Empirical Results Suggest Quasi-Monte Carlo Sampling Increases Accuracy in the Estimation of Annual Energy Production from Operational Data

Jordan Perr-Sauer, Nicola Bodini, Stephen Becker*, Eric Simley, Rob Hammond, Jason Fields

* Stephen Becker is affiliated with the University of Colorado, Boulder.

Introduction:

Annual Energy Production (AEP) is an important performance metric used in the wind energy industry, quantifying how much energy a wind power plant is expected to produce in a year. In previous work, NREL researchers identified several choices made during analysis that can have an impact on this metric (Fields, 2021). OpenOA (Perr-Sauer, 2021) is open-source software which contains an implementation of AEP using a Monte Carlo (MC) method (Bodini, 2020). One drawback of using MC that it requires many iterations to converge to a solution. If the computational cost of each iteration is high, the overall computational cost will also be high. In this work, we augment OpenOA's AEP analysis to use Quasi-Monte Carlo (QMC) sampling to reduce the overall computational burden of AEP estimation. This augmented method is evaluated using open data from Engie Renewables La Haute Borne 4-turbine wind power plant, which contains two years of operational data. Our experiment shows that using QMC provides an order-of-magnitude performance improvement, in a way that is consistent with the theory.

Key Point: **Randomized Quasi-Monte Carlo (RQMC) shows an improved convergence rate when compared to standard Monte Carlo for the estimation of AEP in one dataset.**

Randomized Quasi-Monte Carlo (RQMC):

This section provides some of the foundational equations underpinning QMC theory, using the treatment of (Owen, 2019). We start with the definition of the discrepancy of a sequence. It has been shown by the Koksma-Hlawka inequality that the convergence rate of MC integration is bounded above by a term that depends only on the discrepancy of the sequence and the variation (measure of smoothness) of the function. Notably, the dimensionality of the function does not factor in to this bound, suggesting that QMC methods are most useful in mid-to-high dimension situations. Sobol sequences are a common example of a low-discrepancy sequence. In this work, we use the implementation of Sobol sequences from SciPy's scipy.stats.qmc module (Virtanen, 2020). An example sequence is shown in Figure 1. QMC is a deterministic method which does not produce an uncertainty of the estimate. Therefore, RQMC methods have been developed to provide this uncertainty. An illustration of RQMC is shown in Figure 2.

Discrepancy of a Sequence:

Star Discrepancy

$$D_n^* = D_n^*(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \sup_{\boldsymbol{a} \in [0,1)^d} \underbrace{|\delta(\boldsymbol{a}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)|}$$

Local Discrepancy: $$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\boldsymbol{x}_i \in [\boldsymbol{0}, \boldsymbol{a})} - \prod_{j=1}^{d} a_j$$

Koksma-Hlawka Inequality:

$$\left|\frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{x}_i) - \int_{[0,1]^d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right| \leqslant D_n^*(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) V_{\mathrm{HK}}(f)$$

(15.10, Owen)

Convergence Rates:

Independent Random Sequence (standard Monte Carlo):

$$D_n^* \overset{n \to \infty}{\leq} \frac{\sqrt{\log\log(n)}}{\sqrt{2n}} = O(n^{-1/2})$$ (p11, Owen)

Low-Discrepancy Sequence (e.g., Sobol sequence):

$$D_n^* = O\left(\frac{(\log n)^d}{n}\right) = O(n^{-1+\epsilon}) \text{ for any } \epsilon > 0$$

(p11, Owen)

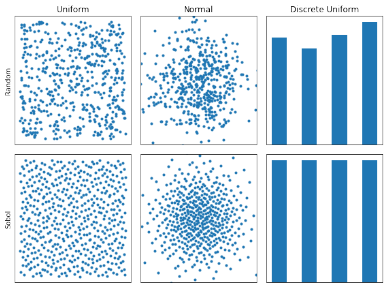

Figure 1: Example of Sobol sampling, which was used to implement RQMC for this work. This sampling has low discrepancy, unlike random sampling, which can be seen here as "clumping" in the sample space. When used to integrate a smooth function, samples with less discrepancy tend to converge more quickly.

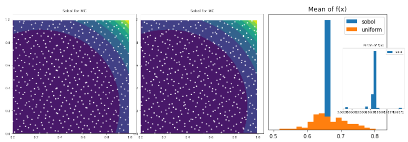

Figure 2: To obtain confidence intervals on QMC estimators, we can vary the initialization of the low-discrepancy sequence generator to obtain multiple sequences. In this figure, you can see two such sequences over the same function, with the mean estimated by RQMC reported in the histogram.

References:

N. Bodini and M. Optis, "Operational-based annual energy production uncertainty: are its components actually uncorrelated?," Wind Energ. Sci., vol. 5, no. 4, pp. 1435–1448, Oct. 2020, doi: 10.5194/wes-5-1435-2020.

M. Fields, M. Optis, J. Perr-Sauer, A. Todd, J. Lee, J. Meissner, E. Simley, N. Bodini, L. Williams, S. Sheng, and R. Hammond. Wind Plant Performance Prediction Benchmark Phase 1 Technical Report. Technical Report NREL/TP-5000-78715, 1826665, MainId:32632, Oct. 2021.

B. Owen, "Monte Carlo Book: the Quasi-Monte Carlo parts," Unpublished, 2019. Accessed: Nov. 29, 2021. [Online]. Available: https://artowen.su.domains/mc/qmcstuff.pdf

J. Perr-Sauer et al., "OpenOA : An Open-Source Codebase For Operational Analysis of Wind Farms," vol. 6, pp. 6–8, 2021, doi: 10.21105/joss.02171.

P. Virtanen et al., and SciPy 1.0 Contributors. (2020) SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods, 17(3), 261-272.

Engie La Haute Born 4 Turbine Dataset. Retrieved from: https://opendata-renewables.engie.com

Experimental Setup:

We augment the AEP analysis method from OpenOA (Figure 3) by including Sobol sampling for the parameters drawn from the input distributions in Table 1. We repeat the analysis several times over sweeps of hyperparameters in two experiments using two years of operational data from Engie Renewables La Haute Borne 4-turbine wind power plant. The "Accuracy" experiment sweeps over two different numbers of MC iterations, performing 30 repetitions each. Its purpose is to observe the distribution of AEP between QMC and MC methods and quantify if there is a bias. The "Scaling" experiment performs only 5 repetitions, but sweeping across seven different numbers of iterations to quantify the asymptotic convergence rate with respect to N.

Parameter	Distribution	Parameters
Meter	Normal	$\mu = 1, \sigma = 0.005$
Loss	Normal	$\mu = 1, \sigma = 0.05$
Windiness	Discrete Uniform	{10,11…, 21}
Loss Threshold	Discrete Uniform	{10,11,…, 21}
Reanalysis Product	Discrete Uniform	{Merra2, ERA5}

Table 1: Input distributions for those parameters sampled in both experiments. Meter is the wind plant energy measurement uncertainty. Windiness is number of years of historical wind data for long-term correction.

Experiment	N (Iterations)	Repetitions	QMC
Scaling	$\{2^2, 2^3 \ldots, 2^8\}$	5	{Yes, No}
Accuracy	$\{2^7, 2^8\}$	30	{Yes, No}

Table 2: Hyperparameters swept in two experiments. The "Accuracy" experiment has more repetitions and can be used to assess bias, while the "Scaling" experiment sweeps over a larger number of iterations.

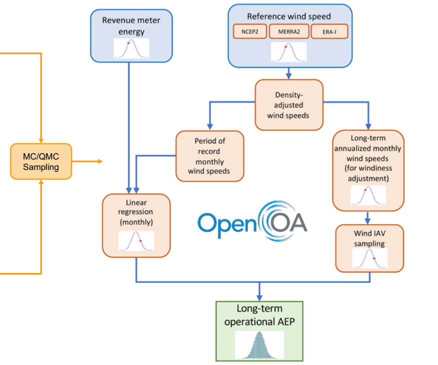

Figure 3: Flowchart of the AEP analysis, reproduced from Bodini and Optis, 2020. In our experiment, sampling of inter-annual variability (IAV) is disabled.

Results:

Our results suggest that QMC sampling produces an order of magnitude improvement in the efficiency of AEP estimation. Figure 4 shows histograms of AEP mean and standard deviation in the "Accuracy" experiment. Observe how the estimates produced by QMC (orange) are both more accurate and unbiased compared to MC (blue), while taking similar compute time. Figure 5 shows the convergence rate for the AEP mean and standard deviation estimates with respect to the number of iterations. This figure is plotted on a log-log scale, and indicates orders of magnitude improvement when using QMC sampling.

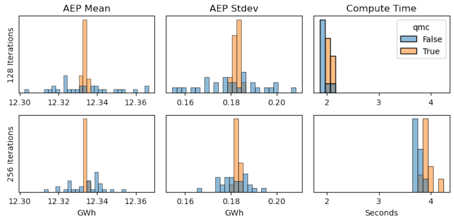

Figure 4: Histograms showing the AEP mean, standard deviation, and compute time, for 128 iterations (top) and 256 iterations (bottom). Results from Sobol sampling in the RQMC method are in orange, whereas random sampling in the MC are in blue. The compute time per iteration is similar between methods.

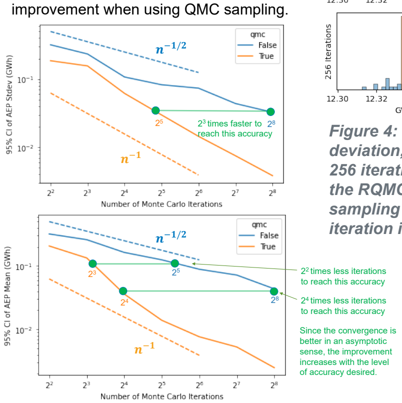

Figure 5: Convergence plots (log-log scale) showing RQMC method converging at an asymptotically faster rate than the random method. Dashed lines are plotted to compare against the convergence rates predicted by theory.

Implementation: OpenOA

OpenOA is an open-source software package, supported by NREL, which implements the AEP Analysis method used in this work (Perr-Sauer 2021). We augment the MonteCarloAEP class with a Boolean parameter "qmc" to enable the Quasi-Monte Carlo method using Sobol sequences as implemented in SciPy (Virtanen, 2020).

```
operational_analysis.analysis.MonteCarloAEP(qmc=True)
```

We plan to merge this in to OpenOA's main fork after the upcoming OpenOA V3 release. For now, the version used in this work is available through the following fork:

https://github.com/jordanperr/OpenOA/tree/mc_explore

This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Wind Energy Technologies Office. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

NAWEA Wind Tech 2022
Newark, Delaware
9/20 – 9/22/2022
NREL/PO-2C00-83810