# Probabilistic Forecasting of Household Loads: Effects of Distributed Energy Technologies on Forecast Quality


Frederik vom Scheidt*
frederik.scheidt@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Xinyuan Dong
dongxinyuan1994@gmail.com
Karlsruhe Institute of Technology
Karlsruhe, Germany

Andrea Bartos
andrea.bartos@student.kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Philipp Staudt
philipp.staudt@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Christof Weinhardt
christof.weinhardt@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany



## ABSTRACT

Distributed energy technologies introduce new volatility to the edges of low voltage grids and increase the importance of short-term forecasting of electric loads at a granular level. To address this issue, first probabilistic forecasting models for residential loads have been developed in recent years. However, knowledge is lacking about how well these models perform for households with different endowments of distributed energy technologies. Therefore, we first create a new semi-synthetic data set which contains not only conventional residential loads, but net loads of 40 households differentiated regarding heating type (electric space heating, no electric space heating), and rooftop solar installation (solar, no solar). Second, we develop a novel probabilistic forecasting model based on Gated Recurrent Units that uses data from weather forecasts and calendar variables as external features. We apply the developed model, and three benchmarks, to the new data set and find that the GRU model outperforms the other models for households with electric heating, with solar, and with both technologies, but not for households without distributed energy technologies.



## CCS CONCEPTS

• **Computing methodologies → Neural networks**.

## KEYWORDS

Probabilistic forecasting, Distributed energy technologies, Gated recurrent units, Net load forecasting


## 1 INTRODUCTION

In electricity systems with increasing numbers of distributed energy technologies (DETs), appropriate short-term forecasting of loads at a granular level gains importance [36, 39]. Probabilistic load forecasting models for household loads provide more information about future uncertainties than point forecasts [39], but have been focused on 'conventional' residential loads, and have largely neglected the influence of distributed energy technologies. Therefore, this study makes the following contributions. First, we develop a new semi-synthetic residential data set which contains the net load profiles of 40 households, differentiated by heating type (electric space heating, no electric space heating), and rooftop solar installation (solar, no solar). This unique data set allows us to analyze how well probabilistic forecasting models perform over various types of households with DETs. Second, we present a probabilistic forecasting model based on Gated Recurrent Units (GRU) that includes data from weather forecasts and calendar variables as external features. We compare this model to three benchmark models, one of them a recently proposed model based on Long Short-term Memory (LSTM) networks [40]. Our work thus sheds light on the so-far neglected role of DETs in residential probabilistic load forecasting, proposes a new forecasting model that is compared to state-of-the-art benchmark models, and provides a new benchmark data set for future research in this area.

## 2 RELATED WORK

Unlike point forecasting which outputs a single predicted value at each time step, probabilistic forecasting makes it possible to express the uncertainty in a prediction, which is a crucial component for optimal decision making [15]. For electrical load, probabilistic forecasting generates a distribution of the future load, thus capturing characteristics of a load profile's volatility. As suggested by Hong and Fan [21], probabilistic load forecasts can be conducted in terms of quantile forecasting, interval forecasting and density forecasting. In recent years, there has been growing interest in probabilistic load forecasting on city level or system level. A structured overview of probabilistic load forecasting studies is provided in Table 3 in the appendix. The overview shows that studies have investigated a wide range of new methods, including kernel methods [4], neural networks [9, 13, 35, 38, 40], Gaussian process [32, 34, 35], additive

[1]Code and data are available at https://github.com/FVS-energy/prob_forecasting

quantile regression [33], and ensemble models [2, 29]. However, most studies have applied these methods to regular households' loads. Regarding household loads influenced by DETs, Van der Meer et al. [35] address probabilistic forecasting of net loads of houses with rooftop solar. They propose a dynamic Gaussian Process that produces sharper prediction intervals at significant lower computational effort than the provided benchmarks. However, there is a trade-off with the ability to capture sharp peaks. The authors also find that indirectly forecasting net demand (i.e. through forecasting both demand and own generation) leads to wider prediction intervals with higher coverage probability. In a following work, Van der Meer et al. [34] find that net load forecasts have improved sharpness and reliability of prediction intervals, when several households are aggregated. This hints at the specific challenges of individual load forecasting. However, neither of the two studies addresses individual net load forecasts for households without solar or with other technologies. Besides probabilistic households load forecasting, studies have developed specialized probabilistic forecasts for flexibilities of electric vehicles [24], or rooftop solar generation [1], but without integrating these into residential load forecasts.

In summary, past research has focused on the development of models for household loads, or specific single technologies. An important gap prevails regarding probabilistic load forecasting for consumers with different types of distributed energy technologies that are about to "disrupt the traditional load profiles" [39].

# 3 METHODOLOGY

This section introduces the structure of the proposed forecasting model. Additionally, it describes benchmark methods, the selected loss metric, hyperparameter tuning, and cross-validation.

## 3.1 Long Short-Term Memory Networks and Gated Recurrent Units

Unlike traditional neural networks, which learn the relationships of inputs and outputs based on provided training data for every instance, recurrent neural networks (RNNs) are able to learn dependencies within sequential input data such as time series. However, conventional RNNs can practically only learn short-term dependencies due to the problem of vanishing gradients.

As a remedy, Long Short-Term Memory networks have been developed, which are able to learn long-term dependencies [14, 19]. There are three gates in an LSTM unit, which control the flow of information in the network. However, LSTMs can suffer from slow training since parameters for three gates have to be estimated.

Therefore, two-gate based Gated Recurrent Units (GRUs) have been developed. Compared to LSTMs, GRUs have one less parameter that needs to be estimated. In other contexts, GRUs have shown similar performance as LSTMs, with shorter computational times [8, 27]. Therefore, in this paper, we present a probabilistic forecasting model based on quantile forecasts with GRUs.

## 3.2 Network structure

The quantile GRU model (QGRU) includes four steps. A visual representation is displayed in the appendix in Figure 2. The first step takes historical load data as input. $t$ refers to the predicted time step. $n_i$ is the number of input steps, i.e. the length of the input time series. $n_o$ denotes the number of output steps, defining the prediction horizon. The previous load profile goes through two GRU layers and one dense layer. This layer passes on the last hidden state $h_t$. In the second step, the calendar data, i.e. the features weekday and time of day, are one-hot encoded. In the third step, weather data is introduced. Each variable, i.e. temperature, wind speed, and relative humidity, is normalized. The fourth step concatenates the output of the three previous steps. Then, the resulting input vector is passed through two fully-connected dense layers, which finally generate five quantile forecasts.

## 3.3 Pinball loss for quantile forecasting

Pinball loss is an established evaluation metric for probabilistic forecasts in the energy sector. It is used as the single deciding error measure in the Global Energy Forecasting Competition [22], as well as in studies on probabilistic household load forecasting [9, 40]. Pinball loss evaluates the forecasts of each quantile individually, as formulated by Equation 1. The core idea of the pinball loss is the asymmetric penalization of forecast errors, depending on the quantile. If a forecasted quantile value is smaller than the actual observation, this will be penalized stronger for higher quantiles, as the quantile loss is the product of the quantile and the absolute error. The proposed QGRU model (and all benchmark models) predict five values at each time step for $q \in [10\%, 25\%, 50\%, 75\%, 90\%]$. The aim of the training process is to minimize the average pinball loss of all five quantiles, as formulated by Equation 2.

$$L_{q,t}(y_t, \hat{y}_t^q) = \begin{cases} (1-q)(\hat{y}_t^q - y_t), & \hat{y}_t^q \geq y_t \\ q(y_t - \hat{y}_t^q), & \hat{y}_t^q < y_t \end{cases} \tag{1}$$

$y_t$: real observation at time step $t$
$\hat{y}_t^q$: the $q$th quantile forecast at time step $t$

$$minL = \sum_q \sum_{t=1}^{T} L_{q,t}, q \in [10\%, 25\%, 50\%, 75\%, 90\%] \tag{2}$$

## 3.4 Benchmarks

To adequately evaluate the performance of the proposed model, we compare it to three other quantile load forecasting models. The first benchmark is a quantile LSTM (QLSTM) model. It has the same network structure as the QGRU model, but employs LSTM layers instead of GRU layers. The second benchmark is a quantile regression neural network (QREGNN) model with four dense layers. Both QLSTM and QREGNN take the same input features as the QGRU. We use these two benchmarks to evaluate the QGRU's performance against other models using the same input data. The third benchmark is a quantile LSTM model without weather input features (QLSTM_noWeather). It corresponds to the model proposed by Wang et al. [40]. We use it to measure how the weather features affect forecast loss and to provide an established benchmark. For all three benchmark models, hyperparameter tuning and cross-validation is performed, to allow for an adequate comparison.

## 3.5 Hyperparameter tuning

Since there are individual models for each household, hyperparameter tuning is done for each model individually. For QGRU, QLSTM, and QLSTM_noWeather, we tune the learning rate, the number of units in the recurrent layers and the number of units in the dense layers. For the QREGNN, we tune the learning rate and the number of units in the dense layers. The tested values are shown in Table 1.

**Table 1: Values of Hyperparameters**

| Hyperparameter | Values |
|---|---|
| Learning rate | 0.001, 0.01, 0.1 |
| Number of units in recurrent layers | 4, 8, 12 |
| Number of units in dense layers | 10, 30, 50 |

### 3.6 Cross-validation

Due to the time series character and the limited time range of the data (one year), a two-fold rolling window approach is conducted to cross-validate the models. First, we train, validate and test our models only on the first 80% of data, i.e. with a train-validation-test split of 40-20-20. Second, we expand the training window, resulting in a 60-20-20 split. For the final evaluation, we average the test losses from step one and two. Rolling window cross-validation is a common approach in energy forecasting [25, 35]. It improves the generalisation of our findings, amongst others because it better captures any seasonality effects.

## 4 DATA

This section introduces the data pre-processing steps and the final input data set for the load forecasting task.

### 4.1 Load data

The residential electricity load data was collected by Commonwealth Edison (ComEd), a large electric utility in the US [10]. The data set contains anonymous smart meter data of residential customers in and around the city of Chicago for the year of 2016. Each smart meter provides half-hourly load data, which leads to 16,128 observations for every household. For each customer ID, the delivery service class is stated, which describes the housing type (single family homes and multi family homes), as well as the heating type (electric space heating and no electric space heating). For more information and other applications of the original data set, we refer to [7, 37]. For our purpose, we focus on households in single family homes. From those, we randomly draw ten customers with electric space heating and ten without electric space heating.

### 4.2 Solar data

For solar data, we use the Python tool pvlib [20]. Following the approach from [5, 6], we simulate power generation from rooftop solar systems based on given weather and irradiation data from 2016. We simulate PV systems with three different azimuths of 135 (south-east), 180 (south), and 225 (south-west) degrees. Each solar system is sized to a capacity of 6.9 kW, following [12]. We randomly assign an azimuth to each household. The assigned solar generation curve is subtracted from the load curve, resulting in the net load[2].

[2]Our approach assumes that households with rooftop solar do not change their electricity consumption behavior, e.g. to specifically use self-generated electricity. This is motivated by the fact that they have no financial incentive to do so under the prevailing flat tariff net metering regulation scheme.

Finally, the data set comprises the net load data for 40 households, i.e. for ten households without electric space heating or solar (Figure 3a in the appendix), for ten households with electric space heating, but no solar (Figure 3b), for ten households without electric space heating, but solar (Figure 3c), and for ten households with both electric space heating and solar (Figure 3d).

### 4.3 Weather data

When using external input features for forecasting, it should be ensured that only data is used that in reality would be available at the time of forecasting [36]. It has been shown that the errors inherent in weather forecasts increase load forecast errors on system level [3]. Therefore, we acquire historical *weather forecast* data for 2016.

The US ”National Oceanic and Atmospheric Institute” provides historical weather forecasts with a sufficiently long history via the Climate Forecast System Version 2 [31]. There, forecasts for more than 50 variables are stored in six hour intervals. We select air temperature, specific humidity and wind speed, since they are the most frequently used weather variables for load forecasting [11] and have a large influence on thermal comfort [18], which presumably play a crucial role for the load forecasts of households with electric heating. The data sets are provided via a HTTPS file server [30]. From this server, the data can be downloaded in the Grib2 format and transformed for use in the forecasting task with the Python Package cfgrib. The grid node with the closest spatial proximity to the smart meter area is selected (42.05°, -87.2°). Last, we interpolate the data to hourly values using a cubic regression spline as proposed by Hyndman and Fan [26].

### 4.4 Calendar data

Electricity consumption patterns on public holidays are usually different from normal days [17]. Past studies have either simply considered all public holidays as weekends [28] or used more sophisticated rules to label public holidays and surrounding days [23]. We follow the latter approach and re-label special days accordingly.

### 4.5 Final input data set

Since other data is given at hourly intervals, the smart meter data sets are aggregated from a half-hourly to an hourly resolution, resulting in 24 observations per day. The final data set for the case study contains three categorical variables (Customer ID, DET setup, Date), and six input variables for the forecast (Hourly net load, Weekday, Time of day, Hourly temperature forecast, Hourly wind speed forecast, Hourly relative humidity forecast).

## 5 CASE STUDY

For the case study, we forecast the one-hour ahead net load. For each of the 40 households, an individual model is trained. We use the last 336 hours (i.e. two weeks) of net load as lagged input features. The proposed models are implemented in Python, using Keras. The models are run on a GPU on Google Colaboratory [16].

To provide a detailed insight into the performance of the models, Figure 1 shows a scatter plot of each household's average pinball loss under QGRU, compared to the benchmark models. The line y=x represents the performance of the QGRU. All points under this line indicate a case in which the QGRU outperforms the respective

Table 2: Average Pinball loss [kWh] of tested methods for different customer types

| | QGRU | QLSTM | QREGNN | QLSTM_noWeather | Average |
|---|---|---|---|---|---|
| Household | 0.1989 | 0.2019 | 0.2373 | 0.1902 | 0.2070 |
| Household with heating | 0.2060 | 0.2116 | 0.2602 | 0.2061 | 0.2211 |
| Household with solar | 0.1366 | 0.1386 | 0.1367 | 0.1387 | 0.1376 |
| Household with heating and solar | 0.1347 | 0.1394 | 0.1564 | 0.1509 | 0.1453 |

benchmark model. 67.5% of points are under the line, demonstrating the overall superior performance of the proposed QGRU model. More specifically, the QGRU model outperforms the QLSTM in 60.0%, the QREGNN in 72.5%, and the QLSTM_noWeather in 70.0% of cases.

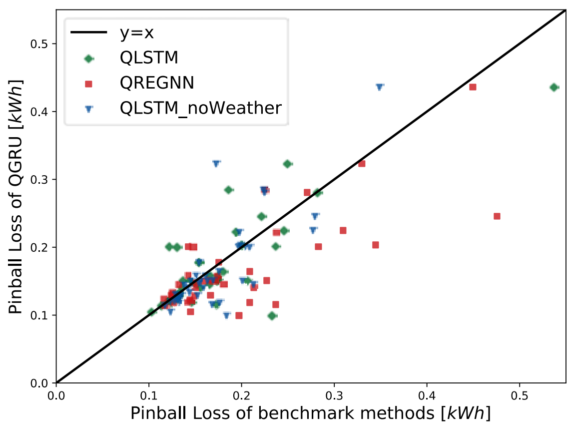



**Figure 1: Comparison of pinball loss between QGRU and benchmark models for all customers**

In Table 2, we present the performance of the proposed QGRU and the three benchmarks methods, averaged across customers. The proposed QGRU achieves the lowest pinball losses overall. It achieves the lowest loss for three of the four customer groups, namely households with electric heating, households with solar, and households with both technologies. Only in the case of households without any technology, the benchmark QLSTM model without additional weather input data (as proposed by Wang et al. [40]) outperforms the QGRU on average. Our results thus confirm the good performance of this model on standard households, but also show that it is outperformed by the QGRU model for households with energy technologies. This underlines the importance of tailoring forecasting models to the specific case.

All models show the highest average loss for households with electric heating. This might be due to the higher total load of those households. Figure 4 in the appendix shows the average pinball loss versus the annual net consumption for each households. It shows that households in the data set with electric heating have exceptionally high loads. Although differences among models exist, all models show a very high pinball loss for at least one customer from the set of households with heating. This finding might indicate that for the tested models, a training set of less than one year which only includes one heating period is inadequate for learning the households' heating behavior, which is important to consider in the development of future models.

All models achieve the lowest pinball loss on the net load profiles of households with solar generation. Notably, this finding seems to hold independent of these households' total annual net loads, as Figure 4 shows. This is surprising and indicates that the forecasting models are able to adequately capture the periodicity of net loads that include solar generation. This notion is supported by the comparison of households with no technology and of households with both solar and heating. These groups have similar mean net consumption, but pinball losses are lower for the latter.

Since no literature on probabilistic forecasts of net loads with DET influence exist, we cannot yet benchmark our results against literature. However, we compare our results for households without DETs to Wang et al. [40] who also use a pinball loss guided LSTM without weather data. We find that the QLSTM_noWeather model in our case study achieves an average pinball loss about twice as high as in the case study in [40]: 0.2019, compared to 0.0963. We assume this difference is due to the higher number of data points per customer in the data set used in [40]: 26,000 data points per customer, compared to 8,783 in our data set. This indicates the positive effect of more training data on forecasting performance.

Future work could enhance our approach by including other distributed energy technologies, such as electric vehicles, and residential batteries. Besides, it can utilize the forecasts by integrating them into the operation of home energy management systems. For this, the code and data published with this study can be used.

## 6 CONCLUSION

In this paper, we argue that increasing adoption of distributed energy technologies will affect the quality of existing forecasting tools for individual households' net loads. We present a pinball loss guided GRU model that produces quantile forecasts of net loads. We develop a new, semi-synthetic residential net load data set that includes standard customers without distributed energy technologies as well as customers with electric heating, rooftop solar, and both technologies. We apply the proposed model and three benchmark models to this data set. We find that the proposed quantile GRU model outperforms the benchmark models for customers with distributed energy technologies, independent of technology. However, the quantile GRU model is outperformed for the group of standard households by a quantile LSTM model that ignores weather data. All models perform best for households with own solar generation, and worst for households with electric heating. We thus provide first fundamental insights for probabilistic forecasting of household loads under the influence of distributed energy technologies.

# ACKNOWLEDGMENTS

We thank David Dann for his help on efficiently structuring the code project. We thank Niklas Bäckel for his help with data acquisition.

# REFERENCES

[1] Mousa Afrasiabi, Mohammad Mohammadi, Mohammad Rastegar, and Shahabodin Afrasiabi. 2020. Deep learning architecture for direct probability density prediction of small-scale solar generation. *IET Generation, Transmission & Distribution* 14, 11 (2020), 2017–2025.

[2] M. Afrasiabi, M. Mohammadi, M. Rastegar, L. Stankovic, S. Afrasiabi, and M. Khazaei. 2020. Deep-Based Conditional Probability Density Function Forecasting of Residential Loads. *IEEE Transactions on Smart Grid* 11, 4 (2020), 3646–3657. https://doi.org/10.1109/TSG.2020.2972513

[3] Alireza Khotanzad, Reza Afkhami-Rohani, Tsun-Liang Lu, Alireza Abaye, Malcolm Davis, and Dominic J. Maratukulam. 1997. ANNSTLF-a neural-network-based electric load forecasting system. *IEEE Transactions on Neural Networks* 8, 4 (1997), 835–846. https://doi.org/10.1109/72.595881

[4] Siddharth Arora and James W Taylor. 2016. Forecasting electricity smart meter data using conditional kernel density estimation. *Omega* 59 (2016), 47–59.

[5] Patrick R Brown and Francis M O'Sullivan. 2020. Spatial and temporal variation in the value of solar power across United States electricity markets. *Renewable and Sustainable Energy Reviews* 121 (2020), 109594.

[6] Scott Burger. 2019. Rate Design for the 21st Century: Improving Economic Efficiency and Distributional Equity in Electricity Rate Design.

[7] Scott P Burger, Christopher R Knittel, Ignacio J Pérez-Arriaga, Ian Schneider, and Frederik Vom Scheidt. 2020. The efficiency and distributional effects of alternative residential electricity rate designs. *The Energy Journal* 41, 1 (2020). https://doi.org/10.5547/01956574.41.1.sbur

[8] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555* (2014).

[9] A. Elvers, M. Voß, and S. Albayrak. 2019. Short-Term Probabilistic Load Forecasting at Low Aggregation Levels Using Convolutional Neural Networks. In *2019 IEEE Milan PowerTech*. 1–6. https://doi.org/10.1109/PTC.2019.8810811

[10] Exelon. 2018. *Our companies - ComEd*. http://www.exeloncorp.com/companies/comed

[11] Eugene A Feinberg and Dora Genethliou. 2005. Load forecasting. In *Applied mathematics for restructured electric power systems*. Springer, 269–285.

[12] David Feldman, Vignesh Ramasamy, and Robert Margolis. 2021. *U.S. Solar Photovoltaic BESS System Cost Benchmark Q1 2020 Report*. Technical Report. National Renewable Energy Laboratory. https://data.nrel.gov/submissions/158

[13] Dahua Gan, Yi Wang, Ning Zhang, and Wenjun Zhu. 2017. Enhancing short-term probabilistic residential load forecasting with quantile long–short-term memory. *The Journal of Engineering* 2017, 14 (2017), 2622–2627.

[14] Felix A. Gers, Jürgen Schmidhuber, and Fred Cummins. 1999. Learning to forget: Continual prediction with LSTM. *9th International Conference on Artificial Neural Networks: ICANN '99* (1999).

[15] Tilmann Gneiting and Matthias Katzfuss. 2014. Probabilistic forecasting. *Annual Review of Statistics and Its Application* 1 (2014), 125–151.

[16] Google. 2021. Colaboratory. https://colab.research.google.com/

[17] Wan He. 2017. Load forecasting via deep neural networks. *Procedia Computer Science* 122 (2017), 308–314.

[18] Henrique Steinherz Hippert, Carlos Eduardo Pedreira, and Reinaldo Castro Souza. 2001. Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on power systems* 16, 1 (2001), 44–55.

[19] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[20] William F Holmgren, Clifford W Hansen, and Mark A Mikofski. 2018. pvlib python: A python package for modeling solar energy systems. *Journal of Open Source Software* 3, 29 (2018), 884.

[21] Tao Hong and Shu Fan. 2016. Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting* 32, 3 (2016), 914–938.

[22] Tao Hong, Pierre Pinson, Shu Fan, Hamidreza Zareipour, Alberto Troccoli, and Rob J Hyndman. 2016. Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond.

[23] Tao Hong, Jason Wilson, and Jingrui Xie. 2013. Long term probabilistic load forecasting and normalization with hourly information. *IEEE Transactions on Smart Grid* 5, 1 (2013), 456–462.

[24] Julian Huber, David Dann, and Christof Weinhardt. 2020. Probabilistic forecasts of time and energy flexibility in battery electric vehicle charging. *Applied Energy* 262 (2020), 114525. https://doi.org/10.1016/j.apenergy.2020.114525

[25] Julian Christoph Huber. 2020. *Engineering User-Centric Smart Charging Systems*. Ph.D. Dissertation. Karlsruher Institut für Technologie (KIT). https://doi.org/10.5445/IR/1000122898

[26] Rob J Hyndman and Shu Fan. 2009. Density forecasting for long-term peak electricity demand. *IEEE Transactions on Power Systems* 25, 2 (2009), 1142–1153.

[27] Kang Ke, Sun Hongbin, Zhang Chengkang, and Carl Brown. 2019. Short-term electrical load forecasting method based on stacked auto-encoding and GRU neural network. *Evolutionary Intelligence* 12, 3 (2019), 385–394.

[28] Peter Lusis, Kaveh Rajab Khalilpour, Lachlan Andrew, and Ariel Liebman. 2017. Short-term residential load forecasting: Impact of calendar effects and forecast granularity. *Applied Energy* 205 (2017), 654–669.

[29] Joakim Munkhammar, Dennis van der Meer, and Joakim Widén. 2021. Very short term load forecasting of residential electricity consumption using the Markov-chain mixture distribution (MCM) model. *Applied Energy* 282 (2021), 116180. https://doi.org/10.1016/j.apenergy.2020.116180

[30] National Oceanic and Atmospheric Institute. 2019. *Modeldata*. https://nomads.ncdc.noaa.gov/modeldata/cfsv2_forecast_ts_9mon/

[31] Suranjana Saha, Shrinivas Moorthi, Xingren Wu, Jiande Wang, Sudhir Nadiga, Patrick Tripp, David Behringer, Yu-Tai Hou, Hui ya Chuang, Mark Iredell, Michael Ek, Jesse Meng, Rongqian Yang, Malaquías Peña Mendez, Huug van den Dool, Qin Zhang, Wanqiu Wang, Mingyue Chen, and Emily Becker. 15 Mar. 2014. The NCEP Climate Forecast System Version 2. *Journal of Climate* 27, 6 (15 Mar. 2014), 2185 – 2208. https://doi.org/10.1175/JCLI-D-12-00823.1

[32] Mahmoud Shepero, Dennis van der Meer, Joakim Munkhammar, and Joakim Widén. 2018. Residential probabilistic load forecasting: A method using Gaussian process designed for electric load data. *Applied Energy* 218 (2018), 159–172. https://doi.org/10.1016/j.apenergy.2018.02.165

[33] Souhaib Ben Taieb, Raphaël Huser, Rob J Hyndman, and Marc G Genton. 2016. Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression. *IEEE Transactions on Smart Grid* 7, 5 (2016), 2448–2455.

[34] Dennis W. Van der Meer, Joakim Munkhammar, and Joakim Widén. 2018b. Probabilistic forecasting of solar power, electricity consumption and net load: Investigating the effect of seasons, aggregation and penetration on prediction intervals. *Solar Energy* 171 (2018b), 397–413.

[35] Dennis W. Van der Meer, Mahmoud Shepero, Andreas Svensson, Joakim Widén, and Joakim Munkhammar. 2018a. Probabilistic forecasting of electricity consumption, photovoltaic power generation and net demand of an individual building using Gaussian Processes. *Applied energy* 213 (2018a), 195–207.

[36] Frederik vom Scheidt, Hana Medinová, Nicole Ludwig, Bent Richter, Philipp Staudt, and Christof Weinhardt. 2020. Data Analytics in the Electricity Sector–A Quantitative and Qualitative Literature Review. *Energy and AI* (2020), 100009.

[37] Frederik vom Scheidt, Philipp Staudt, and Christof Weinhardt. 2019. Assessing the Economics of Residential Electricity Tariff Selection. In *2019 International Conference on Smart Energy Systems and Technologies (SEST)*. 1–6. https://doi.org/10.1109/SEST.2019.8849143

[38] Julian Vossen, Baptiste Feron, and Antonello Monti. 2018. Probabilistic forecasting of household electrical load using artificial neural networks. In *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*. IEEE, 1–6.

[39] Y. Wang, Q. Chen, T. Hong, and C. Kang. 2019. Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges. *IEEE Transactions on Smart Grid* 10, 3 (2019), 3125–3148. https://doi.org/10.1109/TSG.2018.2818167

[40] Yi Wang, Dahua Gan, Mingyang Sun, Ning Zhang, Zongxiang Lu, and Chongqing Kang. 2019. Probabilistic individual load forecasting using pinball loss guided LSTM. *Applied Energy* 235 (2019), 10–20.

[41] Shu Zhang, Yi Wang, Yutian Zhang, Dan Wang, and Ning Zhang. 2020. Load probability density forecasting by transforming and combining quantile forecasts. *Applied Energy* 277 (2020), 115600.

# A APPENDIX

**Table 3: Overview of Probabilistic Individual Load Forecasting Literature**

| Study | Main method(s) | Input features | Data | Evaluation metric | Load scenarios |
|---|---|---|---|---|---|
| *This study* | *GRU, LSTM* | *net load, calendar, weather* | *1 year at 1 h resolution* | *Average pinball loss* | *Household (HH), HH with solar, HH with heating, HH with solar + heating* |
| Munkhammar et al. 2021 | Markov-chain mixture distribution model | load | 3 years at 30 min resolution | Reliability MAE, PINAW, normalized CRPS | HH |
| Afrasiabi et al. 2020 | Ensemble of CNNs, GRU, MDN | load, weather | 1 year at 30 min resolution | RMSE, MAPE, CRPS, CE | HH |
| Zhang et al. 2020 | Ensemble of GRU, GBRT, RF, LightGBM | load, calendar | 1.5 years at 30 min resolution | CRPS | HH |
| Elvers et al. 2019 | CNN | load, calendar, weather | 2 years at 15-60 min resolution | Pinball loss | HH |
| Wang et al. 2019 | LSTM | load, calendar | 1.5 years at 30 min resolution | Average pinball loss | HH |
| Shepero et al. 2018 | Gaussian process, log-normal process | load, calendar | 3 years at 30 min resolution | MAE, RMSE, PINAW, PICP | HH |
| Van der Meer et al. 2018a | Static + dynamic Gaussian Process | net load | 3 years at 30 min resolution | MAE, MAPE, NRMSE, PICP, PINAW, NCRPS | HH with solar |
| Van der Meer et al. 2018b | Dynamic Gaussian Process, Quantile regression | net load | 3 years at 30 min resolution | PICP, PINAW, NCRPS | HH with solar |
| Vossen et al. 2018 | MDN, Softmax Regression Networks | load, calendar | three different data sets | CRPS | HH |
| Gan et al. 2017 | LSTM | load | 500 d at 30 min resolution | Average quantile score | HH |
| Taieb et al. 2016 | Boosting additive quantile estimation | load | 1.5 years at 30 min resolution | CRPS | HH |
| Arora and Taylor 2016 | Conditional kernel density estimation | load | 8 mo at 30 min resolution | CRPS, unconditional coverage | HH |

CE: Cross-entropy, CNN: Convolutional neural network, GBRT: Gradient boosting regression tree, RF: random forest, LGBM: Light gradient boosting machine, CPRS: Continuous Ranked Probability Score, MAE: Mean average error, MAPE: Mean average percentage error, MDN: Mixed density networks, NRMSE: Normalized root mean square error, PICP: Prediction interval coverage probability, PINAW: Prediction interval normalized average width, NCRPS: Normalized continuous ranked probability score, RMSE: Root mean square error

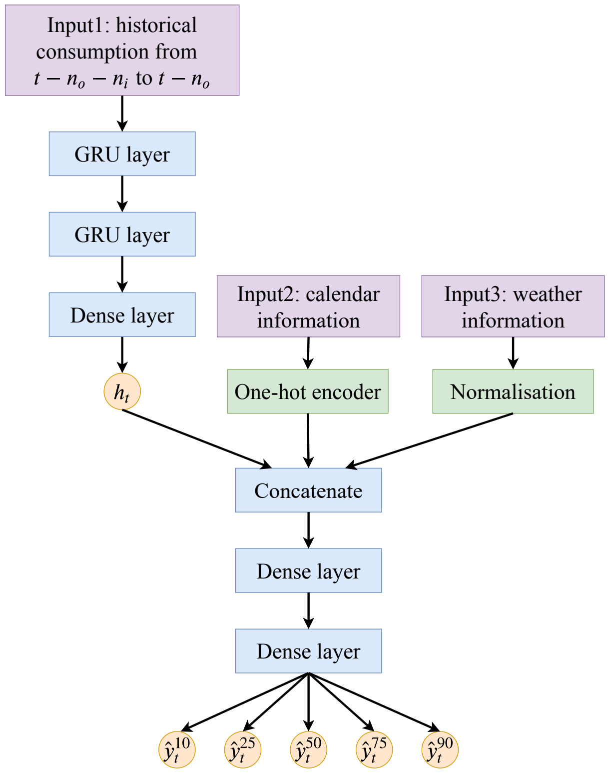



**Figure 2: Network structure of the QGRU forecasting model**

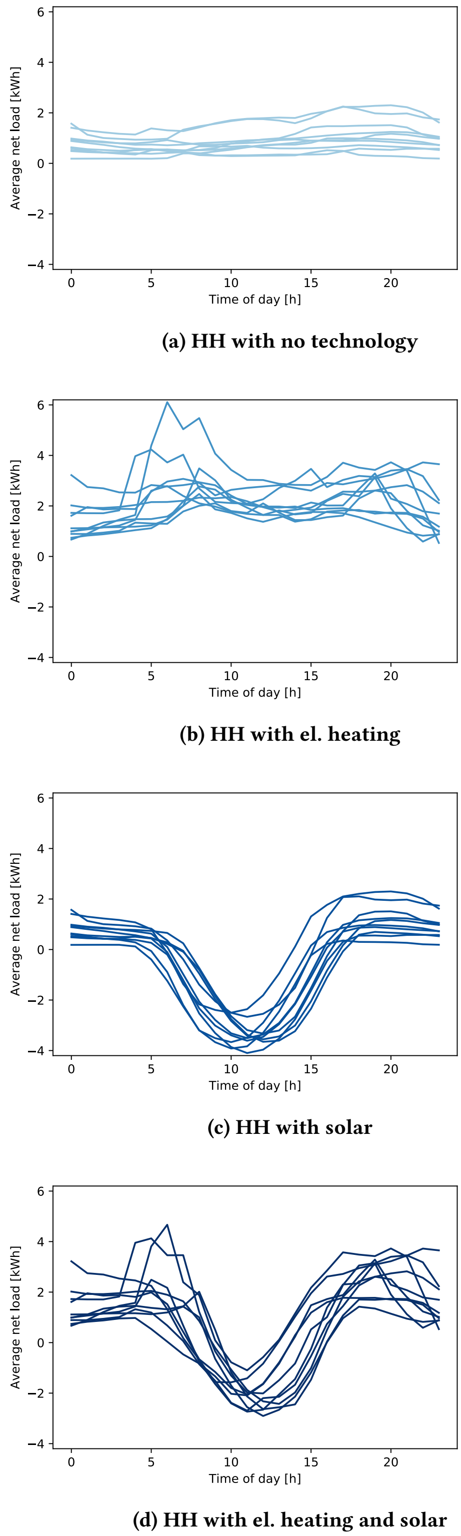
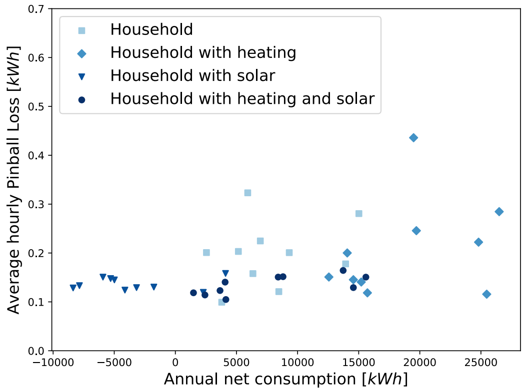
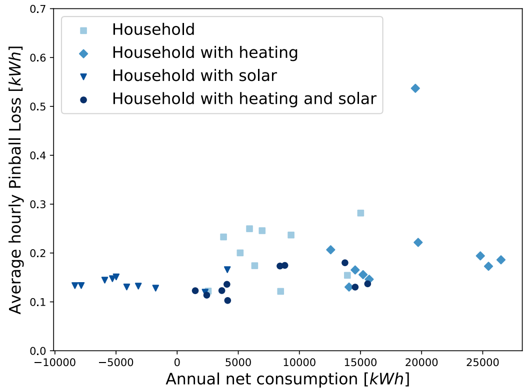


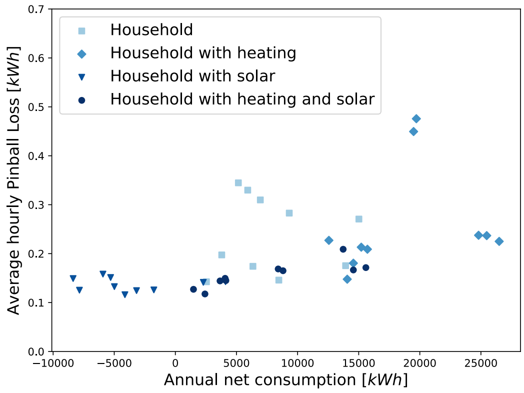

**Figure 3: Average daily net load curves of the ten households in each group**



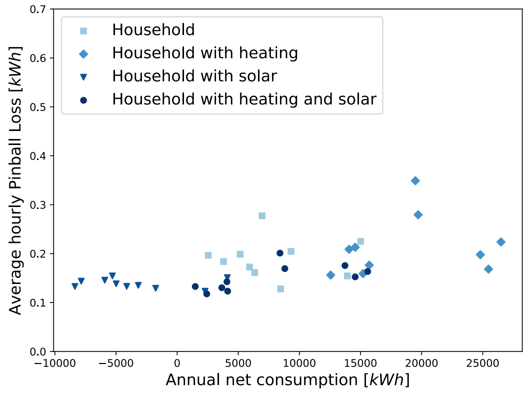

**(a) QGRU**




**(b) QLSTM**




**(c) QREGNN**




**(d) QLSTM_noWeather**

**Figure 4: Net load and pinball loss**