Abstract

Electricity demand profiles of dwellings are mainly composed of various known (deterministic) and unknown (stochastic) processes. Effective data processing approaches (such as time series decomposition) are mainly used to simplify underlying patterns in the complex stochastic processes by fragmenting the different layers of hidden processes (referred as components of time series). This paper will demonstrate the performance of state-of-the-art STL (a Seasonal-Trend decomposition procedure based on Loess) techniques (Cleveland, Cleveland, McRae, & Terpenning, 1990), embedded within the framework of the HMM-GP model, in simulating dynamics of high-resolution electricity demand data. The method is applied to the case studies located in the Findhorn community.

Introduction

The modelling and analysis of the electricity demand patterns are essential for developing sustainable plans, policies, management of infrastructure and strategies for optimisation of resource utilisation. The electricity demand modelling is mainly used to study the variation in the short/medium/long-term electricity usage patterns either for an individual dwelling or at the community/national scale. Unlike the physical models which can be calibrated using the building specific physical parameters and relies on the expert knowledge, statistical/computational models are purely based on the historically observed records. Recently, in the application areas of the energy demand modelling, considerably large amount of novel statistical and computational approaches has been developed. A thorough literature review of some of the widely applied data-centric approaches such as time series models, regression models, decomposition models, ARIMA type models, hybrid ANN models including support vector regression), Fuzzy logic models, Genetic algorithm, etc., including an overview on conventional energy system models (e.g. MARKAL/TIMES/LEAP), can be referred somewhere else (Suganthi & Samuel, 2012; Bhattacharyya & Timilsina, 2009; Bhattacharyya & Timilsina, 2010).

The electricity demand patterns are known to be impacted by a range of socio-economic and environmental factors. The complex dynamic of the electricity demand time series is, therefore, the consequence of complex interaction and influences imparted by these multiple factors. The degree of influence of different factors could vary significantly, for example some of these factors can be realised as a constant element that varies gradually over the longer period (referred as “Trend”) whereas some of the other factors could vary in the form of repetitive periodic patterns (referred as “Seasonal”).

These intrinsic features embedded in the time series can be explored in the form of determinism and stochasticity through the application of time series decomposition methods. A range of time series decomposition approaches has been developed that aiming to segregate deterministic and stochastic features from the time series and has been thoroughly reviewed elsewhere (Rios & De Mello, 2012). Some of the most commonly applied approaches in this category are Fourier Transform (FT) (Stoica & Moses, 2004), Wavelet Transform (WT) (Qian, 2002), and Empirical Mode Decomposition (EMD) [(Huang, et al., 1998) (Rios & De Mello, 2016)]. These approaches are thoroughly reviewed elsewhere (Rios & De Mello, 2012). Another widely applied categories of the time series decomposition approaches are based on the process of de-constituting time series into a set of non-observables (latent) components (Dagum E. B., 2010).

According to (Persons, 1919), time series are mainly composed of four such components: i) Long-term trends; ii) Cyclic movements - often superimposed on the long-term trends (also referred as cyclic-trends); iii) Seasonal movements (which should not be confused with the four typical meteorological seasons of a year); iv) Residual/random variations. The latent components can be associated with different types of temporal variation and thus the presence of a latent component in a series strongly depending on the nature of the process and on the temporal resolution of the series. For example, some system/processes demonstrate seasonality due to certain specific periodic activities occurring at different temporal scales, e.g. at some specific time of day (daily), at some specific day of the week (weekly), monthly, quarterly, etc.

One of the simplest approaches for extracting latent component is the Classical Decomposition (CD) method. The CD method is based on the application of the method of moving averages to extract trend cycles and then utilises averaging of the detrended time series over a specified time period to extract a constant seasonal factor,
Some of the key features of the STL approach that makes this method suitable for application in the electricity demand decomposition are: i) the STL approach can be applied to any type of seasonality (daily, weekly, monthly, quarterly, annual, etc.); ii) the STL approach allows seasonal components to vary over time; iii) it provide flexibility for selecting smoothing parameter for trend extraction; iv) the method is robust to outliers. One of the key limitations of the STL approach is that it is only suitable for the application with the additive time series and therefore any multiplicative time series need to be appropriately transformed before application of the STL decomposition.

The high-resolution electricity demand profiles are essential for understanding detailed energy consumption behaviour. For example, certain activities that are occurring at fine temporal resolutions and could have a significant impact across the communities of dwellings (such as the use of a kettle at a certain time, say after a popular TV program nationwide could trigger a sharp spike in aggregated profile). To access the impact of such events a large amount of individual demand profiles at high resolution is required and as such current information at this level is not widely available. Thus one of the interesting application of the proposed modelling approach could be the synthetic simulation of high-resolution electricity demand profiles. Other potential applications could be in context-sensitive smart data-infilling, short-term projections for developing effective demand response strategies, long-term projections for developing future planning, policies and management strategies.

**Case study**

The paper will demonstrate the application of the proposed methodology (detailed in the next section) through the application to a few selected case study buildings located in the Findhorn Foundation Eco-Village (FFEV) community. The FFEV is located in the North East of Scotland near to the small town of Forres and maintains a sustainable lifestyle (FINDHORN FOUNDATION, 2019). Most of the buildings in the FFEV utilises modern insulation, triple glazed windows, intelligent design processes to maximise passive solar gain for optimising overall energy efficiency while effectively using environment-friendly recycled materials where possible, such as straw bales, recycled whiskey barrels and grass roofing. In addition to this, the community uses renewable energy technologies such as solar and wind. The building in the FFEV community can be grouped under four categories based on building energy system type: i) Air Source Heat Pump (ASHP) or Ground Source Heat Pump (GSHP); ii) Gas boiler (GB); iii) Electric Boiler (EB); iv) District Heating using Biomass (DHB). Under the construction type, the buildings in FFEV can be grouped under seven categories: i) Findhorn Construction (FC); ii) Findhorn + Construction (F+C); iii) Soillse Construction (SC); iv)
Centinis Construction (CC); v) Whins Construction (including, Ecomobile) (WC); vi) Stonebuilt Construction (SC); vii) Parkhome Construction.

Data: Organisation and Processing

This section aims to provide a brief overview of the data acquisition, organisation and processing for the present study. The FFEV has approximately 211 buildings. As part of the previous and on-going research projects (e.g. ORIGIN (ORIGIN project website, n.d.), CESI (Centre for Energy Systems Integration (CESI) (Grant: EP/P001173/1), n.d.) approximately 55 buildings were fitted with smart meters that collected data at 5 minutes resolution over a two year period. To demonstrate the effectiveness of the proposed methodology and to ensure a robust model development suitable for potential applications, the dataset needs to be measured over a continues time-period of considerable length (in this case, due to limited availability of high-resolution data measured over a continues long-period and for a diverse range of buildings, we focus on to capture seasonal variations at weekly levels). A thorough investigation examining data quality suggested that for approximately 14 dwellings, good quality (i.e. with less than 5% missing data) five minutely electricity demand profiles can be organised for a period of up to six weeks from (15 Feb 2015 to 28 March 2015). Please refer to Table 1 for details on various aspects of these 14 dwellings including missing data.

Table 1: Case study building in FFEV used in this study

<table>
<thead>
<tr>
<th>Building ID</th>
<th>Energy System</th>
<th>Construction Type</th>
<th>% of missing data</th>
</tr>
</thead>
<tbody>
<tr>
<td>A03</td>
<td>GB</td>
<td>FC</td>
<td>0.29</td>
</tr>
<tr>
<td>A18</td>
<td>DHB</td>
<td>SC</td>
<td>0.37</td>
</tr>
<tr>
<td>B02</td>
<td>GB</td>
<td>WC</td>
<td>0.03</td>
</tr>
<tr>
<td>B08</td>
<td>GB+ST</td>
<td>FC</td>
<td>0.05</td>
</tr>
<tr>
<td>C02</td>
<td>GB+SPV</td>
<td>F+C</td>
<td>1.62</td>
</tr>
<tr>
<td>C14</td>
<td>GB</td>
<td>F+C</td>
<td>5.69</td>
</tr>
<tr>
<td>C19</td>
<td>GB+ST</td>
<td>F+C</td>
<td>2.95</td>
</tr>
<tr>
<td>C24</td>
<td>GB</td>
<td>F+C</td>
<td>1.90</td>
</tr>
<tr>
<td>C34</td>
<td>EB</td>
<td>CC</td>
<td>1.39</td>
</tr>
<tr>
<td>D08</td>
<td>ASHP/GSHP+SPV</td>
<td>WC</td>
<td>1.23</td>
</tr>
<tr>
<td>D15</td>
<td>ASHP/GSHP+SPV</td>
<td>WC</td>
<td>4.19</td>
</tr>
<tr>
<td>D20a</td>
<td>ASHP/GSHP+SPV</td>
<td>WC</td>
<td>0.21</td>
</tr>
<tr>
<td>D25b</td>
<td>ASHP/GSHP+SPV</td>
<td>WC</td>
<td>0</td>
</tr>
<tr>
<td>F27</td>
<td>ASHP/GSHP+SPV</td>
<td>WC</td>
<td>3.31</td>
</tr>
</tbody>
</table>

Some addition symbols used in the table are ST for Solar Thermal and SPV for Solar PV.

Missing data: for the cases when there is only one data point is missing, data were infilled using standard interpolation approaches. For instance, with more than one consecutive missing data points, a strategic approach based on the preliminary data analysis that focuses on the temporal patterns of the long time series has been used. Such data points were infilled with corresponding values observed either a week before/after. The data infilling strategy is aligned to the further pieces of evidences of seasonal element observed at weekly scale in a later section (please refer to Figure 4 and 5).

Figure 1: Boxplot of 5 minutely electricity demand for 14 buildings over the study period (15th Feb 2015 – 28th March 2015).

A thorough preliminary exploratory data analysis comprising of five summary statistics for all 14 case studies are illustrated in Figure 1. The five summary statistics displayed in the boxplot includes: i) 5th percentile; ii) 25th percentile; iii) 50th percentile; iv) 75th percentile; v) 95th percentile. Maximum values are displayed over the individual boxplots in the black colour text. Figure 1 clearly demonstrates that the statistical characteristics of the 14 buildings included in the study are considerably varying. For the illustration purpose, two buildings with considerably distinct statistical characteristics have been selected to demonstrate the effectiveness of the proposed methodology.

Methodology

A time series is usually referred to as a set of observations that are collected at equally spaced successive points in the time (Box, Jenkins, Reinsel, & Ljung, 2015). Let \( X(t) = \{X_t; t \in T\} \) denotes a time series, where \( T = \{t_1, t_2, ..., t_n\} \) are \( n \) equally spaced sequential times stamps. The key idea underpinning a time series decomposition procedure is that a time series can be decomposed into three unobserved (latent) components: Trend \( T(t) \), Seasonal \( S(t) \) and Remainder \( R(t) \) such that, we have:

\[
X(t) = T(t) + S(t) + R(t)
\]  

(1)

\[
X(t) = T(t) \ast S(t) \ast R(t)
\]  

(2)

for additive and multiplicative time series, respectively. For additive decomposition model, the latent components are considered to be mutually independent whereas if
Fit HMM to the random component

The idea underpinning proposed methodology is that the observed patterns of electricity demand profiles are mainly composite of several systematically/repeatedly occurring/influencing physical activities and/or natural processes, which are usually associated with a high element of uncertainty (referred as “Randomness”). Thus, the STL based filter can be utilised to de-constituting the complex dynamic of the process into simplified deterministic ("Trends" and "Seasonality") and stochastic ("Random") components. These simplified patterns can be analysed to understand various factors influencing the demand patterns and can be modelled within the HMM-GP framework for synthetic simulation of the electricity demand profiles. The HMM-GP model is applied mainly to the random component, extracted as part of STL procedure. The application of HMM-GP model is intended to generate multiple random components. The synthetic electricity demand profiles are constructed by recomposing the simulated random components with the Trend and Seasonal components extracted from the observed series.

The proposed STL-integrated HMM-GP methodology is applied to all 14 case-studies for the purpose of constructing aggregated demand profiles (one of the key potential application of the proposed methodology, as discussed in the next section). However, two case studies, specifically A03 and D08, has been selected for the demonstration purpose. The selection of case studies is to illustrate the effectiveness of the proposed methodology for capturing the variations in the time series due to various factors, such as building energy system, the impact of solar PV and construction types as presented in Table 1. To illustrate the potentials of the proposed time series simulation methodology, a thorough analysis of key statistical characteristic (specifically, comparison of percentiles at step-size of 1 including percentiles of peak load, probability density distribution, etc.) of observed and simulated time series has been presented.

**Result and Analysis**

The proposed methodology can be utilised for a range of applications, such as in infilling missing data, to generate synthetic profiles from a small sample to generate user-specified number of synthetic profiles (these synthetic profiles can be aggregated to construct aggregated demand profile to understand/analyse community demand patterns), forecasting short-term to long-term high resolution demand that can be used for future planning of infrastructures, policies development, energy trading, uncertainty analysis, etc. However, to serve these various application areas it is important to demonstrate the efficiency of the proposed methodology in generating synthetic demand profiles with similar/close statistical characteristics such that they can be utilised as a realistically possible alternative of observed profiles. This section is aimed to demonstrates the efficiency of proposed STL based HMM-GP approach in simulating statistically robust synthetic demand profiles.

**The STL decomposition**

The STL decomposition approach as discussed in above sections is applied to deconstruct the five-minutely observed demand profiles of two specified case-study buildings, A03 and D15, selected from the FFEV (referring to Figure 4 and 5 respectively). The selected observed series covers a period of six-weeks starting from

---

**Figure 2: The HMM-GP modelling framework.**

The paper is based on the novel methodology (illustrated in Figure 2), to examine the potentials of the proposed STL based time series decomposing method, that will be fitted within the framework of the HMM-GP model. A brief overview of the Hidden Markov model (HMM) that underpins the HMM-GP methodology is separately illustrated in Figure 3. The HMM-GP modelling framework is originally developed for simulating synthetic flow time series and can be referred elsewhere [Patidar, Allen, Haynes, & Haynes, 2018], (Patidar, Jenkins, & Simpson, 2014], (Patidar, Jenkins, & Simpson, 2016]. However, in the context of synthetic simulation of high-resolution electricity demand profiles, this paper is the first attempt that will test the ability of the proposed STL based HMM-GP modelling framework.

---

**Figure 3: The Hidden Markov Model (HMM), comprising of five elements.**

| Step 1 | Take log of the time-series to transforms an additive time series into a multiplicative series. |
| Step 2 | Apply STL decomposition to segregate the log time-series into three: trend, seasonal and random components. |
| Step 3 | Fit HMM to the random component to generate simulated random component. |
| Step 4 | Construct synthetic demand profiles by adding simulated random components with the seasonal and trend component. |
| Step 5 | Fit Generalised Pareto (GP) distribution to the observed extreme values (i.e. over 95th percentile) and resample simulated extreme values to account in unseen extremes. |

The paper is based on the novel methodology (illustrated in Figure 2), to examine the potentials of the proposed STL based time series decomposing method, that will be fitted within the framework of the HMM-GP model. A brief overview of the Hidden Markov model (HMM) that underpins the HMM-GP methodology is separately illustrated in Figure 3. The HMM-GP modelling framework is originally developed for simulating synthetic flow time series and can be referred elsewhere [Patidar, Allen, Haynes, & Haynes, 2018], (Patidar, Jenkins, & Simpson, 2014], (Patidar, Jenkins, & Simpson, 2016]. However, in the context of synthetic simulation of high-resolution electricity demand profiles, this paper is the first attempt that will test the ability of the proposed STL based HMM-GP modelling framework.
15th February 2015 to 28th March 2015. The STL decomposition approach has been shown to segregates the deterministic (“Trend” and “Seasonal”) and stochastic (“Random”) feature of electricity demand profiles of the two case study buildings. A weekly seasonal window has been selected. The statistics of the STL decomposition can be referred in Table 2 and 3 respectively for A03 and D15, which clearly demonstrate that the average of the random and seasonal components to zero, although these two components appear to capture most of the variance of the observed series.

Similarly, change in trend patterns can be perturbed or modelled in relation to certain future forecasted changes.

![Figure 4: The STL decomposition of five minutely electricity demand profile of case study building A03 conducted over the six-week period from 15th February 2015 to 28th March 2015.](image)

**Table 2: Statistics for STL decomposition A03**

<table>
<thead>
<tr>
<th>Electricity demand</th>
<th>Trend</th>
<th>Seasonal</th>
<th>Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>0.44</td>
<td>0.44</td>
<td>0.00</td>
</tr>
<tr>
<td>Variance</td>
<td>0.41</td>
<td>0.01</td>
<td>0.11</td>
</tr>
</tbody>
</table>

The HMM-GP model is applied to the random component. Application of the HMM-GP model generates simulated random components that capture the statistical variability of the process. The trend and seasonal components are added back to the simulated random components to generated synthetic demand profiles which preserve deterministic features specific to the household of the observed series in the synthetic profiles.

**Table 3: Statistics for STL decomposition D15**

<table>
<thead>
<tr>
<th>Electricity demand</th>
<th>Trend</th>
<th>Seasonal</th>
<th>Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>0.50</td>
<td>0.50</td>
<td>0.00</td>
</tr>
<tr>
<td>Variance</td>
<td>0.41</td>
<td>0.01</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Comparing individual demand profiles

This section is intended to illustrate the abilities of the STL based HMM-GP model in replicating intrinsic features of the electricity demand time series in the simulated synthetic series. Figure 6 and Figure 7 shows the patterns of the observed and synthetic five-minutely electricity demand series for a one-week period (starting from 15th February 2015 to 21st February) for the case study buildings A03 and D15 respectively.

![Figure 5: STL decomposition of five minutely electricity demand profile of case study building D15 conducted over the six-week period from 15th February 2015 to 28th March 2015.](image)

![Figure 6: Comparing five minutely electricity demand profile of case study building A03 with a synthetic demand profile generated using the STL based HMM-GP method for a one-week period from 15th February 2015 to 21st February 2015.](image)

From the visual inspection of Figure 6 and 7, the proposed STL based HMM-GP model appears to capture the essential features, variability and temporal dynamics of the demand time series. To examine the abilities of the proposed method in capturing key statistical features of the observed series a thorough investigation comparing percentile and probability density distributions of the observed and synthetic series has been conducted. For the
purpose of illustration 20 synthetic series are generated using the STL based HMM-GP model for both the case studies, A03 and D15.

Figure 7: Comparing five minutely electricity demand profile of case study building D15 with a synthetic demand profile generated using the STL based HMM-GP method for a one-week period from 15th February 2015 to 21st February 2015.

Comparing Percentiles

Percentiles analysis of the 20 synthetic demand profiles, shown in Figure 8, appears to closely follow the patterns of the percentile of the observed series. The selection of 20 synthetics is arbitrary, although methodology can be used to generate any user-specified number of synthetic profiles. The percentile analysis has been conducted for the percentiles starting from 0th to 100th percentile with a step-size of 1 (presented on a scale of 0.0 to 1.0 on the x-axis).

Figure 8: Comparing percentile analysis of observed electricity demand profiles with synthetic demand profiles generated using STL based HMM-GP model for case study building A03 (Upper panel) and D15 (Lower panel). Percentile analysis is conducted over the six-week period from 15th February 2015 to 28th March 2015.

The upper panel demonstrates the percentiles compared for the case study A03 whereas the lower panel demonstrates the percentile analysis for the case study D15. The percentiles match is considerably good for most of the values below the 85th percentile for A03 (75th percentile for D15), with some slight variations noted for higher demand values. One possible reason for this could be the large variations of the demand values in the successive high percentiles. However, the model appears to perform considerably well in estimating extreme peak demand values (i.e. values over 90th percentile). This is due to the fact that the proposed model utilises the potential of Generalised Extreme Values (GEV) distributions, specifically Generalised Pareto (GP) distribution, to effectively model the extreme peak demand values over the 95th percentiles.

Comparing probability density distribution (pdd)

The pdd are considered as one of the key statistical property of a time series. To further assess the suitability of the proposed approach, the pdd of the observed electricity demand profiles are compared with the pdd of 20 synthetic demand profiles for both the case studies in Figure 9. The pdd comparison shows a close match between the observed and the simulated synthetic profiles with a slight variation noted for the second peak that is occurring at low demand values in the synthetic profiles in comparison to the observed demand. This feature indicates the possibility of a consistent underestimation of the high load values (over 1 kW) by the model. This type of issues can be generally overcome by means of statistical bias correction approach and will be possibly considered as part of future work.

Figure 9: Comparing probability density distribution (pdd) of observed electricity demand profiles with synthetic demand profiles generated using STL based HMM-GP model for case study building A03 (Upper panel) and D15 (Lower panel). Percentile analysis is conducted over the six-week period from 15th February 2015 to 28th March 2015.

Comparing aggregated demand profiles (pdd)

One of the potential application of the proposed methodology will be in generating aggregating demand profiles. Figure 10 illustrates the dynamical behaviour of aggregated electricity demand profiles, observed for all the 14 case studies. The aggregated demand profiles are constructed for the observed and corresponding synthetic
demand by adding five-minutely profiles over the six-week period from 15th Feb 2015 – 28th March 2015. The synthetic demand profiles are generated through the application of the STL based HMM-GP model and one synthetic profile corresponding to each of 14 case studies is randomly selected for constructing an aggregated synthetic demand profile.

Figure 10: Comparing five-minutely aggregated electricity demand profiles for 14 case-studies, observed with corresponding 14 synthetic electricity demand profiles. Comparison is conducted over the six-week period from 15th February 2015 to 28th March 2015.

Conclusion

A range of data-driven statistical and computation techniques has been developed recently for modelling electricity demand patterns. This paper presented a thorough literature review of recent developments in the area of time series decomposition approaches including their application in the context of the energy demand simulation.

The paper investigated the potentials of an STL based time series decomposition approach in enhancing the capabilities of the HMM-GP modelling framework. The HMM-GP modelling framework is designed for modelling of high-resolution electricity demand patterns that integrates a Generalised Pareto (GP) distribution for effective simulation of extreme values within the framework of HMM. The application of STL based time series decomposition procedure is intended to develop an effective data pre-processing strategy with an aim for enhancing the overall capabilities of the HMM-GP models. To demonstrate the effectiveness of the proposed modelling schematics, a thorough analysis of various statistical properties including percentile analysis and probability density distribution of the simulated and original time series has been conducted.

The results presented in the paper shows that the proposed methodology is considerably effective in capturing the key dynamics of the process and therefore can be seen as a potential tool with several applications that require synthetic simulation of electricity demand profiles of high resolution. One of the key features of the proposed approach is its ability to effectively capture the extreme (peak) loads. Another direct benefit of the proposed methodology is that it is purely based on the observed records and can be applied to a small/large data set (in this case for 6 week’s data). One of the current limitations of the proposed methodology, considering the potential for wide-scale application, is that the approach needs to be robustly validated for various potential applications and by accounting in a range of building types, geographic locations, temporal resolutions, etc. The method has the potential to be further developed to simulate the impacts of the various factors impacting electricity demands (such as climate, occupancy, life-style, etc.). In addition to all this, there are lots of scope for enhancing the overall technical elements of the proposed methodology, e.g., possibility of integrating other advance time series decomposition methods, use of different distributions within the HMM for effective simulation of higher percentile load values, etc..

Acknowledgement

This work is supported by two EPSRC funded projects: UK Centre for Energy Systems Integration (Centre for Energy Systems Integration (CESI) (Grant: EP/P001173/1), n.d.) and Community Energy Demand Reduction in India (Community-scale Energy Demand Reduction in India (CEDRI) (Grant EP/R008655/1), n.d.).

Bibliography


*Community-scale Energy Demand Reduction in India (CEDRI)* (Grant EP/R008655/1). (n.d.). Retrieved from www.cedri.hw.ac.uk


