# Stochastic Unit Commitment: Stochastic Process Modeling for Load and Renewables


David L. Woodruff, UC Davis

Sarah M. Ryan, Iowa State

Jean-Paul Watson, Sandia

# Project Goals

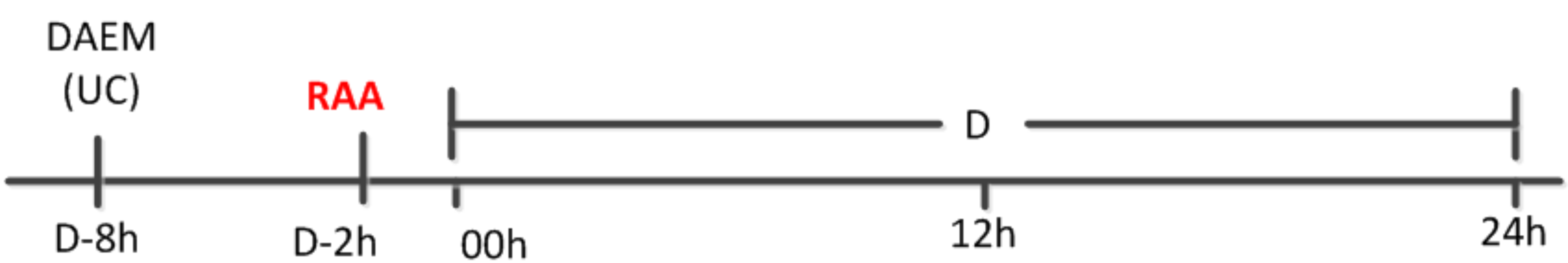
- Execute _stochastic_ unit commitment (UC) ***at scale***, ***on real-world data sets***
  - Stochastic UC state-of-the-art is very limited (tens to low hundreds of units)
  - Our solution must ultimately be useable by an ISO
- Produce solutions ***in tractable run-times***, ***with error bounds***
  - Parallel scenario-based decomposition
    - For both upper and lower bounding (Progressive Hedging and Dual Decomp.)
  - Quantification of uncertainty
    - Rigorous confidence intervals on solution cost
- Employ high-accuracy stochastic process models
  - Leveraged to achieve computational tractability while maintaining solution quality and robustness
- Demonstrate ***cost savings on an ISO-scale system*** at ***high renewables penetration levels***

# Day-Ahead Unit Commitment (SCUC D-8h)

- Day-Ahead Energy Market (DAEM or DAM)
- Clears **demand bids** and **supply offers** at 1600h on the day prior to the operating day
- Produces:
  - Hourly schedules for the next operating day for market participants (i.e., generation and demand)
  - Hourly interchange schedules
  - Hourly day-ahead Locational Marginal Prices (LMPs)
- **No reserve requirements**

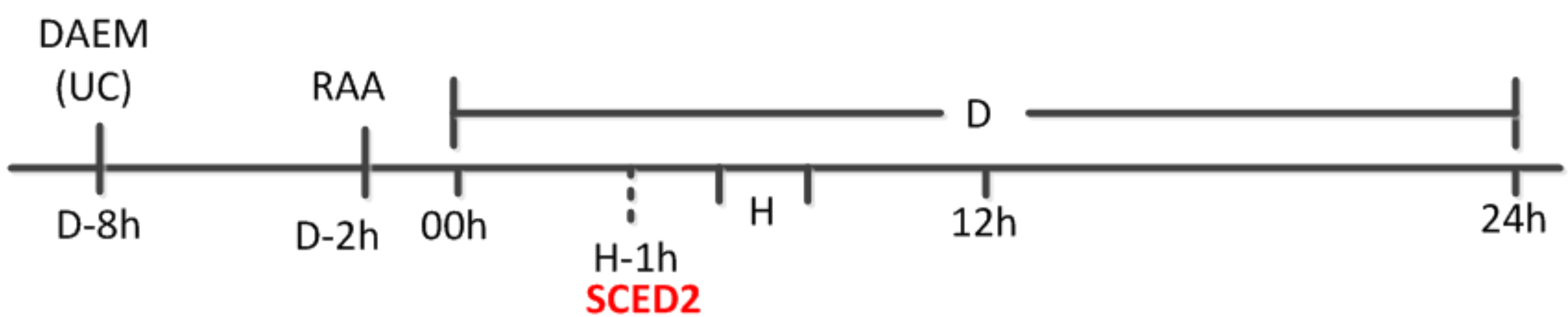
# Reliability Unit Commitment - RUC (SCUC D-2h)

- Reliability Assessment (Reserve Adequacy Analysis - RAA)
- Minimize additional start-up and no load costs to provide sufficient capacity to satisfy the forecasted load plus the **operating and replacement reserve requirements**
- Clears ISO **forecasted load** at 2200h
- DAM commitments are respected
- Produces:
  - Additional commitments
  - Updated generator dispatch points



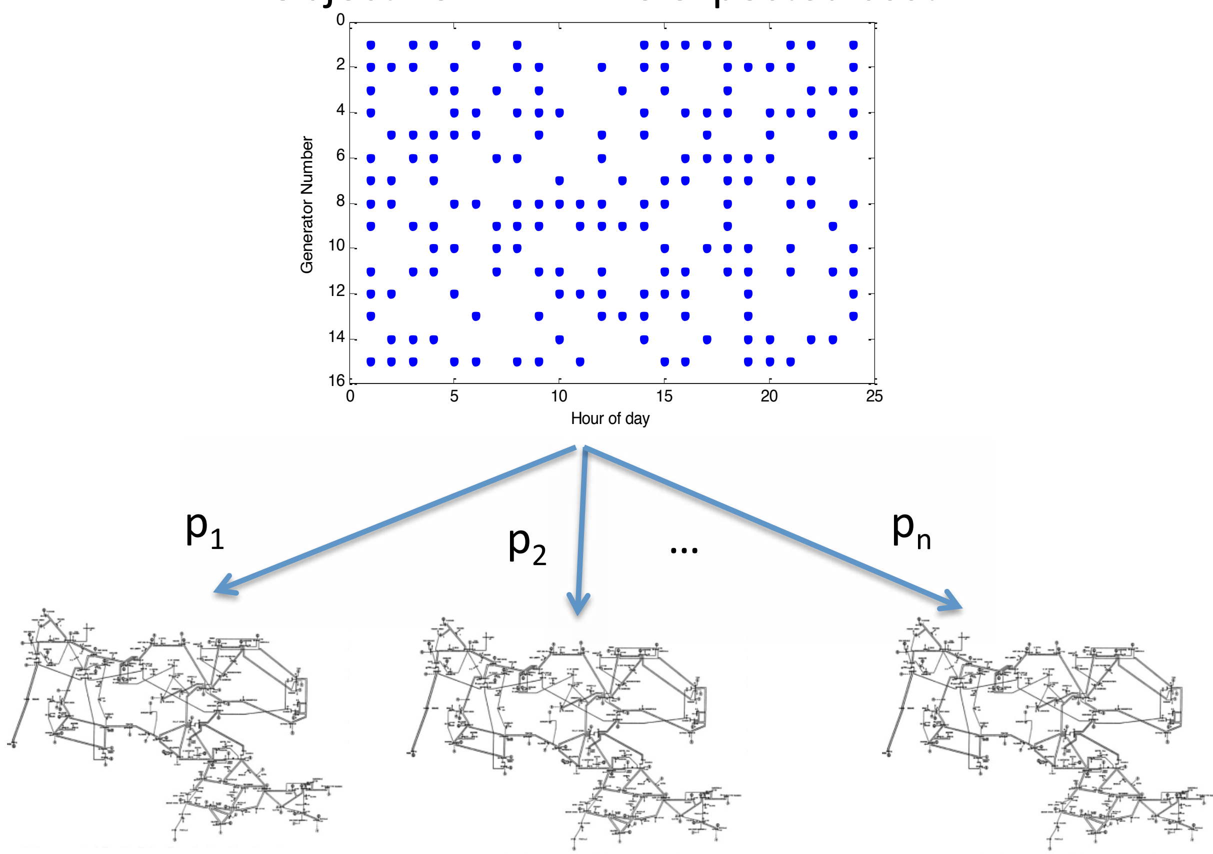
# Look-Ahead SCED (H-1h)

- SCED with ability to bring online fast start resources
- Intended to meet intra-hour reserve requirements
- Updated load and variable generation forecasts
- It produces:
  - Generator setpoints
  - Commitment of fast start units



# General UC Model Structure

Objective: Minimize expected cost




Scenario 1 Scenario 2 ... Scenario n

First stage variables:

- Unit On / Off (*per hour*)

Nature resolves uncertainty

- Renewables output
- Forced outages

Second stage variables (*per hour*):

- Generation levels
- Power flows
- Voltage angles
- ...

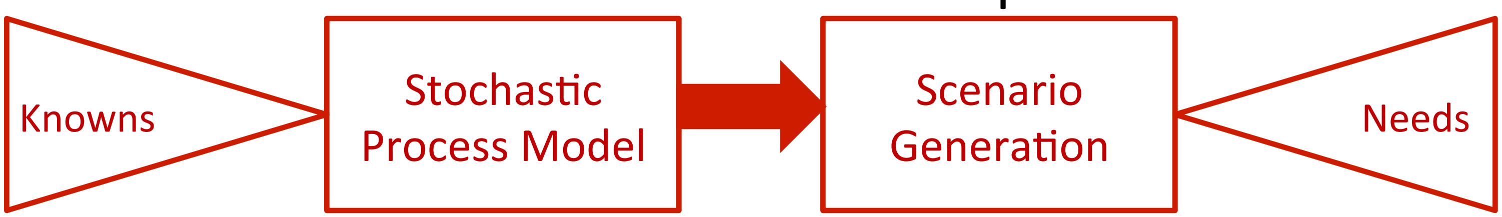
# RUC: Knowns vs. unknowns

**Afternoon of day *D*-1, ISO knows:**

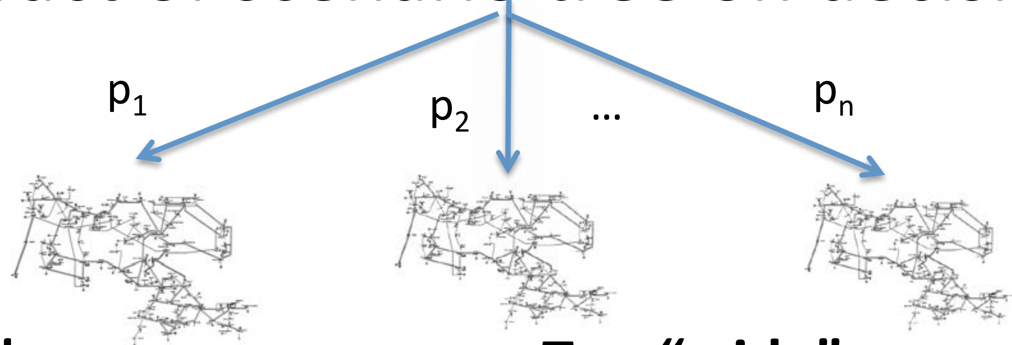
- Past hourly weather
- Past hourly loads
- Past hourly renewable generation
- Weather forecast for day *D*
- "Type" of day *D*

**For stochastic UC model, ISO needs:**

- Multiple scenario paths: 24 (at least) hours' worth of
  – Loads → focus of this talk
  – *Variable generation availability*
  – *Thermal unit availability*
- Probabilities of occurrence
  – *Joint for multiple variables*
- Temporal tree structure



# Impact of scenario tree on decisions



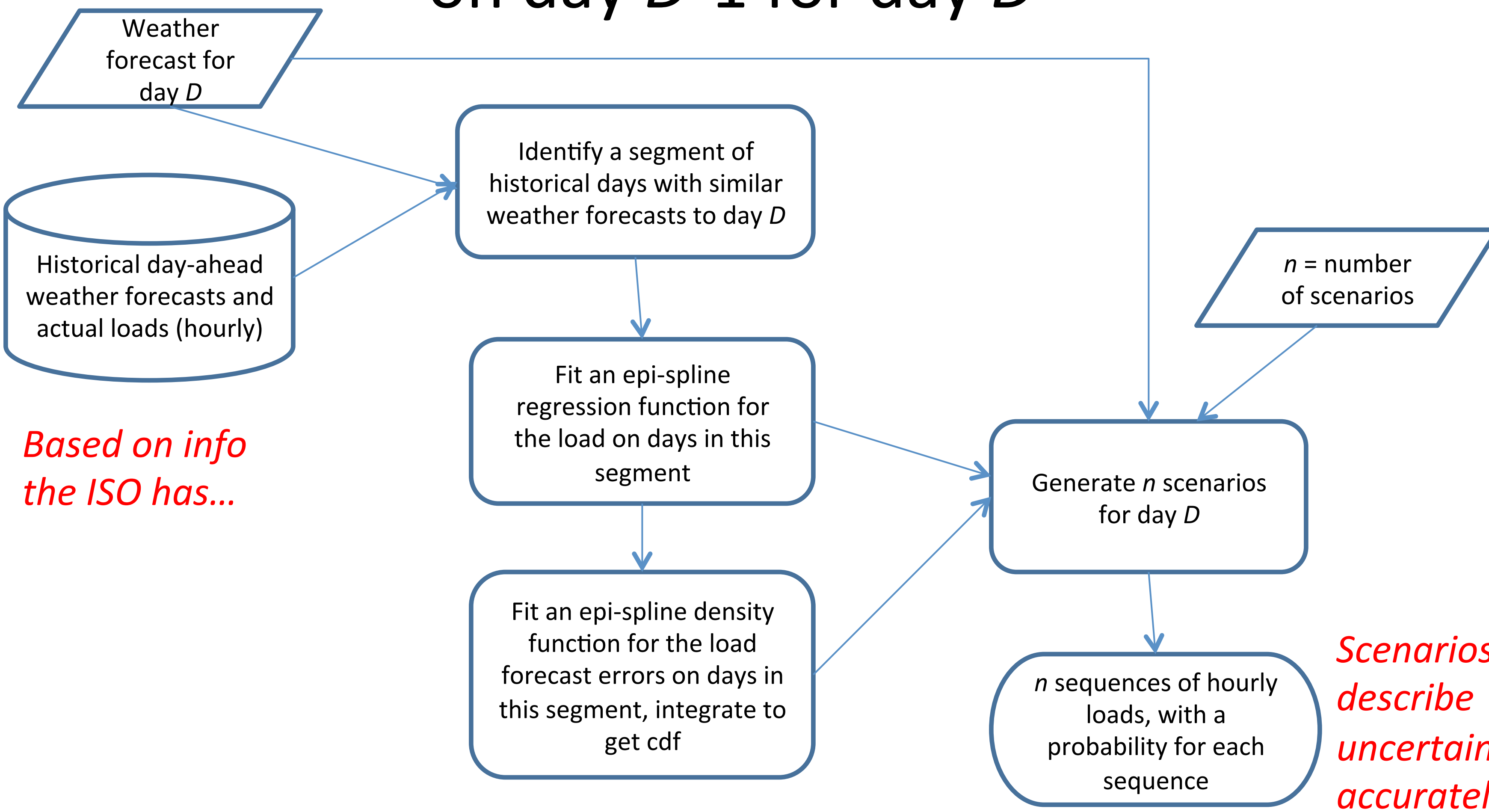

**Too "narrow"**

- Optimization fails to account for actual risks
- Too few low-cost units committed
  - Cost: Start up additional high-cost units
  - Reliability: Shed load

**Too "wide"**

- Optimization result is too risk-averse
- Too many low-cost units committed
  - Cost: Excessive no-load cost of committed units
  - Environmental: curtail variable generation

# Overview of process to generate load scenarios on day *D*-1 for day *D*

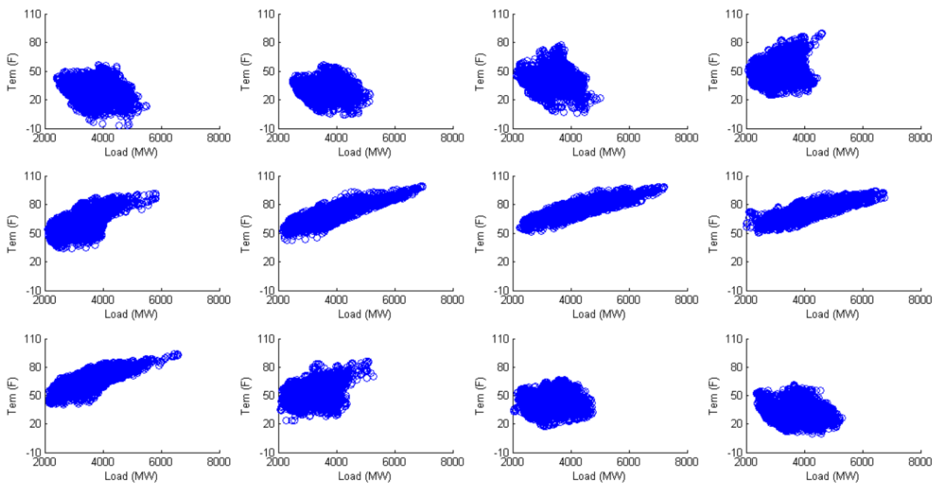



*Based on info the ISO has…*

*Scenarios describe uncertainty accurately*

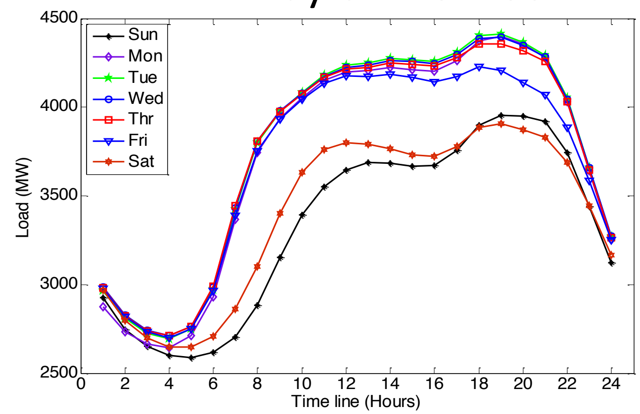
# Influences on load

## Season: A/C or heat, diurnal light patterns



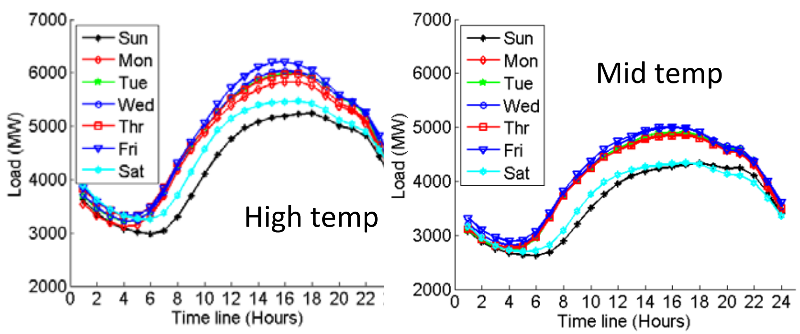

Scatter plots of load vs. temperature by month in CT

## Day of the week

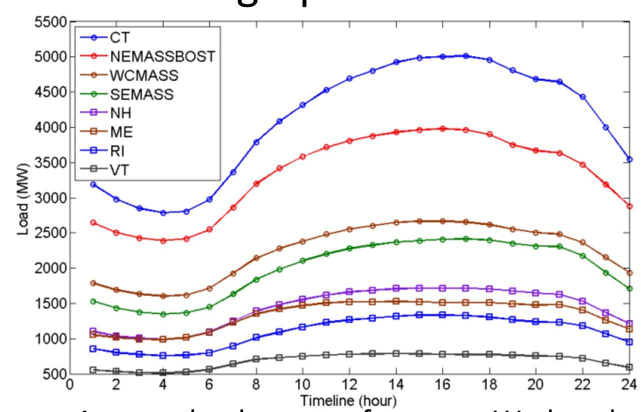



Average load curves of summer days in CT

## Weather, mainly temperature



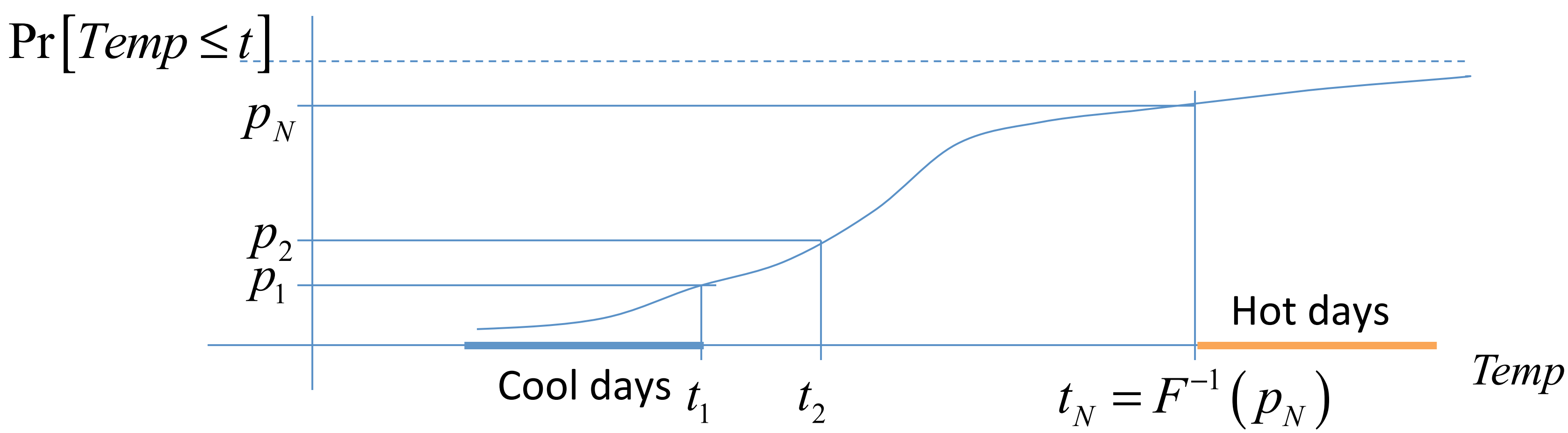

## Geographic location




Average load curves of summer Wednesdays

# Identify data segment

- Separate by seasons = date ranges
  - Diurnal light patterns, heating vs. cooling, impact of wind and humidity (RealFeel temperature)
- Within a season
  - Transform every day to “Wednesday” based on average load patterns
  - Transform zones to a “master zone”
  - Segment days by partition of temperature forecast distribution (average, or a particular hour)



# Fit epi-spline regression function

- Data: (forecast temp, forecast dewpt, actual load) in hour *h*, day *j* $\left(t_h^j, d_h^j, l_h^j\right)$
- Model: $l(r) = z^t(r)t(r) + z^d(r)d(r)$

  where *l*(*r*), *t*(*r*), *d*(*r*) are *actual* load and *forecast* weather variables as functions of continuous time, *r*, and the regression functions $z^t(r), z^d(r)$ are twice-differentiable.
- Approximate $z^t(r), z^d(r)$ with epi-splines $s^t(r), s^d(r)$ that have piecewise constant second derivatives on $\{0,\ldots,r_{k-1},r_k,\ldots,r_N \equiv 24/\delta\}$
- Integrate twice to get

$$s(r) \equiv s_0 + v_0 r + \delta \sum_{i=1}^{k-1}(r - r_i + \delta/2)a_i + (1/2)(r - r_{k-1})^2 a_k,\ r \in (r_{k-1}, r_k]$$

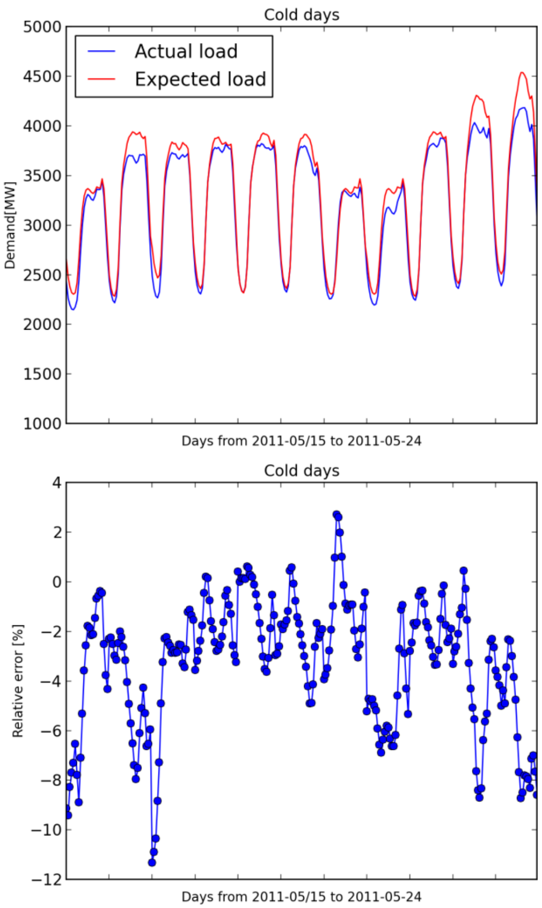
- Measure errors as $e_h^j = l_h^j - s^t(h)t_h^j + s^d(h)d_h^j$
- Optimization problem

$$\min_{s_0, v_0, \{a_i, i=1,\ldots,N\}} \left\| e = \left(e_h^j, h = 1,\ldots,24; j \in J\right) \right\|_p$$

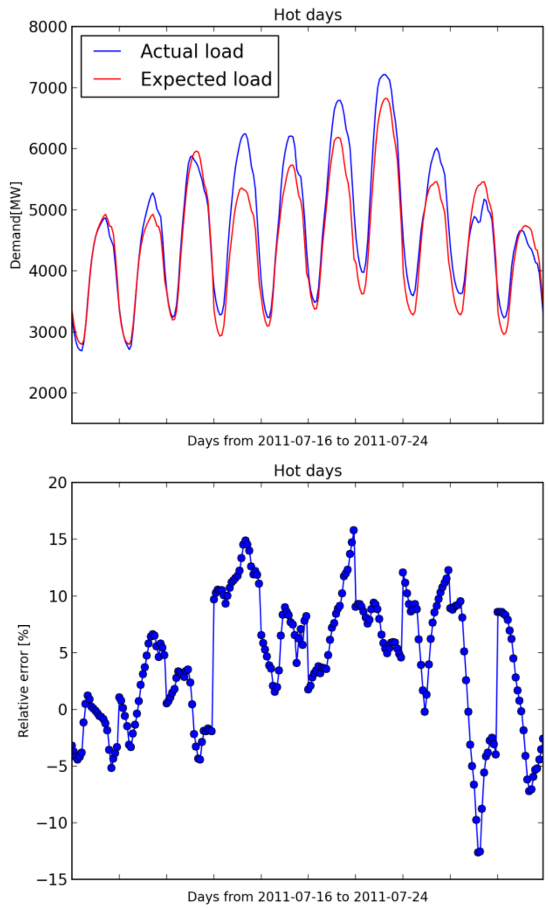
# Advantages of the epi-spline regression

- Rich family of possible curves, not just polynomials
- Nonparametric estimation of hourly load patterns
- Does not involve lagged loads
- No assumptions about error distributions; e.g., white noise
- Can add constraints based on "soft information" – compensate for small segmented data sets
  - Values: do not underestimate peak loads
  - Slopes: understand daily patterns of increase/decrease
  - Curvature: (so far, bounds have not had much impact)

# Load fitting results




May, 2011, in CT




Aug, 2011, in CT

Absolute daily error

$$e^j = \sum_h \left| e_h^j \right|$$

Relative daily error

$$e_{rel}^j = e^j \Big/ \sum_h l_h^j$$

Average relative daily error

$$\mathrm{ARDE} = \sum_{j \in J} e_{rel}^j \Big/ |J|$$

Root mean squared deviation

$$\mathrm{RMSD} = \sqrt{\tfrac{1}{|J|} \sum_{j \in J} \left( e_{rel}^j - \mathrm{ARDE} \right)^2}$$

| Season | ARDE (%) | RMSD |
|---|---|---|
| Spring | 1.0952 | 0.9021 |
| Summer | 3.4248 | 1.9614 |
| Fall | 2.6202 | 3.2381 |
| Winter | 2.5809 | 1.5652 |

# Obtain error distribution for each hour

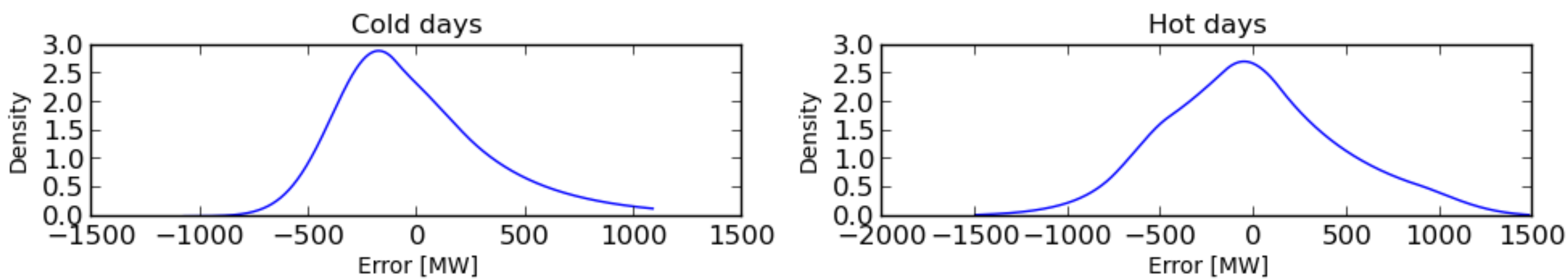
- For hour *h*, compute mean and standard deviation of errors in the segment
- Let $\alpha = \min\left\{\left(e^j, j \in J\right), \bar{e} - 3\sigma_e\right\}, \beta = \max\left\{\left(e^j, j \in J\right), \bar{e} + 3\sigma_e\right\}$
- Approximate error density as

$$f(x) = e^{-w(x)}, x \in [\alpha, \beta]$$

where *w*(*x*) is an epi-spline having piecewise constant second derivatives $a_k \in [0, \kappa]$

- For numerical reasons, translate domain to [0, *b*-*a*] and then rescale to [0,1]
- Maximize likelihood of the observed errors
  - Convex objective function
  - Linear constraints
- Integrate density to obtain cumulative distribution function

# Error densities for an hour




- Unimodal by constraint
- Not necessarily symmetric or centered at zero
- Vary by day segment and hour
- Incorporate both weather forecasting and load modeling errors

# Scenario Generation Workflow

David L. Woodruff
Graduate School of Management
UC Davis
Davis CA USA 95616
DLWoodruff@UCDavis.edu

with
Roger J-B Wets
Jean-Paul Watson
Kai Spürkel
Sarah Ryan
Ignacio Rios
Yonghan Feng

# Simulation Support

- We provide support for a deterministic simulation of stochastic optimization applied to 2011 to estimate the potential energy cost savings versus deterministic optimization.
- We will give you a flavor of the work to obtain and evaluate scenarios.
- We hope to publish some scenarios in the next few months.

# Load, Wind, Outages

- We can forecast load based on weather forecasts and we use that forecast technology to generate load scenarios.
- We use 2011 actuals as actuals.

- We use 3Tier Analogs to get wind power scenarios.
- We use EWITS 2011 wind power as actuals.

- We generate outage scenarios based historic outage rates.
- We simulate actuals.

# Left and right sides

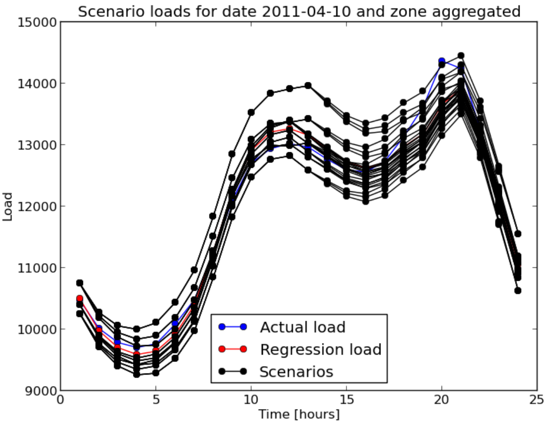
- Left side: build forecast models and analyze errors
- Right side: create scenarios

# Scenario generation

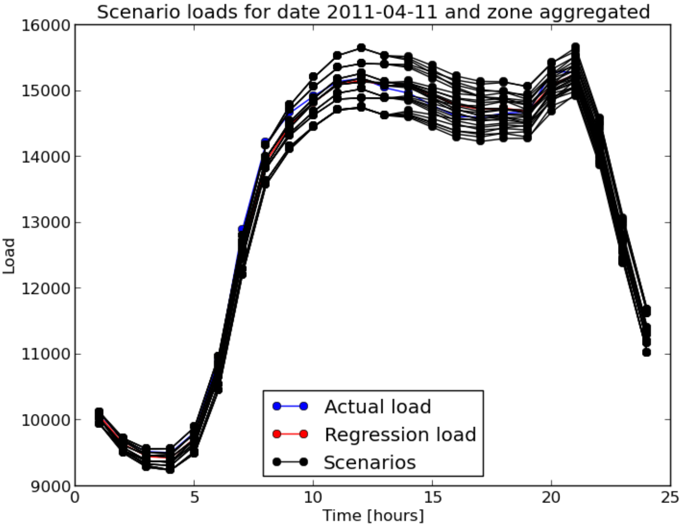
## Objectives

- Small number of scenarios
- Cover almost all possible outcomes
- Have a distribution similar to the observed distribution
- (so these are not Monte-Carlo samples)

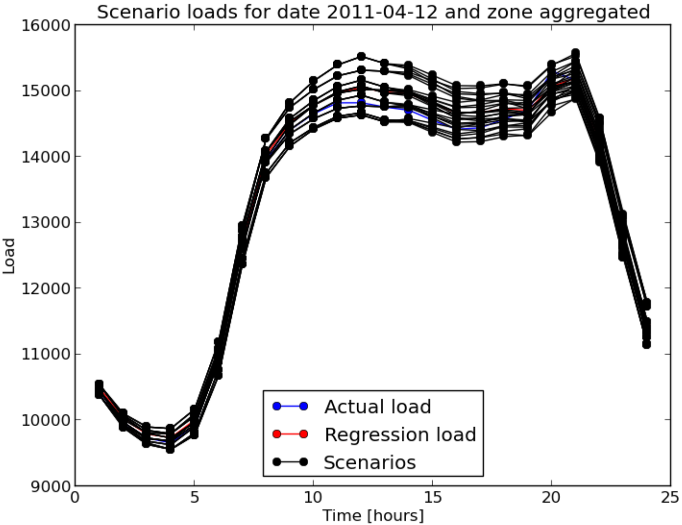
# Example with 27 Scenarios



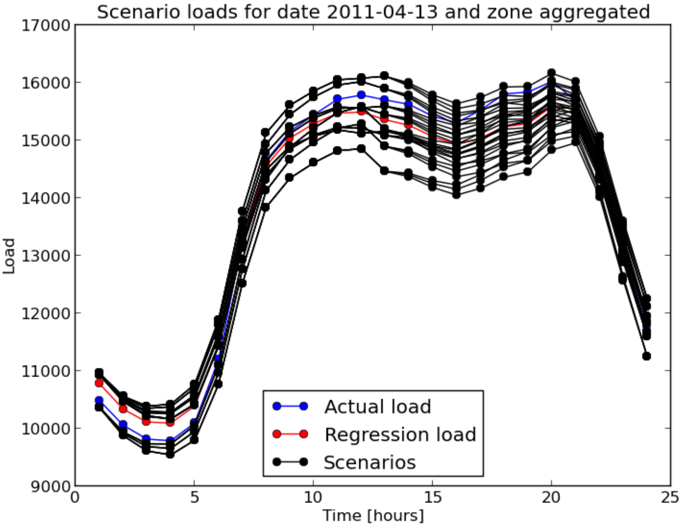
# Another Example with 27 Scenarios



# Yet Another Example with 27 Scenarios

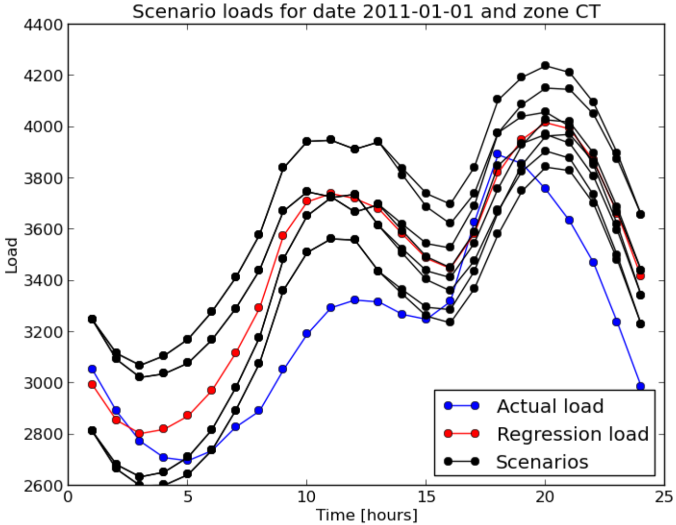


# Last Example with 27 Scenarios



# 8 Scenarios for January 1

(but ignore its special nature, so these are not very good)



# Measuring quality of scenarios

## Introduction

The next issue is to measure the quality of our scenarios $S$.

- Compare different methods to generate scenarios
- Compare different sets of parameters
- Find the minimum amount of scenarios that are sufficient to obtain similar results
- Many other...

# Measuring quality of scenarios

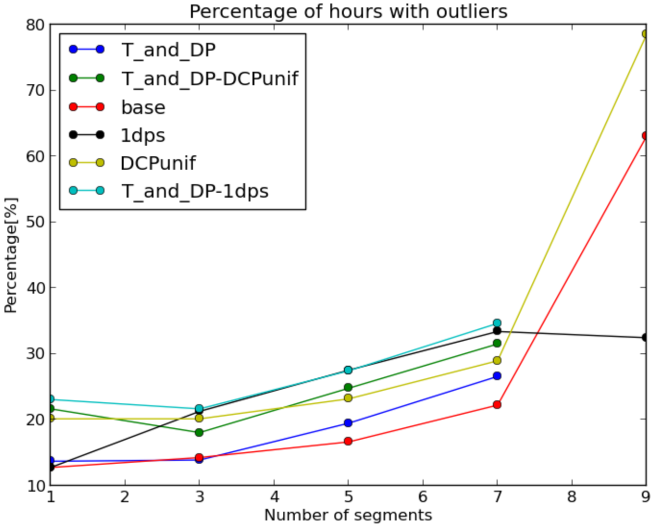
## Introduction

How do we measure the quality of our scenarios? Looking at scenario properties:

- Distance between the scenarios and the observed distribution
- Number of outliers
- Distance from the observed value to the closest scenario
- Distance from the observed value to the farthest scenario

# Sample of Outlier Study

## Spring



# Conclusion

- We have given an overview of the methods that an ISO would use to generate scenarios for day-ahead hourly demand.
- We have briefly touched on how we evaluate those scenarios.
- Our immediate goal is to provide input to a stochastic SCUC that is similar to the input that an ISO would provide in practice.