

A Neural Ordinary Differential Equations Based Approach for Demand Forecasting within Power Grid Digital Twins

Xiang Xie*, Ajith Kumar Parlikad* and Ramprakash Srinivasan Puri[†] *Institute for Manufacturing, University of Cambridge, Cambridge, United Kingdom.

E-mail: {xx809, aknp2}@cam.ac.uk [†]Machine Learning Group, Calculus Energy Limited, Coventry, United Kingdom.
E-mail: ramprakash.puri@calculusenergy.com

Abstract

Over the past few years, deep learning (DL) based electricity demand forecasting has received considerable attention amongst mathematicians, engineers and data scientists working within the smart grid domain. To this end, deep learning architectures such as deep neural networks (DNN), deep belief networks (DBN) and recurrent neural networks (RNN) have been successfully applied to forecast the generation and consumption of a wide range of energy vectors. In this work, we show preliminary results for a residential load demand forecasting solution which is realized within the framework of power grid digital twin. To this end, a novel class of deep neural networks is adopted wherein the output of the network is efficiently computed via a black-box ordinary differential equation (ODE) solver. We introduce the readers to the main concepts behind this method followed by a real-world, data driven computational benchmark test case designed to study the numerical effectiveness of the proposed approach. Initial results suggest that the ODE based solutions yield acceptable levels of accuracy for wide range of prediction horizons. We conclude that the method could prove as a valuable tool to develop forecasting models within an electrical digital twin (EDT) framework, where, in addition to accurate prediction models, a time horizon independent, computationally scalable and compact model is often desired.

I. INTRODUCTION

As one of the most critical civil infrastructures, the electricity power grid has witnessed a significant transition from a simply physically interconnected network towards the smart grid, which allows for seamless incorporation of new technologies including advanced metering, automation, communication, distributed generation, and distributed storage into the grid infrastructures [1]. In particular, the concept of electrical digital twin (EDT) is gaining popularity within smart grid applications - which essentially provides a dynamic replica of power grid and is capable of well mimicking its real-world behaviors. According to the UK National Infrastructure Commission's report on Data for public good [2], *UK needs a digital framework for data on infrastructure to harness the benefits from sharing better quality information about its infrastructure; how it is used, maintained and planned*. Typically, a well-designed EDT framework [3] integrates sensors, cloud computing, artificial intelligence and machine learning techniques to create living digital simulation models that could realize the true value of heterogeneous data. Such integrated digital approaches enable actionable insights which drive better informed decisions for smart grid applications. Further, according to the Gemini principles [4] proposed by Centre for Digital Built Britain (CDBB), the vision for higher-level digital twin (i.e. EDT) is not that this will yield a large, singular digital twin of the entire target. Rather, it is envisaged to consist of federations of digital twins (e.g. several building digital twins) loosely coupled together via shared data such as local weather conditions. As shown in Figure:[1], advanced metering infrastructures (AMIs) within power grid digital twins, transmit streams of sensor information in real time, which is collected, extracted, transformed and then analysed to provide for example robust forecasting and control based solutions. These inter-operable models within EDT are capable of performing economic analysis of the grid through forecasted future energy production, market prices and hourly consumption values. Statistical and deep learning techniques serve as knowledge engines (KEs) that could leverage massive data to generate insights and deeper understanding of the demands. As an example, with the support of the KEs, the power grid is empowered to deal with the challenges posed by higher penetration of renewable energy, provide better designs and triggers of demand side response (DSR) actions, facilitate active pricing strategies and provide valuable input to the planning for the changes of the distribution network [5].

From the perspective of the practical application of EDT, demand forecasting (or the so called *Residential Power Load forecasting* when applied at household level) can be divided into four distinct bins: (a) very short-term load forecast when the prediction horizon is in the order of the upcoming few minutes (vSTLF), (b) short-term load forecast (STLF), usually from several minutes to one week ahead, (c) medium-term forecast, usually from a week to a year ahead, and (d) long-term forecast, usually longer than a year ahead. In this work, we restrict our analysis to short-term load forecasting for residential household demands ranging from several minutes up to few hours into the future. Such short-term demand prediction is often required for the EDT-supported applications, such as scheduling power system operation, energy balancing or energy market trading [6]. Shortly speaking, STLF plays a significant role in enhancing the practical operation of an EDT within a smart grid setup.

Computational techniques, utilizing past data, forecasted weather parameters incorporating daily activities and appliance usages, have demonstrated reasonable forecasting accuracy for some residential and commercial building level data. These techniques range

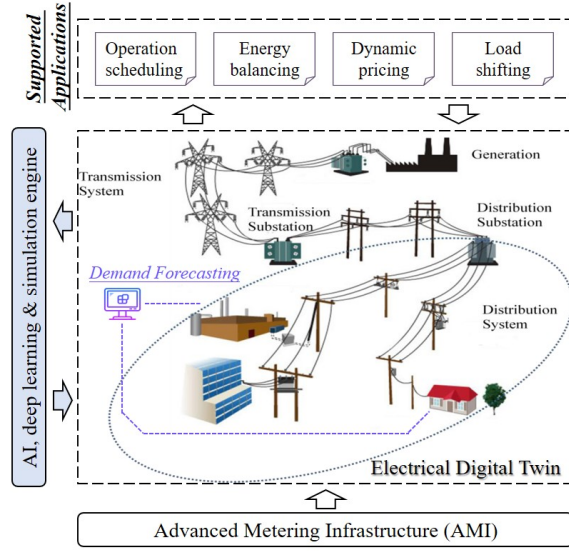


Fig. 1. A conceptual framework for Electrical Digital Twin (EDT).

from simple linear and nonlinear regression type models, decision trees, support vector machines (SVM) [7], [8], autoregressive moving average models [9], fuzzy-logic based approaches [10], gradient boosting methods [11], artificial neural networks [12]–[14] to name a few. Essentially, conventional models try to estimate the parameters from an assumed sequential model. However, these models are generally not expressive enough for describing the dynamics behind complex real-life data such as household level electricity consumption. With the advent of modern computing architectures such as graphical processing units (GPUs) and cheap cloud computational power, it has now become common practice to employ machine learning based models.

Indeed, over the past few years, deep learning techniques such as DNN, DBN and RNNs have become an active research field within STLTF for generation and consumption at building level, which are mainly adopted to learn layers of meaningful latent representations [15]. More recently, in [16], a continuous-time, neural ODE based generative approach to modeling demand time series has been proposed within an variational autoencoder (VAE) architecture. In this novel approach, each time series is represented by a latent trajectory. This unique trajectory is determined from a local initial state and the latent dynamics is shared across the time series. Using the time observation vector and the initial latent state as input, a user-defined black box ODE solver computes a vector of latent state variables, which sufficiently describes the latent dynamics at each observation point. In the final step, the neural network outputs the approximated variables using the ODE computed latent vector and parameters as input. The resultant generative neural ODE framework is memory efficient and is able to trade numerical speed/accuracy with problem size/complexity. The explicit use of an ODE solver means that the continuous-time demand forecast model is independent from time horizons, which adequately solves data inconsistency problems between power grids and buildings in addition to numerically ill-posed problems arising from external factors such as missing data. These properties, which are unique to the neural ODE framework, are certainly desirable within an EDT framework for smart grid applications.

The remainder of the paper is organized as follows. In section II, we formulate the proposed model based on the generative neural ODE approach. The input parameters, reverse mode differentiation via the ODE solver, and the distinct advantages of the proposed approach are outlined. In section III, the a real world, benchmark test case is used to validate the proposed forecasting approach is described. The effects of varying the hyper-parameters on the prediction results is studied, followed by a short discussion of the results obtained via the generative approach. Section IV concludes this paper and proposes future works within this new family of deep learning.

II. NEURAL ODE BASED DEMAND FORECASTING

In this section, the neural network architecture integrating the black-box neural ordinary differential equation (ODE) within variational autoencoder architecture (VAE) is presented. Variational autoencoder (VAE) is a variational inference based autoencoder that learns the latent space representation of time series data in an unsupervised manner [17]. In order to better model complex real-world electricity load demand time series, the aforementioned neural ordinary differential equation is integrated into VAE architecture by replacing the feed-forward network in the latent space of VAE with neural ODE. Here, neural ODE can introduce temporal dependencies of time series data into VAE in a continuous time manner [6].

A. Neural Ordinary Differential Equations

Recurrent neural network (RNN) with residual connections is a well-known network architecture for modeling sequential data [18]. RNNs are considered as deep networks with shared parameters in each layer when unfolded on time, which aims at describing the evolution of the hidden state through time. Specifically, residual connected RNN adopts a feed-forward architecture by recursively transforming a neurons internal state or so-called *memory* into the next neurons input along a temporal sequence. This kind of network architecture provides the capability of not only retaining sequential information but also introducing additional information at each time step. Inheriting the basic idea of residually connected RNN, in neural ordinary differential equations, the entire neural network is propagated through a continuous block of computation with infinitesimally small time steps rather than a discrete sequence of hidden layers in RNN [16]. In the limit, the continuous dynamics of hidden layers can be parameterized using an ordinary differential equation (ODE):

$$\frac{d\mathbf{h}(t)}{dt} = f(\mathbf{h}(t), t, \theta) \quad (1)$$

where function $\mathbf{h}(t)$ is the value of hidden state evaluated at time t , which is unrolled as a continuous parameterisation of layer depth. Shortly speaking, the values of the hidden states can be calculated by a black-box ODE solver, utilizing a wide range of techniques - ranging from simple Euler's methods to higher-order variants of the Runge-Kutta method [19]. All these numerical methods obtain the solution sequentially by evaluating the continuous-defined hidden state dynamics f :

$$\mathbf{h}(t_n) = \text{ODEsolver}(\mathbf{h}(t_0), t_0, t_1, \dots, t_n, f, \theta) \quad (2)$$

The derivative function f is implemented using a simple neural network parameterized by θ , which assumes to be independent from time in this application.

An important aspect of implementing the neural ODE approach is to optimize the choice of parameters θ by back-propagating through solving a second, augmented ODE backwards using the adjoint method. The adjoint method is able to efficiently compute the gradient of the loss function $\mathcal{L}(\text{ODEsolver}(\mathbf{h}(t_0), t_0, t_1, \dots, t_n, f, \theta))$ with respect to the parameters θ without incurring a high memory footprint. The implementation details of this back-propagation algorithm is beyond the scope of this paper and the reader is referred to [16] and the references therein.

B. Neural ODE Based Time Series Forecasting Model

Leveraging the neural ODE approach mentioned above, the change of time series can be summarized as a smooth trajectory starting from a local initial state in latent space, and temporal information is recursively processed using an ODE solver in form of a sequence of latent random variables.

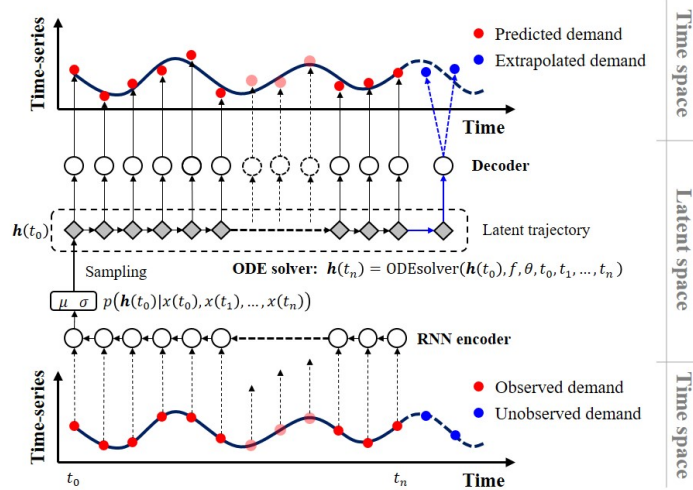


Fig. 2. Illustration of a neural ODE based time series forecasting model.

Figure 2 depicts this unrolled architecture of neural ODE based encoder-decoder modules used for time series forecasting. Note that an encoder-decoder structure is integrated into the architecture to provide an end-to-end mapping between the input time series and its corresponding forecasted output time series.

Consider the input vectors $X = \{x(t_0), x(t_1), \dots, x(t_n)\}$ of the time series forecasting model, where the multidimensional inputs $x(t_0), x(t_1), \dots, x(t_n) \in \mathbb{R}^m$ are combinations of past values of the targeted time series and the most influential and

highly correlated $(m - 1)$ exogenous variables at each observation time instance $t \in \{t_0, t_1, \dots, t_n\}$. A RNN based recognition network reads each input time series X sequentially backwards in time, from t_n back to t_0 . In this way, the recurrent neurons can approximate the posterior $q(\mathbf{h}(t_0) | x(t_0), x(t_1), \dots, x(t_n), \phi)$, where $\mathbf{h}(t_0) \in \mathbb{R}^h$ represents the local initial state for generating corresponding latent trajectory in latent space, while h denotes the dimension of latent space and ϕ is the parameter of RNN. It has been reported that when modeling complex temporal behaviours, the performance of deep neural networks would significantly improve when uncertainty is inherently included in their hidden states. Therefore, to enhance the effectiveness of the proposed time series forecasting model, the variational inference model is incorporated and latent random variable \mathbf{h} is defined, which makes the model more robust to the uncertainties contained in the input time series. Mathematically speaking, for arbitrary distribution in h dimensional space, the targeted distribution can be generated by taking a set of h normally distributed variables and mapping them through a sufficiently complicated function. Thus, it can be assumed that the latent random variable $\mathbf{h}(t_0)$ is sampled from a Gaussian distribution, that is $\mathbf{h}(t_0) \sim \mathcal{N}(\mu, \sigma)$, where the mean μ and variance σ are derived from the hidden state of RNN at time t_0 . Starting from the local initial state $\mathbf{h}(t_0)$, the subsequent latent trajectory $\mathbf{h}(t_1), \dots, \mathbf{h}(t_n), \mathbf{h}(t_{fore})$ be uniquely determined using a numerical differential equation solver with desired accuracy. Thus, incorporating the ODE solver allows us to make predictions ($t_i \in [t_0, t_n]$) and extrapolations ($t_{fore} \in [t_n, +\infty)$) at arbitrary time points on a continuous timeline.

It is easy to see from existing literature that most of current time series forecasting methods adopt a recursive framework, which means that the predictions are treated as inputs to generate a forecast for the next time step. Studies have shown that the forecasting performance often deteriorates significantly after few time steps. On the other hand, the proposed neural ODE approach in this paper provides a flexible time series forecasting model that is capable of forecasting the time series value on arbitrary time horizons. We show this horizon independent property in Section III. Finally, the decoder directly reconstructs the time series $\hat{x}(t_0), \hat{x}(t_1), \dots, \hat{x}(t_n), \hat{x}(t_{fore}) \in \mathbb{R}$ using the latent trajectory $\mathbf{h}(t_0), \mathbf{h}(t_1), \dots, \mathbf{h}(t_n), \mathbf{h}(t_{fore})$ with parameters α . It is worth noting that the output time series is only a partial reconstruction of the original input time series.

After specifying the neural network architecture, the encoder, decoder and latent dynamics are trained simultaneously by optimizing the evidence lower bound (ELBO) of the marginal probability, defined as the following equation:

$$ELBO = \sum_{i=0}^n \log p(x(t_i) | \mathbf{h}(t_i), \alpha) + \log p(\mathbf{h}(t_0)) - \log q(\mathbf{h}(t_0) | \{x(t_i), t_i\}_i, \phi) \quad (3)$$

where the first term denotes the sum of the expected log likelihood of the time series data, while the second and third terms represent the Kullback-Leibler divergence between the variational distribution and the prior of the local initial state $\mathbf{h}(t_0)$, and the prior $p(\mathbf{h}(t_0))$ follows a standard Gaussian distribution.

In summary, a neural ODE based autoencoder with partial reconstruction is proposed to realize the time series forecasting. Practically speaking, it is reasonable that the dimensionality of input time series is larger than the that of output time series. This means that all the available information is visible to the encoder, while the designed neural network only needs to focus on capturing and activating key information that is closely relevant to the selected output time series.

C. Key Parameters in Electricity Demand Forecasting

Inspired by the neural ODE based time series forecasting model proposed in section II-B, the electricity load demand forecasting problem in smart grids is studied in this section. Electricity load demands are normally recorded in form of regularly sampled time series in smart grids. $L = \{l(t_0), l(t_1), \dots, l(t_n)\}$ denotes the electricity load demands recorded from t_0 to t_n , where the time interval $[t_0, t_n]$ is the input width of sliding time window for forecasting future demand $\hat{l}(t_{fore})$ within the output time range:

$$\begin{cases} \text{RNN} : \{l(t_i) \in \mathbb{R}^m : t_i \in [t_0, t_n]\} \rightarrow \mathbf{h}(t_0) \in \mathbb{R}^h \\ \text{N-ODE} : \mathbf{h}(t_0) \in \mathbb{R}^h \rightarrow \mathbf{h}(t_0), \dots, \mathbf{h}(t_{fore}) \in \mathbb{R}^h \\ \text{Decoder} : \mathbf{h}(t_{fore}) \rightarrow \hat{l}(t_{fore}) \in \mathbb{R} \end{cases} \quad (4)$$

As shown in the figure 3, in this work, a sliding window based multiple-input-multiple-output (MIMO) mapping framework. To this end, the two significant hyper-parameters related to the demand forecasting model used in this study are the input/output sliding window width w_i/w_o and dimensionality of latent space h . The input width of the sliding time window w_i is the set of lagged values used to build the future forecasts within the output sliding window width w_o . Further, to avoid overfitting or underfitting the electricity demand data, the latent space dimension needs to be optimized. In the next section, the effect of output sliding window width w_o as well as latent space dimension h on demand forecasting performance is studied.

III. COMPUTATIONAL TEST CASE

To evaluate the accuracy of the demand forecasting model, a publicly available benchmark dataset is utilized. This section provides details on these datasets alongside a description of the tuning parameters used within the neural ODE framework to fit and extrapolate the load time series. Solution accuracy, error quantities and strategies to trade numerical accuracy for speed are

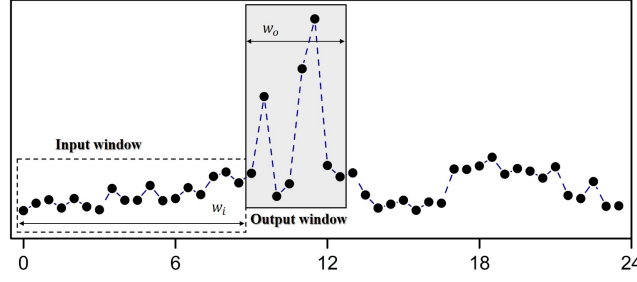


Fig. 3. An example of sliding window based demand forecasting framework.

discussed. All computations described in this work were carried out on an Ubuntu Linux server with PyTorch [20] and torchdiffeq [16] compiled against a 16GB PNY NVIDIA Tesla P100 Card with CUDA 9.0 and cuDNN support.

A. Dataset

The chosen benchmark dataset arises from energy consumption readings for a sample of 5,567 London Households that took part in the UK Power Networks (UKPN) led Low Carbon London project [21]. This archive contains 36,460 measurements gathered between November 2011 and February 2014. Readings were taken at half hourly intervals. i.e values are reported in kWh/hh (per half hour). Households have been allocated to a CACI Acorn group (2010). The customers in the trial were recruited as a balanced sample representative of the Greater London population. The consumption of a single household, identified with ID: MAC000006 [22] belonging to the CACI Acorn group [23], *Adversity*, utilizing a *Standard* tariff is used in this study. We use this dataset rather than aggregated or commercial building data because short-term load forecasting (STLF) at such individual household is challenging due to significant uncertainty and volatility inherent in this type of data [24], [25]. Literature on residential load forecasting is limited and thus insights cannot be readily gained from existing research. Due to the decentralized nature of modern grid system, short-term household-level demand forecasting remains an open and an important research problem.

It is worth noting that forecasting for large demands such as commercial building, city or the entire grid has been achieved with relatively high accuracy and prominently appears within the building simulation and power systems literature [26], [27]. However, forecasting demands at a local, disaggregated level, such as the studied individual household in this case, is considered a fairly complex problem [28] since it is difficult to entirely reconstruct the original signal and perform time-series extrapolation at this level. It is often assumed that this final time evolution of the demand states is a superposition of stable components of the signal resulting from daily activity patterns of the household and unstable components resulting from stochastic events [29].

According to the rule-of-thumb applied within the machine learning community, the available dataset is split into two non-overlapping groups, namely training set and testing set. 70% of available data is utilized to develop the forecasting model, while 30% of that is used to verify the forecasting accuracy. Generally, the demand time series forecasting should be complemented by other high-relevant economic, environmental and behavioural factors [30], such as occupancy patterns, household incomes, temperature, humidity, wind speed and solar radiance. However, in this work, we restrict ourselves to a purely data driven approach itself, wherein, only the past electricity demands are used as inputs to forecast future energy consumption. Thus, we aim to capture the more salient components of the time series signal.

B. Performance evaluation index

In this benchmark test, data samples of lagged electricity demand within 24 hours ($w_i = 48$) is used to forecast the demands 6 hours ahead ($w_o = 12$) in horizon, when the latent space dimensionality h takes value of 128, 256 or 512 respectively. Specifically, the lagged input can be expressed as:

$$l_{in} = \{l(t_1), \dots, l(t_{w_i})\}, w_i = 48 \quad (5)$$

and the forecasted output can be expressed as:

$$\hat{l}_{out} = \{\hat{l}(t_{w_i+1}), \dots, \hat{l}(t_{w_i+w_o})\}, w_o = 12 \quad (6)$$

In order to assess and compare the performance of the proposed demand forecasting method under different parameter combinations, the normalized root mean square error (NRMSE) is used as the error metric:

$$\text{NRMSE}(\%) = 100 \frac{\sqrt{\frac{1}{n_t} \sum_{j=1}^{n_t} (\hat{l}_j(t_i) - l_j(t_i))^2}}{l_{max} - l_{min}} \quad (7)$$

where $\hat{l}_j(t_i)$ and $l_j(t_i)$ correspond to the forecasted and true electricity demand respectively, l_{max} and l_{min} are the maximum and minimum electricity demand within the period, and n_t denotes the size of the testing dataset. The use of NRSME is made to reduce their influence of large variations in the time series data.

C. Results and discussion

Figure:[4] shows the computed forecasting accuracy of training and testing set in terms of NRMSE under different latent dimension, h . From these results, it is evident the forecast performance of the neural ODE based method depends on the latent space dimensionality. Typically, increasing the dimensionality contributes to a higher expressiveness capability of the model, but inevitably increases computational cost and higher risk of overfitting at the same time. Overall, the forecasted values show fairly accurate predictions on household demands. Finally, the testing results suggest that for this particular test case, a latent dimension of 512 is an appropriate choice. One could attribute this higher latent dimension to the fact that unaggregated values were used in this study.

Figure:[5] illustrates the true and forecasted 1 step (half an hour) and 12 steps (6 hours) ahead demands for a latent space dimensionality of 512. It is evident that the household-level demand profiles are stochastic and random spikes in demands occur within short period of time, which in-turn is difficult to capture and reconstruct. These spikes in consumption are attributed to the volatile nature of the individual resident, such as operating different appliances. Note that in this work, we do not perform any level of aggregation. Indeed, it is well known that such aggregation reduces the inherent variability in electricity consumption resulting in increasingly smooth load profiles [31], which are easier to reconstruct, leading to lower forecasting errors (see for example [32], [33]). On the other hand, when aggregation is not utilized, large errors have been reported in literatures (e.g. [34], [35]). Apparently, such forecasting errors vary from household to household, and has been attributed to dwelling lifestyle, regularity of appliance usage and even personal preferences [36]. To this end, the current work demonstrates that the neural ODE based approach is capable of capturing the salient futures within the time series data with a constant error rate for both one-step and multi-step predictions.

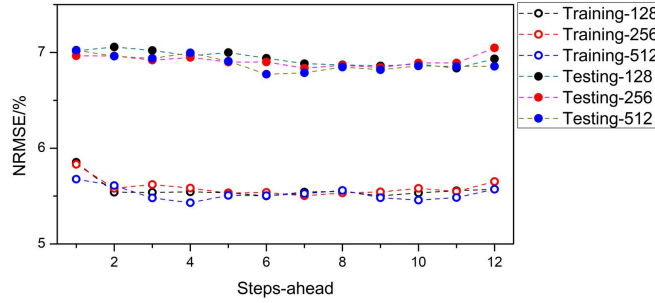


Fig. 4. Household electricity consumption forecasting performance.

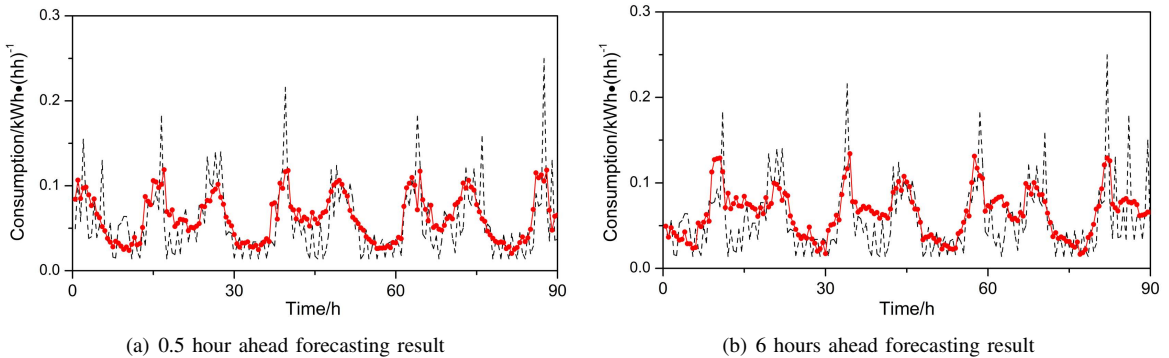


Fig. 5. Household electricity consumption forecasting results ($h = 512$). Dotted lines: actual values; red line: a series of forecasted quantities.

Thus, it is possible to obtain relatively stable model with a balanced forecasting accuracy within the targeted output window w_o .

IV. CONCLUSION AND FUTURE WORK

In this paper, we have shown preliminary results from a novel class of deep learning framework, namely the neural ODE, which is integrated within variational autoencoder architecture to realize short-term load forecast. Within this neural ODE framework, an

encoder-decoder structure is further integrated into the framework to provide an end-to-end mapping: a RNN based encoder is used to generate the local initial state of the latent trajectory, while a decoder is used to reconstruct the forecasts from the extended latent trajectory. A household-level benchmark dataset is utilized to verify the effectiveness of the proposed method. It is shown that the proposed architecture achieves stability and balanced forecasting performance when the hyper-parameters are chosen wisely, given the volatile nature of the individual residential level electricity loads. In addition to this, the continuous timeline property of the proposed architecture contributes towards efficiently working with data inconsistency problems arising from data sharing between building and electrical digital twins.

The next phase of this on-going work focuses on enhancing model generalization, integration of exogenous factors and studying the influence of model depth on prediction errors. These studies will enable utilizing multi-horizon prediction algorithms to control physical assets belonging to different energy vectors within an electrical digital twin framework. A straightforward extension of the approach presented in this paper is also envisaged for forecasting the generation of renewable energy production assets such as wind and solar.

ACKNOWLEDGEMENT

This research that contributed to this paper was funded by the EPSRC/Innovate UK Centre for Smart Infrastructure and Construction (CSIC) and Centre for Digital Built Britain (CDBB) at the University of Cambridge.

REFERENCES

- [1] M. L. Tuballa and M. L. Abundo, "A review of the development of smart grid technologies," *Renewable and Sustainable Energy Reviews*, vol. 59, pp. 710–725, 2016.
- [2] "National infrastructure commission, data for the public good," <https://www.nic.org.uk/publications/data-public-good>, accessed: 2019-05-01.
- [3] Q. Lu, A. K. Parlikad, P. Woodall, G. Don Ranasinghe, and J. Heaton, "Developing a dynamic digital twin at a building level: Using Cambridge campus as a case study," in *International Conference on Smart Infrastructure and Construction (ICSIC)*, 2019.
- [4] CDBB, "Gemini principles," <https://www.cdbb.cam.ac.uk/DFTG/GeminiPrinciples>, accessed: 2019-07-28.
- [5] P. Mirowski, S. Chen, T. K. Ho, and C.-N. Yu, "Demand forecasting in smart grids," *Bell Labs technical journal*, vol. 18, no. 4, pp. 135–158, 2014.
- [6] M. Q. Raza and A. Khosravi, "A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings," *Renewable and Sustainable Energy Reviews*, vol. 50, pp. 1352–1372, 2015.
- [7] M. G. De Giorgi, S. Campilongo, A. Ficarella, and P. M. Congedo, "Comparison between wind power prediction models based on wavelet decomposition with least-squares support vector machine (LS-SVM) and artificial neural network (ANN)," *Energies*, vol. 7, no. 8, pp. 5251–5272, 2014.
- [8] J. Wu and D. Niu, "Short-term power load forecasting using least squares support vector machines (LS-SVM)," in *2009 Second International Workshop on Computer Science and Engineering*, vol. 1, 2009, pp. 246–250.
- [9] J. W. Taylor, "Triple seasonal methods for short-term electricity demand forecasting," *European Journal of Operational Research*, vol. 204, no. 1, pp. 139 – 152, 2010.
- [10] K. B. Song, Y. S. Baek, D. H. Hong, and G. Jang, "Short-term load forecasting for the holidays using fuzzy linear regression method," *IEEE transactions on power systems*, vol. 20, no. 1, pp. 96–101, 2005.
- [11] S. B. Taieb and R. J. Hyndman, "A gradient boosting approach to the kaggle load forecasting competition," *International Journal of Forecasting*, vol. 30, no. 2, pp. 382 – 394, 2014.
- [12] C. Sun, F. Sun, and S. J. Moura, "Nonlinear predictive energy management of residential buildings with photovoltaics & batteries," *Journal of Power Sources*, vol. 325, pp. 723 – 731, 2016.
- [13] M. Beccali, M. Cellura, V. L. Brano, and A. Marvuglia, "Forecasting daily urban electric load profiles using artificial neural networks," *Energy Conversion and Management*, vol. 45, no. 18, pp. 2879 – 2900, 2004.
- [14] K. Siwek, S. Osowski, and R. Szupiluk, "Ensemble neural network approach for accurate load forecasting in a power system," *International Journal of Applied Mathematics and Computer Science*, vol. 19, no. 2, pp. 303–315, Jun. 2009.
- [15] M. Långkvist, L. Karlsson, and A. Loutfi, "A review of unsupervised feature learning and deep learning for time-series modeling," *Pattern Recognition Letters*, vol. 42, pp. 11–24, 2014.
- [16] T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, "Neural ordinary differential equations," in *Advances in Neural Information Processing Systems*, 2018, pp. 6571–6583.
- [17] M. Jang, S. Seo, and P. Kang, "Recurrent neural network-based semantic variational autoencoder for sequence-to-sequence learning," *Information Sciences*, 2019.
- [18] B. Yue, J. Fu, and J. Liang, "Residual recurrent neural networks for learning sequential representations," *Information*, vol. 9, no. 3, p. 56, 2018.
- [19] L. F. Shampine, I. Gladwell, L. Shampine, and S. Thompson, *Solving ODEs with matlab*. Cambridge university press, 2003.
- [20] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," *NIPS 2017 Workshop Autodiff Program Chairs*, 2017.
- [21] "Low Carbon London," [http://innovation.ukpowernetworks.co.uk/innovation/en/Projects/tier-2-projects/Low-Carbon-London-\(LCL\)/](http://innovation.ukpowernetworks.co.uk/innovation/en/Projects/tier-2-projects/Low-Carbon-London-(LCL)/), accessed: 2019-01-04.
- [22] "SmartMeter Energy Consumption Data in London Households," <https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households>, accessed: 2019-01-04.
- [23] "CACI (2010) ACORN consumer classification," <https://acorn.caci.co.uk/downloads/ACORN-User-guide.pdf>, accessed: 2019-01-04.
- [24] H. Shi, M. Xu, and R. Li, "Deep learning for household load forecasting a novel pooling deep RNN," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5271–5280, 2018.
- [25] S. Humeau, T. K. Wijaya, M. Vasirani, and K. Aberer, "Electricity load forecasting for residential customers: Exploiting aggregation and correlation between households," in *2013 Sustainable Internet and ICT for Sustainability (SustainIT)*, Oct 2013, pp. 1–6.
- [26] M. W. Ahmad, M. Mourshed, and Y. Rezgui, "Trees vs neurons: Comparison between random forest and ann for high-resolution prediction of building energy consumption," *Energy and Buildings*, vol. 147, pp. 77 – 89, 2017.
- [27] J. W. Taylor and P. E. McSharry, "Short-term load forecasting methods: An evaluation based on european data," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 2213–2219, Nov 2007.

- [28] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan 2019.
- [29] C. N. Yu, P. Mirowski, and T. K. Ho, "A sparse coding approach to household electricity demand forecasting in smart grids," *IEEE Transactions on Smart Grid*, vol. 8, no. 2, pp. 738–748, 2017.
- [30] L. Zhang and J. Wen, "A systematic feature selection procedure for short-term data-driven building energy forecasting model development," *Energy and Buildings*, vol. 183, pp. 428–442, 2019.
- [31] R. Sevlia and R. Rajagopal, "A scaling law for short term load forecasting on varying levels of aggregation," *International Journal of Electrical Power & Energy Systems*, vol. 98, pp. 350 – 361, 2018.
- [32] K. Siwek, S. Osowski, R. Szupiluk, P. Wojewnik, and T. Zabkowski, "Blind source separation for improved load forecasting in the power system," in *Proceedings of the 2005 European Conference on Circuit Theory and Design, 2005.*, vol. 3. IEEE, 2005, pp. III–61.
- [33] R. Szupiluk, P. Wojewnik, T. Zabkowski, A. Niepodleglosci, P. T. Cyfrowa, and A. Jerozolimske, "Combining forecasts with blind signal separation methods in electric load prediction framework," in *Artificial Intelligence and Applications*, 2006, pp. 439–444.
- [34] K. Gajowniczek and T. Zbkowski, "Data mining techniques for detecting household characteristics based on smart meter data," *Energies*, vol. 8, no. 7, pp. 7407–7427, 2015.
- [35] F. Javed, N. Arshad, F. Wallin, I. Vassileva, and E. Dahlquist, "Forecasting for demand response in smart grids: An analysis on use of anthropologic and structural data and short term multiple loads forecasting," *Applied Energy*, vol. 96, pp. 150–160, 2012.
- [36] K. Gajowniczek and T. Zabkowski, "Electricity forecasting on the individual household level enhanced based on activity patterns," *PloS one*, vol. 12, no. 4, p. e0174098, 2017.