University of Reading

Importance subsampling: improving power system planning under climate-based uncertainty

Article

Accepted Version

Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Hilbers, A. P., Brayshaw, D. J. ORCID: https://orcid.org/0000-0002-3927-4362 and Gandy, A. (2019) Importance subsampling: improving power system planning under climate-based uncertainty. Applied Energy, 251. 113114. ISSN 0306-2619 doi: 10.1016/j.apenergy.2019.04.110 Available at https://centaur.reading.ac.uk/83771/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1016/j.apenergy.2019.04.110

Publisher: Elsevier

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

University of Reading

CentAUR

Central Archive at the University of Reading

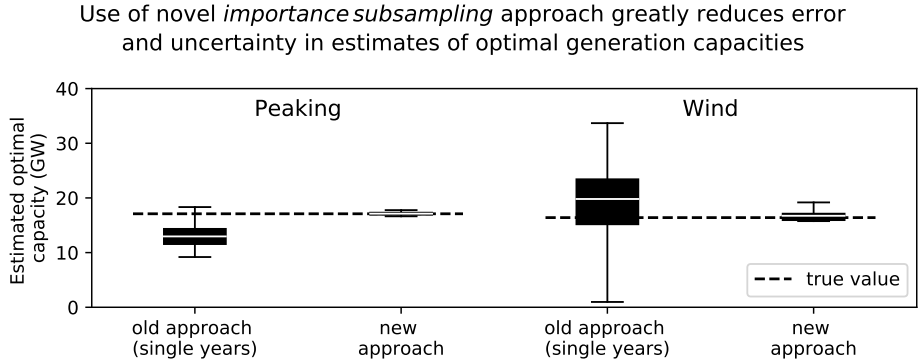
Reading's research outputs online

Importance Subsampling: Improving Power System Planning Under Climate-based Uncertainty

Adriaan P Hilbers[a,*], David J Brayshaw[b,c], Axel Gandy[a]

[a] *Department of Mathematics, Imperial College London*
[b] *Department of Meteorology, University of Reading*
[c] *National Centre for Atmospheric Science, University of Reading*

Abstract



Recent studies indicate that the effects of inter-annual climate-based variability in power system planning are significant and that long samples of demand & weather data (spanning multiple decades) should be considered. At the same time, modelling renewable generation such as solar and wind requires high temporal resolution to capture fluctuations in output levels. In many realistic power system models, using long samples at high temporal resolution is computationally unfeasible. This paper introduces a novel subsampling approach, referred to as *importance subsampling*, allowing the use of multiple decades of demand & weather data in power system planning models at reduced computational cost. The methodology can be applied in a wide class of optimisation-based power system simulations. A test case is performed on a model of the United Kingdom created using the open-source modelling framework *Calliope* and 36 years of hourly demand and wind data. Standard data reduction approaches such as using individual years or clustering into representative days lead to significant errors in estimates of optimal system design. Furthermore, the resultant power systems lead to supply capacity shortages, raising questions of generation capacity adequacy. In contrast, *importance subsampling* leads to accurate estimates of optimal system design at greatly reduced computational cost, with resultant power systems able to meet demand across all 36 years of demand & weather scenarios.

Keywords: energy modelling, power systems, renewable energy, weather variability, timeseries reduction, modelling methods

*Corresponding author
Email address: aph416@ic.ac.uk (Adriaan P Hilbers)

1. Introduction

1.1. Context

In the face of climate change, worldwide efforts are being undertaken to reduce carbon emissions (IPCC, 2018). A common roadmap to sustainability is to decarbonise electricity supply and electrify other sectors such as as transport and heating (Staffell, 2017). In many countries, an essential part of this strategy is an increased use of variable renewable energy (VRE) generation such as solar and wind (Jacobson et al., 2015; National Grid, 2017). Hence, while electricity demand has always exhibited some weather-dependence (see e.g. Thornton et al., 2017), the same increasingly holds for supply.

To aid in questions of energy strategy, decision-makers frequently employ energy system models (ESMs), computer programs that simulate the energy transitions in a given geographical region (Ventosa et al., 2005). Power system models (PSMs) form a subset concerned primarily with the electricity sector. A common use of PSMs is to calculate the optimal *generation mix* by minimising the sum of installation and generation costs while meeting demand (Stoft, 2002; Bazmi & Zahedi, 2011; Ringkjøb et al., 2018). Models considering renewables require coherent weather timeseries such as windspeeds or solar irradiances as inputs.

1.2. The computational cost of weather & climate variability

Recent studies indicate that robust power system planning under natural climate variability requires long samples of demand and weather data (spanning multiple decades). In particular, characteristics of power systems, such as the optimal installed capacities of different generation technologies, may be highly dependent on which year of data is used. Hence, power systems designed using single-year simulations may be suboptimal in the long run (Bloomfield et al., 2016; Pfenninger, 2017; Zeyringer et al., 2018; Wohland et al., 2018). The degree of this dependence is expected to increase as more VRE generation is employed (Staffell & Pfenninger, 2018; Collins et al., 2018).

The need for long samples poses a computational challenge since accurately modelling systems with significant shares of VRE generation also requires high temporal resolution. Previous studies indicate that models with low resolution fail to capture VRE output fluctuations and underestimate required flexible and dispatchable generation capacity (De Jonghe et al., 2011; Poncelet et al., 2016; Collins et al., 2017; Kotzur et al., 2018). For many realistic PSMs, it is computationally unfeasible to solve the optimisation problem using both a long sample of data and a high temporal resolution (Pfenninger, 2017).

1.3. Established data reduction approaches

Various approaches to reduce computational cost without lowering temporal resolution exist. One strategy is the *soft-linking* of a long-term planning model with a more detailed simulation operating on a shorter scale (Ringkjøb et al., 2018; Zeyringer et al., 2018; Collins et al., 2017). Another is to run a model with a smaller number of representative periods (e.g. days or weeks) obtained by clustering the full dataset. Numerous studies explore the efficacy of such approaches in reproducing model outputs at reduced computational expense (de Sisternes & Webster, 2013; Pfenninger, 2017; Nahmmacher et al., 2016; Kotzur et al., 2018; Härtel et al., 2017; Poncelet et al., 2017). They arrive at a number of common conclusions. Firstly, the reduction approach must not remove extremes (e.g. by "averaging away" peaks) since including them ensures the model determines power system design able to meet demand in such scenarios. For this reason, heuristic adjustments such as including the maximum demand day are sometimes employed. Furthermore, clustering typically works poorly when applied to multiple decades of timeseries data since small changes in approach may lead to large spreads in model outputs. For example, Pfenninger (2017) clusters

25 years of demand & weather data and finds that optimal wind capacity is anywhere between 0.8 and 2.8 times peak demand depending on the choice of clustering algorithm and heuristic adjustment, giving the user virtually no indication of a good investment strategy. In addition, clustering does not generalise easily to models taking a large number of input timeseries. Consider a model with hourly resolution for 10 demand regions, 5 wind farms and 5 concentrated solar power plants. Clustering the days requires clustering vectors of length 480 (=24·(10+5+5)).

1.4. This paper's contribution

This paper introduces a novel subsampling approach, called *importance subsampling*, that can be applied to multiple decades of demand & weather timeseries data. PSM outputs evaluated using subsamples reliably estimate those found using the full timeseries at greatly reduced computational cost. The methodology is introduced in full generality and can be either directly applied or straightforwardly generalised to a wide class of optimisation-based PSMs.

A test case is performed using a model of the United Kingdom power system created using the open-source energy modelling framework *Calliope* (see Pfenninger & Pickering, 2018) and 36 years of UK-wide demand and wind data. Model outputs using *importance subsampling* reproduce those using all 36 years of data more reliably than alternative approaches such as using individual years or clustering days.

This paper is structured as follows. Section 2 provides motivation, outlining the risks in designing power systems using short samples of demand & weather data. Section 3 describes the *importance subsampling* approach in detail and full generality. Section 4 evaluates its use in a test case, comparing results to those found using individual years or by clustering timeseries into a number of representative days. Section 5 discusses the results' implications and recommends potential extensions. In the appendix (Section 6), the reader finds full descriptions of the PSM and datasets employed.

2. Motivation: climate variability in power system planning

This section provides a concrete illustration of the problem of inter-year climate-based variability in power system planning. A PSM of the UK is considered along with 36 years of hourly demand and wind levels (over the period 1980 to 2015). The demand timeseries is detrended so that the only differences between years are climate-driven (e.g. demand being higher when it is cold, see Section 6.2.1 for details).

The box-and-whisker plot in figure 1(a) shows the distribution of optimal generation capacities of 4 technology types across the 36 individual years of historic data (i.e. for 1980, 1981, ... , 2014, 2015). The white horizontal line inside the box indicates the median and the wider dashed horizontal line shows the optimal capacity when all 36 years of data considered at once: the best estimate for the "truly" optimal design. Two phenomena stand out. The first is the considerable inter-year spread, particularly for wind. Using 2010 data indicates almost no wind at all should be built (1.0GW), whereas the optimal configuration for 1986 has more wind capacity than any other technology (33.7GW). The second is the bias between the median of individual-year model runs and the 36-year optimal capacities (seen by comparing the white line on the box plots with the wider dashed black line). For example, 92% of individual-year model runs underestimate optimal peaking capacity, with a median underestimation of around 20%. The opposite is true for wind, for which the median individual-year wind capacity overestimates the 36-year optimum by around 15%.

Furthermore, a power system designed using one year's optimal capacities may lead to supply capacity shortages in another. Figure 1(b) shows the distribution of the number of hours with supply capacity shortages (insufficient capacity to meet demand) as a function of the year used to determine the optimal capacities (in other

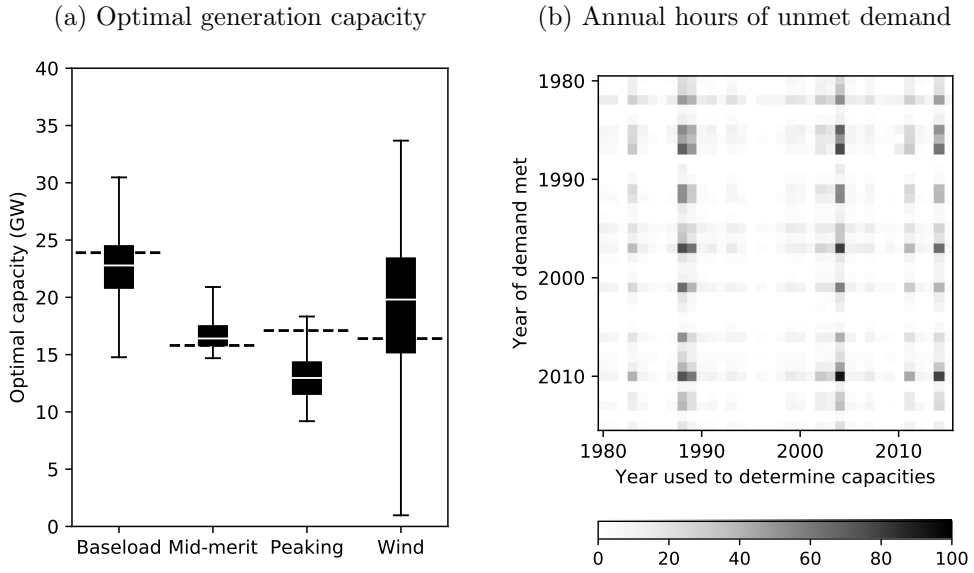


Figure 1: (a) Distribution of optimal capacities across individual-year simulations from 1980 to 2015. The box shows the 25th, median and 75th percentile, and the whiskers show the minimum and maximum. The optimal capacities for a simulation using all 36 years of data are shown as a wider dashed line for comparison and serve as the best guess for the "true" optima. (b) Number of hours with supply capacity shortages. The x-axis indicates the year of data used to determine the optimal capacities and the y-axis the year of demand & weather data in which to meet demand. For example, the top right square shows the number of hours in which a system designed using 2015's optimal capacities fails to meet demand when used in the year 1980.

words, the system is "designed" on basis of the data of the year indicated on the x-axis and asked to meet conditions of the year on the y-axis). The risk of picking the "wrong year" on which to base system design is high: one designed using 2004's capacities fails to meet 93 hours of demand in 2010, and an average of 30 hours annually across all 36 years. Across the whole dataset, a power system designed using a random choice of year has a 50% chance of at least one hour of supply capacity shortage in another, and a 35% chance of shortage for at least 3 hours. This highlights the risks of informing power system strategy on single years and illustrates the need to consider longer samples.

3. Methodology

This section introduces the precise *importance subsampling* methodology in full generality. Like the clustering strategies discussed in Section 1.3, it is based on time-series compression through subsampling. The specific application to a power system planning problem is presented in Section 4.

3.1. Intuition

The *importance subsampling* methodology works by subsampling timesteps from the full dataset. They are always selected in full; if one timeseries value at timestep t is sampled, then so are all others. This ensures correlations between timeseries (e.g. spatial windspeed correlations or demand-wind correlations) are automatically accounted for.

Determining which timesteps to sample can be complicated. In Section 2, a system designed using 2004 data leads to supply capacity shortages in other years because all of 2004's high-demand timesteps have high wind levels. As a result, the 2004 optimal design underestimates the amount of required backup (dispatchable) generation capacity. This highlights the fact that cost-optimal power system design will never have

excess capacity; it will *just* be able to meet demand in the "worst" timestep (the one requiring the largest supply). Indeed, this is the reason that heuristic adjustments such as including the maximum demand day are sometimes employed in the established subsampling approaches outlined in Section 1.3. The disadvantage of heuristic methods is that, when VRE generation is involved, they may fail to identify those timesteps truly required to ensure generation capacity adequacy. The peak demand in 2004 is not lower than in other years, so including the maximum demand day fails to mitigate the underestimation of required generation capacity. In this case, it may be necessary to sample instead a day with a slightly lower demand but virtually no wind. Determining the correct trade-off between high demand and low wind (e.g. is a timestep with 50GW demand and 0.1 wind capacity factor "worse" than one with 55GW peak demand and 0.2 wind capacity factor?) is difficult *a priori* and will differ between PSMs.

The *importance subsampling* approach identifies the essential timesteps in a systematic way by assigning each one a measure of the difficulty in meeting demand, referred to as its *importance*. This has two advantages. Firstly, it one-dimensionalises timeseries inputs: a timestep with many values (e.g. demand levels and renewable outputs across multiple regions) is assigned a single *importance*. This allows timesteps to be ranked objectively irrespective of the PSM's structure. Secondly, the ranking means a selection of timestep(s) with the highest *importance* can be included forcibly into the modelling sample to ensure generation capacity adequacy in the resultant power system.

A timestep's *importance* may depend on the power system design. In single-region models, the *net demand* (residual demand after all renewable generation is used) is a good candidate since it is equal to the required dispatchable generation. The net demand, however, depends on the installed renewable capacity. For example, a timestep with a demand of 50GW and wind capacity factor 0.1 has a higher net demand than one with a demand of 55GW and wind capacity factor of 0.3 exactly when the installed wind capacity is less than 25GW. Since the installed wind capacity is itself a model output, a "Catch 22" situation occurs: determining which timesteps to use in a simulation requires an estimate for each timestep's *importance*, which itself requires the model outputs.

A two-stage approach is therefore proposed. A stage 1 optimisation run, using a random sample of timesteps, gives a rough indication of optimal power system design. This is used to estimate each timestep's *importance*. A stage 2 sample is subsequently created by including a number of the timesteps with the highest *importance* and a random selection of those remaining. This is used in a second model run to to estimate the model outputs found using the full dataset. The approach is dependent on the choice of *importance* function, which should be a proxy for the "difficulty" in meeting demand (in terms of requiring a large amount of generation capacity). In the case study (Section 4), a timestep's *variable cost* is used and this choice is applicable in a large class of PSMs. However, others are possible, and the choice may be tailored through consideration of the PSM or from expert knowledge.

3.2. Precise methodology

The following terminology is employed throughout this section:

- $\mathcal{T}^F = \{1, \ldots, N^F\}$: full set of timesteps to be sampled from.
- ts_data_t: timeseries input data in timestep t. For example, in a model with demand levels and wind capacity factors, $\text{ts_data}_t = (\text{demand}_t, \text{wind}_t)$.
- weight_t: weight assigned to timestep t in a PSM. For example, when clustering timeseries into representative periods (see Section 1.3), periods are typically weighted for their relative cluster sizes.

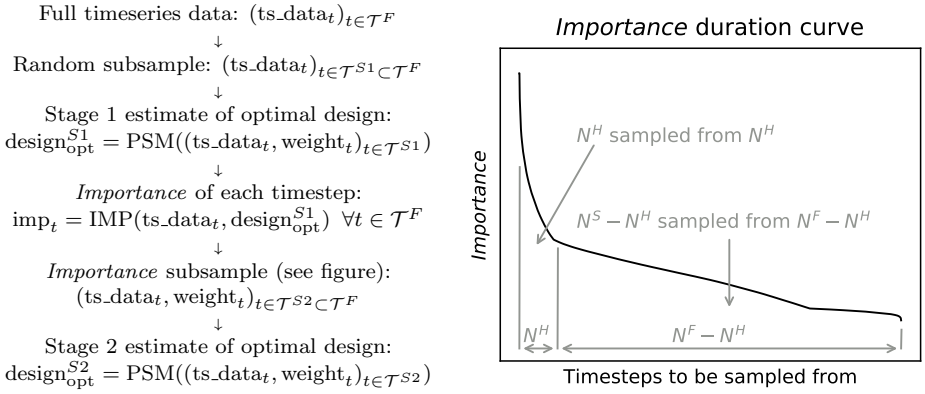


Figure 2: *Importance subsampling* methodology. The *importance* duration curve is constructed by plotting timesteps in descending order of *importance* in direct analogy to the *load duration curve* common across power system analysis. The *importance subsample* contains, from the full timeseries, the N^H timesteps with the highest *importance* and a random selection of size $N^S - N^H$ from those remaining. Timesteps are weighted to reflect their relative proportions across the full dataset as described in Section 3.2.

- $\text{design}_{\text{opt}}$: optimal power system design, e.g. a vector with each generation technology's installed capacity.

One furthermore needs two functions:

- A power system model with an optimiser that returns, given some demand & weather timeseries data and timestep weights indexed by timesteps $\mathcal{T}$, the optimal power system design:

$$\text{design}_{\text{opt}} = \text{PSM}((\text{ts_data}_t, \text{weight}_t)_{t\in\mathcal{T}}). \tag{1}$$

- An *importance* function that assigns to each timestep a measure of the difficulty in meeting demand. This may depend on the power system design:

$$\text{imp}_t = \text{IMP}(\text{ts_data}_t, \text{design}_{\text{opt}}). \tag{2}$$

Importance subsampling works as follows. Suppose one wants to determine, given some (equally weighted) timeseries data $(\text{ts_data}_t, \text{weight}_t)_{t\in\mathcal{T}^F}$, the optimal power system design $\text{design}^F_{\text{opt}} = \text{PSM}((\text{ts_data}_t, \text{weight}_t)_{t\in\mathcal{T}^F})$. Evaluating this function may be too computationally expensive, so $\text{design}^F_{\text{opt}}$ is estimated as follows:

1. Randomly sample N^S timesteps from the N^F to create a stage 1 subsample $\mathcal{T}^{\text{S1}} \subset \mathcal{T}^F$ of length N^S with equal weights.
2. Determine optimal design for stage 1 subsample:

$$\text{design}^{\text{S1}}_{\text{opt}} = \text{PSM}((\text{ts_data}_t, \text{weight}_t)_{t\in\mathcal{T}^{\text{S1}}}). \tag{3}$$

3. Calculate *importance* of each timestep in the full dataset using stage 1 design:

$$\text{imp}_t = \text{IMP}(\text{ts_data}_t, \text{design}^{S1}_{\text{opt}})\ \ \forall\, t \in \mathcal{T}^F. \tag{4}$$

4. Create a stage 2 sample of size N^S using the N^H timesteps with the highest *importance* and a random selection of those remaining:

$$\mathcal{T}^{\text{S2}} = \{\underbrace{t^{\text{S2}}_1, t^{\text{S2}}_2, \ldots, t^{\text{S2}}_{N^H-1}, t^{\text{S2}}_{N^H}}_{\substack{\text{Timesteps with} \\ \text{highest } \textit{importance}}}, \underbrace{t^{\text{S2}}_{N^H+1}, t^{\text{S2}}_{N^H+2}, \ldots, t^{\text{S2}}_{N^S-1}, t^{\text{S2}}_{N^S}}_{\substack{\text{Random selection} \\ \text{of those remaining}}}\} \subset \mathcal{T}^F \tag{5}$$

with associated weights

$$\text{weight}_t^{\text{S2}} = \begin{cases} \frac{1}{N^F} & t \in \{t_1^{\text{S2}}, \ldots, t_{N^H}^{\text{S2}}\} \\ \frac{N^F - N^H}{N^F(N^S - N^H)} & t \in \{t_{N^H+1}^{\text{S2}}, \ldots, t_{N^S}^{\text{S2}}\}. \end{cases} \tag{6}$$

Figure 2 provides an illustration of this step.

5. Determine optimal design for stage 2 subsample:

$$\text{design}_{\text{opt}}^{S2} = \text{PSM}((\text{ts_data}_t, \text{weight}_t)_{t \in \mathcal{T}^{S2}}). \tag{7}$$

$\text{design}_{\text{opt}}^{\text{S2}}$ is the estimate of $\text{design}_{\text{opt}}^{F}$.

Each evaluation of $\text{design}_{\text{opt}}^{\text{S2}}$ involves 2 evaluations of optimal design (stage 1 and stage 2) using N^S timesteps each. Since the computational cost of the other steps (sampling and evaluating *importance*) is typically negligible, the cost of evaluating $\text{design}_{\text{opt}}^{S2}$ is twice that of a single evaluation with sample size N^S.

The weights in step 4 account for the relative proportions of the two bins. For example, suppose an *importance subsample* of size 120 timesteps is created using the 60 timesteps with the highest *importance* and a random selection of 60 from those remaining. Half of this subsample consists of timesteps with a high *importance* even though they represent a much smaller proportion of the full dataset. The weights are chosen to cancel out this oversampling and prevent overengineering for extreme scenarios. The N^H timesteps with the highest *importance* represent a proportion $\frac{N^H}{N^F}$ of the dataset and the other $N^S - N^H$ represent $\frac{N^F - N^H}{N^F}$. The weights account for this:

$$\sum_{k=1}^{N^S} \text{weight}_{t_k^{\text{S2}}}^{S2} = \sum_{k=1}^{N^H} \frac{1}{N^F} + \sum_{k=N^H+1}^{N^S} \frac{N^F - N^H}{N^F(N^S - N^H)} \tag{8}$$

$$= \underbrace{\frac{N^H}{N^F}}_{\substack{\text{Correct proportion} \\ \text{of total dataset}}} + \underbrace{\frac{N^F - N^H}{N^F}}_{\substack{\text{Correct proportion} \\ \text{of total dataset}}} = 1 \tag{9}$$

4. Test case: UK power system model

4.1. Overview

In this section, the performance of the *importance subsampling* approach is compared with other subsampling strategies when applied to a PSM based on the United Kingdom (UK) and created in the open-source energy modelling framework *Calliope* (see Pfenninger & Pickering, 2018). This model is designed as an idealised test case and should not be viewed as a realistic representation of the UK power system or used to inform policy or strategy. It is employed only since it uses a similar linear programming formulation as many PSMs popular in the energy community (see e.g. Trutnevyte, 2016; Bazmi & Zahedi, 2011; Hall & Buckley, 2016) and results on this model can be reasonably expected to generalise to those models also.

Hourly UK-wide demand levels and wind capacity factors are estimated over the 36-year period from 1980 until 2015 and serve as the model's timeseries inputs. Long-term anthropogenic trends such as economic growth and efficiency improvements are removed so that different years of demand data can be fairly compared. Model outputs are the optimal installed capacities of 4 possible generation technologies. The first 3 (baseload, mid-merit and peaking) are generic dispatchable technologies that differ only in their installation and generation costs. The 4th technology, wind, has 0 generation cost but output capped by time-varying wind levels. The PSM and input timeseries are discussed fully in the appendix (Section 6).

A "perfect model" framework is assumed: the optimal capacities across the 36 years are taken to be those that minimise cost under the "true" distribution of demand and wind. The 36-year capacities (for baseload, mid-merit, peaking and wind technologies) hence serve as targets:

$$[\text{cap}_b^{36\text{y}}, \text{cap}_m^{36\text{y}}, \text{cap}_p^{36\text{y}}, \text{cap}_w^{36\text{y}}] = \text{PSM}((d_t, w_t, \text{weight}_t)_{t\in\mathcal{T}^{36\text{y}}}) \tag{10}$$

where d_t is the demand level and w_t is the wind capacity factor in timestep t. Each timestep is assigned equal weight. A subsample $\hat{\mathcal{T}} \subset \mathcal{T}^{36\text{y}}$ generates estimators of the optimal capacities as defined by

$$[\hat{\text{cap}}_b, \hat{\text{cap}}_m, \hat{\text{cap}}_p, \hat{\text{cap}}_w] = \text{PSM}((d_t, w_t, \text{weight}_t)_{t\in\hat{\mathcal{T}}}). \tag{11}$$

Subsampling strategies are evaluated on a number of criteria. The estimators defined by equation (11) should have both a low variation across samples generated by the same process and a low bias (understood as median error) when compared to the targets in equation (10). Furthermore, suppose a power system is designed using the estimated optimal capacities in equation (11). For this (hypothetical) system, two statistics are calculated:

- *hours of unmet demand*: number of hours in the full sample in which the power system has insufficient generation capacity to meet demand. For example, the optimal power system design for 1980 data may be unable to meet demand for some hours in the period 1981-2015, leading to hypothetical supply shortages.

- *extra system cost*: additional cost (sum of installation and generation costs) of meeting the 36-year demand using a suboptimal power system. A PSM user might be aware of the risks in designing systems using short samples and want to compensate. The cheapest way to do this *a priori* (i.e. without more optimisation) is to use extra peaking capacity to meet any unmet demand. Define the *extra system cost* of the sample $\hat{\mathcal{T}}$ by

$$\text{extra system cost}^{\hat{\mathcal{T}}} = (\text{system cost}^{\hat{\mathcal{T}}} - \text{system cost}^{\text{opt}}) \tag{12}$$

  where $\text{system cost}^{\hat{\mathcal{T}}}$ is the cost of a system designed using sample $\hat{\mathcal{T}}$ (with extra peaking capacity and generation to ensure no unmet demand) and $\text{system cost}^{\text{opt}}$ is the cost of the 36-year optimal system. Since it is by definition cost-optimal, equation (12) is nonnegative.

Computational cost increases with timeseries length, so samples should be made as short as possible. The computational expense in solving the optimisation problem in this model scales linearly in the number of timesteps. In some PSMs (e.g. mixed-integer linear programs), the computational effort increases faster than linearly, in which case a reduction in simulation length causes a proportionally larger decrease in computational cost.

Four subsampling strategies are investigated: using individual years, random sampling of timesteps, k-medoids clustering into representative days and *importance subsampling*. In k-medoids clustering, samples are generated by clustering the 48-dimensional vectors of each individual day's (normalised) hourly demand and wind levels into k clusters. Each cluster's representative day is the day in the full timeseries whose vector is closest to the cluster mean. The model is run on the k representative days weighted by the number of days their cluster contains.

4.2. Importance subsampling: setup

Recall from Section 3 that *importance subsampling* requires a PSM that can determine optimal system design. The test case PSM is an example of this. Here, the design output is the vector of optimal installed capacities of the four technologies.

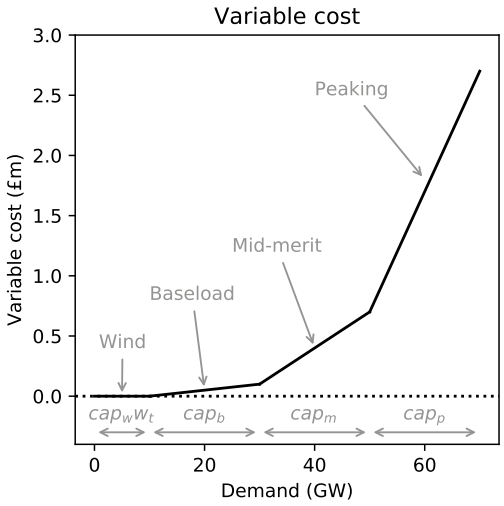


Figure 3: A timesteps's variable cost as a function of demand. The first $\text{cap}_w w_t$ (installed wind capacity times wind capacity factor) GW are met by wind (with no contribution to variable cost), followed by baseload, mid-merit and peaking. The variable cost increases more quickly as progressively more expensive generation technologies are employed.

It is determined by minimising the sum of installation and generation costs while meeting demand. Input data is hourly UK-wide demand and wind capacity factors:

$$\text{ts_data}_t = (d_t, w_t) \tag{13}$$

$$\text{design}_{\text{opt}} = [\text{cap}_b, \text{cap}_m, \text{cap}_p, \text{cap}_w] \tag{14}$$

where d_t and w_t are the demand level and wind capacity factor in timestep t respectively. The second requirement is an *importance function*. A timestep's *variable cost* is proposed for this purpose:

$$\text{imp}_t = \text{varcost}_t = f_b \text{gen}_{bt} + f_m \text{gen}_{mt} + f_p \text{gen}_{pt} + f_w \text{gen}_{wt} \tag{15}$$

$$= f_b \text{gen}_{bt} + f_m \text{gen}_{mt} + f_p \text{gen}_{pt} \tag{16}$$

where f_i is technology i's cost per unit electricity generated (in £/GWh) and gen_{it} is the amount of electricity generated by technology i in timestep t (in GWh). The costs per unit electricity of the different technologies are shown in Table 2 in the appendix. Wind is assumed to generate electricity at 0 marginal cost: $f_w = 0$.

A timestep's variable cost as a function of its demand is shown in Figure 3. Demand is met by *merit-order stacking* of technologies in ascending order of generation cost, with all available wind power used first, followed by baseload, mid-merit and peaking. The first $\text{cap}_w w_t$ GW (the installed capacity times the timestep's wind capacity factor) of demand are met by wind power with no contribution to the variable cost. The next cap_b GW are met by baseload, followed by cap_m GW of mid-merit and cap_p GW of peaking (any unmet demand is also assumed to be met by peaking). The variable cost increases more quickly as technologies with progressively higher generation cost are used. It depends implicitly on the timestep's demand level and wind capacity factor, as well as the installed capacity of the different generation technologies, through the generation terms in equation (16) and as summarised in Figure 3. It is a convenient proxy for the "difficulty" in meeting demand: timesteps with a high variable cost are exactly those requiring the most dispatchable generation capacity.

The optimal capacities across samples of varying total size N^S are determined. Each sample includes the $N^H = 60$ timesteps with the highest estimated variable cost

and a random sample of those remaining. For additional clarity, the steps introduced in Section 3 applied to the test case model are explicitly presented below.

1. Randomly sample N^S timesteps from the 36-year dataset to create an equally weighted stage 1 subsample $(d_t, w_t, \text{weight}_t)_{t\in\mathcal{T}^{\text{S1}}}$
2. Determine optimal capacities:
$$[\text{cap}_b^{\text{S1}}, \text{cap}_m^{\text{S1}}, \text{cap}_p^{\text{S1}}, \text{cap}_w^{\text{S1}}] = \text{PSM}((d_t, w_t, \text{weight}_t)_{t\in\mathcal{T}^{\text{S1}}}). \tag{17}$$
3. Estimate variable cost of each timestep in full dataset using equation (16). The variable cost is the specific choice of *importance* function:
$$\begin{aligned}\text{varcost}_t &= \text{varcost}(d_t, w_t, [\text{cap}_b^{\text{S1}}, \text{cap}_m^{\text{S1}}, \text{cap}_p^{\text{S1}}, \text{cap}_w^{\text{S1}}]) &(18)\\ &= f_b\text{gen}_{bt} + f_m\text{gen}_{mt} + f_p\text{gen}_{pt}. &(19)\end{aligned}$$
4. From the full dataset, construct stage 2 subsample by including the $N^H = 60$ timesteps with the highest variable cost and a random sample of size $N^S - N^H$ from those remaining. Weight the timesteps in each of the 2 bins to reflect their relative proportions throughout the full dataset:
$$\text{weight}_t = \begin{cases} \frac{1}{315360} & 60 \text{ timesteps with highest variable cost} \\ \frac{315360-60}{315360(N^S-60)} & \text{otherwise.} \end{cases} \tag{20}$$
where $315360 = 8760{\cdot}36$ is the total number of hourly timesteps in the 36-year dataset.
5. Determine optimal capacities for the stage 2 subsample:
$$[\text{cap}_b^{\text{S2}}, \text{cap}_m^{\text{S2}}, \text{cap}_p^{\text{S2}}, \text{cap}_w^{\text{S2}}] = \text{PSM}((d_t, w_t, \text{weight}_t)_{t\in\mathcal{T}^{\text{S2}}}). \tag{21}$$
This is the estimate for the optimal system design.

4.3. Results

Figure 4 shows the distribution of optimal capacites across samples generated via different subsampling schemes. The 36-year optima (the targets) are shown as dashed black lines. Plots compare capacities obtained at equal computational cost. Since determining a single set of capacities using *importance subsampling* requires two optimisation runs and for which the computational expense scales linearly in the number of timesteps, the associated plots use half the sample length. For example, in Figure 4(c), the box-and-whiskers plots for individual years, random sampling and k-medoids clustering correspond to single simulations using 8760 hourly timesteps, while the plot for *importance subsampling* uses two runs of half this length each. Individual years have only one possible sample length of 8760 hourly timesteps. Since k-medoids clustering subsamples representative days, k is the sample size divided by 24 (e.g. 480 timesteps correspond to 20 representative days).

Figure 4(a) shows the distribution of optimal capacities across simulations with computational expense equivalent to a PSM run of length 480 timesteps. While the variabilities across capacities determined via random sampling and *importance subampling* are roughly equal, the former induces a larger bias (understood as the difference between the median, indicated by the white line inside the box, and the 36-year optimum, indicated by the dashed black line). k-medoids clustering into 20 representative days leads to low variabilities but large biases, with underestimation of optimal baseload and peaking capacity of 5 and 8GW respectively and overestimation of optimal wind capacity by more than 10GW.

Figure 4(b) shows the same plots for computational expense equivalent to 1920 timesteps (just under 3 months of hours). *Importance subsampling* performs better than random sampling and clustering, which again underestimate (overestimate) optimal peaking (wind) capacity respectively.

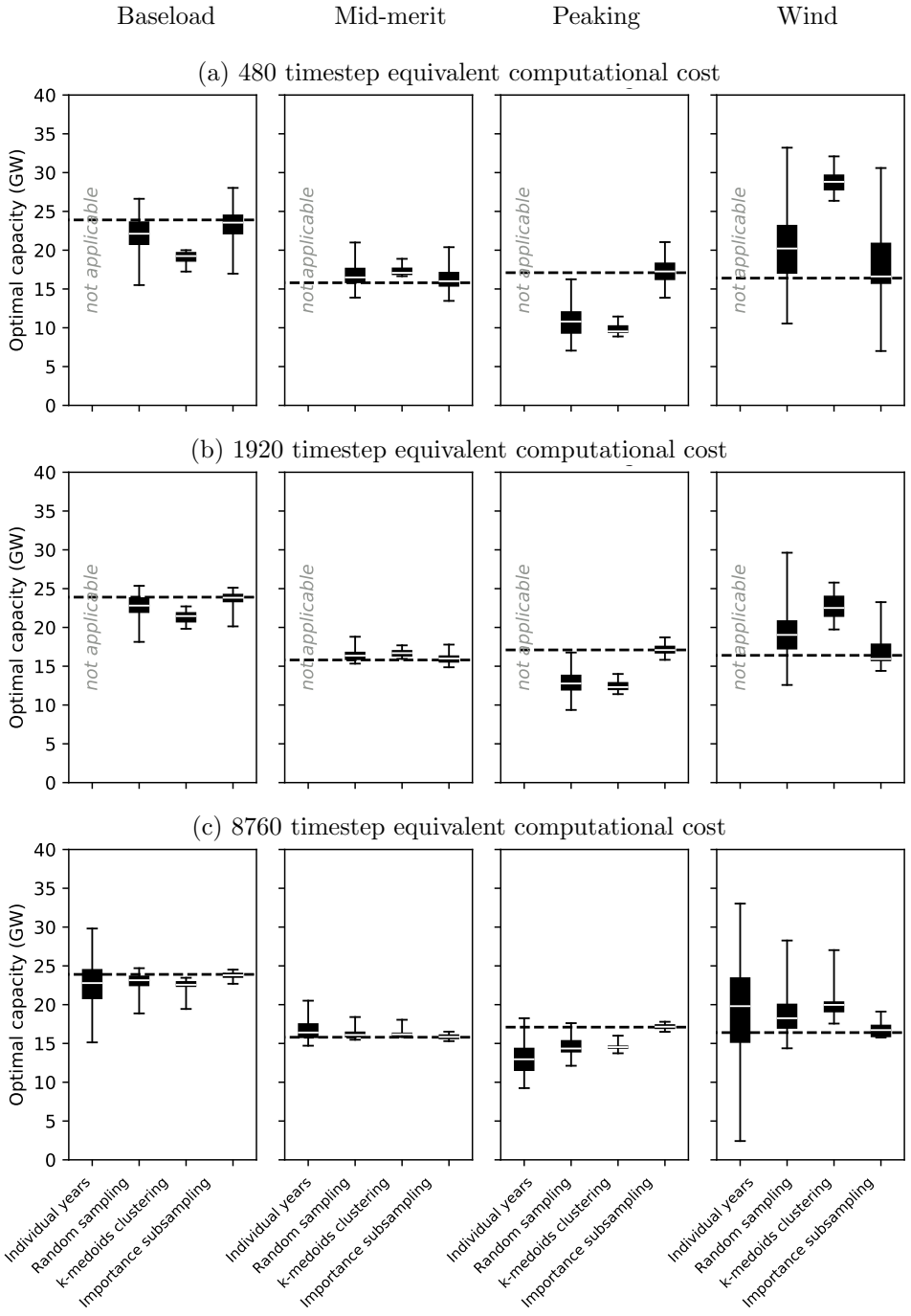


Figure 4: Distribution of optimal capacities for different subsampling methodologies. The box shows the 25th, 50th (median) and 75th percentiles, while the whiskers show the 2.5th and 97.5th. (a) corresponds to a computational cost equivalent to a single PSM run using 480 timesteps, while (b) and (c) correspond to 1920 and 8760 timesteps respectively. The dashed line indicates the optimal capacities across all 36 years of data: the best estimate of the "true" optima and the target under subsampling.

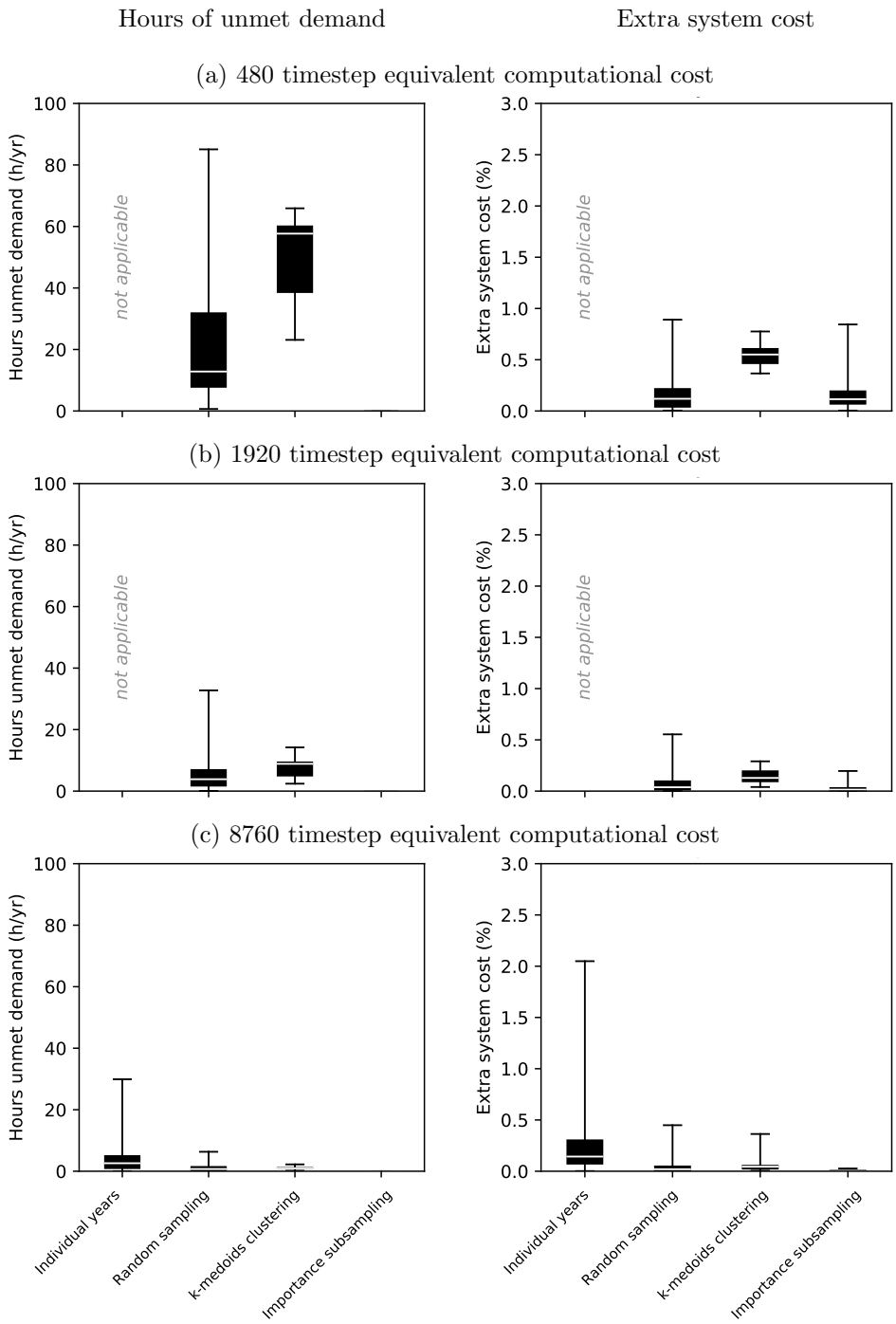


Figure 5: Distribution of *hours of unmet demand* and *extra system cost* for different subsampling methodologies. The box shows the 25th, 50th (median) and 75th percentiles, while the whiskers show the 2.5th and 97.5th. (a) corresponds to a computational cost equivalent to a single PSM run using 480 timesteps, while (b) and (c) correspond to 1920 and 8760 timesteps respectively and the target under subampling.

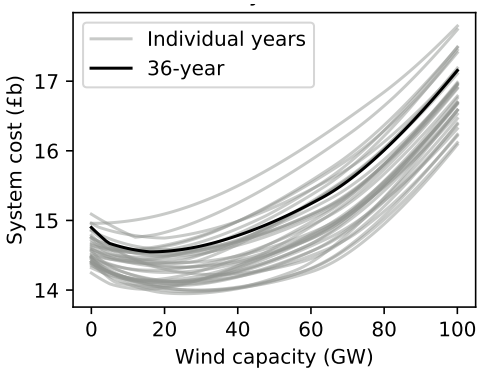


Figure 6: Optimal system cost as a function of prescribed wind capacity, for each of the 36 years and an annualised 36-year model run.

For simulations corresponding to full years of computational cost (8760 timesteps, Figure 4(c)), *importance subsampling* performs markedly better than the other schemes. A 95% prediction interval across individual-year simulations ranges from 2.5GW to 33.0GW, giving the user virtually no indication of optimal design. Results are slightly better under random sampling and k-medoids clustering. However, for all 3 schemes, the majority of samples underestimate optimal peaking capacity and overestimate optimal wind capacity. For *importance subsampling*, bias is low (the median of optima lies almost exactly on the target) and variation is greatly reduced.

Figure 5 shows the distribution of the number of hours of annual unmet demand. Power systems designed using individual years, random sampling and k-medoids clustering have a high probability of leading to supply capacity shortages. For example, one designed using k-medoids clustering into 20 days (480 timesteps) has a 75% change of unmet demand in at least 40 hours annually. Similarly, a power system designed using a random choice of individual year has a 35% chance of failing to meet demand in more than 3 hours. In contrast, *importance subsampling* leads to virtually no unmet demand, even for short samples lengths.

Figure 5 also shows the distribution of *extra system cost*. Extra cost for *importance subsampling* are similar to those for random sampling for 480 timesteps worth of computational expense but levels are lower than other samplers for longer simulation lengths. At computational cost equivalent to a 1920-timestep simulation, 95% of simulations lead to cost overruns of no more than 0.2%, and for 8760 timesteps any extra costs are negligible.

The percentage errors in optimal capacities for individual technologies are much larger than those in the total system cost. For example, the 2010 optimal wind capacity is 1.0GW, a 94% error when compared to the 36-year optimum of 16.4GW. However, the resultant power system (with extra peaking to ensure no unmet demand) costs just 2.3% more. Across all years, the mean absolute percentage error in installed wind capacity is 34%, but the resultant systems have a mean cost just 0.3% higher. The reason for this is twofold. Firstly, calculating the extra system cost assumes additional peaking (to meet unmet demand) is installed at the same "standard" price — there is no additional penalty for the unanticipated need for this additional capacity. A stronger penalisation for unmet demand (such as a loss-of-load cost, typically much higher than the costs used here) would greatly increase the additional expense. Secondly, there is a flat optimum with respect to installed wind capacity, shown Figure 6. This curve is constructed by imposing a fixed wind capacity of x and optimising the capacities of the remaining technologies in the usual way. The curve's y-value hence shows the minimum system cost as a function of prescribed wind capacity. There is one curve for each individual year along with an annualised

version using all 36 years at once. In most years, choosing a wind capacity anywhere between 0 and 40GW has a small effect on cost. This explains the large inter-year variability in optimal wind capacity; changes to the annual distribution of demand and wind change the shape of the cost curves slightly and may perturb the x-value of the minimum significantly with a comparatively modest effect on system cost.

5. Discussion

Previous studies indicate that, to reliably estimate optimal system design under natural climate variability, power system models (PSMs) with a high proportion of variable renewable energy require both a high temporal resolution and long samples of demand & weather data. These combined requirements typically exceed computational limitations, and timeseries reduction approaches such as subsampling are required. The accuracy of PSM outputs using subsampled data is highly dependent on whether a number of "extreme" timesteps are included in modelling samples. However, identifying them is difficult *a priori* since which timesteps are "extreme" frequently themselves depends on what the model is trying to determine (a chicken-and-egg situation).

This paper introduces a novel subsampling approach, called *importance subsampling*, that systematically identifies such timesteps and includes them in modelling samples. The main idea is to use a first stage simulation to obtain a rough indication of optimal system design. This is used, in conjunction with an *importance* function, to determine which timesteps must be sampled. A second stage model run, including these timesteps, provides estimates for the model outputs found using all available data.

A test case is performed on a model of the United Kingdom and 36 years of historic demand & weather data. Optimal system design for individual years is found to be highly dependent on the choice of year as well as underestimating (overestimating) optimal dispatchable (renewable) capacity respectively. Furthermore, resultant power systems lead to supply capacity shortages when applied to different weather years, and the long-term optimal power system design cannot be reliably determined by taking averages across multiple individual-year simulations.

Estimates of optimal system design using random samples or data clustered into representative days have lower variability and slightly smaller errors. Interestingly, randomly sampling 8760 hourly timesteps more accurately estimates optimal capacities than individual years. Hence, in models where timesteps may be scrambled (e.g. ones ignoring ramping or storage constraints), random sampling may be preferred over selecting long contiguous periods. However, errors in estimation of optimal system design and associated generation capacity shortages or cost overruns are present across both random sampling and clustering.

In contrast, *importance subsampling* consistently estimates, at reduced computational cost, the model outputs found using all available data. Even for short sample lengths, the bias between the median of capacities found using *importance subsamples* and the long-run optima is very small. This means a user can estimate optimal capacites by the median across multiple PSM runs with short *importance subsamples*, something that would be impossible with the other subsampling approaches due to their biases.

The *importance subsampling* approach is introduced in full generality and can be applied to a wide class of optimisation-based PSMs by modifying the choice of *importance* function. In the test case, this is a timestep's variable cost, and this choice generalises naturally to any model with such a notion. For example, in more detailed models, the variable cost can be summed across regions or calculated considering more sophisticated techno-economic factors. A judicious choice of *importance* function may also be inspired by expert knowledge. For example, if the user has a good idea of which

demand & weather scenarios are essential to ensure generation capacity adequacy, the *importance* function may be chosen to identify such events.

A drawback of the proposed approach is that it scrambles the order of timesteps and hence can only be directly applied to models with a small level of continuous dependence such as storage or ramping. Using the same approach to sample longer time periods such as days or weeks allows the modelling of such phenomena within periods, which may be sufficient in many settings. This may require additional refinements such as combining *importance subsampling* with clustering to reduce output variability.

The extra cost from designing suboptimal systems (and adjusting to ensure no unmet demand, see Section 4.3) is typically much smaller than the errors in estimates of optimal generation capacities. One of the causes, a flat optimal cost (particularly with respect to installed wind capacity), highlights an important distinction. If the only goal is the cheapest power system, a single-year simulation may suffice provided some extra capacity is installed to ensure demand can always be met. However, if one seeks to use minimisation of total system cost as a means of determining optimal system design, the impact of climate uncertainty is significant and approaches incorporating many years of climate information are required. In addition, one should be careful when invoking arguments of the type "wind should not be installed since the model indicates optimal design includes almost no wind". A more robust argument requires the investigation of the degree of suboptimality of power systems with perturbed design.

There are a number of possible extensions to this investigation. One is to relax the "perfect model" assumption that the 36-year simulation determines the "true" optimal capacities. In reality, the 36 years are themselves a sample from the "true" climate and the associated model outputs can be uncertain or biased. Using techniques from extreme value statistics or a climate model to generate a larger sample of hypothetical demand & weather data offer a more robust approach. Another extension is to conduct the same investigation using a different PSM or geographical region. On the one hand, Woollings (2010) notes that United Kingdom's climate exhibits a large degree of inter-year variability (so that sampling from multiple decades may not be necessary for other locations). On the other, the PSM employed in this investigation aggregates demand and wind levels into a single value for the whole country, "averaging out" spatial variations (and hence also some of the inter-year variability). For this reason, the degree of inter-year variability and both the need for and efficacy of *importance subsampling* for other PSMs should be explored. Finally, one could investigate its use in sampling longer representative periods (e.g. days or weeks). This may require additional adjustments such as combining the method with clustering approaches to further reduce variance in model outputs.

In line with the *open energy modelling* movement (see Hilpert et al.; Pfenninger et al., 2017) the data and all files required for the construction of the model (in the framework *Calliope*) are publicly available at `https://github.com/ahilbers/2019_importance_subsampling`.

6. Appendix

6.1. Power system model

The PSM is a simple representation of the UK power system, viewing it as a single node with inelastic UK-wide hourly demand levels and wind capacity factors. A linear optimiser determines the optimal capacities across 4 possible technologies by minimising the sum of installation and generation costs. The first 3, generically called *baseload*, *mid-merit* and *peaking*, differ only in their fixed (investment, £/GW) and variable (generation, £/GWh) costs. The fourth option, *wind*, has 0 generation cost but a supply level capped by the installed capacity times the wind capacity factor. Its

Sets		Timeseries data, timestep t	
$\mathcal{I}$	Technologies: baseload, mid-merit, peaking, wind	d_t	Demand (GWh)
$\mathcal{T}$	Timesteps (hourly)	w_t	Wind capacity factor
		λ_t	Timestep weight

Parameters		Decision variables	
c_i	Annualised installation cost, technology i (£m/GWyr)	cap_i	Installed capacity, technology i (GW)
f_i	Generation cost, technology i (£m/GWh)	gen_{it}	Electricity generated, technology i, timestep t (GWh)

Table 1: Nomenclature used to describe the PSM in Section 6.

Technology	Installation cost (£m/GWyr)	Generation cost (£m/GWh)
Baseload	$c_b = 300$	$f_b = 0.005$
Mid-merit	$c_m = 100$	$f_m = 0.035$
Peaking	$c_p = 50$	$f_p = 0.1$
Wind	$c_w = 100$	$f_w = 0$

Table 2: Assumed installation and generation costs for the 4 technologies used in the power system model employed in this paper.

installation cost refers to *rated capacity*: the maximum power output that a wind farm can produce at wind speeds just under the *cut-out point* when turbines are turned off to avoid damage. The model takes two timeseries inputs: hourly UK-wide demand levels and wind capacity factors.

6.1.1. Assumed costs

Table 2 shows each technology's assumed costs. They are chosen to be roughly indicative of the possible generation technologies available in the UK but should not be interpreted as realistic estimates of future build or generation cost. Each technology's installation cost is annualised by dividing by the expected plant lifetime.

Mid-merit prices are based on the CCGT H class reactor from (Department for Business, Energy and Industrial Strategy, 2016), which has a build cost of £100/kW per year of plant lifetime, corresponding to c_m = £100m/GWyr. The generation cost is determined assuming a wholesale gas price of 60 pence per therm. Using the conversion 1 therm $\approx$ 30kWh and assuming a generation efficiency of 55% gives a generation cost $f_m \approx$ £35t/GWh. The costs for baseload and peaking are chosen either side of these values.

Wind is assumed to have no generation cost. The construction cost is taken from the *medium* value for an *Onshore UK >5MW* wind farm in (Department for Business, Energy and Industrial Strategy, 2016): £1200/kW, corresponding to £50/kWyr over a 24-year lifetime. Adding fixed O&M costs of £23k/kWyr gives £73m/GWyr. This is revised upwards to £100m/GWyr to reflect infrastructure costs and the fact that offshore wind is more expensive than onshore.

6.1.2. Mathematical setup

The optimiser chooses capacities by minimising the sum of installation and generation costs while meeting an inelastic demand timeseries, phrased as a linear optimisation problem with decision variables $\{\text{cap}_i, \text{gen}_{it},\, i \in \mathcal{I}, t \in \mathcal{T}\}$. Timesteps are 1 hour in length.

$$\min_{\{\text{cap}_i, \text{gen}_{it},\, i\in\mathcal{I},\, t\in\mathcal{T}\}} z = \sum_{i\in\mathcal{I}} \left[c_i \text{cap}_i + 8760 \sum_{t=1}^{T} f_i \lambda_i \text{gen}_{it} \right] \tag{22}$$

subject to

$$\text{gen}_{it} \leq \text{cap}_i \quad \forall\, i \in \mathcal{I}\backslash\{w\},\, \forall\, t \in \mathcal{T} \tag{23}$$

$$\text{gen}_{wt} \leq \text{cap}_w w_t \quad \forall\, t \in \mathcal{T} \tag{24}$$

$$\sum_{i \in \mathcal{I}} \text{gen}_{it} = d_t \quad \forall\, t \in \mathcal{T} \tag{25}$$

$$\text{cap}_i, \text{gen}_{it} \geq 0 \quad \forall\, i \in \mathcal{I},\, \forall\, t \in \mathcal{T}. \tag{26}$$

For definitions of letters and symbols, see Table 1. The constant factor 8760 (= the number of hours in a year) is included since weights are chosen to sum to 1.

Constraint (23) ensures that conventional (baseload, mid-merit and peaking) generation levels never exceed their capacity. Constraint (24) ensures that wind generation never exceeds installed capacity times the wind capacity factor. Constraint (25) ensures that supply meets demand in each timestep. Given installed capacities, generation levels are determined by merit-order stacking of technologies in ascending order of variable cost.

The PSM is built and solved in the open-source modelling framework *Calliope* (see Pfenninger & Pickering, 2018).

6.2. Data

This investigation uses two timeseries: hourly UK-wide electricity demand levels $(d_t)_{t\in\mathcal{T}}$ and wind capacity factors $(w_t)_{t\in\mathcal{T}}$ over the period 1980-2015. They are modified versions of those found in (Bloomfield et al., 2016)

6.2.1. Demand model

The demand timeseries is based on a regression between weather and demand data collected from two different sources. UK-wide daily mean temperature is obtained from the MERRA reanalysis (Rienecker et al., 2017) for the period 1980-2015. Metered UK-wide demand data is obtained from National Grid, the UK transmission operator, over 2006-2015 (see National Grid, 2018). For this period, in which there is overlapping meteorological and demand data, a regression model is run for daily demand:

$$\begin{aligned}\text{Demand}(t_d) = \alpha_1 + \alpha_2 t_d + \alpha_3 \sin(\omega t_d) + \alpha_4 \cos(\omega t_d) + \alpha_5 \text{Te}(t_d) + \alpha_6 \text{Te}^2(t_d) \\ + \sum_{k=7}^{13} \alpha_k \text{WD}_k(t_d) + \alpha_{14} \text{HOL}(t_d) + \epsilon(t_d)\end{aligned} \tag{27}$$

where

- $\text{Demand}(t_d)$ is the totaly daily demand in day t_d.
- $\alpha_1 + \alpha_2 t_d$ accounts for anthropogenic demand trends (e.g. economic growth or efficiency improvements).
- ω is year^{-1}.
- $\sin(\omega t_d)$ and $\cos(\omega t_d)$ account for year-long demand cycles.
- $\text{Te}(t_d)$ is the *effective temperature*: a slightly time-offset daily mean temperature to account for the fact that demand is influenced by temperature at a time lag. It is defined recursively via $\text{Te}(t_d) = \frac{1}{2}\text{Te}(t_d - 1) + \frac{1}{2}\text{T}(t_d)$, where $T(t_d)$ is the UK-wide daily average temperature on day t_d, and initialised via $\text{Te}(t_d =$ 1-1-2006$) = \text{T}(t_d =$ 1-1-2006$)$.
- $\text{WD}_7(t_d)$ and $\text{WD}_{13}(t_d)$ are indicator variables that take value 1 only on Monday through Sunday respectively, but not on bank holidays.

- HOL(t_d) is an indicator variable that takes value 1 only on bank holidays.
- $\epsilon(t_d)$ is an error term.

For values of the parameters, see (Bloomfield et al., 2016). The distribution of the error term $\epsilon(t_d)$ is symmetrical with a sharp mode at 0 and a standard deviation of 34.2GWh. There is no trend in average daily error across years or months.

After determining the coefficients, daily demand is extrapolated to 1980-2015 by truncating $\epsilon(t_d)$ and using the daily temperatures to extend the timeseries back. Four diurnal curves, one for each of the 3-month periods DJF, MAM, JJA, SON, are used to upsample the timeseries to hourly resolution, with each day given a weighting of two diurnal curves. For example, December 1 is given a 50/50 weighting of the diurnal curves for SON and DJF.

Anthropogenic demand trends are removed by setting α_2=0 and replacing α_1 by $\hat{\alpha}_1 = \alpha_1 + \alpha_2 t|_{t=1-1-2011}$, representing the middle of the period 2006-2015. This ensures that the differences between the demand distribution in different years is weather-driven.

6.2.2. Wind model

The wind model is identical to that used in the *HIGH wind* scenario in (Bloomfield et al., 2016), which is in turn based on two other studies. First, from the MERRA re-analysis (Rienecker et al., 2017), hourly windspeeds at 2, 10 and 50m elevation are obtained at gridpoints throughout the UK. These are fed through a logarithmic profile to estimate windspeeds at 80m, the average height of UK wind turbines.

Windspeeds at MERRA gridpoints are linearly interpolated to the locations of all wind farms in operation or in the construction pipeline as of April 2014 (see Drew et al., 2015). At each location, wind speeds are transformed to capacity factors using an adjusted power curve from a Siemens 2.3MW turbine (see Cannon et al., 2015, for details). An average over the wind farm configuration, weighted by each farm's capacity, determines the UK-wide capacity factor. In the PSM, adding capacity happens in proportion to the wind farms' relative sizes. For example, doubling wind capacity implicitly means doubling the capacity of each wind farm.

6.3. Supplementary material

The demand timeseries used in the test case (Section 4) as introduced in Section 6.2.1 is based on a regression. In particular, in obtaining 36 years of demand data, the error term in equation (27) is truncated. This has the effect of reducing variability. For this reason, the performance of *importance subsampling* is also examined on a real demand timeseries, where only a long-term demand trend is removed but there is no truncation of the error. Figures 4 and 5 are recreated using the altered demand timeseries. Results are broadly identical as those found using regression data and are available as supplementary material at `github.com/ahilbers/2019_importance_subsampling`.

Acknowledgements

Adriaan Hilbers was funded through the EPSRC CDT in Mathematics for Planet Earth. David Brayshaw was partially supported by the PRIMAVERA project, funded by the European Union's Horizon 2020 programme, Grant Agreement no. 641727. The authors thank Hannah Bloomfield for providing demand and wind timeseries and 3 anonymous referees for their constructive comments.

References

Bazmi, A., & Zahedi, G. (2011). Sustainable energy systems: Role of optimization modeling techniques in power generation and supply — a review. *Renewable and Sustainable Energy Reviews*, *15*, 3480–3500. doi:10.1016/j.rser.2011.05.003.

Bloomfield, H., Brayshaw, D., Shaffrey, L., Coker, P., & Thornton, H. (2016). Quantifying the increasing sensitivity of power systems to climate variability. *Environmental Research Letters*, *11*, 124025. doi:10.1088/1748-9326/11/12/124025.

Cannon, D. J., Brayshaw, D. J., Methven, J., Coker, P. J., & Lenaghan, D. (2015). Using reanalysis data to quantify extreme wind power generation statistics: A 33 year case study in Great Britain. *Renewable Energy*, *75*, 767 – 778. doi:10.1016/j.renene.2014.10.024.

Collins, S., Deane, J. P., Poncelet, K., Panos, E., Pietzcker, R. C., Delarue, E., & Ó Gallachóir, B. (2017). Integrating short term variations of the power system into integrated energy system models: A methodological review. *Renewable and Sustainable Energy Reviews*, *76*, 839 – 856. doi:10.1016/j.rser.2017.03.090.

Collins, S., Deane, P., Ó Gallachóir, B., Pfenninger, S., & Staffell, I. (2018). Impacts of inter-annual wind and solar variations on the European power system. *Joule*, *2*, 2076 – 2090. doi:10.1016/j.joule.2018.06.020.

De Jonghe, C., Delarue, E., Belmans, R., & D'haeseleer, W. (2011). Determining optimal electricity technology mix with high level of wind power penetration. *Applied Energy*, *88*, 2231 – 2238. doi:10.1016/j.apenergy.2010.12.046.

de Sisternes, F. J., & Webster, M. D. (2013). Optimal selection of sample weeks for approximating the net load in generation planning problems. *ESD Working Papers*, (pp. 1–12).

Department for Business, Energy and Industrial Strategy (2016). *Electricity Generation costs*. Technical report. Department for Business, Energy and Industrial Strategy.

Drew, D. R., Cannon, D. J., Brayshaw, D. J., Barlow, J. F., & Coker, P. J. (2015). The impact of future offshore wind farms on wind power generation in Great Britain. *Resources*, *4*, 155–171. doi:10.3390/resources4010155.

Hall, L., & Buckley, A. (2016). A review of energy systems models in the UK: Prevalent usage and categorisation. *Applied Energy*, *169*, 607–628. doi:10.1016/j.apenergy.2016.02.044.

Härtel, P., Kristiansen, M., & Korpås, M. (2017). Assessing the impact of sampling and clustering techniques on offshore grid expansion planning. *Energy Procedia*, *137*, 152–161. doi:10.1016/j.egypro.2017.10.342.

Hilpert, S., Kaldemeyer, C., Krein, U., Günther, S., Wingenbach, C., & Pleßmann, G. (). The open energy modelling framework (OEMOF) — a novel approach in energy systems modelling. *Preprint.*, . doi:10.20944/preprints201706.0093.v1.

IPCC (2018). *Global warming of 1.5C: Summary for Policymakers.*. Policy report. World Meteorological Organization, Geneva, Switzerland.

Jacobson, M. Z., Delucchi, M. A., Cameron, M. A., & Frew, B. A. (2015). Low-cost solution to the grid reliability problem with 100% penetration of intermittent wind, water, and solar for all purposes. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 1506015065. doi:10.1073/pnas.1510028112.

Kotzur, L., Markewitz, P., Robinius, M., & Stolten, D. (2018). Impact of different time series aggregation methods on optimal energy system design. *Renewable Energy*, *117*, 474–487. doi:10.1016/j.renene.2017.10.017.

Nahmmacher, P., Schmid, E., Hirth, L., & Knopf, B. (2016). Carpe diem: a novel approach to select representative days for long-term power system modeling. *Energy*, *112*, 430–442. doi:10.1016/j.energy.2016.06.081.

National Grid (2017). *Future Energy Scenarios 2017*. Technical report. National Grid.

National Grid (2018). *Data Explorer*. Dataset. National Grid. https://www.nationalgrid.com/uk/electricity/market-operations-and-data/data-explorer.

Pfenninger, S. (2017). Dealing with multiple decades of hourly wind and PV time series in energy models: a comparison of methods to reduce time resolution and the planning implications of inter-annual variability. *Applied Energy*, *197*, 1–13. doi:10.1016/j.apenergy.2017.03.051.

Pfenninger, S., DeCarolis, J., Hirth, L., Quoilin, S., & Staffell, I. (2017). The importance of open data and software: Is energy research lagging behind? *Energy Policy*, *101*, 211–215. doi:10.1016/j.enpol.2016.11.046.

Pfenninger, S., & Pickering, B. (2018). Calliope: a multi-scale energy systems modelling framework. *Journal of Open Source Software*, *3(29)*, 825. doi:10.21105/joss.00825.

Poncelet, K., Delarue, E., Six, D., Duerinck, J., & D'haeseleer, W. (2016). Impact of the level of temporal and operational detail in energy-system planning models. *Applied Energy*, *162*, 631 – 643. doi:10.1016/j.apenergy.2015.10.100.

Poncelet, K., Höschle, H., Delarue, E., Virag, A., & D'haeseleer, W. (2017). Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning problems. *IEEE Transactions on Power Systems*, *32*, 1936–1948. doi:10.1109/TPWRS.2016.2596803.

Rienecker, M. et al. (2017). MERRA — NASA's modern-era retrospective analysis for research and applications. *Journal of Climate*, *24 (14)*, 3624–3648. doi:10.1175/JCLI-D-11-00015.1.

Ringkjøb, H.-K., Haugan, P. M., & Solbrekke, I. M. (2018). A review of modelling tools for energy and electricity systems with large shares of variable renewables. *Renewable and Sustainable Energy Reviews*, *96*, 440 – 459. doi:10.1016/j.rser.2018.08.002.

Staffell, I. (2017). Measuring the progress and impacts of decarbonising British electricity. *Energy Policy*, *102*, 463–475. doi:10.1016/j.enpol.2016.12.037.

Staffell, I., & Pfenninger, S. (2018). The increasing impact of weather on electricity supply and demand. *Energy*, *145*, 65–78. doi:10.1016/j.energy.2017.12.051.

Stoft, S. (2002). *Power System Economics*. Wiley-IEEE Press.

Thornton, H., Scaife, A., Hoskins, B., & Brayshaw, D. (2017). The relationship between wind power, electricity demand and winter weather patterns in Great Britain. *Environmental Research Letters*, *12*, 064017. doi:10.1088/1748-9326/aa69c6.

Trutnevyte, E. (2016). Does cost optimization approximate the real-world energy transition? *Energy*, *106*, 182–193. doi:10.1016/j.energy.2016.03.038.

Ventosa, M., Baíllo, Á., Ramos, A., & Rivier, M. (2005). Electricity market modeling trends. *Energy Policy*, *33-7*, 897 – 913. doi:10.1016/j.enpol.2003.10.013.

Wohland, J., Reyers, M., Märker, C., & Witthaut, D. (2018). Natural wind variability triggered drop in German redispatch volume and costs from 2015 to 2016. *PLoS ONE*, *13*, 1. doi:10.1371/journal.pone.0190707.

Woollings, T. (2010). Dynamical influences on European climate: an uncertain future. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *368*, 3733 – 3756. doi:10.1098/rsta.2010.0040.

Zeyringer, M., Price, J., Fais, B., Li, P.-H., & Sharp, E. (2018). Designing low-carbon power systems for Great Britain in 2050 that are robust to the spatiotemporal and inter-annual variability of weather. *Nature Energy*, *3*, 395–403. doi:10.1038/s41560-018-0128-x.