Deep and Efficient Impact Models for Edge Characterization and Control of Energy Events

Grigore Stamatescu[1,2], Rahim Entezari[1,2], Kay Römer[1], Olga Saukh[1,2]
[1]*Institute of Technical Informatics, Graz University of Technology, Austria*
[2]*Complexity Science Hub Vienna, Austria*
{gstamatescu, entezari, roemer, saukh}@tugraz.at

***Abstract*—Network control in microgrids is an active research area driven by a steady increase in energy demand, the necessity to minimize the environmental footprint, yet achieve socio-economic benefits and ensure sustainability. Reducing deviation of the predicted energy consumption from the actual one, softening peaks in demand and filling in the troughs, especially at times when power is more affordable and clean, present challenges for the demand-side response. In this paper, we present a *hierarchical energy system architecture* with embedded control. This architecture pushes prediction models to edge devices and executes *local control loops* to address the challenge of managing demand-side response *locally*. We employ a two-step approach: At an upper level of hierarchy, we adopt a conventional machine learning pipeline to build load prediction models using automated domain-specific feature extraction and selection. Given historical data, these models are then used to label prediction failure events that force the operator to use backup energy sources to stabilize the network. On a lower level of hierarchy, computed labels are used to train *impact models* realized by LSTM networks running on edge devices to infer the probability that the power consumption of the player contributes to the upper level prediction failure event. The system is evaluated on clustered and aggregated energy traces from a public data set of academic buildings. The results show the benefits of the proposed hierarchical energy system architecture in terms of impact prediction with 55 % accuracy. This allows minimizing the number of prediction failure events by 11.69 % by executing targeted local control.**

***Index Terms*—hierarchical energy system, local control, energy prediction, from local to global, impact models, smart buildings**

I. Introduction

The effective management of large-scale energy systems offers a potential for outreaching economic, social and environmental impact. Smart buildings are becoming key players in such energy management strategies. Due to the increasing instrumentation of buildings with networked sensors, we now have access to rich data traces of their operation. Forecasting various parameters of these operational traces is of paramount importance for the reliable operation of the infrastructure and its efficient integration with the grid. Current research related to the electrical energy usage of buildings falls into one of two categories. The first research direction focuses on the *short-term load forecasting* (STLF) problem over hourly, daily and weekly time horizons and its impact on balancing supply and demand [1]. The second research line focuses on *anomaly identification* and early warning systems based on energy traces. Anomalies are usually understood in terms of conspicuous energy traces produced by misbehaving consumers or energy equipment [2]. However, neither research direction alone helps to manage demand-side response and minimize the gap between the predicted energy consumption and the real data. The main reason for this is that both topics reside at different levels of data aggregation: An energy trace of a broken appliance is masked by the usage patterns of other devices. We tackle the problem by constructing a novel model,—referred to as *impact model*,—to predict the contribution of an *individual behavior* to the *aggregated behavior*. This approach is a prerequisite for a targeted *local control* to soften energy peaks and fill in the troughs in order to support the demand-size response at a new level.

Predicting energy consumption is an active field of research boosted by the availability of good-quality public data sets of building data, including detailed electrical energy consumption traces. A common pipeline for building prediction models involves preprocessing, feature extraction, feature selection, model training and evaluation [3]. In this work we train typical machine learning models described in the literature [4], such as regression models [5] and tree-based algorithms [6], to predict the energy consumption at a building level. These have been shown to offer a good trade-off between forecasting accuracy, ease of implementation, training, and explainability to end-users [7]. We adapt a more general definition of *anomaly* as an *inconsistency between the actual and the predicted energy consumptions*. If the inconsistency is significant, the network operator will have to stabilize the network by activating usually expensive energy reserves. Given a good-quality prediction model, anomalies can be described as a *stochastic process*. Anomalies may be caused by transient equipment faults and malfunctions, or persistent changes in energy usage behavior influenced by external factors such as weather and seasonality [8]. Despite stochasticity of anomaly events which makes them impossible to predict better than by chance, we focus on modelling the dependency between the local and the aggregated energy traces at consecutive levels of hierarchy by means of LSTM networks. Such impact models allow locally managing energy consumption to minimize the chance of anomaly occurrence.

Although a classical grid is often seen as a hierarchical structure, the grid operator makes decisions on its aggregated view. In this work, we introduce intermediate levels of hierarchy at the level of individual buildings, campuses and districts. We refer to our grid architecture as a *hierarchical energy system*. We require an energy consumption measurement device, such as a smart meter, at every level of hierarchy capable of running an energy prediction algorithm. Moreover, every device estimates the impact of its local subnetwork to the aggregation trace at the upper level. If the impact model predicts that the local behavior may cause an anomaly, the device may apply *local control* to mitigate the risk of anomaly occurrence. Examples of local control include lowering the temperature in a certain part of the building by a few degrees, turning off auxiliary capacities, and rescheduling non-critical tasks. Since demand-side response management occurs in real time, it is essential that all model inferences and local control loops are executed *on the edge in real time*.

The contributions of this paper are summarized as follows:

- We propose a concept of a hierarchical energy system with local control, which pushed energy prediction models and control to edge devices. We describe the overall architecture in Sec. II.
- Sec. III and Sec. IV introduce a generalized view of anomalies and describe the structure of the LSTM impact models that predict the contribution of the local behavior to the aggregated one.
- Sec. V presents evaluation results of the model performance on a public data set comprising one year of energy traces from academic buildings. We show that the hierarchical energy system with local control allows minimizing the number of energy anomalies at the building cluster level by by 11.69 % and thus considerably improves the demand-side response management.

Finally, Sec. VI surveys the state-of-the-art and Sec. VII concludes this paper.

II. Hierarchical Energy System with Local Control

Large-scale energy systems are typically organised in hierarchical layers of monitoring and control. In a reference three level architecture presented in Fig. 1, a central coordinator at the top level is responsible for balancing electricity generation and consumption within a high-level planning framework. The intermediate level coordinates locally interacting major entities such as large energy consumers and distributed renewable energy generation facilities that operate at reduced reaction times. Finally, low-level commercial entities, such as buildings or small-scale industrial or manufacturing facilities, generally manage their own consumption through a mix on energy incentives and penalties corresponding to individually agreed contracts with the energy supplier. The contracts regulate energy pricing, penalties for overconsumption and static caps on the total amount of energy that can be drawn from the grid. Incentives can also be provided to large consumers for drawing surplus energy from the grid during certain time periods.

We adopt this grid structure, yet argue that a hierarchical energy system needs a *local energy forecasting and local control* through intelligent energy management running on embedded devices on the edge. The case study is based on

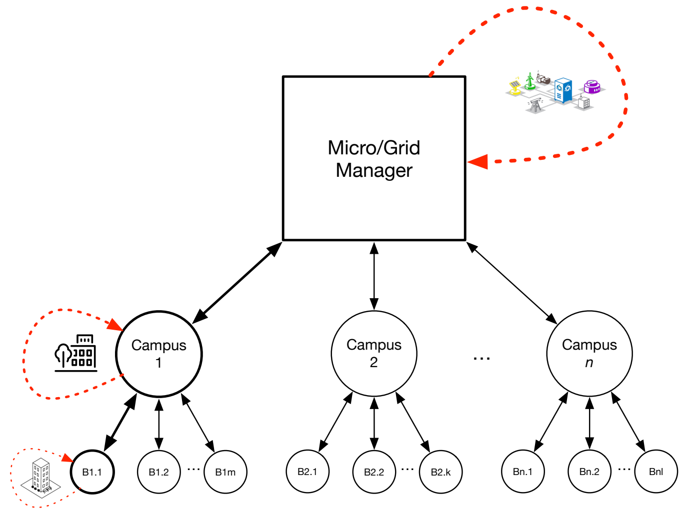

Fig. 1: Hierarchical energy system with three layers

data traces from individual consumers, *e.g.*, large commercial buildings, grouped into clusters, *e.g.*, neighborhood and campus energy management units, that are coordinated by a top-level grid dispatcher entity. Deploying prediction models and running inference on the edge can improve local control loops, minimize reaction times and increase robustness of the overall system. Upstream aggregation of the measured and predicted energy time series results in the information loss by obfuscating small-scale energy events. Also sampling and control periods for predicting electricity demand and classifying events increase as we move up in the hierarchy: from tens of seconds or minutes on the edge to tens of minutes and hours. By building prediction models from energy traces, the resulting anomalies,–defined as inconsistent differences between actual and expected values,–can be swiftly propagated toward the decision center and included into the large-scale and slow control loops as perturbation channels. This will improve the energy management of the whole microgrid, reduce energy costs, and minimize the unnecessary provisioning of environment-damaging energy resources.

Our case study and evaluation is focused on two bottom levels of the diagram depicted in Fig. 1 that cover the individual consumer entities, medium-to-large commercial buildings with over 5000 sqm of floor space and widely varying usage types, grouped together and coordinated by an energy manager at a cluster level. The main idea explored in this paper is twofold: 1) we argue that each building is capable of locally estimating the extent to which its own consumption contributes to the known anomalies reported on the aggregated data in the corresponding cluster, and 2) we can take a local remedial action to curtail positive or negative impact without outside influence. Control actions can be taken for example by modulating large local consumers of electricity. Conversely, negative anomalies can be compensated by storing electricity in on-site storage systems or building thermal buffers through the HVAC system [9]. We assume that the number of consumers in a cluster is reasonably low so that anomalies can be evaluated by the members with sufficiently high confidence levels.

III. Predicting Energy Consumption

The first stage in the process consists of building *short-term load forecasting* (STLF) models using a typical conventional machine learning pipeline. The steps involve pre-processing of individual time-series, feature extraction, feature selection, model training and model selection. For the purpose of this study, the offline procedure is carried out on a reference building energy data set acquired in the Building Data Genome project [10]. The data represents hourly whole-building energy meter measurements from 507 mostly academic buildings from various climate areas. The data set is particularly suitable for testing the hierarchical energy system since it allows for a straightforward and natural clustering of co-located energy consumers, *e.g.*, buildings in the same campus. We map the clusters to the two bottom layers of our hierarchical energy aggregation architecture presented in Fig. 1.

We automate the feature extraction and selection steps by leveraging a new time series modelling toolbox `tsfresh` [11], with wide applicability across domains. It is successfully used in industrial applications such as machinery lifespan estimation and product quality forecasting, which we currently extend to energy pattern analysis and forecasting. Moreover, it is suitable for the analysis of data streams and handling time series in batches with the output features being fed to other frameworks such as `scikit-learn` and `tensorflow`. Several hundred features are automatically computed, sorted by their relevance and discarded by means of the statistical significance test which evaluates their contribution to the quality of the prediction model. Feature extraction and selection on the full data set took 25 h on Intel Xeon server-class system[1].

Multiple machine learning models have recently been used for prediction in various energy related tasks. Among these we have selected a subset of models that we apply to our initial building energy load forecasting. Selected features were thus used to train the following models: linear regression with lasso parameter regularization [5], linear regression with ridge regularization [12] and boosted regression trees (`xgboost`) [6]. Plain linear regression (LR) is also considered, mostly in order to quantify the performance of the other models to a fixed baseline. Prediction accuracy is evaluated in terms of the Mean Absolute Prediction Error (MAPE), a relative metric which enables comparison across various levels of absolute energy consumption. MAPE is defined as:

$$MAPE = \frac{1}{n}\sum_{1}^{n}\left|\frac{Y_t - Yp_t}{Y_t}\right| \cdot 100\%, \tag{1}$$

where n represents the number of samples, Y_t and Yp_t stand for the actual data and predicted data, respectively. We achieve MAPE values between 0.5% and 4.5% across the modelled data set, depending on the specific building electricity consumption patterns, energy efficiency, location and local weather.

[1]We used: `tsfresh` 0.11.2, `scikit-learn` 0.20.2, `python` 2.7, `keras` 2.2.4 with `tensorflow` 1.13.1.

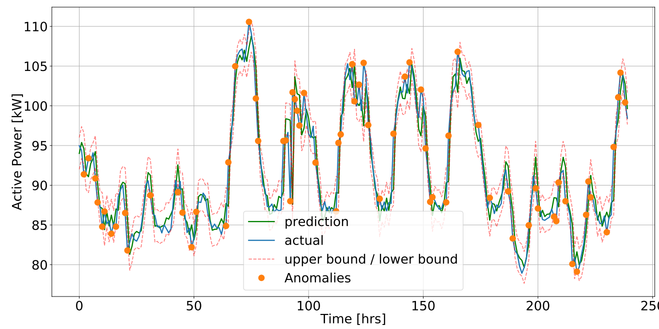

Fig. 2: Building energy prediction and anomaly labelling using machine learning models

Once individual prediction models are validated, we implement a dynamic thresholding mechanism to compare predicted values to the actual consumption trace. We label all threshold violations as anomaly events. This method of finding anomalies is justified by the limited availability of manual labels by human experts and is used when building many domain-specific ML models [8]. To implement dynamic thresholding, we first compute a cross validation score for each time series using the *negative mean absolute error* (MAE):

$$MAE = \frac{1}{n}\sum_{1}^{n}\left|Y_t - Yp_t\right|. \tag{2}$$

Specificly to time series data, cross-validation is performed on a rolling basis with new validation folds being built as supersets of the previous folds in order to preserve the structure of the time series fragments. The `tscv` package is used for implementation. The upper (UB) and lower (LB) bounds are finally computed as follows:

$$\begin{aligned} UB &= q + (MAE + 1.96 \cdot \sigma) \\ LB &= q - (MAE + 1.96 \cdot \sigma), \end{aligned} \tag{3}$$

where q is a predicted energy consumption. The value 1.96 in the above formulas serves as a scaling factor and corresponds to the 95 % confidence interval in our implementation. The value can be tuned to make anomaly labelling more conservative or more relaxed. To ensure data consistency when building impact models, the linear regression with lasso parameter shrinkage is used as a reference when labelling anomalies across all individual and aggregated time series predictions.

Fig. 2 illustrates the outcome of the energy prediction step for a sample building in the data set. The graphic shows the actual values, the predicted values and the confidence intervals over a sample of ten consecutive days. The MAPE metric in this particular case is 1.88 % over the test set. The anomalies are marked with orange dots and their indices and relative magnitudes are recored for further use. Overall, around 15 % to 20 % of the data points are labelled as anomalies with the current parametrisation of the models.

Fig. 3 shows the contribution of the extracted features to the prediction model output. In this specific case, the top features that contribute to the prediction output are related to the Fast

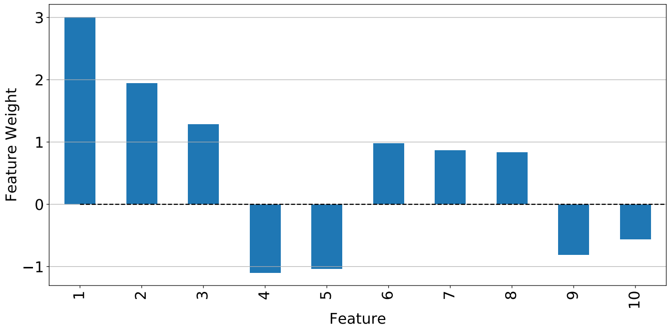

Fig. 3: Model features ranked by their respective contributions (Features: 1 = 'cwt_coeff_widths_(2,5,10,20)_coeff_9_w_2', 2 = 'quantile_q_0.9', 3 = ' mean_second_derivative_central', 4 = ' fft_coeff_4_attr_real', 5 = 'fft_coeff_3_attr_real', 6 = 'minimum', 7 = 'fft_coeff_2_attr_imag', 8 = 'quantile_q_0.4', 9 = 'fft_coeff_5_attr_real', 10 = 'f_agg_mean_isAbs_true')

Fourier Transform (FFT) and Continuous Wavelet Transform (CWT) coefficients of time series fragments as well as to basic metrics such as median values. This type of analysis helps to reduce the feature space for embedded implementation as well as to understand the determining factors in the forecast.

IV. Event Prediction at the Edge

At the lower level of hierarchy in Fig. 1, we use the previously extracted anomaly labels to define a supervised learning task of anomaly impact classification using deep neural networks. The *impact models* run by each building and output a binary indicator *if a consumer at the lower level is the cause of the anomaly in the aggregated trace at the upper level*. If this is the case, it will have to adjust its consumption pattern in the first place by taking *local control* actions. Given the input time dependencies and sequence characteristics a LSTM model was chosen and its parameters were selected in terms of the number of layers, number of units per layer, activation functions and optimisation method. The model takes the latest 24 h of data which led to an event, *i.e.*, 24 time steps for both individual consumption and aggregated consumption at the corresponding upper layer. The anomaly labels have been previously calculated according to a prediction model as described in Sec. III. We first compute overshoots and undershoots of the prediction inconsistency in the aggregated trace, and then assign the contribution of individual buildings to the identified anomalies as a percentage value. The building with maximum impact is assigned 1, while other buildings are assigned 0. This is done while accounting for the fact that some buildings may actually contribute to the mitigation of an anomaly by underconsuming at peak times. In this case, we ignore the deviations that have a different sign than the sum in the calculation of the contributions. One limitation of this work is that we treat both positive and negative deviations in a similar manner, whereas in a realistic scenario negative deviations would result in an incentive towards the building to actually consume more.

Fig. 4 illustrates the high-level input-output structure of the RNN LSTM model that we use at the individual building level. The implementation is based on the 'Vanilla LSTM'

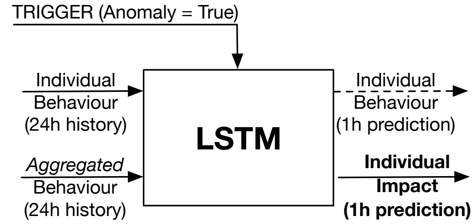

Fig. 4: Architecture of the anomaly impact model (LSTM)

[13] unit as building blocks. For each cell, a time series $\{x\}$ of length $n \in \mathbb{N}$ is processed by a single LSTM layer and yields two outputs: a hidden state h and a cell state c. The initial state of the network (c_0, h_0) is input to the first cell at initialisation along with the first time step of the sequence x_1. The first output h_1 is computed and the new cell state c_1 is propagated to the next computational cell in an iterative fashion. The output state at time step t is achieved by means of combining the current output of the LSTM layer with the cell state information that accounts for the previously extracted information [14].

There are four gate types in an LSTM cell: the input gate i which controls the level of cell state update, the layer update g which controls the added information to the cell state, the forget gate f which controls the removal of information from the cell state, and the output gate o which control the effect that the cell state has on the output. During training an LSTM network, the conventional NN parameters input weights $\mathbf{W}$, and the bias $\mathbf{b}$ are learned along with an extra set of parameters W, R and b in the form of the recurrent weights R. This set of parameters is achieved through the concatenations of input weights, recurrent weights, and biases at the component-wise level. The value of the LSTM network thus comes from its ability to propagate information towards the output over longer time sequences, while offering the internal mechanisms to prioritise certain patterns during training.

The models are triggered and run whenever an anomaly is reported by the upper level of hierarchy. In this case, the local LSTM model is provided with the aggregated energy consumption trace from the upper level. The history of the own individual energy consumption is available locally. Optionally, the model can also predict the own individual energy consumption for the next hour. The prediction can be used to plan local control.

As an example, Fig. 5 illustrates the individual and cumulative active power draw over a ten day period from three reference buildings 1-3, aggregated into a cluster 1, also see Fig. 1. For consistency of timestamps, construction specifics and climate influences, all buildings stem from the same academic campus located in Zurich, Switzerland. The potential for improving energy management and control resides in the differences between the low-level consumers managed at the cluster level. The usage patterns across the buildings are heterogeneous since these include offices, classrooms and laboratory spaces. They can be identified in their source data set with their textual descriptions: 'Office_Travis', 'UnivClass_Terri' and 'UnivLab_Tracy' respectively. The aggregated trace in Fig. 5 represents the total energy consumption of three buildings. The floor area is also reported as being 9244 sqm for Building 1, 12796 sqm for Building 2 and 5968 sqm for Building 3, which allows further energy efficiency profiling.

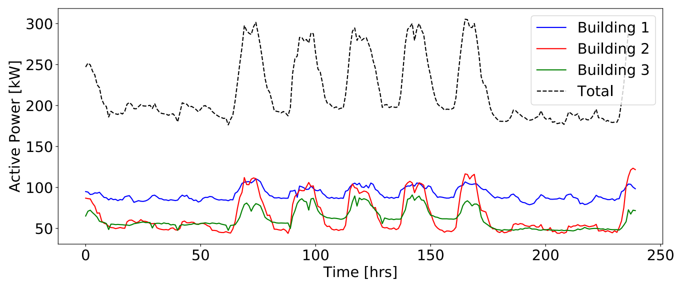

Fig. 5: Individual and aggregated traces of 3 campus buildings

Table I shows an anomaly impact results on a sample campus comprising three buildings. In this example, Building 1 is the dominant contributor to the upper-level anomalies in the aggregated trace.

Anomaly ID	Building 1	Building 2	Building 3
1	38	36	26
2	70	30	0
3	59	25	16

TABLE I: Impact of a building [%] on the reported anomalies

A slice of a training time series for the LSTM for Building 1 is presented in Table II.

Input Individual Behaviour	[90,211 89,3 ... 90,125 90,48]
Input Aggregated Behaviour	[180,836 181,17 ... 184,75 185,48]
Output Impact	1

TABLE II: LSTM training example with 24 time steps

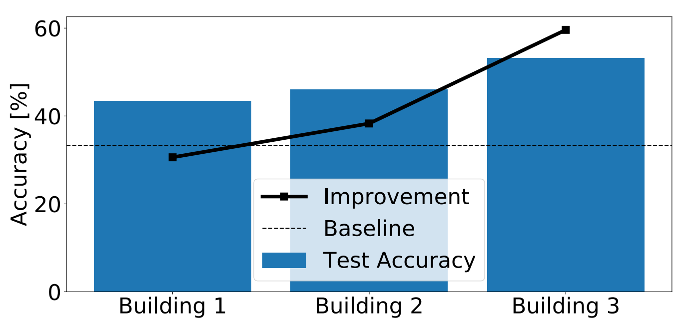

Fig. 6: Test accuracy and improvement over the baseline

We evaluate the performance on impact models by comparison to a random choice. This means, once an anomaly occurs at the campus level, we use a random choice to decide which building contributed most to the anomaly. The baseline random choice accuracy is thus considered to be 50 % for two buildings in a cluster, 33.33 % for three buildings, 25 % for four buildings, and $\frac{1}{N} \cdot 100\%$ for N buildings in a cluster. Fig. 6 presents the first results obtained for our campus comprising three buildings. The accuracy values correspond to the absolute accuracy achieved by three trained LSTMs for each building on a test data set. The shown *improvement* is computed by dividing the test accuracy by the baseline. The average test accuracy across the three buildings is 47% and the average improvement compared to random baseline is 43%. In this case, the LSTMs are trained using 2622 time series drawn from the data, while 463 time series are used for testing. A randomised test where we select three buildings from the dataset with aligned timestamps, yielded an average test accuracy of 40% and an average improvement of 21% over the baseline.

V. Evaluation

A one-layer LSTM implementation with 50 blocks has been used to evaluate the results over multiple cluster sizes and hierarchies.

For the LSTM training, the inputs are first normalised to correct for the difference in absolute levels between individual consumers and the aggregated series. This is more important for larger clusters, where the aggregated values become much larger than individual values, thereby biasing the network decision. The data set is partitioned into 70 %, 15 %, 15% for training, validation and testing the model respectively. Our LSTM block uses ReLU activation functions expressed as $g(z) = \max(0, z)$. This type of activation function has been shown in the literature to speed up neural network training by accelerating gradient descent, especially for larger datasets [15]. Another benefit is the reduced likelihood of the gradient vanishing to improve convergence.

The output layer implements a binary classifier whether the local behaviour of a consumer causes an anomaly in the aggregated behavior at the campus level. The binary classifier is implemented using a sigmoid activation functions $g(z) = 1/(1+e^{-z})$ and the cross-entropy loss for performance assessment:

$$L = -\Big(y \log(p) + (1 - y) \cdot \log(1 - p)\Big), \tag{4}$$

where y is a binary indicator reflecting if the class label is a correct classification for the respective training example, and p is the predicted probability that the observation belongs to the respective class. The selected method for solving the iterative optimisation problem is Stochastic Gradient Descent (SGD) with Adaptive Moment Estimation (ADAM) parameter optimisation. An average network is trained over 50 epochs. In this implementation context, we tested the LSTM performance in several scenarios that we detail below.

A. Impact of multiple buildings per cluster

In this scenario we evaluate the impact of the varying number of consumers per cluster. We vary the number between two and six and refer to the corresponding clusters as C2 to C6 respectively. Similarly to Fig. 6, we compute the average test accuracy and improvement over the baseline for each of these clusters in Fig. 7. Boxplots show statistical variance of the computed impact over the set of buildings in each

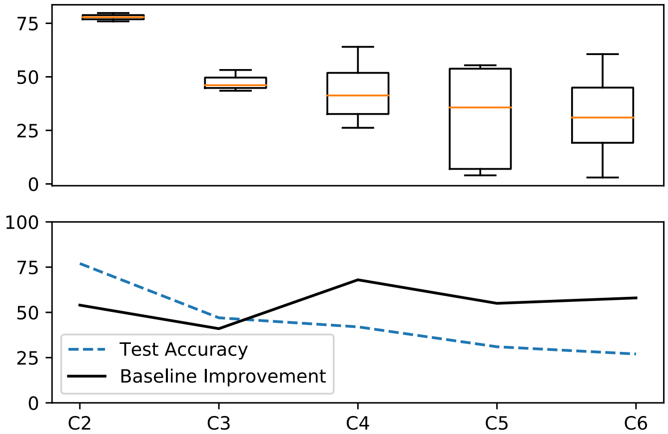

Fig. 7: Analysis of clusters with growing number of buildings

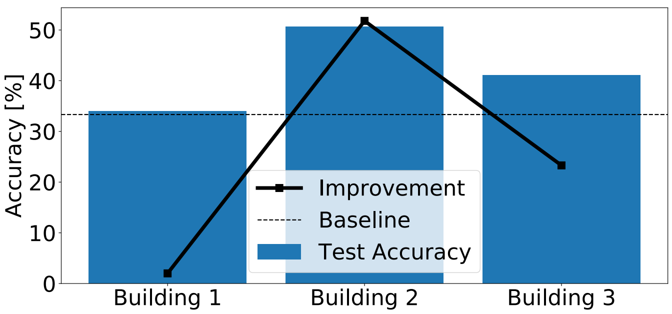

Fig. 8: Test accuracy and improvement over baseline for a cluster of buildings with similar energy usage patterns

cluster. Naturally, the test accuracy gradually decreases with the increase in the number of buildings per cluster. The relative improvement, however, appears to be stable at around 55 % on average. This happens for a decreasing baseline linked to the random choice of a building in a cluster, from 50 % in C2 down to 16.67 % in C6. This finding is important and highlights the steady performance of our impact models, even on this limited data set. With an increase in the number of buildings per cluster we can also observe larger variability at the individual consumption levels. This variability can be partially explained by the limited information carried by the aggregate behaviour when combining multiple local behaviors.

B. Impact of different consumption patterns

This scenario compares two three-building clusters which were composed based on their dominant energy usage pattern. In contrast to the results presented above for cluster C3, the other cluster contains three buildings with the same dominant usage pattern, namely only laboratories. The laboratories can be identified in the original data set by their textual descriptions: 'UnivLab_Taylor', 'UnivLabTami_Terri' and 'UnivLab_Terrie' respectively. We refer to the latter cluster as C3b and the results are illustrated in Fig. 8. The average accuracy and improvement metrics are lower than in Fig. 6, by 5 % and 25 % respectively. This can be attributed to the fact that similar usage patterns in the input time series make their impact to the aggregated trace less distinguishable and makes it more difficult for the individual LSTMs to make a better impact prediction than by chance.

C. Performance over multiple levels of hierarchy

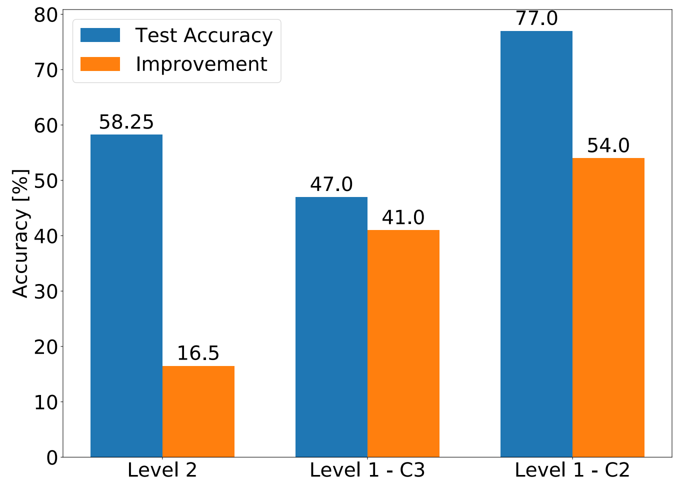

Fig. 9: Performance drop with the number of hierarchy levels

Considering the aggregation at the top level for two clusters C3 and C3b, we report the results over two levels of hierarchy by combining the resulting energy time series. Fig. 9 presents the comparison of the average metrics for low-level clusters of three and two buildings respectively to them being combined in the second-order cluster. The results show a considerable decrease in the relative improvement compared to the C2 cluster at the lower level. This is caused by the obfuscation of fine-grained interesting and distinguishing energy patterns by the higher-level subsequent aggregation. Note that the temporal resolution of the data set is hourly values, which is low compared to 15 min values provided by conventional energy meters. We can expect better performance of the impact models and their scalability with the number of aggregation levels as the temporal resolution of the input consumption data is improved.

D. Electrical energy control at the building level

We finally analyse the potential control effect that our modelling approach might have on eliminating aggregated anomalies through better local assessment of consumption impact. We are interested in determining to what extent an individual consumer should self-regulate (mostly curtail or shift) its electricity consumption in order to mitigate its impact on the aggregated anomalies within a cluster. The actual reduction of energy consumption is strongly related to the consumer load profile and usage type. One reference classification according to [16] divides the load to the base load (between 0 % and 40 % of peak load), partially flexible load (between 40 % and 70 %), and flexible load (over 70 %). This offers a safe area of load curtailment of up to 30 % of peak load for each building.

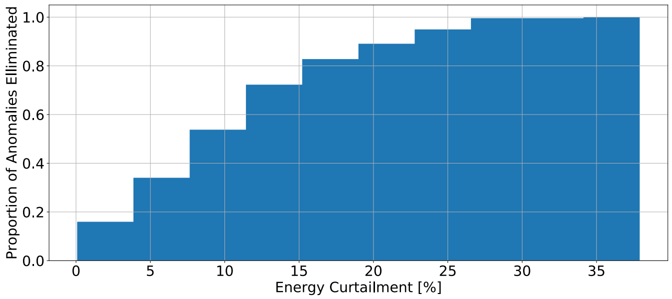

Fig. 10: Cumulative distribution of energy reduction

We use the previously defined cluster C3 with three buildings 1-3 as an example. The data we use is summarized in Table III.

	Dominant Impact	Random	Impact Model	Δ
Building 1	465	155	203	48
Building 2	1866	622	860	238
Building 3	754	251	401	150

TABLE III: Breakdown of anomaly counts by dominant building for cluster C3 = 3085

For compiling the data we use a simplifying assumption that the full impact of the anomaly is attributed to the building which contributes the most to the prediction inconsistency. The 'Random' column lists the number of anomalies that would be identified with a probability of 33 % while for the 'Impact Model' the test set accuracy is predicted by LSTM models in each building. The Δ column shows the maximum number of anomalies that we could correctly assign to each building which also represents the upper bound of the improvement potential compared to a random choice. The relative magnitude of the anomaly deviation in the total energy consumption for the dominant building is subsequently computed for a random subset of the instances of size Δ. We wish to identify the upper bound on the energy curtailment needed for which the most anomalies are eliminated across all buildings in the cluster.

The cumulative distribution of the energy reduction for Building 2 from cluster C3 is illustrated in Figure 10. This allows establishing a decision boundary for which the energy reduction is feasible in conjunction with the target reduction of the total number of anomalies. In this case a 7.64 % reduction in energy consumption would eliminate 53.78 % of the anomalies generated by this building. An 11.42 % reduction would eliminate 72.2 % of the anomalies while a 15.2 % reduction would eliminate 82.77 %. For the latter case if we apply this improvement across all buildings in the cluster we would achieve a total reduction in the number of anomalies of 39.27 %, out of which 11.69 % would be attributable to our new method compared to the baseline. The control loop consists of locally reducing power consumption in conjunction with the desired anomaly avoidance potential according to a lookup table populated with these values.

The process can be repeated across buildings and cluster hierarchies. This also offers tuning knobs for more sophisticated optimisation procedures that can adjust the individual energy reduction in conjunction with dynamically varying objectives.

VI. Related Work

Related works on *hierarchical microgrid control strategies* cover high-level architectures for data collection and control, including intelligent edge devices [17], [18]. At the lower level, accurate consumption forecasts for consumers are required to balance the generation and demand in a microgrid [19]. With regard to techniques for *building short-term load forecasting*, Support Vector Regression is analysed in [20] by applying the techniques for multi-family residential buildings. The study identifies the optimal spatiotemporal granularity of input data for floor-level hourly measurements. In this setting, the Coefficient of Variation value is reported at 2.16 % with a standard error of 0.26 %. A time series multivariate methodology for feature extraction is presented in [21]. The study revolves around selecting the best feature subset for the prediction task in iterative manner. The evolutionary search task aims at minimising the root-mean-square error (RMSE) and mean-square error (MSE) in the feature selection phase with positive impact on the posterior model prediction test error. Improvements between 30-40% are reported in terms of the test error compared to no feature selection depending on the number of future steps in the prediction horizon. The authors of [22] present an extensive study of regularised regression, including lasso and ridge regularizations, for investigating the driving factors for annual electrical energy consumption. A recent study [23] investigates the applicability of autoencoder type networks for automated feature engineering in building energy models in an unsupervised manner.

Recent contributions have focused on leveraging and extending prediction models for *anomaly classification in energy time series*. Most notably, the Rimor system [24] describes and implements a method for residential load analytics with the goal to identify abnormal patterns in energy consumption. The system is evaluated against several other methods and reflects an average improvement of 15 % in detection accuracy. Four residential data sets are used for validation to show how increasing the number of contextual features input to the model improves the accuracy evaluation metric, the symmetric MAPE in this case. In the context of anomaly detection, the impact of the time series chunk input size W is evaluated together with the proportion of the observed samples S behaving abnormally in a chunk. A similar two stage approach to anomaly detection in an energy consumption time series is presented in [25]. For prediction a hybrid ARIMA and Neural Network model is built using t lagged values of the time series as input features. The performance is evaluated based on the RMSE, MAPE and MSE metrics on various spans of training data collected from the local smart meters. A statical thresholding strategy is deployed by comparing the model prediction to the actual values within a margin defined by 2 standard deviations that persist over more than 5 min. A classification of building energy anomalies and outliers is provided by [26]. The work described in [27] uses conventional ARMA prediction models that run on edge devices with a prototype running on Raspberry Pi.

This work combines elements from these two areas of research and proposes a new approach for energy anomaly impact assessment based on deep learning methods.

VII. Conclusion and Future Work

We present a method and associated evaluation for improving energy management strategies and control of hierarchical energy systems on the edge. This is achieved through localised impact models which enable individual consumers to assess and act upon their effect on the larger grid. Using current computational intelligence techniques for short-term load forecasting and anomaly detection enables the online impact modelling at the local level. Practical evaluation on a reference data set has shown improvements over baseline accuracy of 55 % with a reduction in the total number of anomalies of 11.69 % in comparison to the baseline method. The work is beneficial for microgrid managers and end-users (building operators in our case study). For the microgrid managers this is due to improved control latency and the overall predictability of the consumption patterns. By anticipating curtailment actions the end users can avoid sudden, potentially damaging, actions. We aim for a computationally-efficient inference on embedded computing platforms. The proposed system is dedicated to medium- to large-sized buildings but is flexible and can be adapted to support other types of energy data such as many energy-intensive processes in the industry.

Future work is focused on mitigating some of the limitations of the current study, in particular with regard to LSTM model structure and evaluation diversity on better temporally resolved datasets. In particular we are interested in deploying the models, which currently have 10,651 parameters in our implementation, on embedded computing platforms and studying the effect of real-time and control constraints.

Acknowledgments

This work is funded by the Austrian Academy of Sciences, Joint Excellence in Science and Humanities (JESH) program.

References

[1] Maheen Zahid, Fahad Ahmed, Nadeem Javaid, Raza Abbasi, Hafiza Zainab Kazmi, Atia Javaid, Muhammad Bilal, Mariam Akbar, and Manzoor Ilahi. Electricity price and load forecasting using enhanced convolutional neural network and enhanced support vector regression in smart grids. *Electronics*, 8(2):122, Jan 2019.

[2] Jian LUO, Tao HONG, and Meng YUE. Real-time anomaly detection for very short-term load forecasting. *Journal of Modern Power Systems and Clean Energy*, 6(2):235–243, Mar 2018.

[3] Alexander Kell, A. Stephen McGough, and Matthew Forshaw. Segmenting residential smart meter data for short-term load forecasting. In *Proceedings of the Ninth International Conference on Future Energy Systems*, e-Energy '18, pages 91–96, New York, NY, USA, 2018. ACM.

[4] Jui-Sheng Chou and Duc-Son Tran. Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. *Energy*, 165:709 – 726, 2018.

[5] F. Ziel. Forecasting electricity spot prices using lasso: On capturing the autoregressive intraday structure. *IEEE Transactions on Power Systems*, 31(6):4977–4987, Nov 2016.

[6] Huan Zheng and Yanghui Wu. A xgboost model with weather similarity analysis and feature engineering for short-term wind power forecasting. *Applied Sciences*, 9(15), 2019.

[7] Clayton Miller. What's in the box?! towards explainable machine learning applied to non-residential building smart meter classification. *Energy and Buildings*, 199:523 – 536, 2019.

[8] J. Pereira and M. Silveira. Unsupervised anomaly detection in energy time series data using variational recurrent autoencoders with attention. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 1275–1282, Dec 2018.

[9] Y. Ma, F. Borrelli, B. Hencey, A. Packard, and S. Bortoff. Model predictive control of thermal energy storage in building cooling systems. In *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, pages 392–397, Dec 2009.

[10] Clayton Miller and Forrest Meggers. The building data genome project: An open, public data set from non-residential building electrical meters. *Energy Procedia*, 122:439 – 444, 2017.

[11] Maximilian Christ, Nils Braun, Julius Neuffer, and Andreas W. Kempa-Liehr. Time series feature extraction on basis of scalable hypothesis tests (tsfresh – a python package). *Neurocomputing*, 307:72 – 77, 2018.

[12] M.R. Braun, H. Altan, and S.B.M. Beck. Using regression analysis to predict the future energy consumption of a supermarket in the uk. *Applied Energy*, 130:305 – 313, 2014.

[13] Alex Sherstinsky. Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network, 2018.

[14] Cristina Nichiforov, Grigore Stamatescu, Iulia Stamatescu, and Ioana Fagarasan. Evaluation of sequence learning models for large commercial building load forecasting. 2019.

[15] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In Geoffrey Gordon, David Dunson, and Miroslav Dudík, editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pages 315–323, Fort Lauderdale, FL, USA, 11–13 Apr 2011. PMLR.

[16] Anil Kumar, Elena Mocanu, Muhammad Babar, and Phuong H Nguyen. Collaborative learning for classification and prediction of building energy flexibility. In *2019 IEEE PES Innovative Smart Grid Technologies Europe, ISGT Europe*. IEEE, 2019.

[17] Swaminathan Ganesan, Sanjeevikumar Padmanaban, Ramesh Varadarajan, Umashankar Subramaniam, and Lucian Mihet-Popa. Study and analysis of an intelligent microgrid energy management solution with distributed energy sources. *Energies*, 10(9):1419, Sep 2017.

[18] Iulia Stamatescu, Nicoleta Arghira, Ioana Făgărăşan, Grigore Stamatescu, Sergiu Iliescu, and Vasile Calofir. Decision support system for a low voltage renewable energy system. *Energies*, 10(1):118, Jan 2017.

[19] Xiaolong Jin, Jianzhong Wu, Yunfei Mu, Mingshen Wang, Xiandong Xu, and Hongjie Jia. Hierarchical microgrid energy management in an office building. *Applied Energy*, 208:480 – 494, 2017.

[20] Rishee K. Jain, Kevin M. Smith, Patricia J. Culligan, and John E. Taylor. Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. *Applied Energy*, 123:168 – 178, 2014.

[21] Aurora Gonzalez-Vidal, Fernando Jimenez, and Antonio F. Gomez-Skarmeta. A methodology for energy multivariate time series forecasting in smart buildings based on feature selection. *Energy and Buildings*, 2019.

[22] Aven Satre-Meloy. Investigating structural and occupant drivers of annual residential electricity consumption using regularization in regression models. *Energy*, 174:148 – 168, 2019.

[23] Cheng Fan, Yongjun Sun, Yang Zhao, Mengjie Song, and Jiayuan Wang. Deep learning-based feature engineering methods for improved building energy prediction. *Applied Energy*, 240:35 – 45, 2019.

[24] Haroon Rashid, Nipun Batra, and Pushpendra Singh. Rimor: Towards identifying anomalous appliances in buildings. In *Proceedings of the 5th Conference on Systems for Built Environments*, BuildSys '18, pages 33–42, New York, NY, USA, 2018. ACM.

[25] Jui-Sheng Chou and Abdi Suryadinata Telaga. Real-time detection of anomalous power consumption. *Renewable and Sustainable Energy Reviews*, 33:400–411, 2014.

[26] Wenqiang Cui and Hao Wang. A new anomaly detection system for school electricity consumption data. *Information*, 8(4):151, Nov 2017.

[27] X. Chang, W. Li, C. Xia, J. Ma, J. Cao, S. U. Khan, and A. Y. Zomaya. From insight to impact: Building a sustainable edge computing platform for smart homes. In *2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS)*, pages 928–936, Dec 2018.