A Realistic Dataset Generator for Smart Grid Ecosystems with Electric Vehicles

Georgios Charalambidis[1], Charilaos Akasiadis[2], Emmanouil S. Rigas[3] and Georgios Chalkiadakis[1]

1. School of Electrical & Computer Engineering, Technical University of Crete, Greece
2. Institute of Informatics & Telecommunications, NCSR "Demokritos", Greece
3. School of Medicine, Aristotle University of Thessaloniki, Greece

GitHub repository: `https://github.com/tuc-intelligence-energy/smartgrid-dataset-generator`

ACM e-Energy 2022

Motivation

• The Electric Vehicles market is rapidly expanding – different geographical regions have different characteristics.
• We need to test and verify methods for optimization, scheduling, pricing, etc. using real data.
• Privacy issues hinder researchers' access to data.
• Dataset generators must be realistic and privacy preserving.
Our solution:
◦ We incorporate several anonymized datasets.
◦ We put forward generation mechanisms that fit to the existing data and produce new, that is different in absolute values, but adhere to the statistics and principles of the original.
◦ Four methods to compare and select based on user needs.
◦ Additional smoothing and summarization and visualization capabilities.
◦ Data for microgrid level generation, consumption, and EV usage.
◦ Configurable constraints for consistency control

Input Data

Data related to EV comes from many different publicly available sources:

Characteristics of EVs and Chargers:
"Mendeley Data platform", "Spirit Energy", "Tesla"

Drivers Behavior:
"My Electric Avenue"

Aggregate generation levels per production type, as well as aggregate demand for all European regions for the recent years:
"ENTSOE Transparency Platform"

Preprocessing

As the original data has different formats, we need to apply specific transformations, and then combine it properly in order to end up with the final input files. As a first step of the generation process, we preprocess the acquired data files to assign them appropriate timestamps consistent across all generated files, i.e., *Date, Year, MonthOfYear, DayOfWeek, TimeOfDay*. Regarding the data related to the trip events, the file has the same format as that of the charging events: each row represents a trip event/action and each column, the exact time of a trip event, the duration of the trip, the trip distance and the power consumed during each trip respectively.

For the other data category, energy production and total consumption, we convert the timestamp to the appropriate format and combine measurements into a common file. A new column is inserted, that holds the imbalance between the aggregates of consumption and production.

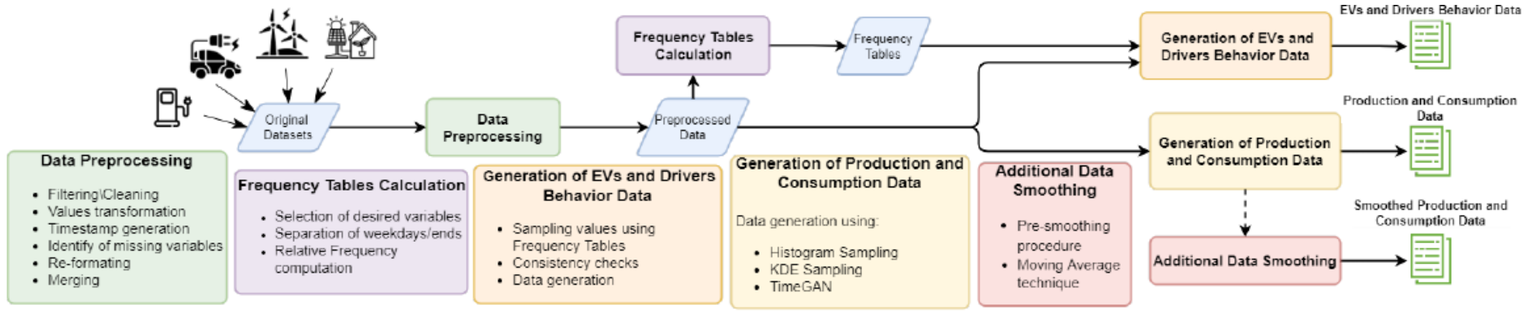
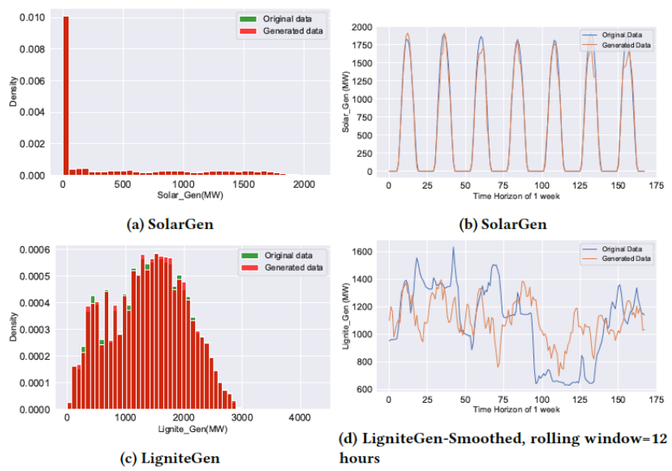


Figure 1: Schematic overview of workflow components in our system

Data Generation Techniques

Our dataset generation process uses different statistical methods to produce synthetic data. There are two desired types of data to generate: ◦ Energy Consumption and Production, ◦ EVs and Drivers Behavior data. We adopt three different generation methods that can be used interchangeably:
◦ *Histogram Sampling* , *KDE Sampling* , *TimeGAN*
Since the drivers' behavior data do not come in a time series format, the application of these methods is not meaningful. Therefore, we also put forward a *Frequency Tables Method*, which requires no such assumption for the input.

Results



(a) SolarGen (b) SolarGen (c) LigniteGen (d) LigniteGen-Smoothed, rolling window=12 hours

Figure 2: Histograms and Time-Series of Production and Consumption Data (HIST method)

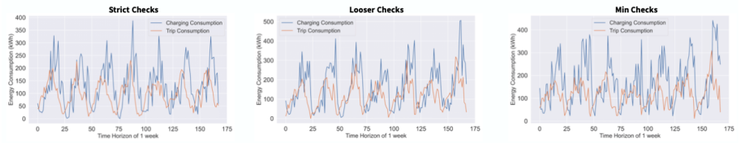


Figure 3: **EV Consumption**

K-S statistic of Production & Consumption Data

Output Data time-horizon	*Method*	*LigniteGen*	*FossilGasGen*	*SolarGen*	*WindGen*	*TotalLoad*	*Average*
4 years of data	HIST	0.005	0.005	**0.03**	**0.004**	0.005	**0.01**
	KDE	**0.004**	**0.004**	0.036	**0.004**	**0.004**	**0.01**
	TimeGAN	0.08	0.038	0.459	0.06	0.035	0.134
1 year of data	HIST	0.011	0.011	**0.031**	**0.009**	0.009	**0.014**
	KDE	**0.001**	**0.006**	0.038	0.01	**0.007**	**0.014**
	TimeGAN	0.052	0.072	0.464	0.068	0.044	0.14

Difference of EVs & Drivers Behavior Data Frequency Tables

Method	*DailyCharges*	*ChargeStartHour*	*StartingSoC*	*EndingSoC*	*DailyTrips*	*TripStartHour*	*Average*
StrictChecks	0.419	**0.057**	0.079	0.231	0.821	**0.345**	0.325
LooserChecks	0.095	0.232	0.077	0.227	0.619	0.465	0.286
MinChecks	**0.072**	0.173	**0.067**	**0.221**	**0.057**	0.453	**0.26**

Future Work

In terms of future work, we aim to evaluate the application of additional data analysis techniques for datasets that exhibit some periodicity in their values; and at the same time, to explore ways to more accurately simulate data that does not appear to be periodic. Additionally, our dataset generator could also serve as a tool towards the verifiability of various multiagent systems. Finally, we intend to extend our dataset generator to cover the V2G mode of EV operation, which is crucial for the efficient integration of renewables into the Grid.