

# Robust transmission expansion assessment using minimal samples

Erich Trieschman

Stanford University, Department of Statistics

etriesch@stanford.edu

## Abstract

## 1. Introduction

In re-structured electricity markets in the United States, transmission system planning is conducted through the joint collaboration between regulatory agencies and an Independent System Operator (ISO) [6]. Through Integrated Resource Planning (IRP) proceedings, Load Serving Entities (LSE's) and regulatory agencies use capacity expansion models to optimize sequential investment decisions in thermal, renewable, and storage plant capacity over decades into the future [7]. Capacity expansion and production cost planning models require forecasts of load, resource availability, and costs to serve as constraints to be met through the dispatch of resources over a transmission network. However, current planning methods often rely on a single or few scenarios of load and resource availability forecasts. While deterministic models can help communicate a single system cost estimates, their results are highly sensitive to model inputs and ignore uncertainty in future conditions.

To address this challenge, methods in decision-making under uncertainty have been developed for electricity transmission system planning. The Transmission Economic Assessment Methodology (TEAM) [1] framework offers a solution to address uncertainties in transmission planning. One pillar of TEAM seeks to manage uncertainty by stress-testing transmission alternatives across potential future scenarios in a stochastic production cost simulation with high nodal resolution. The explicit uncertainty analysis proposed in TEAM includes fuel costs, renewable resource availability, thermal unit availability, and future load to produce an output distribution of system costs for which risk metrics can be derived. However, these approaches are computationally expensive given the calculation of the operational expenses under many stochastic variables and the high-resolution network model needed to characterize transmission economic benefits adequately.

In this project, we focus on methods to select a sub-

set of optimal load, wind, and solar forecasts to characterize the cost-benefit distribution of a transmission expansion project. Given a distribution of correlated load, wind, and solar scenarios, we use experimental design and active learning techniques to minimize the subset of forecasts required to characterize the distribution of system costs of an expansion decision. Our primary objective is to evaluate these methods under a single transmission expansion scenario, which we will optimize under the expected value of the forecasts.

We use a 97-bus synthetic network of the Western Electricity Coordinating Council (WECC) to assess the benefits of developing a High Voltage Direct Current transmission line to integrate floating offshore wind from California's north coast.

We are able to approximate the cost-benefit distributions after only 25-50 samples. Assuming that an adequate distribution can be generated with 500 samples, this represents computational savings of 90-95%. We also find that the optimal methods we explore approximate these cost-benefit distributions significantly better than baselines of randomly selected sample points.

## 2. Methods

This work aims to efficiently generate distributions of economic benefits from transmission expansion scenarios using a nodal production cost simulation (PCS). To achieve this, we use techniques from experiment design and active learning to fit surrogate models to the production cost simulation outputs.

To generate our target empirical distribution of benefits (hereafter "true distribution"), we first determine the optimal transmission expansion under the expected value of the solar, wind, and load profiles. We then run a large sample of model inputs through the cost model. This is the computationally expensive step, which we wish to improve upon.

To generate our estimated empirical distribution of benefits (hereafter "surrogate distribution"), we fit surrogate models to a subset of cost model runs and run a large sample of model inputs through the surrogate model, which we expect to realize significant computational gains. See Fig-

Figure 1 for an overview of our approaches, which we describe in the subsections below

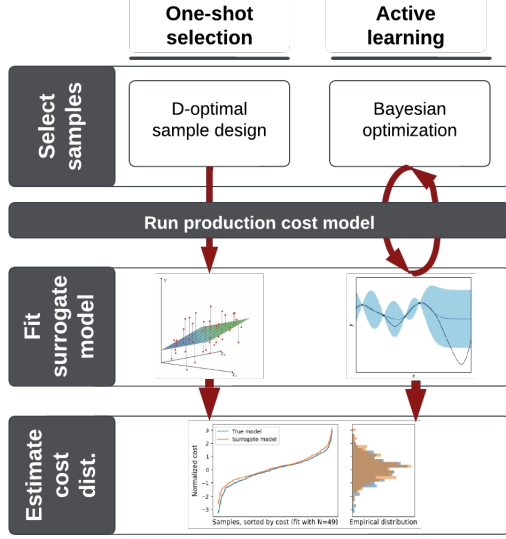


Figure 1. Methods used to estimate empirical distributions

## 2.1. Production cost simulation and model inputs

The PCS is built on the pypsa-usa Breakthrough Energy synthetic Western Electricity Coordinating Council (WECC) network model [15]. The PCS is an operational simulation of the electricity transmission system that provides information on the techno-economic performance of a power system, such as system operating costs, market prices, transmission congestion, and more. Figure 2 shows the network we model.

$$\text{Benefits} = \underbrace{\tilde{\lambda}_{n,t} \cdot P_{\text{load}}}_{\text{HVDC Expansion}} - \underbrace{\lambda_{n,t} \cdot P_{\text{load}}}_{\text{Pre-expansion}}$$

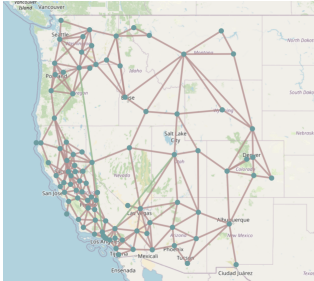


Figure 2. pypsa-usa 96-node WECC network

We generate stochastic load, wind, and solar profiles using the mean-reversion stochastic process method developed by the CAISO [8, 9]. This method uses a stochastic

process-based method to generate profiles representing deviations from an expected base profile. We use the data embedded in the Breakthrough Energy network from 2016 as the "base" hourly load, wind, and solar profiles.

Estimating system costs for just a small subset of nodes over a single year results in a massive number of model inputs (8760 hours in a year  $\times$  3 time series  $\times$  20 nodes  $>$  500,000 inputs). To overcome the curse of dimensionality, we use Principal Component Analysis (PCA) to encode the input data and drastically reduce the sample space. Our goal is to reduce the number of dimensions to a target of 10 dimensions, while retaining the important features of the original data.

PCA is a widely used data reduction technique that identifies linear combinations of input features that capture the most significant variation in the data. This method begins with an eigendecomposition of the covariance matrix of the dataset. The eigenvectors represent the directions along which the data varies the most, while the eigenvalues represent the amount of variation along each of these directions. The eigenvectors with the largest eigenvalues capture the most variation in the data, and are referred to as the principal components once they are normalized. Encoding with PCA involves projecting the original data onto a subset these principal components, which reduces the dimensionality of the dataset while preserving as much of the original variation as possible.

Encoded data,  $Y \in \mathbb{R}^{n \times D}$ , can be formed as

$X^T X = Q \Lambda Q^{-1}$ , the covariance eigendecomposition

$$U \in \mathbb{R}^{n \times n}, \text{ s.t. } u_i = \frac{q_i}{\|q_i\|} \text{ for } Q = [q_1, \dots, q_n]$$

$$Y = XU_D, \text{ for } U_D = [u_1, \dots, u_D]$$

Operation of the production cost simulation (hereafter "cost model") and generation of the encoded stochastic load, wind, and solar profiles (hereafter "model inputs") is conducted outside the scope of this project.

## 2.2. Scenario selection

To evaluate the performance of our scenario selection methods, we compare their surrogate distributions to the true distribution. We also compare each to a baseline, which is trained with a random sample of model inputs; we expect the baseline surrogate distributions to differ more from the true distribution.

We propose two optimal scenario selection methods. These methods are used to generate surrogate models, which approximate the computationally expensive cost model. Surrogate distributions can then be generated by the entire sample of forecasts through the inexpensive surrogate models. The first optimal scenario selection approach is taken from the literature of experiment design. We refer to this approach as "One-shot selection" as it selects all

samples at the outset of the surrogate model training. The second optimal scenario selection approach is taken from the literature of active learning and we refer to it as such. Active learning selection sequentially selects optimal forecasts based on uncertainty in the surrogate model output being fit.

### 2.2.1 One-shot selection

In this approach we assume a linear relationship between forecasts and transmission cost outputs and use the surrogate cost model

$$y_i = x_i^T \beta + \epsilon_i, \text{ with } y_i \in \mathbb{R}, x_i, \beta \in \mathbb{R}^n, \epsilon \text{ white noise}$$

We estimate  $\hat{\beta}$  using maximum likelihood estimation and minimize the uncertainty of our estimator  $\hat{\beta}$  by minimizing the inverse covariance of our input data. Assuming a fixed set of  $p$  input data combinations,  $v_j$ , and  $M$  desired scenarios to be selected, we may select target input data to minimize the inverse covariance by optimizing the following problem:

$$\begin{aligned} \min_m \quad & (\text{over } S_+) \quad \left( \sum_{j=1}^p m_j v_j v_j^T \right)^{-1} \\ \text{s.t.} \quad & m \succeq 0, m^T \mathbf{1} = M, m \in \{0, 1\}^n \end{aligned}$$

Here,  $m \in \mathbb{R}^p$  is an indicator variable whose  $j^{\text{th}}$  entry determines whether scenarios  $v_j$  is selected. We choose  $m \in \{0, 1\}^n$  instead of the canonical formulation,  $m \in \mathbb{Z}^n$ , because our cost model is deterministic and we only need to concern ourselves with a single selection for each scenario.

We scalarize the problem with the D-optimal design formulation as described in Boyd et. al. [3]. We also relax the binary constraint using the L1-norm heuristic, replacing  $m$  with  $\lambda$  in our formulation to represent this. We choose the L1-norm constraint instead of the canonical constraint  $\lambda^T \mathbf{1} = \alpha M$  to maximize sparsity in the resulting solution, since we seek a solution with up to one choice per input scenario.

$$\begin{aligned} \min_{\lambda} \quad & -\log \det \left( \sum_{j=1}^p \lambda_j v_j v_j^T \right) \\ \text{s.t.} \quad & 0 \preceq \lambda \preceq \mathbf{1}, \quad \|\lambda\|_1 \leq \alpha M \end{aligned}$$

This relaxed problem may yield a solution  $\lambda_i \in [0, 1]$ . We choose the input data scenarios with the  $M$  largest values in  $\lambda$  to run through the cost model.

There are two design choices for this problem, which we care to explore in future work. First, we must choose how to discretize continuous input data,  $x_i$ , into finite input data scenarios,  $v_j$ . We choose 10 data bins for each dimension and a uniform split across bins. Second, we must choose hyperparameter  $\alpha$  to control the impact of the L1-heuristic constraint; we use  $\alpha = 2$ .

### 2.2.2 Active learning selection

Active learning design is a sequential approach to selecting optimal scenarios for the cost model. In this approach, we iteratively select forecasts to run through the cost model based on the previous forecasts and the cost model outputs.

One popular method for active learning selection is Bayesian optimization, which we use here. This approach estimates a probability distribution over cost function outputs, conditional on all of the previous model outputs; it does so through a Gaussian Process (GP) model, which is updated after each evaluation of the function. The selection of the next forecast is determined by an "acquisition function", defined by the objective we seek to optimize [4] [14].

We define the Gaussian Process surrogate model as

$$f(x) \sim \mathcal{GP}(0, k(x, x'))$$

Where  $x$  are the model inputs, and  $k(x, x')$  is a pre-defined covariance function, often the squared exponential function,  $k(x^{(i)}, x^{(j)}) = \exp(-\frac{1}{2}\|x^{(i)} - x^{(j)}\|_2^2)$ . Assuming 0-mean is common in Bayesian optimization.

A property of the Gaussian Process is that any finite collection of model outputs follows a multivariate Normal distribution. Namely

$$\begin{bmatrix} f(x^{(1)}) \\ \vdots \\ f(x^{(n)}) \end{bmatrix} := \mathbf{f}^{(1:n)} \sim N(0, \mathbf{K})$$

Where  $\mathbf{K} \in \mathbb{R}^{n \times n}$  is the covariance matrix of the model inputs with  $\mathbf{K}_{ij} = k(x^{(i)}, x^{(j)})$ .

Using the Sherman-Morrison-Woodbury formula and properties of conditional multivariate Normal distributions, we may generate the predictive probability of our surrogate cost model, given a model input and the evidence:

$$\begin{aligned} f^{(n+1)} \mid \mathbf{x}^{(1:n+1)} &\sim \mathcal{N}(\mu_{n+1}, v_{n+1}), \text{ where} \\ \mu_{n+1} &:= \mathbf{k}(x^{(n+1)})^T \mathbf{K}^{-1} \mathbf{f}^{(1:n)} \\ v_{n+1} &:= k(x^{(n+1)}, x^{(n+1)}) - \mathbf{k}(x^{(n+1)})^T \mathbf{K}^{-1} \mathbf{k}(x^{(n+1)}) \\ \mathbf{k}(x^{(n+1)}) &:= [k(x^{(n+1)}, x^{(1)}), \dots, k(x^{(n+1)}, x^{(n)})] \end{aligned}$$

In our setting, we use Maximum Entropy Search (MES) as the acquisition function, which allows us to identify the forecast connected to the point of highest entropy in the surrogate cost model. Entropy is a measure of uncertainty about a given model input, so sequentially selecting maximal entropy points allows us to reduce overall uncertainty across our surrogate model [12] [10]. We first derive an analytical form of the entropy equation,  $H(f \mid \mathbf{x}^{1:n}, x)$ :

$$\begin{aligned} H(f \mid \mathbf{x}^{(1:n)}, x) &:= - \int N(f \mid \mu_x, v_x) \log [N(f \mid \mu_x, v_x)] df \\ &= - \int N(f \mid \mu_x, v_x) \left[ -\frac{1}{2} \log(2\pi v_x) - \frac{(f - \mu_x)^2}{2v_x} \right] df \\ &= \frac{1}{2} \log(2\pi v_x) + \frac{1}{2} \end{aligned}$$

Using kernel  $k(x^{(i)}, x^{(j)}) = \exp(-\frac{1}{2l^2}\|x^{(i)} - x^{(j)}\|_2^2)$ , the point of maximum entropy in the surrogate model,  $x^*$ , is therefore

$$\begin{aligned} & \operatorname{argmax}_x \frac{1}{2} \log(2\pi v_x) + \frac{1}{2} = \operatorname{argmax}_x v_x \\ & = \operatorname{argmax}_x k(x, x) - \mathbf{k}(x)^T \mathbf{K}^{-1} \mathbf{k}(x) \\ & = \operatorname{argmin}_x \sum_{i=1}^n \sum_{j=1}^n \mathbf{K}_{ij}^{-1} \mathbf{k}(x)_i \mathbf{k}(x)_j \\ & = \operatorname{argmin}_x \sum_{i=1}^n \sum_{j=1}^n C_{ij} \exp\left(-\frac{1}{l^2}\|x - \frac{1}{2}(x_i + x_j)\|_2^2\right) \\ & \text{where } C_{ij} = \mathbf{K}_{ij}^{-1} \exp\left(-\frac{1}{4l^2}\|x_i - x_j\|_2^2\right) \end{aligned}$$

This problem is clearly not convex, however, we employ several search strategies to approximate  $x^*$  with  $\hat{x}^*$ . First we consider a brute force approach where we calculate entropy for a fixed set of sample points,  $x \in \mathcal{D}$ .  $\hat{x}^*$  is simply the sample point yielding the maximal entropy. We also consider sequential convex programming (SCP) techniques using second order Taylor approximations and particle methods; these local methods leverage convex optimization theory.

In SCP we form a convex relaxation to the objective function,  $f(x) = \sum_{ij} C_{ij} \exp(-\|x - \frac{1}{2}(x_i + x_j)\|_2^2)$ , at each step of the optimization, and we iterate until local convergence. The update step in this optimization routine becomes

$$x^{(n+1)(k+1)} = \operatorname{argmin}_x \hat{f}^{(k)}(x) \text{ s.t. } x \in \mathcal{T}^{(k)}$$

Where  $\mathcal{T}^{(k)} := \{x \mid |x_i - x_i^{(n+1)(k)}| \leq \rho_i^{(k)}\}$  is the convex trust region for the problem. Hyperparameter,  $\rho^{(k)}$  is updated through the canonical trust region update:

$$\begin{aligned} \rho^{(k+1)} &= \begin{cases} \beta^{succ} \rho^{(k)} & \text{if } \delta \geq \alpha \hat{\delta} \\ \beta^{fail} \rho^{(k)} & \text{else} \end{cases}, \text{ for} \\ \delta &:= f\left(x^{(n+1)(k)}\right) - f\left(x^{(n+1)(k+1)}\right) \\ \hat{\delta} &:= f\left(x^{(n+1)(k)}\right) - \hat{f}^{(k)}\left(x^{(n+1)(k+1)}\right) \end{aligned}$$

Practitioners often set  $x^{(n+1)(k+1)} \leftarrow x^{(n+1)(k)}$  when  $\delta < \alpha \hat{\delta}$ ; we implement this as well.

Using second-order approximation methods,  $\hat{f}^{(k)}(x)$  is defined as the convex part of the second order Taylor expansion of our objective function described above, evaluated at point  $x^{(n+1)(k)}$ :

$$\begin{aligned} \hat{f}(x) &\cong f\left(x^{(n+1)(k)}\right) + \nabla f\left(x^{(n+1)(k)}\right)^T \left(x - x^{(n+1)(k)}\right) + \\ &\quad \left(x - x^{(n+1)(k)}\right)^T P \left(x - x^{(n+1)(k)}\right) \end{aligned}$$

For

$$\begin{aligned} \nabla f(x) &= \sum_{ij} C_{ij} \exp\left(-\frac{1}{l^2}\|x - \frac{1}{2}(x_i + x_j)\|_2^2\right) \frac{1}{l^2} (-2x + (x_i + x_j)) \\ P &= \left(\nabla^2 f\left(x^{(n+1)(k)}\right)\right)_+ \\ \nabla^2 f(x) &= \sum_{ij} C_{ij} \exp\left(-\frac{1}{l^2}\|x - \frac{1}{2}(x_i + x_j)\|_2^2\right) \\ &\quad \times \frac{1}{l^2} \left(4 \left[x - \frac{1}{2}(x_i + x_j)\right] \left[x - \frac{1}{2}(x_i + x_j)\right]^T - 2I\right) \end{aligned}$$

Using the particle method,  $\hat{f}^{(k)}(x)$  is defined by a quadratic function fit to a random sample of data,  $z_1, \dots, z_L \in \mathcal{T}^{(k)}$  evaluated with our non-convex objective function,  $f(x)$

$$\hat{f}^{(k)}(x) = \left(x - x^{(n+1)(k)}\right)^T P \left(x - x^{(n+1)(k)}\right) + q^T \left(x - x^{(n+1)(k)}\right) + r$$

Where

$$(P, q, r) = \operatorname{argmin}_{P \succeq 0, q, r} \sum_{i=1}^L \left( \left( z_i - x^{(n+1)(k)} \right)^T P \left( z_i - x^{(n+1)(k)} \right) + q^T \left( z_i - x^{(n+1)(k)} \right) + r - f(z_i) \right)^2$$

The particle method approach simplifies computation and allows us to flexibly select where in the trust region,  $\mathcal{T}^{(k)}$ , to sample.

There are several design choices for our active learning approach, which we care to explore in future work. First, we choose how to characterize our Gaussian process. We described above the specific kernel we consider, but several options exist. Within the definition of the kernel there also exist hyperparameters that may be tuned for optimal performance. In our setting, we used  $l = 2$  and observe that outcomes are highly sensitive to this hyperparameter.

Second, we choose an acquisition function for selecting each subsequent sample point. We use MES, but other approaches that balance exploration and exploitation could also be considered.

Third, within our SCP approach we choose which function approximation to use; we explore both Taylor and particle method approximations. We also choose trust region and trust region update hyperparameters. For each subsequent sample, we initialize  $\rho_i^{(0)} = \gamma \min X_i^T$  with  $\gamma = 0.01$  and we define  $(\beta^{succ}, \beta^{fail}) = (1.1, 0.5)$  and  $\alpha = 0.1$ .

### 3. Results

We have two major takeaways from this work. First, we find that both of the surrogate models are able to approximate the cost model after only 25-50 samples. Assuming that an adequate true distribution can be generated with 500 samples, this represents computational savings of 90-95%. While the linear model is not able to simulate the tails of

the true distribution well, the Gaussian process model successfully approximates the outer percentiles (5th and 95th).

Second, we find that optimal sample selection methods for fitting both surrogate models can generate surrogate distributions significantly closer to the true distribution than random sampling for fitting surrogate models.

Table 1 summarizes the results for all of our trials. Empirical distributions are generated with 500 input data samples. Surrogate models are fit with 50 samples (either optimally or randomly selected). The numbers in the table represent the absolute difference between the statistic calculated for the true distribution and the statistic calculated for the surrogate distribution; we want these numbers to be close to zero. We use bootstrapping to generate a confidence interval around these differences<sup>1</sup>.

Table 1. Absolute difference between true distribution and surrogate distribution statistic (mean  $\pm$  std for 100 bootstrap iterations)

N=500 samples=50	One-shot selection		Active learning selection		
	Baseline	D-optimal	Baseline	Taylor	Particle
Min	1.66 $\pm$ 0.46	1.49	0.64 $\pm$ 0.53	0.46 $\pm$ 0.41	0.41 $\pm$ 0.39
5th pctl	0.87 $\pm$ 0.28	0.85	0.68 $\pm$ 0.23	0.24 $\pm$ 0.16	0.23 $\pm$ 0.16
10th pctl	0.59 $\pm$ 0.23	0.53	0.47 $\pm$ 0.19	0.14 $\pm$ 0.12	0.17 $\pm$ 0.14
25th pctl	0.32 $\pm$ 0.16	0.26	0.28 $\pm$ 0.14	0.11 $\pm$ 0.10	0.14 $\pm$ 0.11
50th pctl	0.14 $\pm$ 0.11	0.11	0.11 $\pm$ 0.09	0.09 $\pm$ 0.07	0.14 $\pm$ 0.10
75th pctl	0.31 $\pm$ 0.16	0.34	0.26 $\pm$ 0.13	0.11 $\pm$ 0.09	0.12 $\pm$ 0.09
90th pctl	0.56 $\pm$ 0.23	0.66	0.45 $\pm$ 0.17	0.14 $\pm$ 0.13	0.16 $\pm$ 0.12
95th pctl	0.62 $\pm$ 0.27	0.75	0.47 $\pm$ 0.20	0.18 $\pm$ 0.17	0.18 $\pm$ 0.15
Max	1.34 $\pm$ 0.52	1.35	0.76 $\pm$ 0.46	0.40 $\pm$ 0.45	0.53 $\pm$ 0.37
Mean	0.13 $\pm$ 0.11	0.04	0.11 $\pm$ 0.09	0.09 $\pm$ 0.07	0.11 $\pm$ 0.09
Std	0.47 $\pm$ 0.15	0.50	0.37 $\pm$ 0.09	0.10 $\pm$ 0.07	0.10 $\pm$ 0.07

We present our results in figures below. Figures 3 and 4 compare the true distribution to the surrogate distribution, fit with 50 samples. Figures 5 and 6 show how the difference in statistics presented in Table 1 evolve as the surrogate models are fit with more sample points. Uncertainty bands represent a 95% confidence interval, generated by bootstrapping.

#### 4. Discussion and future work

Our results demonstrate that complex production cost simulation models can be approximated with simple linear or gaussian process models as a function of encoded stochastic weather and load inputs. It remains to be investigated whether cost model outputs can also be modeled as a function of other PCS inputs like fuel costs.

We suspect that the performance we observe is the result of underlying data and model structures. First, the production cost simulation uses a DC Power Flow model, which requires a linear objective and also linearizes the AC power

<sup>1</sup>Active learning requires initialization with a few sample points. We choose these randomly in each iteration, hence the distributions we observe reveal the impact of initial conditions on the approach. D-optimal design is deterministic, so we do not generate a confidence interval for that method

Figure 3. Distributions generated by one-shot selection

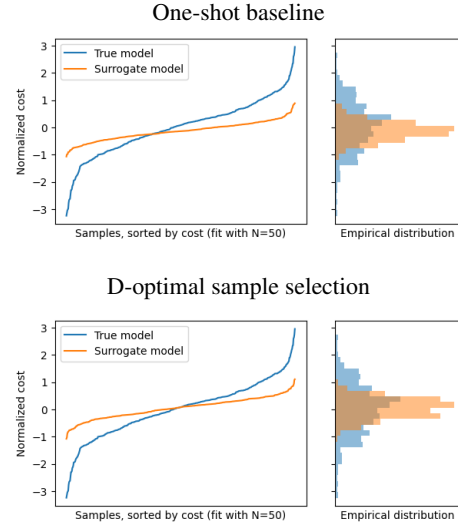
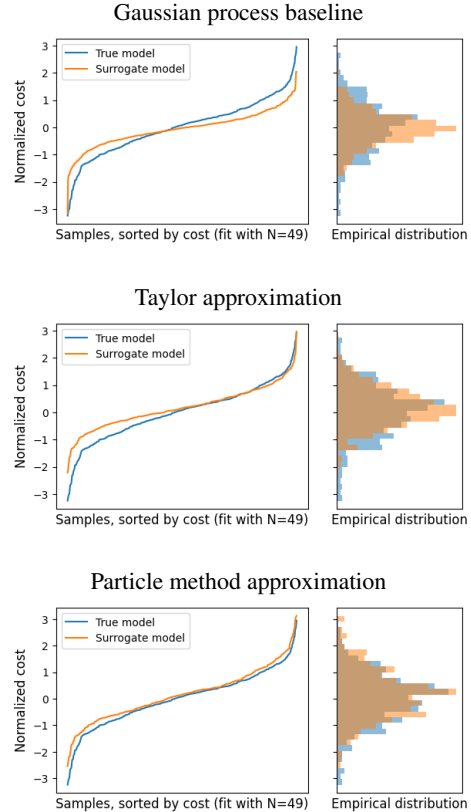
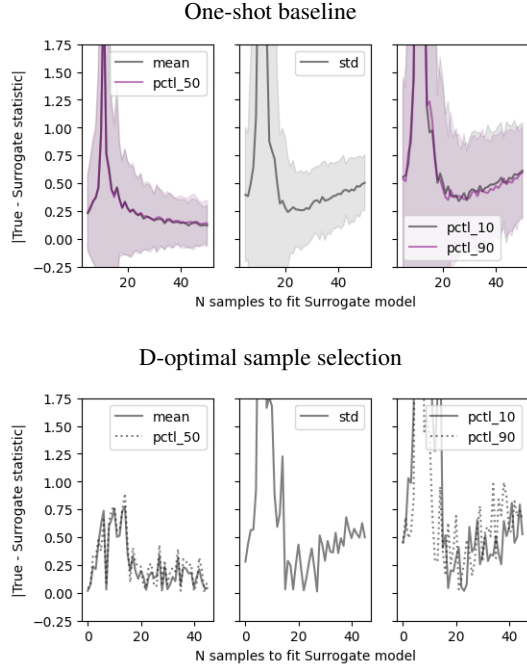


Figure 4. Distributions generated by active learning



flow constraints. Hence, it is reasonable to expect that some linear regression model may adequately capture the model behavior. Second, the stochastic profile method models a

Figure 5. Absolute differences in statistics from distributions generated by one-shot selection



random walk with a mean reverting term. The process generates each forward step by taking the sum of the previous step, a mean-reverting term, and a random error. It is reasonable to expect that such a structure could be exploited by Gaussian process surrogate models. We propose studying the underlying mechanics leading to our project’s outcome in future work.

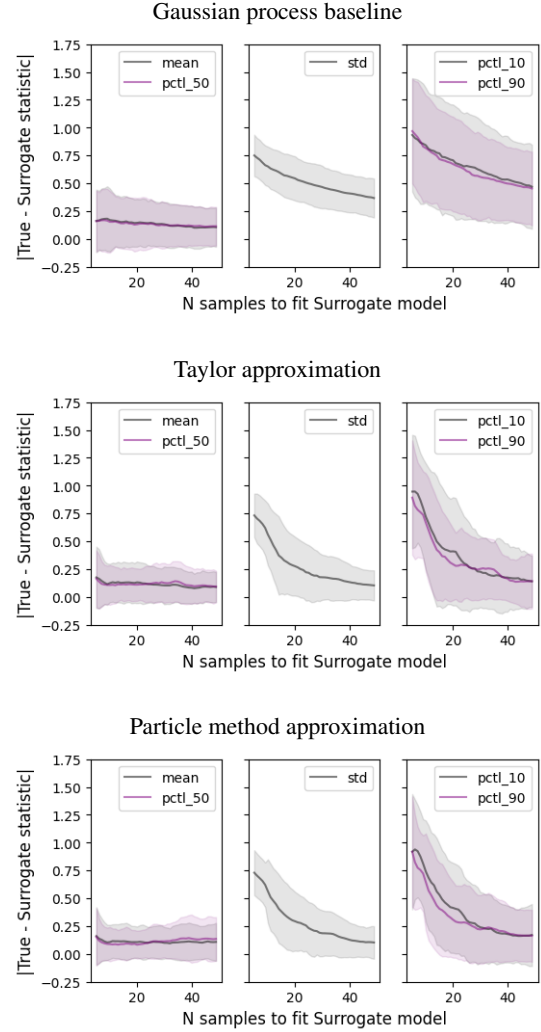
Another curiosity for future work is to explore the characteristics of these optimal input data samples. We would like to explore the nature of these samples and what they reveal about the core drivers of costs in the cost model. This line of research may reveal key nodes in the network or types of weather patterns that the cost model is particularly sensitive to.

**PLACEHOLDER:** Discuss how these methods might be used in a setting where we consider a broader set of expansion scenarios (beyond a single line).

Our PCA encoding strategy captures approximately 2.5% of the total input data variance. We would like to explore how our results change when we are able to capture a higher share of total variance through an encoding. Such encoding strategies could include wavelet transforms and variational autoencoders for time series data as presented in Desai et. al [5].

There are several ways we could test the robustness of our results. First, we propose examining our methods under purely random input data, instead of with well-behaved

Figure 6. Absolute differences in statistics from distributions generated by active learning



stochastic mean-reversion profiles. **TODO:** why would this be helpful? Second, we propose testing these methods under the canonical two-bus, three-bus, and Tripoli **spelling?** networks. Examining our results under these simplified networks will allow us to study the ways in which Gaussian Processes may simulate production cost models.

Lastly, we propose several extensions to improve and fine tune our sample selection procedures:

1. One-shot selection to optimize for quantile regression parameters instead of linear regression parameters as described in Wang et al. (2020) [13]. This strategy could offer a sampling technique to target specific quantiles of the distribution.
2. Strategies for robust hyperparameter tuning within our Gaussian process kernel as described in Rasmussen et



al [11].

3. Modeling a mixture of Gaussian processes instead of a single Gaussian process as outlined in Blei et al [2].

## 5. Contributions and acknowledgements

This work is conducted by Erich Trieschman as part of a larger project in collaboration with Kamran Tehrani. Kamran is responsible for running the production cost model, generating stochastic profile data, and data encoding. All code and results are available on GitHub at <https://github.com/etrieschman/grid-planner>.

## References

- [1] M. L. e. a. Awad. Using market simulations for economic assessment of transmission upgrades. pages 241–270, 2010. 1
- [2] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, apr 2017. 7
- [3] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004. 3
- [4] E. Brochu, V. M. Cora, and N. de Freitas. A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning, 2010. 3
- [5] A. Desai, C. Freeman, Z. Wang, and I. Beaver. Timevae: A variational auto-encoder for multivariate time series generation, 2021. 6
- [6] Grid Lab and Center for Energy Efficiency and Renewable Technologies. Transmission in California. Technical report, Mar. 2023. 1
- [7] B. F. Hobbs, Q. Xu, J. Ho, P. Donohoo, S. Kasina, J. Ouyang, S. W. Park, J. Eto, and V. Satyal. Adaptive Transmission Planning: Implementing a New Paradigm for Managing Economic Risks in Grid Expansion. *IEEE Power and Energy Magazine*, 14(4):30–40, July 2016. Conference Name: IEEE Power and Energy Magazine. 1
- [8] D. S. Liu. Order Instituting Rulemaking to Integrate and Refine Procurement Policies and Consider Long-Term Procurement Plans. 2
- [9] S. Liu, W. Huang, and Y. Zhang. A stochastic production simulation model for renewable integration and system flexibility studies. In *2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pages 1–8, Beijing, Oct. 2016. IEEE. 2
- [10] S. Mussmann and P. Liang. On the relationship between data efficiency and error for uncertainty sampling, 2018. 3
- [11] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 11 2005. 7
- [12] P. Sebastiani and H. P. Wynn. Maximum entropy sampling and optimal bayesian experimental design. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1):145–157, 2000. 3
- [13] H. Wang and Y. Ma. Optimal subsampling for quantile regression in big data, 2020. 6
- [14] J. Wang. An intuitive tutorial to gaussian processes regression, 2022. 3
- [15] Y. Xu, N. Myhrvold, D. Sivam, K. Mueller, D. J. Olsen, B. Xia, D. Livengood, V. Hunt, B. R. d’Orfeuil, D. Muldrew, M. Ondreicka, and M. Bettilyon. U.S. Test System with High Spatial and Temporal Resolution for Renewable Integration Studies, Feb. 2020. arXiv:2002.06155 [physics]. 2