Supplementary Material

Adriaan P. Hilbers, David J. Brayshaw, Axel Gandy

1 Description

This document provides supplementary material for the paper *Importance subsampling: improving power system planning under climate-based uncertainty* (2019). It shows the results of applying the *importance subsampling* methodology to a second demand and wind timeseries, based on metered data instead of a regression.

In the paper, the proposed methodology is applied to 36 years worth of hourly UK-wide demand levels and wind capacity factors. The demand timeseries is based on a regression for which the error term is truncated, and the wind timeseries based on re-analysis wind speed data. Both such approaches typically produce timeseries with an unrealistically low level of variability. Hence, a reasonable question is whether the *importance subsampling* methodology also reliably estimates model outputs when considering timeseries that are less "smooth".

This supplementary material reproduces Figures 4 and 5 with timeseries inputs based on metered data instead of regressions or re-analysis. Metered UK-wide demand data is obtained over 2008-2015 from the *National Grid Data Explorer* (`nationalgrideso.com/balancing-data/data-explorer`). Hourly UK-wide demand levels and wind capacity factors across the period 2008-2015 are then obtained as follows:

- Demand is the value in column `ND` at each whole hour. Long-term anthropogenic demand trends are removed using a linear detrending function with the same slope as that used in the paper
- Wind capacity factor is the value in column `EMBEDDED WIND GENERATION` divided by the value of `EMBEDDED WIND CAPACITY`. Any timesteps for which this value exceeds 1 are replaced by 1.

The purpose of this exercise is purely to obtain a timeseries with a realistic degree of variability. No additional pre-processing is done. The timeseries values themselves are not realisticl. For example, the fact that some timesteps have a wind generation higher than wind capacity indicates individual timestep values are inaccurate. However, these concerns are ignored, since the only goal is to demonstrate that *importance subsampling* reliably estimates the model outputs found using the full timeseries.

2 Results

Plots of the optimal capacities, *hours of unmet demand* and *extra system cost* exactly as in Figures 4 and 5 of the paper are shown below. The results are broadly the same as in the paper, with *importance subsampling* providing unbiased estimates of optimal capcities with low levels of extra system cost and virtually no unmet demand.

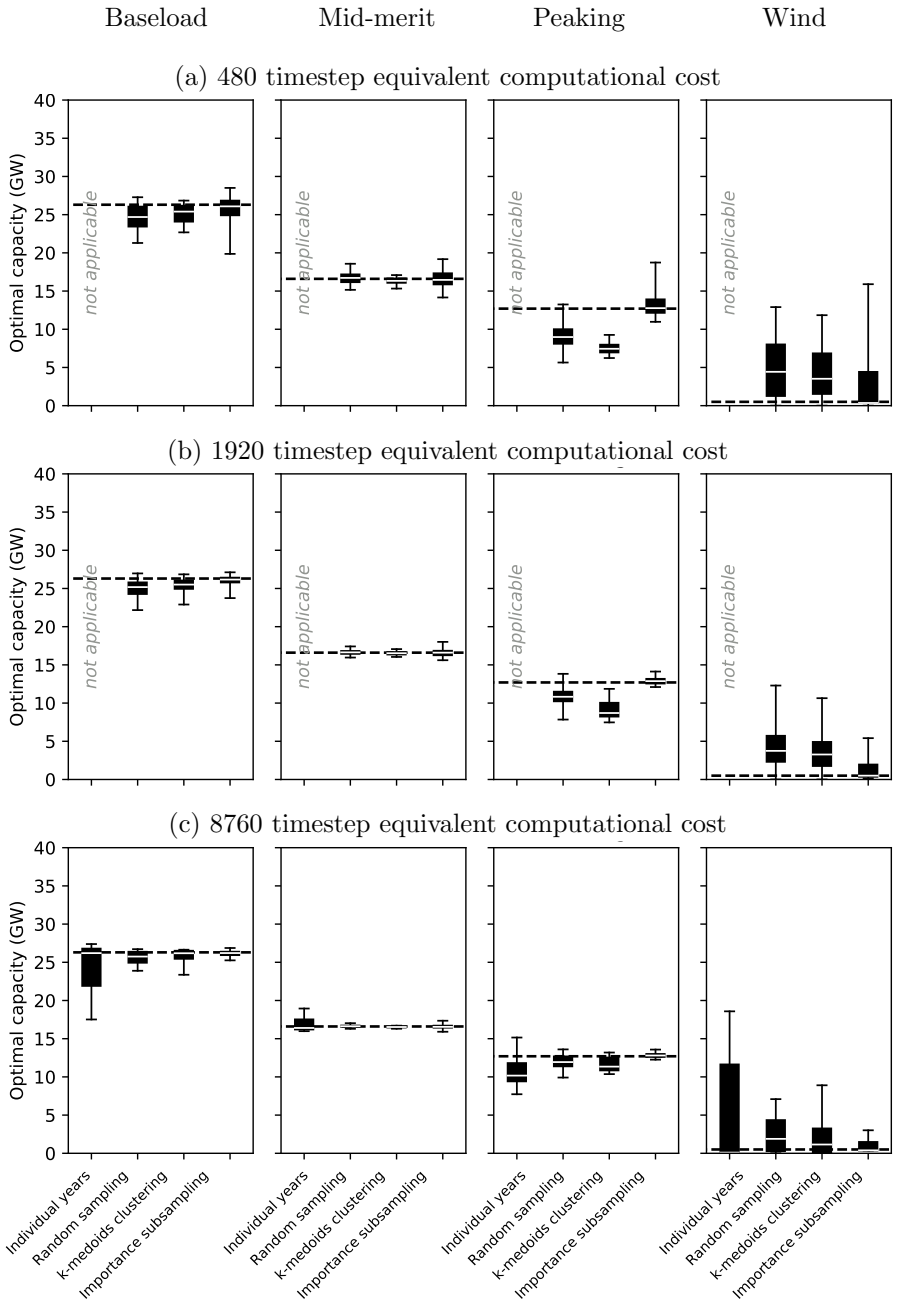

Figure 1: Distribution of optimal capacities for different subsampling methodologies. The box shows the 25th, 50th (median) and 75th percentiles, while the whiskers show the 2.5th and 97.5th. The dashed line indicates the optimal capacities across all 8 years of data: the best estimate of the "true" optima and the target under subsampling.

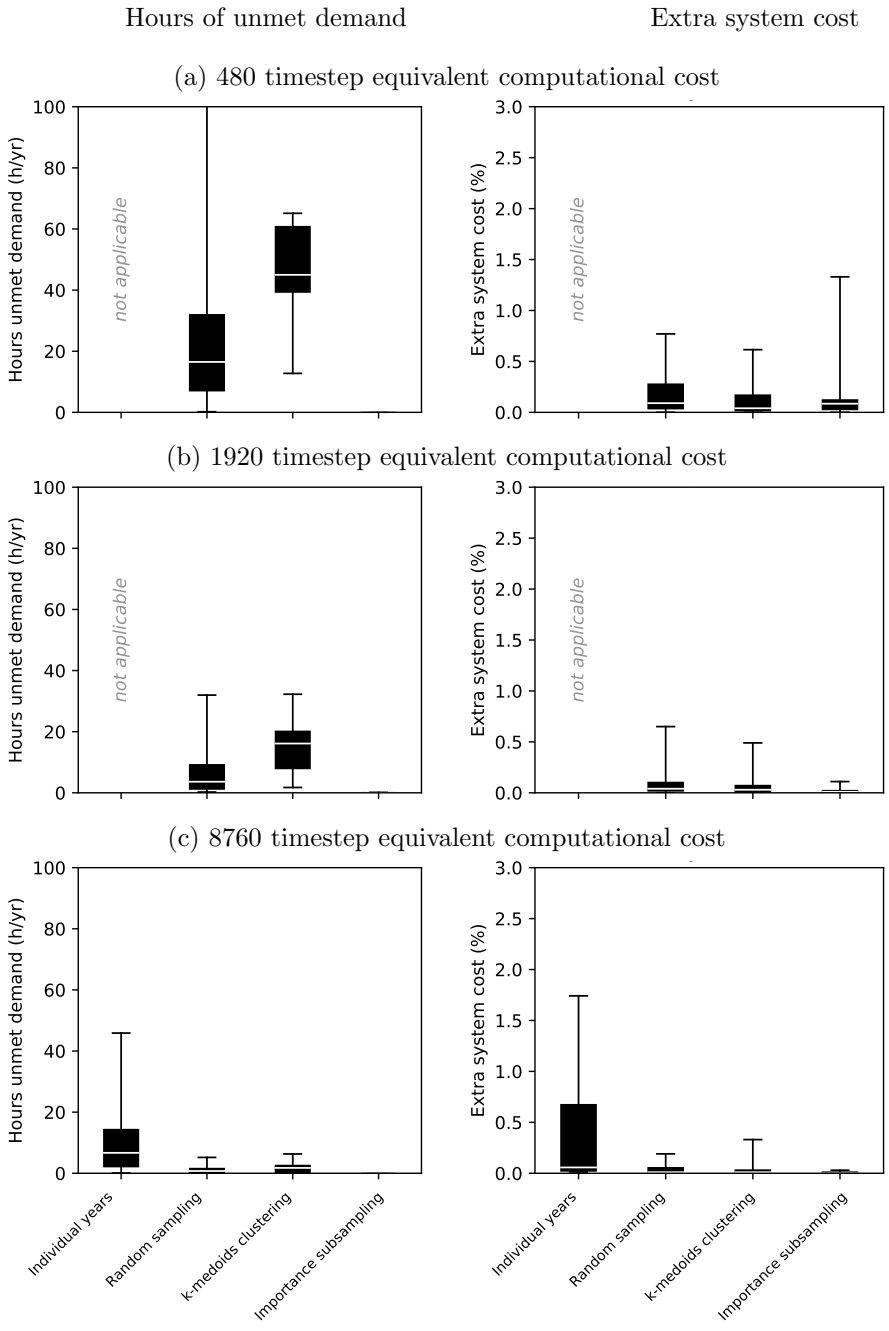

Figure 2: Distribution of *hours of unmet demand* and *extra system cost* for different subsampling methodologies. The box shows the 25th, 50th (median) and 75th percentiles, while the whiskers show the 2.5th and 97.5th.