# Federated Learning for Distributed Load Forecasting: Addressing Data Imbalance in Smart Grids

Alexander Wallis*, Sascha Hauke**, Hannah Jörg**, Konstantin Ziegler*
*Department of Computer Science
**Department of Interdisciplinary Studies
University of Applied Sciences Landshut
Landshut, Germany
e-mail: {alexander.wallis | sascha.hauke | hannah.joerg | konstantin.ziegler}@haw-landshut.de

***Abstract*—The integration of renewable energy resources transforms traditional energy systems, introducing prosumers – entities that both produce and consume energy – as key participants in modern Smart Grids. Effective load forecasting is mandatory for optimizing energy resources and grid stability. Federated Learning (FL) has emerged as a promising approach for distributed training of Machine Learning (ML)-based forecasting models. This enables collaborative model optimization across multiple prosumers while preserving data privacy. However, the impact of unbalanced data sets across participants remains a critical challenge in terms of potentially effecting learning convergence and forecast accuracy. In this work, we define and implement a FL system based on real-world electricity consumption data from a variety of prosumers. Experimental results demonstrate the trade-off between centralized and federated learning approaches, providing insights into addressing data heterogeneity in FL systems. These insights highlight the potential of FL to support the evolution of distributed energy systems while ensuring data-privacy and scalability. Future research directions include other strategies to migrate the effect of data imbalances and further improve the efficiency of federated optimization for dynamic energy systems.**

***Keywords*-Short-Term Load Forecasting; Federated Learning; Smart Grid; Data Privacy; Distributed Data.**

## I. INTRODUCTION

Accurate load forecasting is mandatory for stable and reliable Smart Grid (SG) operation. But, the accuracy of load forecasting models, in particular ML based models, highly depends on the amount and quality of available training data [1]. Especially on smaller grid levels, e.g., low-voltage grids, or even residential household levels, the available electricity consumption data are very limited. But, with the rise of *prosumers* – consumers also able to produce electricity – prediction models on exactly this grid level is crucial for network management tasks [2].

Even if households are able to record and transmit electricity consumption data through smart meter utilization, the grid operator needs sufficient data storage and computational resources to process the data. Otherwise, the gathered data must be transferred for further processing. This transfer rises data privacy concerns and is even prohibited by law, e.g., General Data Protection Regulation [3]. The ability of information and behavior retrieval based on leakage of electricity consumption data has already been shown in the past [4]–[6].

Here, FL seems to be a promising approach to develop a single ML model for electricity consumption forecasting with distributed data sets – and at the same time satisfying data privacy regulation [7]. In contrast to the traditional approach, where the training of ML model is done centralized, this task is shifted to each user individually.

In [8], FL was first used by McMahan *et al.* to train prediction models on mobile devices through users' keyboard inputs. Afterwards, applications with FL were proposed in various fields, e.g., medical and health care, industrial engineering, finance, transportation [9]–[11].

For SG development, various FL approaches were proposed, too. In [12], FL is used for anomaly detection in terms of energy usage with a detection rate compared to centralized approaches. The authors in [13] present a conceptual framework for secure FL usage in SG environments with focus on vertical and horizontal data distribution over the clients. A detailed overview of further interesting FL researches in the field of SGs is given in [14].

Although, FL can be a promising approach for distributed load forecasting, the impact of unbalanced data sets among the clients is unclear. To evaluate FL in the context of prosumer-level load forecasting, we present the following contributions in this work:

- Definition and implementation of FL system composed of a variety of prosumer based on real-world electricity consumption data.
- Comparison of forecast accuracy between a centralized and a federated learning approach for model optimization.
- Investigation of the influence of unbalanced data sets within a federation on the learning convergence and the overall forecasting error.

This work is organized as following. First, the necessary background information as well as notation and terminologies are given in Section II. Second, the proposed FL approach is described in detail and the different experiments conducted are described in Section III. Third, the experiment results are presented, compared, and subsequently evaluated and discussed w.r.t. forecasting accuracy in Section IV. Fourth and last, the insights gained from the experiments' results are summarized and starting points for further research are given in Section V.

**ISBN: 978-1-68558-242-5**

## II. BACKGROUND

Before further detailing the conducted experiments in Section III, we give the respective problem formulation (Section II-A) and background information on FL (Section II-B) as well as an overview of related work (Section II-C).

### A. Problem Formulation

Basically, the load forecasting problem can be categorized into three groups based on the forecast horizon: (i) short-term, (ii) middle-term and (iii) long-term load forecasting. In this work, attention is paid on Short-Term Load Forecasting (STLF), since we are interested in a household's next day electricity consumption.

Let $\mathbf{x}_d = (x_d^{(0)}, ..., x_d^{(T)}) \in R^T$ be a household's consumption of day $d$ divided into $T$ time intervals. Further, let $\mathbf{y}_d = (y_{d+1}^{(0)}, ..., y_{d+1}^{(T)}) \in R^T$ be the next day's electricity consumption, then $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i) | i = 0, ..., D\}$ is the data set composed of input-output pairs for a total of $D$ days. Now, a supervised learning approach approximates a function $\mathbf{y}_d \approx \hat{f}(\mathbf{x}_d)$ for the following optimization problem:

$$\arg\min_{\hat{f}\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^{n} L(\hat{f}(\mathbf{x_i}), \mathbf{y}_i) \tag{1}$$

where $L(\cdot)$ is the desired cost function to be minimized.

Typically, in a centralized learning setting, this is done by collecting each household's data and subsequently by training a combined forecasting model, which is afterwards distributed to every household. Indeed, this rises all of the problems and concerns described earlier (see Section I) and FL is a promising approach to tackle all of them.

### B. Federated Learning

Contrary to the centralized learning, a FL approach guarantees data-privacy by preserving prosumers' consumption data locally. A collaboration of prosumer – a so-called *federation* – trains a STLF model by only exchanging respective model parameters. Typically, the participants within a federation are called clients but in this work the terms clients, prosumers and households are used interchangeably. Let $\mathcal{P} = \{p^{(i)} | i = 0, ..., N\}$ be the set of $N$ prosumers then the FL procedure involves the following steps:

1) **Distribution** of the initial global model to all prosumers which are part of the federation $p \in \mathcal{P}$.
2) **Training** of the global model by adjusting it's parameters based on the local data set of every prosumer.
3) **Returning** the adjusted model parameters to a central unit, e.g., trusted 3rd party, data center, one of the participants.
4) **Aggregation** of all received parameters by a predefined `aggregate`-function and integration into the global model.

This whole procedure, also depicted in Figure 1, is repeated over a defined number of *communication rounds* $r$. Interestingly, reducing the number $C$ of clients participating in every learning round increases the communication efficiency without loss of prediction accuracy [8]. So, in every round a prosumer subset $\mathcal{P}'_r \subseteq \mathcal{P}$ with $|\mathcal{P}'_r| = C$ is randomly chosen to take part in the training task in step 2.

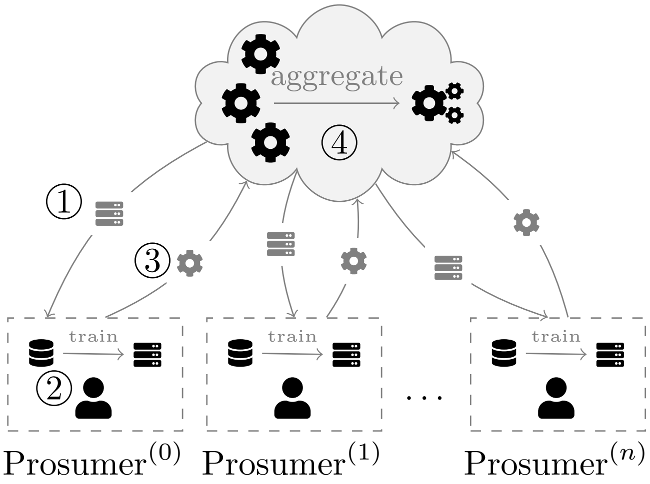

Figure 1. In a Federated Learning approach, all prosumers train their models locally on their own data.

Beside the number of prosumers involved in training, the used `aggregate`-function offers additional flexibility. In [8], the author introduces `FedSGD` and `FedAvg`, where the later is the common approach for solving the FL problem by calculating the (weighted) average (often mean) per parameter. Other aggregation approaches are, e.g., federated adaptive optimizers (`FedAdam`, `FedAdagrad`, `FedYogi`) [15], momentum-based variance-reduced technique (`FAFED`) [16], heterogeneity focused (`FedProx` [17], `SCAFFOLD` [18]). There are plenty more proposed `aggregate`-methods, and the related questions in terms of, e.g., applicability, optimality, generalization, are major research topics.

At this point, it is worth noting that additional security mechanism are needed to guarantee some desired security level. Although, FL offers a framework for data-privacy in distributed learning, data leakage or reconstruction attacks are still possible [19]. Privacy enhancing techniques applicable for FL settings are, e.g., differential privacy and homomorphic encryption [20].

In the next section, we give an overview of existing FL research with focus on STLF.

### C. Related Work

After describing the FL approach in general, we give an overview of existing FL research conducted in the field of residential STLF. Here, we limit the related work explicitly to (i) residential households and (ii) maximum 24-hour forecast horizon.

A comparison between `FedAvg` and `FedSGD` with different forecast horizons (1 h and 24 h) is given in [21]. They showed that their proposed FL model with `FedAvg` reaches higher accuracy than a centralized and a personalized model.

In [22], the authors compare the forecasting accuracy of a FL model on prosumers involved in training and on hold-out prosumers. They choose this approach to evaluate how well the global model fit for non-participating prosumers. Here, the non-participant prosumers fine tune the pre-trained model for 5 epochs locally. They conclude that this fine tuning step improves the forecast accuracy compared to the global model.

In terms of unbalanced client data distribution, Liu *et al.* proposed the closest approach [23]. Here, clients are divided into 5 groups based on the resolution of their available consumption data ranging from $300\,\mathrm{s}$ to $1.800\,\mathrm{s}$.

A hybrid CNN-LSTM model is used in a FL setting in [24]. To handle the consumption heterogeneity, the authors propose a model fine-tuning step after the weight aggregation based on multiple kernel variant of maximum mean discrepancies. Furthermore, all clients are involved in every training and the number of data samples are equal over all clients.

The authors in [25] compare the accuracy of a centralized model with a FL one, a FL plus clustering, and FL plus clustering and subsequently local fine tuning. Here, the last approach reaches the highest accuracy. But, to manage all experiment permutations the evaluations are done with fixed $C = 0.1$.

All of the mentioned related work are summarized with their respective training and model parameters in Table I. It can be seen that the related work in terms of unbalanced data sets is non existing – as far as we know – for the STLF problem on residential prosumer level.

## III. METHODOLOGY

To evaluate our proposed FL approach, different experiments are conducted in this work. Therefore, we build a federation composed of prosumers represented by household data taken from public available real-world electricity records (see Section III-A).

### A. Used Data Set

In this work, residential household data are taken from the `SmartMetersInLondon` [26] data set, which is a refactored version of the "Low Carbon London Project" data. This data set contains electricity consumption records for $5,567$ London households between November 2011 and February 2014. In the following, the conducted data preprocessing and preparation steps as well as the selection of suitable households is described.

*a) Household Selection:* Since the date range differs between prosumers in the data set, only houses with the most overlap are selected. Furthermore, households with more than three consecutive hours of missing values are removed otherwise linear interpolated. In total, 20 households are selected suitable for further usage. The hourly mean electricity consumption is depicted for every day in the training set in Figure 2. Subsequently, the respective consumption data is preprocessed for every selected household in the following.

*b) Data Preprocessing:* Since the date ranges of available data varies tremendously across all prosumers, we select the time between 1st January 2013 and 28th February 2014 with the most overlapping data. This interval is further divided into train and test data ($\mathcal{D}_{\text{train}}$ and $\mathcal{D}_{\text{test}}$), whereas the whole year 2013 is used for training and the remaining data for testing. This leads to $|\mathcal{D}^{(p)}_{\text{train}}| = 8,760$ and $|\mathcal{D}^{(p)}_{\text{test}}| = 1,416$ samples for every prosumer. For every prosumer, both data sets are normalized individually by $x' = \frac{x-\sigma}{\mu}$, where $x'$ is the transformed consumption time series.

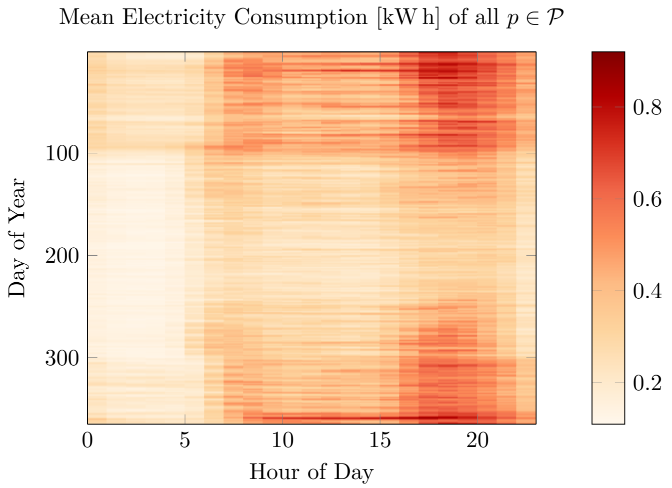

Figure 2. Mean electricity consumption of all selected households from the `SmartMeterInLondon` data set.

*c) Look-back and Forecast Horizon:* The accuracy of time series forecasting depends on both, the chosen look-back window as well as the forecast horizon. In the related work (Section II-C), those parameter are also chosen variously. Here, our proposed forecasting model uses the last $24\,\mathrm{h}$ as input to predict the next $24\,\mathrm{h}$. Although, additional features, e.g., weather, holiday, weekday/weekend, can reduce the forecast error, we restrict our model to the raw consumption values. In [27], we evaluate the FL model with further feature engineering.

After the household selection and necessary preprocessing steps, the used ML model architecture, as well as further detail on the overall development process is given in the next part.

### B. System Setting

In this section, we give all the relevant information about the model architecture and used hyperparameters. Afterwards, a definition for different kinds of learning prosumers within the federation based on the ability to store training data is presented. A description of the used federation, as well as the training procedure is given in the third part.

*a) Model and Hyperparameters:* In this work, we choose a vanilla Multi-Layer Perceptron (MLP) as model architecture, similar to the proposed model in [8]. This architecture allows an easy implementation and training on lightweight devices with limited computational resources. This fully connected MLP has two hidden layers with 200 neurons each and uses a Rectified Linear Unit as activation function.

*b) Weak and Strong Prosumer:* We introduce the terms *strong* and *weak* prosumer, to describe two different types of

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

ISBN: 978-1-68558-242-5

TABLE I. OVERVIEW AND SUMMARY OF RELATED WORK FOR FEDERATED LEARNING (FL) APPROACHES FOR RESIDENTIAL SHORT-TERM LOAD FORECASTING (STLF).

| Related Work | #Clients | $C$ | ML-Model | Data Set | Balanced Data | Aggregation |
|---|---|---|---|---|---|---|
| Taïk and Cherkaoui [22] | 200 | $5, 10$ | LSTM | AUSTIN | yes | `FedAVG` |
| Fekri *et al.* [21] | 19 | 6 | LSTM | non-public | yes | `FedSDG, FedAVG` |
| Liu *et al.* [23] | 50 | 10 | iQGRU | AUSTIN | semi | `FedAVG` |
| Shi and Xu [24] | 10 | 10 | CNN-LSTM | LONDON | yes | `FedAVG` |
| Briggs *et al.* [25] | 100 | 0.1 | LSTM | LONDON | yes | `FedAVG` |

prosumers based on the amount of available training data. The two types are defined the following way:

**Definition 1.** Let $p \in \mathcal{P}$ be a prosumer only able to store training data between two consecutive communication rounds, then it is called a *weak* prosumer $p_{\text{weak}}$.

**Definition 2.** Let $p \in \mathcal{P}$ be a prosumer with no storage limitations, then it is called a *strong* prosumer $p_{\text{strong}}$.

Based on the Definitions 1 and 2, we define the fraction of strong prosumers within a federation as the so-called *strong-prosumer-fraction*:

**Definition 3.** Let $|p_{\text{weak}}|, |p_{\text{strong}}|$ be the number of weak respective strong prosumers in $\mathcal{P}$, then the strong-prosumer-fraction is defined as $\phi = \frac{|p_{\text{strong}}|}{|p_{\text{weak}}|+|p_{\text{strong}}|}$.

This allows a straightforward distinction between prosumers within a federation and introduces another parameter for the overall training procedure.

*c) Training Procedure:* For all conducted experiments, with or without strong and weak prosumers, the respective training procedure takes $r = 100$ communication rounds in total. At $r = 0$ the global model's weights $w$ are randomly initialized. After every round, the global model's weights are updated by a weighted `FedAvg` aggregation function, s.t. $w_{r+1} \leftarrow \sum_{p \in \mathcal{P}'_r} \frac{n_p}{n} w_r^{(p)}$, where $n_p, n$ is the number of sample per prosumer respective the number of all samples. The local weights $w_r^{(p)}$ are calculated locally for every $p \in \mathcal{P}'_r$ in parallel by $w_r^{(p)} \leftarrow w_r - \eta \nabla_w \mathcal{L}(w_r; \mathbf{x}_i, \mathbf{y}_i)$ for a single epoch with a learning rate of $\eta = 0.001$ and the Mean Squared Error (MSE) as loss function $\mathcal{L}(\cdot)$.

To evaluate the proposed FL approach and also to analyze the impact of unbalanced data sets, various experiments are conducted, which are further detailed in the following section.

### C. Experiment Settings

The proposed FL approach for residential STLF is evaluated in different experiments. The evaluation is based on the MSE error metric given as $\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$, where $n$ is the number of test set samples and $\hat{y}_i, y_i$ is the predicted respective actual consumption value. We run the following three experiments:

I **Benchmark** A centralized model – as well as one local model for every prosumer – is trained over $r$ epochs.

II **Number of Learners** Since a new subset of learning prosumers is selected in every round (see Section II-B), we evaluate the model's forecast accuracy for different number of learners $C = \{1, 3, 5, 7, 10, 20\}$.

TABLE II. TEST SET ERROR FOR EXPERIMENT I. MSE IS CALCULATED OVER ALL 20 PROSUMERS

| Model | ↓ MSE ($\mu \pm \sigma$) | min | max | won |
|---|---|---|---|---|
| centralized | $0.181 \pm 0.13$ | 0.030 | 0.545 | 3 out of 20 |
| personalized | $0.166 \pm 0.13$ | 0.021 | 0.514 | 17 out of 20 |

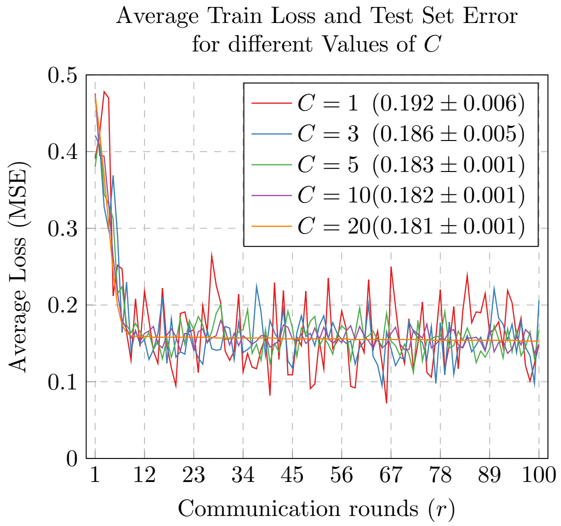

Figure 3. Experiment II: Train loss and test set error with mean and standard deviation over 10 repetitions for different values of $C$.

III **Strong Prosumer Fraction** With the introduction of weak and strong prosumers, we evaluate our FL approach based on unbalanced data sets. For $C = \{1, 10, 20\}$ the strong-prosumer-fraction $\phi = \{0.05, 0.25, 0.5, 0.75, 1\}$ is considered. Here, the unbalanced data set evolves over the communication rounds $r = \{1, 2, \ldots, 100\}$ by:

$$\text{weak:} \quad \mathcal{D}_r^{(p)} = \mathcal{D}_{r-1:r}^{(p)} \tag{2}$$

$$\text{strong:} \quad \mathcal{D}_r^{(p)} = \mathcal{D}_{0:r}^{(p)}. \tag{3}$$

So, for strong prosumer the training samples increase by $n = \lfloor \frac{|\mathcal{D}|}{r} \rfloor$ in every round, whereas for weak prosumer the samples have a fixed size of $n$.

The experiments I-III are repeated for $N = 10$ times to handle the randomness via model initialization and prosumer sampling with $C, \phi$. Our proposed FL approach is implemented in `Python=3.9` with `PyTorch` and model training was

ISBN: 978-1-68558-242-5

executed on a local machine with a Nvidia Geforce RTX 2080 graphic card. The experiments' results are listed in the next section.

## IV. EXPERIMENT RESULTS & DISCUSSION

The results of the various experiments are presented in the same order as defined in Section III-C. The respective results are provided below, followed by a detailed analysis and discussion.

Figure 3 illustrates the training loss across all communication rounds $r$ as well as the test set error in the legend. For the different values of $C = \{1, 3, 5, 10, 20\}$, the test set error is given as mean with standard deviation over all 10 repetitions. Similar to experiment I, the MSE is calculated over all prosumers $p \in \mathcal{P}$ without individual examination.

For experiment III, results are given in two ways. First, the average training loss over all runs is depicted in Figure 4. Second, Table III list the test set errors. In addition to numerical values over all prosumers, the MSE is also calculated separately for the sets of $p_{\text{weak}}$ and $p_{\text{strong}}$. The minimum and maximum MSE values are determined over all 10 runs combined for each combination of $C$- and $\phi$-values.

In this work, a FL approach was proposed for the STLF problem at residential prosumer level. Three experiments were conducted to analyze the impact of unbalanced data distribution among prosumers within the federation.

The first experiment compared a centralized MLP trained on all prosumers' data with a personalized MLP trained individually for each prosumer. Of 20 households in total, 17 times the personalized model reaches a higher accuracy (see Table II). This indicates a strong distribution of consumption behaviour across the prosumers since more data does not guarantee better results.

The second experiment examined the effect of different numbers of learners. As shown in Figure 3, test set errors show minimal variation for $C > 1$, with nearly identical training loss reduction. However, lower $C$-values introduce more variance, emphasizing trade-off between distribution computational resources and learning efficiency.

In real-world scenarios, training data availability varies among prosumers due to recording and storage capabilities as well as temporal offsets in joining the federation. To address this, the third experiment introduced the distinction between weak and strong prosumers, defined by storage capability. The strong prosumer fraction $\phi$ represents the proportion of strong prosumers within a federation. Figure 4 indicates slower training convergence with a decreasing number of strong prosumers, irrespective of $C$-values. However, reducing $\phi$ to 0.75 or 0.5 did not significantly impact training speed or test set error. This finding is relevant for practical applications, suggesting that not all prosumers need to contribute learning resources to maintain overall performance.

## V. CONCLUSION & FUTURE WORK

This work developed a ML-based model for the STLF problem at residential prosumer level. Given that high-resolution electricity consumption data contain behavioral information, data privacy concerns arise when transferring and processing such data. To address this, FL was incorporated as a viable approach to train ML models on distributed data without requiring direct data exchange. Three experiments were designed and conducted to evaluate the proposed FL approach. The results demonstrated that FL can achieve competitive forecasting accuracy while preserving data privacy. The trade-off between the number of learners and computational efficiency was also analyzed, along with the effects of strong and weak prosumers on training convergence and performance.

TABLE III. TEST SET ERROR FOR EXPERIMENT III. ERROR IS GIVEN AS MSE WITH MEAN AND STANDARD DEVIATION OVER ALL 10 REPETITIONS.

| $C$ | $\phi$ | ↓ MSE ($\mu \pm \sigma$) | | | | |
|---|---|---|---|---|---|---|
| | | all | strong | weak | min | max |
| 1 | 0.05 | $\mathit{0.215 \pm 0.14}$ | $0.192 \pm 0.12$ | $0.216 \pm 0.14$ | 0.026 | 0.674 |
| | 0.25 | $\mathbf{0.193 \pm 0.12}$ | $0.209 \pm 0.15$ | $0.188 \pm 0.11$ | 0.039 | 0.597 |
| | 0.5 | $0.194 \pm 0.12$ | $0.202 \pm 0.13$ | $0.186 \pm 0.12$ | 0.037 | 0.565 |
| | 0.75 | $0.196 \pm 0.12$ | $0.194 \pm 0.12$ | $0.199 \pm 0.13$ | 0.038 | 0.587 |
| | 1 | $0.201 \pm 0.13$ | $0.201 \pm 0.13$ | – | 0.036 | 0.626 |
| 10 | 1 | $\mathit{0.223 \pm 0.15}$ | $0.142 \pm 0.07$ | $0.227 \pm 0.15$ | 0.029 | 0.750 |
| | 0.25 | $0.186 \pm 0.12$ | $0.187 \pm 0.13$ | $0.186 \pm 0.11$ | 0.033 | 0.540 |
| | 0.5 | $0.184 \pm 0.12$ | $0.182 \pm 0.12$ | $0.187 \pm 0.12$ | 0.038 | 0.550 |
| | 0.75 | $0.181 \pm 0.12$ | $0.185 \pm 0.12$ | $0.170 \pm 0.10$ | 0.038 | 0.525 |
| | 1 | $\mathbf{0.180 \pm 0.12}$ | $0.180 \pm 0.12$ | – | 0.041 | 0.527 |
| 20 | 1 | $\mathit{0.198 \pm 0.13}$ | $0.205 \pm 0.13$ | $0.198 \pm 0.13$ | 0.034 | 0.711 |
| | 0.25 | $0.190 \pm 0.12$ | $0.193 \pm 0.13$ | $0.189 \pm 0.12$ | 0.035 | 0.591 |
| | 0.5 | $0.183 \pm 0.12$ | $0.173 \pm 0.11$ | $0.192 \pm 0.13$ | 0.040 | 0.546 |
| | 0.75 | $0.181 \pm 0.12$ | $0.172 \pm 0.11$ | $0.208 \pm 0.12$ | 0.038 | 0.523 |
| | 1 | $\mathbf{0.179 \pm 0.11}$ | $0.179 \pm 0.11$ | – | 0.042 | 0.516 |

Note: lowest error is in **bold**, highest in *italic*.

In future work, we will focus on extending and improving the proposed FL approach. This study primarily addressed unbalanced data sets within a federation, adopting constraints such as a lightweight MLP architecture, state-of-the-art `FedAvg` weight aggregation, and the exclusion of external features. To enhance overall forecasting accuracy, these constraints should be revisited. Preliminary results indicate the utilizing more complex Long-Short Term Memory Neural Network (LSTM) models and incorporating weather information can reduce forecasting errors. Additionally, this study did not explicitly implement a security layer. Future research will explore methods to ensure data privacy and prevent information leakage while integrating insights from this study. Furthermore, the potential of Transformer-based models for STLF remains an unexplored area, warranting future investigation.

## ACKNOWLEDGMENT

The research in this work is supported for Alexander Wallis by the Federal Ministry of Education and Research BMBF (FKZ 03FHP212) and for Sascha Hauke by Bayern Innovativ.

## REFERENCES

[1] A. Jain, H. Patel, L. Nagalapatti, N. Gupta, S. Mehta, S. Guttula, *et al.*, "Overview and importance of data quality for machine learning tasks," in *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 2020, pp. 3561–3562.

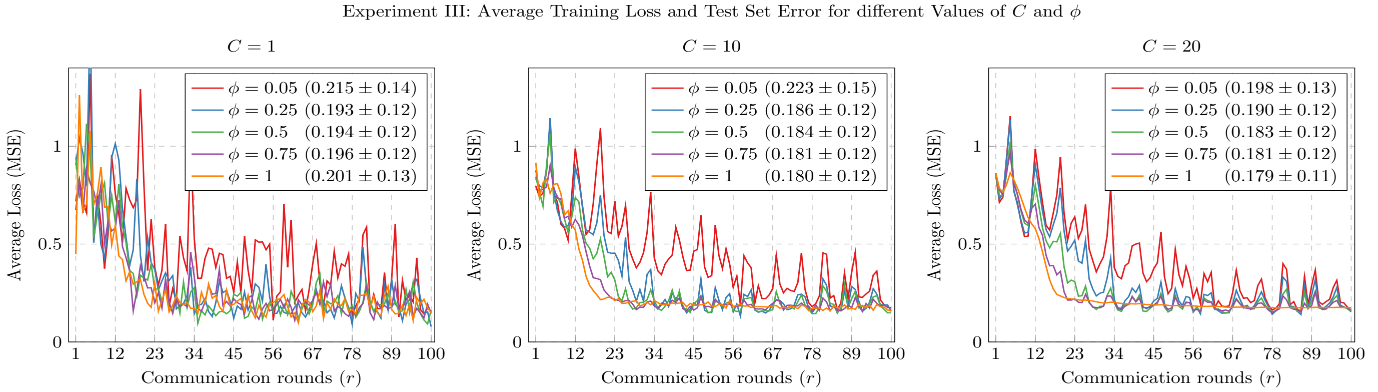

Figure 4. The training loss and test set error for different fractions of strong prosumer $\phi$ evaluated for $C = 1$ (left), $C = 10$ (middle), and $C = 20$ (right).

[2] H. Habbak, M. Mahmoud, K. Metwally, M. M. Fouda, and M. I. Ibrahem, “Load forecasting techniques and their applications in smart grids,” *Energies*, vol. 16, no. 3, p. 1480, 2023.

[3] GDPR, “General data protection regulation,” *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC*, 2016.

[4] P. L. Ambassa, A. V. Kayem, S. D. Wolthusen, and C. Meinel, “Inferring private user behaviour based on information leakage,” *Smart Micro-Grid Systems Security and Privacy*, pp. 145–159, 2018.

[5] G. Wood and M. Newborough, “Dynamic energy-consumption indicators for domestic appliances: Environment, behaviour and design,” *Energy & buildings*, vol. 35, pp. 821–841, 2003.

[6] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin, “Private memoirs of a smart meter,” in *Proceedings of the 2nd ACM workshop on embedded sensing systems for energy-efficiency in building*, 2010, pp. 61–66.

[7] J. Chen, H. Yan, Z. Liu, M. Zhang, H. Xiong, and S. Yu, “When federated learning meets privacy-preserving computation,” *ACM Comput. Surv.*, vol. 56, no. 12, Oct. 2024, ISSN: 0360-0300. DOI: 10.1145/3679013.

[8] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, “Communication-Efficient Learning of Deep Networks from Decentralized Data,” in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, vol. 54, PMLR, Apr. 2017, pp. 1273–1282.

[9] L. Li, Y. Fan, M. Tse, and K.-Y. Lin, “A review of applications in federated learning,” *Computers & Industrial Engineering*, vol. 149, p. 106 854, 2020.

[10] P. M. Mammen, “Federated learning: Opportunities and challenges,” *arXiv preprint arXiv:2101.05428*, 2021.

[11] J. Wen, Z. Zhang, Y. Lan, Z. Cui, J. Cai, and W. Zhang, “A survey on federated learning: Challenges and applications,” *International Journal of Machine Learning and Cybernetics*, vol. 14, no. 2, pp. 513–535, 2023.

[12] J. Jithish, B. Alangot, N. Mahalingam, and K. S. Yeo, “Distributed anomaly detection in smart grids: A federated learning-based approach,” *IEEE Access*, vol. 11, pp. 7157–7179, 2023.

[13] H. Liu, X. Zhang, X. Shen, and H. Sun, “A federated learning framework for smart grids: Securing power traces in collaborative learning,” *arXiv preprint arXiv:2103.11870*, 2021.

[14] X. Cheng, C. Li, and X. Liu, “A review of federated learning in energy systems,” *2022 IEEE/IAS industrial and commercial power system Asia (I&CPS Asia)*, pp. 2089–2095, 2022.

[15] S. Reddi, Z. Charles, M. Zaheer, Z. Garrett, K. Rush, J. Konečný, *et al.*, *Adaptive federated optimization*, 2021. arXiv: 2003.00295 `[cs.LG]`.

[16] X. Wu, F. Huang, Z. Hu, and H. Huang, *Faster adaptive federated learning*, 2023. arXiv: 2212.00974 `[cs.LG]`.

[17] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, “Federated optimization in heterogeneous networks,” *Proceedings of Machine learning and systems*, vol. 2, pp. 429–450, 2020.

[18] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh, “Scaffold: Stochastic controlled averaging for federated learning,” in *International conference on machine learning*, PMLR, 2020, pp. 5132–5143.

[19] C. Ma, J. Li, M. Ding, H. H. Yang, F. Shu, T. Q. Quek, *et al.*, “On safeguarding privacy and security in the framework of federated learning,” *IEEE network*, vol. 34, no. 4, pp. 242–248, 2020.

[20] V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantanha, and G. Srivastava, “A survey on security and privacy of federated learning,” *Future Generation Computer Systems*, vol. 115, pp. 619–640, 2021.

[21] M. N. Fekri, K. Grolinger, and S. Mir, “Distributed load forecasting using smart meter data: Federated learning with recurrent neural networks,” *International Journal of Electrical Power & Energy Systems*, vol. 137, p. 107 669, 2022. DOI: https://doi.org/10.1016/j.ijepes.2021.107669.

[22] A. Taïk and S. Cherkaoui, “Electrical load forecasting using edge computing and federated learning,” in *ICC 2020-2020 IEEE international conference on communications (ICC)*, IEEE, 2020, pp. 1–6.

[23] Y. Liu, Z. Dong, B. Liu, Y. Xu, and Z. Ding, “Fedforecast: A federated learning framework for short-term probabilistic individual load forecasting in smart grid,” *International Journal of Electrical Power & Energy Systems*, vol. 152, p. 109 172, 2023.

[24] Y. Shi and X. Xu, “Deep federated adaptation: An adaptative residential load forecasting approach with federated learning,” *Sensors*, vol. 22, no. 9, p. 3264, 2022.

[25] C. Briggs, Z. Fan, and P. Andras, “Federated learning for short-term residential load forecasting,” *IEEE Open Access Journal of Power and Energy*, vol. 9, pp. 573–583, 2022.

[26] *Smart meters in london*, https://www.kaggle.com/datasets/jeanmidev/smart-meters-in-london, Accessed: 2010-09-30.

[27] A. Wallis, U. Ludolfinger, S. Hauke, and M. Martens, “Towards federated short-term load forecasting,” *Internationale Energiewirtschaftstagung (IEWT 2021)*, pp. 1–9, 2021.