Probabilistic Short-Term Low-Voltage Load Forecasting using Bernstein-Polynomial Normalizing Flows

Marcel Arpogaus [1] **Marcus Voss** [2] **Beate Sick** [3] **Mark Nigge-Uricher** [4] **Oliver Dürr** [1]

Abstract

The transition to a fully renewable energy grid requires better forecasting of demand at the low-voltage level to increase efficiency and ensure a reliable control. However, high fluctuations and increasing electrification cause huge forecast errors with traditional point estimates. Probabilistic load forecasts take future uncertainties into account and thus enables various applications in low-carbon energy systems. We propose an approach for flexible conditional density forecasting of short-term load based on Bernstein-Polynomial Normalizing Flows where a neural network controls the parameters of the flow. In an empirical study with 363 smart meter customers, our density predictions compare favorably against Gaussian and Gaussian mixture densities and also outperform a non-parametric approach based on the pinball loss for 24h-ahead load forecasting for two different neural network architectures.

1. Introduction

The energy sector is the major contributor to greenhouse gas emissions (World Resources Institute, 2020). The take-up of renewable and distributed energy resources transforms the electric energy system to be more decentralized. This increases the role of low-voltage (LV) grids that typically make up the largest part of distribution systems but are still the least monitored and controlled. Accurate short-term load and generation forecasts at the LV level ranging from minutes to days ahead are becoming essential for grid operators, utilities, building- and district operators, and the customers themselves for many applications. Forecasts impact greenhouse gas emissions directly by making the energy systems more efficient and indirectly by enabling a reliable carbon-free energy supply with high utilization of fluctuating renewable energy sources and increasingly unpredictable consumption patterns through electrification of the heating and mobility sectors. Such applications include, for instance, peak load reduction (Rowe et al., 2014) and voltage control (Zufferey et al., 2020). Accurate load forecasts can be used for grid state estimation (Hermanns et al., 2020) rendering a completely measured, hence expensive grid unnecessary. They can also be used for anomaly detection to increase resilience (Fadlullah et al., 2011) or detect energy theft (Fenza G., 2019). Short-term forecasts can inform different participants of services like local and peer-to-peer energy markets (Morstyn et al., 2018), real-time pricing schemes (He et al., 2019) or flexibility applications (Ponoćko & Milanović, 2018) that are emerging with the energy transition. Haben et al. (2021) provide a recent review of LV load forecasting methods and applications.

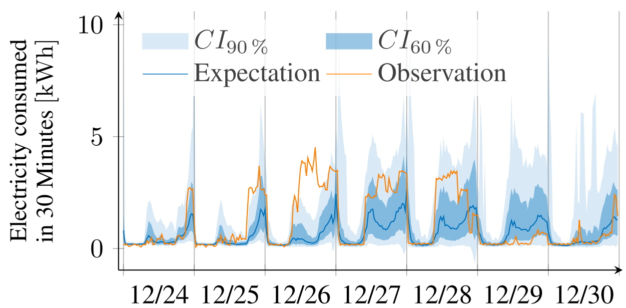

Figure 1. Exemplary load forecast in Christmas week based on data from (Commission for Energy Regulation (CER), 2012).

LV demand typically exhibits higher volatility than the high voltage or more aggregated system demand. For this reason, probabilistic forecasts are more appropriate than point forecasts due to their representation of the underlying uncertainty. Whereas point forecasts model the expected value, probabilistic forecasts estimate the distribution for more informed decision-making (see Hong & Fan, 2016, for a review). The distribution can be modeled fully continuous (Pinto et al., 2017; Arora & Taylor, 2016) or be approximated by quantile estimates (Wang et al., 2019; Gerossier et al., 2018; Elvers et al., 2019). Figure 1 gives an example of a probabilistic forecast for a household that had an extraordinary high consumption during the Irish Christmas

[1] HTWG Konstanz - University of Applied Sciences [2] Technische Universität Berlin (DAI-Lab) [3] EBPI, University of Zurich & IDP, Zurich University of Applied Sciences [4] Bosch.IO GmbH. Correspondence to: Marcel Arpogaus <marcel.arpogaus@htwg-konstanz.de>.

Tackling Climate Change with Machine Learning Workshop at ICML 2021.

holidays. The plot displays the uncertainty information in the form of confidence intervals for 60% and 90%. It still underestimates the unusual high load on the $26th$, however, once this observation is processed, it can be incorporated in the forecasts of the following days. This way probabilistic forecast can reflect future uncertainty.

Probabilistic forecasts at the LV level are challenging, as the time series are multivariate, and the marginal distributions are typically skewed and multi-modal. Other studies proposed parametric approaches like Kernel Density Estimation (Arora & Taylor, 2016; Pinto et al., 2017), a Gaussian distribution (Haben et al., 2019; Wijaya et al., 2015; Salinas et al., 2020) or a Gaussian Mixture Model (GMM) (Vossen et al., 2018) and non-parametric deep architectures minimizing the pinball loss (Elvers et al., 2019; Wang et al., 2019). *Normalizing Flows (NFs)* are flexible parameterized transformations from simple distributions (e.g. Gaussian) to complex ones (cf. (Papamakarios et al., 2019)).

In Section 2 *we propose Bernstein-Polynomial Normalizing Flow (BNF) for short-term density forecasting of LV loads*. In Section 3 we use public data from 363 smart meter customers of the CER dataset (Commission for Energy Regulation (CER), 2012) to compare variants of the Neural Network (NN) architecture used to estimate the BNF parameters, and we compare it to both parametric and non-parametric benchmark approaches in 24h-ahead forecasting. Finally, we conclude this study in Section 4.

2. Normalizing Flows using Bernstein-Polynomials

We tackle the load forecasting problem in the framework of deep probabilistic regression. For the covariates $\mathbf{x}$ such as lagged power consumption at earlier time steps, holiday indicator, or temperature, the marginal Conditional Probability Distributions (CPDs) $p_y(y_t|\mathbf{x})$, of the electric load at timestep t are predicted. We have used three well established probabilistic models beside our Bernstein normalizing flow model, which is described in the following.

The main idea of NF is to fit a parametric bijective function that transforms between a complex target distribution $p_y(y)$ and a simple distribution $p_z(z)$, often $p_z(z) = N(0, 1)$. The change of variable formula allows us to calculate the probability $p_y(y)$ from the simple probability $p_z(z)$ as follows:

$$p_y(\mathbf{y}) = p_z(f(\mathbf{y})) \left|\det \nabla f(\mathbf{y})\right|^{-1} \quad (1)$$

with the Jacobian determinant $\det \nabla f(y)$ ensuring that the probability integrates to one after the transformation (hence the name normalizing flow). Using Eq. 1 the parameters of f can be tuned such that the likelihood of the observed training samples is maximized. Often a combination of K simple transformation functions f_i is used to compose more expressive transformations $f(z) = f_K \circ f_{K-1} \circ \cdots \circ f_1(y)$, while staying computational efficient (Papamakarios et al., 2019).

Up to now NF models have gained the most attention in applications where complex high-dimensional unconditional distributions $p_y(\mathbf{y})$ are modeled, for example for image generation (Kingma & Dhariwal, 2018) or speech synthesis (van den Oord et al., 2017). Probabilistic regression models based on NF, modeling the CPD $p_y(y|\mathbf{x})$ have gained only little attention. However, recent research applied NF for probabilistic regression with very promising results (Rothfuss et al., 2020; Rasul et al., 2020; Trippe & Turner, 2018; Sick et al., 2020; Ramasinghe et al., 2021). Our implementation is based on Sick et al. (2020), but has some significant improvements, to allow a stable inversion of the flow from the latent variable z to the observed y, which are described in the following and illustrated in Figure 2. The core of the transformation are the Bernstein polynomials of order M, which can approximate any function in $y \in [0, 1]$ for $M \to \infty$ but empirically $M = 10$ polynomials are often sufficient (Hothorn et al., 2018). Using higher degree polynomials, however, can increase the expressiveness at no cost to the training stability (Ramasinghe et al., 2021).

Our implementation uses Bernstein polynomials as the second transformation in a chain of three transformations $f_3 \circ f_2 \circ f_1$ (see Figure 2).

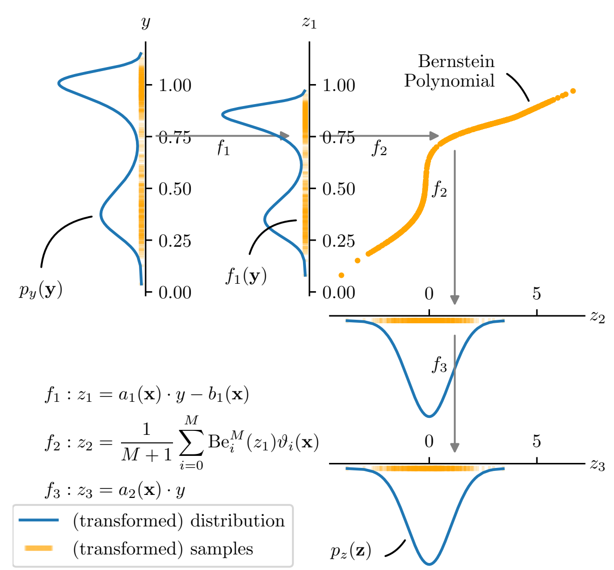

Figure 2. Visual representation of the normalizing flow. The flow transforms a bimodal distribution $p(y|x)$ in the upper left side, via a chain of flows to a standard Gaussian $p(z)$ lower right side. The dependence on the covariate x stems from the x-dependence of the NN controlled flows f_1, f_2, f_3. The flow f_2 uses Bernstein polynomials for maximal flexibility.

The first transformation f_1 scales and shifts y. The second transformation f_2 consists of a Bernstein polynomial with $M+1$ parameters $\theta = (\vartheta_0, \vartheta_1, \dots \vartheta_M)$. The last transformation in the chain, f_3 is the scale of the base distribution.

All the parameters $a_1, b_1, \vartheta_{\mathbf{i}}$ and a_2 are controlled by the outputs of a NN with input $\mathbf{x}$. To ensure the invertibility of the complete flow the individual components need to be strictly monotonous. For f_1 and f_3 we need to ensure that the scale parameters a_1 and a_2 are positive, this is done by applying a $\mathrm{softplus}$ activation function to the output of the network. In order to ensure the strict monotonicity of the Bernstein polynomial f_2 the parameters ϑ_i need to be increasing. Moreover, we restricted the transformation $f_2(z_2)$ to be at least in the range $[c=-3, d=3]$ to be within $\pm 3\sigma$ of the standard Gaussian. To achieve the required restrictions we apply the following procedure to the unconstrained outputs ϑ'_i of the network. We first use the $\mathrm{softmax}$ function to get positive values $\vartheta'' = \mathrm{softmax}(\vartheta'_1, \dots, \vartheta'_{M-1})$ and from these values we determine the coefficients in the Bernstein polynomial as $\vartheta_0 = c$ and $\vartheta_i = c + \sum_{j<i} \vartheta''_j \cdot (d-c)$ for $i = 1, \dots, M-1$. This ensures that the ϑ_is are increasing in $[c, d]$ and leads to $M-1$ parameters in total. Finally, we allow that the transformation is smaller than c and larger than d by subtracting and adding the $\mathrm{softplus}$ of the outputs ϑ'_0 and ϑ'_M of the network to c, and d.

Training of our probabilistic BNF model is done by tuning the parameters of the chained transformations $f_3 \circ f_2 \circ f_1$ to minimize the negative log-likelihood $-\sum_i^N \ln(p_y(y_i))$ of the N training data points, using Equation 1. The training data has been transformed to $[0,1]$. In the case of (equally transformed test-data) being outside this range, the Bernstein polynomials are linearly extrapolated. For simulating scenarios of possible energy consumption time series, we need to sample from the learned model, what requires the inversion of the fitted chain of transformations f^{-1}. Because there is no closed-form solution for the inversion of higher-order Bernstein polynomials, we use a root finding algorithm to determine the inverse (Chandrupatla, 1997).

3. Load Forecasting Simulation Study

This section evaluates the use of BNF approach for load forecasting in an empirical study. First, the forecasting approach and benchmark methods are introduced, then the dataset is described, and finally the results are discussed.

3.1. Probabilistic Forecasting Models

In the study we compare the combinations of two different NN architectures with four different methods to model the CPDs. The networks control the parameters θ needed for the different CPDs (see Figure 3). For the NNs, we compare:

A *fully connected neural network* (FC) with three hidden layers, with 100 units, the ELU activation function (Clevert et al., 2016), and batch normalization (Ioffe & Szegedy, 2015). The historical data was flattened and concatenated with the metadata. This model is not ideal for processing the sequential data but serves as a simple baseline model to compare against the more sophisticated CNN model.

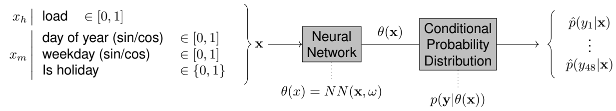

Figure 3. Overall model: The variables on the left side are the input to a NN (FC or 1DCNN). The NN controls the parameters $\theta(x)$ for the respective model (BNF, GM, GMM, or QR) yielding the marginal CPD $p(y_t|\theta(X))$ for 48 time steps t.

A *dilated 1D-Convolution Network architecture* (1DCNN) inspired by the WaveNet architecture (van den Oord et al., 2016). The model was build by stacking 8 1D convolutional layers with doubling dilatation rates $1, 2, 4, \dots, 128$. This results in a model with a receptive field capturing 256 input values. Hence, almost the whole input sequence, consisting of one week of historical load data with a total of 348 input features, can be processed. Each of these dilated convolutions uses the ReLU activation function and has 20 filters. Finally, a regular 1D convolutional layer without dilatation, again ReLU activation and ten filters. The output of this last convolutional layer is then flattened and concatenated with the meta data before it is fed into a final fully connected layer with ELU activation function followed by a last layer without activation function to generate the output. Batch normalization was not used in this model.

For predicting the CPDs this study proposes the *Bernstein-Polynomial Normalizing Flow (BNF)* in which θ are the parameters of the Bernstein polynomial of order $M = 20$ as well as the additional linear transformations making up the flow f (see Figure 2).

We compare the BNF with the following benchmarks:

A *simple Gaussian Model (GM)* as a probabilistic extension of regular regression, predicting not only the conditioned mean $\mu(x)$, but also the conditional variance $\sigma^2(x)$ (e.g., Haben et al., 2019; Wijaya et al., 2015; Salinas et al., 2020).

To model more complex distribution shapes, a *Gaussian Mixture Model (GMM)* mixing three normal distributions was implemented. The output vector θ contains the mean and variance $\mu_k(x), \sigma_k^2(x)$ and the mixing coefficients $\alpha_k(x)$ for $k = 1, 2, 3$ (see, e.g., Vossen et al., 2018).

A *Quantile Regression (QR)* predicting 100 quantiles has been implemented as a typical baseline in probabilistic load forecasting (cf., e.g., Elvers et al., 2019; Wang et al., 2019). The 100 quantiles for each time step have been constrained to be monotonically increasing by applying a softplus activation function and then calculating the cumulative sum. Note that strictly speaking the QR is not a continuous CPD, hence the NLL is not tractable, and instead the pinball loss is minimized.

3.2. Dataset Description

The models were trained on a dataset containing electricity demand information for smart meter customers in Ireland, recorded in the period from *2009/07/14* until *2010/12/31*, in a resolution of 30 Minutes (Commission for Energy Regulation (CER), 2012). All non-residential buildings were dropped, since the stochastic behavior of residential customers was from explicit interest in this study. Additionally, all incomplete records have been removed. A random subset of 10% (363 customers) was extracted. All records until *2010/10/31 23:30:00* have been used for training[1], the remaining readings were left out for testing. The scripts used to preprocess the data set and conduct the experiments are available at GitHub[2]. At runtime, the data is shuffled and batched into mini batches of size 32. Each sample consists of an input tuple $\mathbf{x} = (x_h, x_m)$ containing the *historical* data x_h, with the lagged electric load of the past seven days and *meta* data x_m, with trigonometric encoded time information as done by Andrich van Wyk (2018) and a binary holiday indicator as indicated in Figure 3. The prediction target $\mathbf{y}$ is the load for the next day, with resolution of 30 minutes. Hence, the model predicts 48 CPDs $p(y_1|\mathbf{x}), \dots, p(y_{48}|\mathbf{x})$ for every future time step.

3.3. Discussion of Results

Table 1 summarized our results for the different architectures and probabilistic models. All models were trained with early stopping for maximal 300 epochs with the Adam optimizer (Kingma & Ba, 2017).

Arch	CPD	CRPS	NLL	MAE	MSE
FC	BNF	**0.021**	**−123.2**	0.413	9.430
	GMM	0.024	−116.0	0.368	0.438
	GM	0.951	−77.0	0.742	23.464
	QR	0.026	–	0.390	0.615
1DCNN	BNF	**0.017**	**−132.089**	0.342	0.429
	GMM	0.019	−125.933	0.384	0.45
	GM	0.018	−101.29	0.347	**0.366**
	QR	**0.017**	–	**0.321**	0.399

Table 1. Results of empirical experiments, for both *architectures* and all four CPD models (subsection 3.1). The table shows the Continous Ranked Probability Score (CRPS), the Negative Logarithmic Likelihood (NLL), the Mean Absolute Error (MAE) and the Mean Squared Error (MSE). Lower is better.

[1] the last 10% were used for validation during the development

[2] `https://github.com/MArpogaus/stplf-bnf`

Probabilistic models should be evaluated by a score that takes the whole CPD into account, like the strictly proper scores Continous Ranked Probability Score (CRPS) or Negative Logarithmic Likelihood (NLL) (Bröcker & Smith, 2007; Gneiting & Raftery, 2007). These scores are reported in Table 1, in addition, we also included the common point scores MAE and MSE. We observe two things: First, using dilated convolutional layers helps to process sequential data. For all CPDs the CRPS and the NLL are superior in the 1DCNN architecture compared to the FC network. Second, the scores are better the more flexible the CPD. In general, the GMM is better than the GM and the flexible BNF and QR are always better than the GMM. We speculate that the performance gain of the BNF compared to the QR in the FC case is because the BNF is less prone to overfitting since using the Bernstein basis is smoother compared to estimating 100 quantiles, only restricted by the order. An additional benefit of the BNF over the QR is that it naturally provides a continuous distribution.

4. Conclusion

Forecasting at the LV level is becoming essential for many stakeholders, while more and more applications in low carbon energy systems are explored. Due to high volatility of load profiles, probabilistic load forecasts are an emerging research topic, as they are capable of expressing uncertainties introduced by the fluctuations caused by the increasing penetration of renewable and distributed energy sources. The majority of probabilistic load forecasting literature focuses on parametric or QR approaches for estimating marginal distributions. The conditioned probability distribution remains unknown. Parametric methods estimate the distribution by making assumptions of the underlying distribution, while QR can only provide a discrete approximation of the full distribution. Instead, the proposed probabilistic deep learning model uses a cascade of transformation functions, known as normalizing flows, to directly model the conditioned probability density. Model parameters are obtained by minimizing the NLL directly through gradient descent.

We demonstrated that BNFs are a very powerful and stable method to express complex non-Gaussian distributions, with almost no regularization or special tuning. This makes them a preferential choice over the QRs or Gaussian approaches for probabilistic load forecasts. BNFs are also applicable for other use cases like anomaly detection or generation of synthetic scenarios for grid planning. A possible enhancement might be to take the multivariate nature of the forecast more directly into account. Instead of predicting multiple marginal CPDs for a fixed forecast horizon, a future implementation could benefit from autoregressive architectures for non-fixed forecast horizons or extending the BNF to multivariate versions.

Acknowledgments

Parts of this work have been funded by the Federal Ministry for the Environment, Nature Conservation and Nuclear Safety due to a decision of the German Federal Parliament (AI4Grids: 67KI2012A) and by the German government (WindNODE: 03SIN539).

References

Andrich van Wyk. Encoding Cyclical Features for Deep Learning. https://www.avanwyk.com/encoding-cyclical-features-for-deep-learning/, April 2018.

Arora, S. and Taylor, J. Forecasting electricity smart meter data using conditional kernel density estimation. *Omega*, 59:47–59, 2016.

Bröcker, J. and Smith, L. A. Scoring Probabilistic Forecasts: The Importance of Being Proper. *Weather and Forecasting*, 22(2):382–388, April 2007. ISSN 0882-8156. doi: 10.1175/WAF966.1.

Chandrupatla, T. R. A new hybrid quadratic/bisection algorithm for finding the zero of a nonlinear function without using derivatives. *Advances in Engineering Software*, 28(3):145–149, April 1997. ISSN 0965-9978. doi: 10/djqsx4.

Clevert, D.-A., Unterthiner, T., and Hochreiter, S. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). *arXiv:1511.07289 [cs]*, February 2016.

Commission for Energy Regulation (CER). CER Smart Metering Project - Electricity Customer Behaviour Trial. `http://www.ucd.ie/issda/data/commissionforenergyregulationcer/`, 2012.

Elvers, A., Voß, M., and Albayrak, S. Short-Term Probabilistic Load Forecasting at Low Aggregation Levels Using Convolutional Neural Networks. In *2019 IEEE Milan PowerTech*, pp. 1–6, Milan, Italy, June 2019. IEEE. ISBN 978-1-5386-4722-6. doi: 10.1109/PTC.2019.8810811.

Fadlullah, Z., Fouda, M., Kato, N., Shen, X., and Nozaki, Y. An early warning system against malicious activities for smart grid communications. *IEEE Network*, 25(5):50 – 55, 2011.

Fenza G., Gallo M., L. V. Drift-aware methodology for anomaly detection in smart grid. *IEEE Access*, 7:9645 – 9657, 2019.

Gerossier, A., Girard, R., Bocquet, A., and Kariniotakis, G. Robust day-ahead forecasting of household electricity demand and operational challenges. *Energies*, 11(12), 2018. ISSN 1996-1073.

Gneiting, T. and Raftery, A. E. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102(477):359–378, March 2007. ISSN 0162-1459. doi: 10/c6758w.

Haben, S., Giasemidis, G., Ziel, F., and Arora, S. Short term load forecasting and the effect of temperature at the low voltage level. *International Journal of Forecasting*, 35 (4):1469–1484, 2019.

Haben, S., Arora, S., Giasemidis, G., Voss, M., and Greetham, D. V. Review of low-voltage load forecasting: Methods, applications, and recommendations. *arXiv:2106.00006v1 [stat]*, 2021.

He, B., Li, J., Tsung, F., Gao, Y., Dong, J., and Dang, Y. Monitoring of power consumption requirement load process and price adjustment for smart grid. *Computers & Industrial Engineering*, 137, 2019. ISSN 0360-8352. doi: https://doi.org/10.1016/j.cie.2019.106068.

Hermanns, J., Modemann, M., Korotkiewicz, K., Paulat, F., Kotthaus, K., Pack, S., and Zdrallek, M. Evaluation of different development possibilities of distribution grid state forecasts. *Energies*, 13(8), 2020. ISSN 1996-1073. doi: 10.3390/en13081891.

Hong, T. and Fan, S. Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting*, 32(3):914 – 938, 2016. ISSN 0169-2070. doi: https://doi.org/10.1016/j.ijforecast.2015.11.011.

Hothorn, T., Möst, L., and Bühlmann, P. Most Likely Transformations. *Scandinavian Journal of Statistics*, 45(1):110–134, 2018. ISSN 1467-9469. doi: 10.1111/sjos.12291.

Ioffe, S. and Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv:1502.03167 [cs]*, March 2015.

Kingma, D. P. and Ba, J. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*, January 2017.

Kingma, D. P. and Dhariwal, P. Glow: Generative Flow with Invertible 1x1 Convolutions. *arXiv:1807.03039 [cs, stat]*, July 2018.

Morstyn, T., Farrell, N., Darby, S. J., and McCulloch, M. D. Using peer-to-peer energy-trading platforms to incentivize prosumers to form federated power plants. *Nature Energy*, 3(2):94–101, 2018.

Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S., and Lakshminarayanan, B. Normalizing Flows for Probabilistic Modeling and Inference. *arXiv:1912.02762 [cs, stat]*, December 2019.

Pinto, R., Bessa, R. J., and Matos, M. A. Multi-period flexibility forecast for low voltage prosumers. *Energy*, 141:2251–2263, 2017. ISSN 0360-5442. doi: https://doi.org/10.1016/j.energy.2017.11.142.

Ponoćko, J. and Milanović, J. V. Forecasting demand flexibility of aggregated residential load using smart meter data. *IEEE Transactions on Power Systems*, 33(5):5446–5455, 2018.

Ramasinghe, S., Fernando, K., Khan, S., and Barnes, N. Robust normalizing flows using Bernstein-type polynomials. *arXiv:2102.03509 [cs, stat]*, February 2021.

Rasul, K., Sheikh, A.-S., Schuster, I., Bergmann, U., and Vollgraf, R. Multi-variate Probabilistic Time Series Forecasting via Conditioned Normalizing Flows. *arXiv:2002.06103 [cs, stat]*, February 2020.

Rothfuss, J., Ferreira, F., Boehm, S., Walther, S., Ulrich, M., Asfour, T., and Krause, A. Noise Regularization for Conditional Density Estimation. *arXiv:1907.08982 [cs, stat]*, February 2020.

Rowe, M., Yunusov, T., Haben, S., Singleton, C., Holderbaum, W., and Potter, B. A peak reduction scheduling algorithm for storage devices on the low voltage network. *IEEE Trans. on Smart Grid*, 5:2115–2124, 2014.

Salinas, D., Flunkert, V., Gasthaus, J., and Januschowski, T. Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3):1181–1191, 2020.

Sick, B., Hothorn, T., and Dürr, O. Deep transformation models: Tackling complex regression problems with neural network based transformation models. *arXiv:2004.00464 [cs, stat]*, April 2020.

Trippe, B. L. and Turner, R. E. Conditional Density Estimation with Bayesian Normalising Flows. *arXiv:1802.04908 [stat]*, February 2018.

van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., and Kavukcuoglu, K. WaveNet: A Generative Model for Raw Audio. *arXiv:1609.03499 [cs]*, September 2016.

van den Oord, A., Li, Y., Babuschkin, I., Simonyan, K., Vinyals, O., Kavukcuoglu, K., van den Driessche, G., Lockhart, E., Cobo, L. C., Stimberg, F., Casagrande, N., Grewe, D., Noury, S., Dieleman, S., Elsen, E., Kalchbrenner, N., Zen, H., Graves, A., King, H., Walters, T., Belov, D., and Hassabis, D. Parallel WaveNet: Fast High-Fidelity Speech Synthesis. *arXiv:1711.10433 [cs]*, November 2017.

Vossen, J., Feron, B., and Monti, A. Probabilistic Forecasting of Household Electrical Load Using Artificial Neural Networks. In *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pp. 1–6, June 2018. doi: 10.1109/PMAPS.2018.8440559.

Wang, Y., Gan, D., Sun, M., Zhang, N., Lu, Z., and Kang, C. Probabilistic individual load forecasting using pinball loss guided lstm. *Applied Energy*, 235:10–20, 2019. ISSN 0306-2619. doi: https://doi.org/10.1016/j.apenergy.2018.10.078.

Wijaya, T. K., Sinn, M., and Chen, B. Forecasting uncertainty in electricity demand. In *AAAI-15 Workshop on Computational Sustainability*, 2015.

World Resources Institute. World greenhouse gas emissions in 2016, 2020. URL `https://www.wri.org/data/world-greenhouse-gas-emissions-2016`.

Zufferey, T., Renggli, S., and Hug, G. Probabilistic state forecasting and optimal voltage control in distribution grids under uncertainty. *Electric Power Systems Research*, 188:106562, 2020.