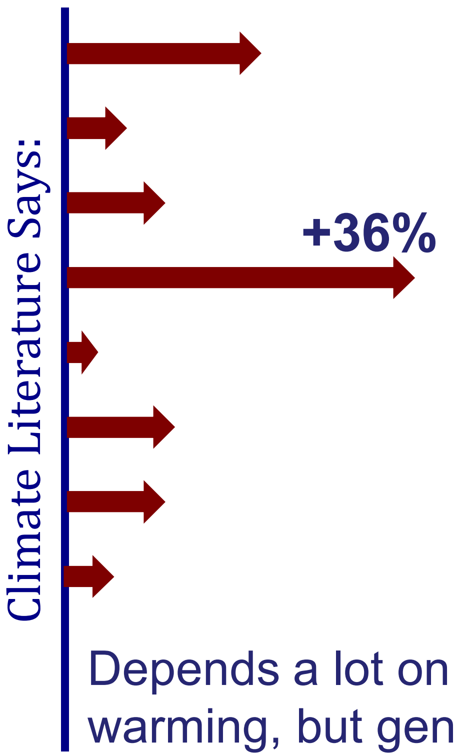
- If we can predict power outages today, can we also predict them for future storms?
  - What will future storms look like?
  - Essential to consider climate change
- Can we anticipate future risks to our power system?

Uncertain Climate – Hurricane Link

Sandia National Laboratories

How will climate change affect North Atlantic tropical cyclones?

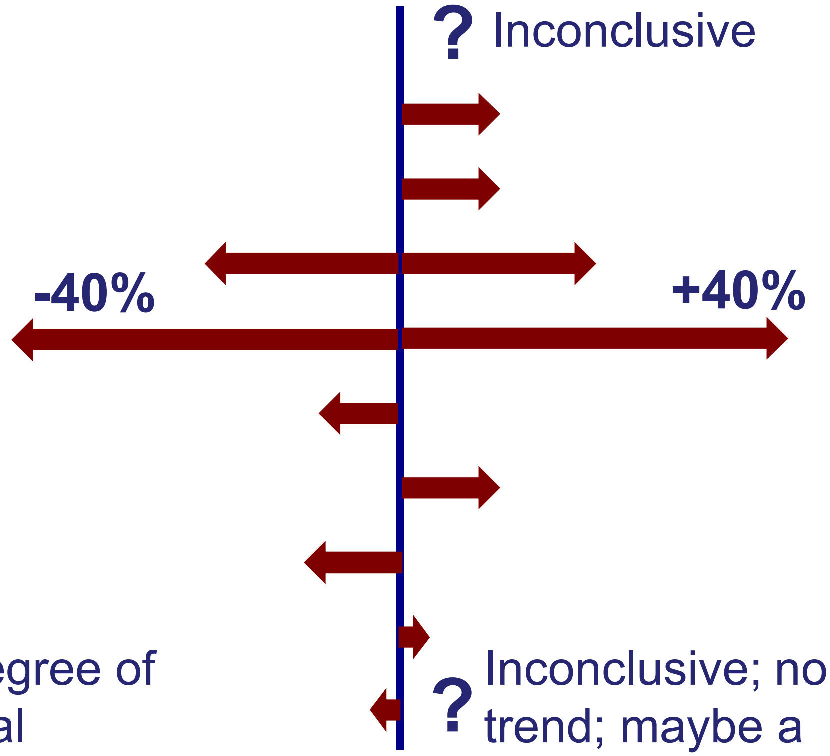
Intensity?



Depends a lot on degree of warming, but general consensus on increasing trend

Frequency?

? Inconclusive



? Inconclusive; no trend; maybe a decrease?
No consensus in the literature

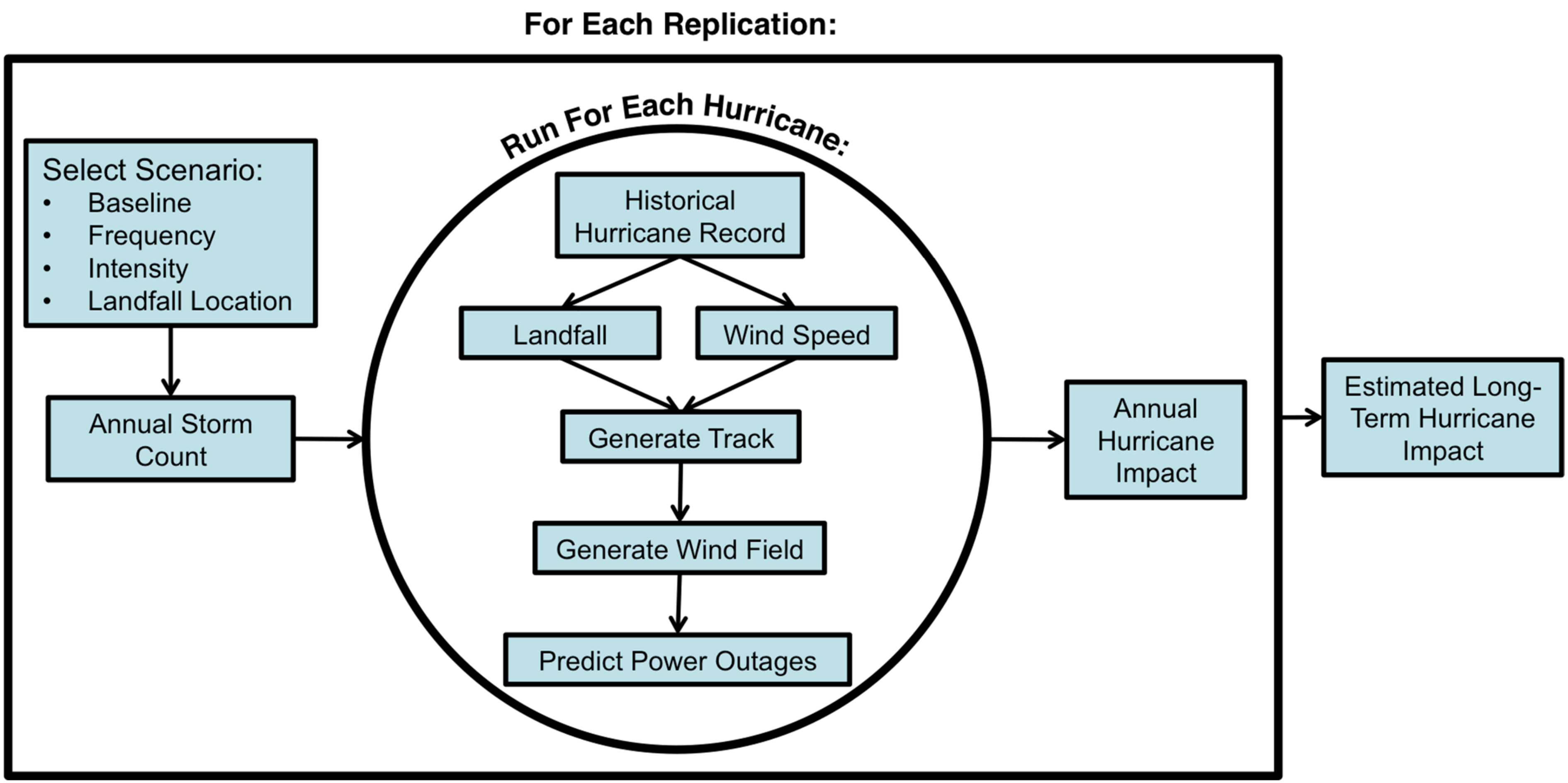
Location?

?

Little to no change; dependent on atmospheric circulation; possible increases in Gulf or West Atlantic

Stochastic Simulation Structure

Sandia National Laboratories

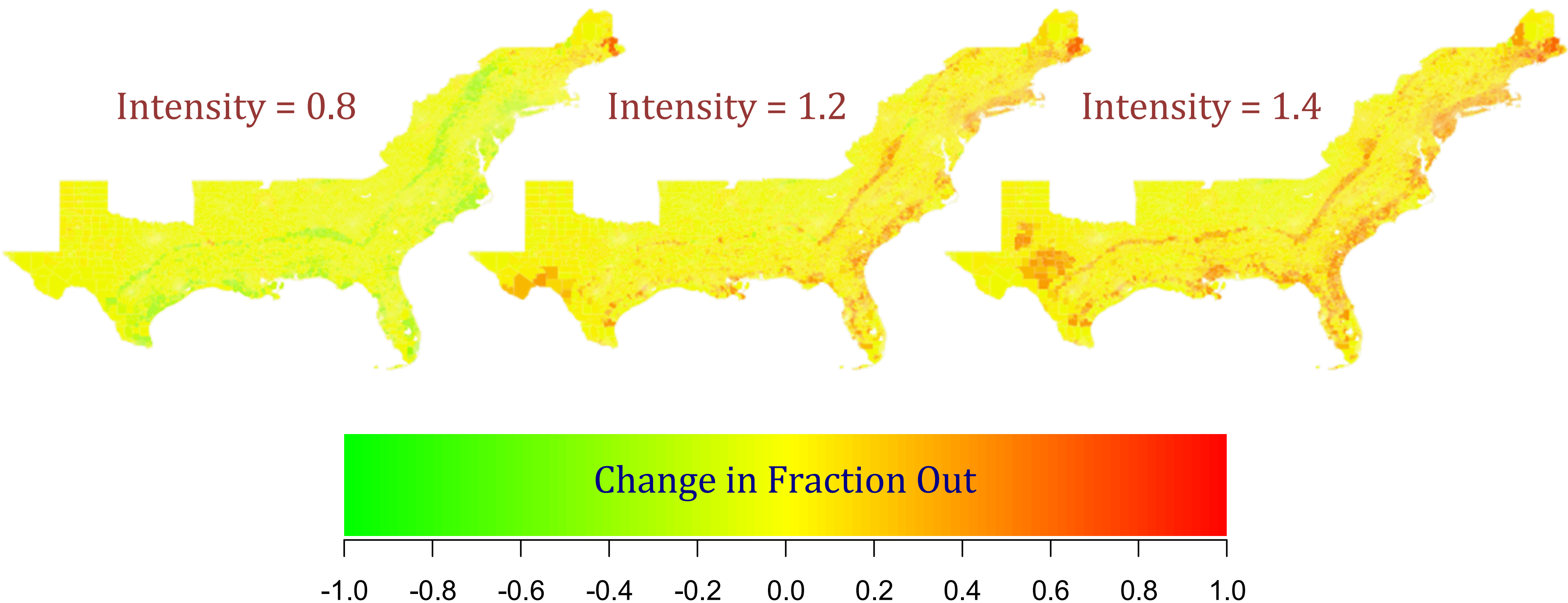


Repeat to reach convergence of the 99th percentile within 1%

100-Year Fraction without Power

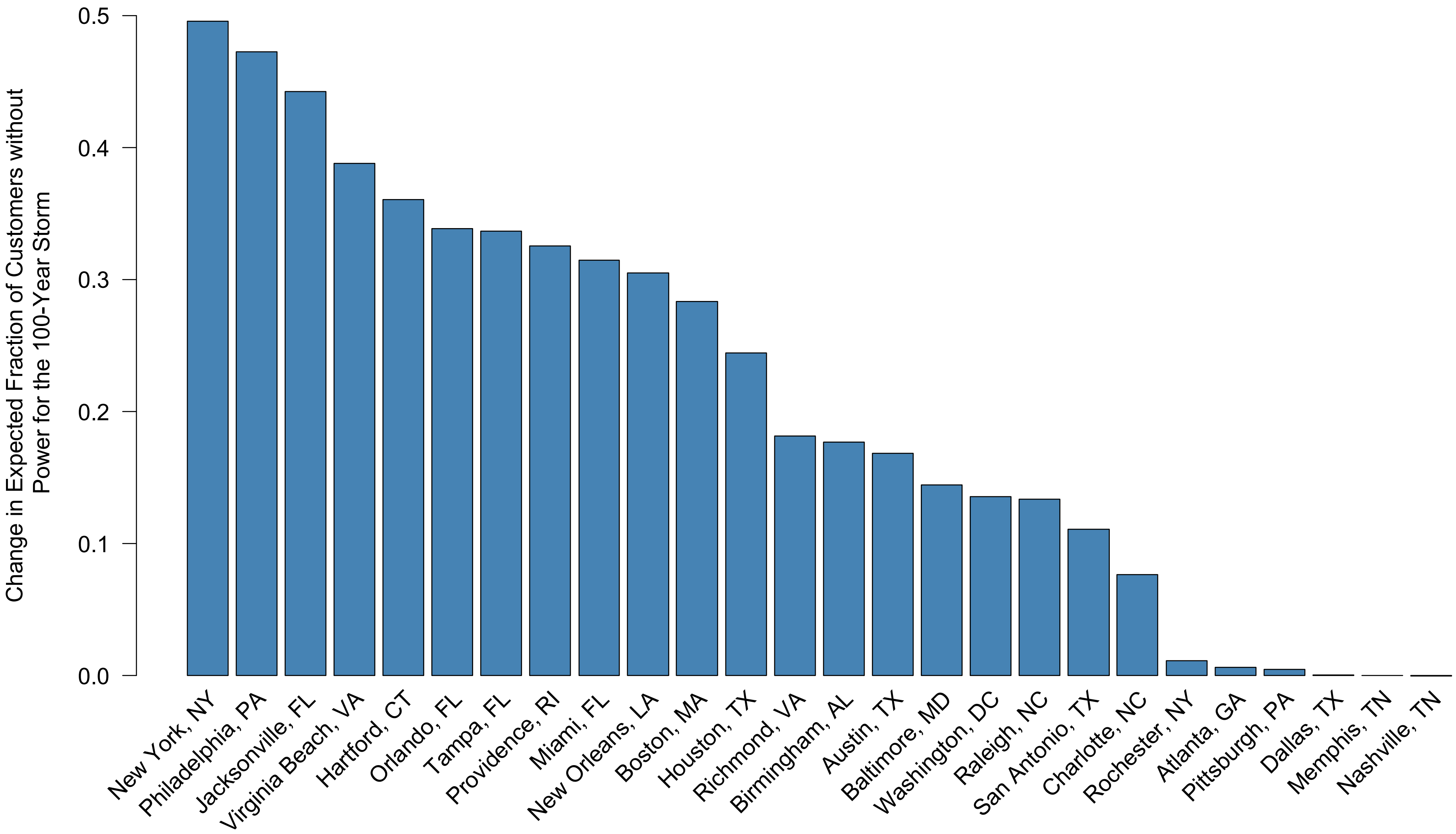
Sandia National Laboratories

Plotting **Difference** From Baseline:



*Independent samples result in sample error in very low population density areas

Sensitivity to Hurricane Intensity

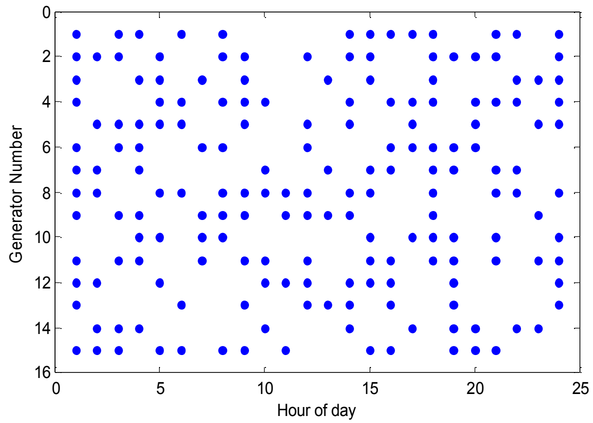


Vignette # 2:

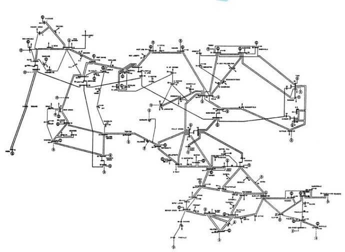
Stochastic Unit Commitment

The General Structure of a Stochastic Unit Commitment Optimization Model

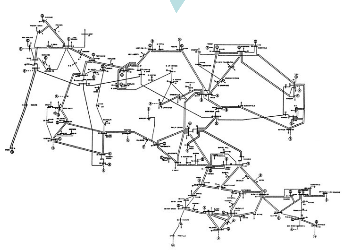
Objective: Minimize expected cost



p_1 p_2 … p_N

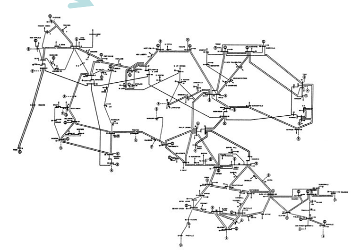


Scenario 1



Scenario 2

…



Scenario N

First stage variables:

- Unit On / Off

Nature resolves uncertainty

- Load
- Renewables output
- Forced outages

Second stage variables (*per time period*):

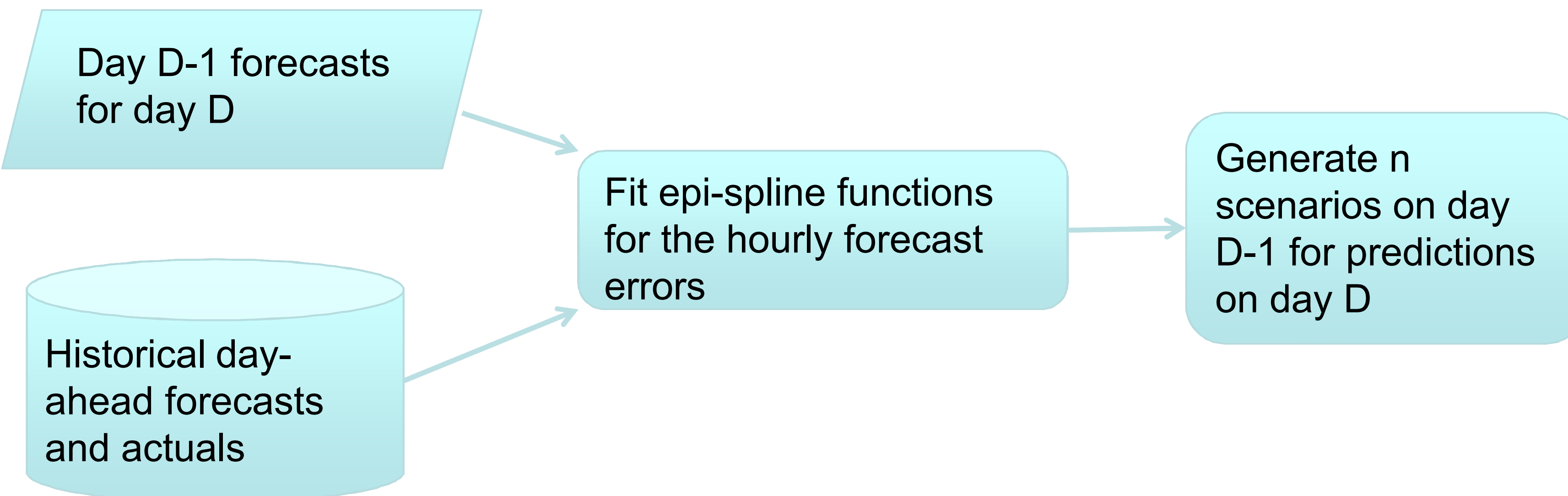
- Generation levels
- Power flows
- Voltage angles
- …

On Scenario Generation for Stochastic Unit Commitment…

- Stochastic programming, like all things algorithmic, operates on the "GIGO" principle (Garbage In, Garbage Out)
  - If you don't get the scenarios right, then the solution will be useless
- Our observation from numerous projects is that scenario generation dominates the time required for algorithm development
  - Roughly an 80%/20% split in "practice"
- There is a huge historical database of forecasted and actual observations, which can be leveraged to create accurate stochastic process models of load, wind, and solar
  - This is operations – no need to pretend that distributions don't exist

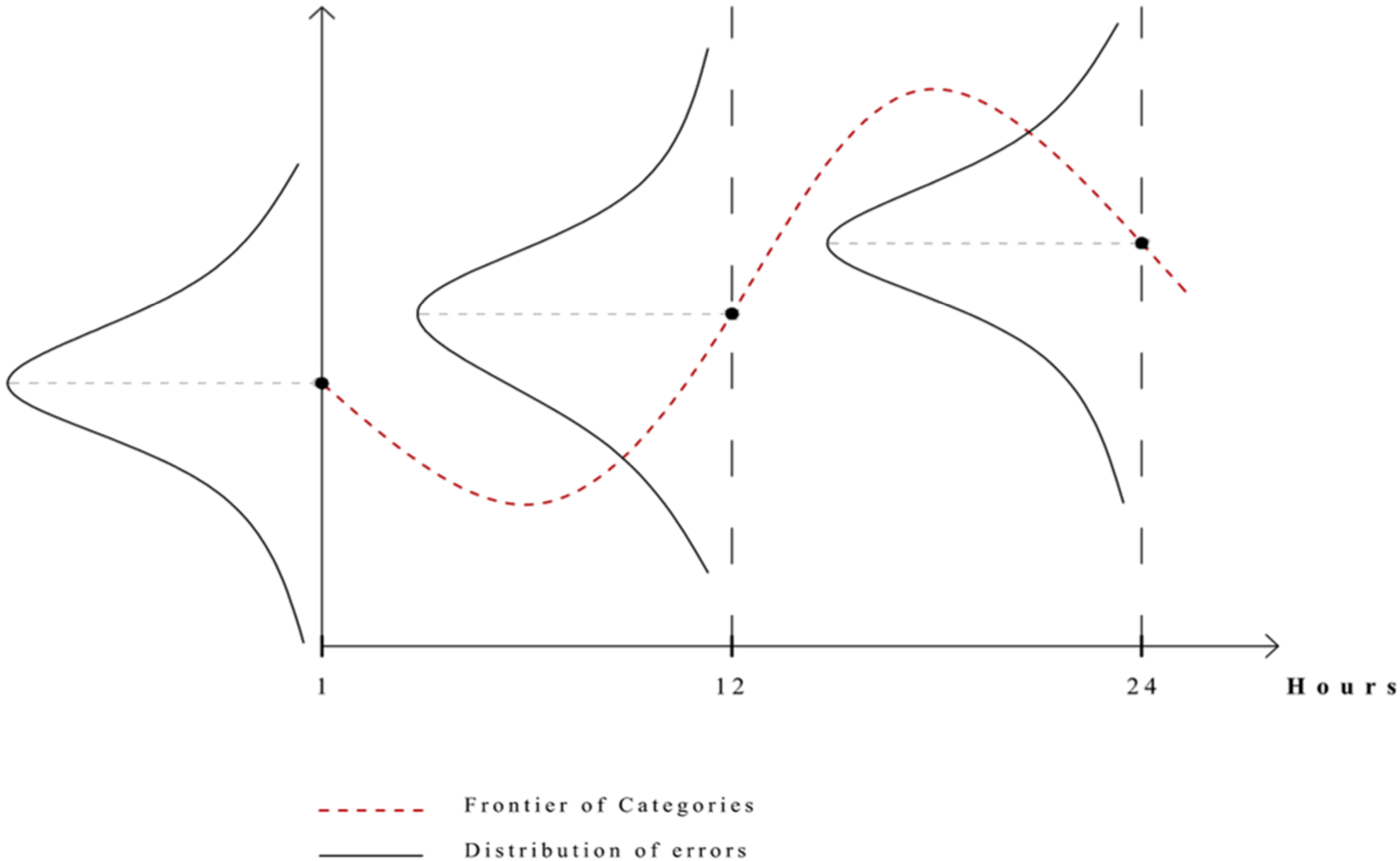
Epi-Splines and Scenario Generation

- We have developed a novel technique for approximating stochastic process models using historical forecast data and corresponding actual realizations
  - Based on epi-spline technologies (Wets et al.)
  - Not Monte Carlo, not AR(I)MA
  - Approximates the distributions – no sampling required!



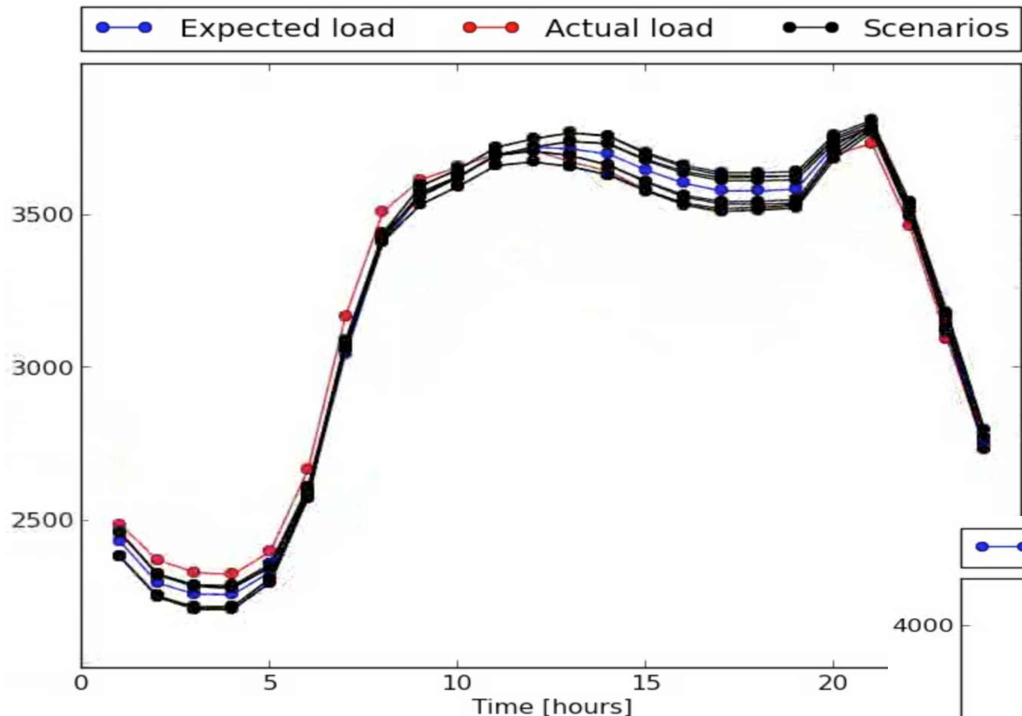
From Error Distributions to Scenarios

Sandia National Laboratories



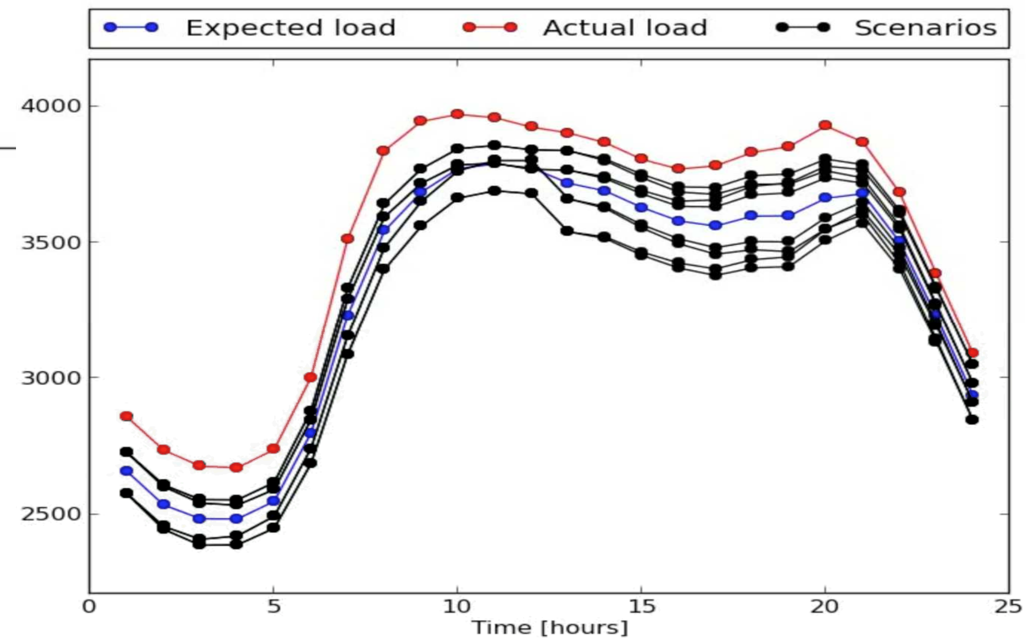
Extensive problem-specific research required to analyze time evolution of forecast errors…

Illustrative Load Scenarios: ISO-NE



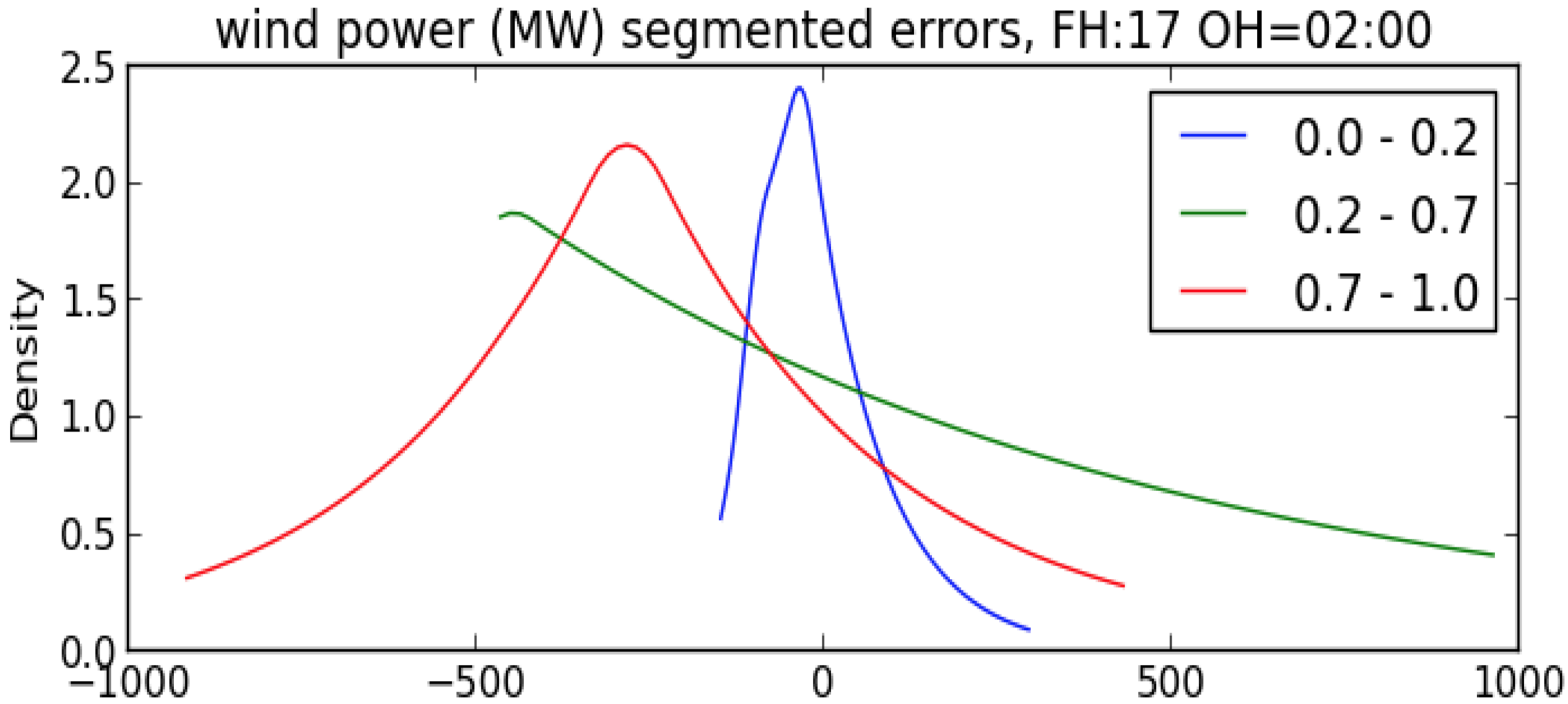
If the historical data indicates no variability, then the scenarios will reflect that consistency

Captures variability in load when present – but predictions are not perfect!



Wind Power Error Distribution Estimation

Aggregate power forecast errors are ***not*** well-represented by standard parametric (e.g., Gaussian) distributions …
… and the qualitative nature of the distribution varies by aggregate power level

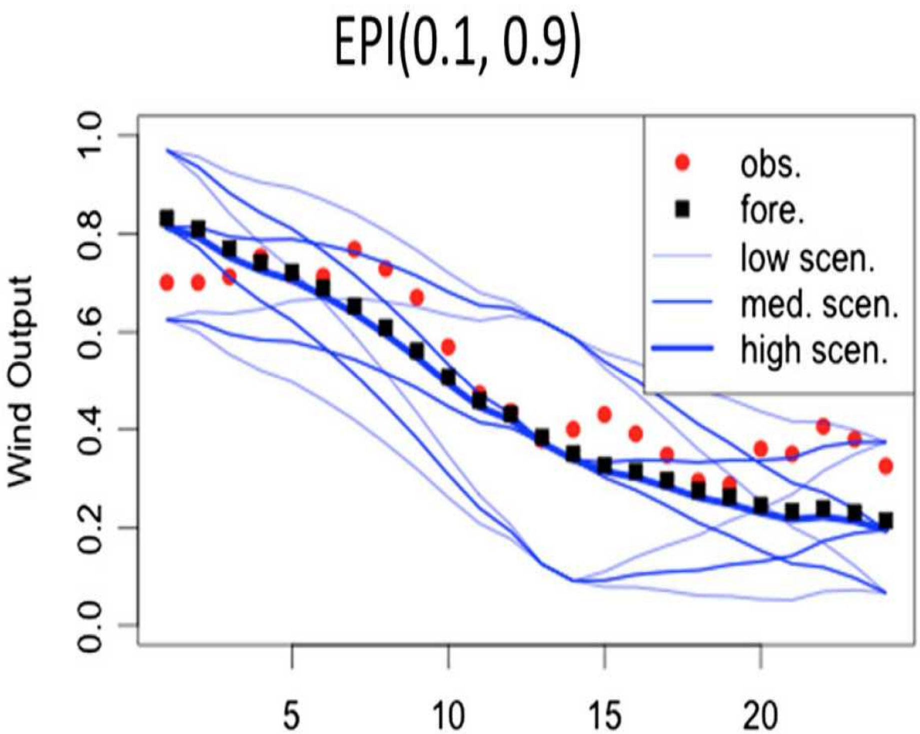


Generated using publicly available BPA data

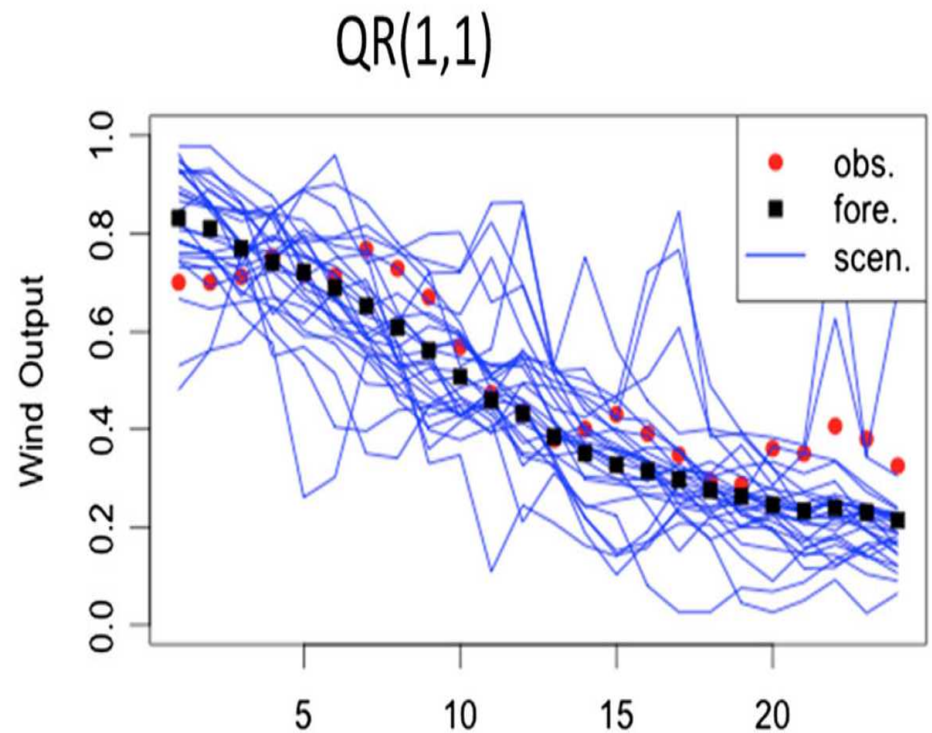
Wind Power Scenario Generation: BPA

Our scenario generation methods do not yield "noisy" scenarios, which are prevalent in scenarios generated by state-of-the-art quantile regression methods

Epi-Spline Scenarios



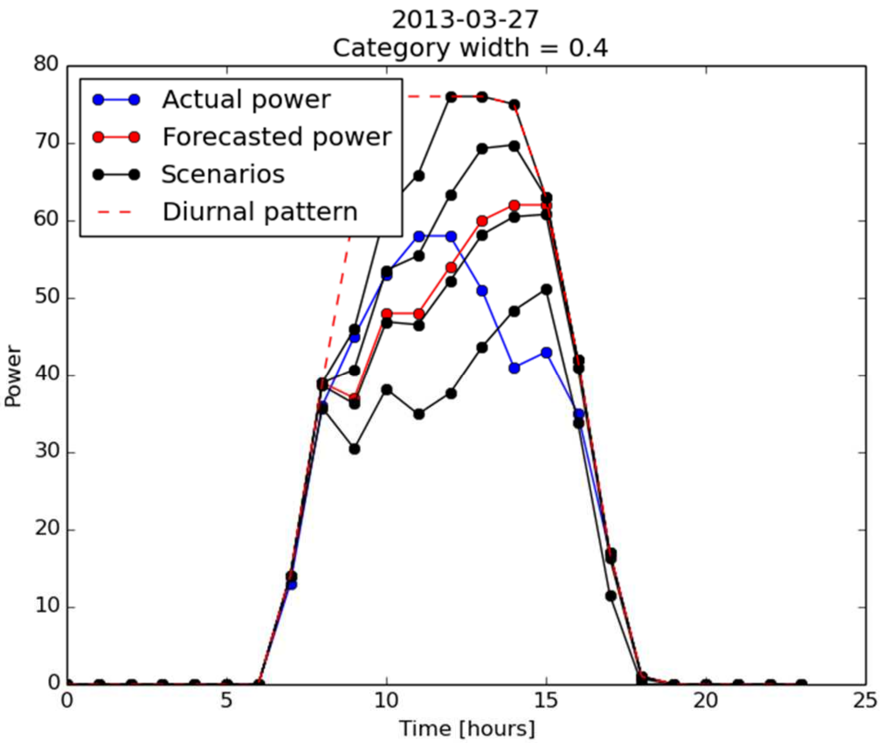
Quantile Regression Scenarios

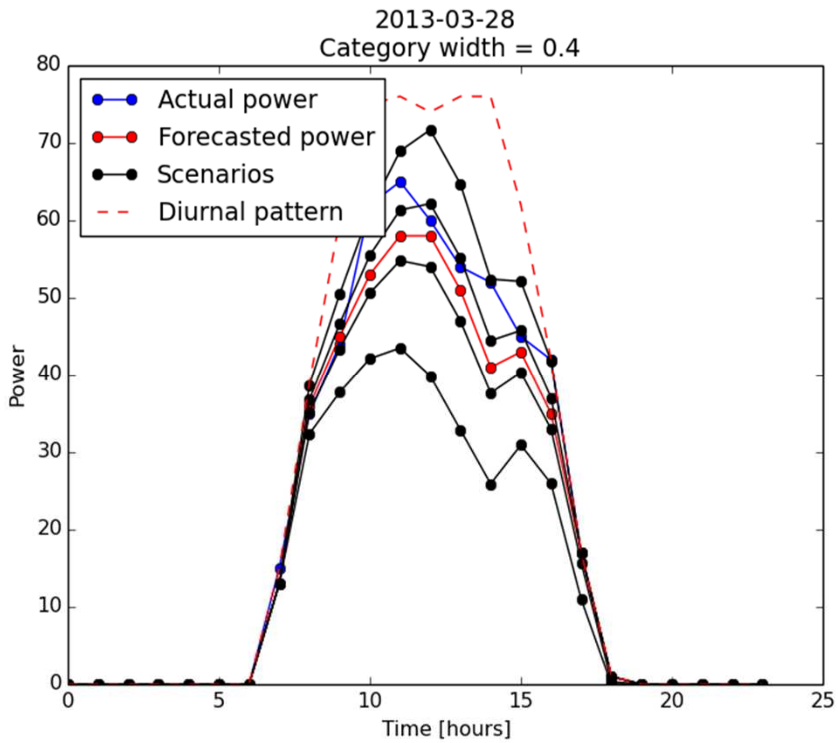


From Sari et al., *Wind Energy* (2015)

Solar Scenario Generation: CAISO

- Using published historical forecast/actual time series





Scenario Generation: Discussion

- We can and should leverage the significant volume of historical data concerning load and renewables forecast / actuals
  - Arguably do ***not*** need stochastic forecasts from vendors
  - Instead, we can build stochastic models from historical point forecasts

- Stochastic process model accuracy can approach that of state-of-the-art point forecasting techniques
  - But in addition represents variability

- Approximation of stochastic process models, rather than Monte Carlo sampling, can yield significant reductions in the number of scenario required for stochastic unit commitment
  - Enabled by epi-spline-based models of stochastic load, wind, and solar

Cutting to the Chase: Cost Savings (1)

- Simulated studies on ISO-NE, using NREL renewables data
- Computed in terms of relative cost increase of deterministic over stochastic
  - Yes, this implies that stochastic does win (but)…
- Results in terms of percentages
  - Q1: 1.52%
  - Q2: 1.31%
  - Q3: 0.89%
  - Q4: 1.23%
- Case corresponds to roughly 20% wind penetration levels
  - Qualitatively different results as wind penetration levels grow to 50% and greater

Cutting to the Chase: Cost Savings (2)

- Translating percentage savings into dollars…
  - Q1: ~$4M per month
  - Q2: ~3M per month
  - Q3: ~$12M per month
  - Q4: ~$2.5M per month
- Overall, the savings in 2011 "would have been" $64.5M
- That is real money!
- Or maybe: That is real money?
  - Full details require (at least) another hour-long talk…

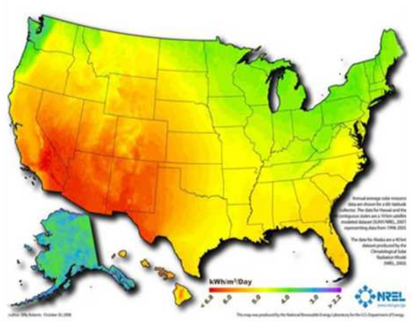
Vignette # 3:

Expansion Planning

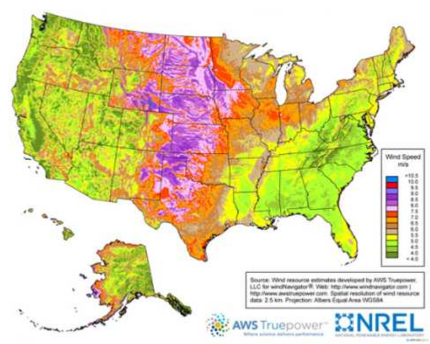
Transmission Expansion: Introduction

Sandia National Laboratories

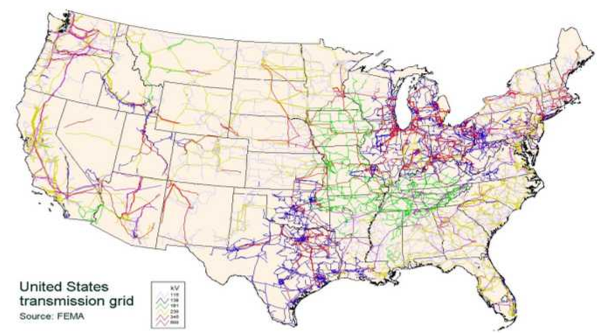
Solar Resources (NREL)



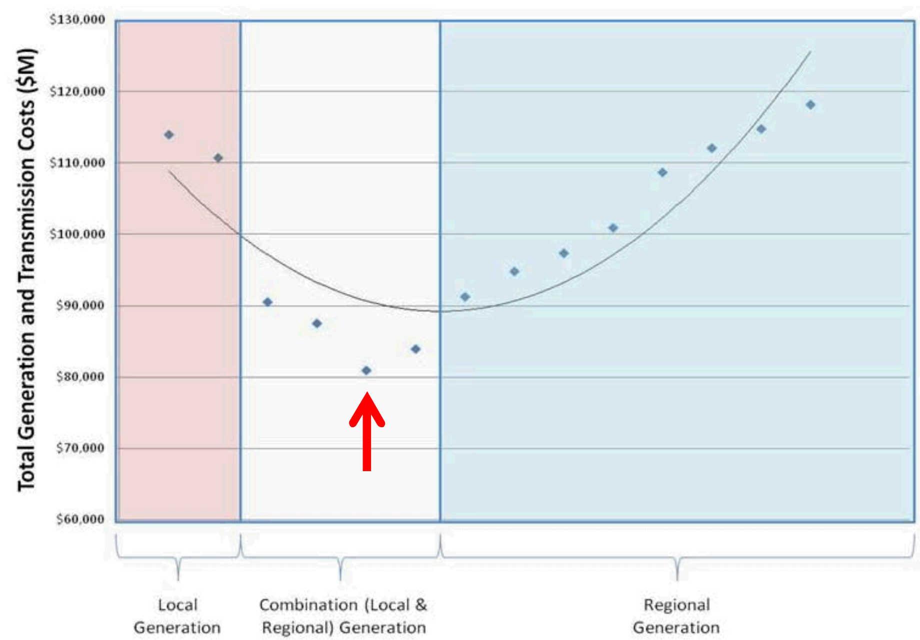
Wind Resources (NREL)



U.S. Transmission System (FEMA)



Zone Scenario Generation and Transmission Cost (MISO, 2010)



Goal:

Identify most cost effective combination of transmission and generation investments to meet:

1) Forecasted demand
2) Renewable and environmental goals

A Stochastic Planning Model

Sandia National Laboratories

Objective: Minimize present worth of capital plus operation costs

Decision variables

- Transmission investments (binary)
- Generation investments (continuous)
- Generation dispatch
- Power flows
- Phase angles
- Load curtailment

Deterministic constraints

- Transmission build limits (max number of circuits per corridor)
- Generation build limits (max capacity per bus, renewable resource potentials)
- Installed reserves (min firm capacity per region, ELCC for renewables)
- RPS constraint (min generation from renewables, dualized, treated as soft constraint)

Scenario-dependent constraints (DC OPF)

- Supply = Demand (KCLs)
- Loop-flow constraints for existing lines (KVLs)
- Loop-flow constraints for candidate lines (disjunctive KVLs)
- Thermal limits
- Max generation limits (use hourly capacity factors from historical data for renewables)

Test Case: WECC 240-bus System

WECC 240-bus system:
(Price & Goodin, 2011)

- 140 Generators (200 GW)
- 448 Transmission elements
- 21 Demand regions
- 28 Flowgates

Renewables data (Time series, GIS)
(NREL, WREZ, RETI)

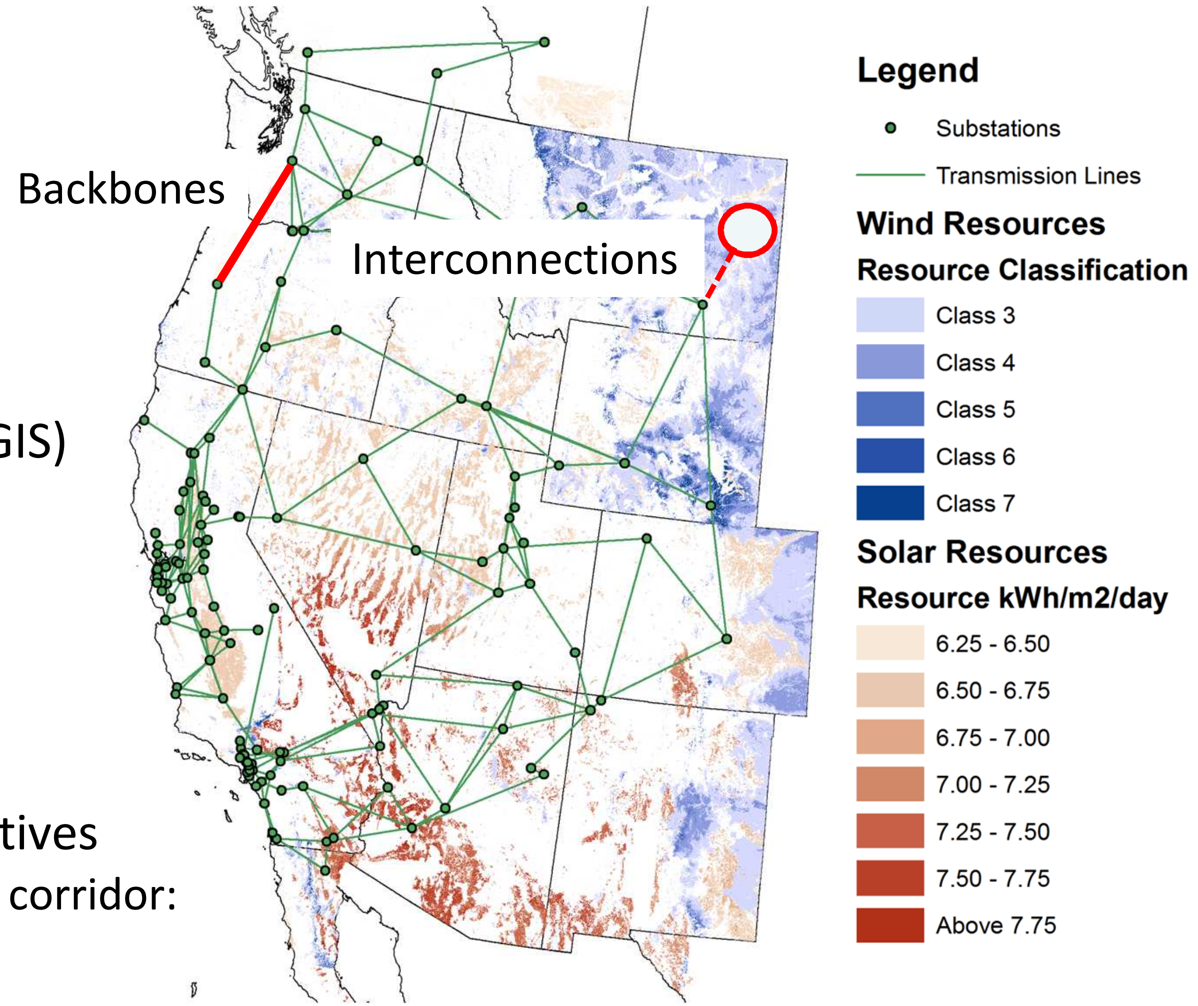
- 54 Wind profiles
- 29 Solar profiles
- 31 Renewable Hubs (WREZ)

Candidate Transmission Alternatives

Maximum number of circuits per corridor:

2 for Backbones

4 for Interconnections to Renewable Hubs

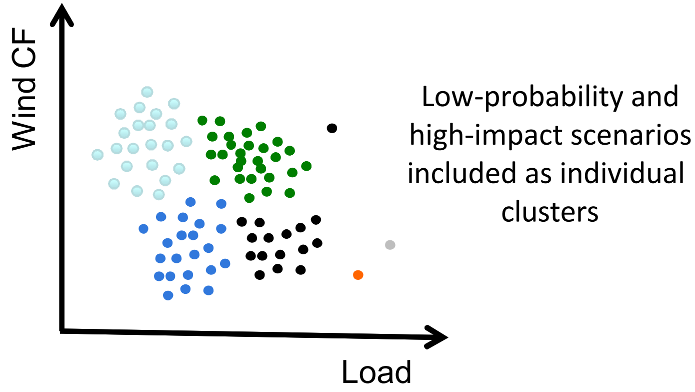


A Scenario Reduction Framework

- The computational requirements for solving stochastic optimization problems are proportional to the number of scenarios considered
- Even with HPC resources, we have difficulty solving for more than 500 scenarios in < hours run times
- So – we need to reduce the number of scenarios

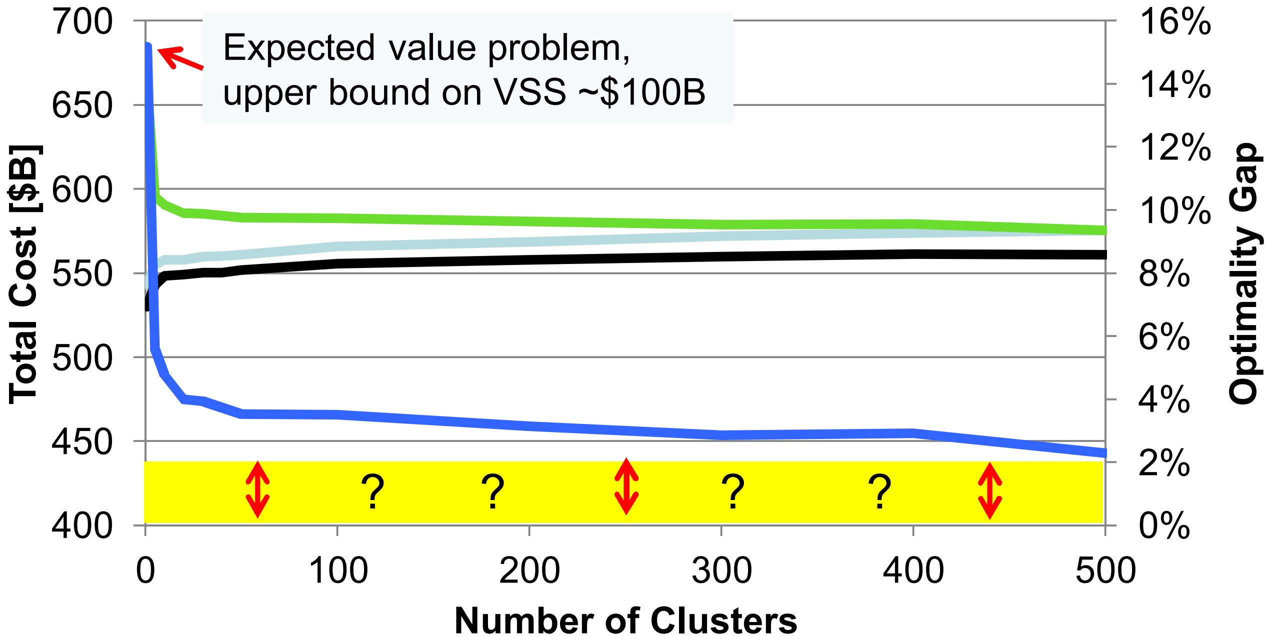
k-means clustering

- Group similar hours with similar loads, wind, solar, and hydro levels.
- Reduced problem provides a lower bound on optimal total system cost because all stochastic parameters are on RHS of constraints (Munoz et al, 2014). The more clusters, the tighter the lower bound.



The Value of Stochastic Expansion

Convergence of upper and lower bounds

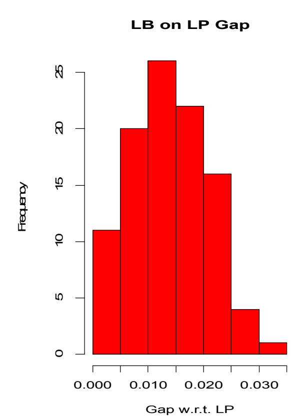


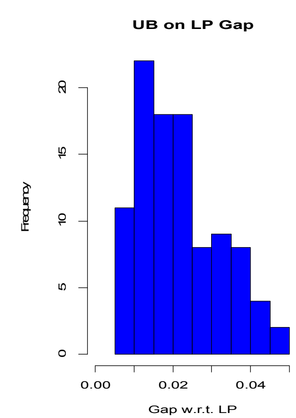
500-scenario problem

- **Red Mesa HPC:** 1.9 hrs.
- **Workstation :** 8.7 hrs.

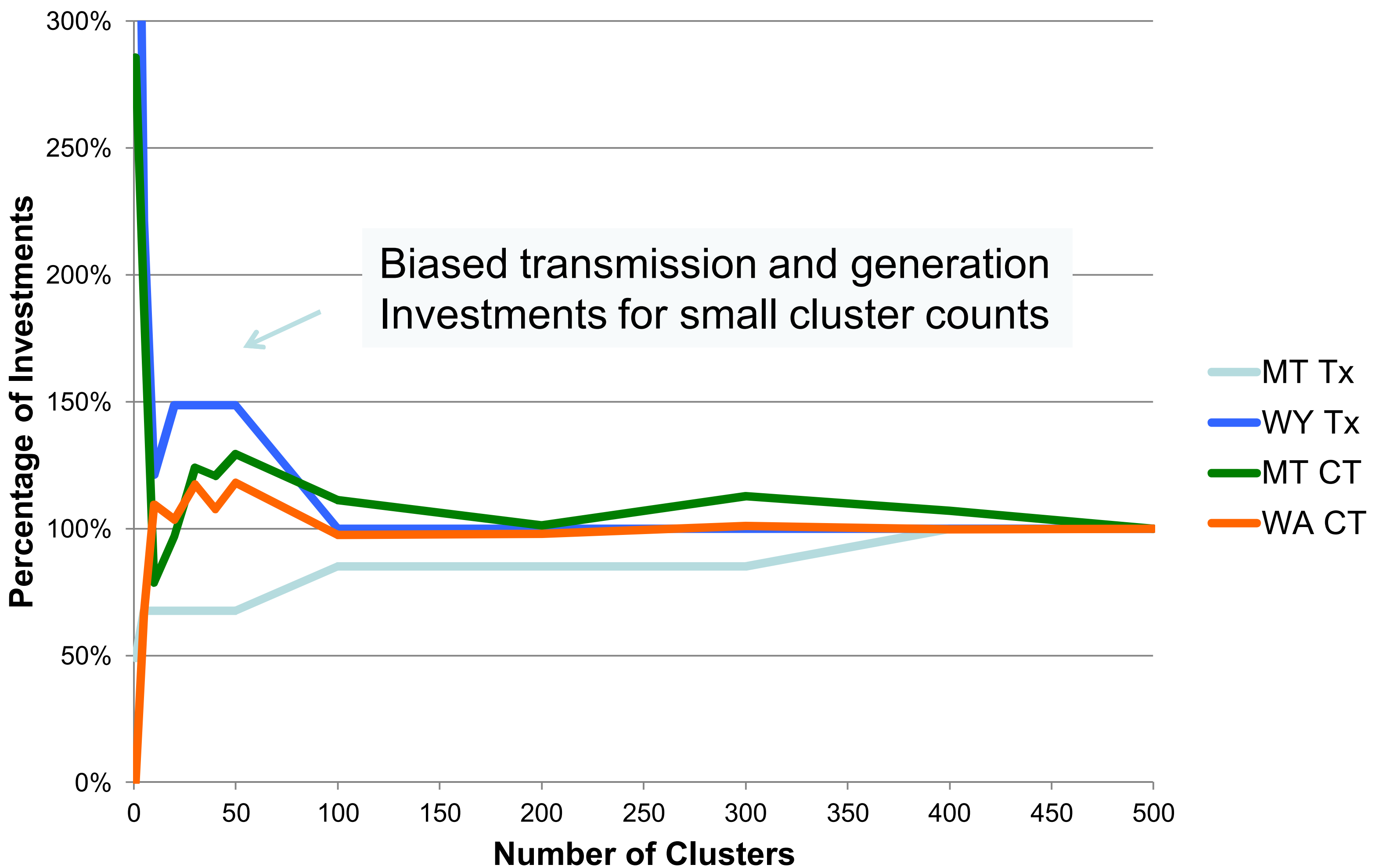
LP relaxation vs optimal solution from MILP:

- Solved 100 single-scenario problems, 0.5% gap and 1hr time limit
- MILP lower bound w.r.t. LP relaxation is **1.4%** (average)
- MILP upper bound w.r.t. LP relaxation is **2.2%** (average)





Investments vs. Time Granularity

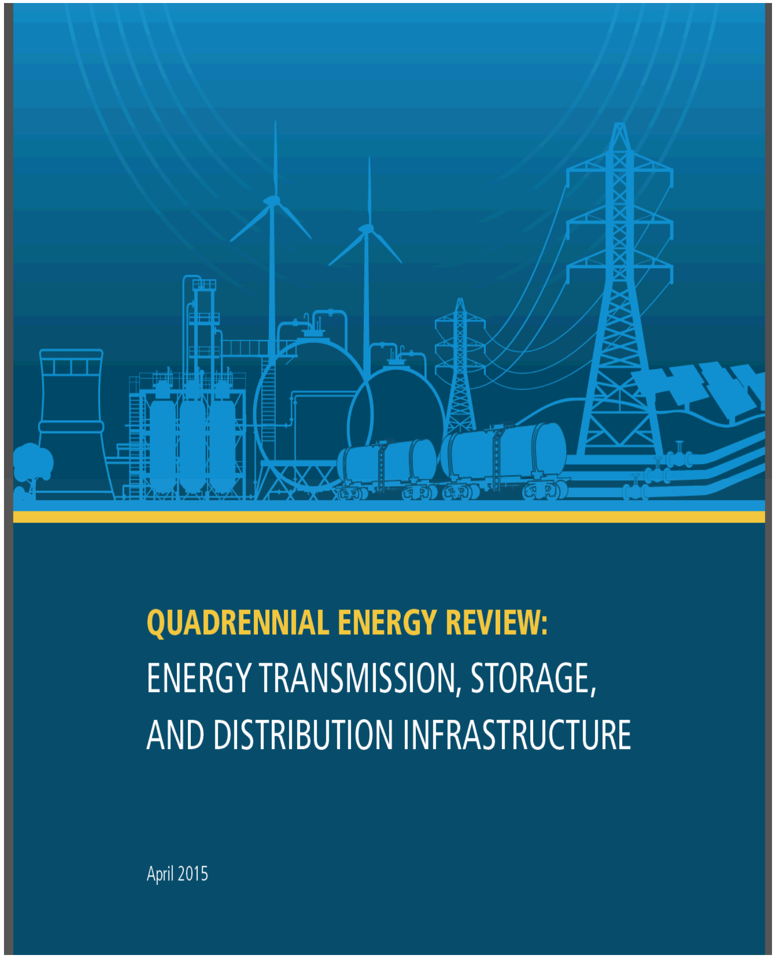


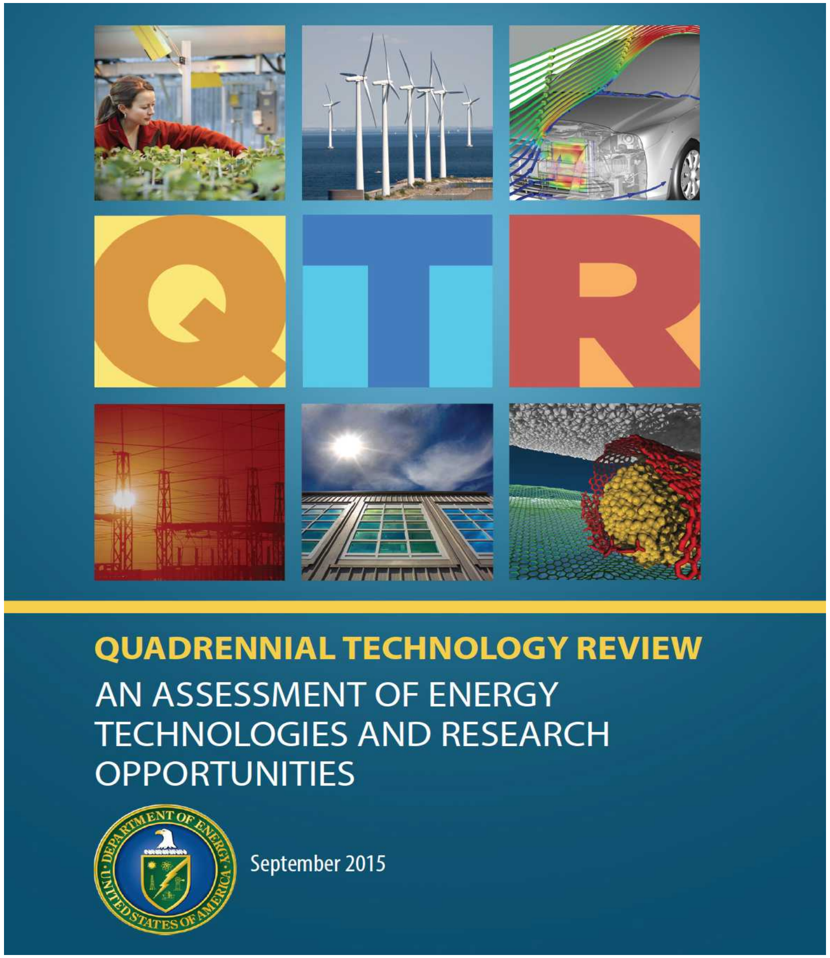
Open question: Is it possible to obtain the solution at the right using far fewer numbers of clusters (scenarios)?

Vignette # 4:

Resiliency Analysis

Electricity Grid Resiliency Analysis





Defining Resilience

Sandia National Laboratories

Presidential Policy Directive (PPD) 21

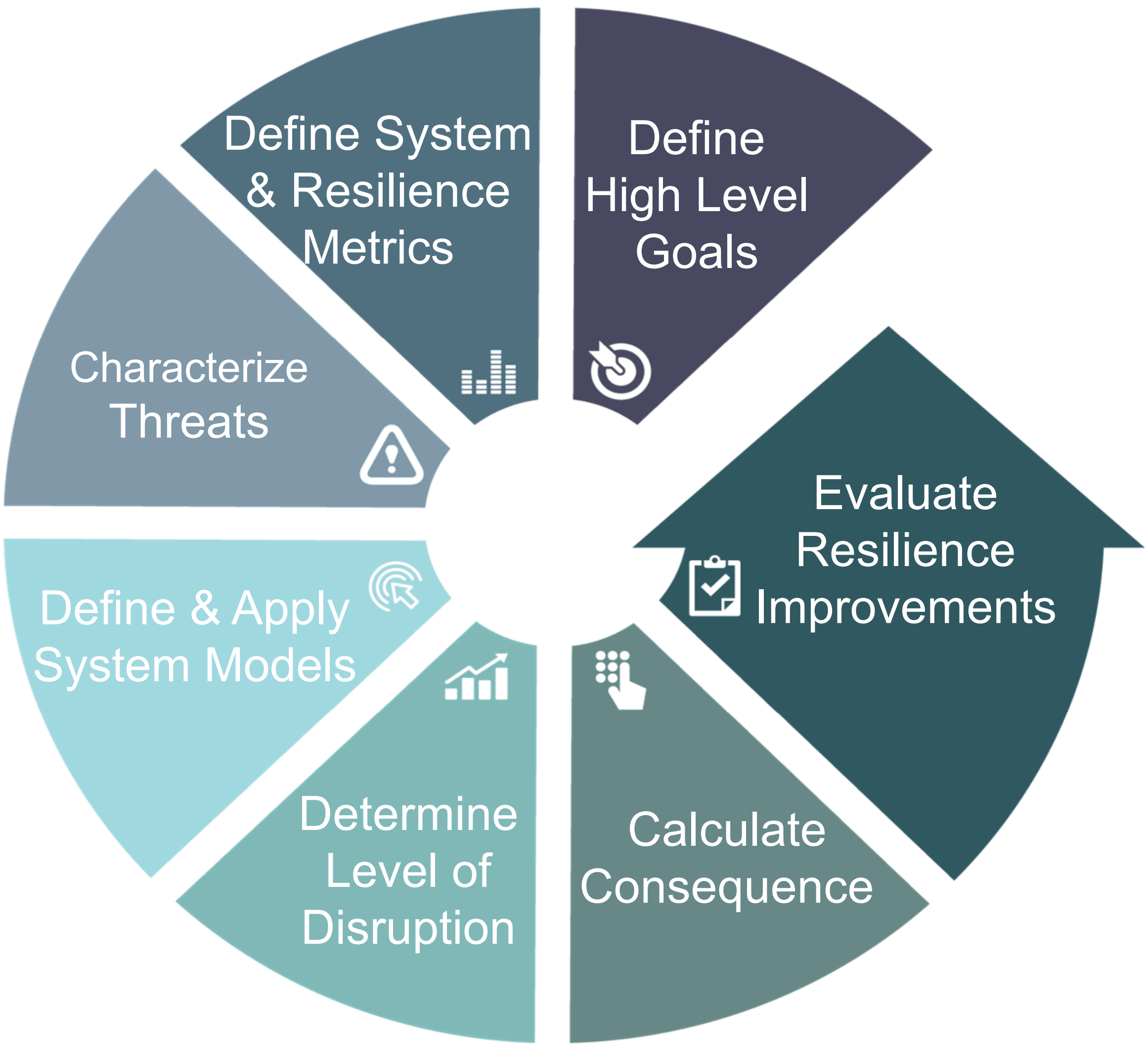
1. "[preserve] infrastructure that are vital to the **public confidence** and the Nation's **safety**, **prosperity**, and **well-being**."
2. "[prevent] debilitating impact on the national **security**, **economic** stability, **public health** and **safety**, or any combination thereof"
3. "…analyze *threats* to, *vulnerabilities* of, and potential *consequences* from all hazards on critical infrastructures".

-PPD-21: Critical Infrastructure Security and Resilience

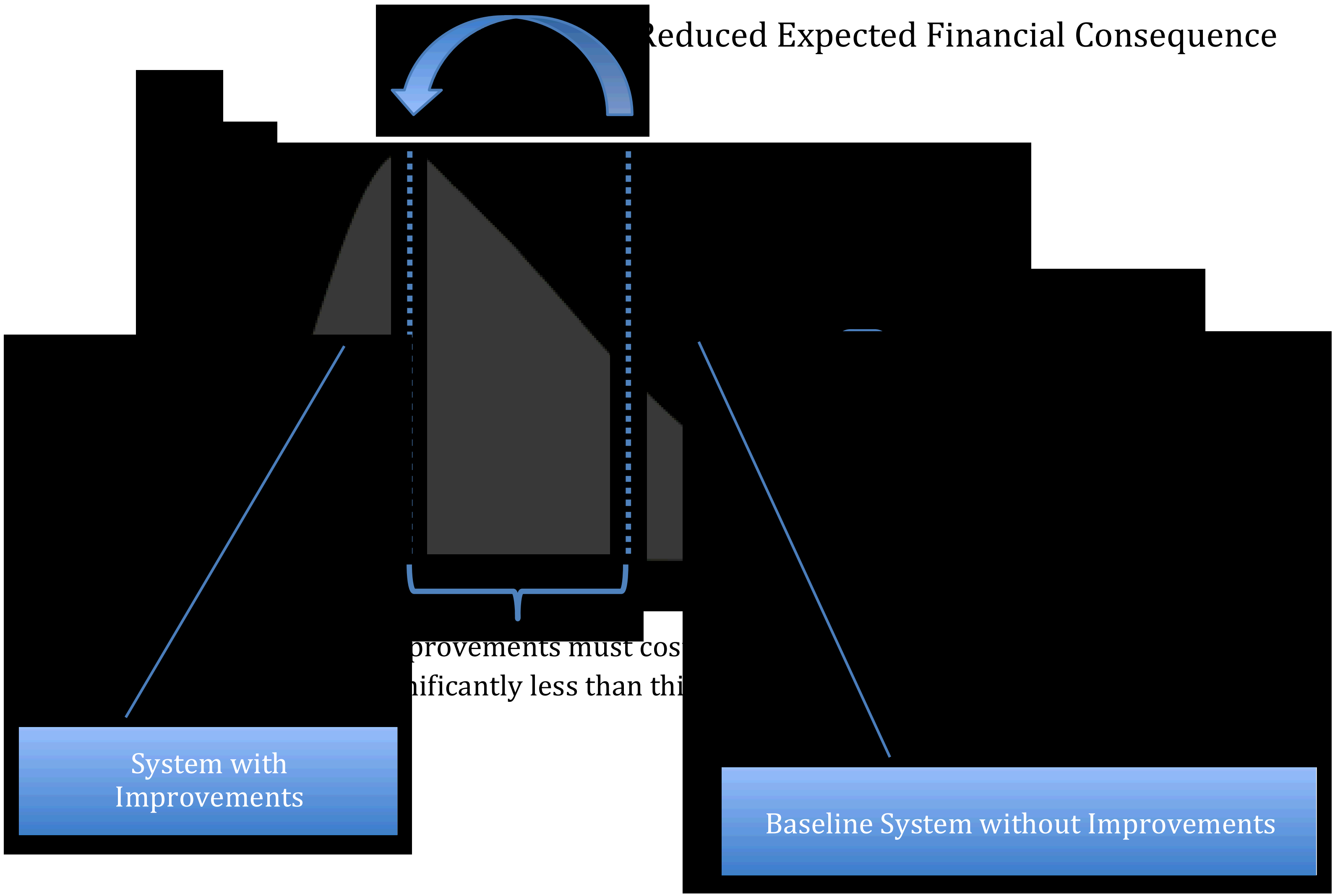
"without some numerical basis for assessing resilience, it would be impossible to monitor changes or show that community resilience has improved. At present, no consistent basis for such measurement exists…"

-Disaster Resilience: A National Imperative, National Academy of Sciences

A Framework for Resilience Analysis



Evaluating System Improvements



Scenario Analysis: Identify Threat Types

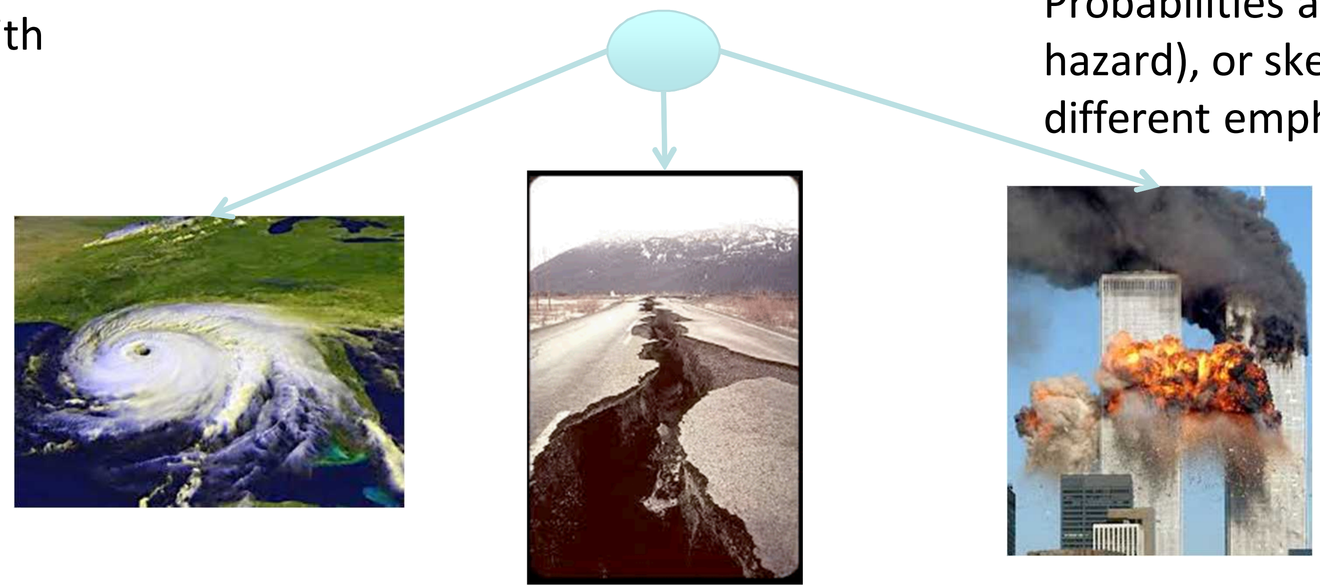
A infrastructure is designed to be resilient to a specific set of possible disruptions

Definition of possible disruptions can proceed via construction of a ***scenario tree***

Alternatives exist, but they are more nuanced in terms of definition

We begin with high-level threat definitions

Probabilities are uniform (all-hazard), or skewed to reflect different emphases



High-level scenario identification is expected to be an output from an iterative and interactive stakeholder-driven process

Scenario Analysis: Characterize Individual Threat

Given high-level threat characterization, the next step is to further refine the description of the specific threats

<u>Historical information and forecast models is used to guide specification of possible events and their relative likelihoods</u>

…

p_1 … p_2 … p_n

Category 2, landfall at high tide

Category 4, north-of-peninsula storm track

Category 5, eye tracks over metropolitan area

Scenario Analysis: Disrupting the System

The final step is to translate disruption events into system impacts

p_1 … p_2 … p_n

Category 2, landfall at high tide

Category 4, north-of-peninsula storm track

Category 5, eye tracks over metropolitan area

… …

Assume uniform probabilities

Damage Realization K

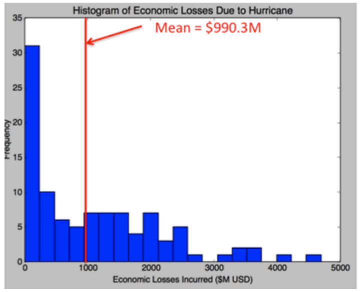
Damage Realization N

Given a specific manifestation of a disruption event, we then specify a *distribution* of infrastructure impacts

For IEEE 118 bus system:

1. Normal distribution of generator failures, with $\upsilon=20$, $\sigma=5$
2. Normal distribution of line failures, with $\upsilon=40$, $\sigma=7$

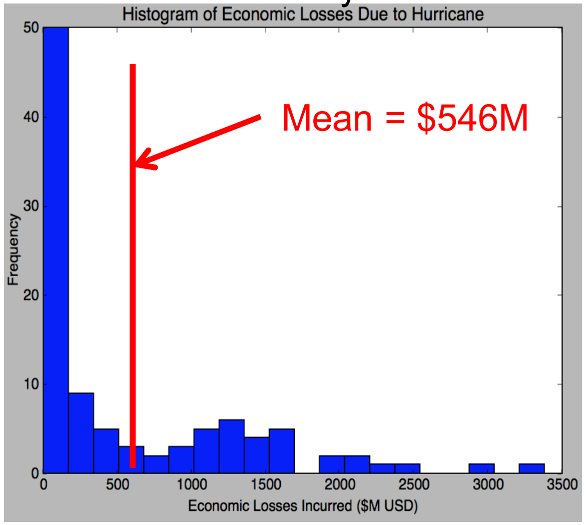
Ex: How Should We Invest $100M?



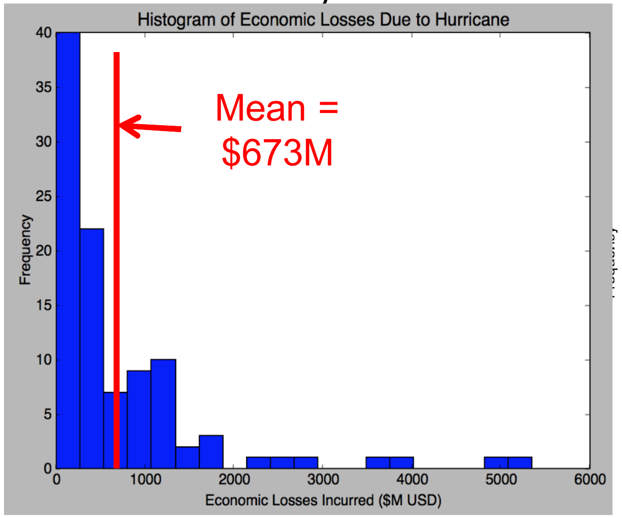
Baseline mean was $990M

Invest the same $100M in both flood walls and burying cables

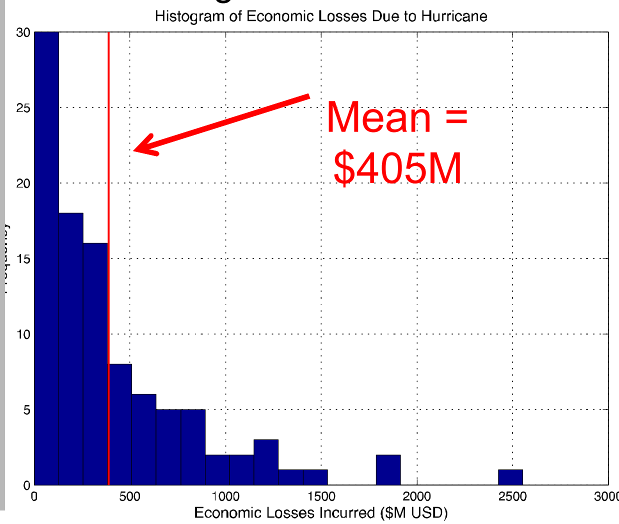
$100M of generator flood walls only



$100M of burying lines only



$100M of burying lines and generator flood walls



Conclusions

- Power grid problems are differentiating for data science due to the intimately linkage with formal decision problems
  - Playing very high in the Davenport and Harris analytic hierarchy
- Stochastic optimizers need to work with data scientists!
  - We can developer solvers and publish papers on our own …
  - … but we can't have any real impact by doing so!

QUESTIONS