Prescient:
software for scenario creation
and production cost modeling under uncertainty

Professor David L. Woodruff
Graduate School of Management
University of California Davis
Davis, CA 95616 USA

High-Level Talk Goals

- Introduce the challenges and one solution to the construction of probabilistic scenarios for key quantities in power systems operations

- Discuss the rigorous evaluation of stochastic optimization approaches to key power systems operations

- Illustrate the relationship between probabilistic scenarios and power system performance using a illustrative case study

A Word From Our Sponsors…

- Grid Modernization Laboratory Consortium (GMLC)
  - Project 1.4.26 – Multi-Scale Production Cost Modeling

- Bonneville Power Administration (BPA)
  - Funded work on high-accuracy probabilistic wind forecasting
  - Provide real-world data sets, publicly available

- Department of Energy’s ARPA-E office
  - Scalable stochastic unit commitment project

Representative Key Collaborators

- Sandia National Laboratories
  - Jean-Paul Watson
  - Andrea Staid

- University of California Davis
  - Roger Wets
  - Dominic Yang

- Purdue University
  - Benjamin A. Rachunok

Part 1: Counterfactual Re-Enactment Methodology

Data

Our interest is in quantifiable results, so we view the world as a generator of data. At time t, the world generates a vector of "observations" $\mathcal{O}(t)$.

The Objective

The goal is to come up with a way to do well in the future, perhaps by taking into account the nature of the world and observations from the past. We quantify "doing well" using a function

$$\tilde{f}_t(x, \{\mathcal{O}(\tau), \tau \geq t\})$$

where x is a vector of decision values made just before time t.

To Use A Computer To Optimize

- One common approximation is to look at parts of $\mathcal{O}(\tau)$ and we will call these vectors $\xi(\tau)$. This is typically done for $\tau \geq t_{now}$, where t_{now} is the time "now."
- One also typically approximates $\tilde{f}_t(x, \{\mathcal{O}(\tau),\ \tau \geq t\})$ using some function $f_t(x, \{\xi(\tau),\ \tau \geq t\})$ that is easier to work with or at least possible to write down.

The Future Matters

- Only the decisions for "now" can be implemented, but typically the decisions impact the world in the future (i.e., $\{\mathcal{O}(\tau), \tau \geq t_{now}\}$ depend on x) so it is usually a good idea to take that into account.

- At time $t = t_{now}$ one might use a computer to find

$$\operatorname*{argmin}_x f_t(x, \{\xi(\tau),\ \tau \geq t\}))$$

and maybe explicitly require $x \in \Omega(\xi)$.

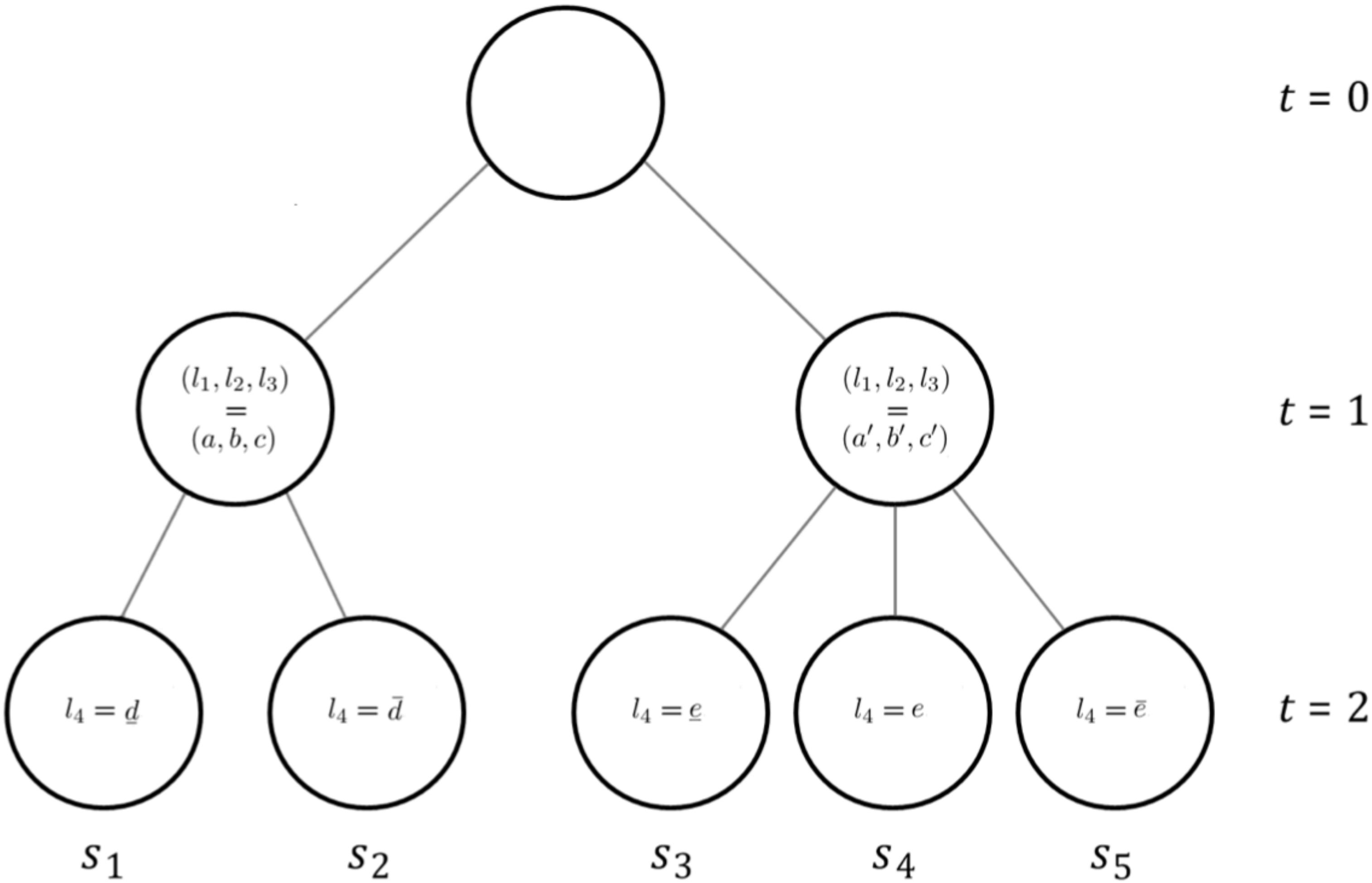
Discrete Scenarios

- Define $\xi \equiv \{\xi(t)\}_{t=1}^{T}$ on a discrete probability space.
- Use Ξ to represent the full set of scenarios.
- Each scenario, ξ, has probability π_{ξ}.
- Write simply ξ to represent the entire scenario.

Scenario Trees

- We organize ξ into a tree with the property that scenarios with the same realization up to stage t share a node at that stage.
- So $\vec{\xi}^{\,t}$ refers also to a node in the scenario tree.
- Let $\mathcal{G}_t$ be the set of all scenario tree nodes for stage t
- Let $\mathcal{G}_t(\xi)$ be the node at time t for a particular scenario, ξ.
- For a particular node $\mathcal{D}$ let $\mathcal{D}^{-1}$ be the set of scenarios that define the node.

A Scenario Tree



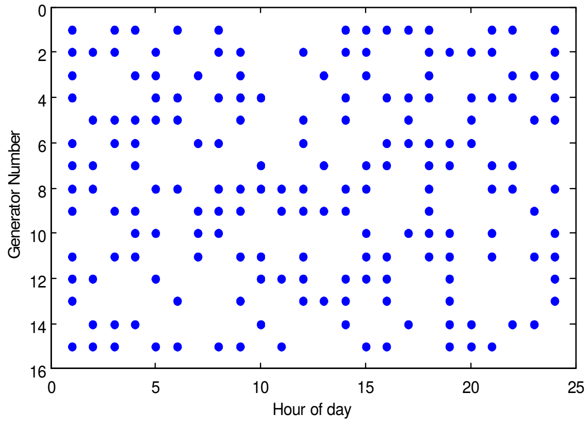
Potential Uses for Scenarios

- Analysis of a plan or policy (x); e.g. simulation
- Optimization

The General Structure of a Stochastic Unit Commitment Optimization Model

Sandia National Laboratories

Objective: Minimize expected cost



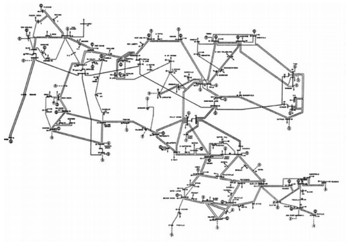
First stage variables:

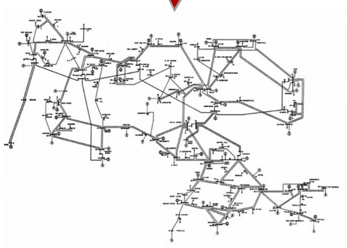
- Unit On / Off

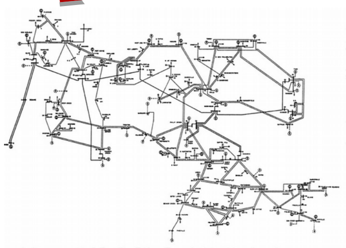
Nature resolves uncertainty

- Load
- Renewables output
- Forced outages

p_1 p_2 … p_N







Scenario 1 Scenario 2 … Scenario N

Second stage variables (*per time period*):

- Generation levels
- Power flows
- Voltage angles
- …

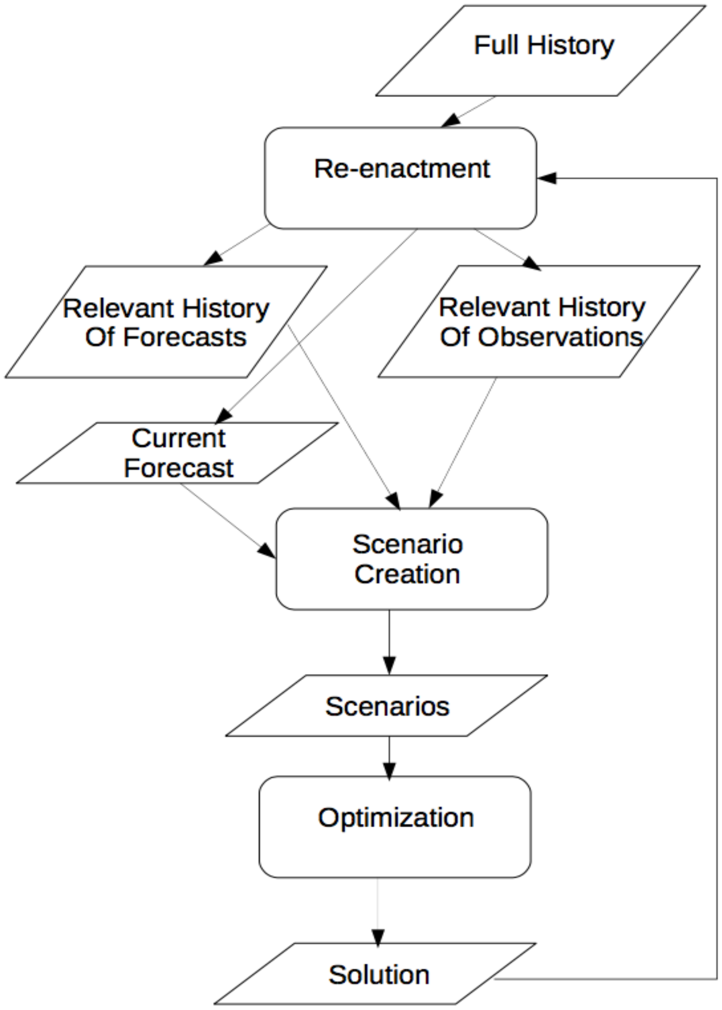
Some Ways to Get Scenarios

- Statistical models, perhaps obtained by data mining (e.g., in Finance)
  - Monte Carlo sampling
  - Moment matching
- Simulations (e.g., in Forest Harvesting with Fire Risk)
- Forecast Error Distributions (e.g., Unit Commitment)

Evaluating Scenarios

- Analyze the statistical properties
- Analyze the solutions obtained
  - Simulation
    - In-sample
    - Out-of-sample
    - Independent
  - Re-enactment

Software Architecture for Counterfactual Re-Enactment



Sketch of Counterfactual Re-Enactment

Sandia National Laboratories

Details are *highly* application specific

- $\tilde{f}(\cdot) \succ \texttt{Eval}(\cdot) \succeq f(\cdot)$
- $\texttt{Eval}(\hat{x}; \mathcal{O}(\tau), \tau = t_{now} + T_{lt}, T)$, ($T$ is end of data)
- T_{oper} = periods of use of the solution; T_0 is first time with data and T_{sc} is needed for scenarios.

1: **Initialization**:
2: **Scenario Creation**:
3: **Optimization:**
4: **Evaluation and Record Keeping:** Compute and store the results of $\texttt{Eval}(\hat{x}; \mathcal{O}(\tau), \tau = t_{now} + T_{lt}, T)$
5: **Iterate:**
6: **Termination:**

This is a platinum standard rolling horizon simulation – *sounds* easy, but the devil is in the details

Part 2: Constructing Probabilistic (Mainly Wind) Scenarios

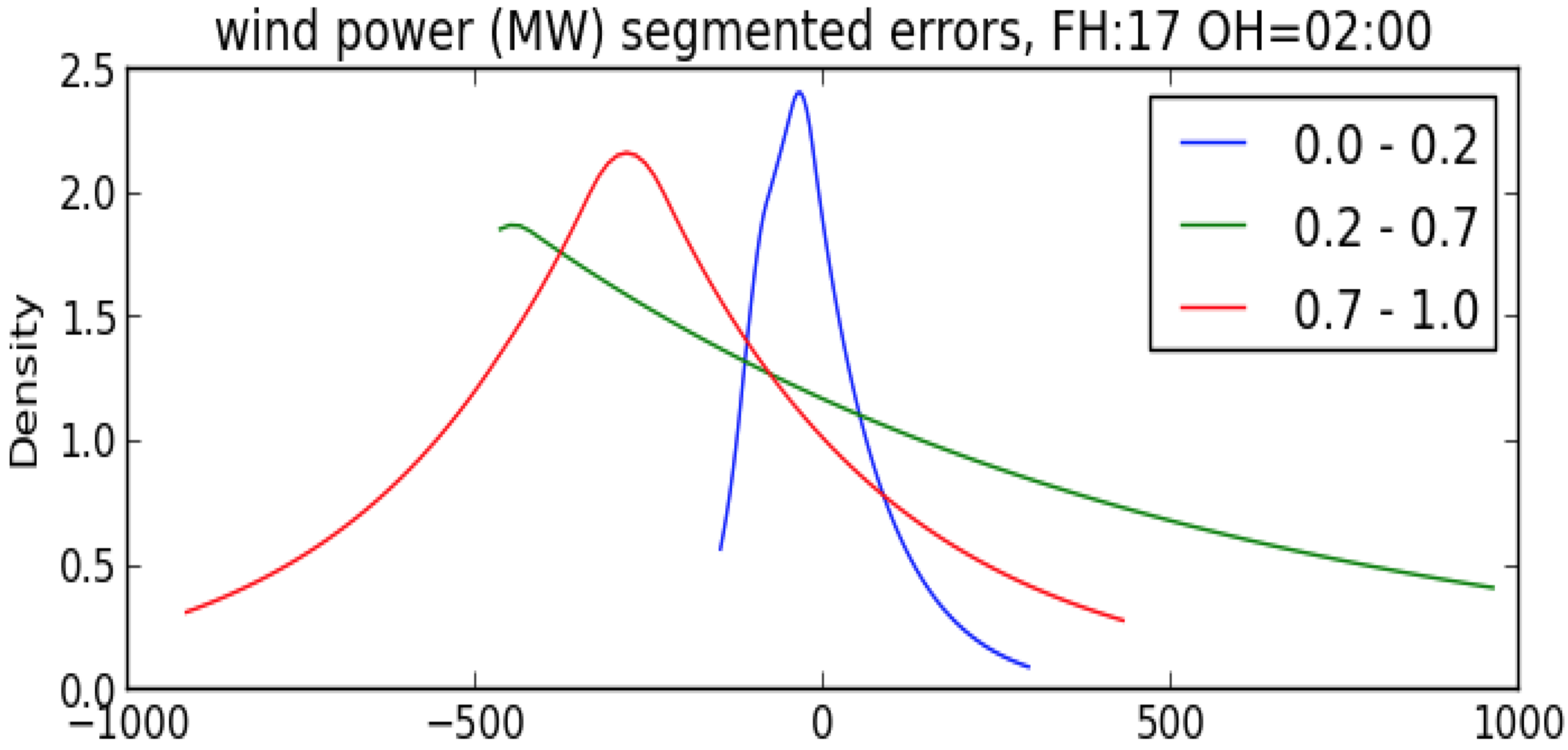
Probabilistic Wind Power Scenarios

- There are several ways to represent the uncertainty associated with wind in planning problems
- One approach is modeling this variable generation stochastically, using scenarios
  - Each scenario represents a possible trajectory of wind power over time and has an associated probability
- We rely on the availability of a point forecast and evaluate historical forecast errors to build up non-parametric distributions of expected future errors
- Epi-spline basis functions are used to allow a user to target specific partitions of the error distribution, thus controlling for scenarios that reach into the tails

Wind Power Error Distribution Estimation

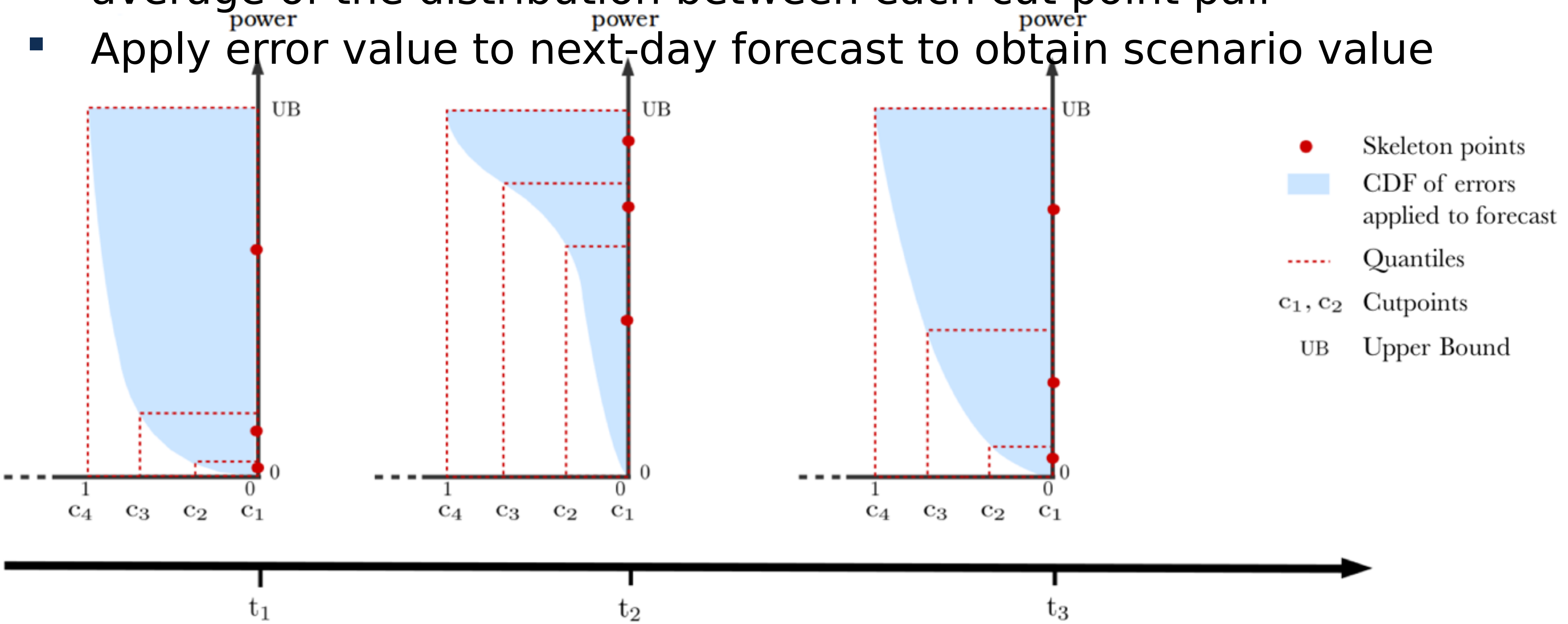
Aggregate power forecast errors are ***not*** well-represented by standard parametric (e.g., Gaussian) distributions …

… and the qualitative nature of the distribution varies by aggregate power level



Epi-Spline Scenario Creation

- For a subset of hours in day (i.e., hours 1, 12, 24), calculate empirical **forecast error** CDF from relevant* historical forecast/actual pairs
  - Correlations in forecast error drop off quickly with time, allowing for independent calculations
- Divide distribution at cut points, and calculate the weighted average of the distribution between each cut point pair
- Apply error value to next-day forecast to obtain scenario value



For full details: Staid A, Watson JP, Wets RJB, Woodruff DL. Generating short-term probabilistic wind power scenarios via nonparametric forecast error density estimators. *Wind Energy*. 2017. DOI: 10.1002/we.2129

Assessing Scenario Quality

- Visual comparisons only get you so far…
- There are a number of proper scoring rules used to evaluate probabilistic forecasts and scenarios
  - Energy Score (has known discrimination issues)
  - Brier Score (event-based, need to know what you care about upfront)
  - Variogram Score (improved discrimination using pairwise differences)
- However, ultimate test of quality is performance in a real-world system
  - More on this in Part 4 of this talk
- But we can say:
  - Scenarios should represent a wide enough range of plausible wind power realizations to ensure a feasible solution as the future unfolds
  - However, too wide of a range will drive costs up

Sandia National Laboratories

RESEARCH ARTICLE

Generating Short-Term Probabilistic Wind Power Scenarios via Non-Parametric Forecast Error Density Estimators

Andrea Staid[1], Jean-Paul Watson[1], Roger J.-B. Wets[2], and David L. Woodruff[2]

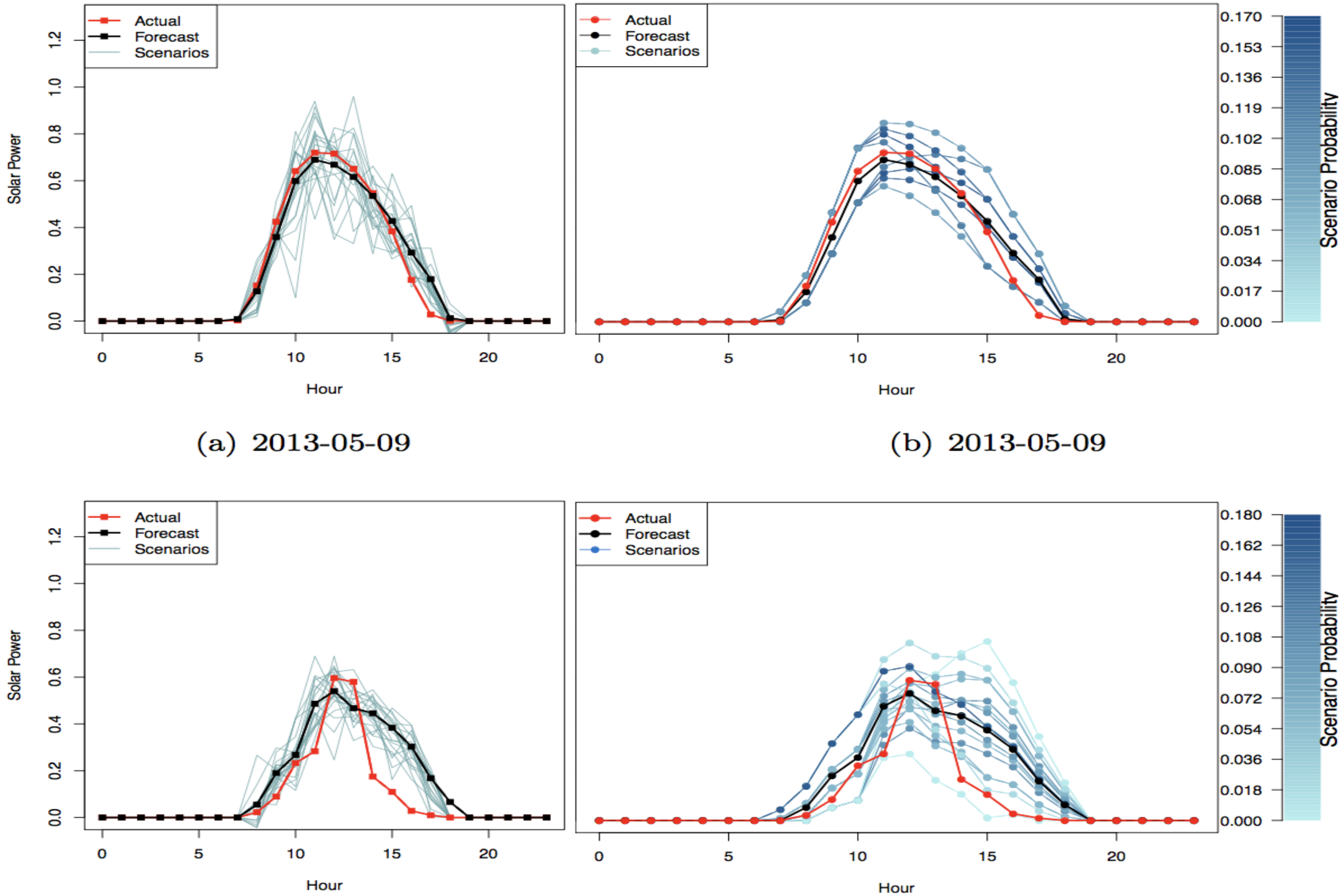
[1] Sandia National Laboratories, Albuquerque, New Mexico, USA
[2] University of California Davis, Davis, California, USA

ABSTRACT

Forecasts of available wind power are critical in key electric power systems operations planning problems, including economic dispatch and unit commitment. Such forecasts are necessarily uncertain, limiting the reliability and cost-effectiveness of operations planning models based on a single deterministic or "point" forecast. A common approach to address this limitation involves the use of a number of probabilistic scenarios, each specifying a possible trajectory of wind power production, with associated probability. We present and analyze a novel method for generating probabilistic wind power scenarios, leveraging available historical information in the form of forecasted and corresponding observed wind power time series. We estimate non-parametric forecast error densities, specifically using epi-spline basis functions, allowing us to capture the skewed and non-parametric nature of error densities observed in real-world data. We then describe a method to generate probabilistic scenarios from these basis functions that allows users to control for the degree to

In Wind Energy (2017)

Probabilistic (Bulk) Solar Scenarios



(a) 2013-05-09 (b) 2013-05-09

(c) 2013-05-15 (d) 2013-05-15

Constructing Probabilistic Scenarios for Wide-Area Solar Power Generation

David L. Woodruff

Graduate School of Management, University of California, Davis, CA 95616-8609, USA

Julio Deride, Andrea Staid, Jean-Paul Watson

Discrete Math and Optimization Department, Sandia National Laboratories, Albuquerque, NM 87185, USA

Gerrit Slevogt

Department of Mathematics, University of Duisburg-Essen, Germany

César Silva-Monroy

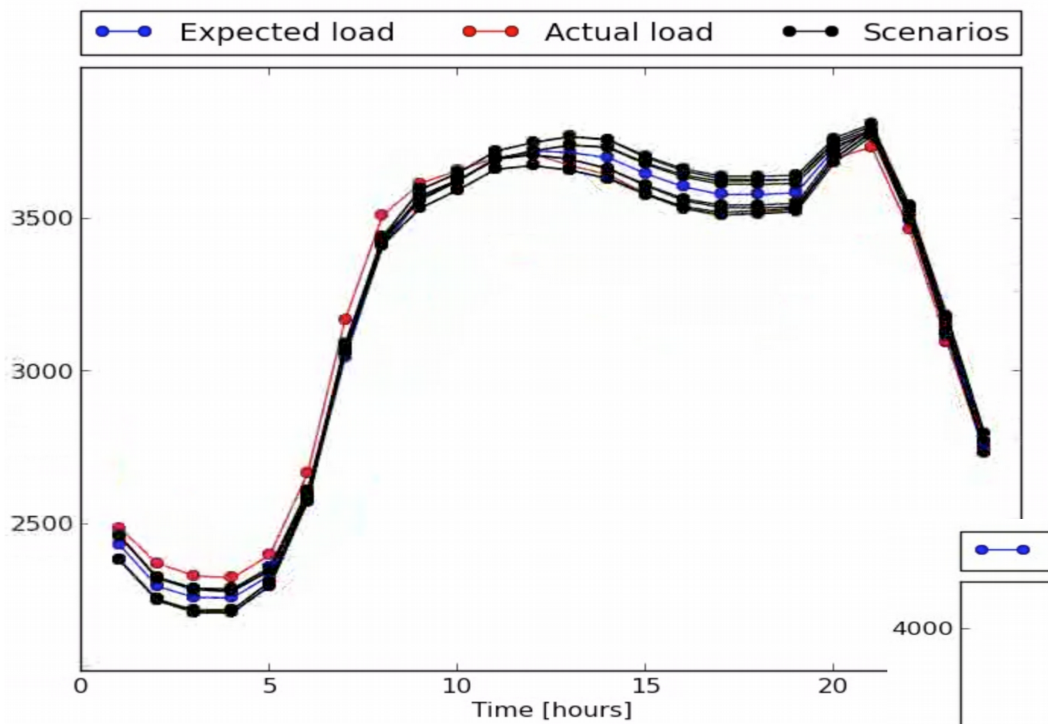
Demand Energy, Liberty Lake, WA 99019, USA

Abstract

Optimizing thermal generation commitments and dispatch in the presence of high penetrations of renewable resources such as solar energy requires a characterization of their stochastic properties. In this paper, we describe novel methods designed to create day-ahead, wide-area probabilistic solar power scenarios based only on historical forecasts and associated observations of solar power production. Scenarios are created by segmentation of historic data, fitting non-parametric error distributions using epi-splines, and then computing specific quantiles from these distributions. Additionally, we address the challenge of establishing an upper bound on solar power output. Our specific application driver is for use in stochastic variants of core power systems operations optimiza-

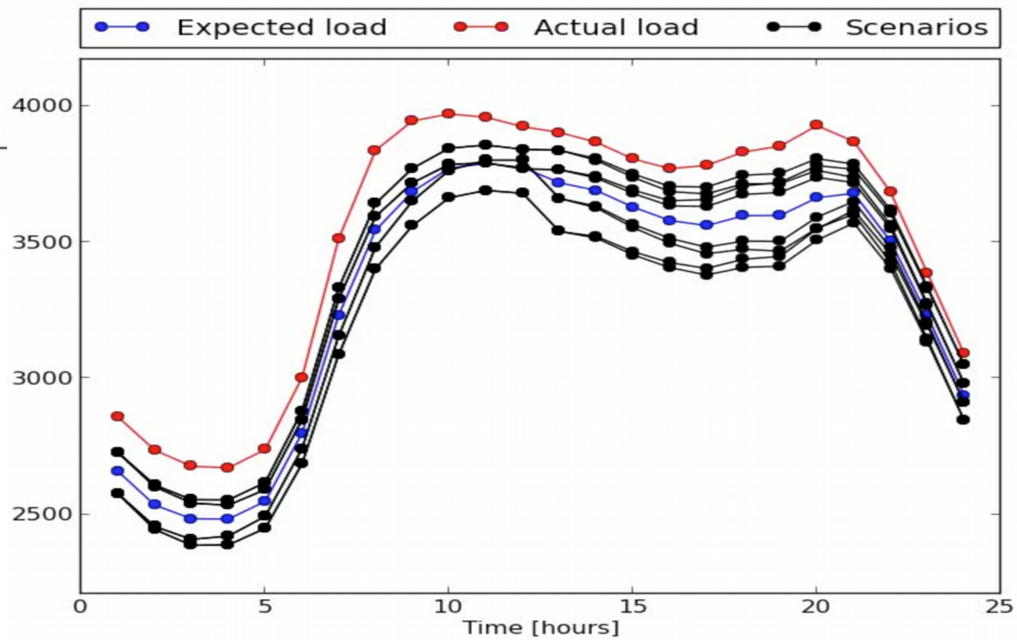
In Press at Solar Energy

Probabilistic Load Scenarios



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!



Toward Scalable Stochastic Unit Commitment

Part 1: Load Scenario Generation

Yonghan Feng · Ignacio Rios · Sarah M. Ryan · Kai Spürkel · Jean-Paul Watson · Roger J-B Wets · David L. Woodruff

Revised: November 15, 2014

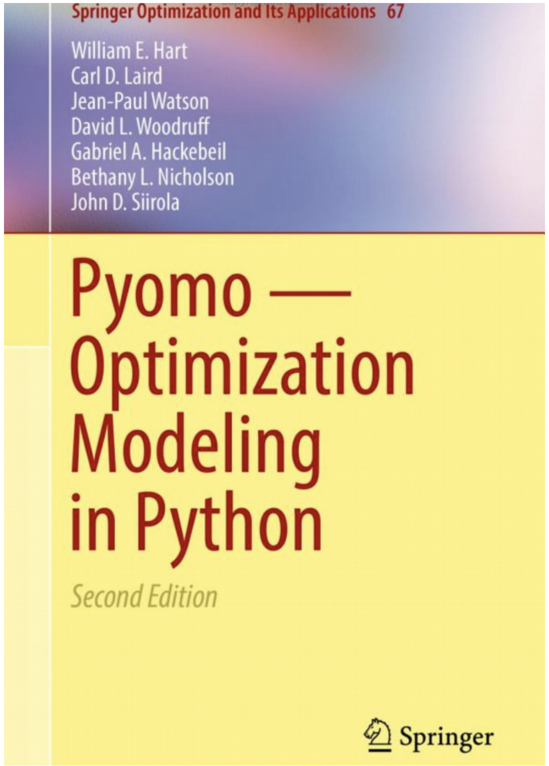
Abstract Unit commitment decisions made in the day-ahead market and during subsequent reliability assessments are critically based on forecasts of load. Traditional, deterministic unit commitment is based on point or expectation-based load forecasts. In contrast, stochastic unit commitment relies on multiple load scenarios, with associated probabilities, that in aggregate capture the range of likely load time-series. The shift from point-based to scenario-based forecasting necessitates a shift in forecasting technologies, to provide accurate inputs to stochastic unit commitment. In this paper, we discuss a novel scenario generation methodology for load forecasting in stochastic unit commitment, with application to real data associated with the Independent System Operator for New England (ISO-NE). The accuracy of the expected scenario generated using our methodology is consistent with that of point forecasting methods. The resulting sets of realistic scenarios serve as input to rigorously test the scalability of stochastic unit commitment solvers, as described in the companion paper. The scenarios generated

In Energy Systems (2015)

Part 3: On Solving Stochastic Unit Commitment

Our Software Environment: Pyomo

- Project homepage
  - www.pyomo.org
- “The Book”



- Mathematical Programming Computation papers
  - Pyomo: Modeling and Solving Mathematical Programs in Python (Vol. 3, No. 3, 2011)
  - PySP: Modeling and Solving Stochastic Programs in Python (Vol. 4, No. 2, 2012)

Toward Scalable Stochastic Unit Commitment

Part 2: Solver Configuration and Performance Assessment

Kwok Cheung · Dinakar Gade · César Silva-Monroy · Sarah M. Ryan · Jean-Paul Watson · Roger J.-B. Wets · David L. Woodruff

Received: April 30, 2014

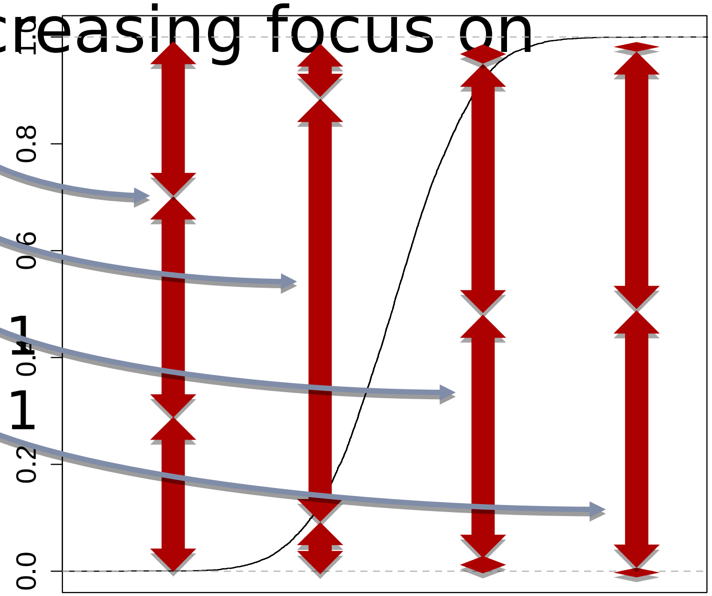
Abstract In this second portion of a two-part analysis of a scalable computational approach to stochastic unit commitment, we focus on solving stochastic mixed-integer programs in tractable run-times. Our solution technique is based on Rockafellar and Wets' progressive hedging algorithm, a scenario-based decomposition strategy for solving stochastic programs. To achieve high-quality solutions in tractable run-times, we describe critical, novel customizations of the progressive hedging algorithm for stochastic unit commitment. Using a variant of the WECC-240 test case with 85 thermal generation units, we demonstrate the ability of our approach to solve realistic, moderate-scale stochastic unit commitment problems with reasonable numbers of scenarios in no more than 15 minutes of wall clock time on commodity compute platforms. Further, we demonstrate that the resulting solutions are high-quality, with costs typically within 1-2.5% of optimal. For

In Energy Systems (2015)

Part 4: The Impact of the Nature of Probabilistic Scenarios on Stochastic Power Systems Operations

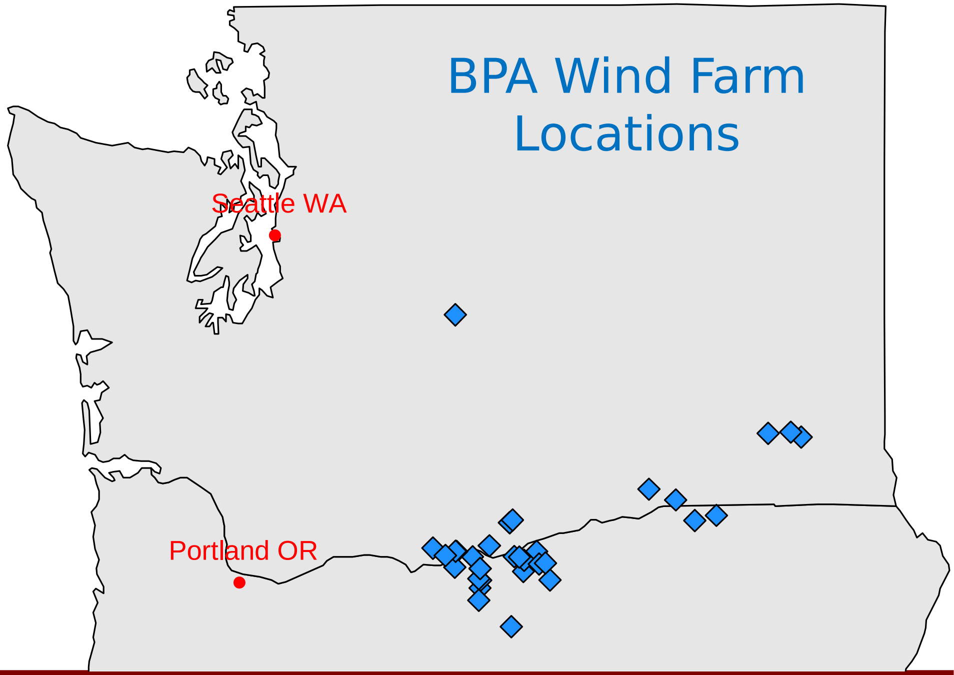
Scenario Set Comparison

- Current state-of-the-art method for scenario generation proposed by Pinson *et al.* uses quantile regression to produce a probabilistic forecast and samples from a Gaussian multivariate random variable
- We compare this to Epi-Spline scenarios using a range of cut point sets with increasing focus on ``tail" events
  - Cut points: 0 – 0.33 – 0.66 – 1
  - Cut points: 0 – 0.1 – 0.9 – 1
  - Cut points: 0 – 0.05 – 0.5 – 0.95 – 1
  - Cut points: 0 – 0.01 – 0.5 – 0.99 – 1



Application and Data

- Generate wind power scenarios using data from Bonneville Power Administration (BPA)
  - BPA has 33 wind farms, with a total capacity of 4782 MW
  - Using vendor-issued forecast data and actual power measurements from November 2015 through May 2017
  - Create day-ahead scenarios of aggregated wind power for balancing area using forecasts issued at 11am on previous day
  - Rolling horizon scenario creation, starting February 1, 2017 (with previous data used for training)

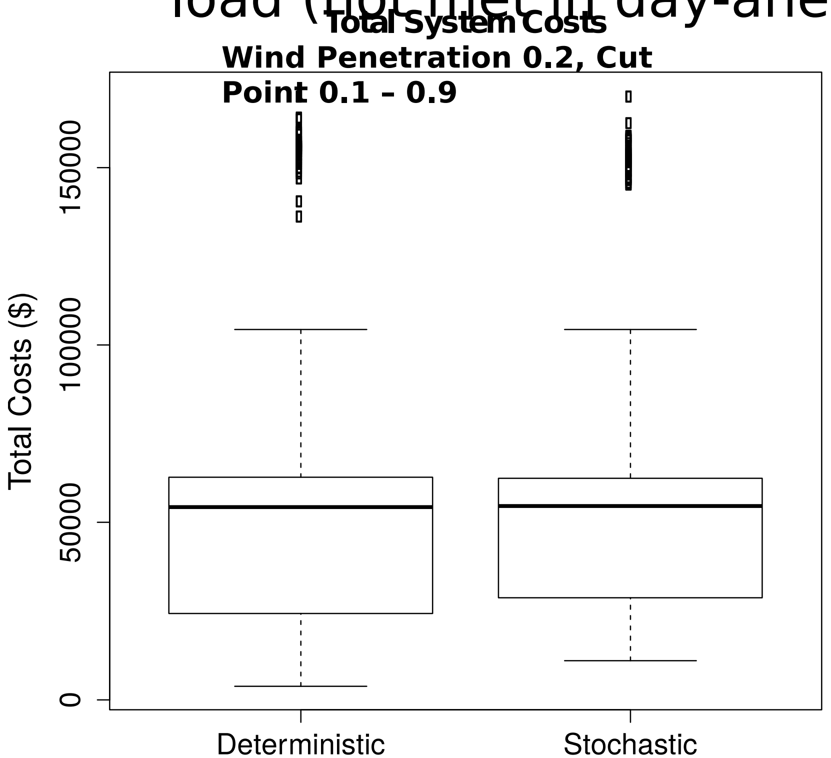


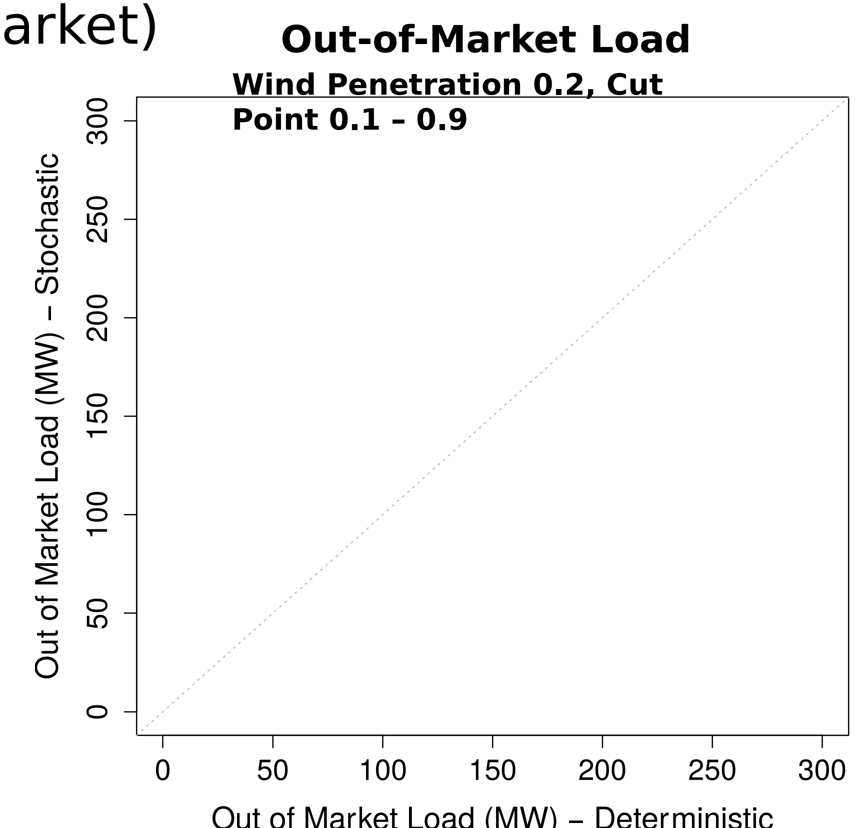
Counterfactual Re-Enactment Methodology: (Some) Details

- Stochastic day-ahead unit commitment optimization model applied to small, five-generator network (Max demand ~1400 MW)
  - Copper plate model, ignoring network flows
  - Hourly, rolling-horizon simulation with economic dispatch on the hour
  - Not carrying additional reserves, as scenarios should capture required flexibility
- Stochastic wind power scenarios use real data from BPA
  - Scale wind power to assess different wind penetration levels
  - Create day-ahead scenarios based on vendor-issued forecast, determine generator commitments, simulate system performance on realized actual wind power values
- Evaluate different scenario sets and wind penetration levels
  - Comparing cost (fixed and variable), renewables used and curtailed, over-generation, and out-of-market load

Unit Commitment Performance

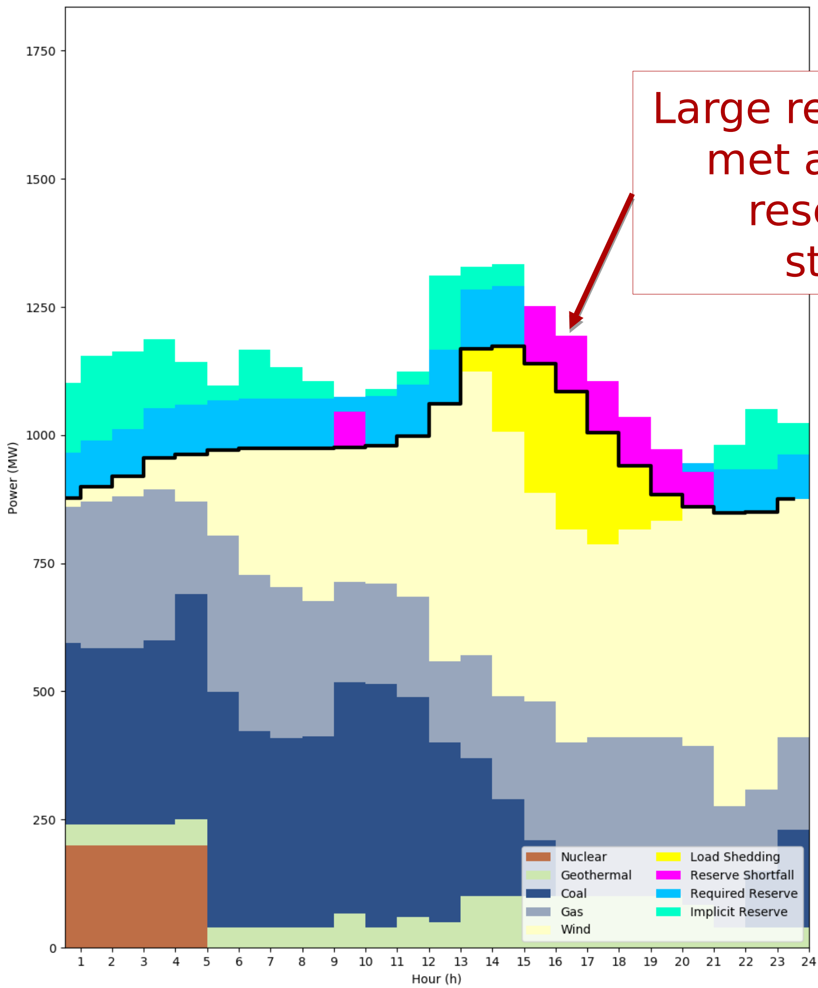
- Costs are comparable in deterministic and stochastic solutions
- However, we do not account for the cost of procuring additional generation in real-time to serve the out-of-market load (not met in day-ahead market)





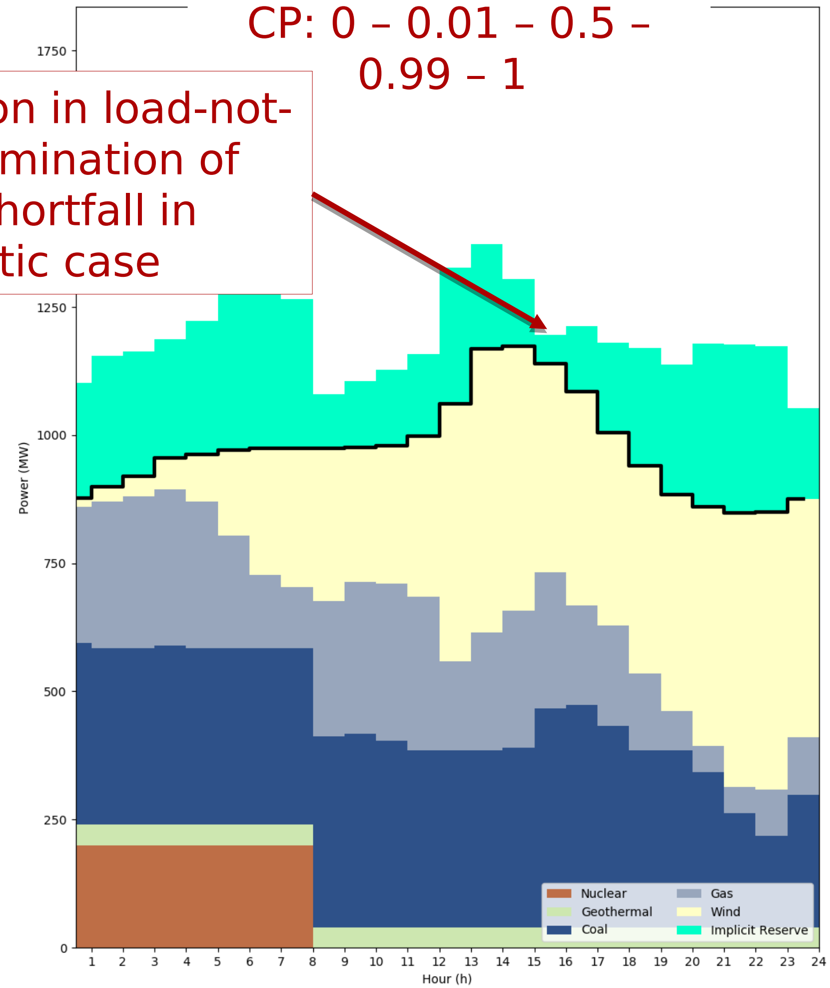
Stochastic vs Deterministic

Deterministic: 2017-03-18
CP: 0 – 0.01 – 0.5 – 0.99 – 1



Stochastic: 2017-03-18
CP: 0 – 0.01 – 0.5 – 0.99 – 1

Large reduction in load-not-met and elimination of reserve shortfall in stochastic case



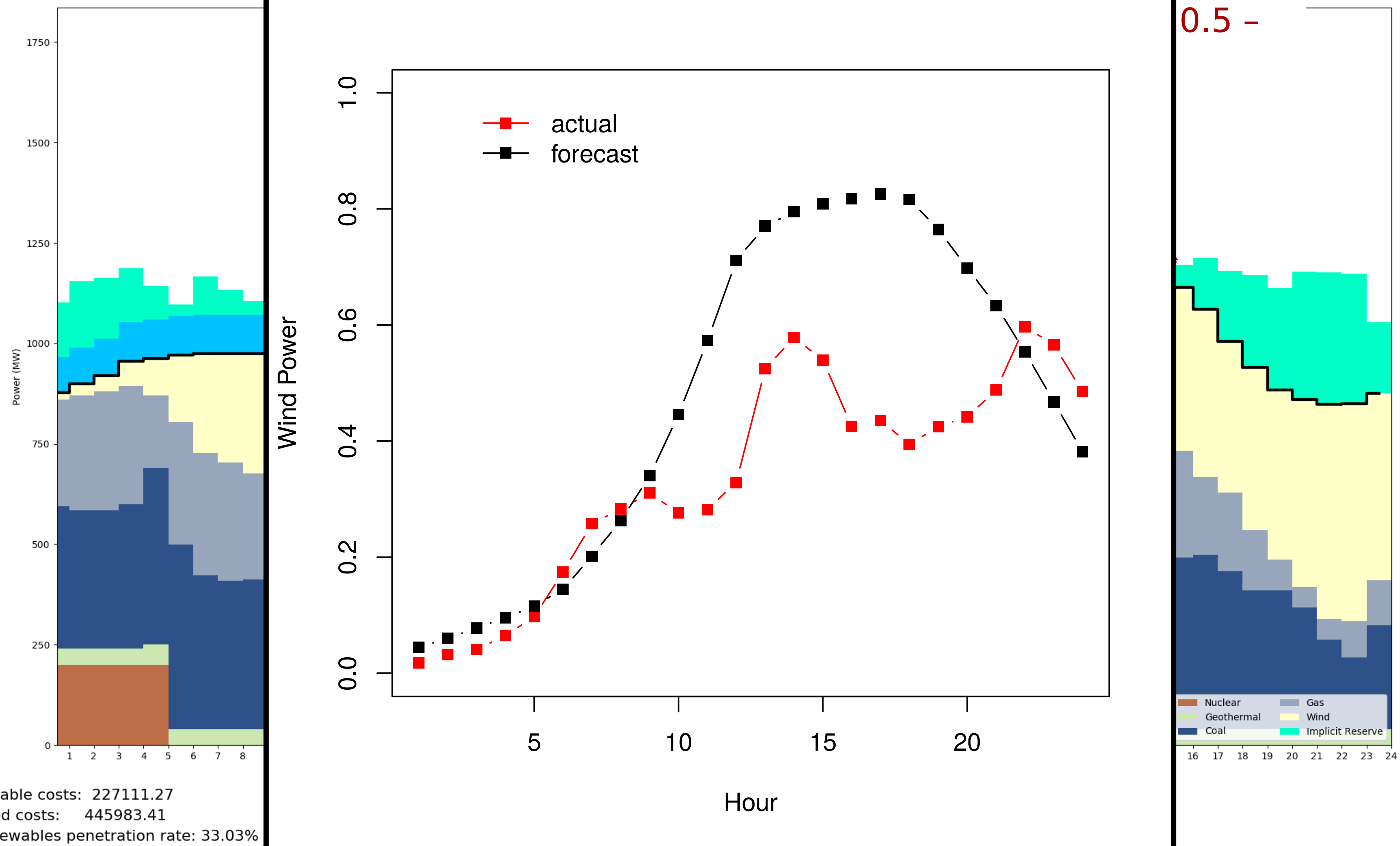
Variable costs: 227111.27
Fixed costs: 445983.41
Renewables penetration rate: 33.03%

Variable costs: 181086.81
Fixed costs: 571981.60
Renewables penetration rate: 32.88%

Stochastic vs Deterministic

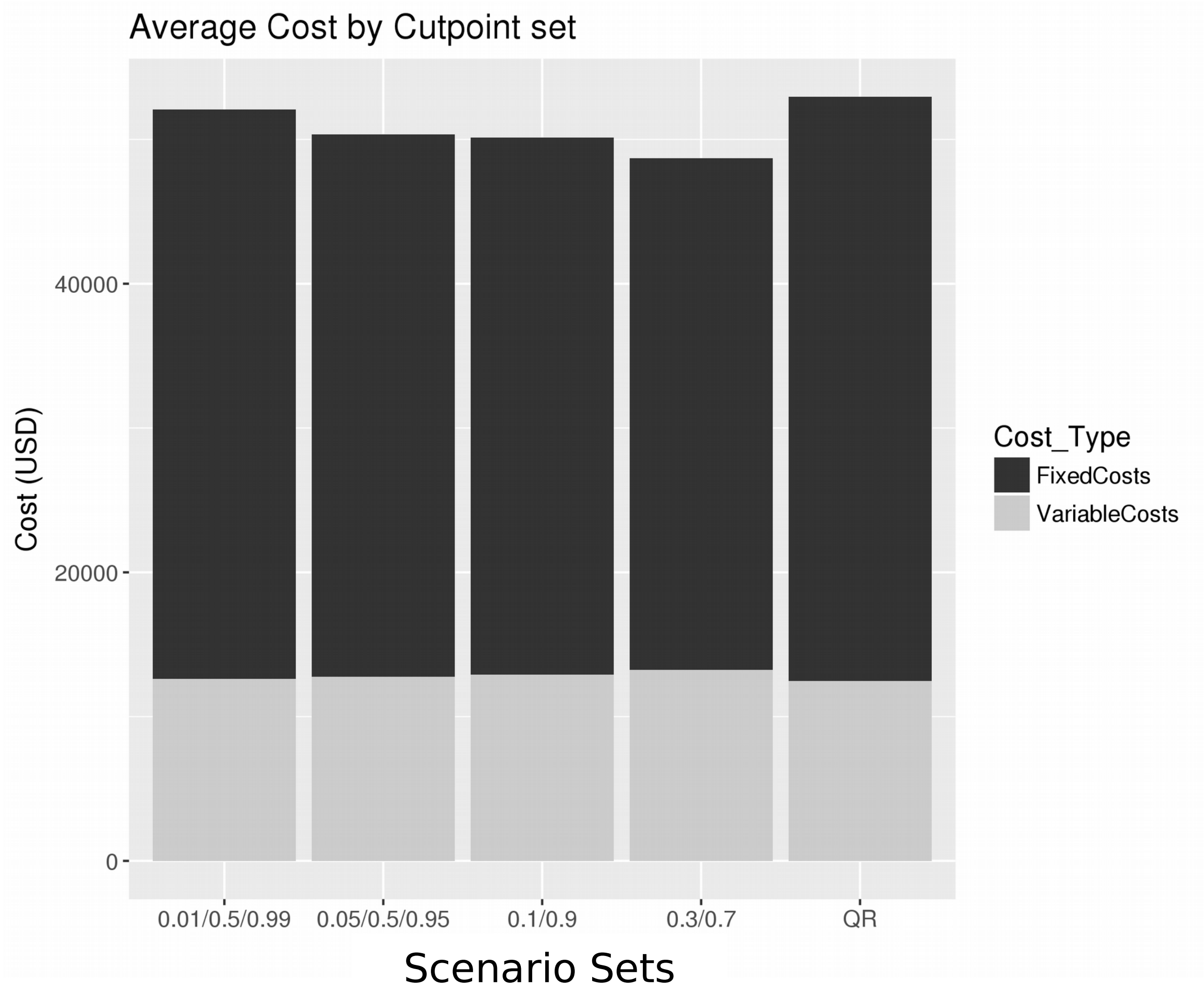
Deterministic: 2017 03 18
CP: 0 – 0.

Stochastic: 2017-03-
0.5 –



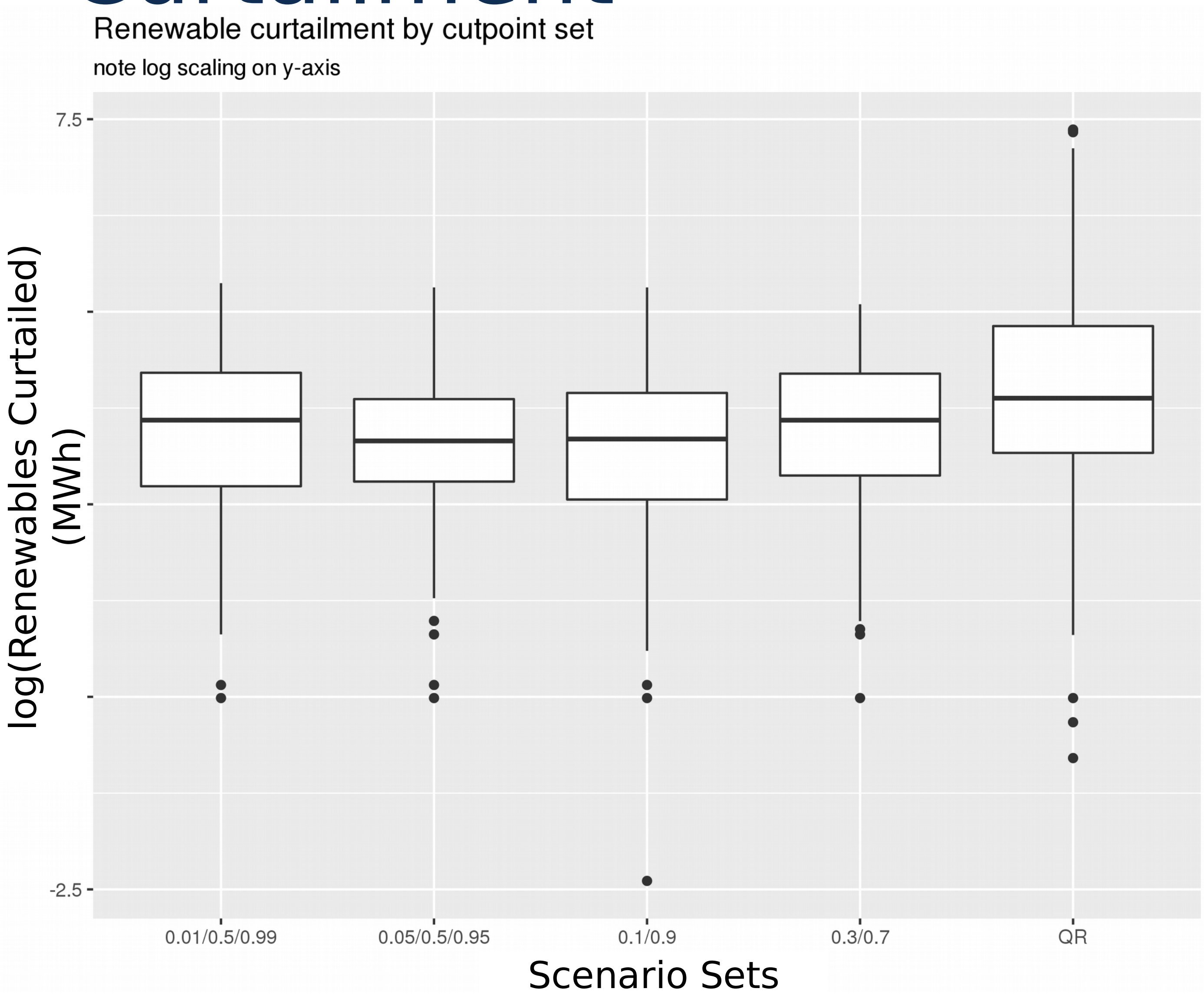
Variable costs: 227111.27
Fixed costs: 445983.41
Renewables penetration rate: 33.03%

Compare Scenario Sets: Cost



- Slight generation cost variation among scenario sets
- Wider sets have higher costs, to deal with the increased variability
- However, this doesn't account for the cost of procuring additional generation that isn't met in day-ahead scheduling

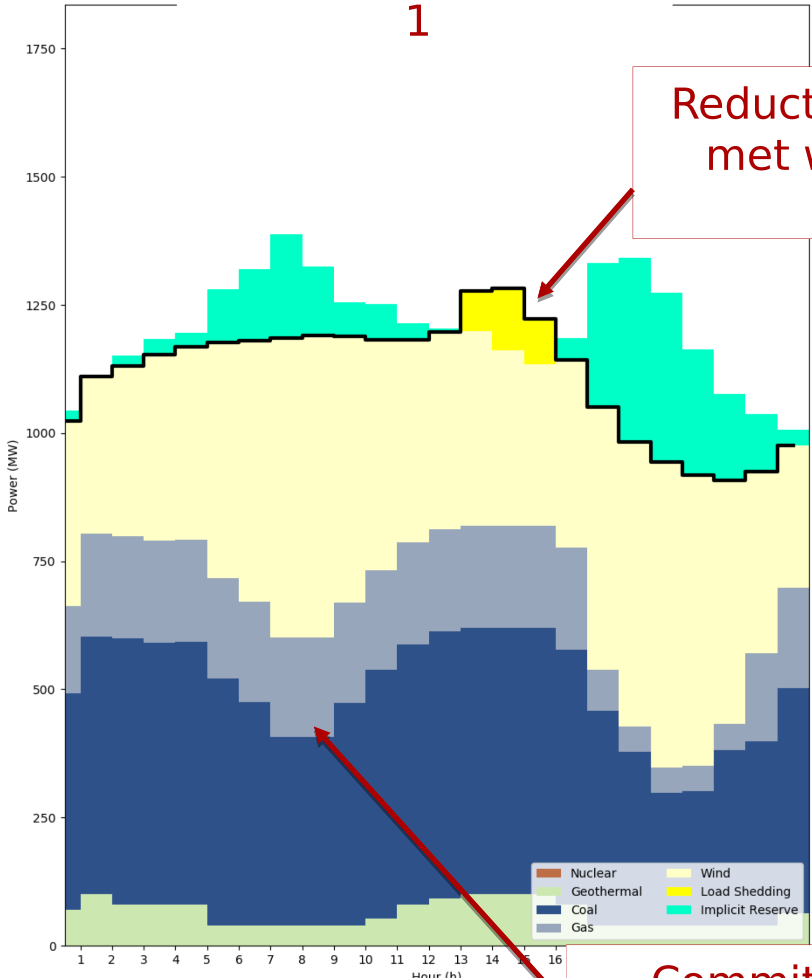
Compare Scenario Sets: Curtailment



- More curtailment with quantile regression scenarios
- Thermal generation often cannot respond fast enough for extreme ramps in wind

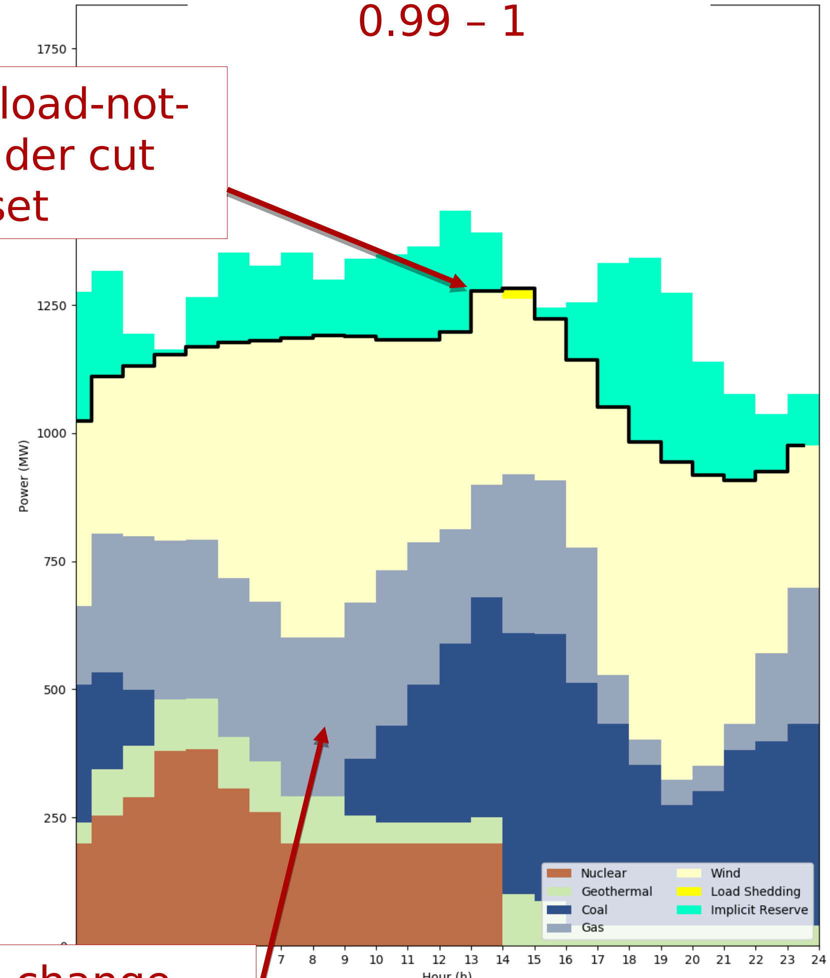
Single Day Commitments

2017-04-02
CP: 0 – 0.33 – 0.66 – 1



Variable costs: 298129.21
Fixed costs: 307123.20
Renewables penetration rate: 38.83%

2017-04-02
CP: 0 – 0.01 – 0.5 – 0.99 – 1



39.05%

Reduction in load-not-met with wider cut point set

Commitments change significantly between cut point sets

Future Work

- Evaluation of additional scenario sets
  - Assess value of scenarios that explicitly incorporate wind power ramp events
  - Look at performance of simple methods used in literature, compare to methods presented here
- Run re-enactment on larger test cases
  - Have started on WECC 240 case
  - Increase wind penetration levels to assess scenario performance at high renewable levels
- Assess performance over a longer date range
  - Incorporate more variability, both in seasonal wind and load
- Different wind dataset, if possible
  - Evaluate scenario creation methodology on additional wind sites, as ramp behavior and wind variability vary by location

Questions?

- Contact:
  - DLWoodruff@UCDavis.edu