Importance subsampling for power system planning under multi-year demand and weather uncertainty

Conference or Workshop Item

Accepted Version


It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.
Published version at: http://dx.doi.org/10.1109/PMAPS47429.2020.9183591
To link to this article DOI: http://dx.doi.org/10.1109/PMAPS47429.2020.9183591

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur
CentAUR

Central Archive at the University of Reading

Reading’s research outputs online
Importance subsampling for power system planning under multi-year demand and weather uncertainty

Adriaaan P. Hilbers
Department of Mathematics
Imperial College London
London, United Kingdom
ORCID: 0000-0002-9882-9479

David J. Brayshaw
Department of Meteorology
University of Reading
Reading, United Kingdom
ORCID: 0000-0002-3927-4362

Axel Gandy
Department of Mathematics
Imperial College London
London, United Kingdom
ORCID: 0000-0002-6777-0451

Abstract—This paper introduces a generalised version of importance subsampling for time series reduction/aggregation in optimisation-based power system planning models. Recent studies indicate that reliably determining optimal electricity (investment) strategy under climate variability requires the consideration of multiple years of demand and weather data. However, solving planning models over long simulation lengths is computationally unfeasible, and established time series reduction approaches induce significant errors. The importance subsampling method reliably estimates long-term planning model outputs at greatly reduced computational cost, allowing the consideration of multi-decadal samples. The key innovation is a systematic identification and preservation of relevant extreme events in modeling subsamples. Simulation studies on generation and transmission expansion planning models illustrate the method’s enhanced performance over established “representative days” clustering approaches. The models, data and sample code are made available as open-source software.

Index Terms—Power system modeling, weather, representative day, subsampling, temporal aggregation

I. INTRODUCTION

A. Demand and weather uncertainty and computational complexity in power system planning models

Many power system planning models estimate optimal electricity investment strategy by minimising a cost function given techno-economic assumptions and constraints [1]. The use of demand and weather time series inputs (e.g. hourly load levels and wind speeds) induces forward propagated demand and weather uncertainty in model outputs, particularly when variable renewable (VR) generation such as solar and wind is considered [2]. This is because an input time series may be viewed as a sample from some underlying demand and weather distribution, and a different sample, leading to different outputs, is equally valid.

Recent studies indicate that inter-annual climate/weather variability induces considerable output uncertainty in planning models. For example, optimal system design may depend strongly on what year of data is used [3]–[7], with some optimal capacities differing as much as 80% [8], [9]. Hence, to reliably determine long-term optimal electricity strategy, multiple years of demand and weather data should be considered.

At the same time, using long samples of time series input data is unfeasible in many power system planning models due to the computational complexity in solving the associated optimisation problem [10]. This effect is amplified since accurately modeling VR generation requires a high temporal and spatial resolution, precluding an increase in step length to reduce the number of time steps [11]–[13].

B. Time series reduction and subsampling

Power system modelers employ various schemes to reduce temporal complexity, as discussed in detail in [14]. A popular approach involves subsampling, often by clustering, into a smaller number of representative periods (typically days) that seek to encode longer time series in reduced form [15]–[19].

Most such reduction approaches are what the authors of [14] call a priori (or input-based) in that they depend only on the input time series but not on the underlying power system model. For example, clustering the same demand and weather data generates the same representative days (ignoring any randomness in the clustering algorithm) irrespective of whether these days are subsequently used in a highly renewable system or a system with only fossil fuel generation. In contrast, a posteriori methods use knowledge of the model’s structure or an estimate of its results. As an example, in [20], the authors run a planning model on each day individually and aggregate days by clustering (with additional modifications) the vectors of associated investment costs (model outputs). In this way, they use a subsample tailored to the model in question.

Many time series aggregation methods, such as clustering, replace individual periods/days by a group average. This “smoothes out” extremes and may lead to (1) overly optimistic predictions of system costs and (2) system design estimates that are not robust to such events. For example, such approaches may underestimate the amount of required backup generation capacity in events with high demand but low renewable generation, leading to systems with a high risk of blackouts [14]. For this reason, users sometimes make a priori heuristic adjustments such as including the maximum demand day [8].

This work was supported by the United Kingdom Engineering and Physical Sciences Research Council (EPSRC) Mathematics of Planet Earth Centre for Doctoral Training, grant number EP/L016613/1.

PMAPS 2020
C. This paper's contribution

This paper generalises significantly a version of the importance subsampling method for time series reduction/aggregation originally introduced in [9]. The contribution is as follows. Section I-A indicates that long demand and weather samples should be considered for robust power system planning under climate/weather variability, but that this is computationally challenging. Furthermore, time series reduction methods that remove extremes (Section I-B) lead to inaccurate estimates of optimal electricity strategy. The importance subsampling method systematically identifies such extremes in the full time series and includes them in a subsample, allowing reliable estimation of optimal multi-year electricity strategy at greatly reduced computational cost. A simulation study demonstrates the method’s improved performance over established clustering methods. The models, data and sample code for the method are available at [21].

This paper extends the approach introduced in [9] to almost general cost-minimisation based power system planning models. The requirements for the original method (analytical formula for generation cost, no link between time steps) restricted applications to models solvable using the (net) load duration curve. These stipulations are all relaxed. Furthermore, the generalised method allows, in addition to sampling individual time steps, the sampling of representative periods such as days or weeks (hence allowing for consideration of intra-period temporal dependencies) and uses clustering to leverage off their extensive use in power system applications.

This paper is structured as follows. Section II introduces the method and the intuition behind its working. Section III provides a case study of the method’s application to two generation and expansion planning (GTEP) models. Section IV discusses results and recommends extensions.

II. IMPORTANCE SUBSAMPLING

A. Overview and intuition

This paper considers planning models that determine optimal system design $D_{opt}$ (e.g. installed capacities of generation and transmission technologies) given some (possibly weighted) demand and weather data. The model $PSM_{plan}$ is viewed as a function from a sample $\mathcal{T}$ (of length $|\mathcal{T}|$) of demand and weather data to $D_{opt}$:

$$D_{opt} = PSM_{plan}(\mathcal{T}).$$  \hfill (1)

The model can also run in operational mode, in which a design $\mathcal{D}$ is specified and the optimal system operation, typically with the lowest generation cost, is determined:

$$\left(O_1, \ldots, O_{|\mathcal{T}|}\right) = PSM_{operate}(\mathcal{T}|\mathcal{D}).$$  \hfill (2)

Here, $O_t$ is the power system operation (generation and transmission levels) at time $t$, for each time step $t = 1, \ldots, |\mathcal{T}|$.

In many planning models (e.g. ones that seek the design minimising the sum of build costs and subsequent minimum generation costs), the operational model is solved as a subproblem. However, this is not necessary for the proposed method.

The setup is then as follows. Suppose $\mathcal{T}$ is a contiguous sample of demand and weather data and the associated optimal system design $D_{opt}$, as defined by (1), is desired. If $\mathcal{T}$ is long, solving the optimisation problem to determine $D_{opt}$ may be computationally unfeasible. The importance subsampling method instead estimates $D_{opt}$ by running the planning model across a shorter (weighted) subsample of time series data $S$. An estimate of $D_{opt}$ in (1) is hence provided by

$$D_{opt}^S := PSM_{plan}(S).$$  \hfill (3)

The goal is that $D_{opt}^S \approx D_{opt}$ with $|S| < |\mathcal{T}|$, so that optimal system design is reliably estimated at reduced computational cost. The key is hence how the sample $S$ is created.

The importance subsampling method works by explicitly including a selection of “extreme” events in a modeling subsample. In this paper, “extreme” refers to time steps in which meeting demand requires a large amount of generation or transmission capacity. In many planning models, omitting such extremes leads to inaccurate estimates of optimal system design (e.g. underestimation of required backup generation capacity, see Section I-A). It is hence essential that such events are identified and preserved when subsampling. At the same time, identifying them a priori is difficult since which time steps are “extreme” may depend on the system design being estimated. As a simple example, in a model containing an unknown wind generation capacity, it is unclear whether a 55MW demand, 0.1 wind capacity factor time step is more “extreme” (in the sense of requiring more backup capacity) than one with 60MW demand and 0.2 wind capacity factor.

Importance subsampling systematically identifies relevant extremes using their importance, calculated using the user-defined function IMP (see Section II-B). This function should be a proxy for how “extreme” a time step is. In many applications (e.g. Section III), a time step’s generation cost performs well since expensive measures (e.g. peaking plants or load curtailment) are activated only in scenarios when there is otherwise a supply shortage. Other choices are possible; electricity price is another candidate in models with such a notion. In practice, expert knowledge of the system should motivate a choice of importance function.

Determining each time step’s importance generally requires the associated generation and transmission levels, determined by running the model in operational mode with some fixed design. Ideally, importance levels are calculated using the true optimal system design, but this is the output being estimated and hence unknown. For this reason, a preliminary planning model run with a standard subsampling scheme is used for a rough estimate of optimal system design. This is subsequently used to determine the importance of each time step in the full time series.

At a high level, the approach works as follows (with correspondence to Section II-B shown in parentheses):

i) Estimate optimal system design through model run with standard subsampling (a priori) procedure (steps 1-2).
Using this estimate, calculate importance for each time step in full time series. In general, this requires running the model in operational mode (steps 3-4).

Create importance subsample by subsampling again while explicitly preserving the relevant extreme events (those with high importance) (step 5).

A model is subsequently run with the importance subsample \( S \) to estimate optimal system design.

### B. Algorithm

**Algorithm: Importance subsampling**

**Inputs:**
- \( \mathcal{T} \): full time series to sample from, of length \( |\mathcal{T}| \)
- \( n_d \): number of days to sample
- \( n_{de} \): number of “extreme” days to sample

**Steps:**
1. Cluster full time series into \( n_d \) weighted representative days to create \( C \).
2. Run planning model to determine \( D^C_{\text{opt}} \), the cluster estimates of optimal system design:
   \[
   D^C_{\text{opt}} = \text{PSM}_{\text{plan}}(C). \tag{4}
   \]
3. Determine system operation \( (O^C_1, \ldots, O^C_{|\mathcal{T}|}) \) (generation and transmission levels at each time in \( \mathcal{T} \)) by running model in operational mode with design \( D^C_{\text{opt}} \):
   \[
   (O^C_1, \ldots, O^C_{|\mathcal{T}|}) = \text{PSM}_{\text{operate}}(\mathcal{T}|D^C_{\text{opt}}). \tag{5}
   \]
4. Assign each time step its importance:
   \[
   \text{imp}_t = \text{IMP}(O^C_t) \quad \forall t \in \mathcal{T}. \tag{6}
   \]
   for user-defined IMP function (see Section II-A).
5. Create \( n_{de} \)-day importance subsample \( S \) containing:
   - \( n_{de} \) most “extreme” days in \( \mathcal{T} \) as ranked by the importance of their highest time step.
   - \( n_d - n_{de} \) days obtained by clustering the remainder of the full time series.

Weight days to account for their proportion across the full time series \( \mathcal{T} \) (see Section II-C).

**Outputs:**
- \( S \): weighted importance subsample

### C. Remarks

The proposed method is an a posteriori time series aggregation scheme that customises the subsample to the model. Most established time series reduction methods (Section I-A) are a priori schemes which do not use knowledge of the underlying model. This corresponds to following just step 1 in Section II-A (steps 1-2 in Section II-B) and using \( D^C_{\text{opt}} \) as the final estimate. Importance subsampling instead uses this estimate to create an additional customised sample to provide a second, more accurate, estimate of optimal system design.

Estimating optimal system design using importance subsampling requires three model runs (with correspondence to Section II-B shown in parantheses):
1. Preliminary planning model run on \( n_d \) days (step 2)
2. Operational model run with fixed capacities on full time series (step 3)
3. Planning model run on \( n_{de} \)-day importance subsample.

The importance subsample is weighted to account for the relative proportions of days throughout the full time series. For example, the 30 most extreme days in a 30-year sample represent one day a year, but will represent 30 days a year if all included a 1-year subsample. The time steps are hence weighted to avoid the estimated optimal system design being over-engineered for extreme events. The \( n_d - n_{de} \) non-extreme days are weighted according to their cluster sizes.

The clustering/subsampling algorithms used in steps 1 and 5 of Section II-B are customisable. Furthermore, periods other than days may be sampled. The advantage of sampling longer periods (e.g. weeks) is that it respects long-term temporal dependencies such as storage to a greater degree. The advantage of sampling shorter blocks is that the total number of periods in a fixed-length subsample is larger, meaning a higher diversity of possible demand and weather scenarios is considered. Hence, the choice of subsample blocks is a tradeoff between a high diversity of scenarios and a low distortion of time step chronology. In models without temporal dependencies, randomly sampling individual time steps (without any clustering) typically gives the best results, but longer periods should be selected in models with ramping or storage constraints.

### III. Simulation studies

#### A. Overview

This section examines the efficacy of the proposed method on sample power system planning models. This is done by comparing optimal system design across importance subsamples with that across the full time series from which samples are drawn. As a benchmark, the model is also run across samples generated by \( k \)-medoids clustering, representing the current industry standard. The reason for choosing \( k \)-medoids clustering (where the real day closest to the cluster mean is used as the representative day, instead of the mean itself) is its widespread use in previous studies and that it typically performs better than \( k \)-means [14], which smooths out extremes by using the cluster mean.
Two metrics indicating the “degree of suboptimality” of estimated system design (compared with the true optimum across the full time series) are also calculated:

- **Peak capacity shortage**: maximum value of demand minus available generation capacity, in MW.
- **Energy unserved**: (annualised) demand impossible to meet with generation capacity, in MWh/yr. Can be turned into monetary value via a value of lost load.

These metrics are determined by running a model, with design estimated across subsamples, in operational mode across the full time series.

### B. Power system planning models

The method is tested on two (generation and transmission expansion) planning models with topologies equal to a renewables-containing version of the IEEE 6-bus system originally introduced in [23], [24]. Optimal system design (generation and transmission capacities across the grid) is determined by minimising the total system cost (sum of install and generation costs) given hourly demand levels and wind capacity factors. The model determines the optimal capacity and generation mix across baseload, peaking and wind technologies. Time series input data consists of hourly demand levels and wind capacity factors in European countries, introduced in [25], [26]. An overview of the model’s topology is provided in Fig. 1. The models and time series data are publicly available at [27], where their mathematical formulations are described.

The two models differ in the complexity of their constraints. The first, the LP model, is a linear program (LP) in which technologies may be installed to any nonnegative capacity and there is no intertemporal dependence between time steps. In the MILP model, baseload capacity is constrained to blocks of 3GW and a ramp rate of 20%/hr, making the model a mixed-integer linear program (MILP) with ramping constraints.

### C. Simulation setup

Details for the simulations are provided in Table I. For the LP model, optimal design across 2008-2017 serves as the target for simulations with \( n_d = 48 \) and 90-day subsamples. For the MILP model, the integer constraints impose (much) longer solution times and the optimal design across 2017 is estimated using subsamples of length \( n_d = 24 \) and 48 days. Since both \( k \)-medoids clustering and importance subsampling contains randomness from the initialisation of cluster means, the distribution of model outputs across 100 subsamples is determined.

Computational times are based on model runs created in the open-source energy modeling framework Calliope [22] and solved using the cbc [28] solver on a 2.7GHz Intel Core i5-5257U processor with 8GB of RAM (with some additional “swap” memory available).

The implementation of the importance subsampling method (Section II-B) is as follows. \( k \)-medoids clustering is used in steps 1 and 5. Each time series is normalised to lie between 0 and 1 so that demand levels and wind capacity factors can be compared fairly, but no additional adjustments are made. For the importance function in step 4, a time step’s generation cost (sum of each technology’s generation cost times generation level) is used since expensive technologies with high generation cost (e.g. peaking plants) are used only in “extreme” scenarios when blackouts would occur otherwise. For \( n_d \), the number of “extreme” days containing the highest generation costs, a value of \( n_d = \frac{1}{4}n_d \) is used as this gives good results in practice, balancing the need to capture a range of “extreme” events while leaving enough of the sample to accurately approximate the more typical days.

### D. Results

Fig. 2 shows the distribution of estimated optimal capacities and suboptimality metrics (peak capacity shortage, energy unserved) across the test simulations. For the same plots among the full range of model outputs (at regional level) and for a larger number of sample sizes, including sampling individual hourly time steps instead of days, visit [21]. Importance subsampling into representative days generally has superior performance to \( k \)-medoids clustering, with more accurate estimates of optimal system design, lower levels of capacity shortage and energy unserved, and lower variability in model outputs. For example, \( k \)-medoids clustering usually un-
**LP model, full time series 2008-2017**

(a) $n_d=48$-day subsample, $n_d=16$ “extreme” days

(b) $n_d=90$-day subsample, $n_d=30$ “extreme” days

**MILP model, full time series 2017**

(c) $n_d=24$-day subsample, $n_d=8$ “extreme” days

(d) $n_d=48$-day subsample, $n_d=16$ “extreme” days

![Diagram showing distribution of estimated optimal capacities and suboptimality metrics across 100 samples generated by k-medoids clustering and importance subsampling. The box-and-whiskers show the 2.5%, 25%, median, 75% and 97.5% percentiles. The optima across the full time series (targets) are shown as a dashed black line.](image)

This paper’s main contribution is a significant generalisation of the importance subsampling method. This approach can reliably estimate long-term planning model outputs (considering multiple decades of demand and weather data) at greatly reduced computational cost, outperforming industry standard time series aggregation methods. As discussed in Section I-A, the consideration of such long samples is essential for...
robust power system planning under natural climate/weather variability. This method also allows the consideration of climate model data in power system planning (see e.g. [30]–[32]), where differences in climate have a weak signal-to-noise ratio under natural variability and hence require long samples to identify statistically significant effects. The models, data and sample code are available at [21].

The key innovation behind the proposed methodology is a systematic identification and inclusion of relevant “extreme” events in subsamples through a user-defined importance function. This function may be tailored to the specific application, but in many settings a time step’s generation cost or electricity price will suffice (Section II-C). The systematic identification of relevant extremes may also be used in its own right, for example by using the identified “extreme” days for subsequent reliability analysis at higher spatiotemporal resolution.

Importance subsampling is an a posteriori method (see Section I-B) in which properties of the model and/or its results are used to improve time series aggregation. The method can be compared with that in [20], where the time series aggregation is performed, along with a number of other modifications, across the vectors of model outputs (costs at technological and regional levels) for each individual day. This is another promising approach to a similar problem. Importance subsampling is simpler, though a more thorough comparison of the two methods would be interesting.

Two potential extensions to the method are further refinements to model interdependence between the sampled days (as in e.g. weekly/seasonal storage), and dimensionality reduction when the number of grid locations becomes very large.

REFERENCES