Advanced Stochastic Solutions for Management of Uncertainty: Incorporating Storage and Scenario Generation

C. Lindsay Anderson
Luckny Zephyr
Laurie L. Tupper

CERTS R&M Review
Washington DC
June 9-10, 2016

CERTS
Consortium for Electric Reliability Technology Solutions

CORNELL UNIVERSITY FOUNDED A.D. 1865

Motivation

With increasing participation of variable and uncertain resources on both sides of the power system, operational decisions require stochastic methods. Challenges include:

- Characterizing uncertainty, scenario selection
- Computational tractability, large networks
- Flexibility,
  - for different types of uncertainty (wind, solar, responsive demand)
  - for inter-temporal constraints of energy storage

Overview

Scenario Selection
(Band Depth Clustering)

Scalable Stochastic Unit Commitment
(Chance-constrained Programming)

Incorporating Storage Decisions in Dispatch
(Approximate Dynamic Programming)

Presentation Outline

We then will present progress on:

1. Integration of storage through stochastic dual dynamic programming (SDDP); and,
2. Testing of scenario reduction algorithms using the Multi-period Optimal Scheduling Tool (MOST).

Energy Storage

With the growth of renewable generation, energy storage has been proposed as a method for managing uncertainty.

- Most of previous research:
  - Deals with a single storage unit and a single wind farm; and
  - Do not consider power network constraints.
- It is clear that:
  - network constraints increase the size of the problem, and therefore its complexity; and
  - the number of storage and wind farm units increase the search space;
  - a critical issue, in particular in the context of the intertemporal constraints, which are best dealt with by *stochastic dynamic programming*.

Economic Dispatch Problem with Storage

We consider an *economic dispatch* problem for a network comprising $|G|$ conventional generators, $|M|$ wind farms and $|S|$ storage. We aim to find the generation and storage policy that minimizes the operating cost over of finite horizon of T hours:

$$\min \left\{ \mathbb{E} \left[\sum_{t=1}^{T} c_t \left(\cdot \right) \right] \right\}$$

Subject to, for $1 \leq t \leq T$:

$\Phi(p_t, d_t, w_t) = 0$	Power balance equations
$\Theta(\theta_t, e_t) = 0$	Power flow equations
$\Xi(s_t, \Delta_t) = 0$	Storage dynamics
$\underline{\Psi}_t \leq \Psi_t \leq \overline{\Psi}_t$	Box and ramping constraints

CERTS
Consortium for Electric Reliability Technology Solutions

Representation under Stochastic Dynamic Programming

- Previous problem is a sequential decision problem.

$$\boxed{\text{Current state}} \xrightarrow{\text{New information}} \boxed{\text{Decisions}} \xrightarrow{\text{Transition filter}} \boxed{\text{Next state}}$$

Figure: Illustration of the information decision structure

- Stochastic dynamic programming (SDP) is then suited to the problem.
- SDP decomposes (by period) the problem into subproblems in a coordinated way.
- For each possible state, an optimization problem is solved seeking for the best trade-off between utilizing the resources "today" and leaving them for the future.

CERTS
Consortium for Electric Reliability Technology Solutions

SDP Formulation

For $t = T, T-1, \cdots, 1$:

$$F_t(s_t, p_{t-1}, w_{t-1}) = \min\left\{c_t(\cdot) + \mathbb{E}\left[F_{t+1}(s_{t+1}, p_t, \tilde{W}_t)\right]\right\}$$

$$\begin{aligned} \text{S.t. } & \Phi(p_t, d_t, w_t) = 0 \\ & \Theta(\theta_t, e_t) = 0 \\ & \Xi(s_t, \Delta_t) = 0 \\ & \underline{\Psi}_t \leq \Psi_t \leq \overline{\Psi}_t \end{aligned}$$

- $F_{t+1}(s_{t+1}, p_t, w_t)$, called *cost-to-go* is the cost from period $t+1$ through the end of the horizon.
- Two sources of complexity:
  - Computation of an expectation
  - Optimization step for **each** *state value* (s_t, p_{t-1}, w_{t-1})

Curse of dimensionality

- The problem cannot be solved for all discrete state values (s_t, p_{t-1}, w_{t-1}). Example:
    - With 5 wind turbines, 5 storage devices and 5 generators;
    - Each dimension discretized into 10 levels (in each time period);
        - In total $10^5 \times 10^5 \times 10^5 = 10^{15}$ grid points.

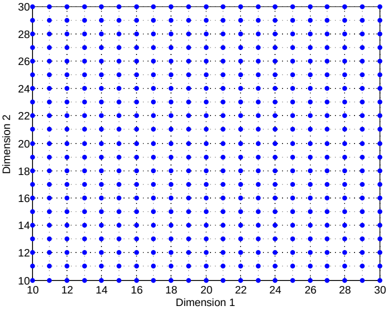


Figure: Example of a two-dimension grid

Stochastic Dual Dynamic Programming

- SDDP does not discretize the state space, rather samples it.
- Replace $\mathbb{E}\left[F_{t+1}(s_{t+1}, p_t, \tilde{W}_t)\right]$ (assumed to be convex) with some **lower bound** $\hat{V}_{t+1}(s_{t+1}, p_t, w_t)$.

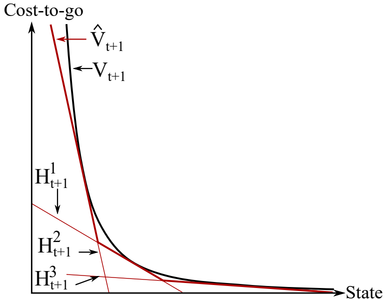


Figure: Illustration SDDP approximation

- $\hat{V}_{t+1}(s_{t+1}, p_t, w_t) := \max_i \{H^i_{t+1}(s_{t+1}, p_t, w_t) | i \in I\}$

SDDP Approximation of the Problem

For $t = T, T-1, \cdots, 1$

$$\hat{F}_t(s_t, p_{t-1}, w_{t-1}) = \min \left\{ \sum_{t=1}^{T} c_t(\cdot) + \rho_{t+1} \right\} \tag{1}$$

$$\text{S.t. } \Phi(p_t, d_t, w_t) = 0 \tag{2}$$

$$\Theta(\theta_t, e_t) = 0 \tag{3}$$

$$\Xi(s_t, \Delta_t) = 0 \tag{4}$$

$$\underline{\Psi}_t \leq \Psi_t \leq \overline{\Psi}_t \tag{5}$$

$$\rho_{t+1} \geq \tilde{c}^i_{t+1} + \tilde{g}_{s^i_{t+1}} s_{t+1} + \tilde{g}_{p^i_t} p_t + \tilde{g}_{w^i_t} w_t, \ 1 \leq i \leq I \tag{6}$$

$[\tilde{c}^i_{t+1}, \tilde{g}_{p^i_t}, \tilde{g}_{w^i_t}]$ are computed in period $t+1$ using dual prices.

CERTS
Consortium for Electric Reliability Technology Solutions

Algorithm Scheme

Sequence of backward and forward passes:

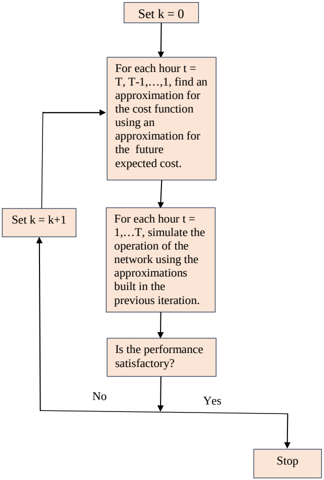


Figure: SDDP algorithm to build the approximations and sample the state space

Results Overview

- Illustration on IEEE 9-bus network
- Comparison with SDP on IEEE 9-bus network
- Scalability testing on different network sizes

Optimal storage strategy

IEEE 9-bus network

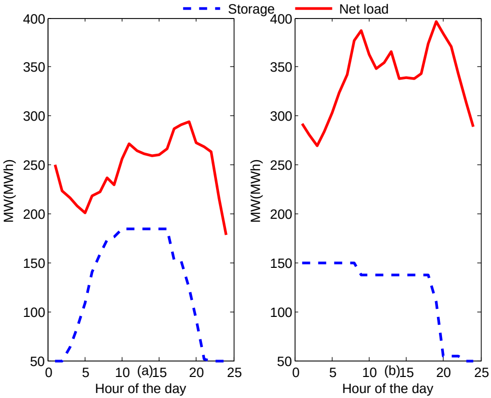


Figure: Example of storage trajectory when charging (discharging) cost is low

Optimal storage strategy (cont.)

IEEE 9-bus network

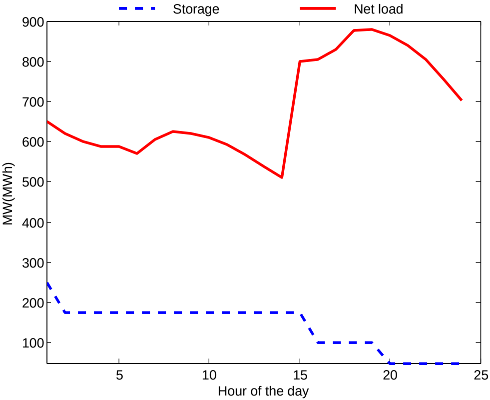


Figure: Example of storage trajectory when charging (discharging) cost is high

Comparison with SDP

IEEE 9-bus network : 100 simulations each run

Table: CPU time in seconds: SDDP and SDP

Method	Run 1	Run 2
SDDP	499.35	1 876.22
SDP	13 038.16	12 726.88

Table: Solution cost: SDDP and SDP

Method	Run 1			
	Min	Max	Mean	Stand. deviation
SDDP	78 622.29	162 215.38	126 872.83	21 812.53
SDP	75 392.76	161 872.07	125 968.26	22 922.10
Method	Run 2			
	Min	Max	Mean	Stand. deviation
SDDP	88 691.21	164 333.09	134 267.42	19 210.77
SDP	85 882.05	164 333.09	133 981.13	19 657.52

Scalability Testing Networks:

Table: Test networks' characteristics

Case	# of buses	# of generators	# of trans. lines
1	30	6	41
2	57	7	80
3	89	12	210
4	118	54	186
5	300	69	411

Example of Optimal Storage Strategy

IEEE 118 bus-network

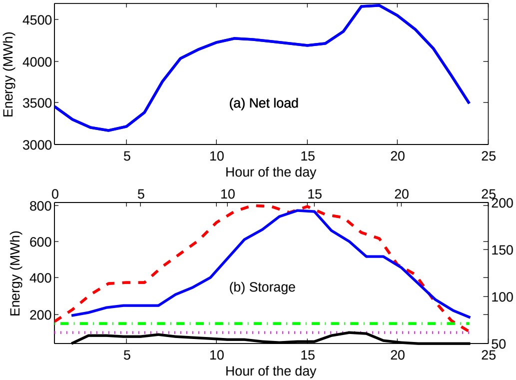


Figure: Mean storage trajectory (over 100 simulations) : five storage units and one wind farm

Optimal strategy: Use the most efficient (charging/discharging or storage) batteries.

Solution time as a function of network sizes and state space dimension

Table: Computation time in seconds for different number of buses, storage facilities and wind farms

# buses	$\lvert S\rvert$	$\lvert M\rvert$	Time	# buses	$\lvert S\rvert$	$\lvert M\rvert$	Time
30	1	1	1 229.80	118	1	1	2 399.35
30	5	1	1 582.67	118	5	1	2 444.99
30	5	5	1 323.88	118	5	5	2 453.89
57	1	1	1 388.09	118	10	5	2 179.62
57	5	1	1 454.47	118	20	10	2 248.39
57	5	5	1 396.26	300	1	1	4 159.16
57	10	5	1 597.71	300	5	1	4 234.72
89	1	1	1 570.67	300	5	5	4 570.01
89	5	1	1 709.68	300	10	5	4 617.65
89	5	5	1 575.09	300	20	10	5 036.37
89	10	5	1 737.32				

The solution time seems to remain reasonable as the size of the network and the dimension of the state space increase.

Summary

- SDDP allows optimization of the trade-off between here-and-now reward against the value of future flexibiity
- SDDP manages the dimensionality problem, to allow computational tractability,
- Is relatively easy to implement,
- Overall, the approach allows a fair trade-off between solution time and accuracy

Testing of Scenario Selection via Band Depth Clustering

The goal of scenario selection is the identification of a relatively low cardinality set of representatives, that sufficiently represent the uncertainty set. In this case,

- Each scenario will be a time series describing the wind speed over the course of a 24-hour day
- Each time series should be a plausible observation (not smoothed)
- The set of time series should reflect a range of possible wind behaviors

A data-driven approach is taken to this problem.

Clustering Overview

- Clustering is performed on a set of previous observations
  - One representative scenario is taken from each cluster
  - Cluster sizes suggest probabilities to assign to each scenario
- This process requires both a clustering algorithm and a measure of similarity/distance between observations
  - Conventional measures of distance like the L^p norm do not perform well in high dimensions, or in the presence of nonconstant variance and skewed data.

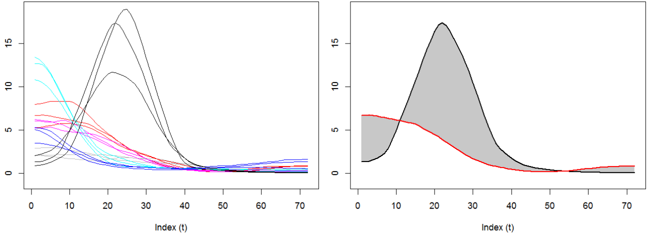
Depth statistics

We turn to depth statistics in order to examine observations relative to one another, not an external model.

- Depth statistics give centrality of an observation relative to a given set
- One-dimensional version: median and quantiles
- Many extensions to different types of data

The band depth

- *Modified band depth*: Lopez-Pintado and Romo (2009)
- Let $\mathbf{x}$ represent an observation, $x(t)$ represent the value at time point t, and X denote the entire set of observations
- Each band $\mathbf{b}$ is defined by a set of observations drawn from X
- The limits of the band at each t are defined by the max and min of the observations defining the band



CERTS
Consortium for Electric Reliability Technology Solutions

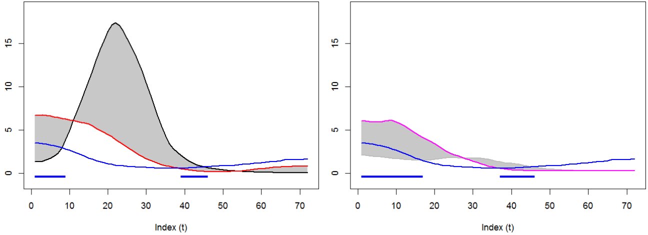
Band Depth Clustering

Previoulsy, we described the use of band depth clustering, which

- uses band depth as a measure of similarity between observations
- measures distance between observations based on trajectory shape, instead of overall level
- becomes a proven distance metric

The band depth

- Any other observation $\mathbf{x} = \{x(t)\}$ from the dataset can be compared to this band
- Let $T^b(\mathbf{x})$ be the set of all t for which $x(t)$ is in the band $\mathbf{b}$
- Given $\mathbf{x}$ and $\mathbf{b}$, consider the proportion of time $\mathbf{x}$ falls within the band: $T^{\mathbf{b}}(\mathbf{x})/T$
- We can calculate $T^{\mathbf{b}}(\mathbf{x})/T$ for any band $\mathbf{b}$

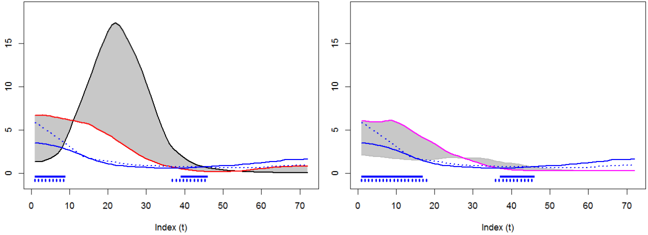


Depth to distance

- Averaging over all bands gives a measure of depth, or centrality, for $\mathbf{x}$ relative to the set X
- To do unsupervised clustering, we must extend depth to *pairwise* distances
- For any $\mathbf{x}$ and $\mathbf{y}$ from the dataset, find a distance $D_{\mathbf{xy}}$
- We introduce the *band distance*: Tupper, Matteson, and Anderson (2015)

The band distance

- Given two observations $\mathbf{x}$ and $\mathbf{y}$ from the simulated dataset, similar observations will fall into each band at similar times
- Compare $T^b(\mathbf{x})$ and $T^b(\mathbf{y})$ using Jaccard similarity (a set similarity measure)



CERTS
Consortium for Electric Reliability Technology Solutions

Cornell University
Founded A.D. 1865

The band distance

- Given observations $\mathbf{x}$ and $\mathbf{y}$ and a band $\mathbf{b}$, we get a bandwise similarity score:

$$s^{\mathbf{b}}_{\mathbf{xy}} = \frac{|T^{\mathbf{b}}(\mathbf{x}) \cap T^{\mathbf{b}}(\mathbf{y})|}{|T^{\mathbf{b}}(\mathbf{x}) \cup T^{\mathbf{b}}(\mathbf{y})|}$$

The band distance

- Given observations $\mathbf{x}$ and $\mathbf{y}$ and a band $\mathbf{b}$, we get a bandwise similarity score:

$$s^{\mathbf{b}}_{\mathbf{xy}} = \frac{|T^{\mathbf{b}}(\mathbf{x}) \cap T^{\mathbf{b}}(\mathbf{y})|}{|T^{\mathbf{b}}(\mathbf{x}) \cup T^{\mathbf{b}}(\mathbf{y})|}$$

- This is converted to a distance score for each band:

$$d^{\mathbf{b}}_{\mathbf{xy}} = 1 - s^{\mathbf{b}}_{\mathbf{xy}}$$

The band distance

- Given observations $\mathbf{x}$ and $\mathbf{y}$ and a band $\mathbf{b}$, we get a bandwise similarity score:

$$s_{\mathbf{xy}}^{\mathbf{b}} = \frac{|T^{\mathbf{b}}(\mathbf{x}) \cap T^{\mathbf{b}}(\mathbf{y})|}{|T^{\mathbf{b}}(\mathbf{x}) \cup T^{\mathbf{b}}(\mathbf{y})|}$$

- This is converted to a distance score for each band:

$$d_{\mathbf{xy}}^{\mathbf{b}} = 1 - s_{\mathbf{xy}}^{\mathbf{b}}$$

- The overall distance is the average score over $\mathbf{b} \in B_{\mathbf{xy}}$, the set of all bands containing $\mathbf{x}$ or $\mathbf{y}$ at any time:

$$D_{\mathbf{xy}} = \frac{1}{|B_{\mathbf{xy}}|} \sum_{\mathbf{b} \in B_{\mathbf{xy}}} d_{\mathbf{xy}}^{\mathbf{b}}$$

Properties of the band distance

- Considers observations only relative to the rest of the dataset
- Handles heteroskedasticity
- Is invariant under transformations that preserve observations' order at each t
- Involves no user-defined parameters
- Can be applied to any vectorized data
- Meets definition of a distance metric

Testing Scenario Selections "In Situ"

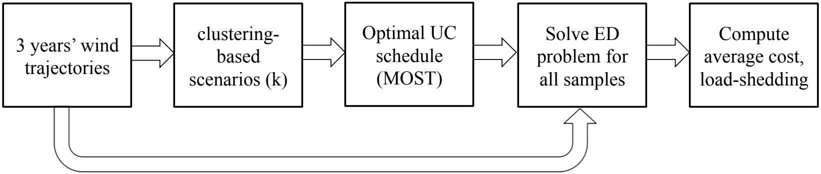
The measure of a scenario selection tool, is the ability to make better decisions with the same number of scenarios. To test this, we use

- The IEEE 30-bus network with one generator assumed to be wind
- Hourly wind speed data are drawn from NREL's EWITS database
- Matpower Optimal Scheduling Tool (MOST)
  - Divide the year into six two-month periods to minimize effect of seasonality
  - Three years of data provide approx. 180 observations per season

CERTS
Consortium for Electric Reliability Technology Solutions

Testing Scenario Selection

We then test the efficacy of each scenario set in characterizing the wind behavior of that season.



Better unit commitment decisions will lead to lower average dispatch costs and/or lower probability of lost load.

CERTS
Consortium for Electric Reliability Technology Solutions

Cornell University Founded A.D. 1865

Testing Scenario Selection

Results show that using band distance leads to greater reliability with comparable cost to Euclidean distance.

- Average difference in dispatch cost favors BDC, but is very small, in less constrained situations
- The system is most restricted in September-October season,
  - For $k = 10$, 15, and 20, Euclidean distance leads to loss of load on 17, 17, and 11 days
  - Band distance leads to loss of load on 0, 14, and 0 days

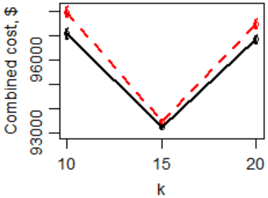
For a risk-averse operator, this is a strong argument for using the band distance

CERTS
Consortium for Electric Reliability Technology Solutions

Cornell University Founded A.D. 1865

Testing Scenario Selection

As a result of higher LOLE, the average cost of system dispatch is lower for solutions based on BDC than on the more common Euclidean distance metric.



Average daily system generation costs, incorporating lost load penalty, across all days in September-October. Dashed line shows the cost when using Euclidean distance; solid line, with band distance. Error bars show 1 standard deviation of pairwise differences.

Summary of Band Depth Testing

Initial results indicate that BDC is

- an improved selection approach,
- provides more secure solutions with the same number of scenarios and similar cost
- is flexible and can incorporate multiple wind farms, allowing for correlation among sites

Concluding Remarks

Optimizating Interaction of Renewables and Storage (SDDP)

- Incorporates realism of dynamic programming decisions
- Allows multiple wind farms and storage units
- Reasonable scaling of computational cost

Data Driven Scenario Selection (Band Depth Clustering)

- More effective selection of scenarios
- Improved stochastic programming solutions
- Flexible to incorporate multiple correlated wind farms

References

- Tupper, Laura L., David S. Matteson, and C. Lindsay Anderson. Band Depth Clustering for Nonstationary Time Series and Wind Speed Behavior. Submitted, 2015. `http://arxiv.org/abs/1509.00051`
- Luckny Zephyr, C. Lindsay Anderson. Integrating Storage in Power Systems Management. `http://arxiv.org/abs/1604.08189`

Thanks to: the MOST working group, Cornell University; Applied BioPhysics, Troy, NY; Palo

CERTS
Consortium for Electric Reliability Technology Solutions