Article

Short-Term Probabilistic Load Forecasting in University Buildings by Means of Artificial Neural Networks

Carla Sahori Seefoo Jarquin *,†, **Alessandro Gandelli** † [ID], **Francesco Grimaccia** † [ID] **and Marco Mussetta** † [ID]

Department of Energy, Politecnico di Milano, 20133 Milan, Italy; marco.mussetta@polimi.it (M.M.)
* Correspondence: carlasahori.seefoo@mail.polimi.it
† These authors contributed equally to this work.

Abstract: Understanding how, why and when energy consumption changes provides a tool for decision makers throughout the power networks. Thus, energy forecasting provides a great service. This research proposes a probabilistic approach to capture the five inherent dimensions of a forecast: three dimensions in space, time and probability. The forecasts are generated through different models based on artificial neural networks as a post-treatment of point forecasts based on shallow artificial neural networks, creating a dynamic ensemble. The singular value decomposition (SVD) technique is then used herein to generate temperature scenarios and project different futures for the probabilistic forecast. In additional to meteorological conditions, time and recency effects were considered as predictor variables. Buildings that are part of a university campus are used as a case study. Though this methodology was applied to energy demand forecasts in buildings alone, it can easily be extended to energy communities as well.

Keywords: energy forecasting; probabilistic forecasting; time series analysis; singular value decomposition; clustering

1. Introduction

The world is currently undergoing substantial changes, and uncertainty is part of our day to day lives. In today's context, energy has proven to be critical to sustain and foster development. Every country shall strive for a balanced trilemma [1], where energy security, energy equity and environmental sustainability are met. Optimization in energy operations are important both for cost and energy performance reasons, with related effects on the environment in the next future energy transition scenario [2]. To achieve this, understanding how, why and when the energy demand will change is crucial. In this regard, energy forecasts provide a great service. Depending on business needs, different forecast horizons must be considered, which can be coarsely divided into very short-term, in the range between 1 and 30 min [3]; short-term, for intraday forecasting [4]; day ahead forecasting [5]; and long-term forecasting, for horizons greater than one day. This classification can change from author to author [6], roughly dividing the forecasts into short-term load forecast (STLF) or long-term load forecast (LTLF), with a cut-off horizon of two weeks. The time horizon choice depends on business needs or applications. For instance, STLF might be of interest for demand response, hour-ahead or day-ahead scheduling, or energy trading or unit commitment, whereas for system planning or energy policy making, one could be more interested in scales of years or decades. Table 1 depicts the load forecast applications and classification according to [6].

This research concerns forecasts for the day ahead, therefore lying within the STLF. There exist two main perspectives regarding how to generate forecasts: weather point forecast—also known as deterministic—or probabilistic forecast. The first one, literally, provides a prediction at each step of the time horizon based on the most likely future. It assumes every event to be the inevitable result of antecedent causes, while the probabilistic

electric load forecasting (PLF) provides a probability density function (PDF) or a prediction interval (PI), thus spanning the five inherent dimensions of a forecast: three dimensions, time and probability. For instance, in [7], a point forecast model was trained employing an ensemble of regression trees, then suitable input scenarios created for obtaining a series of forecasts, and accordingly, the percentiles. This resulted in a more valuable aid tool for decision makers throughout the energy supply chain. Indeed, unlike deterministic forecasts, PLF provides a range of plausible values and their associated probability, allowing to express the associated uncertainty of the process, thus providing a broader knowledge of the predictions and enabling the stakeholders to make better informed decisions [8].

Table 1. Load forecasting applications and classification adapted from [6].

Category	Business Needs	Cut-Off Horizon
STLF	Demand response, hour or day-ahead scheduling, energy trading, unit commitment	≤2 weeks
LTLF	System planning, energy policy	>2 weeks

Recent studies consider ambient weather conditions, building components and occupants' behaviors in order to take into account the complexity of factors that can affect the building's energy model development and its computation [9].

Additionally, several mechanisms have been proposed in the literature showing how feature selection allows to reduce the forecast bias. For instance, in [10], a big data approach to variable selection was taken, allowing the algorithm to select a large amount of lagged and moving average temperature variables to enhance the forecast accuracy.

There are different techniques that may be adopted to generate a short-term probabilistic load forecast. They can be divided into statistical or artificial intelligence approaches. Within the first group, one can find models such as multiple linear regression (MLR), semiparametric additive, autoregressive moving average and exponential smoothing. In the second group, there are Artificial Neural Networks (ANN), Fuzzy Regression (FR), Support Vector Machine (SVM) regression, Gradient Boosting (GB) and Long Short-Term Memory (LSTM) [6].

The first types are highly interpretable, they correlate the load with influencing variables, thus providing insights among the relationship between load and its drivers, e.g., regression analyses are statistical processes for estimating relationships among variables. In MLR models, the load is found in terms of explanatory variables, such as weather and nonweather variables, which influence the electrical load; their influence is then identified on the basis of correlation analysis on the explanatory variables with the load [11].

The second type, on the other hand, does not need the forecaster to designate the functional form of the relationship between load and drivers. They are able to represent more complex relationships; they can take multiple variables to perform the forecast, and they are more resistant to overfitting [12].

Among these, Autoregressive Integrated Moving Average (ARIMA) models provide an effective linear model of stationary time series, as they are capable of modeling unknown processes with the minimum number of parameters and are often used for load forecasting.Nevertheless, they have the main drawback of accounting only for past and current data points, while ignoring other influential elements such as weather or calendar variables [13], which have proven to be key predictors.

FR is a fuzzy variation of classical regression analysis, providing fuzzy functional relationship between dependent and independent variables, where the phenomenon under consideration not only has stochastic variability but vagueness is also present in some form, which is contrary to linear regression, which assumes that the observed deviations between the true value and the estimated one are related to errors in measurement or observations.

FR assumes that the errors are due to the indefiniteness of the system structure; these models were successfully applied for STLF, particularly focused on holidays in Korea [14].

The SVM framework is another of the most popular approaches for supervised learning: SVM-based models have the flexibility to represent complex functions and still resist overfitting; they have been succesfully applied to STLF in the last two decades [15].

GB produces a prediction model in the form of an ensemble of weak learners (independent variables with limited predictive information) into a single prediction model; it has an interesting feature of performing intrinsic variable selection. There are different GB-based predictor models reported in the literature, which are able to produce highly competitive predictions in STLF applications [16].

Advanced machine learning models can also be driven by the electricity consumption pattern associated with the day and time; in [9], the authors demonstrate benchmarking results derived from three different machine learning algorithms, namely SVM, GB and LSTM, which were trained by using 1-year datasets with subhourly (30 min) temporal granularity. Several other comparisons of traditional methods such as Linear Regression and SVM versus Multilayer Perceptron (MLP) and LSTM can be found in the literature, all applied to day-forward predictions in a context of microgrid clusters. Despite their success, LSTM-based architectures struggle when complex and multiscale temporal dependencies are present [17].

The aim of this research is to generate probabilistic load forecasts for the day ahead in buildings by taking a dynamic ensemble approach [18]. The results are intended to explore the generation of probabilistic forecasts based on temperature scenarios and to study the main drivers for energy consumption in the studied buildings. In this work, in particular, ten different models based on artificial neural networks are proposed. The generated forecasts employ Recurrent Neural Network (RNN) models; these are advantageous because of the relative ease to train them, the availability of open-source tools to do so, their capacity to represent complex relationships and the available computational power [19]. Furthermore, for the sake of an open-source approach, this study was performed with the Python 3.8 interpreter.

The paper is structured as follows: In the following Section 2, the considered case study is presented. In Section 3, the implemented steps to generate the probabilistic load forecast are illustrated. Section 4 reports the discussion of obtained key results and, lastly, in Section 5, the conclusions and final remarks are discussed.

2. Case Study

Gas-heated buildings that are part of Politecnico di Milano were used as the case study; they are located in Bovisa campus, Milano, Italy. The analyzed cases are buildings CT4, B12, B22 and B23, as shown in Figure 1, as well as their aggregated consumption. The first one is a technological station and the second is mainly in use for teaching activities, while in the last two, offices and laboratories are present.

The datasets range from 1 May 2016 to 28 February 2017. Consumption measurements were collected in one-hour intervals and are shown in Figure 2. Weather datasets, on the other hand, were collected in 10-second intervals (and were further downsampled to hourly intervals), consisting of seven time series: relative humidity $H_\%$ (%), outdoor (ambient) temperature T_a (°C), wind speed V_a (m/s), wind direction V_a, dir (°), global solar irradiance G (W/m^2) and precipitation P_{mm} (mm). Figure 3 shows the daily average of each meteorological variable.

These time series were collected from the meteorological station in the Department of Energy at Politecnico di Milano in Bovisa campus. The location of the meteorological station is also shown in Figure 1.
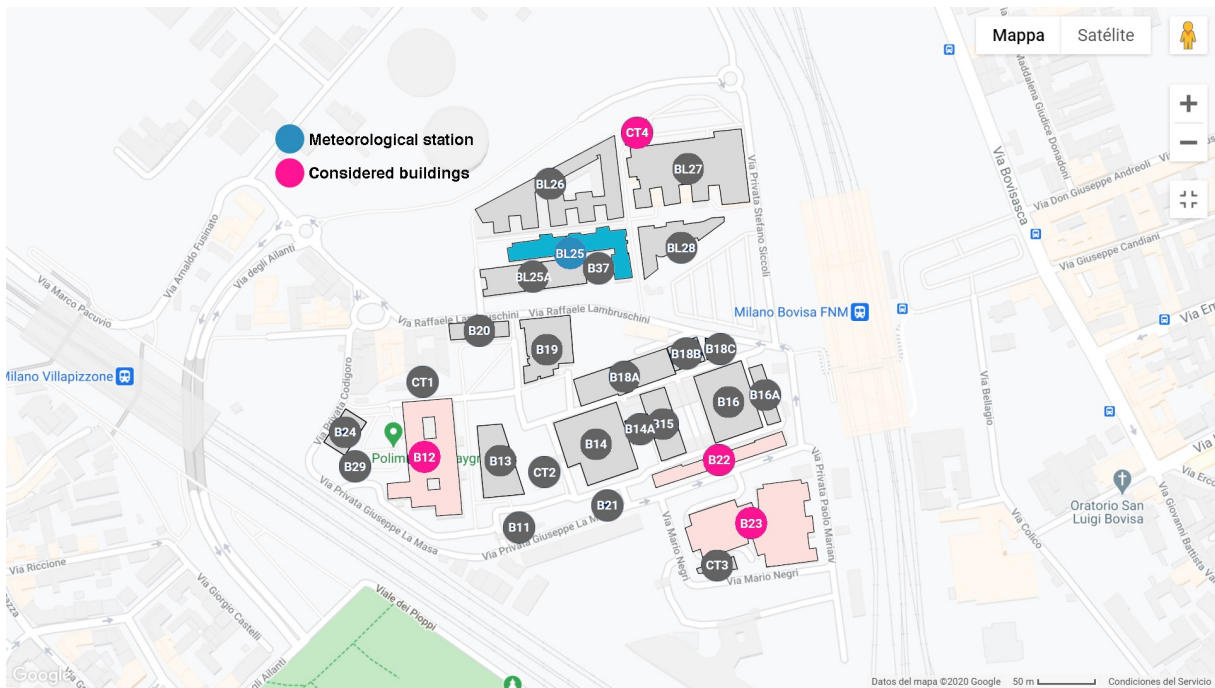
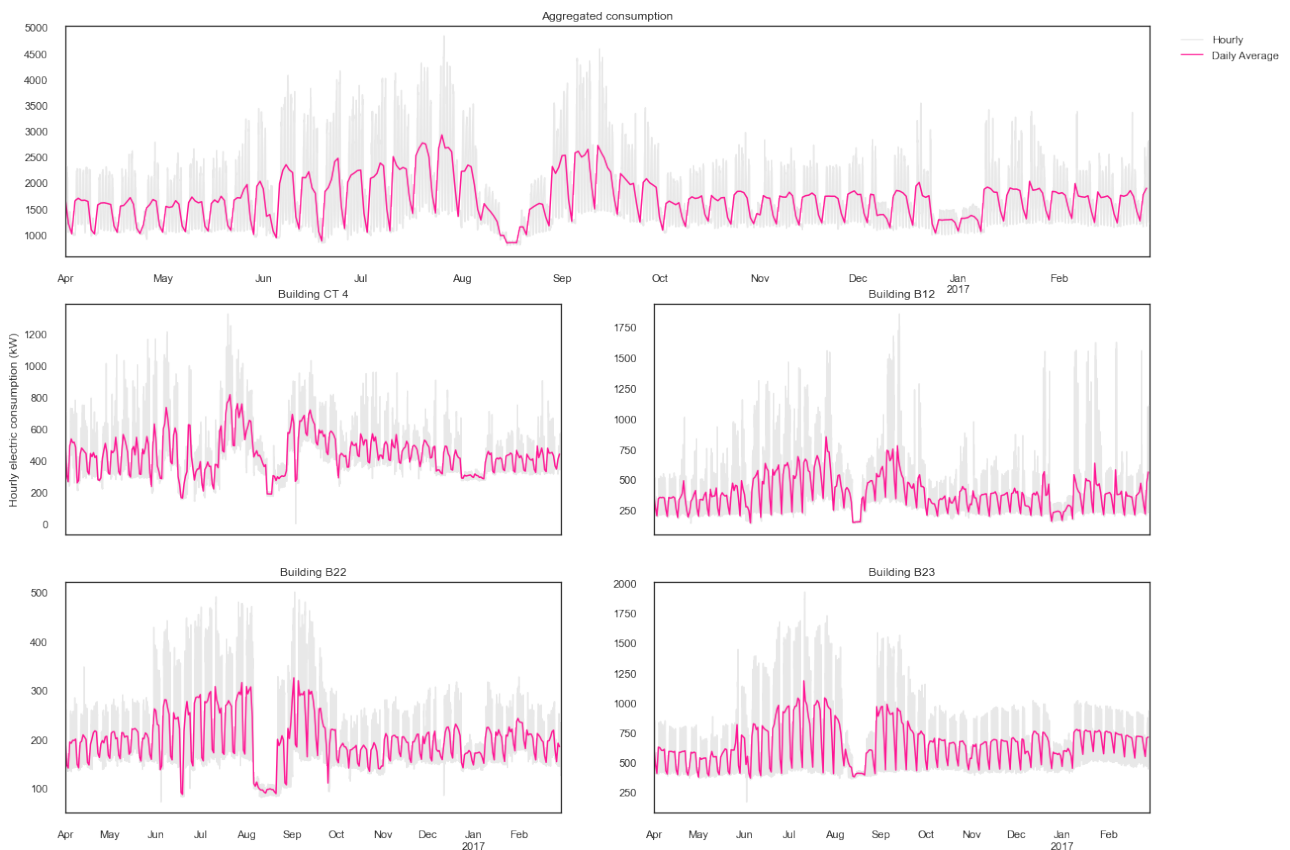**Figure 1.** Location of buildings in the case study at the university campus.

Figure 2. Electricity consumption in each building: The gray line shows the electricity consumption per hour (kW), while the dark line presents the daily average (kWh/d) for each time series.
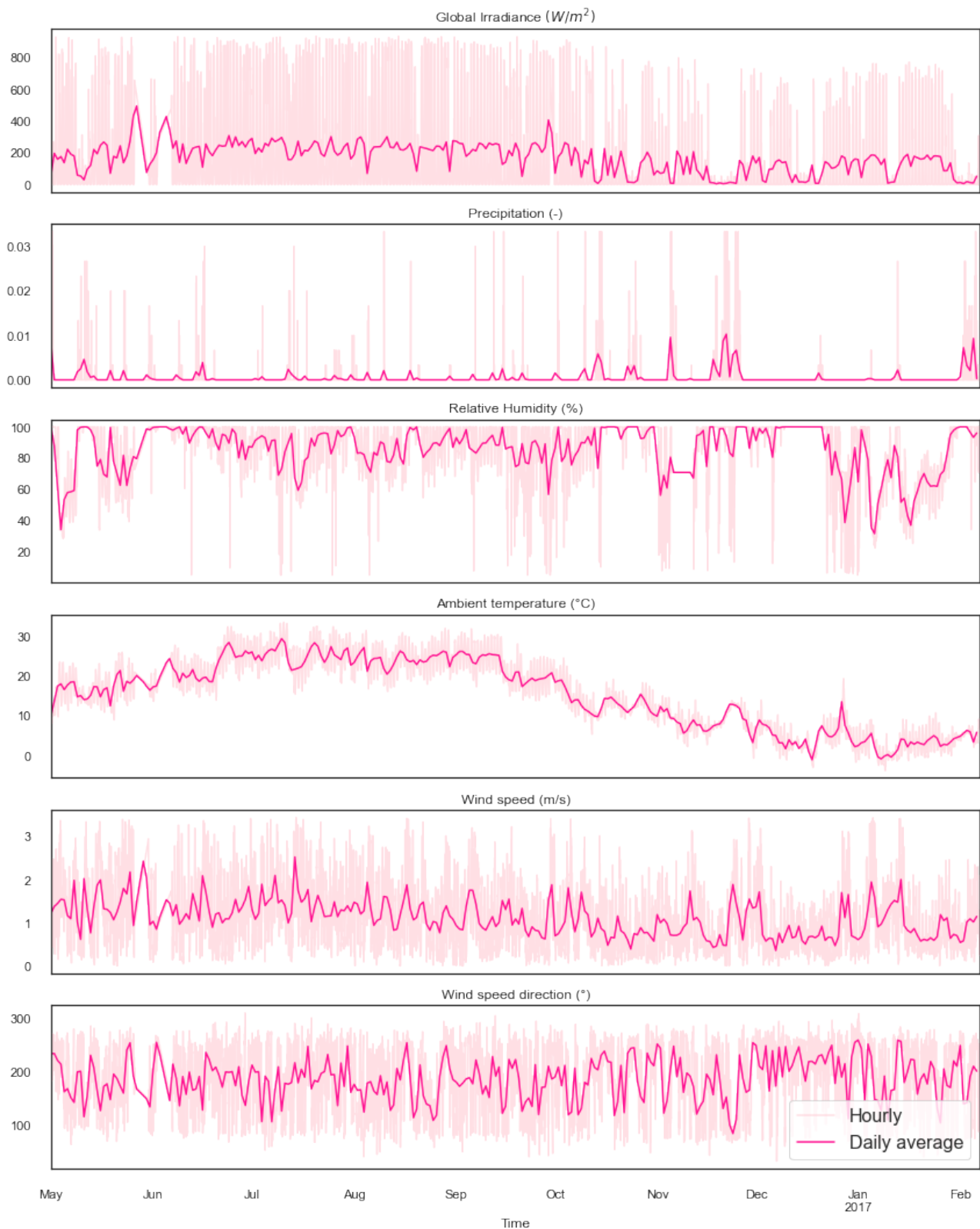
Figure 3. Hourly and daily average metereological variables for the studied time range.

3. Methods Implementation

To generate the probabilistic load forecasts, different steps were undertaken. One can visualize it as a five-stage process: the similar day method, where the data were preprocessed to find consumption phenotypes; nonparametric feature selection; point

forecast through artificial neural networks; and, finally, a dynamical ensemble to generate the probabilistic load forecast. This section briefly describes the methods and their implementation.

3.1. Similar Day Method

Data are observations of real-world phenomena [20]. They are the bridge between the model and the physical world [21]. Preprocessing is imperative to improve models' accuracy when representing reality, assuming the future will, somehow, look similar to the past. The approach taken in this work was to study the historical data to find the most representative patterns which underlined the hidden structures in the data. In biology, the term *phenotype* refers to observable properties of an organism, which include appearance, development and behavior.

Some examples are wing length or hair color. In energy engineering, this term is adopted to describe the properties of daily consumption profiles [22]. To find the consumption phenotypes, the load time series was transformed into strings by means of the Symbolic Aggregate ApproXimation (SAX) transformation [23]. Like so, it was possible to take advantage of robust text mining techniques and classify the daily profiles into motifs and discords, which are the terms for frequent and infrequent patterns in data mining. The patterns with a frequency of less than 1% of the total available data were identified as discords and discarded. The motifs, on the other hand, were further aggregated into clusters by means of the k-means approach. By choosing the number of clusters $k = 3$, the related profiles were identified as low-, medium- and high-consumption phenotypes.

3.2. Nonparametric Feature Selection

One of the research goals was to evaluate the importance of different effects as predictors of the load. Typically, in the literature, meteorological variables, particularly temperature, and load for the day before are employed.

Within the research framework, some other effects were explored, such as day–night and summer–winter shifts, together with the recency effect. The weather effects were enclosed by the hourly meteorological measurements. Day–night and seasonal shifts were captured by a cosine transformation of the respective hour and month indexes. The recency effect corresponds to a cognitive bias that makes people better remember the things said last in a speech. In the scope of energy demand, the recency effects explain the fact that electricity consumption is affected by temperatures of the preceding hours. This last effect is captured by the first and second derivatives of the meteorological variables.

The aforementioned effects were translated into a set of 43 candidate features. In order to select those with the greatest importance to predict the load and minimize the curse of dimensionality effects [24], feature selection was undertaken. Mutual information (MI), which is a nonparametric approach, was applied. This kind of method provides details on the shared information among two variables. It is a measure of the uncertainty reduction of a target variable, while knowing the probability distribution function of another variable. The mutual information will be zero if, and only if, the two variables do not share any information at all. A feature subset of the ten highest MI variables was explored as input features on the artificial neural networks. Note that for all the cases, the highest MI was reported by the load for the day before, which indeed made a good predictor. However, it was left out of the models in order to avoid overfitting and to better appreciate the effects of adding the other predictors.

From all 43 candidates, models using from four up to ten of the most informative features were trained for each surveyed case. The symbols correspond to those listed in Table 2: L, the load; T_a, ambient temperature; m, month index; G, global solar irradiance; h, hour index; d, day index; $H_\%$, relative humidity; and W_s, wind speed. For instance, the target day is denoted by (d), whereas $(d-1)$ indicates the day before. The first and second derivatives are denoted by $'$ and $''$, respectively. For example, for predicting the hourly load $L(d, h)$ of the d-th day for building CT4, the first most important input is the load of the

day before $L(d-1,h)$ at the same hour h; then, in the second place, the temperature on the target day $T_a(d,h)$; then, in the third place, the temperature for the day before, $T_a(d-1,h)$, at the same hour; and so on. Not surprisingly, the consumption and temperature of the day before come in the first places.

Table 2. Overview of available inputs.

	Symbols	Input	Units
1	L	Electricity consumption	kW
2	G	Global solar irradiance	W/m^2
3	P_{mm}	Precipitation	mm
4	$H_\%$	Relative humidity	%
5	T_a	Ambient temperature	°C
6	W_s	Wind speed	m/s
7	W_d	Wind direction	°
8	h	Hour index	h
9	d	Day index	–
10	m	Month index	–
11	w	Weekend index	–

3.3. Point Forecast

The tagged consumption days were split into train, validation and test sets of sizes 80%, 10% and 10%, respectively. Ten different models based on artificial neural networks were proposed per phenotype. Table 3 summarizes the models. M1 is a naive persistence model for the point forecasts, and M2 is used as a benchmark for the probabilistic forecasts. Note that the real measurements for the target day were used as proxy for an available temperature forecast. Additionally, when referring to a model trained for a specific phenotype, the subindexes L, M and H were used to refer to low, medium or high consumption. For example, $M_{10,H}$ would refer to M_{10}, trained, validated and tested on the high-consumption datasets.

Table 3. Overview of studied artificial neural network (ANN) models.

Model	Candidate Features
M_1	$T_a(d-1)$
M_2	$T_a(d-1), T(d)$
M_3	$M_2 + h, d, m$
M_4	2nd to 4th most informative features
…	…
M_{10}	2nd to 10th most informative features

All the models are feed-forward neural networks with back-propagation learning, specifically using the Adam optimizer with a learning rate equal to 0.001 [25]. Each one had one hidden layer with rectified linear unit (ReLU) as the activation function between the input and hidden layer, and the linear transfer function ($purelin(x) = x$) between the hidden and output layer. The batch size is equal to 48. The training and validation process was run 100 times. Additionally, the training was stopped if one of the following conditions was verified: (1) the number of epochs was greater than 500 or (2) the cost function was not improving for more than 10 consecutive epochs. For each model, the networks structure and the number of hidden neurons were found from employing the time series cross-validation line search, as depicted in [26]. While this is not the most efficient way to fine-tune, it gives a sense of which direction to go; the results from this step are shown in Table 4. The maximum number of neurons in the models for all the cases did not exceed 112, not even when 10 inputs were fed into the neuron. This raised the question on whether or not the feed-forward artificial neural networks were going to be able to fully capture complex relationships among the inputs.

Table 4. Number of hidden units employed per proposed model according to results from line search (continued).

Model	Number of Neurons		
	Low	**Medium**	**High**
1	22	18	22
2	21	15	30
3	66	66	60
4	44	44	44
5	57	57	57
6	66	66	66
7	68	82	82
8	76	92	68
9	66	102	84
10	72	112	62

The point forecast performance was assessed by means of the root mean squared error (RMSE) [27], coefficient of variance of the root mean squared error (CVRMSE) [28], enveloped mean absolute error (EMAE) [29] and the well-known forecast skill (S). Let us define the residual, e_h, given by the difference between an observed value y and its forecast $\hat{y}_h$, mathematically expressed as [30]

$$e_h = y_h - \hat{y}_h \tag{1}$$

Then, the aforementioned metrics are depicted as

$$RMSE = \sqrt{\frac{1}{N} \sum_{h=1}^{N} e_h^2} \tag{2}$$

$$CVRMSE = \frac{RMSE}{\bar{Y}} \cdot 100 \tag{3}$$

$$EMAE = \frac{\sum\limits_{h=1}^{N} |e_h|}{\sum\limits_{h=1}^{N} \max(y_h, \hat{y}_h)} \cdot 100 \tag{4}$$

$$S = 1 - \frac{U}{V} \approx 1 - \frac{RMSE}{RMSE_p} \tag{5}$$

where N is the forecasts number, $\bar{Y}$ is the mean of the test set load, U is the uncertainty, V is the variability of the load and $RMSE_p$ is the RMSE achieved by the persistence method. In particular, RMSE is widely used but has the drawback of being scale-dependent. Therefore, EMAE and CVRMSE were studied as well. Particular attention was paid to the latter, as the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) defines as an acceptable accuracy a value of CVRMSE $< 30\%$ for hourly predictions. Finally, the forecast skill score S is often used to evaluate whether a model is superior compared with the naive method.

3.4. Probabilistic Load Forecast

In this work, to generate probabilistic load forecasts, a dynamical ensemble approach was adopted by considering a standard of 100 simulation runs for each considered model. These kinds of ensembles are obtained by evolving an ensemble of different initial conditions [5]. Thus, the ensembles were constructed by adding perturbations on the original temperature profiles from the test set and making predictions with them. Herein, the perturbed scenarios were created by means of the truncated singular value decomposition (TSVD) [31]. First, the temperature time series was recast into a matrix $\vec{X}^{h \times d}$, where $h = 24$,

and d corresponds to the days the time series spans. Then, it was decomposed with the singular value decomposition into the product of three matrices:

$$\vec{X} = \vec{U}\vec{\Sigma}\vec{V}^{\mathrm{T}} \tag{6}$$

where $\vec{U}$ and $\vec{V}$ are orthonormal unitary matrices, known as left and right singular matrices, respectively. The column vectors from $\vec{U}$ and $\vec{V}^{\mathrm{T}}$ are known as the left and right singular vectors. Finally, $\vec{\Sigma}$ is a diagonal matrix, with the singular values on its diagonal, which correspond to the square root of the eigenvectors from the correlation matrix $\vec{X}\vec{X}^{\mathrm{T}}$.

By keeping only the leading singular values and the corresponding singular vectors to rebuild the time series with a low-rank approximation, the truncated singular value decomposition was adopted. Besides the first right singular vector, the rest were disturbed by means of Gaussian noise before going back to the time domain. Defining the standard deviation of the blended noise as $S_v = \varepsilon^2$, the added noise was gradually varied by varying ε from 0 to 0.2 in 0.01 increments. For each day of the test set, 100 scenarios were created. That way, temperature profiles were simulated for each day of the test day, allowing both to evaluate the uncertainty propagation into the point forecasts coming from faulty weather forecasts, to evolve the initial conditions of the analysis and, ultimately, create the dynamical ensembles.

After obtaining the different predictions, the percentiles from 1 to 99th were calculated and provided the prediction intervals for the probabilistic load forecast. The ANNs' performance was assessed through pinball loss, a metric specifically designed for quantile forecasts [6] and formulated as

$$L_\tau(\hat{y}_\tau, y) = \begin{cases} (1-\tau)|\hat{y}_\tau - y|, & \text{if } \hat{y}_\tau \leq y \\ \tau|\hat{y}_\tau - y|, & \text{if } \hat{y}_\tau > y \end{cases} \tag{7}$$

where $\hat{y}_\tau$ is the point forecast for quantile τ, and y is the observation. They can be computed for each quantile and summed over the forecast horizon to attain the pinball loss. A low pinball loss score indicates an accurate probabilistic model.

4. Results

The aforementioned methods were implemented to forecast the electricity consumption for the day ahead of five different cases: four buildings and their aggregated consumption. Each case was studied separately, the main drivers and consumption phenotypes were identified per case, and ten models per phenotype were proposed. The mean performance from the 100 simulation run for the dataset was plotted. Figure 4 shows the datasets' distribution along the time spanned for one of the studied buildings. The hue indicates consumption level: the darker the color, the higher the consumption. The pink shades depict training data; the validation days are in gray; and the test days are in blue. The figure allows us to observe that in the case of building CT4, low consumption levels coincide with weekends and holidays, and medium-level days are mostly in autumn and winter, thus in heating season, but also span weekends during the summer. However, the high-consumption days correspond mainly to working days in the cooling season. In fact, similar behaviors were found for the other studied cases.

Before proceeding, let us start by providing some notation used to refer to the input variables. The target day is denoted by d, while the day before corresponds to $d-1$. The ambient temperature is depicted by T_a; G corresponds to the global solar radiation; $H_\%$ refers to the relative humidity; L corresponds to the electric load; h to the hour; and m to the month. The first and second derivatives are depicted by ' and ", respectively. For example, $L'(d-1)$ would refer to the first derivative of the load from the day before.

The top ten features for each surveyed case suggest a strong dependency on seasonal effects, day–night shifts, solar heat gains through solar radiation, thermal inertia (particularly with $G(d-1)$) and the recency effect through the first and/or second derivative of the solar radiation. Interestingly, for the oldest building among the studied cases, B12,

the solar heat gains were more informative than temperature, after the hour of the day. In fact, by capturing these effects in most of the cases, the model's accuracy was enough not only to outperform M_1 and M_2 but also to comply with ASHRAE's benchmark and recent literature results with popular machine learning algorithms, such as RNN [12,19] and LSTM [9].
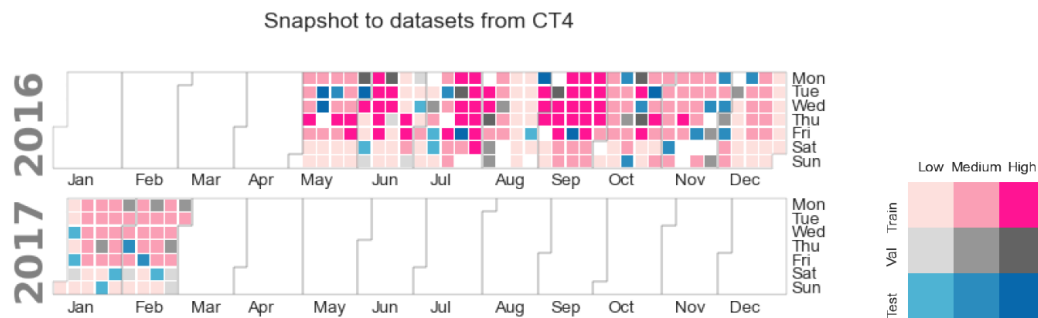

Figure 4. Composition of training, validation and test sets of building CT4.

The most accurate models on a CVRMSE basis are shown in Table 5, highlighting their predictor variables. It can be observed that most of them received calendar variables, solar radiation for the target day and the day before, as well as their respective first and second derivatives. Considering the average accuracy of the best models per phenotype, it was somewhat surprising to find low and medium consumption to have similar accuracies, with CVRMSEs = 13.83% and 13.33%, respectively. The high-consumption models, on the other hand, were behind, with CVRMSE = 17.95%. It should be noted that there were more days tagged as medium than the other phenotypes. Thus, with more training examples, it should be possible to improve the results.

Specifically looking at each level, $M_{3,L}$ was the most accurate among the proposed models, with an average CVRMSE = 12% for all the cases, as reported in Table 6, which reports the average test performance of the ten models on the aggregated consumption of the four surveyed buildings.

This is a particularly relevant result if compared with popular machine learning algorithms, such as [12,19], where similar results have been obtained by means of an RNN on the same dataset, and it is better than [9], where an LSTM approach resulted in a CVRMSE equal to 18.9%.

This indicates, indeed, that the best predictors for low-consumption days are calendar variables, such as day of the week, month of the year and hour of the day, but there is also little improvement compared with the naive benchmark. Since these days were mostly weekends and holidays, it may be worth to explore using dummy variables providing this information as predictors. In the case of medium- and high-consumption phenotypes, mainly M_9 and M_{10} were the most accurate. As commented before, these models captured information on the weather, calendar variables, solar heat gains, thermal inertia and recency effects, suggesting a greater complexity to predict the demand, as more and diversified information was needed. To provide an example on the predicted values, Figure 5 presents a test day from each consumption phenotype.

Through gradually perturbing the temperature profiles to generate predictions, the error propagation was studied. The results suggest that unless the temperature forecast completely missed the real conditions, the models' performance was not significantly affected, thus maintaining similar accuracy to when real temperature measurements were employed for the load predictions. On the one hand, this may indicate the models to be robust to uncertainty from the weather forecast. On the other hand, it may also implicate low weights assigned to the temperature input; thus, exploring a different method to generate the prediction intervals should be considered.
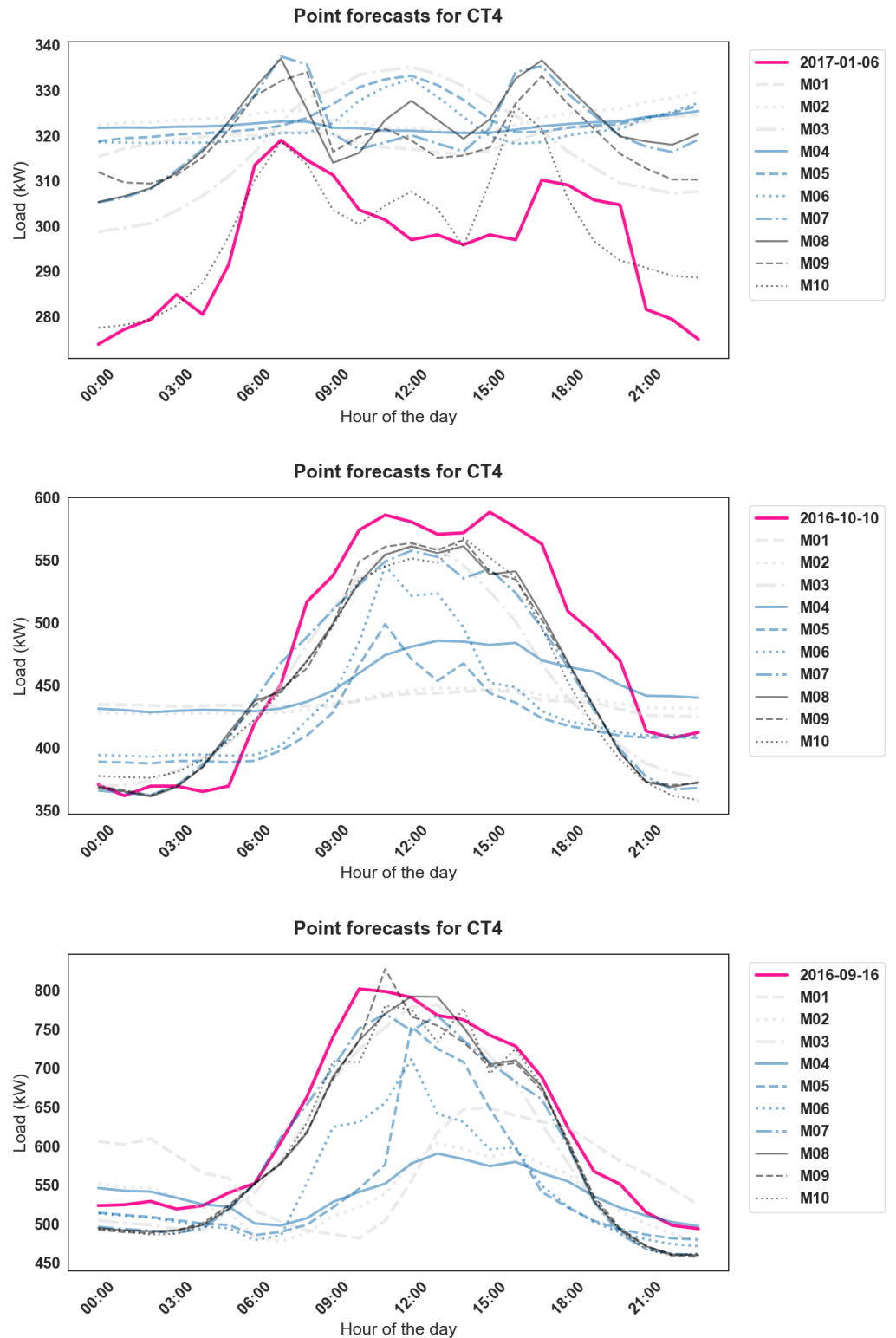
Figure 5. Point forecasts for different consumption levels of building CT4: low (**top**), medium (**middle**) and high (**bottom**). The pink line depicts the true values.

By adopting the dynamical ensemble approach, it was possible to generate prediction intervals, which better captured the true values compared with the point predictions. Figure 6 presents an example of some of the generated probabilistic forecasts for a building among the studied cases.
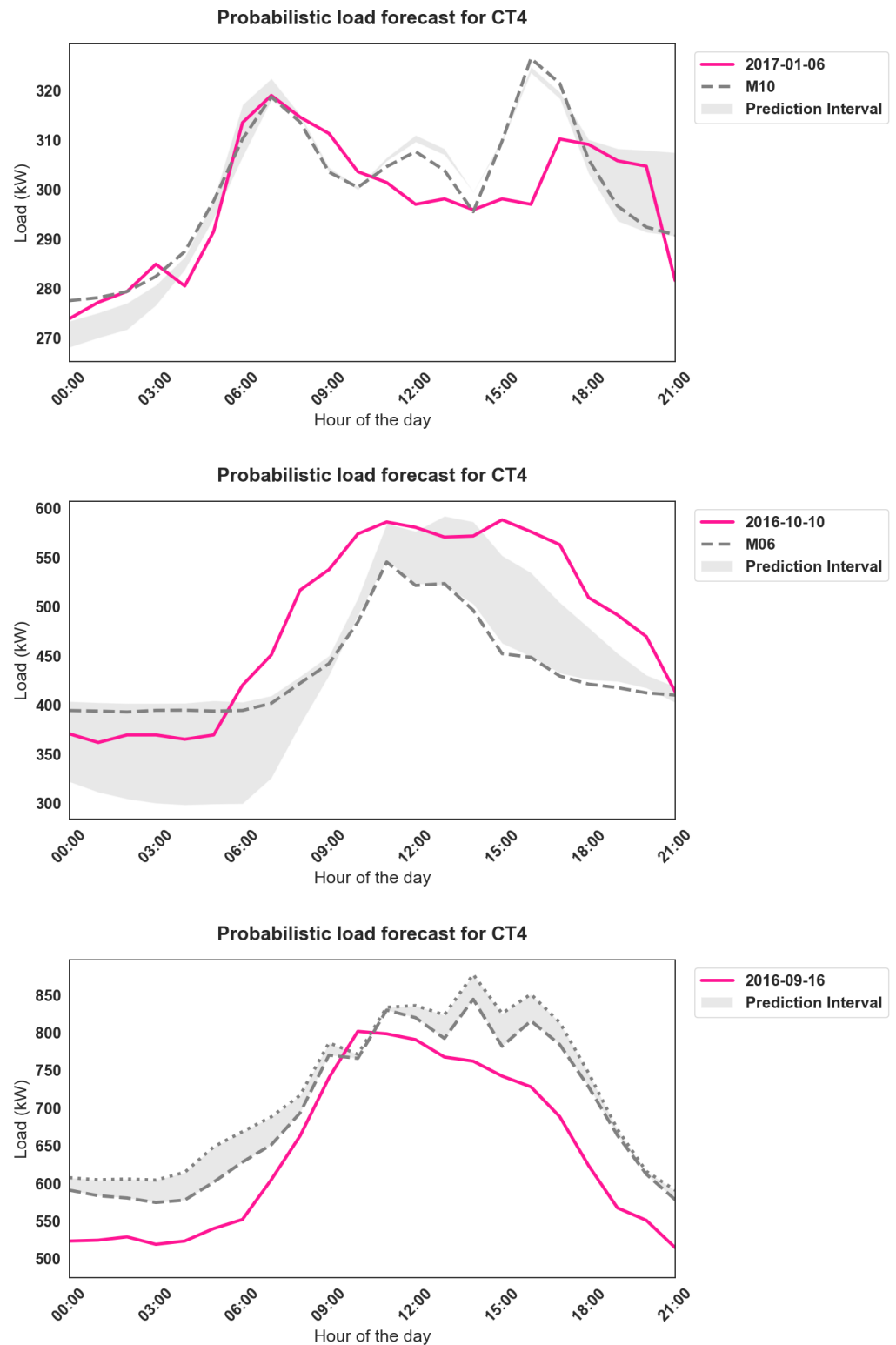
Figure 6. Probabilistic forecasts for different consumption levels of building CT4: M10, L (**top**), M6, M (**middle**) and M10, H (**bottom**). The pink line depicts the true values

From the pinball loss scores, an interesting result is remarked. Four out of five cases show the lowest loss for the low-consumption days, followed by medium and high last. Moreover, by adding information on the main drivers, the models outperformed—in most cases—the benchmark M_2. In this case, however, the pinball loss cannot be readily compa-

rable among the cases, as it is scale-dependent. After examining the results per phenotype, the following is commented. The low-consumption forecasts required simpler models than their medium- and high-consumption counterparts to achieve the lowest pinball loss. For instance, $M_{3,L}$ was among the best performing ones, adding only calendar variables to the ambient temperature. While for the other classifications, capturing additional details such as solar heat gains and recency effects was needed.

Table 5. Summary of models with the highest accuracy per phenotype of each surveyed case depicting the used predictor variables.

Building	Level	Model	Ta(d)	Ta(d-1)	m	G(d)	G(d-1)	h	G'(d)	d	RH(d)	G'(d-1)	L'(d-1)	G"(d)
	L	$M_{5,L}$	x	x	x	x								
CT4	M	$M_{10,M}$	x	x	x	x	x	x	x	x	x			
	H	$M_{8,H}$	x	x	x	x	x	x	x					
	L	$M_{3,L}$	x	x	x			x		x				
B12	M	$M_{7,M}, M_{8,M}$	x	x		x	x	x	x			o		
	H	$M_{9,H}$	x	x	x	x	x	x	x			x		
	L	$M_{2,L}$	x	x										
B22	M	$M_{10,M}$	x	x	x	x	x	x	x			x		x
	H	$M_{9,H}$	x	x	x	x	x	x	x			x		
	L	$M_{3,L}$	x	x	x			x		x				
B23	M	$M_{6,M}, M_{8,M}$	x	x	x	o	x	x	o					
	H	$M_{10,H}$	x	x	x	x	x	x	x			x	x	
	L	$M_{3,L}$	x	x	x			x		x				
All	M	$M_{10,M}$	x	x	x	x	x	x	x			x		x
	H	$M_{10,H}$	x	x	x	x	x	x	x			x		x

Table 6. Test performance of models for aggregated consumption.

Model	RMSE (kW)			CVRMSE (%)			EMAE (%)		
	L	M	H	L	M	H	L	M	H
M1	177.50	560.82	755.21	15.19	32.86	32.73	11.59	23.55	22.82
M2	165.32	548.01	723.00	14.14	32.11	31.33	10.55	23.13	21.11
M3	138.87	269.80	331.62	11.88	15.81	14.37	9.62	9.71	10.90
M4	175.38	256.62	357.10	15.01	15.04	15.47	11.37	9.22	11.38
M5	178.86	261.83	361.10	15.30	15.34	15.65	11.14	9.40	11.41
M6	185.31	264.58	345.00	15.85	15.50	14.95	11.41	9.50	11.20
M7	169.13	261.70	333.36	14.47	15.34	14.45	10.36	9.17	10.57
M8	171.28	244.53	326.53	14.65	14.33	14.15	10.45	8.60	10.67
M9	171.44	240.56	312.89	14.67	14.10	13.56	10.49	8.50	10.27
M10	169.98	230.34	305.43	14.54	13.50	13.24	10.45	8.18	9.92

5. Conclusions

Five different cases of university buildings were studied. Point and probabilistic load forecasts for the day ahead were generated by means of feed-forward neural networks. In order to train the neural networks, the similar day method was adopted. The data available were classified into three daily consumption phenotypes, identified as low, medium and high. Ten models per phenotype were proposed and evaluated, two were used as benchmarks. Besides for the high-consumption models of one building, the rest not only outperformed the benchmark models but showed acceptable accuracies according to ASHRAE's benchmark. A similar behavior was found from the probabilistic approach: providing information on different electricity consumption drivers, such as seasonal, day–night shifts and recency effects, improved the models' performance against the benchmark. This indicates that using the most informative features indeed improved the predictions.

From the point forecast approach, some of the best results were from M_3, which enclosed day–night and summer–winter shifts, thus highlighting the strong seasonal and opening hours dependency of the studied buildings, especially for weekends and holidays. However, working days tagged as medium- or high-consumption required more effects,

such as heat gains and recency effects; indeed, models M_9 and M_{10} were often times the most accurate.

Results from the error propagation analysis confirm the point forecasts to be robust to faulty weather forecast. Although employing a dynamical ensemble approach allowed to generate prediction intervals which better captured the true values in comparison with the point predictions, the calculated prediction intervals were too narrow, and overall, the considered approach did not provide useful prediction intervals. Since this method was applied unsuccessfully, it may be worth exploring a different approach to generate prediction intervals, such as fitting a probability density function on the residuals instead of the dynamic ensemble. Another option would be to generate scenarios based on solar heat gains rather than on temperature.

Finally, regarding future work, this research can be extended in multiple directions. The methodology was applied only for electric load, but provided there are available data, it could be employed for heat demand. Moreover, its application for different use types such as residential, commercial or industrial facilities could be explored. Furthermore, applying it within the smart cities paradigm and as a tool for energy communities could be viable as well. In fact, within the collective action approach, it could be possible to perform hierarchical forecasting, thus predicting the individual demand per unit and conforming the community to then forecast the aggregated consumption as a whole.

References

1. Šprajc, P.; Bjegović, M.; Vasić, B. Energy security in decision making and governance-Methodological analysis of energy trilemma index. *Renew. Sustain. Energy Rev.* **2019**, *114*, 109341. [CrossRef]
2. Grimaccia, F.; Niccolai, A.; Mussetta, M.; D'Alessandro, G. ISO 50001 Data Driven Methods for Energy Efficiency Analysis of Thermal Power Plants. *Appl. Sci.* **2023**, *13*, 1368. [CrossRef]
3. Niccolai, A.; Orooji, S.; Matteri, A.; Ogliari, E.; Leva, S. Irradiance Nowcasting by Means of Deep-Learning Analysis of Infrared Images. *Forecasting* **2022**, *4*, 338–348. [CrossRef]
4. Veeramsetty, V.; Chandra, D.R.; Grimaccia, F.; Mussetta, M. Short Term Electric Power Load Forecasting Using Principal Component Analysis and Recurrent Neural Networks. *Forecasting* **2022**, *4*, 149–164. [CrossRef]
5. Pretto, S.; Ogliari, E.; Niccolai, A.; Nespoli, A. A New Probabilistic Ensemble Method for an Enhanced Day-Ahead PV Power Forecast. *IEEE J. Photovoltaics* **2022**, *12*, 581–588. [CrossRef]
6. Hong, T.; Fan, S. Probabilistic electric load forecasting: A tutorial review. *Int. J. Forecast.* **2016**, *32*, 914–938. [CrossRef]
7. Khoshrou, A.; Pauwels, E.J. Short-term scenario-based probabilistic load forecasting: A data-driven approach. *Appl. Energy* **2019**, *238*, 1258–1268. [CrossRef]
8. Van der Meer, D.W.; Munkhammar, J.; Widén, J. Probabilistic forecasting of solar power, electricity consumption and net load: Investigating the effect of seasons, aggregation and penetration on prediction intervals. *Sol. Energy* **2018**, *171*, 397–413. [CrossRef]
9. Huang, J.; Algahtani, M.; Kaewunruen, S. Energy Forecasting in a Public Building: A Benchmarking Analysis on Long Short-Term Memory (LSTM), Support Vector Regression (SVR), and Extreme Gradient Boosting (XGBoost) Networks. *Appl. Sci.* **2022**, *12*, 9788. [CrossRef]

10. Wang, P.; Liu, B.; Hong, T. Electric load forecasting with recency effect: A big data approach. *Int. J. Forecast.* **2016**, *32*, 585–597. [CrossRef]

11. Moghram, I.; Rahman, S. Analysis and evaluation of five short-term load forecasting techniques. *IEEE Trans. Power Syst.* **1989**, *4*, 1484–1491. [CrossRef] [PubMed]

12. Mansoor, M.; Grimaccia, F.; Leva, S.; Mussetta, M. Comparison of echo state network and feed-forward neural networks in electrical load forecasting for demand response programs. *Math. Comput. Simul.* **2021**, *184*, 282–293. [CrossRef]

13. Zhang, X.; Wang, J.; Zhang, K. Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by Cuckoo search algorithm. *Electr. Power Syst. Res.* **2017**, *146*, 270–285. [CrossRef]

14. Song, K.B.; Baek, Y.S.; Hong, D.H.; Jang, G. Short-term load forecasting for the holidays using fuzzy linear regression method. *IEEE Power Eng. Soc. Gen. Meet.* **2005**, *2*, 96–101. [CrossRef]

15. Chen, B.J.; Chang, M.W.; Lin, C.J. Load forecasting using support vector machines: A study on EUNITE Competition 2001. *IEEE Trans. Power Syst.* **2004**, *19*, 1821–1830. [CrossRef]

16. Lloyd, J.R. GEFCom2012 hierarchical load forecasting: Gradient boosting machines and Gaussian processes. *Int. J. Forecast.* **2014**, *30*, 369–374. [CrossRef]

17. Giglio, E.; Luzzani, G.; Terranova, V.; Trivigno, G.; Niccolai, A.; Grimaccia, F. An Efficient Artificial Intelligence Energy Management System for Urban Building Integrating Photovoltaic and Storage. *IEEE Access* **2023**, *11*, 18673–18688. [CrossRef]

18. Nespoli, A.; Leva, S.; Mussetta, M.; Ogliari, E.G.C. A Selective Ensemble Approach for Accuracy Improvement and Computational Load Reduction in ANN-Based PV Power Forecasting. *IEEE Access* **2022**, *10*, 32900–32911. [CrossRef]

19. Valinataj Bahnemiri, P.; Grimaccia, F.; Leva, S.; Mussetta, M. Optimization of Neural Network-based Load Forecasting by Means of Whale Optimization Algorithm. In *Proceedings of the ELECTRIMACS 2022: Selected Paper*; Pierfederici, S., Martin, J.P., Eds.; Springer International Publishing: Cham, Switzerland, 2023; Lecture Notes in Electrical Engineerings, Volume 993.

20. Zheng, A.; Casari, A. *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2018.

21. Ranjan, K.G.; Prusty, B.R.; Jena, D. Review of preprocessing methods for univariate volatile time-series in power system applications. *Electr. Power Syst. Res.* **2021**, *191*, 106885. [CrossRef]

22. Miller, C.; Nagy, Z.; Schlueter, A. Automated daily pattern filtering of measured building performance data. *Autom. Constr.* **2015**, *49*, 1–17. [CrossRef]

23. Keogh, E.; Lin, J.; Fu, A. Hot SAX: Efficiently finding the most unusual time series subsequence. In Proceedings of the 5th IEEE International Conference on Data Mining (ICDM'05), IEEE, Houston, TX, USA, 27–30 November 2005; p. 8.

24. Chen, L. Curse of dimensionality. In *Encyclopedia of Database Systems*; Springer: Berlin/Heidelberg, Germany, 2009.

25. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.

26. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016. Available online: http://www.deeplearningbook.org (accessed on 12 April 2023).

27. Van der Meer, D.W.; Widén, J.; Munkhammar, J. Review on probabilistic forecasting of photovoltaic power production and electricity consumption. *Renew. Sustain. Energy Rev.* **2018**, *81*, 1484–1512. [CrossRef]

28. Bagnasco, A.; Fresi, F.; Saviozzi, M.; Silvestro, F.; Vinci, A. Electrical consumption forecasting in hospital facilities: An application case. *Energy Build.* **2015**, *103*, 261–270. [CrossRef]

29. Dolara, A.; Grimaccia, F.; Leva, S.; Mussetta, M.; Ogliari, E. Comparison of training approaches for photovoltaic forecasts by means of machine learning. *Appl. Sci.* **2018**, *8*, 228. [CrossRef]

30. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice, 2nd edition*; OTexts, Melbourne, Australia: 2018.

31. Glorot, X.; Bordes, A.; Bengio, Y. Deep sparse rectifier neural networks. In Proceedings of the 14th International Conference on Artificial Intelligence and Statistics, Fort Lauderdale, FL, USA, 11–13 April 2011; pp. 315–323.