

# Managing Duck Curve Type Energy Imbalances with Variational Recurrent Autoencoder-Based Clustering

by Alkiviadis Savvopoulos, Christos Alexakos and Athanasios Kalogeras (Industrial Systems Institute ATHENA Research Center)

**Peak residential energy demand does not always coincide with peak production times. This energy imbalance is known as the “duck curve”. Variational recurrent autoencoders can normalise the duck curve, optimise consumption profile clustering, and acquire useful insights for managing energy demand.**

The “duck curve” problem is a phenomenon that occurs in energy consumption, with significant imbalances occurring between renewable energy production yield and peak energy demand, owing to their short or non-existent temporal correlation [1]. The term “duck curve” refers to the collective visualisation of the production of renewable energy, the actual energy demand, and the energy supply from non-renewable sources, all as a function of time throughout the day. An abrupt demand surge coincides with the end of daily “green” solar energy production.

Strategies to solve this problem tend to focus on energy storage and management demand. However, algorithmic approaches may accelerate such efforts. The duck curve may be regarded as a specification of a unit commitment problem, where each building connected to the grid contributes to the formulation of the daily period of peak energy demand. Nevertheless, since not all energy grid participants’ daily peak demand coincides with total aggregate peak demand, the problem can be mitigated through algorithmic disambiguation between grid users, with demand peaking during the high renewable energy supply period and users heavily tolling the grid during aggregate high demand hours.

The proposed method analyses time series corresponding to residential building energy consumption profiles to determine their load demand characteristics. The sheer volume of data and exceptionally high sampling rate pose a challenge for data handling and analysis. With the aim of achieving an acceptable trade-off between dimensionality reduction and reconstruction loss, we elected to compress the data. Variational autoencoders of the unsupervised learning algorithmic category represent a compression algorithm, utilising state-of-the-art solutions with

renown abilities in robust pattern handling of vast amounts of datasets such as neural networks [2]. Moreover, the sequential nature of energy consumption profiles, whose values are partially dependent on the immediately preceding values, means that, due to the aforementioned temporal dependency, steep fluctuations are alleviated mitigated. We address this with long-short term memory (LSTM) neural networks, which take the place of the hidden layers inside the encoding and decoding modules of the variational autoencoders, to handle non-stationary input data and non-coherent energy patterns. The total process of the proposed method is shown in Figure 1.

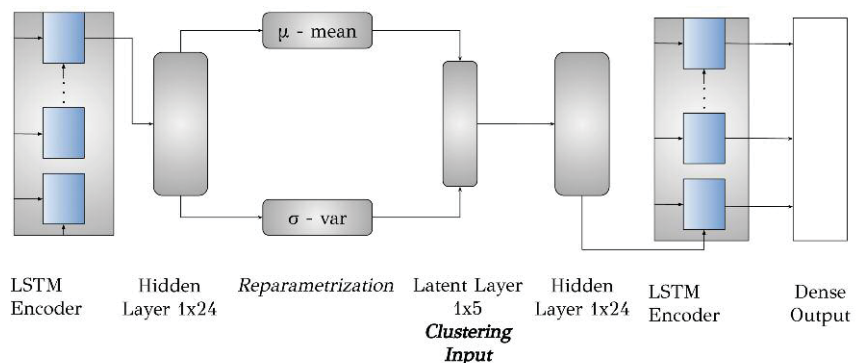


Figure 1: Variational recurrent neural network layer stack.

After the schema hyper-parameters are fine-tuned, the variational autoencoders are deployed, with an outer layer of 48 nodes corresponding to the half-hourly sampling rate of the dataset used, a hidden layer of 24 dimensions, and a double latent layer of five dimensions. This is where the reparameterisation technique of the variational autoencoders takes place, leading to the final compressed representation. The decoding module is symmetric and contains the same layer dimensions. Adaptation of the LSTM inside the schema, mandates the replacement of the classic feed forward neural networks, making the last layer of the

LSTM stack, the previous hidden layer that feeds the latent. Similarly, for decoding, the LSTM feeds a final outer dense layer.

After ensuring a much smaller feature space through the encoding step of the algorithmic pipeline, the analysis and clustering of these encoded daily load profiles can be implemented at much greater velocities. The next step of the procedure is to divide the compressed energy consumption profiles into different groups with similar characteristics. The clustering method utilised, in this work which represents this work’s its main contribution, is the creation of custom cluster representatives

as a reference point for each encoded time series. The fundamental concept of this method is that in order to disambiguate between residential building consumption patterns, two clusters are created: one that correlates with daily patterns whose peak demand resides during the hours of high renewable energy supply, and a second that comprises all time series not compliant with “green” energy.

In order to create these representations, which will be compared to each compressed load profile and compute their corresponding similarities, daily solar radiance is taken into consid-

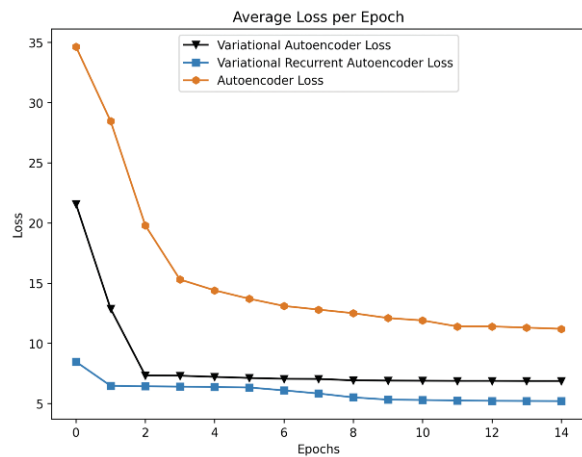


Figure 2: Neural network training results comparison.

eration. The first cluster is constructed maintaining the dataset scaled median value as its maximum, throughout the first morning hours until late afternoon, when solar output is at its highest. The second cluster possesses values at a similar scale, in a sense, complementary, as the evening and night hours must be represented. Then, vectors are encoded, scaled and compared with each time series, so as to determine its

class and assign to the corresponding cluster.

Using daily energy consumption measurements of 5,567 London Households [L1] as experimental data, variational recurrent autoencoders outperformed both simple variational autoencoders with classic encoding modules, and vanilla autoencoders without reparametrisation, in terms of loss during

training. In order to introduce experimental robustness, the aforementioned neural networks were also tested; however, due to the structure of the household energy data, the proposed schema captured time-dependent intricacies more efficiently, as seen in Figure 2. With reference to clustering results, as depicted in Figure 3, our method performed an efficient separation of time series according to peak demand, into a renewable energy compliant cluster, and a second cluster containing load profiles that must be dealt with in order to solve the duck curve issue.

In summary, the proposed algorithm clusters the dataset's input points according to their peak demand time index throughout the day, while simultaneously outperforming classic clustering approaches such K-means and spectral clustering. The algorithm does not surpass traditional clustering techniques in terms of classic metrics such as precision and recall. Rather, its strength lies in its ability to create load profile groups that can produce meaningful insights, and efficiently detect whether a profile correlates with renewable energy production standards, in real time.

We acknowledge support of this work by the project "Enabling Smarter City in the MED area through Networking-ESMARTCITY" (3MED171.1M2022), which is implemented under the "Interreg Mediterranean" programme, co-financed by the European Regional Development Fund (ERDF), Instrument For Pre-Accession Assistance (IPA) and national sources.

#### Link:

[L1]: <https://kwz.me/h4j>

#### References:

- [1] H.O.R. Howlader, et al.: "Optimal thermal unit commitment for solving duck curve problem by introducing CSP, PSH and demand response". IEEE Access, 2018.
- [2] D. P. Kingma, & M. Welling: "Auto-encoding variational bayes", ICLR2014.

#### Please contact:

Alkiviadis Savvopoulos  
Industrial Systems Institute, ATHENA  
Research Center, Greece  
savvopoulos@isi.gr

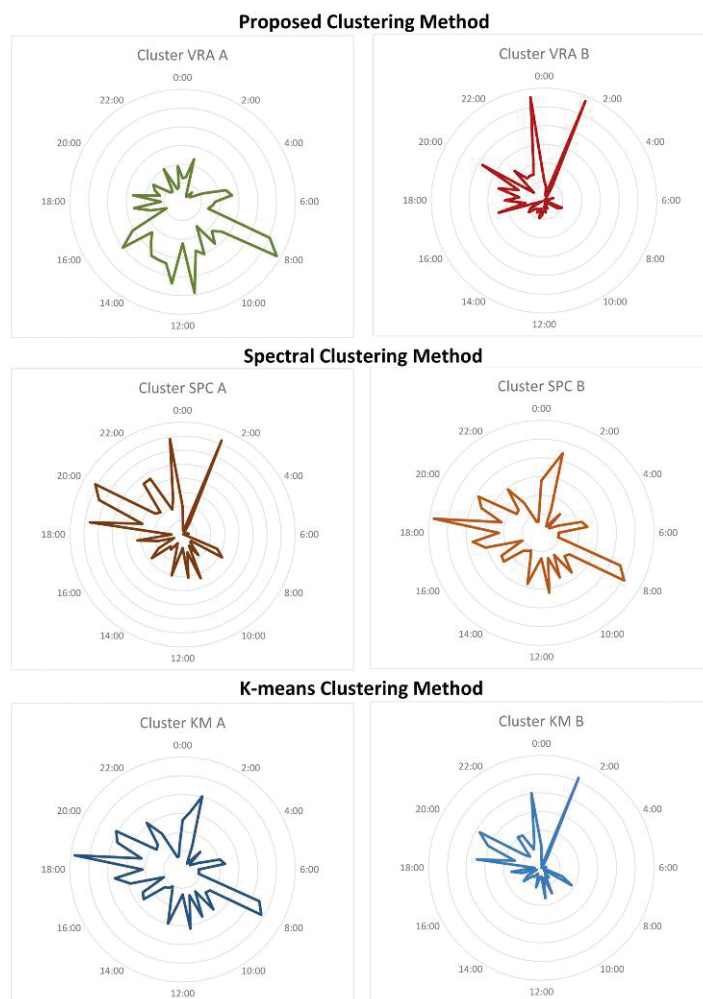


Figure 3: Experimental clustering results. The figure presents the categorisation of the analysed time series in two clusters, a "green" compatible cluster and a non-compatible cluster, according to their consumption peak times, utilising three approaches: the proposed VRNNL, the spectral and k.-means algorithms.