

ARTICLE



<https://doi.org/10.1038/s41467-022-31942-9>

OPEN

# Data-driven load profiles and the dynamics of residential electricity consumption

Mehrnaz Anvari<sup>1,10</sup>✉, Elisavet Proedrou<sup>2,10</sup>, Benjamin Schäfer<sup>3,4,5,10</sup>, Christian Beck<sup>3,6</sup>,  
Holger Kantz<sup>7</sup> & Marc Timme<sup>8,9</sup>✉

The dynamics of power consumption constitutes an essential building block for planning and operating sustainable energy systems. Whereas variations in the dynamics of renewable energy generation are reasonably well studied, a deeper understanding of the variations in consumption dynamics is still missing. Here, we analyse highly resolved residential electricity consumption data of Austrian, German and UK households and propose a generally applicable data-driven load model. Specifically, we disentangle the average demand profiles from the demand fluctuations based purely on time series data. We introduce a stochastic model to quantitatively capture the highly intermittent demand fluctuations. Thereby, we offer a better understanding of demand dynamics, in particular its fluctuations, and provide general tools for disentangling mean demand and fluctuations for any given system, going beyond the standard load profile (SLP). Our insights on the demand dynamics may support planning and operating future-compliant (micro) grids in maintaining supply-demand balance.

<sup>1</sup>Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, P.O. Box 60 12 03, D-14412 Potsdam, Germany. <sup>2</sup>DLR Institute for Networked Energy Systems, Oldenburg, Germany. <sup>3</sup>School of Mathematical Sciences, Queen Mary University of London, London, UK. <sup>4</sup>Faculty of Science and Technology, Norwegian University of Life Sciences, 1432 Ås, Norway. <sup>5</sup>Institute for Automation and Applied Informatics, Karlsruhe Institute for Technology, Karlsruhe, Germany. <sup>6</sup>The Alan Turing Institute, London, UK. <sup>7</sup>Max Planck Institute for the Physics of Complex Systems, D-01187 Dresden, Germany. <sup>8</sup>Chair for Network Dynamics, Center for Advancing Electronics Dresden (cfaed) and Institute for Theoretical Physics, Technical University of Dresden, 01062 Dresden, Germany. <sup>9</sup>Lakeside Labs, 9020 Klagenfurt am Wörthersee, Austria. <sup>10</sup>These authors contributed equally: Mehrnaz Anvari, Elisavet Proedrou, Benjamin Schäfer. ✉email: [anvari@pik-potsdam.de](mailto:anvari@pik-potsdam.de); [marc.timme@tu-dresden.de](mailto:marc.timme@tu-dresden.de)

Electrical energy is an essential part of daily life and is continuously generated, transmitted, stored and, finally, consumed. Generation and storage of electricity shall match the dynamic consumption of residential, industrial and other sectors at all times. To maintain the balance between the electricity generated by energy providers, and the electricity consumed by consumers, energy suppliers need to estimate the electricity required by all consumer sectors on a broad range of time scales from seconds to days. Estimating the typical variations in the electricity demand over the course of a day yields a load profile, which can be attained either through the definition of a methodology to extract a load profile from empirical data, the creation of a model or a combination of both.

Research aiming at developing load profiles goes back to at least the 1940s<sup>1</sup>, however, the issue of finding a precise, high-resolution load profile is becoming more and more urgent due to the increase of population, electrical heating systems, electrical vehicles for transportation, solar home systems as well as the increasing share of fluctuating renewable energy (RE) feed-in and the construction of distributed power grids, especially smart grids. In 1999, the first methodologically systematic German household load profile, known as the H0 Standard Load Profile (H0 SLP), was developed<sup>2</sup> and has since been in use without alterations in at least Germany and Austria<sup>3</sup>, see also Supplementary Note 6 for translations of these German language references.

We focus here on the residential sector consuming around 29% of all electricity in the European Union<sup>4</sup>, and it is expected that its electricity consumption will considerably increase because of the electrification of both the transport and the heating system. Although initially household load profiles used a temporal resolution of around one hour, newer models have a temporal resolution up to and including one second, due to the recent availability of highly resolved datasets of electricity consumption and the usage of smart metres in selected houses, see Supplementary Note 1. These new datasets allow the grid operators to record the electricity consumption of individual houses at high temporal resolutions. Several previous studies<sup>5–7</sup> analysed high temporal resolution datasets and reported the presence of extreme and significant peaks in the loads which have not been reported for datasets with temporal resolutions of 15 min to 1 h. However, a structured approach to translate these high-resolution datasets into usable load profiles is not readily available to date. Therefore, in this work, we first analyse highly resolved electricity consumption data for groups of houses in Austria, Germany and the UK. Our data-driven analysis indicates the potential for the presence of strong fluctuations and high levels of unpredictability in the distribution grids, see “Complex demand dynamics—the necessity of new load profiles”. Based on this analysis, we introduce a load model where we disentangle an average load profile from the fluctuations on top of the baseline. We thereby obtain a new load profile that is consistent with high-resolution electricity consumption data.

As mentioned above, one important reason of having a high-resolution load profile is the increasing share of RE feed-in. In contrast to electric energy generated from burning fossil fuels, RE feed-in is weather-dependent, intermittent, and highly variable<sup>8–10</sup>. It thus becomes harder to balance supply and demand. As the share of renewable feed-in is increasing, recent and current research focuses on gaining a deeper understanding of the electricity generated by RE as well as advanced approaches of balancing demand and supply, e.g., by load-shifting<sup>11,12</sup>. In contrast, the dynamics of electricity consumption is far from understood, in particular on the residential level, partially because electricity consumption data are difficult to obtain.

In this article, we offer a method for (a) extracting a new type of load model from a given set of consumption data to

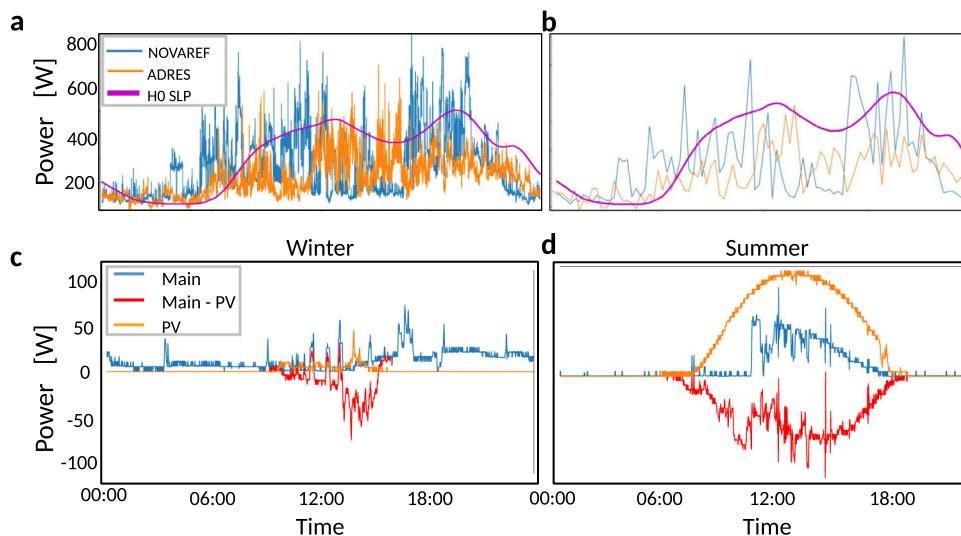
disentangle the baseline demand dynamics over the course of a day (average load profile, ALP) and the fluctuations on top of that baseline (stochastic fluctuation profile, SFP) and (b) better understanding the consumer demand dynamics. Taking into account available microscopic data, the model is capable of characterising (i) the demand of already moderately sized sets of consumers with (ii) high temporal resolution and (iii) applicable to a range of data, including those recorded in the future (and not restricted to those analysed in this article). Together, these features and capabilities go far beyond what the standard load profile (H0 SLP) offers and our approach is distinct from more recent complementary approaches since our model does not require microscopic parameters for consumer behaviour, consumer appliances, house infrastructures or other features that other models depend on.

After reviewing the need for a new demand model in more technical detail (see the section “Complex demand dynamics—the necessity of new load profiles”), we analyse the electricity consumption data of Germany, Austrian and UK houses measured for several weeks. We disentangle the variations in the consumption dynamics into two main factors, the average load profile (ALP) and the statistics of short-time fluctuations of the ALP. First, through the application of the empirical mode decomposition (EMD)<sup>13</sup>, we extract the average load profile from the time series data. This extracted trend captures the demand much more accurately than the often-used H0 SLP (see the section “Demand trend: mode decomposition”). Next, we use superstatistics to model the fluctuations around this trend into a stochastic fluctuation profile (SFP). Combining the trend and fluctuation analysis, we successfully reproduce a synthetic high-resolution load profile, yielding a full data-driven load profile (DLP). Our modelling approach readily transfers to use on other datasets. To facilitate such a transfer, we are providing executable code (see the section “Demand fluctuations: stochastic model”). We thereby develop a load profile methodology that is applicable to existing power grids and datasets, and also provide the tools for extracting load profiles in different regions and under different boundary conditions, for instance for microgrids with high shares of on-site generation or electric cars.

## Results

**Complex demand dynamics—the necessity for new load profiles.** Notwithstanding recent advances, energy suppliers still mostly use the older load profiles, such as the H0 standard load profile, which only has a 15-min temporal resolution. In Germany, the standard load profile (H0 SLP), was introduced in 1999. Ninety percent of the residential load data used in its creation were measured in the 1970s or earlier with an hourly temporal resolution. Only 10% of the measured load data had a temporal resolution of 15 min. Here, we review some of the recent advances towards a modern load profile and touch on the persistent need for a generally applicable consumption framework which we provide in the next sections.

Three well-known model classes exist to describe a household load profile. Top-down or conditional demand analysis models are downward models that use the total electricity consumption estimates of multiple households as well as macro-variables to model the dynamics of household energy consumption by generating household load profiles<sup>14,15</sup>. Bottom-up models use micro-variables as input, such as the number of active occupants, the appliances’ energy demand and usage time etc. They also often use Markov chains to generate household load profiles<sup>16,17</sup>. Finally, hybrid models employ a combination of the techniques used in top-down and bottom-up models to build up a Statistical Adjusted Engineering (SAE) model<sup>18</sup>. A detailed analysis of how



**Fig. 1 Systematic, asymmetric and intermittent deviations of empirical power consumption dynamics and German standard load profile (HO SLP).**

**a** We compare the HO SLP with the averaged real consumption data of a single day in winter for the NOVAREF project, the German dataset recorded between 2013 and 2016 including 12 houses, and ADRES project, the Austrian dataset recorded between 2009 and 2010, including 30 houses. The averaged real consumption data are created by taking the recorded electricity consumption of all houses in the measured time interval and, then, evaluating  $P_{avg}(t) = \frac{1}{N} \sum_i^N P_{h_i}(t)$ , where  $P_{h_i}$  is the measured electricity consumption of each house and  $N$  is the number of houses, which is respectively 12 and 30 for the NOVAREF and ADRES data. All three datasets were upsampled to a sampling rate of 2 s (0.5 Hz). The HO SLP not only fails to capture the correct daily trend, but we also observe large fluctuations of the consumption at short time scales. **b** Shows the same datasets at 15-min resolution to emulate the temporal resolution that was used to generate the HO SLP. Here the failure of the HO SLP to capture the correct daily trend is even more pronounced. **c, d** show respectively the recorded household power consumption in winter (January 1, 2017) and summer (June 30, 2017) with and without photovoltaics as well as the electricity generated by the PV module installed on the roof of a house at a sampling rate of one minute (PV data source<sup>70</sup>). Note the smaller spikes visible in panels **(a, b)** because the data is averaged over 12 and 30 houses, respectively, for NOVAREF and ADRES, while in panel **c** and **d** only a single household is shown.

these models have been applied was recently reviewed in ref. <sup>19</sup>. At present, many demand analysis models exist that can generate daily residential electricity load profiles, see Supplementary Note 1,<sup>1,20</sup> for details. Only a few of these models use a high temporal resolution of the order of seconds and those require a lot of micro-parameters, which still leaves us with the need for an accurate, high-resolution, easy-to-use load profile to be developed.

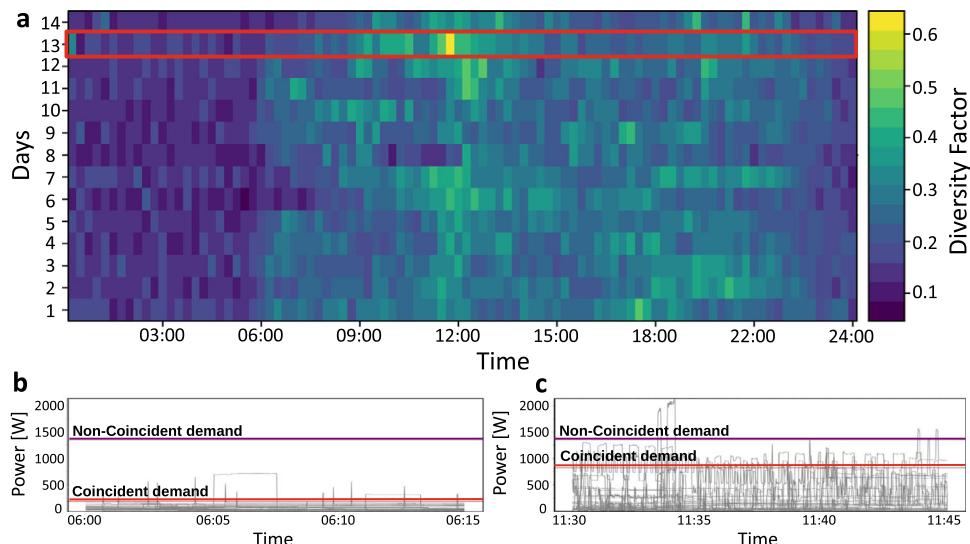
A focus on the higher temporal resolution is necessary to fully understand modern consumption patterns and respond quickly, for instance the disturbances caused by input fluctuations or regulatory or trading anomalies<sup>21</sup>. Recent statistical research on power consumption (see for example, refs. <sup>6</sup> and <sup>7</sup>) demonstrates that substantial differences exist between the statistical features of the highly resolved power consumption and consumption on a 15-min time scale. Analysing power consumption on the short time scale of seconds to one minute reveals extreme consumption spikes, which are completely ignored in the 15 min load profile<sup>6</sup>. In Fig. 1a, b, a comparison of the HO SLP with the load profiles of two residential datasets of high temporal resolution, measured in Germany and Austria is shown. The Austrian dataset was recorded in 2009–2010 during the ADRES project<sup>22</sup>. The German dataset was recorded between 2013 and 2016 during the NOVAREF project<sup>23</sup>. In Supplementary Note 2, we report further on analysis datasets related to 70 households, recorded in August and September 2019 in Germany during the ENERA project.

As can be seen in Fig. 1a, b, the averaged single-day load profiles of both the ADRES and the NOVAREF datasets, strongly deviate from the HO SLP. Specifically, the trend (mean) of the load profile of the measured data is not well described by the HO SLP and the fluctuations on short times scales are significant (see

“Demand trend: mode decomposition”). Furthermore, the analysis of the statistics of the highly resolved power consumption reveals the presence of non-Gaussian, intermittent and asymmetric fluctuations, which need to be taken into account when designing demand side management and control mechanisms, see “Demand fluctuations: stochastic model” and Supplementary Note 4.

As Fig. 1a, b illustrates, the HO SLP data clearly deviates from the consumption pattern indicated by the NOVAREF and ADRES data. There are three main reasons for that. First, due to missing input of microscopic data, the HO SLP is hardly capable of accurately capturing the electricity consumption. Second, the HO SLP is based on data that were measured mostly in the 1960s and 1970s with a small fraction measured towards the end of the 1990s. Since then electricity consumption behaviours and devices have changed drastically. Third, moderate numbers of consumers (e.g., a group of 12 houses considered in this work) do not typically align with an overall average for a larger group of consumers (that might produce a result whose consumption pattern might be closer to the HO SLP).

Some of the observed stochastic properties could be explained by the increased usage of new power generation and new consumption devices, for instance rooftop photovoltaic panels (PVs) and electronic devices such as electrical heating systems, smartphones, tablets, robot vacuum cleaners, electric vehicles etc.<sup>24</sup>. The usage of such devices will influence and alter the load dynamics and in particular its time-dependent average—the load profile. As shown in Fig. 1c, d, the consumption can even reach negative values if houses are directly connected to local RE resources<sup>25</sup>. However, these negative values are less or even reach zero during the winter time because of the reduction of the PV, or the usage of electrical heating. Local fluctuations in solar and



**Fig. 2** Indicating the interaction between households' energy consumption by considering the diversity factor. **a** shows the diversity factor or the ratio between coincident demand,  $P_{cd}$ , and non-coincident demand,  $P_{ncd}$ , of 30 households belonging to the ADRES dataset every 15 min and for 14 days. As it is clear, there are some time intervals during a day whose diversity factor reaches 0.6. **b, c** respectively show the energy consumption trajectory of all 30 households from 6:00 to 6:15 and 11:30 to 11:45. It is clear from the figures that the energy consumption of all houses is low around 6:00 and then is gradually increasing together around 11:00 and, consequently the difference between  $P_{cd}$  and  $P_{ncd}$  is decreasing proving the obvious interaction between houses consuming the electricity.

wind generation may further lead to coincident load profile spikes, both positive and negative. Therefore, precisely analysing and modelling the power consumption of households on a short time scale is essential to ensure the stability of distributed grids. In addition to installing more on-site generations in the residential sector, replacing fossil-fuelled cars with electric ones, which are charged using household electricity, results in changes in the shape of the load profile of the households, especially when organising consumers in microgrids<sup>26</sup>, see also Supplementary Note 2. Since we expect an increasing number of such electrical cars, as well as an increasing penetration of strongly fluctuating RE generating electricity, a new data-driven load profile based on modern and highly resolved measurements is urgently needed.

As mentioned above, extreme consumption spikes are completely ignored in the 15 min load profile as they are averaged out. However, these spikes are of particular importance to lower voltage distribution grids, where coincident consumption can dominate the consumption patterns locally or on a country scale, due to e.g., synchronised activity during major (e.g., sports) events, such as TV pickup in Britain<sup>27</sup> as well as German and U.S. power system frequency variation, respectively, during the 2010 World Cup semi-final game and Super Bowl<sup>28,29</sup>. A well-known simple measure to quantify the coincident electricity consumption between households and moreover to see how this coincidence depends on the number of households is the diversity factor between households<sup>30</sup>. To determine the diversity factor between households, we first evaluate the maximum coincident demand,  $P_{cd}$  and non-coincident demand,  $P_{ncd}$  respectively in 15 min and 1-day time windows. For this purpose, we sum the maximum power demand of all houses every 15 min and then divide the  $P_{cd}$  with the sum of the maximum power demand of all houses over the course of a day, i.e.,  $P_{ncd}$ .

The diversity factor varies from zero to one, where zero indicates no coincident electricity consumption between households, while a diversity factor equal to one shows strong coincidence. As an example, the diversity factor of the ADRES dataset, sampled every 15 min for 14 days, is shown Fig. 2a. To clearly indicate the interaction between houses during a day, (i) the energy consumption trajectory of all 30 households (in grey);

(ii) the maximum coincident demand (in red) and (iii) the non-coincident demand (in purple) between 6:00 to 6:15 (Fig. 2b) and 11:30 to 11:45 (Fig. 2c) for day 13 are shown in Fig. 2b, c.

Looking at the households trajectories, it is clear that the energy consumption of all houses is low at the beginning of the day and then it is gradually increasing until around 11:00. Thus, the value of the diversity factor, which is the ratio between coincident demand ( $P_{cd}$ ), and the non-coincident demand ( $P_{ncd}$ ) increases during daytime. This proves the interaction between the demand of the households during the day and is the reason why the significant spikes in average energy consumption do not disappear after averaging.

To show the relationship between the value of the diversity factor and the number of households, we calculate the diversity factor for the NOVAREF and ENERA datasets, which are composed of 12 and 70 households, respectively, in Supplementary Note 2. Our results demonstrate that the values of the diversity factor are not zero during a day, even for the 70 ENERA houses and for some time intervals the diversity factor ranges between 0.3 and 0.4 or even larger (maximum 0.6). This indicates that regardless of the number of houses coincident demand will take place during certain time intervals (e.g., lunch and dinner time) and spikes during that time will not be averaged out. This will result in a spiky load profile, such as the single-day averaged ENERA load profile, see also Supplementary Note 1.

In the next sections, we introduce a generally applicable methodology for creating residential load profiles and demonstrate its applicability through comparisons with power consumption datasets measured in Germany and Austria, as well as the industry standard load profile. In the next section ("Demand trend: mode decomposition"), we present a methodology to extract the averaged load profile (ALP).

**Demand trend: mode decomposition.** In this section, we present a methodology with which the Averaged Load Profile (ALP) from high temporal resolution datasets is extracted, and the load profile of a full week is created. The latter explicitly takes into consideration not only the differences existing between workdays and

weekends as much as the H0 SLP does, but also the differences between different weekdays unlike the H0 SLP. The main advantage of the ALP compared to the H0 SLP is that it only requires a few weeks of high-resolution measured data and can, therefore, be applied to both small neighbourhoods and whole cities. Furthermore, its only input is the electricity consumption of the houses, unlike most existing models, which rely on numerous micro- or macroscopic parameters. As a result, it can easily be used to analyse both present and future residential power systems, which might include new technologies and devices (see also Supplementary Note 5 for applications on a different time of the year and a different dataset).

Due to the high temporal resolution of the datasets we use, the extracted ALP has a higher level of accuracy than the H0 SLP. It is, therefore, a good alternative to be used instead of the H0 SLP and could provide better and more accurate results, especially when investigating or operating microgrids (a detailed discussion on the subject can be found in “Discussion”). In its present form, the ALP can capture the weekly trend of the load profile of a group of houses (anywhere from 12 to 70 houses have been tested).

To produce the ALP, we first extract the consumption trend from four consecutive weeks (or seven consecutive weekdays or weekends) from high-resolution electricity consumption data measurements, as explained in Supplementary Notes 3 and 5.

In this section, we present our results for the NOVAREF dataset (temporal resolution = 2 s).

The methodology we use to determine the ALP is the following, and it is illustrated in Fig. 3. Using the Empirical Mode Decomposition (EMD)<sup>13</sup>, we split the dataset into multiple modes and, thereby, separate the long-term trends (high-number modes) from the short-term fluctuations (low-number modes). Since the EMD extracts all signals present in the data, the original dataset can be recreated 1-to-1 without any loss of information by summing up all the modes. Next, the data is split into training, validation and test datasets. The training data is used to generate the load trend by summing a certain number of long-term trend modes. During the validation step, the optimal required number of modes is determined to avoid both under- and overfitting. Finally, the extracted ALP is used to compute a test error from the test set, using the optimal number of modes obtained from the previous step. A detailed discussion of the adaptive time-frequency data analysis via the EMD can be found in “Methods”, Supplementary Note 3 and in a recent paper applying EMD in a stochastic context<sup>31</sup>.

Before training, we apply the EMD method on the entire dataset and, thereby, extract 17 individual and independent modes. We then divide these modes into two categories: (a) high-frequency; and (b) low-frequency modes. The high-frequency modes contain information on the stochastic consumption fluctuations (see “Demand fluctuations: stochastic model”), while the low-frequency modes are almost free of stochastic fluctuations and are instead dominated by deterministic effects. To begin the training step, we first choose 4 chronologically consecutive weeks of the measured NOVAREF dataset containing no data gaps, i.e., 01.07–01.14, 01.14–01.21, 02.04–02.11, 02.11–02.18, which were recorded during the winter time between 2013 and 2016. These 4 weeks are the training set. Then, we calculate the average of these four weekly consumption data and, correspondingly, we average the last  $N$  low-frequency modes of these weeks.

In order to get the best performing ALP, we must determine the optimal number of modes  $N$ , i.e.,  $N_{optimal}$ , from the validation set. If we sum too few modes, the extracted ALP fails to track the high-resolution consumption better than the H0 SLP. In contrast, if we sum too many modes it will overfit the dataset and lead later to an inaccurate consumption trend for the test set. To determine the  $N_{optimal}$  and quantify the performance of the optimised ALP, we use the mean-squared error (MSE) on the validation set,

which consists of 9 randomly chosen weeks of the data measured during winter time (see “Methods” for details on MSE and the Supplementary Note 5 for a discussion of a different error metric, i.e., MAE). We must stress here that we used 9 weeks of data because these were the data that were available to us. For the purpose of applying this method on other, shorter, datasets only one additional week of data sound suffice to determine the optimum ALP.

The results of the above analysis determine that  $N_{optimal} \approx 6\dots 8$  for the NOVAREF validation dataset (see the optimal result in Fig. 3b). As seen in Fig. 3b, even for values of  $8 < N < 12$ , the generated ALP outperforms the H0 SLP in terms of demand prediction. Consequently, the ALP determines the daily and weekly electricity consumption trend of the 12 NOVAREF households much more accurately compared to the H0 SLP. However, as mentioned before, for  $N > N_{optimal}$  the ALP overfits the data, and hence later gives an inaccurate results for the test set.

Finally, we use 4 randomly chosen winter weeks to test the extracted ALP. We find that the ALP outperforms the H0 SLP when the  $N_{optimal}$  is chosen, as is evident from its lower average value of MSE, which is reported in the text above the graph in Fig. 3c. It should be noted that the ALP does not perform as well for the smaller or larger number of modes. Figure 3c shows the results for a single day of a week of the test set. We continue the evaluation of the ALP performance in the next section, i.e., “Demand fluctuations: stochastic model”.

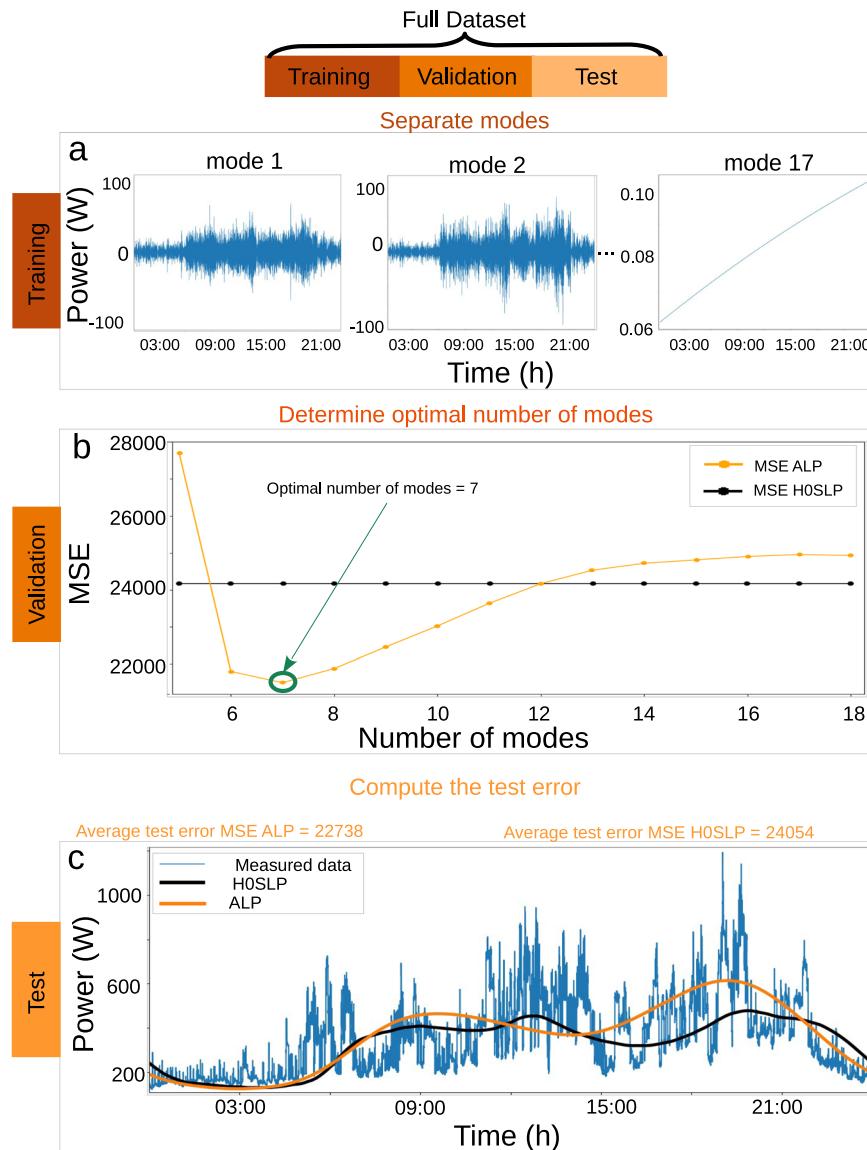
Here, we focus on the analysis of the NOVAREF dataset. We determined the optimum ALP for the winter weeks by training it and determining the related  $N_{optimal}$ . In addition, we investigate the seasonality of the data in more detail in Supplementary Note 5. In particular, we demonstrate that the optimum ALP for the summer weeks still follows the demand trend visible in the winter weeks, and outperforms the H0 SLP in the winter season. We show also the applicability of our methodology by applying it on a UK household dataset in Supplementary Note 5.

The hereby introduced ALP has several major advantages compared to both the H0 SLP and the available demand models. Firstly, it is data-driven and is based on modern, high temporal resolution (2 s) measurements, whereas the H0 SLP is based mostly on hourly resolution data measured before 1999, with a small subset of 15-min resolution data measured between 1990 and 1999<sup>2</sup>. Secondly, because only four chronologically consecutive weeks of data are necessary to generate the weekly ALP, and only a few additional weeks of data are necessary to determine and validate the  $N_{optimal}$  it can be used to extract the averaged load profile of both small and large groups of houses and does not need the plethora of the micro-and-macro-parameters that presently available demand models require. Note that here, we introduce a methodology with which the baseline demand dynamics over the course of a day can be determined from a measured electricity consumption. In “Demand fluctuations: stochastic model”, a stochastic model to generate the stochastic fluctuations on top of this baseline is presented.

**Demand fluctuations: stochastic model.** Having extracted the predominantly deterministic trend of consumption, we now turn to the stochastic fluctuations around this trend. Specifically, we split the power consumption into trend  $P^{(trend)}$  and fluctuations  $P^{(fluc.)}$ :

$$P = P^{(trend)}(t) + P^{(fluc.)}(t). \quad (1)$$

First, we investigate the statistical properties of these consumption fluctuations. Analysing the histograms, we notice that the fluctuations are skewed, i.e., asymmetric, and heavy-tailed, so large deviations are much more likely than if they were characterised by a simple Gaussian distribution, see Fig. 4.



**Fig. 3 Training and validation are used to compute the optimal number of modes, avoiding over- and underfitting.** The full dataset is split into training, validation and test datasets (shown at the top of the Figure). **a** Using the EMD, we decompose the training set into its empirical modes. **b** Next, the validation set is used to optimise the number of modes, avoiding both under- and overfitting. Including too many modes overfits the data but yields poor results on the test datasets. Vice versa, a model with too few modes underfits the data and fails to capture its main trend. Hence, we systematically determine the optimal number of modes by minimising the mean-squared error (MSE). **c** Finally, using the  $N_{optimal}$ , we extract ALP for the test set. Computing the MSE of the ALP obtained for each week demonstrates that the ALP performs better than the HO SLP, since it tracks the measured energy consumption more closely compared to the HO SLP. As shown in panel (c), the average value of the MSE ALP for all weeks is lower than that of the MSE HO SLP. For the purpose of clarity only a single day is shown here (28.05), see also Supplementary Note 5.

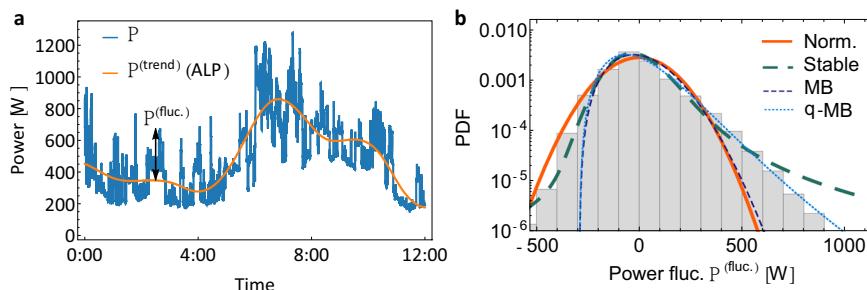
Next, we construct a stochastic model, which describes the observed consumption fluctuation statistics by applying superstatistics methods<sup>21,32–35</sup>. Dividing the full trajectory into several shorter trajectories allows us to characterise each local distribution with a simpler distribution, such as a Gaussian or an exponential distribution. In the case of consumption fluctuations, the fluctuations within each time window of approximately  $T \approx 2000$  s follow a distinct Maxwell–Boltzmann distribution, see Fig. 5. See Supplementary Note 4 for details on the superstatistical procedure, as well as the evaluation of this long-time scale  $T$  (see also Supplementary Fig. 13). Each local Maxwell–Boltzmann distribution has its distinct scale parameter  $\sigma_{MB}$  and offset from zero  $\mu_{MB}$ . When we now move our analysis from one local distribution to the next one, we observe a slow time dynamics of these Maxwell–Boltzmann parameters  $\sigma_{MB}$  and  $\mu_{MB}$  and thereby a

time dynamics of the local Maxwell–Boltzmann distribution itself. Superimposing these time-varying local distributions, we re-obtain the full aggregated statistics as approximately a q-Maxwell–Boltzmann distribution, see Fig. 4.

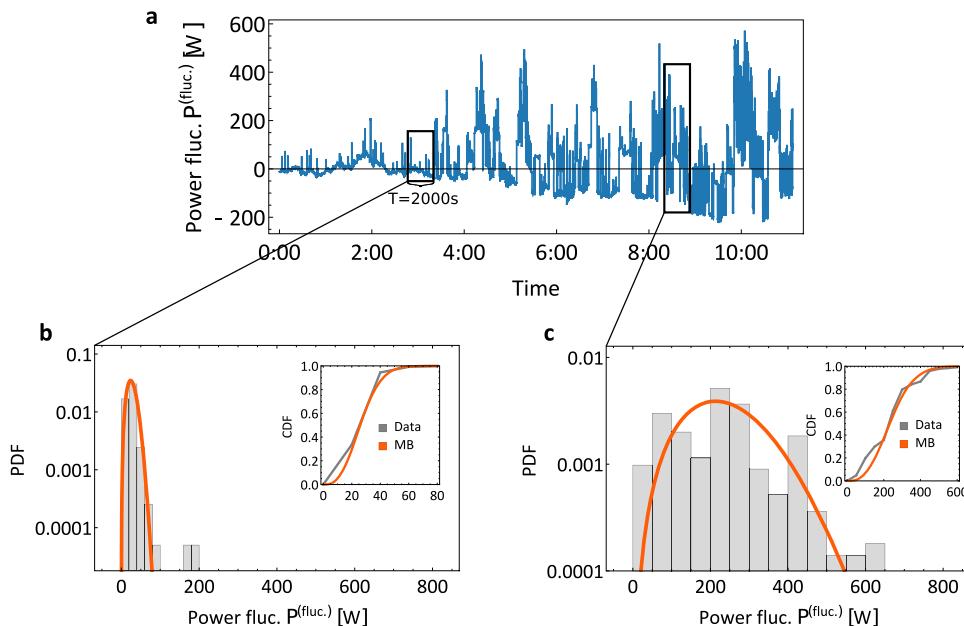
Mathematically, we formulate stochastic equations of motion for the fluctuations leading to local Maxwell–Boltzmann distributions as follows: We define auxiliary variables  $x_i$ , with  $i \in \{1, 2, \dots, J\}$ , each following a simple Ornstein–Uhlenbeck process, based on independent Wiener processes  $W_i$ :

$$dx_i(t) = -\gamma x_i(t)dt + \epsilon dW_i, \quad (2)$$

with damping  $\gamma$  and fluctuations amplitude  $\epsilon$ . Hence, the  $x_i$  are identical but independently distributed Gaussian random variables with a mean 0 and standard deviation  $\sigma = \frac{\epsilon}{\sqrt{2\gamma}}$ . Then, the demand



**Fig. 4 The intermittent characteristic of power consumption fluctuations.** **a** The total power consumption  $P$  is a sum of the trend consumption  $P^{\text{trend}}$ , obtained by the EMD method described in the previous section and fluctuations  $P^{\text{fluc.}}$ . We record the difference between trend and real demand as the fluctuation trajectory. For the purpose of clarity only a single day is shown here. **b** The probability density function (PDF) of the consumption fluctuation does not follow a Gaussian distribution but is better described by a  $q$ -Maxwell-Boltzmann distribution, especially on the right flank. The histogram uses the whole NOVAREF dataset. Alongside the Gaussian, we show a stable and Maxwell-Boltzmann (MB) distribution. The  $q$ -Maxwell-Boltzmann parameters are determined by the methods of moments. See also Supplementary Note 4 for further details.



**Fig. 5 Local Maxwell-Boltzmann distributions of power demand fluctuations.** **a** Using only the high-frequency modes from the empirical mode decomposition (EMD), we plot the consumption fluctuations  $P^{\text{fluc.}}$  over time. **b, c** Using superstatistical methods<sup>32, 33</sup>, we find a long-time scale  $T \approx 2000$  s, see also Supplementary Note 4. On this local time scale  $T$ , we observe local Maxwell-Boltzmann distributions, which can be very narrow (**b**) or broad (**c**). The histograms contain only 1000 (length of  $T$ , given a 2 s resolution) data points of  $P^{\text{fluc.}}$  as local snapshots and we plot MB fits as references. In addition to the PDF, we compare the cumulative probability functions (CDF) of MB distribution with the data in the inset. All plots use the NOVAREF data from a test set in April. The time period and day are chosen for illustration purposes only.

fluctuations  $P^{\text{fluc.}}$  are obtained by aggregating these Gaussian distributions and applying the observed shift from zero  $\mu_{\text{MB}}$ :

$$P^{\text{fluc.}}(t) = \sqrt{(x_1(t))^2 + (x_2(t))^2 + \dots + (x_J(t))^2} + \mu_{\text{MB}}. \quad (3)$$

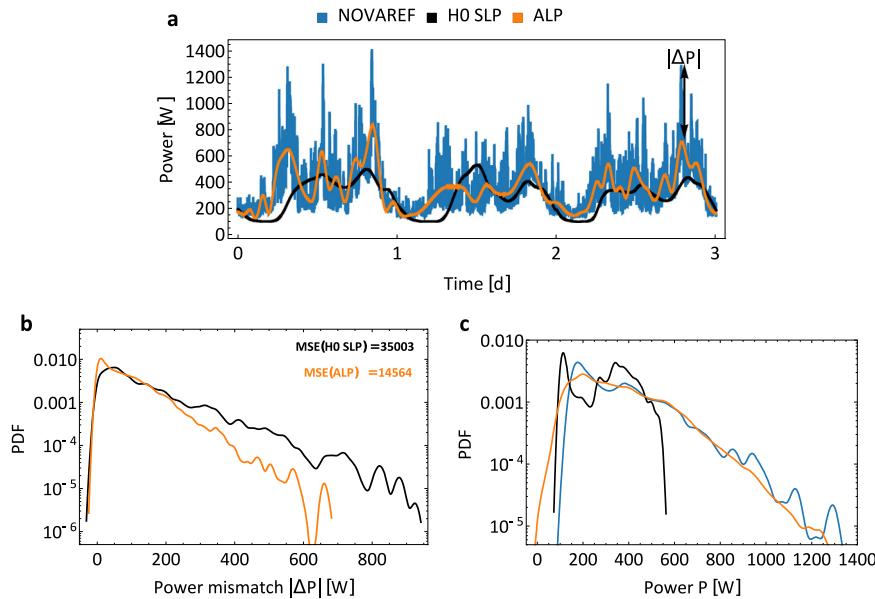
As is known from statistical physics<sup>36</sup>, choosing  $J=3$  yields exactly Maxwell-Boltzmann distributions in the probability density  $p$ :

$$p(P^{\text{fluc.}}) = \frac{1}{\sigma_{\text{MB}}^3} \sqrt{\frac{2}{\pi}} (P^{\text{fluc.}} - \mu_{\text{MB}})^2 \exp\left[-\frac{(P^{\text{fluc.}} - \mu_{\text{MB}})^2}{2\sigma_{\text{MB}}^2}\right], \quad (4)$$

where the shape parameter  $\sigma_{\text{MB}}$  of the local Maxwell-Boltzmann distribution is identical to the standard deviation of the independent Gaussian variables  $\sigma_{\text{MB}} = \sigma$ . We do consider cases of  $J \neq 3$  in Supplementary Note 4 and find that  $J=3$  is the best fit to the data. Indeed, the approach of three independent Gaussian

variables is very convenient for computational application. Alternative modelling approaches using a mathematically simple 1-D process would require more complex dynamics.

In applying the superstatistical approach, we implicitly assumed separation of time scales. Here, we have the long-time scale, on which we locally observe Maxwell-Boltzmann distributions of the power fluctuations of  $T \approx 2000$  s (Fig. 5). Furthermore, we estimate the short time scale, on which each local distribution relaxes to its equilibrium, based on the autocorrelation decay of the data as  $\tau = 1/\gamma \approx 300 \dots 400$  s, see also Supplementary Note 4. Comparing these two time scales, we observe a clear time separation between long-time scales  $T$  and short time scales  $\tau$ , which differ by a factor of 20. Hence, each local Maxwell-Boltzmann distribution relaxes much faster towards its equilibrium (with rate  $1/\tau = \gamma$ ) than the overall process changes towards a new Maxwell-Boltzmann distribution (which happens with a rate of  $1/T$ ).



**Fig. 6 Synthetic power demand in agreement with empirical data.** **a** We display a brief trajectory of the real (blue), HO SLP (black) and the new ALP (orange). The ALP curve overall keeps closer to the real consumption values than the HO SLP does. **b** The histogram of the power mismatch  $\Delta P = |P(\text{real}) - P(\text{model})|$  shows higher deviations of the demand for the HO SLP compared to the ALP. We also report the mean-squared error (MSE) of the HO SLP and the ALP. **c** Combining the trend extraction via EMD to obtain the ALP and the superstatistical model for demand fluctuations (DLP) approximates the real consumption histogram very well, in particular for large consumption values. All trajectories use the same time stamps, starting 15 min past midnight on 15th of April of the NOVAREF data from test set. The histograms return almost identical results for other weeks.

Finally, we combine the EMD-based trend of the demand with the stochastic fluctuation model, obtaining a data-driven load profile (DLP) and, then, compare it to the original NOVAREF consumption data. The model enables us to make some rough general predictions for the near future connected to these training weeks. Long-term forecasting is outside our scope as there is insufficient data available. We notice that while the precise trajectories are not identical (by construction), the stochastic properties align very well with drastically reduced error compared to the standard H0 SLP model (see Fig. 6). In “Methods”, we provide a link to the software and pseudo-code to generate these kinds of trajectories for other demand regions and datasets.

## Discussion

Summarising, we have shown that modern residential electricity load profiles strongly differ from the H0 standard load profile (HO SLP), which is widely used in the industry. We set out to replace or supplement the existing standard load profile and obtained three main results: First, using the empirical mode decomposition (EMD), we develop a model to extract baseline demand dynamics (ALP) over the course of a day. Second, using superstatistics, we identify characteristics of short-time fluctuations around this baseline and, hence, we developed a simple stochastic fluctuation profile (SFP) to describe the non-Gaussian fluctuations of the demand around the trend. Finally, combining both trend and fluctuations yields a full data-driven load profile (DLP).

Hereby, the introduced model specifies a typical range of power consumption at a given time of a day and allows us to study the variation of probabilities of certain extreme power consumption events. Knowledge of the expected demand is critical for energy providers to calculate how much power is needed by each household within a given time period. Simultaneously, knowledge of how much the demand might fluctuate around this trend is also essential, to have sufficient balancing and backup power at hand. According to the Bundesverband der Energie- und Wasserverswirtschaft e. V. (BDEW) (German Association of Energy and

Water Industries)<sup>2</sup>, the electricity load of a household can be expected to deviate between 10 and 20% at any given time. Our fluctuation model provides a better quantitative estimate for these fluctuations to avoid an overestimation of the demand and too much usage of expensive quickly dispatchable generation<sup>37</sup>. On the other hand, we have to prevent an underestimation of the demand as this could easily lead to a collapse of the system.

Another strong point of our modelling approach is its flexibility and broad applicability: We do not present a fixed load profile but a methodology to extract trend (ALP) and fluctuations (SFP) for any present or future power system. This data-driven approach can be applied to different consumer groups, regions or even continents without requiring external or microscopic knowledge. Almost all alternative models require a lot of micro-parameters or fine-grain measurements to achieve the load profile, such as the Markovian model and Hidden Markov Model introduced respectively in refs. <sup>38</sup> and <sup>39</sup>. It is worth to mention that our model is distinct from most machine learning models, as they: (i) Typically require huge training datasets and aim for pattern prediction of data outside the training set and moreover; (ii) do not contribute to the insights our approach offers, e.g., in terms of baseline vs fluctuations (and their relevant time scales). As such our approach is complementary to existing approaches employing machine learning. Maybe even more importantly, the extracted profiles can easily be updated based on recent developments in consumption, e.g., due to additional PV installation or adaptation of electrical cars. Similarly, new load profiles can be created for newly built micro or smart grids<sup>40</sup> or, even existing grids, such as UK. Many existing grids suffer from network congestion and, therefore, either require network reinforcements or operators need to utilise the flexibility of domestic loads to relieve grid load. We demonstrate how our approach is applied to a different group of households located in UK in Supplementary Note 5.

The H0 SLP focused mainly on the time scale of 15 min and reports much smaller fluctuations than our new load profiles. How can we explain this? First, we note that the H0 SLP used

older data, with a temporal resolution much lower than the 1 s of our data sources. More relevant is however the power system perspective: Generation and demand are scheduled for fixed time intervals, such as 15 min and all fluctuations and deviations within each interval are taken care by control mechanisms<sup>41</sup>. This view also emphasises the role of the high-voltage transmission grid where some demand fluctuations, which we observe on the distribution grid, might not occur. Our temporally highly resolved model and focus on the distribution grid become increasingly relevant: Conventional generators and their stabilising inertia are removed from the grid and renewable generators are often directly coupled with households. Hence, the balance of supply and demand has to be present also on the distribution level and on an increasingly fast time scale. Consistently, high-resolution datasets, such as the now-published NOVAREF data are becoming even more important. Any future model, forecast or simulation relies on updated and high-resolution data, even more so when applying data-driven approaches<sup>42</sup> or machine learning<sup>43</sup>. If no such data are available, any prediction will have an increasing error. Both households and the increasing share of renewables introduce fluctuations and without updated highly resolved data, we would face increasing uncertainties on both the expected demand and generation. These uncertainties would likely affect energy prices as well, making them more volatile as predictions of the power balance become less accurate. Finally, we note that not only do both households and renewables introduce fluctuations in the distribution power grid but their fluctuations share similarities. In particular, the heavy tails of consumption fluctuations and its slowly decreasing power spectrum are also observed in renewable generation<sup>10</sup>.

Our model is also especially useful in the case of microgrids which are often powered either partially or fully by renewable resources combined with smart metres. One such renewable resource is photovoltaic panels, which can be installed on rooftops, gardens<sup>44</sup>, walls<sup>45</sup>, walkways<sup>46</sup> and over roads<sup>47</sup>. Two more examples are small vertical wind turbines<sup>48</sup> and bladeless wind turbines<sup>49</sup>, both of which can be mounted on any flat, sturdy surface. They are very well suited for use in residential areas due to their small size and low noise output, especially the newer designs such as Flower Tulips<sup>50</sup> and the Vortex Bladeless Turbine<sup>49</sup>. These renewable energy sources can be utilised not only by households<sup>51</sup> but also by farms, small and large businesses<sup>52</sup> as well as industries<sup>53</sup> since they only require a sturdy flat surface and can be used in a stand-alone mode or in combination with one another<sup>54,55</sup>, and allow them to go partially or entirely off-grid.

In the past, small autonomous grids were comprised mainly of remote, usually rural villages or households powered by fossil fuel generators. In recent years, the concept has been reconfigured to help overcome the challenges of integrating intermittent renewable energy production with the main energy grid<sup>56,57</sup> as well as provide electricity to populations in both remote, island<sup>58</sup> and urban regions as well as business parks, industrial parks and resorts<sup>59</sup> while replacing fossil fuels with renewable resources, in both developed<sup>59</sup> and developing countries<sup>60,61</sup>. Such grids, however, always face the problem of matching the intermittent energy generation with the, as we have shown in “Complex demand dynamics—the necessity of new load profiles”, also the intermittent power consumption of the households that comprise them.

A task that is made harder by the fact that many microgrids are composed of less than the 332 households that the H0 SLP assumes. As shown in “Complex demand dynamics—the necessity of new load profiles” and “Demand trend: mode decomposition”, the consumption behaviour of a small number of houses deviates significantly from the H0 SLP and displays a strong stochastic component. Which technologies can be used to

achieve a successful balancing between the generation and consumption<sup>56,62</sup>, which protection schemes<sup>63</sup> and control systems can be used to ensure the stable and secure operation of these grids<sup>64–66</sup> and even more scheduling and operation issues<sup>57,67,68</sup> are all active areas of research.

Furthermore, our model could allow the administrators of an autonomous or semi-autonomous micro-grid equipped with smart metres to use the collected data to create a general rough prediction of the electricity consumption of the microgrids units in the following weeks and take the appropriate measures to ensure a continuous stable operation, e.g., by ensuring enough balancing power is available. At present, the H0 SLP, which does not differentiate between the electricity consumption of a handful of houses and a small city, makes this impossible. Because our approach is entirely dependent on data with no prior assumptions being made, it could theoretically be used to extract the averaged load profile not only of households but also of small or large-scale of industry or, even, full countries. The current lack of freely available data makes this a task for the future. For example, we wish to apply our model on consumption data of other regions, and compare it to the local equivalent of the H0 SLP. Regions like Canada, where a large number of smart metres are already installed, are particularly promising. While we focused here on household demand, the presented methods should also be applicable to extract the demand trend and fluctuations of industry consumption or combinations of very different consumers. Furthermore, the fluctuation modelling could be extended to include longer-term correlations or match the precise increments in the observed datasets. This would move the study of consumption fluctuations closer to the well-researched state of fluctuations in renewable generation. Given sufficient high-resolution data in the future, other data-driven methods, such as recurrent neural networks, weighted-nearest-neighbour predictors or multilayer perceptrons could be trained on such data. These methods typically require constant data input, while our ALP provides a semi-static prediction. All data used here were taken prior to the COVID-19 pandemic, during which the time people spent at home typically increased substantially. How demand patterns might have changed due to COVID-19 and how well our framework could still capture the dynamics remains an interesting question for future research.

## Methods

**EMD.** The Empirical mode decomposition (EMD) is a well-known method used for nonlinear, non-stationary datasets. It decomposes the data into a finite number of intrinsic mode functions based on the local properties of the data. Therefore, this method is not restricted to linear or stationary time series, as is the case with other methods, such as Fourier spectral analysis. To determine the empirical mode functions, one applies the following steps on a dataset: (1) first the envelope of both local maxima and minima in a dataset are defined separately. For instance, all local maxima are connected by cubic spline lines for the upper envelope and then the same procedure is repeated for the lower envelope. Consequently, all data are confined between upper and lower envelopes. It is worth mentioning that in this method there is no need for the data to have zero crossings, therefore all values can be just positive or negative. (2) Defining the upper and lower envelopes, their mean value,  $m_1$  has to be calculated and then subtracted from the original data, i.e.,  $X(t) - m_1 = h_1$ , where  $X(t)$  is the original data and  $h_1$  is called the first component. (3) In the next step, the first component,  $h_1$  is converted to the intrinsic mode function (IMF), i.e.,  $h_1$  should have the same number of extrema and zero crossings plus the symmetry of upper and lower envelopes around zero (for details, see ref. <sup>13</sup>). After finding the IMF,  $c_1$ , we subtract it from the original data:  $X(t) - c_1 = r_1$ . This  $c_1$  is the first mode and contains the shortest period of the original data  $X(t)$ . (4) Steps (1)–(3) are repeated until  $r_n$  becomes a monotonic function, and it becomes impossible to define any envelop for that. If we sum up all IMFs and the residue, we reproduce again the original data, i.e.

$$X(t) = \sum_{i=1}^n c_i + r_n; \quad (5)$$

Consequently, we can obtain all the intrinsic oscillation modes of the dataset with the EMD method.

**Trend extraction.** In this section, we present the adaptive time-frequency data analysis we used to determine the ALP. It is based on a normalised version of the one-step prediction mean-squared error (MSE)<sup>69</sup> of an estimator.

In order to generate the ALP, we separate firstly four chronologically consecutive weeks of high-resolution electricity consumption dataset with no data gaps as the training set. In this case, weeks 01.07–01.14, 01.14–01.21, 02.04–02.11, 02.11–04.18 of the NOVAREF dataset, which were recorded between 2013 and 2016. For accurate and meaningful results we must use data from the same group of houses whose demand trend is to be determined. The first step in this process is calculating the average of the 4 weeks of data:

$$E_{mean}(t) = \frac{1}{4} \sum_{i=1}^4 E_i(t) \quad (6)$$

where  $E_i(t)$  is the electricity consumption of each week and  $i$  is the number of weeks used in the calculation ( $i = 4$ ). Four weeks are the approximate maximum number of weeks such that by averaging them, the significant spikes occurring during a day will not be missed. Substantially beyond 4 weeks, the extreme values will have been removed by the averaging and adding extra weeks will not make a difference.

For the next step, we apply the EMD on the averaged weekly electricity demand profile  $E_{mean}$ . In the case of the NOVAREF data, 17 individual and independent modes are extracted. Finally to calculate the ALP, we sum the last N low-frequency modes.

$$ALP = \sum_{i=1}^N M_{i+s} \quad (7)$$

where  $M_{i+s}$  is the highest-frequency mode still to be included. In the case of the NOVAREF data, we can have, as an example,  $N = 8$  and  $s = 9$ , hence all modes from  $i + s = 10$  up to  $i + s = 17$  are summed up to obtain the ALP.

To determine the optimal number of low modes, i.e.  $N_{optimal}$  to sum and quantify the performance of the optimised ALP, we apply the mean-squared error (MSE) on the validation set. The MSE measures the average squared difference between the estimated values and the actual values<sup>69</sup> and we calculate it as follows:

$$MSE = \frac{1}{L} \sum_{i=1}^L [P(ALP, t_i) - P(measurement, t_i)]^2 \quad (8)$$

where  $P(ALP, t_i)$  is the ALP obtained from summing up the certain N modes in the training set,  $P(measurement, t_i)$  is the measured consumption time series belonging to the validation set,  $t_i$  is the time in seconds,  $0 < i < L$  and L is the length of data used to normalise the ALP. For the NOVAREF dataset L is 7 days of data in 2 s increments.

The MSE is calculated for both the ALP ( $MSE_{ALP}$ ) and the H0 SLP ( $MSE_{H0SLP}$ ) and compared (see Fig. 3 in “Demand trend: mode decomposition”). The comparison reveals that:

$$MSE_{ALP} < MSE_{H0SLP} \quad (9)$$

for  $5 < N < 12$ . Regardless of the number of modes summed to create the ALP there is always a minimum which fulfils the  $MSE_{ALP} < MSE_{H0SLP}$  condition (see Fig. 3b). This minimum gives the optimal number of low modes, i.e.,  $N_{optimal}$  that must be summed to create the optimal ALP for a given set of chronologically consecutive 4 weeks. This minimum varies with the set of weeks used, though for the NOVAREF dataset  $N_{optimal} = 7$  for the majority of the weeks investigated. A more detailed analysis of the model training, validation and test datasets used, can be found in Supplementary Note 3.

The non-randomness of our results was also verified by calculating the fraction of the  $MSE_{ALP}$  and the  $MSE_{H0SLP}$

$$MSE_{fraction} = \frac{MSE_{ALP}}{MSE_{H0SLP}} \quad (10)$$

for the  $N \approx N_{optimal}$   $MSE_{fraction} < 1$ .

## Data availability

The ADRES household consumption data were provided by the TU Wien and are available for research purposes upon request [https://www.ea.tuwien.ac.at/projects/adres\\_concept/EN/](https://www.ea.tuwien.ac.at/projects/adres_concept/EN/). The IDEAL household dataset is available in <https://datashare.ed.ac.uk/handle/10283/3647?show=full>. The ENERA household consumption data that support the findings of this study were made available by the DLR Institute for Networked Energy Systems and the EWE, respectively. Restrictions apply to their availability as they were used under license for the current study, and so are not publicly available. The data are, however, available from the authors upon reasonable request and with permission of the DLR Institute for Networked Energy Systems and the EWE. The NOVAREF and IDEAL household consumption data that support the findings of this study are available at [https://osf.io/yu2dm/?view\\_only=685370675ca145eb88234031158fc32c](https://osf.io/yu2dm/?view_only=685370675ca145eb88234031158fc32c) for download under the CC-BY-4.0 license.

## Code availability

The code used to calculate the EMD, the ALP and the stochastic fluctuation analysis and generation is available for download on the OSF: [https://osf.io/yu2dm/?view\\_only=685370675ca145eb88234031158fc32c](https://osf.io/yu2dm/?view_only=685370675ca145eb88234031158fc32c), under the CC-BY-4.0 license.

Received: 14 October 2020; Accepted: 11 July 2022;

Published online: 06 August 2022

## References

- Grandjean, A., Adnot, J. & Binet, G. A review and an analysis of the residential electric load curve models. *Renew. and Sustain. Energy Rev.* **16**, 6539–6565 (2012).
- Bitterer, R. & Schieferdecker, B. Repräsentative VDEW-Lastprofil. VDEW Materialien, M-32/99, Frankfurt, (1999).
- E-Control. Sonstige Marktregeln Strom Kapitel 6 Zählwerte, Datenformate und standardisierte Lastprofile. Tech. Rep. Energie-Control Austria für die Regulierung der Elektrizitäts- und Erdgaswirtschaft, Rudolfspl. 13A, 1010 Wien, Austria. <https://www.e-control.at/recht/marktregeln/sonstige-marktregeln-strom> (2019).
- Schilling, S. Final energy consumption by sector and fuel. Tech. Rep., European Environment Agency. <https://www.eea.europa.eu/data-and-maps-indicators/final-energy-consumption-by-sector-9-assessment-4> (2018).
- Monacchi, A., Egarter, D., Elmenreich, W., D'Alessandro, S. & Tonello, A. M. Greend: an energy consumption dataset of households in Italy and Austria. in *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)* 511–516 (IEEE, 2014).
- Wright, A. & Firth, S. The nature of domestic electricity-loads and effects of time averaging on statistics and on-site generation calculations. *Appl. Energy* **84**, 389–403 (2007).
- Marszal-Pomianowska, A., Heiselberg, P. & Larsen, O. K. Household electricity demand profiles—a high-resolution load model to facilitate modelling of energy flexible buildings. *Energy* **103**, 487–501 (2016).
- Liu, J., Krogh, B. H. & Ydstie, B. E. Passivity-based robust control for power systems subject to wind power variability. In *Proceedings of the 2011 American Control Conference*, 4149–4154 (IEEE, 2011).
- Milan, P., Wächter, M. & Peinke, J. Turbulent character of wind energy. *Phys. Rev. Lett.* **110**, 138701 (2013).
- Anvari, M. et al. Short term fluctuations of wind and solar power systems. *New J. Phys.* **18**, 063027 (2016).
- López, M. A., De La Torre, S., Martín, S. & Aguado, J. A. Demand-side management in smart grid operation considering electric vehicles load shifting and vehicle-to-grid support. *Int. J. Electrical Power. Energy Syst.* **64**, 689–698 (2015).
- Logenthiran, T., Srinivasan, D. & Shun, T. Z. Demand side management in smart grid using heuristic optimization. *IEEE Transact. Smart Grid* **3**, 1244–1252 (2012).
- Wu, Z. & Huang, N. E. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Adv. Adaptive Data Anal.* **01**, 1–41 (2009).
- Parti, M. & Parti, C. The total and appliance-specific conditional demand for electricity in the household sector. *Bell J. Econ.* **11**, 309–21 (1980).
- Aigner, D., Sorooshian, C. & Kerwin, P. Conditional demand analysis for estimating residential end-use load profiles. *Energy J.* **5**, 81–97 (1984).
- Capasso, A., Grattieri, W., Lamedica, R. & Prudenzi, A. A bottom-up approach to residential load modeling. *IEEE Transact. Power Syst.* **9**, 957–964 (1994).
- Paatero, J. V. & Lund, D. A model for generating household electricity load profiles. *Int. J. Energy Res.* **30**, 273–290 (2006).
- Train, K., Herriges, J. & Windle, R. Statistically adjusted engineering models of end use load curves. *Energy* **10**, 1103–11 (1985).
- Esther, B. P. & Kumar, K. S. A survey on residential demand side management architecture, approaches, optimization models and methods. *Renew. Sustain. Energy Rev.* **59**, 342–351 (2016).
- Proedrou, E. A comprehensive review of residential electricity load profile models. *IEEE Access* **9**, 12114–12133 (2021).
- Schäfer, B., Beck, C., Aihara, K., Withaut, D. & Timme, M. Non-gaussian power grid frequency fluctuations characterized by lévy-stable laws and superstatistics. *Nat. Energy* **3**, 119 (2018).
- Einfalt, A. et al. ENERGIE DER ZUKUNFT Publizierbarer Endbericht, ADRES-Concept. Tech. Rep., TU Wien. [https://www.ea.tuwien.ac.at/fileadmin/te/ea/projekte/ADRES\\_Concept/PublizierbarerEndberichtADRES\\_815674.pdf](https://www.ea.tuwien.ac.at/fileadmin/te/ea/projekte/ADRES_Concept/PublizierbarerEndberichtADRES_815674.pdf) (2012).
- Lange, M. & Zobel, M. NOVAREF, Erstellung neuer Referenzlastprofile zur Auslegung, Dimensionierung und Wirtschaftlichkeitsberechnung von Hausenergieversorgungssystemen. Tech. Rep. (NEXT)ENERGY, Carl-von-Ossietzky-Strasse 15, 26129, Oldenburg, 2016).
- (Destatis), S. B. Wirtschaftsrechnungen einkommens- und verbrauchsstichprobe ausstattung privater haushalte mit ausgewählten gebrauchsgütern und versicherungen. Tech. Rep., Statistisches Bundesamt (Destatis). [https://www.destatis.de/GPStatistik/servlets/MCRFileNodeServlet/DEHeft\\_derivate\\_00027268/2152603139004.pdf](https://www.destatis.de/GPStatistik/servlets/MCRFileNodeServlet/DEHeft_derivate_00027268/2152603139004.pdf) (2018).

25. Braun, M., Büdenbender, K., Magnor, D. & Jossen, A. Photovoltaic self-consumption in Germany—using lithium-ion storage to increase self-consumed photovoltaic energy. In *24th European Photovoltaic Solar Energy Conference, 21–25 September 2009, Hamburg, Germany*, 3121–3127 (24th European photovoltaic solar energy conference (PVSEC), 2009).
26. Muratori, M. Impact of uncoordinated plug-in electric vehicle charging on residential power demand. *Nat. Energy* **3**, 193–201 (2018).
27. News, B. Can you have a big 'switch off'? [http://news.bbc.co.uk/2/hi/uk\\_news/magazine/6981356.stm](http://news.bbc.co.uk/2/hi/uk_news/magazine/6981356.stm) (2007).
28. Chen, L., Markham, P., Chen, C.-f. & Liu, Y. Analysis of societal event impacts on the power system frequency using fnet measurements. in *2011 IEEE Power and Energy Society General Meeting*, 1–8 (IEEE, 2011).
29. Chester Energy and Policy. Super bowl sunday and electricity demand: what happens in cities with super bowl teams and host cities? <http://chesterenergyandpolicy.com/2018/01/31/super-bowl-sunday-and-electricity-demand-what-happens-in-cities-with-super-bowl-teams-and-host-cities/> (2018).
30. Kersting, W. H. *Distribution System Modeling and Analysis* (CRC Press, 2007).
31. Kampers, G. et al. Disentangling stochastic signals superposed on short localized oscillations. *Phys Lett. A* **384**, 126307 (2020).
32. Beck, C. & Cohen, E. G. Superstatistics. *Physica A: Stat. Mechan. Appl.* **322**, 267–275 (2003).
33. Beck, C., Cohen, E. G. & Swinney, H. L. From time series to superstatistics. *Phys. Rev. E* **72**, 056133 (2005).
34. Weber, J. et al. Wind power persistence characterized by superstatistics. *Sci. Rep.* **9**, 1–15 (2019).
35. Williams, G., Schäfer, B. & Beck, C. Superstatistical approach to air pollution statistics. *Phys. Rev. Res.* **2**, 013019 (2020).
36. Laurendeau, N. M. *Statistical Thermodynamics: Fundamentals and Applications* (Cambridge University Press, 2005).
37. Wood, A. J., Wollenberg, B. F. & Sheble, G. B. *Power Generation, Operation, and Control* (John Wiley & Sons, 2013).
38. Ardakanian, O., Keshav, S. & Rosenberg, C. Markovian models for home electricity consumption. in *Proceedings of the 2nd ACM SIGCOMM workshop on Green networking*, 31–36 (ACM, 2011).
39. Hemmatifar, A. & Mogadali, M. Household energy disaggregation based on difference hidden Markov model. <http://cs229.stanford.edu/proj2016spr/report/067.pdf> (2016).
40. Fang, X., Misra, S., Xue, G. & Yang, D. Smart grid—the new and improved power grid: a survey. *IEEE Commun. Surv. Tutor.* **14**, 944–980 (2011).
41. Machowski, J., Bialek, J. W. & Bumby, J. *Power System Dynamics: Stability and Control* (John Wiley & Sons, 2011).
42. Gorjão, L. R. et al. Open database analysis of scaling and spatio-temporal properties of power grid frequencies. *Nat. Commun.* **11**, 1–11 (2020).
43. Kruse, J., Schäfer, B. & Witthaut, D. Predictability of power grid frequency. *IEEE Access* **8**, 149435–149446 (2020).
44. Spirited energy. <https://www.spiritedenergy.co.uk/kb-pv-ground-mounted-solar-panels> (2020).
45. Solstex, by elemex®architectural facade systems. <https://elemex.com/news/solar-panels-move-off-the-roof-on-to-exterior-walls-with-solstex/> (2020).
46. Platio solar pavement. <https://platirosolar.com/> (2020).
47. Road-integrated photovoltaics. <https://www.ise.fraunhofer.de/en/key-topics/integrated-photovoltaics/road-integrated-photovoltaics-ripv.html> (2020).
48. Adefarati, T. & Bansal, R. Chapter 2 – energizing renewable energy systems and distribution generation. in *Pathways to a Smarter Power System* (eds Tasciakaraoglu, A. & Erdinç, O.) 29–65 (Academic Press, 2019).
49. Vortex bladeless sl. <https://vortexbladeless.com/technology-design/> (2020).
50. Flower turbines. <https://flowerturbines.com/> (2020).
51. Begeno, bürger energie bremen. <https://flowerturbines.com/https://www.begeno.de/buergerstrombremen/%C3%B6kostrom-f%C3%BCr-ihren-haushalt/> (2020).
52. Begeno, bürger energie bremen. <https://www.begeno.de/buergerstrombremen/%C3%B6kostrom-f%C3%BCr-gewerbekunden-100000kwh-2/> (2020).
53. Begeno, bürger energie bremen. <https://www.begeno.de/buergerstrombremen/%C3%B6kostrom-f%C3%BCr-gewerbekunden-100000-kwh/> (2020).
54. Stroe, D.-I., Zaharoff, A. & Iov, F. Power and energy management with battery storage for a hybrid residential PV-wind system—a case study for Denmark. *Energy Procedia* **155**, 464–477 (2018).
55. Prashanth, B., Pramod, R. & Kumar, G. V. Design and development of hybrid wind and solar energy system for power generation. *Mater. Today: Proc.* **5**, 11415–11422 (2018).
56. Olivares, D. E. et al. Trends in microgrid control. *IEEE Transactions on Smart Grid* **5**, 1905–1919 (2014).
57. Lasseter, R. H. Microgrids. in *2002 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No.02CH37309)* Vol. 1, 305–308 (IEEE, 2002).
58. Renewable microgrids: Profiles from islands and remote communities across the globe. [https://rmi.org/wp-content/uploads/2017/04/Islands\\_Microgrid\\_Profiles\\_Islands\\_Global\\_Remote\\_Communities\\_CaseStudy\\_2015.pdf](https://rmi.org/wp-content/uploads/2017/04/Islands_Microgrid_Profiles_Islands_Global_Remote_Communities_CaseStudy_2015.pdf) (2020).
59. What are renewable energy communities and how will they grow? <https://microgridknowledge.com/renewable-energy-communities/> (2020).
60. The rise of microgrids in developing countries. <https://hivepower.tech/the-rise-of-microgrids-in-developing-countries/> (2020).
61. 21 intriguing microgrid projects to watch in 2021. <https://microgridknowledge.com/microgrid-projects-to-watch-in-2021/> (2020).
62. Katiraei, F., Iravani, R., Hatzigaryiou, N. & Dimeas, A. Microgrids management. *IEEE Power Energy Magazine* **6**, 54–65 (2008).
63. Sortomme, E., Venkata, S. S. & Mitra, J. Microgrid protection using communication-assisted digital relays. *IEEE Transact. Power Deliv.* **25**, 2789–2796 (2010).
64. Farrokhabadi, M., Cañizares, C. A. & Bhattacharya, K. Frequency control in isolated/islanded microgrids through voltage regulation. *IEEE Transact. Smart Grid* **8**, 1185–1194 (2017).
65. Delille, G., Francois, B. & Malarange, G. Dynamic frequency control support by energy storage to reduce the impact of wind and solar generation on isolated power system's inertia. *IEEE Transact. Sustain. Energy* **3**, 931–939 (2012).
66. Katiraei, F., Iravani, M. R. & Lehn, P. W. Micro-grid autonomous operation during and subsequent to islanding process. *IEEE Transact. Power Deliv.* **20**, 248–257 (2005).
67. Hirsch, A., Parag, Y. & Guerrero, J. Microgrids: a review of technologies, key drivers, and outstanding issues. *Renew. Sustain. Energy Rev.* **90**, 402–411 (2018).
68. Mumtaz, F. & Bayram, I. S. Planning, operation, and protection of microgrids: an overview. *Energy Procedia* **107**, 94–100 (2017).
69. Schelter, B., Winterhalder, M. & Timmer, J. Local and cluster weighted modeling for time series prediction. In *Handbook of Time Series Analysis:Recent Theoretical Developments and Applications*, (eds David, E. & Ulrich, P.) 39–65 (Wiley Online Library, 2006).
70. Household data. [https://data.open-power-system-data.org/household\\_data/latest/](https://data.open-power-system-data.org/household_data/latest/) (2020).

## Acknowledgements

The authors would like to thank the NOVAREF, ADRES and ENERA projects for the data that they provided, without which this paper would not have been possible. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 840825, the German Ministry for Education and Research (BMBF) under grant no. 03ET4027A for the DYNAMOS project and no. 03EF3055F for the CoNDyNet2 project and the Helmholtz Association under grant no. VH-NG-1727.

## Author contributions

M.A., E.P., B.S., C.B., H.K. and M.T. conceived and designed the research. E.P. and M.A. detrended the data, E.P. developed the deterministic model and B.S., C.B. developed the stochastic model. All authors contributed to discussing and interpreting the results and writing the manuscript.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-31942-9>.

**Correspondence** and requests for materials should be addressed to Mehrnaz Anvari or Marc Timme.

**Peer review information** *Nature Communications* thanks Ralf Metzler, I. Safak Bayram and the other anonymous reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022