Scenario Selection for Risk Quantification in Day-Ahead Power Grid Operational Planning

Guillermo Terrén-Serrano[1,2], Michael Ludkovski[3]

[1]Environmental Studies, [2]Environmental Markets Lab (emLab), [3]Statistics and Applied Probability
University of California Santa Barbara

Motivation, Research Question & Goals

- Independent System Operators (ISOs) carry out day-ahead Unit Commitment (UC) to schedule generators and Economic Dispatch (ED) to ensure the balance between energy supply and demand.
- Forecasts errors drive a wedge between the day-ahead projections and real-time realizations, acting as a key contributor to "day-ahead operational risk":
  - Higher than expected generation costs (G).
  - Insufficient operating reserves — reserve shortfall (RS).
  - Load shedding (LS).
  - Renewable generation curtailment (VC).
- Need to predict events potentially triggering above without running ED hundreds of times

Problem

UC and ED account for grid transmission and congestion constraints preserving the temporal structure (i.e., multi-hour ramping of thermal units).

- Hour-by-hour mismatches between forecasts and actuals are not sufficient to identify operational riskiness.
- Risk managers must treat scenarios holistically as *curves* indexed in time.

Algorithm to pre-screen for scenarios leading to operational *risk*?

Software & Experimental Setup

clnSim Tool

Generates joint probabilistic asset-level **scenarios** across hundreds of VRE assets and load zones [1].

- Simulations are conditional to given DA forecasts.
- Seasonally calibrated.
- **Case study:** 1000 hourly scenarios across 8 ERCOT load zones, and 185 solar & wind assets ($24 \times (8 + 185) = 4632$ dim.)

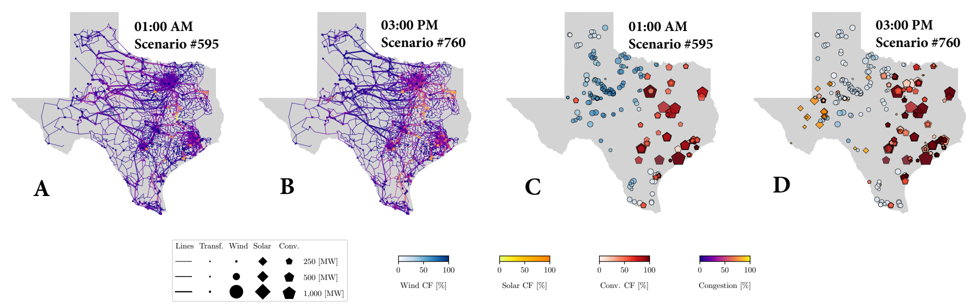

Figure 1. Texas-7k operations simulation on Oct 4. This day has VC at 1 am in scenario 595, and LS at 3 pm in scenario 760. In the VC event, a line in east Texas (C) creates a load pocket that prevents the flow of energy from high wind generation in northwest Texas (A). The LS event is triggered by low wind generation which causes congestion (B) preventing the energy generated by conventional assets from flowing to North Central region (D).

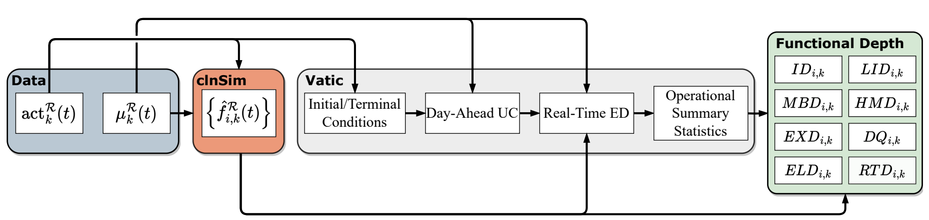

Figure 2. Workflow to screen *clnSim* scenarios for day-ahead power grids operational planning in *Vatic*.

Statistical Toolbox

Consider 8 **functional depth** notions [3] for defining riskiness in day-ahead operational planning of the grid:

- Integrated Depth (ID).
- Modified Band Depth (MBD).
- Extremal Depth (EXD).
- Extreme Rank Length Depth (ERLD).
- L-Infinity Depth (LID).
- h-Mode Depth (HMD).
- Directional Quantile (DQ).
- Random Tukey Depth (RTD).

- The intuitive characterization of operational risk is scenarios that are "outlying", i.e., are far from the bulk and the forecast ("central" scenario).
- The scenarios are fed into ED simulator (*Vatic* [2]), and the results are aggregated across scenarios yielding to the probability distribution of grid operational characteristics (G, RS, LS, and VC).
- Relies on the Texas-7k grid from ARPA-E PERFORM Data Plan.

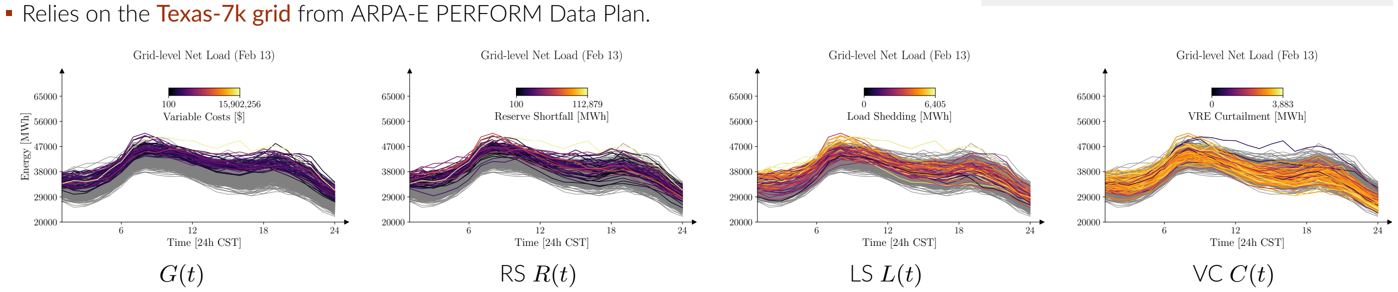

Figure 3. Operational risk against aggregated NL load $f_i^{\mathcal{N}}(t)$ on Feb 13. We highlight the top $m = 50$ scenarios (the more risky, the brighter its color).

- Run grid ED simulations over 25 test days across the calendar year.
- A baseline approach is to implement screening based on summary statistics (e.g., peak load, ramp rate, or peak-valley difference).
  - Scalarizing the scenarios via a single screening metric loses too much information and yields poor screening.
  - System Net Load (NL) is the simplest predictor of risk. Works for predicting generation costs and reserves shortfall but fails for load shedding and VRE curtailment.
- Need more automated approach that gives a smart aggregation of high-dimensional scenarios
- Analyze statistical functional analysis methods to define and predict "extreme" scenarios.
- Outlyingness may be defined by aggregating hourly ranks or be based on distance (in magnitude/functional shape) to the scenarios' core.

Solution & Performance Analysis

Magnitude of Reserve Shortfall

- Identify the most likely scenarios to be in the top 5% (i.e., $m = 50$ out of 1000).
- We devise a 2-stage procedure to select $75 = 1.5m$ scenarios:
  1. n_1 scenarios ($n_1 < N$) are selected by greatest AUC in NL.
  2. n_2 scenarios ($n_2 < n_1$) are selected by lowest functional depth.
- Functional depth improves the proportion of false positive and false negatives (see bottom-right and top-left quadrant in Fig. 5).
- Predicting extreme RS is more challenging during summer when larger shortfalls occur (Fig. 6 left).

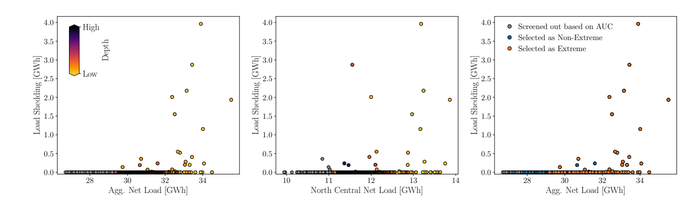

Figure 4. 3-stage detection based on merging grid and zonal level functional depths on Feb 26. Average net load on the x-axis and total load shedding on the y-axis. The left panel shows aggregated grid-level selection and the middle panel shows aggregated zonal-level selection. Scenarios screened out by AUC are in gray, the rest are color-coded (brighter is more extreme) according to ERD functional depth. The right panel shows the merging of the two criteria, the n_1 most extreme scenarios in terms of their combined depth ranking.

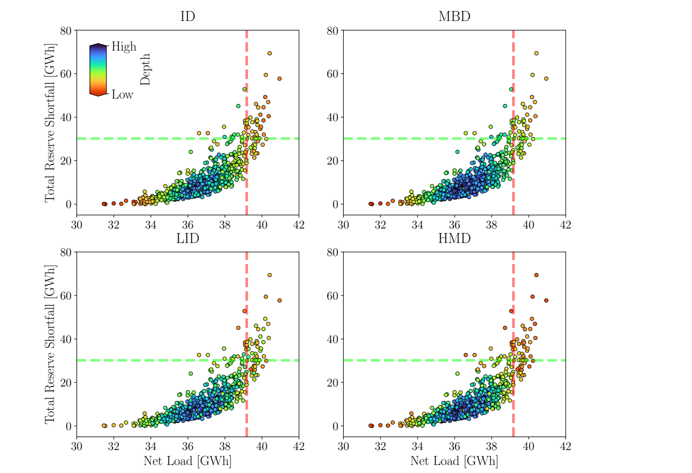

Figure 5. Detecting SC via 4 different functional depth metrics on Feb 13. Aggregated daily NL on the x-axis vs. daily RS on the y-axis. Scenarios are color-coded according to the respective depth metric: the brighter the color gradient, the less deep the scenario. The green horizontal (vertical red) line at 30.17 GWh (resp. 39.18 GWh) shows the threshold for the top 50 highest RS (resp. top 50 highest NL).

VRE Curtailment

- VC extremality: scenarios with daily curtailment ≥ 100 MWh. Exclude days with no VC or more than 25% likelihood of VC.
- Most accurate screening accuracy is 44.69% (DQ) using grid-level VRE generation and NL as predictors.
- Augmenting to zonal information increases accuracy to 61.58% (LID). See Fig. 6 left panel.
- In summer/fall, Far West zone load is a predictor of VC. The best approach is pre-screening to keep $n_1^{VCZ} = 450$ scenarios according to grid-level NL AUC and then apply DQ (90.19%).
- None of the proposed depth metrics help for cold-weather VC.

Load Shedding

- Most scenarios have zero $L_i = 0$.
- Any positive LS is operationally extreme; The number of scenarios with positive LS $L_i > 0$ varies day-by-day. There are $Ave_k(|\mathcal{E}_k|) = 149.48$ scenarios with LS on average. We select $Ave_k(n_2(k)) \approx 225$ scenarios (50% margin).
- LS primarily occurs on days with high grid-level NL during the morning peak and high North Central zonal NL in the evening.
- The functional depths of grid-level NL provide limited predictive power; augment with zonal-level NL via a 3-stage screening procedure (see Fig. 4).
- EXD is best depth for screening LS (93.73%) when using adaptive $n_2(k)$, $n_1^{LSZ} = 650$ and North Central zone NL. See Fig. 6 middle panel.

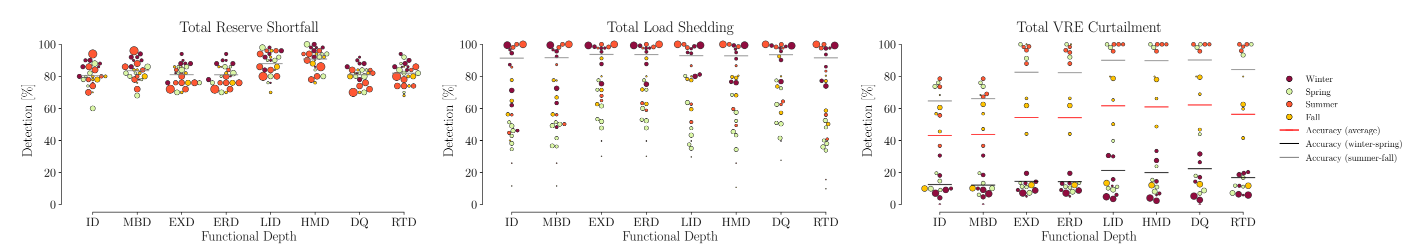

Figure 6. Detection accuracy for identifying RS (left panel), LS (middle panel), and VC (right panel). The bubble swarm plots correspond to different functional depth metrics $\mathcal{D}$. with each dot representing the accuracy $P_k^{\mathcal{D}}$ achieved for the 25 given simulation dates and the gray horizontal bars denoting the respective mean accuracy $Ave_k(P_k^{\mathcal{D}})$. Symbol size denotes the magnitude of the respective RS and the colors represent the seasons.

Take Aways

- Statistical measures of functional depth offer a novel and effective approach to screen scenarios for operational risk assessment.
- Work with depth metric based on grid- and zonal-level generation and net load.
- Load shedding and VRE curtailment events tend to be associated with congestion, and hence depend on grid topology.
- These complex phenomena cannot be predicted based on grid-level covariates and need statistical tools.
- Anticipate additional gains through customization of the functional depth measures.
- Open problem how to effectively merge multiple depth rankings.
- See **arXiv.org** for more details [4]

Acknowledgement

This research was partially founded by the ARPA-E PERFORM grant DE-AR0001289. Use was made of computational facilities of the UCSB Center for Scientific Computing funded by CNS-1725797 and supported by MRSEC (NSF DMR 2308708). We thank X. Yang, G. Swindle, R. Sircar, and R. Carmona for insightful discussions, and M. Grzadkowski, A. Fang for computing support in deploying *Vatic*.

References

[1] Ludkovski M, Swindle G, Grannan E. Large Scale Probabilistic Simulation of Renewables Production. arXiv preprint arXiv:220504736. 2022;.

[2] Grzadkowski M, Fang JA, Yang X, Solomon A, Ronnie-Sircar. PrincetonUniversity/Vatic: Texas-7k 2030 and Python 3.11. Zenodo; 2023. Available from: https://doi.org/10.5281/zenodo.8381228.

[3] Wang JL, Chiou JM, Müller HG. Functional data analysis. Annual Review of Statistics and Its Application. 2016;3:257–295.

[4] Terrén-Serrano G, Ludkovski M. Extreme Scenario Selection in Day-Ahead Power Grid Operational Planning. arXiv preprint arXiv:230911067. 2023;.