energies

Article

Global vs. Local Models for Short-Term Electricity Demand Prediction in a Residential/Lodging Scenario

Amedeo Buonanno [1,*], Martina Caliano [1], Antonino Pontecorvo [1], Gianluca Sforza [2], Maria Valenti [1] and Giorgio Graditi [2]

1 Department of Energy Technologies and Renewable Energy Sources, ENEA, 80055 Portici, Italy; martina.caliano@enea.it (M.C.); antonino.pontecorvo@enea.it (A.P.); maria.valenti@enea.it (M.V.)
2 Department of Energy Technologies and Renewable Energy Sources, ENEA, 00123 Rome, Italy; gianluca.sforza@enea.it (G.S.); giorgio.graditi@enea.it (G.G.)
* Correspondence: amedeo.buonanno@enea.it

Abstract: Electrical load forecasting has a fundamental role in the decision-making process of energy system operators. When many users are connected to the grid, high-performance forecasting models are required, posing several problems associated with the availability of historical energy consumption data for each end-user and training, deploying and maintaining a model for each user. Moreover, introducing new end-users to an existing network poses problems relating to their forecasting model. Global models, trained on all available data, are emerging as the best solution in several contexts, because they show higher generalization performance, being able to leverage the patterns that are similar across different time series. In this work, the lodging/residential electricity 1-h-ahead load forecasting of multiple time series for smart grid applications is addressed using global models, suggesting the effectiveness of such an approach also in the energy context. Results obtained on a subset of the Great Energy Predictor III dataset with several global models are compared to results obtained with local models based on the same methods, showing that global models can perform similarly to the local ones, while presenting simpler deployment and maintainability. In this work, the forecasting of a new time series, representing a new end-user introduced in the pre-existing network, is also approached under specific assumptions, by using a global model trained using data related to the existing end-users. Results reveal that the forecasting model pre-trained on data related to other end-users allows the attainment of good forecasting performance also for new end-users.

Keywords: residential load forecasting; machine learning; nanogrid; time series analysis

check for updates

Citation: Buonanno, A.; Caliano, M.; Pontecorvo, A.; Sforza, G.; Valenti, M.; Graditi, G. Global vs. Local Models for Short-Term Electricity Demand Prediction in a Residential/Lodging Scenario. *Energies* **2022**, *15*, 2037. https://doi.org/10.3390/en15062037

Academic Editors: Alfredo Vaccaro, Andrea Mariscotti, Fabrizio de Caro and Tek Tjing Lie

Received: 28 January 2022
Accepted: 9 March 2022
Published: 10 March 2022

1. Introduction

The electrification of energy consumption associated with a constant increase in the use of Renewable Energy Sources (RES) plays a central role in the European energy transition for the achievement of the decarbonization objectives of the overall energy system, thanks to the intrinsic efficiency of the electricity sector and the technological maturity of RES. The trends in electrification and the increase in RES have already been underway for several years in many OECD countries. According to IEA's semi-annual Electricity Market Report [1], global electricity demand is constantly growing: after reducing by around 1% in 2020 due to the COVID-19 pandemic, it was set to grow by around 5% in 2021 and will rise by another 4% in 2022. At the same time, electricity generation from RES is considered to grow worldwide by more than 6% in 2022.

The transformation required is not zero-impact for the electricity system and implies a series of challenges to be faced so that the energy transition process can be carried out in a decisive and effective manner, maintaining the current high levels of service quality and, at the same time, avoiding excessive costs for citizens.

The growing integration of non-programmable renewable generation contributes to increasing the variability associated with electrical loads, significantly affecting the network management activities of Transmission System Operators (TSOs), based on always balancing the generation and demand of electricity for guaranteeing to citizens a safe, constant and reliable supply of energy.

In this context, electrical load forecasting can play a key role in effectively optimizing the use of energy resources for the purposes of energy system operation, also being able to contribute to energy management and to improve the decision-making process related to the generation and import of electricity and to the planning of the construction of energy infrastructure [2].

Electric load forecasting can also play an important role in smart grid environments, where demand-side management strategies represent essential aspects [3] for the proper design and operation. Due to the nonlinear nature of electrical loads, accurate forecasting is often challenging and can require much effort to be properly addressed [4].

In the last few years, many authors have dealt with the electrical load forecasting in the smart grid context, using conventional methods [5,6] or AI-based methods [7–9], such as Recurrent Neural Network (RNN) [10], Support Vector Regression (SVR) [11,12], Long Short-Term Memory (LSTM) [13], hybrid methods [14] and eXtreme Gradient Bosting (XGBoost) [15]. In a previous work [16], we have implemented different data-driven approaches, such as Persistence (PER), several Linear Regression (LR) methods, Feed Forward Neural Networks (FFNN), Convolutional Neural Networks (CNN), LSTM, XGBoost (XGB) and SVR, to forecast the electrical loads of individual households in a nanogrid environment. The main results of the analysis show that, for specific use cases, all methods tested have similar performance, and the methods that work best are Multivariate Linear Regression (MLR), FFNN and XGB.

In the smart grid context, the forecasting of multiple time series can be complex due to the potential large number of users involved. In this case, two approaches can be used: train one single model for each time series with the parameters that are learned separately (local method) or train a single model where its parameters are learned using all the available time series (global method) [17].

In the last few years, the local methods have been the most used approach, but, recently, the great availability of data and the new empirical and theoretical results have shown the high potential of global models. In fact, when the number of users is high, creating a predictive model for each user could be prohibitive in terms of training time, deployment and maintenance of the solution. Moreover, since the global models can leverage the patterns that are similar across different time series, they are less prone to overfitting than local models, resulting in improved generalization.

The global approaches have been often applied to the demand forecasting of products [18], including thousands of products over different sites, and have emerged as the winning solutions in different forecasting competitions, such as M4 [19] and M5 [20].

The main assumption related to the use of global methods is that the time series come from data generating processes that are similar or related; however, recent results show that the forecasting performance is good also when the considered time series are not [21]. This makes the global approaches more interesting, as it is always possible to obtain heterogeneous time series to improve the performance [22,23].

Recent works [22,24] have investigated these aspects and some interesting insights emerged. In particular, [22] evidenced that, independently of the heterogeneity of the time series, a global model that can perform as well as local models (or outperform them) always exists. This result is very relevant because it refutes the first impression that a global model is more limited and the idea that the relation among time series is fundamental for the effectiveness of the global approach.

However, such a global model is not simple to construct and, hence, it is interesting to understand how these insights could be useful in a smart grid context.

Managing more data, the global models can be more complex than local ones, continuing to obtain better generalization performance. The complexity of global models can be increased with more lags as inputs or non-linear, non-parametric models, or using data partitioning [22,24].

A data-driven approach to regression problems that is particularly useful when data are not sufficient to train new prediction models is transfer learning [25]. With this approach, models pre-trained on a large dataset can be customized and reused without having to train another model for the new dataset from scratch.

In the context of building energy demand prediction, transfer learning comes in handy when extensive historical data of power consumption are not available, as is the case of new buildings.

Several studies in the literature have recently assessed the value of transfer learning in predicting building energy demand for different building types (e.g., commercial, residential) over different time horizons [26–30]. These case studies often revealed an increase in prediction accuracy using data from additional buildings, compared with a model that used only a small target dataset. This happens especially when the source and target data share some characteristics, such as belonging to similar building types (but different distributions) or to the same climate zone (but different locations). In this study, the pre-trained models investigated are directly reused for a new case, without adaptations, as we assume for the new case that no data are available.

Focusing on lodging/residential energy demands, this work aims to show how, in the smart grid context, the forecasting of multiple time series could be tackled using global models with good results.

In particular, the work addresses the comparison of local and global approaches with the minimum manual intervention on the training of the models. To reach this goal, we trained the models using the same hyperparameters. In fact, in a real scenario, searching for the optimal set of hyperparameters is computationally expensive, as well as possibly hampering the prompt deployment of the models in the production system.

The forecasting performance obtained with the global models is similar to the performance obtained by using the local ones, while presenting simpler deployment and maintainability. A recent work has investigated similar aspects, evaluating the benefits of the cross-learning approach [31]. Differently from the mentioned study, this work does not use external features, being based entirely on historical energy demand, and considers some state-of-the-art methods frequently used for forecasting problems.

Furthermore, the work aims to show that, under specific assumptions, the forecast of a new user's energy demand can be approached using a global model trained using data from the existing users with an acceptable loss in performance.

The rest of the paper is structured as follows: in Section 2, the considered approach, the residential user dataset, the evaluation method and the performance metrics are described. The results are shown in Section 3 and discussed in Section 4. Finally, Section 5 reports the main conclusions of the work.

2. Materials and Methods

2.1. Models

Among several forecasting models, for the objectives of this study, we have selected the most used and promising forecasting approaches, namely the Linear Regression Model (Linear), LSTM [13], Temporal Convolutional Network (TCN) [32], Neural Basis Expansion Analysis Time Series Forecasting (NBEATS) [33], Light Gradient Boosted Model (LGBM) [34] and Transformer [35]. As a baseline, the Persistence model (Persistence) has also been implemented.

2.2. Dataset Description

In 2019, the Great Energy Predictor III (GEPIII) challenge [36] was organized by ASHRAE through the Kaggle platform. The hourly energy consumption is gathered from the energy meters (electricity, chilled water, steam and hot water) of 1448 buildings distributed on 16 unknown sites worldwide. The complete dataset covers three years from

2016 to 2018, but only the energy measurements related to year 2016 were provided to the competitors. For this reason, in this work, only 2016 has been considered. This situation is frequently encountered in a real situation where the measurement campaign has been implemented for less than a year.

The buildings are grouped based on their primary use (e.g., education, office, public services, lodging/residential, etc.).

In our work, we have focused on the electricity measurements of the buildings belonging to site 15 since it contains the highest number of buildings with lodging/residential primary use (28). Most of the buildings in site 15 have 15% of data missing (especially from the middle of February to the end of March).

Figure 1 shows the electricity consumption of a generic week of all 28 buildings of site 15, whereas Figure 2 shows the average and standard deviation for the whole period of observation. From the figures, it can be noted that the scales and patterns among buildings are very different even if all users belong to the same group.

Figure 1. Hourly electricity consumption for the 28 considered buildings in a generic week of the measuring period.

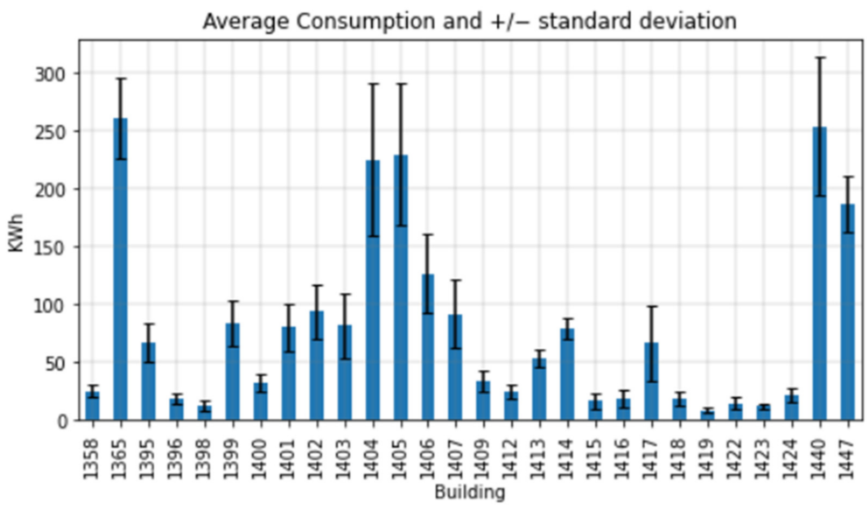

Figure 2. Hourly electricity consumption for the 28 considered buildings.

2.3. Experimental Setting

This work presents two types of experiments related to the 1-h-ahead forecasting, consisting of comparing local and global models and reusing pre-trained forecasting models.

The first experiment aims to compare the local and global models' performance (trained with same hyperparameters) in forecasting the energy demand of a single building. The local model for building *i* is trained using only the energy consumption of building *i*. Therefore, one model for each considered building is produced. The global model is trained using the energy consumption of all 28 buildings, resulting in only one model for all considered buildings.

The second experiment aims to understand if a global model, trained using the energy consumption data related to some buildings, can be used to forecast the energy demand of an additional building not seen before. For this purpose, four buildings that present different scales and/or patterns have been selected (Figure 3) as additional buildings and, for each one, a global model is trained using the energy consumption data of all other 27 buildings. The only information assumed known about the additional building is its average energy consumption, which is estimated from the training set in this work. However, if historical electricity data are not available, it could be estimated making some assumptions on the nominal power, the average consumption in similar cases, etc.

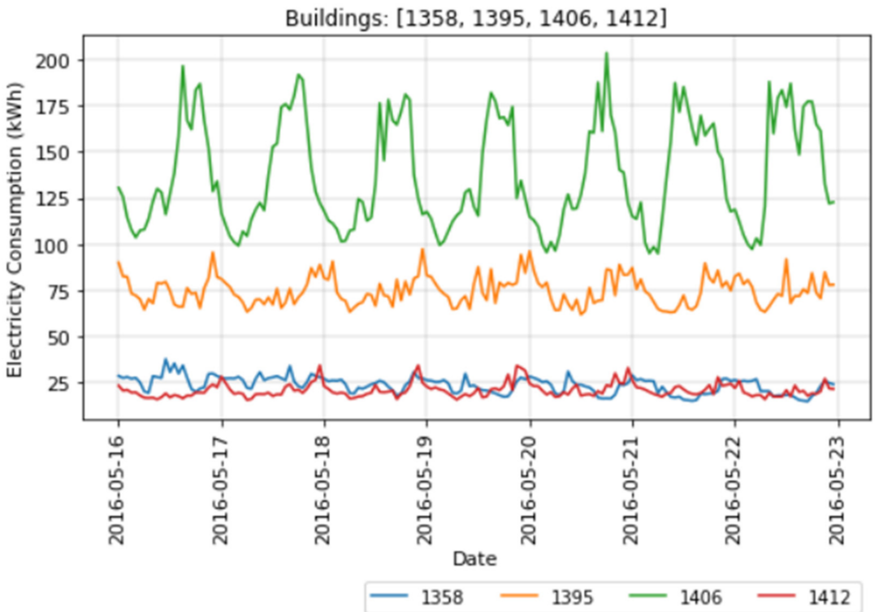

Figure 3. Hourly electricity consumption for buildings 1358, 1395, 1406, 1412 in a generic week of the measuring period, showing variations in scale and profile.

2.4. Preprocessing

From Figures 1 and 2, it is evident that the time series related to different buildings present different scales (due to the different nominal power, user behavior, etc.). For the local models, this aspect can be negligible, but for a global model, it could be problematic [37]. For this reason, each time series is properly normalized, dividing by its average consumption (mean-scale normalization) [38].

The missing data are resolved with a simple average imputation.

2.5. Model Training

In order to forecast 1-h-ahead electricity consumption for the 28 considered buildings, the available data have been divided into training and test sets. Eight months are used for the training set (from 1 January to 31 August) and 4 months for the test set (from 1 September to 31 December). A validation set has been extracted from the training set

(1 July to 31 August) for tuning the number of training epochs and to avoid the overfitting phenomenon (early stopping technique). Once the number of training epochs has been selected, the model is refitted using the complete training set. Each model uses the last 24 h of measurement to forecast the next hour. The mean value of each time series, used by mean-scale normalization, is estimated from the training set.

In Table 1, the main hyperparameters chosen for the considered models are listed: most hyperparameters are set to typical values for simulating the usage of the off-the-shelf models without high hyperparameterization. Other hyperparameters are set to the same value across all the models, in order to define the same conditions for all algorithms (e.g., optimizer, learning rate, batch size, number of lags, maximum number of epochs, etc.).

Table 1. Main parameters chosen for the considered models.

Model Name	Main Chosen Hyperparameters
Linear	Fit Intercept: True
LSTM	Batch Size: 1024; Hidden Size: 25 Optimizer: Adam with Learning Rate 1×10^{-3}; Maximum Number of Epochs: 200.
TCN	Batch Size: 1024; Dilation: 1; Kernel Size: 3; Number of Filters: 25; Dropout: 0.2 Optimizer: Adam with Learning Rate 1×10^{-3}; Maximum Number of Epochs: 200.
NBEATS	Batch Size: 1024; Number of stacks: 30, Number of blocks: 1, Number of fully connected layers: 4, Number of neurons for each fully connected layer: 256, Expansion Coefficient: 5 Optimizer: Adam with Learning Rate 1×10^{-3}; Maximum Number of Epochs: 200.
LGBM	Number of estimators: 100; Learning Rate:0.1
Transformer	Batch Size: 1024; Dropout: 0.1; Number of multi head attention: 4; Number of encoding layers: 3; Number of decoding layers: 3; Dimension of the feed-forward network model: 512 Optimizer: Adam with Learning Rate 1×10^{-3}; Maximum Number of Epochs: 200.

The models have been implemented using Python with the following libraries: Darts [39], NumPy and pandas. The experiments have been performed using a PC with CPU Intel Core i7-9700 @ 3.00GHz–8 cores (Santa Clara, CA, USA), 16GB of RAM, GPU NVIDIA GeForce GTX 1050Ti (Santa Clara, CA, USA), O.S. Microsoft Windows 10 Pro (Redmond, WA, USA).

2.6. Model Performance Evaluation

For model performance evaluation, the Coefficient of Variation (CV) and the Root Mean Squared Error ($RMSE$) are computed, as follows:

$$CV = \frac{\sqrt{\frac{1}{N}\sum_{t=1}^{N}(y(t) - \hat{y}(t))^2}}{\overline{y}} * 100 \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{N}(y(t) - \hat{y}(t))^2}{N}} \quad (2)$$

where $y(t)$ indicates the target value, $\hat{y}(t)$ the predicted value, $\overline{y}$ the average value of the target and N the number of values considered. The CV is a dispersion index that allows the comparison of different methods and/or different datasets, and it is expressed in percentage. The $RMSE$ is expressed here in kWh, referring to the energy load.

The performance has been evaluated discarding imputed values from the test set.

3. Results

3.1. Comparison between Local and Global Models

In the following, the results obtained from comparing the performance of the local and global models on the test set are presented for each building.

In detail, Figures 4–15 show the comparison and the difference between the *RMSE* (even numbered figures) and *CV* (odd numbered figures) obtained by the global and local models for the Linear, LSTM, TCN, LGBM, NBEATS and Transformer methods, respectively. In the figures, the difference is indicated with green (or red) bars, meaning that the *RMSE* or *CV* value obtained using the local model is higher (or lower) than the *RMSE* or *CV* obtained using the global model for the specific building.

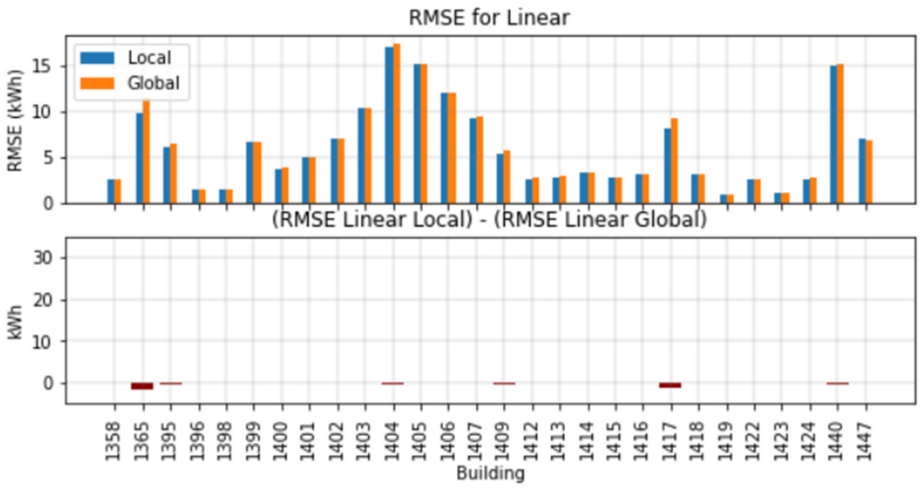

Figure 4. The upper figure shows the comparison between the *RMSE* obtained from the local and global Linear models on test set. The lower figure contains the difference between the *RMSE* obtained for the local and global Linear models on test set.

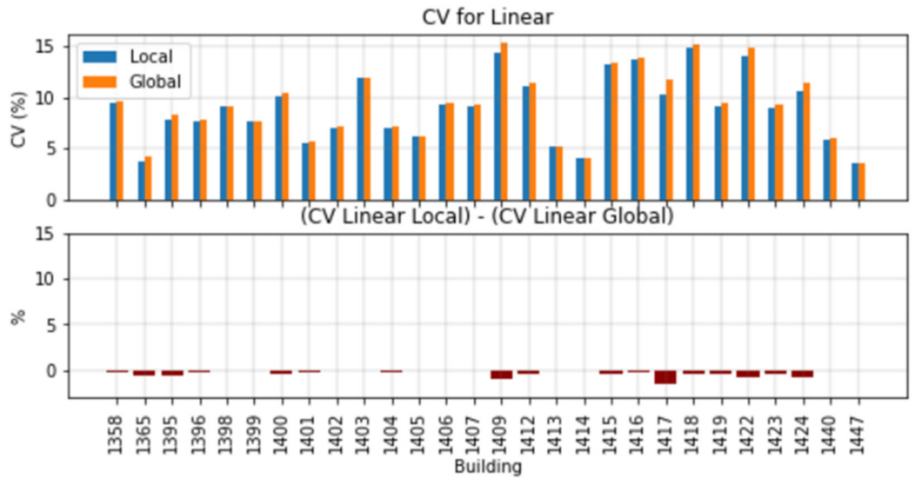

Figure 5. Comparison (in the upper figure) and difference (in the lower figure) between the *CV* obtained from local and global Linear models on test set.

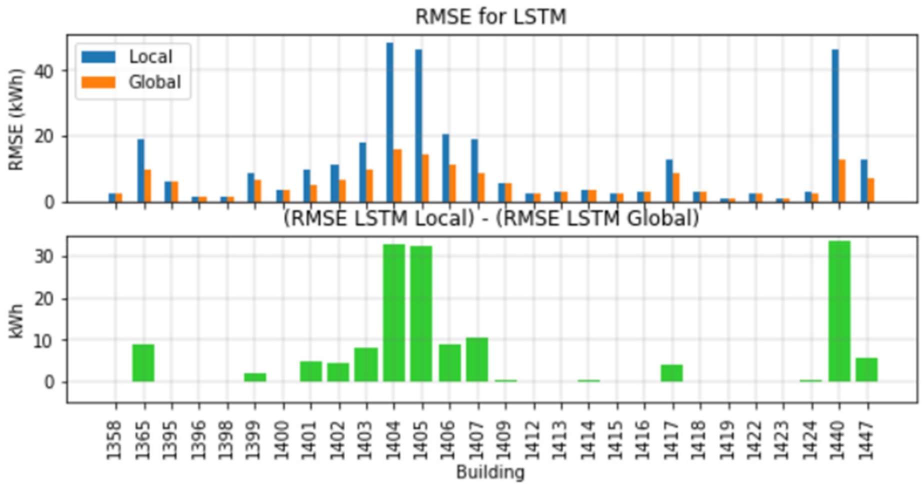

Figure 6. Comparison (in the upper figure) and difference (in the lower figure) between the *RMSE* obtained from local and global LSTM models on test set.

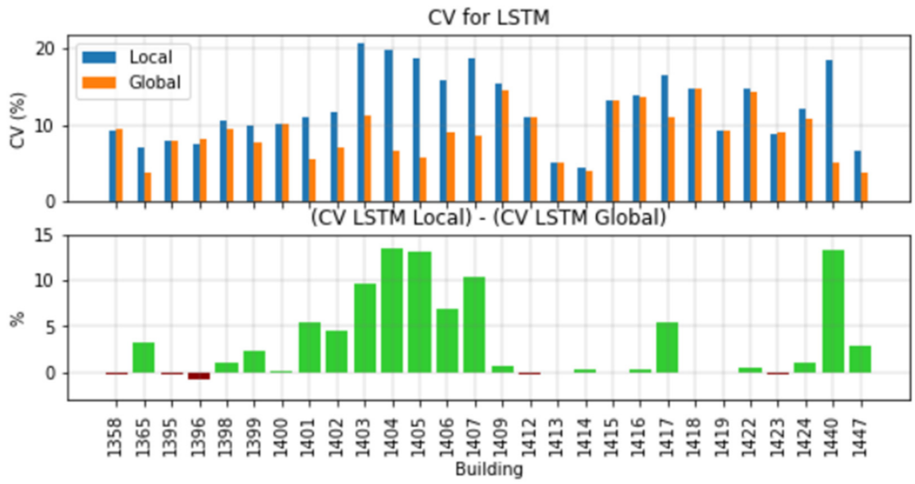

Figure 7. Comparison (in the upper figure) and difference (in the lower figure) between the *CV* obtained from local and global LSTM models on test set.

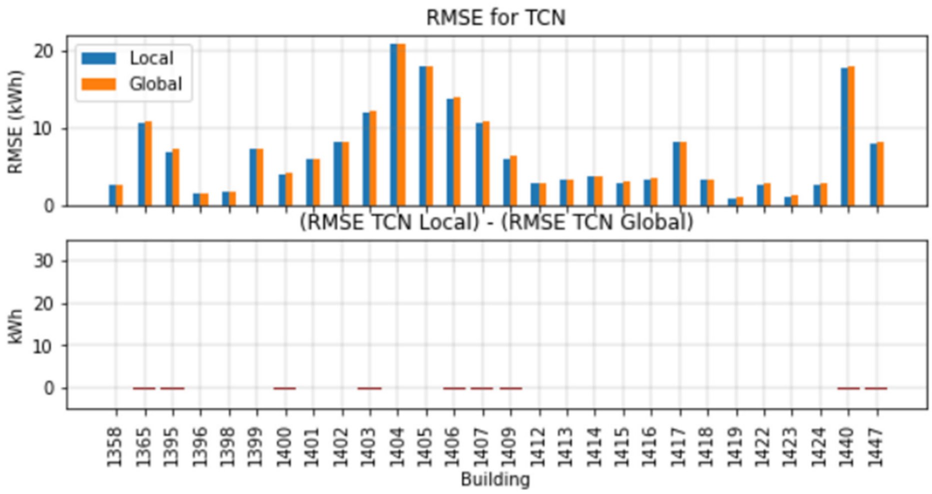

Figure 8. Comparison (in the upper figure) and difference (in the lower figure) between the *RMSE* obtained from local and global TCN models on test set.

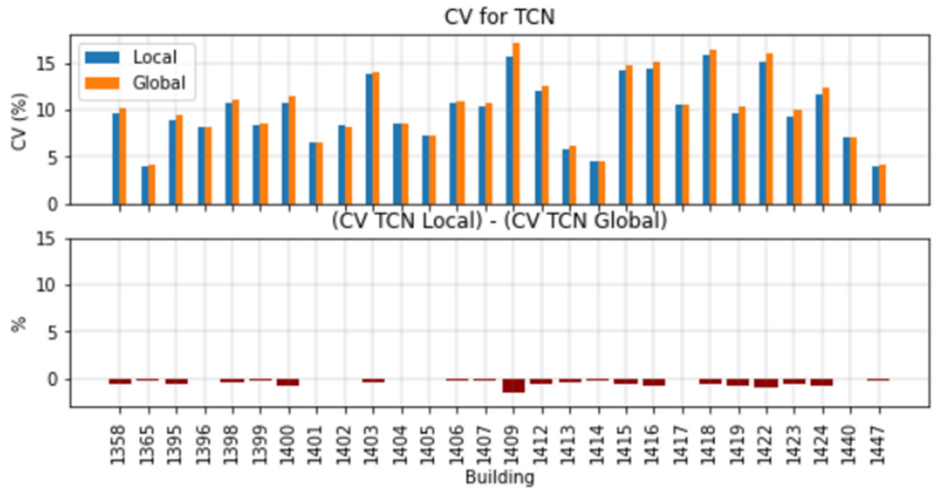

Figure 9. Comparison (in the upper figure) and difference (in the lower figure) between the *CV* obtained from local and global TCN models on test set.

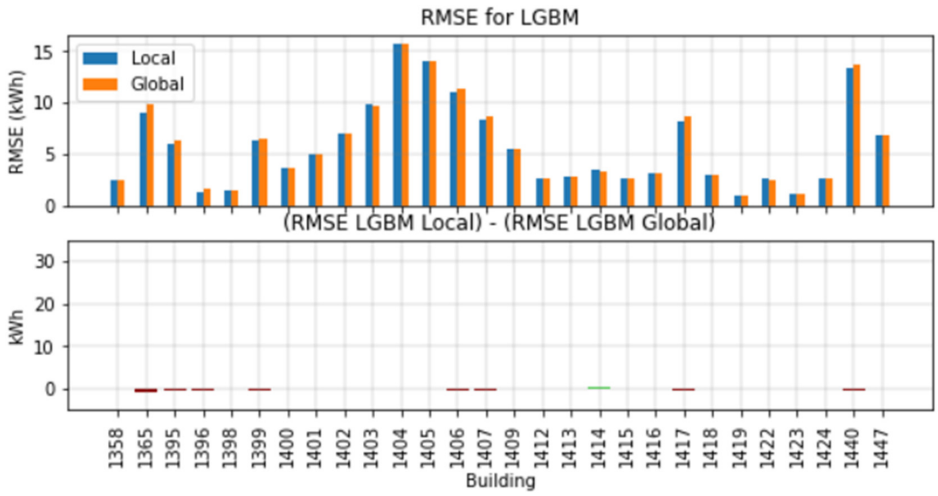

Figure 10. Comparison (in the upper figure) and difference (in the lower figure) between the *RMSE* obtained from local and global LGBM models on test set.

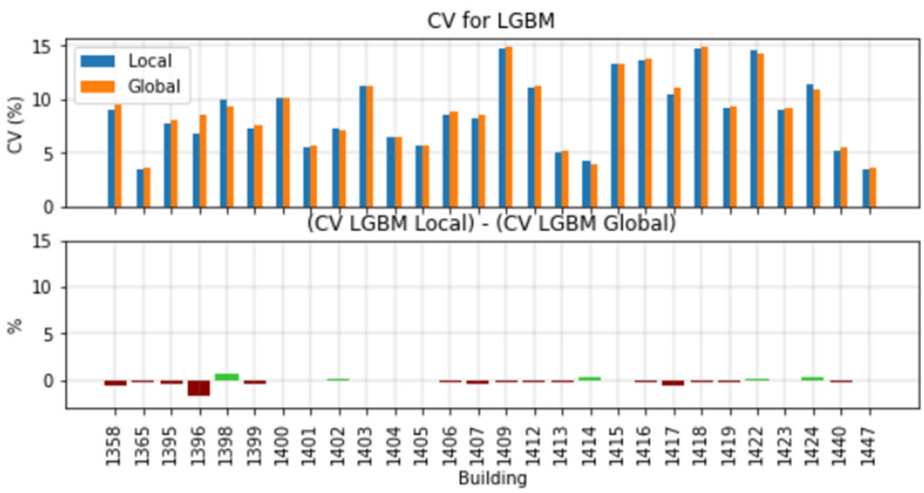

Figure 11. Comparison (in the upper figure) and difference (in the lower figure) between the *CV* obtained from local and global LGBM models on test set.

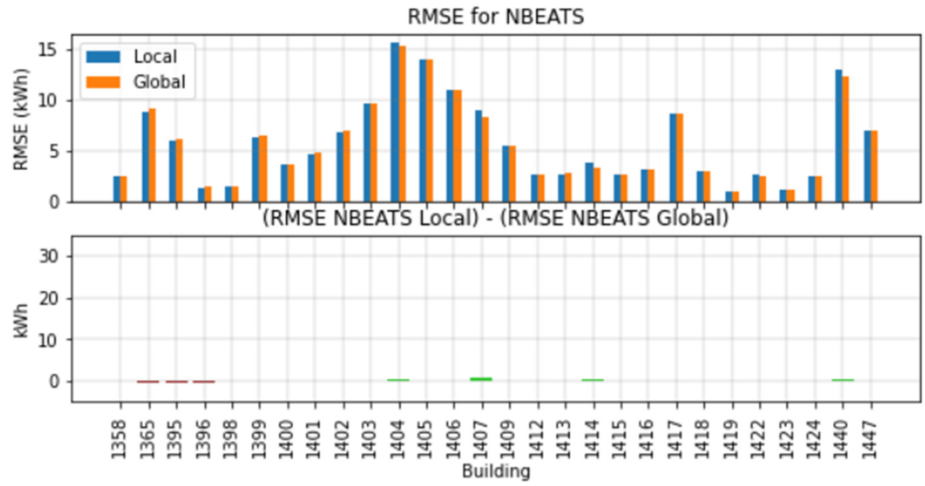

Figure 12. Comparison (in the upper figure) and difference (in the lower figure) between the *RMSE* obtained from local and global NBEATS models on test set.

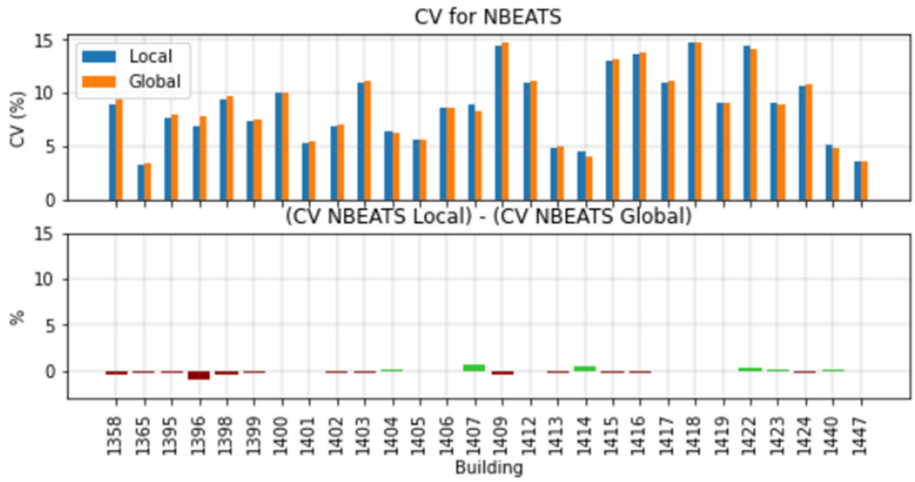

Figure 13. Comparison (in the upper figure) and difference (in the lower figure) between the *CV* obtained from local and global NBEATS models on test set.

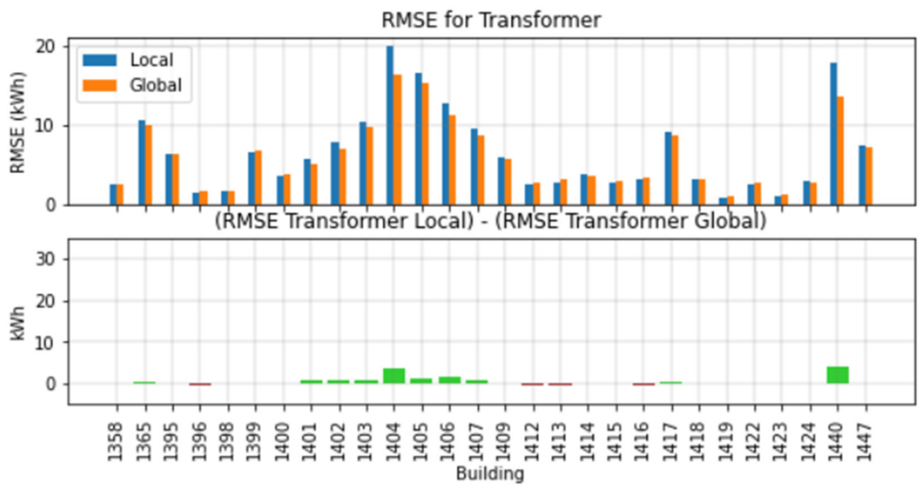

Figure 14. Comparison (in the upper figure) and difference (in the lower figure) between the *RMSE* obtained from local and global Transformer models on test set.

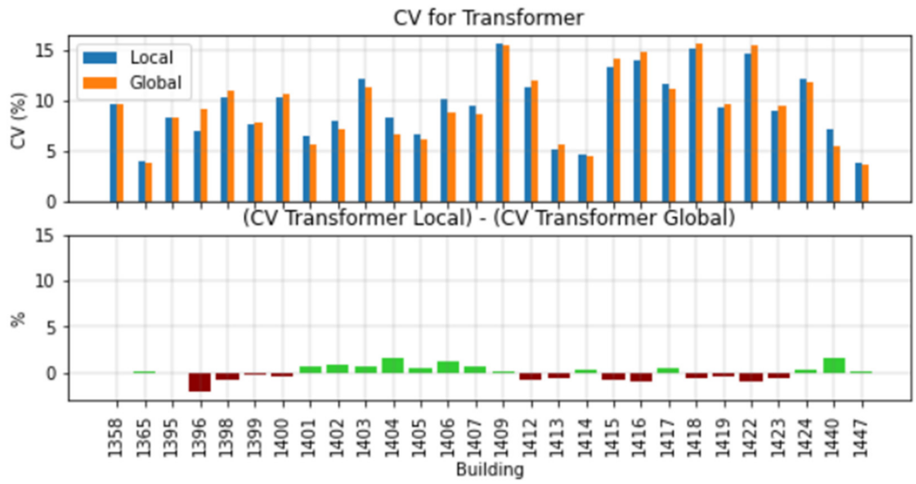

Figure 15. Comparison (in the upper figure) and difference (in the lower figure) between the *CV* obtained from local and global Transformer models on test set.

To clarify the comparison among local and global models for the considered methods, Figure 16 shows the forecasting results, in terms of *CV*, of the tested models across the considered buildings. The details are given in Table 2: it shows that the local and global approaches work similarly (LSTM excluded), with a small performance decrease for the global models, which is probably acceptable in a real scenario.

Figure 16. Comparison between the *CV* obtained from the considered local and global models on test set. The Persistence model has been added as a baseline.

Table 2. In the table are reported the forecasting results of tested models in terms of median *CV* and the 25th and 75th percentiles (in parentheses). The Wilcoxon signed rank is used to test the null hypothesis that the paired *CV* samples for local and global models come from the same distribution. The (*) indicates that the null hypothesis was rejected for p-value < 0.05. The last column represents the variation in performance (*CV*) of a global model with respect to its local counterpart: if negative, the performance decreased.

Model Type	Local (%)	Global (%)	Variation(Local-Global)/Local * 100
LSTM (*)	11.29 (9.06, 15.55)	9.00 (6.31, 11.08)	8.85 (−0.12, 45.13)
TCN (*)	9.58 (7.96, 11.72)	10.23 (8.01, 12.39)	−3.49 (−5.79, −1.84)
LGBM (*)	8.78 (6.25, 11.14)	9.01 (6.23, 11.10)	−1.09 (−3.67, 0.43)
NBEATS	8.93 (6.22, 10.98)	8.73 (6.13, 11.05)	−0.89 (−2.20, 0.59)
Transformer	9.32 (7.07, 11.76)	9.26 (6.55, 11.45)	0.16 (−5.44, 6.56)
Linear (*)	9.11 (6.76, 10.76)	9.23 (6.87, 11.54)	−2.01 (−4.14, −0.69)

3.2. Reuse of Pre-Trained Forecasting Models

Figure 17 shows the comparison between the forecasting results obtained for building 1358 by using LGBM local and global models as described in Section 3.1, and an LGBM global model trained using the training set of time series of all buildings excluding building 1358 (Global-except).

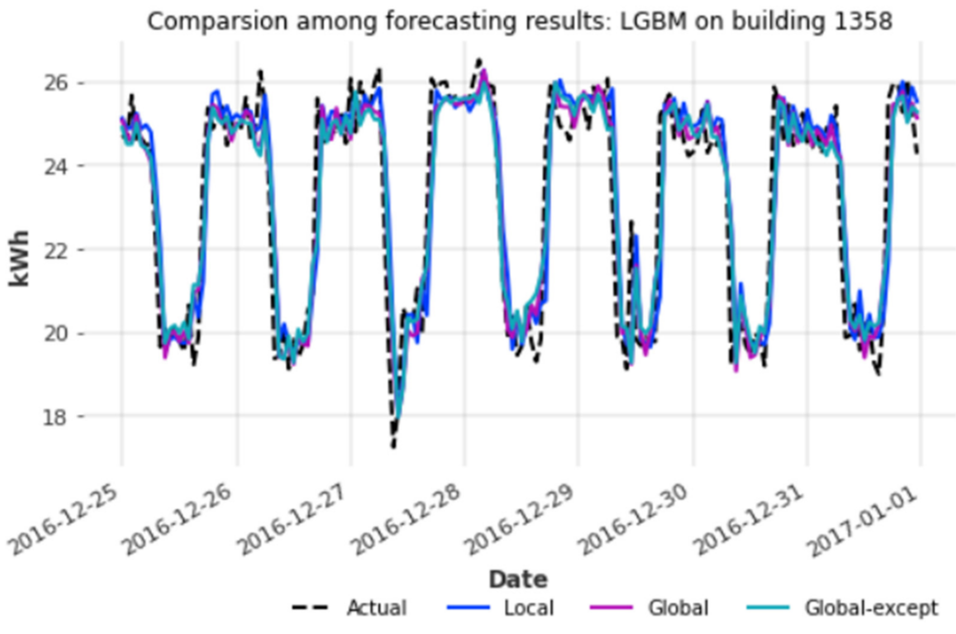

Figure 17. Comparison between the forecasting results using LGBM model for building 1358. Actual: ground-truth; Local: local model trained using training set of building 1358; Global: global model trained using training set of all buildings; Global-except: global model trained using training set of all the buildings excluding building 1358.

The Figures 18–23 show, for each selected building (1358, 1395, 1406, 1412), the prediction performance of the local model, of the global model trained using all the available data and of the global model trained using all the available data except the energy consumption related to the selected building.

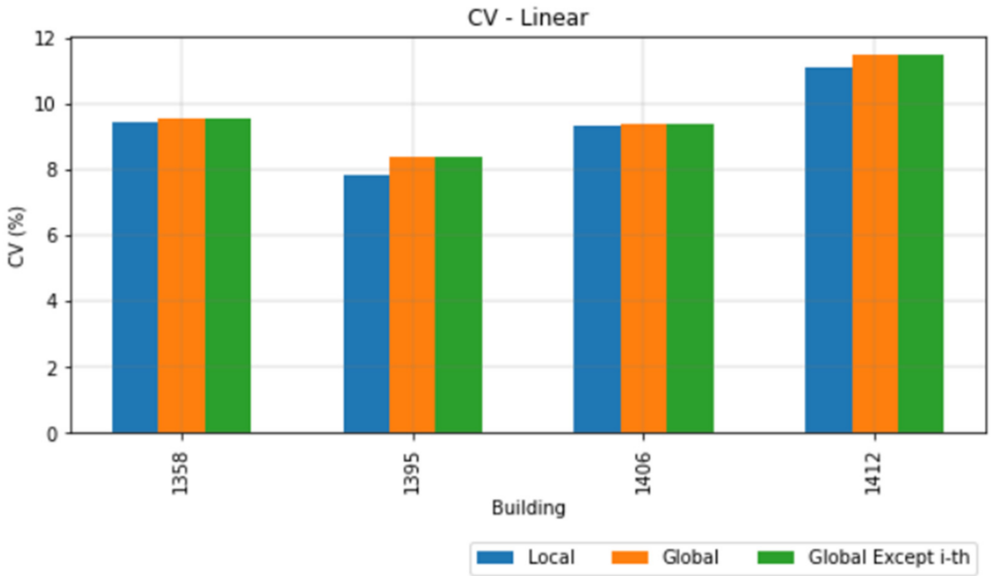

Figure 18. Comparison between the *CV* on test set obtained using Linear method with different modalities: local, global and global trained using all the available data except the energy consumption related to the selected building.

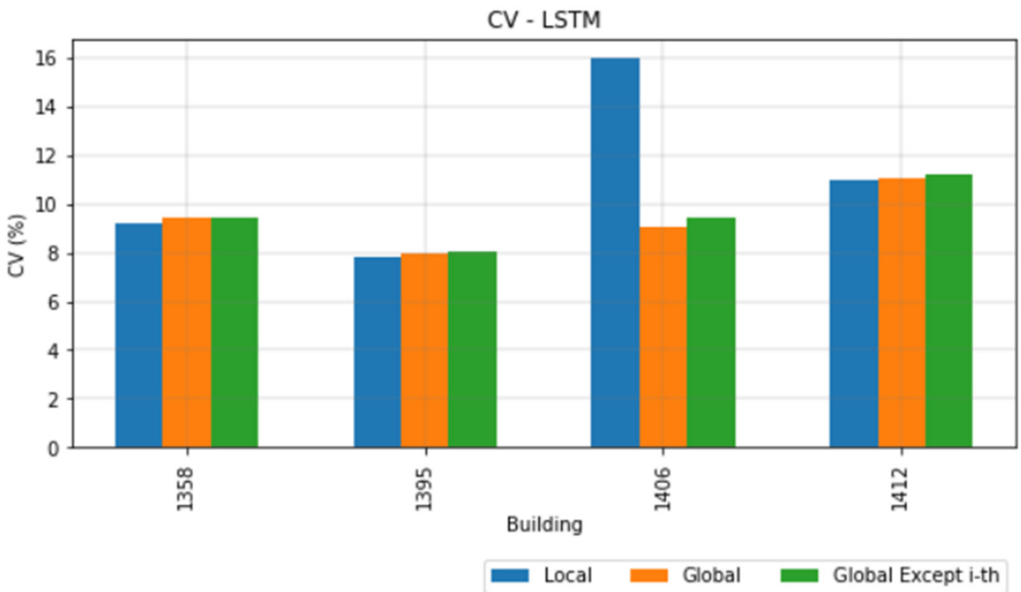

Figure 19. Comparison between the *CV* on test set obtained using LSTM method with different modalities: local, global and global trained using all the available data except the energy consumption related to the selected building.

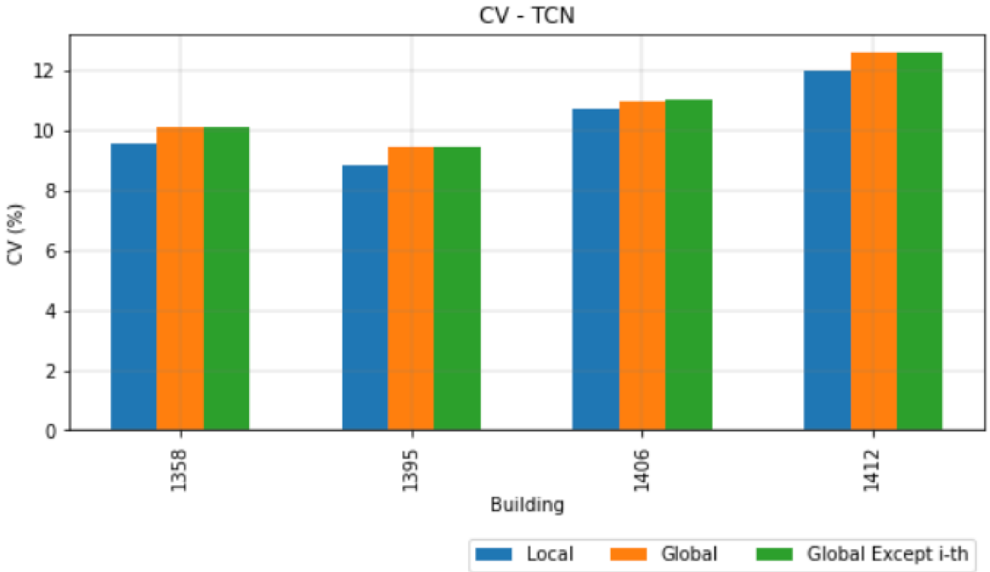

Figure 20. Comparison between the *CV* on test set obtained using TCN method with different modalities: local, global and global trained using all the available data except the energy consumption related to the selected building.

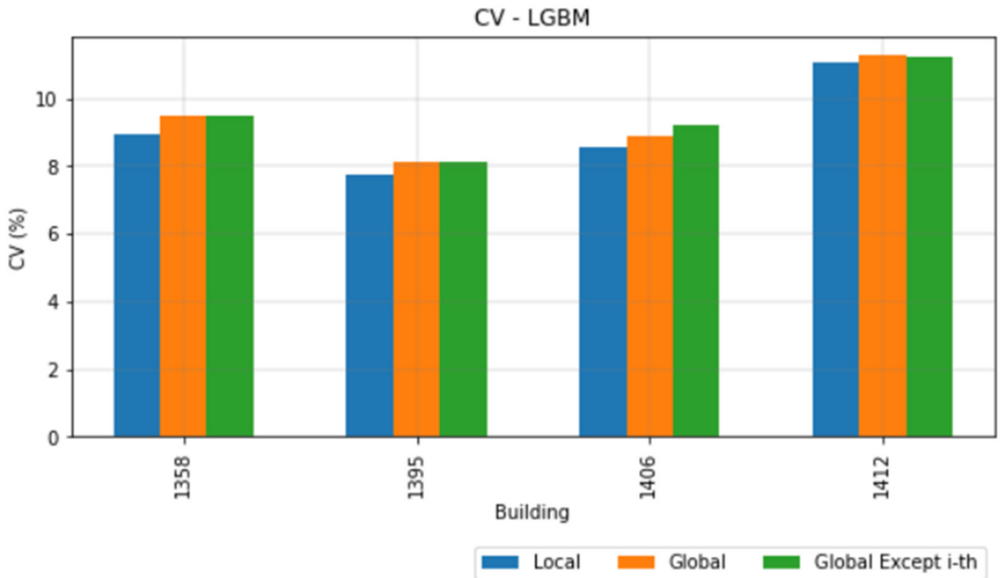

Figure 21. Comparison between the *CV* on test set obtained using LGBM method with different modalities: local, global and global trained using all the available data except the energy consumption related to the selected building.

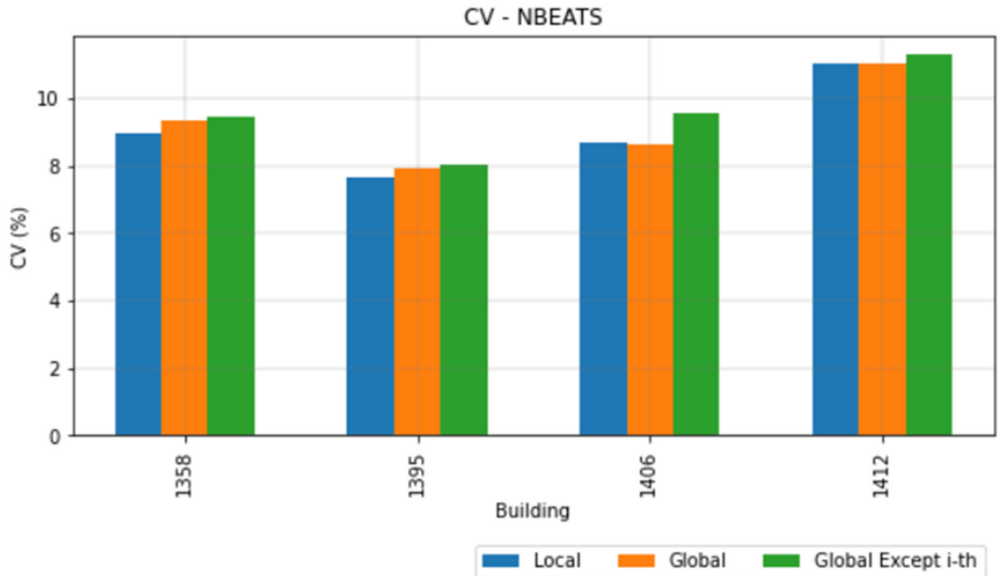

Figure 22. Comparison between the *CV* on test set obtained using NBEATS method with different modalities: local, global and global trained using all the available data except the energy consumption related to the selected building.

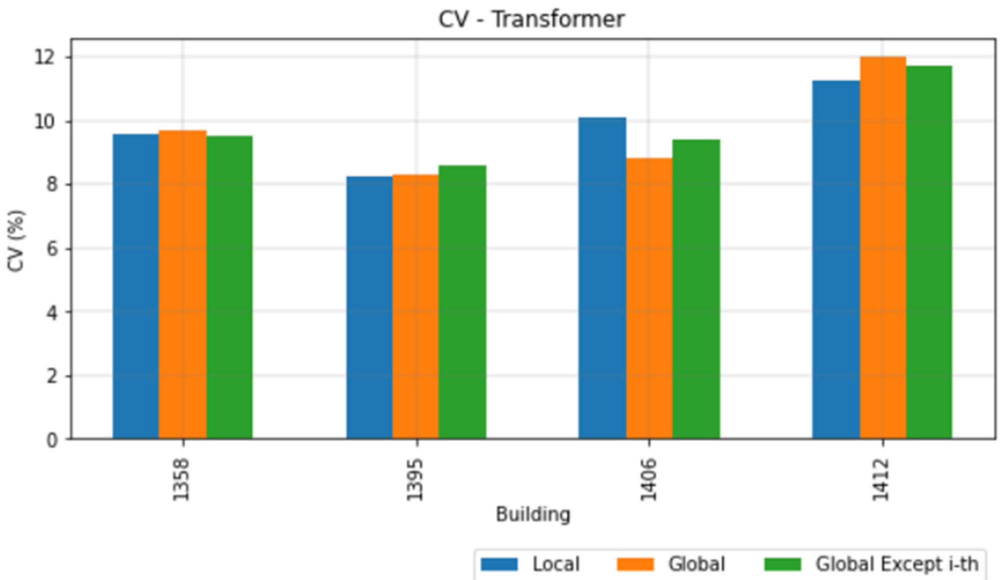

Figure 23. Comparison between the *CV* on test set obtained using Transformer method with different modalities: local, global and global trained using all the available data except the energy consumption related to the selected building.

4. Discussion

As shown in Figures 6 and 7 for the LSTM method, the global model can perform, on average, better than the local ones (Figure 16). This is probably related to the training phase of the local models, when the maximum number of epochs is reached before the minimum of validation loss is observed. As shown in Figures 14 and 15, for the Transformer method, the global model is able to perform slightly better for some buildings. This could depend on the advantage that the global models receive in having access to more data for the training phase. Concerning the other algorithms, the global models have the same or slightly lower forecasting performance than the local ones.

From Figure 16, reporting the comparison between the *CV* obtained from the considered local and global models on the test set, a clear indication about the best-performing method does not emerge, but the LGBM, NBEATS and Linear methods outperform the LSTM and TCN ones.

It is worth noting that global models can be more complex than local ones before encountering the overfitting problem [22] and likely need to have sufficient complexity in order to achieve high performance for the prediction of different building profiles. This means that global models are likely to be outperformed by the local ones using the same set of hyperparameters, if their complexity is not sufficient.

In our study, we found that local models obtained a statistically significantly lower CV than global models (3 out of 5 models, excluding LSTM). However, the observed difference is minimal, thus supporting the use of global models in conditions lacking historical data, with simplified deployment and maintenance.

The performance of the Persistence method defines a baseline that is outperformed by almost all the other approaches, except for the LSTM-Local (as discussed before). This indicates that the 1-h-ahead forecasting on this dataset cannot be solved by simply using a Persistence model, and more complex approaches are required, justifying the necessity of the learning, deployment and maintenance of the selected model.

The experiments carried out show that the global models, already known as good approaches for reducing the maintainability and deployment effort in the real context, could also be a valuable alternative in terms of performance.

The experiments on the reuse of pre-trained forecasting models, reported in Figures 18–23, show that the forecasting of the energy demand of a completely new building could be efficiently solved using a model already trained on energy demand data observed for other buildings. The only information needed is the scale of the energy demand for the new user, which can be estimated using a priori knowledge, such as the nominal power, user's category, information related to the average consumption in similar cases, etc. The impact of a wrong estimate and the assessment of the robustness of the algorithms to the uncertainty present in the estimated scale are important aspects that we plan to investigate in a future work.

5. Conclusions

In this work, several approaches have been tested to forecast 1-h-ahead electricity consumption for 28 lodging/residential buildings, by considering both local and global models. For each considered approach, a local model has been produced for each building, whereas only one global model has been produced considering the 28 buildings all together.

Two different experiments have been carried out consisting of comparing the local and global models' performance in forecasting the energy demand of a single building and forecasting—using a global model called Global-except—the energy demand of a building using the energy consumption data of the other 27 buildings. The approaches used for the experiments are the Linear Regression Model, Long Short-Term Model, Temporal Convolutional Network, NBEATS, LightGBM, Transformer and Persistence. The performance of each approach has been evaluated by means of the Coefficient of Variation and the Root Mean Square Error.

Results highlight that the global models represent a valuable alternative to local models in predicting energy consumption, presenting at the same time benefits in terms of reducing the complexity of deployment and maintainability of the forecasting solutions. Moreover, the results show the efficacy of the Global-except model. This result is remarkable as it reveals how, without any assumption on the characteristics of the time series involved, the forecasting results obtained on a completely new building could be obtained using a global model previously trained on existing buildings, providing a significant advantage to the smart grid/energy community manager.

Author Contributions: Conceptualization, A.B., M.C., A.P. and G.S.; Data curation, A.B.; Investigation, A.B., M.C., A.P. and G.S.; Methodology, A.B., M.C., A.P. and G.S.; Software, A.B.; Supervision, M.V. and G.G.; Validation, A.B., M.C., A.P. and G.S.; Visualization, A.B.; Writing—original draft, A.B., M.C., A.P., G.S. and M.V.; Writing—review and editing, A.B., M.C., A.P., G.S., M.V. and G.G. All authors have read and agreed to the published version of the manuscript.

Funding: The project has been jointly funded by the European Union and Italian Research and University Ministry (MIUR) under the Programma Operativo Nazionale "Ricerca e Innovazione" 2014–2020 (PON "R&I" 2014-2020) Grant Number ARS01_01259.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data used for the experiments are publicly available on the site of the Great Energy Predictor III competition (the website of the competition is: https://www.kaggle.com/c/ashrae-energy-prediction, accessed on 1 July 2021; the direct link to the data is: https://www.kaggle.com/c/ashrae-energy-prediction/data, accessed on 1 July 2021).

Acknowledgments: The work is part of the Research and Innovation Project "Community Energy Storage: Gestione Aggregata di Sistemi d'Accumulo dell'Energia in Power Cloud (ComESto)"—cod. ARS01_01259.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. IEA. *Electricity Market Report*; IEA: Paris, France, 2021.
2. Cao, Z.; Han, X.; Lyons, W.; O'Rourke, F. Energy management optimisation using a combined Long Short-Term Memory recurrent neural network–Particle Swarm Optimisation model. *J. Clean. Prod.* **2021**, *326*, 129246. [CrossRef]
3. Burgio, D.; Menniti, N.; Sorrentino, A.; Pinnarelli, Z.L. Influence and Impact of Data Averaging and Temporal Resolution on the Assessment of Energetic, Economic and Technical Issues of Hybrid Photovoltaic-Battery Systems. *Energies* **2020**, *13*, 354. [CrossRef]
4. Zheng, J.; Xu, C.; Zhang, Z.; Li, X. Electric load forecasting in smart grids using Long-Short-Term-Memory based Recurrent Neural Network. In Proceedings of the 2017 51st Annual Conference on Information Sciences and Systems (CISS), Baltimore, MD, USA, 22–24 March 2017.
5. Zhao, H.; Magoulès, F. A review on the prediction of building energy consumption. *Renew. Sustain. Energy Rev.* **2012**, *16*, 3586–3592. [CrossRef]
6. Zhang, N.; Li, Z.; Zou, X.; Quiring, S.M. Comparison of three short-term load forecast models in Southern California. *Energy* **2019**, *189*, 116358. [CrossRef]
7. Daut, M.A.M.; Hassan, M.Y.; Abdullah, H.; Rahman, H.A.; Abdullahm, M.P.; Hussin, F. Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review. *Ren. Sust. Energy Rev.* **2017**, *70*, 1108–1111. [CrossRef]
8. Liang, Y.; Niu, D.; Hong, W.C. Short term load forecasting based on feature extraction and improved general regression neural network model. *Energy* **2019**, *166*, 653–663. [CrossRef]
9. Mandal, P.; Senjyu, T.; Urasaki, N.; Funabashi, T. A neural network based severalhour-ahead electric load forecasting using similar days approach. *Elec. Power Energy Syst.* **2006**, *28*, 367–373. [CrossRef]
10. Shi, H.; Xu, M.; Li, R. Deep Learning for Household Load Forecasting—A Novel Pooling Deep RNN. *IEEE Trans. Smart Grid* **2018**, *9*, 5271–5280. [CrossRef]
11. Zhang, F.; Deb, C.; Lee, S.E.; Yang, J.; Shah, K.W. Time series forecasting for building energy consumption using weighted support vector regression with differential evolution optimization technique. *Energy Build.* **2016**, *12*, 94–103. [CrossRef]
12. Zhang, X.M.; Grolinger, K.; Capretz, M.A.M.; Seewald, L. Forecasting Residential Energy Consumption: Single Household Perspective. In Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications, Orlando, FL, USA, 17–20 December 2018.
13. Kong, W.; Dong, Z.Y.; Hill, D.J.; Luo, F.; Xu, Y. Short-Term Residential Load Forecasting Based on Resident Behaviour Learning. *IEEE Trans. Power Syst.* **2018**, *33*, 1087–1088. [CrossRef]
14. Rafati, A.; Joorabian, M.; Mashhour, E. An efficient hour-ahead electrical load forecasting method based on innovative features. *Energy* **2020**, *201*, 117511. [CrossRef]
15. Abbasi, R.A.; Javaid, N.; Ghuman, M.N.J.; Khan, Z.A.; Rehman, S.U. Short Term Load Forecasting Using XGBoost. In *Web, Artificial Intelligence and Network Applications. WAINA 2019. Advances in Intelligent Systems and Computing*; Barolli, L., Takizawa, M., Xhafa, F., Enokido, T., Eds.; Springer: Cham, Switzerland, 2019; Volume 927.

16. Caliano, M.; Buonanno, A.; Graditi, G.; Pontecorvo, A.; Sforza, G.; Valenti, M. Consumption based-only load forecasting for individual households in nanogrids: A case study. In Proceedings of the 12th AEIT International Annual Conference, Web-Conference AEIT, Online, 22–25 October 2020.
17. Januschowski, T.; Gasthaus, J.; Wang, Y.; Salinas, D.; Flunkert, V.; Bohlke-Schneider, M.; Callot, L. Criteria for classifying forecasting methods. *Int. J. Forecast.* **2020**, *36*, 167–177. [CrossRef]
18. Wagner, N.; Michalewicz, Z.; Schellenberg, S.; Chirac, C.; Mohais, A. Intelligent techniques for forecasting multiple time series in real-world systems. *Int. J. Intell. Comput. Cybern.* **2011**, *4*, 284–310. [CrossRef]
19. Makridakis, S.; Spiliotis, E.; Assimakopoulos, V. The M4 Competition: 100,000 time series and 61 forecasting methods. *Int. J. Forecast.* **2020**, *36*, 54–74. [CrossRef]
20. Makridakis, S.; Spiliotis, E.; Assimakopoulos, V. The M5 Accuracy competition: Results, findings and conclusions. *Int. J. Forecast.* 2022, *corrected proof*. [CrossRef]
21. Laptev, N.; Yosinski, J.; Erran Li, L.; Smyl, S. Time-series Extreme Event Forecasting with Neural Networks at Uber. In Proceedings of the International Conference of Machine Learning, Sydney, Australia, 6–11 August 2017.
22. Montero-Manso, P.; Hyndman, R.J. Principles and algorithms for forecasting groups of time series: Locality and globality. *Int. J. Forecast.* **2021**, *37*, 1632–1653. [CrossRef]
23. Herzen, J. Training Forecasting Models on Multiple Time Series with Darts. *Unit8*, 6 July 2021. Available online: https://unit8.com/resources/training-forecasting-models/(accessed on 1 December 2021).
24. Hewamalage, H.; Bergmeir, C.; Bandara, K. Global models for time series forecasting: A Simulation study. *Pattern Recognit.* **2022**, *124*, 108441. [CrossRef]
25. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
26. Fan, C.; Sun, Y.; Xiao, F.; Ma, J.; Lee, D.; Wang, J.; Tseng, Y.C. Statistical investigations of transfer learning-based methodology for short-term building energy predictions. *Appl. Energy* **2020**, *262*, 114499. [CrossRef]
27. Mocanu, E.; Nguyen, P.H.; Kling, W.L.; Gibescu, M. Unsupervised energy prediction in a Smart Grid context using reinforcement cross-building transfer learning. *Energy Build.* **2016**, *116*, 646–655. [CrossRef]
28. Ribeiro, M.; Grolinger, K.; El Yamany, H.F.; Higashino, W.A.; Capretz, M.A. Transfer learning with seasonal and trend adjustment for cross-building energy forecasting. *Energy Build.* **2018**, *165*, 352–363. [CrossRef]
29. Ahn, Y.; Kim, B. Prediction of building power consumption using transfer learning-based reference building and simulation dataset. *Energy Build.* **2022**, *258*, 111717. [CrossRef]
30. Wu, D.; Wang, B.; Precup, D.; Boulet, B. Multiple Kernel Learning-Based Transfer Regression for Electric Load Forecasting. *IEEE Trans. Smart Grid* **2020**, *11*, 1183–1192. [CrossRef]
31. Genov, E.; Petridis, S.; Iliadis, P.; Nikopoulos, N.; Coosemans, T.; Massagie, M.; Camargo, L. Short-Term Load Forecasting in a microgrid environment: Investigating the series-specific and cross-learning forecasting methods. *J. Phys. Conf. Ser.* **2021**, *2042*, 012035. [CrossRef]
32. Bai, S.; Kolter, J.Z.; Koltun, V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv* **2018**, arXiv:1803.01271.
33. Oreshkin, B.N.; Carpov, D.; Chapados, N.; Bengio, Y. N-BEATS: Neural Basis Expansion Analysis for Interpretable Time Series Forecasting. *arXiv* **2020**, arXiv:1905.10437.
34. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T. LightGBM: A highly efficient gradient boosting decision tree. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
35. Vaswani, N.; Shazeer, N.; Parmar, J.; Uszkoreit, L.; Jones, A.; Gomez, N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 6000–6010.
36. Miller, P.; Arjunan, A.; Kathirgamanathan, C.; Fu, J.; Roth, J.; Young Park, C.; Balbach, K.; Gowri, Z.; Nagy, A.D.F.; Haberl, F. The ASHRAE Great Energy Predictor III competition: Overview and results. *Sci. Technol. Built Environ.* **2020**, *26*, 1427–1447. [CrossRef]
37. Sen, R.; Yu, H.F.; Inderjit, D. Think Globally, Act Locally: A Deep Neural Network Approach to High-Dimensional Time Series Forecasting. In Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, BC, Canada, 10–12 December 2019.
38. Bandara, K.; Shi, P.; Bergmeir, C.; Hewamalage, H.; Tran, Q.; Seaman, B. Sales demand forecast in e-commerce using a long short-term memory neural network methodology. In *Neural Information Processing. ICONIP 2019. Lecture Notes in Computer Science*; Gedeon, T., Wong, K., Lee, M., Eds.; Springer: Cham, Switzerland, 2019; Volume 11955.
39. Herzen, J.; Lässig, F.; Piazzetta, S.G.; Neuer, T.; Tafti, L.; Raille, G.; Van Pottelbergh, T.; Pasieka, M.; Skrodzki, A.; Huguenin, N.; et al. Darts: User-Friendly Modern Machine Learning for Time Series. *arXiv* **2021**, arXiv:2110.03224.