information

Article

Design of a Meaningful Framework for Time Series Forecasting in Smart Buildings

Louis Closson [1,*], Christophe Cérin [2,*], Didier Donsez [3] and Jean-Luc Baudouin [1]

1 Adeunis, Grenoble Computer Science Laboratory, Université Grenoble Alpes, 621 avenue Centrale, 38400 Saint-Martin-d'Hères, France; jl.baudouin@adeunis.com
2 LIPN/CNRS UMR 7030, University Sorbonne Paris Nord (Paris 13), 99 avenue Jean-Baptiste Clément, F-93430 Villetaneuse, France
3 LIG ERODS Team, Grenoble University, 38041 Grenoble, France; didier.donsez@imag.fr
* Correspondence: l.closson@adeunis.com (L.C.); christophe.cerin@sorbonne-paris-nord.fr (C.C.)

Abstract: This paper aims to provide discernment toward establishing a general framework, dedicated to data analysis and forecasting in smart buildings. It constitutes an industrial return of experience from an industrialist specializing in IoT supported by the academic world. With the necessary improvement of energy efficiency, discernment is paramount for facility managers to optimize daily operations and prioritize renovation work in the building sector. With the scale of buildings and the complexity of Heating, Ventilation, and Air Conditioning (HVAC) systems, the use of artificial intelligence is deemed the cheapest tool, holding the highest potential, even if it requires IoT sensors and a deluge of data to establish genuine models. However, the wide variety of buildings, users, and data hinders the development of industrial solutions, as specific studies often lack relevance to analyze other buildings, possibly with different types of data monitored. The relevance of the modeling can also disappear over time, as buildings are dynamic systems evolving with their use. In this paper, we propose to study the forecasting ability of the widely used Long Short-Term Memory (LSTM) network algorithm, which is well-designed for time series modeling, across an instrumented building. In this way, we considered the consistency of the performances for several issues as we compared to the cases with no prediction, which is lacking in the literature. The insight provided let us examine the quality of AI models and the quality of data needed in forecasting tasks. Finally, we deduced that efficient models and smart choices about data allow meaningful insight into developing time series modeling frameworks for smart buildings. For reproducibility concerns, we also provide our raw data, which came from one "real" smart building, as well as significant information regarding this building. In summary, our research aims to develop a methodology for exploring, analyzing, and modeling data from the smart buildings sector. Based on our experiment on forecasting temperature sensor measurements, we found that a bigger AI model (1) does not always imply a longer time in training and (2) can have little impact on accuracy and (3) using more features is tied to data processing order. We also observed that providing more data is irrelevant without a deep understanding of the problem physics.

Keywords: smart building datasets analyses; enabling technologies for the IoT; Building Information Modeling (BIM); knowledge generalization; data evaluation

Citation: Closson, L.; Cérin, C.; Donsez, D.; Baudouin, J.-L. Design of a Meaningful Framework for Time Series Forecasting in Smart Buildings. *Information* **2024**, *15*, 94. https://doi.org/10.3390/info15020094

Academic Editors: Pascal Bouvry, Mostapha Zbakh, Olivier Debauche and Caesar Wu

Received: 18 January 2024
Revised: 1 February 2024
Accepted: 5 February 2024
Published: 7 February 2024

1. Introduction

According to the International Energy Agency (IEA) [1], building operations accounted for 30% of global final energy consumption in 2021 and 27% of global CO_2 emissions. The stringency of building performance criteria is increasing. However, its pace is hardly matching low-carbon policies, and the energy demand is still on the rise, as well as the complexity of newly created buildings. The IEA consequently noted a 4% increase in energy demand from 2020 to 2021. Tackling this evolution to get on track with the 2050 *Net Zero*

Emissions Scenario requires ambitious and scalable policies involving all stakeholders to save on operating costs, especially on Heating, Ventilation, and Air Conditioning (HVAC).

In France, laws evolve to enforce greener building policies with the BACS and Tertiary decrees (https://www.adeunis.com/en/bacs-decree-buildings/, accessed on 1 January 2020). The former implements mandatory Building Automation and Control Systems, while the latter focuses on energy savings for their operation with a target of 60% energy savings in 2050 compared to 2020. Mandatory reports on the *Operat* official online platform track the performances and adequacy of new policies put forward by companies, but they require much information about buildings. Monitoring companies' efforts to cut energy consumption under the threat of shaming and fines highlights the scale of a nationwide effort. As 85% of European buildings are older than 2001 and 85% of current buildings will still be there in 2050, according to the 2020 European Commission report to the European parliament, facility managers need solutions adapted to already existing facilities.

Internet-of-Things (IoT) companies, like Adeunis, with its IoT battery-powered sensors, offer expert tools and flexibility adapted to the iterative monitoring of existing buildings. Many services and tools exist, with nearly 460 companies associated to create the Smart Building Alliance (https://www.smartbuildingsalliance.org/en/home/, accessed on 1 January 2020). Facility managers can then save on operating costs, ensure they meet state policies, and plan for the most-efficient renovation to prioritize.

In this context, occupancy policies hold a frequent interest, as current and future buildings are dynamic systems serving a spectrum of users with their purposes and schedules. One such is using flex-office, which consists of sharing a part of the space between users at the price of scheduling their working hours so that they do not interfere, but the place occupation rate reaches its full capacity. It synergies well with the increase in homeworking following the COVID crisis, which is currently still highly valued by job applicants and is considered a way to tackle the skyrocketing prices of square meters in the most-demanded places. Such dynamic systems are also subjected to fast evolution and often a loss of knowledge linked to occupant turnover and the progressive deterioration of equipment.

Moreover, the study of indoor and outdoor parameters aims at creating building models by the use of expert systems of black-box AI algorithms. These models may also have been made already by software like EnergyPlus, but they can be lost or inadequate after changes during the construction. It is also recommended by the International Performance Measurement and Verification Protocol (IPMVP-2022) to choose efficient building improvement strategies [2]. The modeling process is also opening to a transition from automation to Model Predictive Control (MPC) [3], leading to further accurate energy demand in building operations under the condition of adequate effort to answer big data challenges.

On the big data side, challenges are summarized by the 5V, standing for *Volume, Velocity, Variety, Veracity, and Value*, popularized by the paper of Cees de Laat, Demchenko Yuri, and Peter Membrey [4]. *Volume* refers to the staggering amount of data, with sensors monitoring indoor/outdoor environments, energy consumption, and users' behaviors, as well as the concurrent information of schedules. *Velocity* refers to the use of data for online analysis to attend to the real-time control of building operations. *Variety* encompasses the heterogeneity of data and the challenge of data fusion and analysis to understand the complex interplay between various factors affecting building performance. *Veracity* addresses the quality and reliability of data, which is necessary for the trust and accuracy of any Predictive Control system. The final V of *Value* represents the ultimate goal of data analysis in smart buildings. Organizations unlock significant values by extracting meaningful insights and predictions, driving efficiency and occupant satisfaction. Time series forecasting and Model Predictive Control optimize energy consumption, reduce costs, and create a comfortable environment.

Our work focuses on Variety and Veracity to study the scalability of smart building AI methods and the relevance of data and observations made in this field. To this end, we studied the ability of AI to forecast HVAC power consumption, CO_2, and temperature.

As a result, we hope to help people implement all of the aforementioned use cases through informed choices. We believe that the discussion, insights, and results we present can serve as a general frame to address the AI and IoT issues in the context of smart buildings.

The paper's organization is as follows. Section 2 describes the dataset used and the monitored building. In Section 3, we draw parallels between our work and related works in forecasting time series and related performance generalization. Section 4 explains our approach to defining and analyzing forecasting issues inherent to the smart building sector. There, we propose a pragmatic metric to evaluate the utility of AI over a simple lag. It emphasizes our effort to match the modeling complexity with reasonably sized AI architectures and analyses performance to define challenges raised by building properties. Section 5 provides more experiments to refine further the specific need for data in smart building forecasting. Indeed, we observe that a bigger architecture does not necessarily need more time to be trained and can have little impact on accuracy, as well as the dominance of data processing order along with feature selection. In Section 6, we specifically interpret the experimental results to provide insights regarding the datasets for future use, pointing out the importance of understanding physics problems over providing a huge amount of data. Section 7 concludes the paper. We also give an Appendix A to exemplify our implementations and the reproducibility concerns.

2. Presentation of the Dataset and the Studied Building

The monitored building [5] is a 1200 m^2 building mainly serving R&D purposes. The building, constructed on top of the ground floor, has a small first floor holding a meeting room. Around 30 employees occupy the building 75% of the working time on average, as remote working remains an ordinary policy.

The building is monitored with 28 LoRaWAN wireless sensors:

- 7 sensors for temperature and humidity;
- 13 other sensors for temperature, humidity, and CO_2 on each;
- 3 sensors for the opening and closing of 2 windows and the break room door;
- 3 sensors for the use of meeting rooms by quantifying motion detected over time and ambient light, which are not used and not shared;
- 3 sensors for the concentrations in total volatile organic components (TVOCs) and particulate matter (PM1.0, PM2.5, PM10);
- 5 sensors for 5 HVAC on the power meter.

The ground plan of the building with the sensors' positions is provided in Figure 1. Only one HVAC from the Tech zone is monitored, the one far from the windows.

The experiments in this paper did not use all the sensors, as side experiments made us change several sensors. The occupancy sensors were inconvenient to exploit. The two sensors put aside have a red cross on them. Only sensors working concurrently during most of the experiment were used to prevent huge gaps in data collection. Like other indoor environment data, the temperature data lasted 232 days, from +2022 October 24 13:20:00+ to +2023 June 14 07:30:00+. Consumption data started later, as they needed specific skills and raised compatibility problems on the electric panel, lasting 90 days from +2023 March 15 16:20:00+ to +2023 June 14 07:40:00+. Multivariate data were based on the intersection of datasets.

To cope with the quantity of data and the complexity of irregular building behavior, for hourly predictions and the base observation of 1-hour sequences, the following data were used for training and inference:

- 15 weeks of data to train in temperature forecasting;
- 6 weeks to evaluate temperature forecasting performances;
- 8 weeks of data to train in HVAC energy consumption forecasting;
- 6 weeks to evaluate HVAC energy consumption forecasting performances.

All sensors have a time step of 10 min, except weather data, which have a time step of 1 h. Weather data, from outdoors, encompasses outdoor temperature, pressure, humidity, and cloudiness. The choice of 10 min has been made considering the use of battery sensors and, thus, the cost of high-frequency monitoring at the scale of a building. It was a reasonable choice, considering the thermal inertia of the building and the scheduling of workers, with continuous workloads and meetings ranging from 1 to 4 h. A subscription to OpenWeatherMap permits access to the weather dataset, but it cannot be shared for licensing reasons.

Despite efforts to limit gaps in data collection, some remain because of network failures at the scale of a sensor or the whole network. However, AI models, as presented in this paper, need regular data with a constant step. We chose to keep 10 min and filled small holes by linear interpolation, which allowed us to keep data at the price of reasonable interpolation error. The gaps were filled up to 40 min by the left for datasets with 10 min.

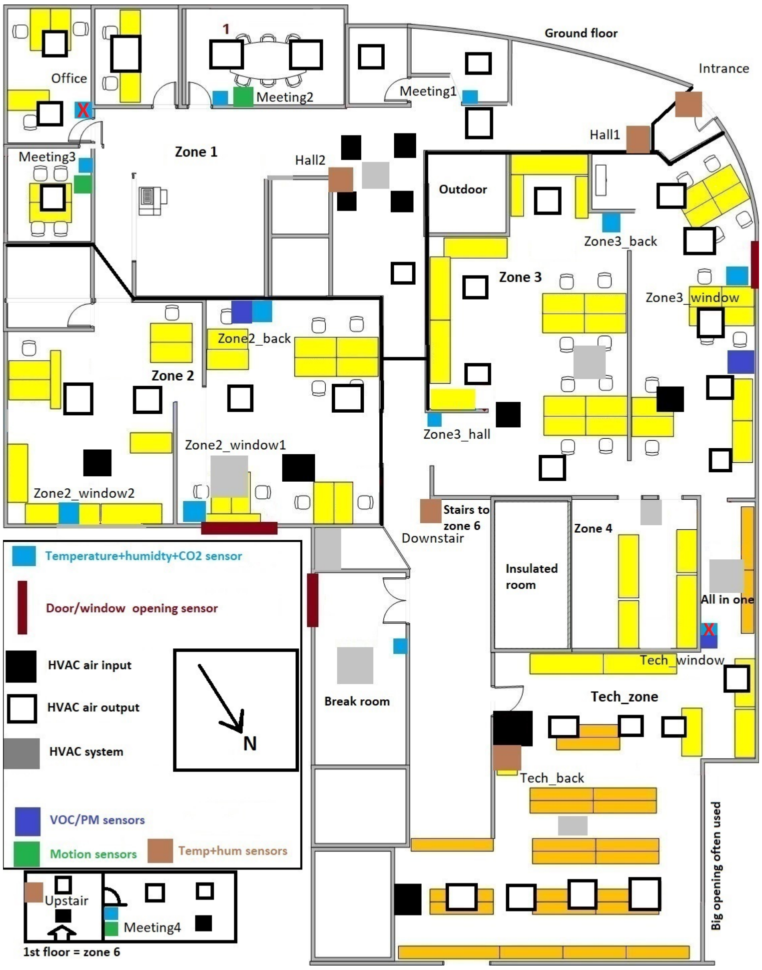


Figure 1. Map of the instrumented building.

3. Related Work

Time series forecasting has been a hot topic for decades for a wide variety of subjects like stock exchanges [6], traffic jams [7], or smart building energy consumption [2], respectively aiming at more profitability, decreasing traffic jams, and decreasing operating cost while increasing indoor comfort.

Traditional linear algorithms, such as the Auto-Regressive Integrated Moving Average (ARIMA) and exponential smoothing models, are still used for time series forecasting, but often cannot capture complex nonlinear patterns that deep learning algorithms can analyze. These linear algorithms prove reliable in some cases [8], where data aggregated over a large territory appear to follow much more-regular patterns. For a wider diversity of data, using deep learning to tackle high-dimensional correlated time series is a significant trend emphasized by the resulting paper of the data forecasting competition M4 [9], where winners efficiently combined deep learning techniques and preprocessing steps into hybrid algorithms.

The preprocessing steps have substantial impacts on forecasting results [10], but the performances are attributed to adaptable deep learning algorithms like Long the Short-Term Memory (LSTM) network, which outperform linear algorithms [9,11]. Another path to improve AI algorithms lies in combining different architectures, with relevant examples applied to energy forecasting like in [12], where the authors combine LSTM with a Convolutional Neural Network (CNN) to improve the LSTM performances. A latter article co-written by the winner of the M4 competition also focuses on the LSTM part to further improve its results [13].

The paper [11] made impressive use of exponential smoothing to identify data seasonality, where others try to increase the dimension of the space using Empirical Mode Decomposition (EMD) [10]. The issue of seasonality remains in the case of smart buildings, but it is secondary. This field intersects with the indoor Internet of Things (IoT) for smart buildings, which is reliable and provides timestamps with each communication. Furthermore, calendars are easy to establish to monitor building occupants' behaviors and provide enough temporal information, which may be missing in other domains [14]. Since contextual encoding of hours, days, weeks, weekends, and holidays is adequate for work schedules, seasonality detection is a secondary issue in smart buildings. EMD can be used along with seasonality detection to reduce noise in data like in the article [15], which discusses its abilities and makes comparisons to ensemble methods and wavelet transformation techniques. It also raises issues about the strong constraints to efficiently apply EMD to data, further developed in the article [16], which explains the highly data-driven performances of EMD and the Dynamic Wavelet Transform (DWT). This is also supported by [17], which specifically compares LSTM Neural Networks, ARIMA, and an EMD hybrid method for energy load forecasting at different scales, showing better results at the scale of the electric network than at the scale of a building.

On the same basis of the CNN-LSTM architecture of [12], researchers also focus on improving AI generalizability. The article [18] put forward a Bayesian model for indoor air quality forecasting, selecting data based on their respective probabilistic contribution to the forecast quality, while [19] focuses on the optimal search of hyperparameters with genetic algorithms to forecast energy loads. The authors of [20] also consider Bayesian models fusion, to estimate indoor temperature in rooms without sensors, while [21] uses XGBoost AI models with the same purpose. These two articles explore a new perspective on building data exploitation. In the specific case of smart building data, Ref. [22] exerted good accuracy in forecasting energy consumption with an attention layer, designed to select the most-relevant features and temporal data among a very big dataset. LSTM enhanced by an attention layer was also used in [23] to study HVAC Predictive Control, using a real highly monitored building. The purpose of attention layers is then to tackle problems arising from the huge amount of data to target the most-relevant information.

Time series forecasting genuinely enables Model Predictive Control (MPC) for smart building control, which focuses much more on decision-making and considers anomalous or irregular events as the next step after building automation. The first challenge lies in forecasting indoor parameters such as temperature [24], CO_2 [25], and energy consumption [26] and establishing a compromise regarding occupant perception [27].

However, few works have widened their study to an open-loop HVAC optimization, considering the scale of the building and the variety of existing HVAC systems. Ref. [28] uses the EnergyPlus expert simulation software for HVAC control simulation, while [29] could access real HVAC control. The complexity of infrastructures for MPC, with humans in the loop [30], is concurrently becoming a main issue, as emphasized by [31,32]. The final step toward an industrial-scale AI solution will undoubtedly be to address this complexity in data by enabling the MPC to select relevant data and detect anomalies [33,34]. Such solutions at industrial scales must also be prepared for scarce data since building with as many sensors as in previous studies is uncommon, which is underlined by future directions quoted in the review [35].

Our paper aims to show a realistic use of AI for data forecasting in smart buildings, embracing a meaningful diversity of performances as expected in a real scenario. The building presented here, with its amount of sensors and data granularity, is the result of an industrial project aiming at the most-common case. Consequently, the fine-tuning of complementary methods to LSTM was out of scope and may not cover such realistic cases. We also show concern for reproducibility, to give insights into enabling Machine Learning Operations (MLOps) for the field of smart buildings.

4. Forecasting Neural Network Evaluation

4.1. Pragmatic Metric to Evaluate AI Performances

The literature in the field of time series forecasting outlines the performances of Recurrent Neural Networks (RNNs) over linear models, with the focus switched to the size of architectures and the quality and quantity of data. Thus, the authors decided to provide insight into the LSTM applied to the forecasting task. Indeed, compared to common architectures like random forest, ARIMA, and the Multi-Layer Perceptron (MLP), LSTM was more accurate for most hyper-parameters and consistent in its superiority. Considering the work schedules of occupants in the building, the study was narrowed to hourly forecasting only.

We also propose a new evaluation metric of the forecasting algorithm performances that allows us to compare to a no-forecast case. Indeed, we considered a no-forecast as a mere delay, where the best estimation for future data is the feature's current value. The Root-Mean-Squared Error (RMSE) expresses the error, with a much higher penalty for the biggest errors in forecasting. However, it is hard to directly link the error of prediction to the potential of AI in forecasting tasks.

Thus, we define performances as the ratio of reduction of the Root-Mean-Squared Error (*RMSE*) of prediction when compared to a delay:

$$score = \frac{RMSE_{delay} - RMSE}{RMSE_{delay}}$$

The upper bound is 1, and a negative value means the algorithm is not worth a simple delay and should not be chosen to reduce the *RMSE*. This metric is intuitive, cross-domain, and questions the very use of algorithms in forecasting issues. We expect it to be used to move forward with a reasonable use of AI for industrial issues with a high potential for improvement.

4.2. Forecasting Neural Network Complexity

LSTM is a recurrent Neural Network composed of one repeating cell, trained with the gradient descent algorithm of *back-propagation through time*. The training process links data sequences to future targeted data after several iterations of the LSTM cell, thus literally learning how to process data through time, as shown in Figure 2a. The LSTM cell includes four gates, which are Multi-Layer Perceptrons (MLPs) with different activation functions in yellow in Figure 2b. The dimension of the LSTM Neural Network refers to the size of the MLP that composes the cell, which is also the size of the latent space. The figure also emphasizes the ability of LSTM to exert long-term memory thanks to the upper path, while the short-term memory and output are transmitted through the lower path. Illustrations are given on Colah's blog, where an in-depth explanation is available (https://colah.github.io/posts/2015-08-Understanding-LSTMs/, accessed on 1 January 2020).

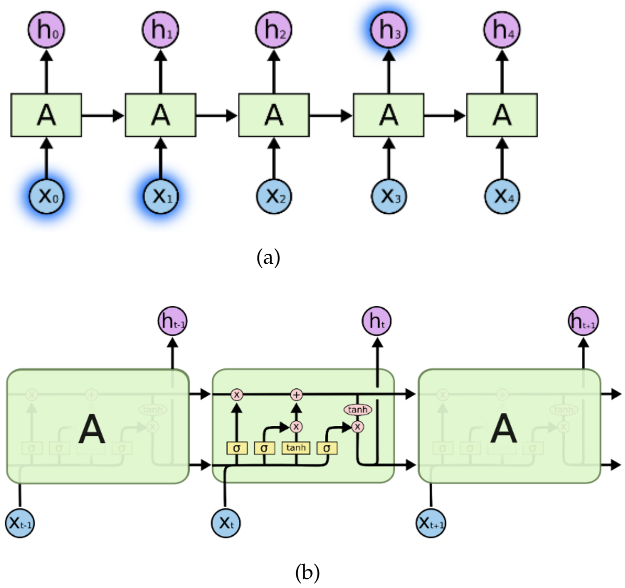
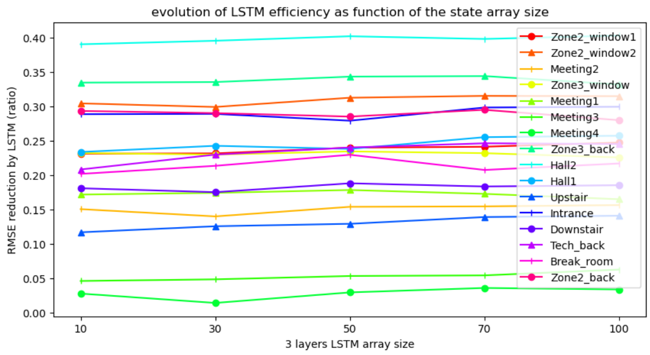


Figure 2. LSTM Neural Network principle and architecture. (**a**) LSTM principle; (**b**) LSTM architecture.

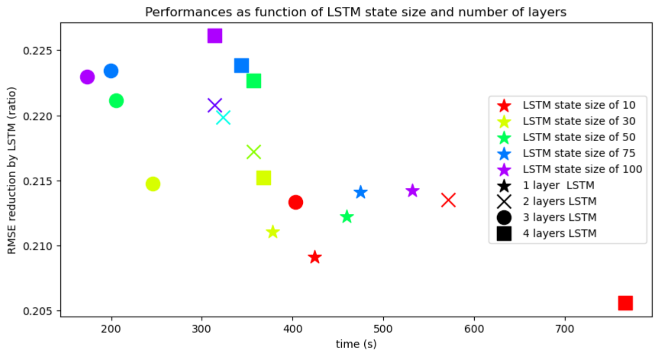
We considered concurrently the number of LSTM stacked layers and the dimension of the arrays processed. Without other information, we used an LSTM network with three layers and a state space of size 100, over 1 h of temperature data (so seven samples including the present value) to forecast 1 h (so six steps ahead). The day of the week and the hour of the day on weekends or holidays were one-hot-encoded. They constitute an array of dimension 33 with '0' and '1', concatenated with the output array of the LSTM. The final array is processed by a Multi-Layer Perceptron (MLP) of 10 neurons, then a neuron without an activation function. The MLP, thus, provides a latent space resulting from a non-linear combination of LSTM monovariate output and calendar data. The final forecast is the result of a linear combination of the 10 dimensions of this latent space by the Perceptron. We conducted multiple tests, which are not shown here, and highlight the drop in accuracy when using other calendar encodings. We expected it to result from time discontinuities in numerical encoding.

Figure 3a depicts the performances of the three layers of the LSTM with several state space sizes for each temperature sensor. Experiments on forecasting HVAC consumption used a training set of 8 weeks and a test set of 3 weeks, increased, respectively, to 15 and 6 weeks when the consumption feature is not used. It appears here that the dimensions of the latent space have little impact on the accuracy of the Neural Network for temperature forecasting. However, it accelerates the mean training speed, as seen in Figure 3b over our 16 temperature sensors. Thus, the selection of three layers and a state space of size 100 also holds the potential to handle more-complex data with more features used during the forecasting.

Since the number of epochs during the training varies a lot for each experiment, with an impacting stochastic part, we introduced a stopping condition: if no better prediction appears in the next 50 epochs of training, the best weights found are returned. A hard cap was put at 2000 epochs to return the best result, which was rarely reached. The choice of a 50 epochs was made after observing that plateaus lasting more than 50 epochs, which seldom existed in the training curves.



(a)



(b)

Figure 3. Overall performances of LSTM for temperature forecasting on all sensors. (**a**) Performance as function of architecture used; (**b**) mean performances and training duration for several LSTM architectures.

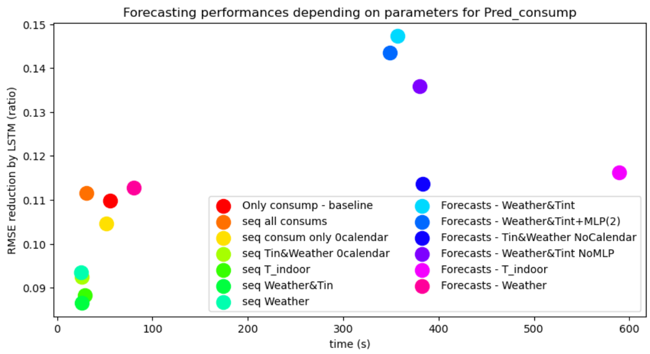
4.3. Forecasting Policies

Calendar data could be used as multivariate input sequences in the LSTM network or concatenated with the output and analyzed by a Multi-Layer Perceptron (MLP), like any other feature. Calendar data provide information on the moment to predict, like the day of the week, the hour of the day, and the 10-min slice of the hour. It also provides information on holidays, but one could eventually expect more case-specific details, like events or occupancy in the building. We chose to directly use the date of the forecasted data instead of a past sequence because this information is known in our case. The same choice arose for multi-feature forecasts. On the one hand, we used common multivariate input arrays. On the other hand, each monovariate feature was processed beforehand by a different LSTM network of the same size alone:

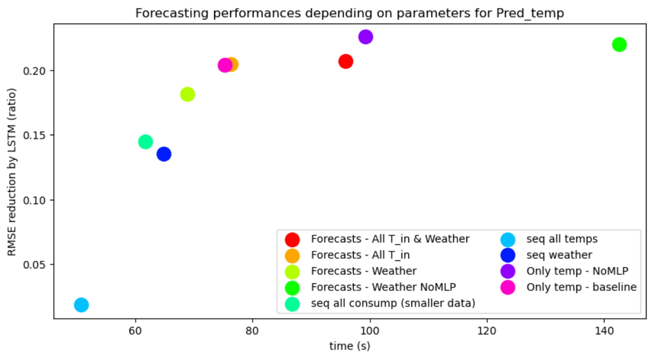
- Single feature monovariate time series processed by a single LSTM with the output concatenated with the calendar data;
- Multivariate time series processed by a single LSTM with the output concatenated with the calendar data;
- Monovariate time series processed by several LSTMs, with all the output concatenated with the calendar data.

A single layer of 10 neurons processes all of them, and then, a linear neuron provides the monovariate output.

The comparison is provided in Figure 4. Each point is the mean value of an experiment's efficiency over HVAC consumptions for Figure 4a and temperature sensors for Figure 4b. The keyword `seq` stands for the use of multivariate sequences, while the keyword `forecasts` stands for independent forecasts before analysis by the MLP. The keyword `NoMLP` describes the use of the LSTM and Perceptron only, without MLP, and `MLP(2)` stands for the use of a two-layer MLP instead of a single one. Other labels refer to the different features used.



(a)



(b)

Figure 4. Performances as a function of selected forecasting policy and features. (**a**) HVAC consumption prediction performances; (**b**) indoor temperature prediction performances.

In Figure 4a, the main increases seem strongly related to using several monovariate time series compared to a single multivariate sequence. Such improvement is not as relevant for temperature forecasting in Figure 4a, but a high reduction in training time is noticeable, at the price of a little drop in accuracy while using independent forecasts.

From the shape of the graphs, accuracy and training time seem generally correlated. This observation is the direct consequence of the exploratory power of the algorithm during training, which is linked to reachable performances. With multivariate sequences and MLP for temperature forecasting, the rise in complexity and the decrease in accuracy could be the reason for the faster training.

Doing so, we underline the ambiguity of training duration in time series forecasts, as using a pure delay would be incredibly faster. According to the accuracy metric we utilized, the comparison to a delay must enable comparative studies based on usefulness concerning the cost of AI.

5. Data Relevance

Retrieving more features and data may improve forecasting performances and increase the cost of devices and data engineering. Any prior insight may save a lot of time, and this is investigated next.

5.1. Feature Selection

The variety of experiments shown in Figure 4 allows us to select architectures and features of interest. For HVAC consumption in Figure 4a, the best accuracy resulted from the simultaneous forecast of weather, indoor temperature, and HVAC energy consumption, concatenated with a one-hot encoded calendar, processed by a Multi-Layer Perceptron.

The absence of a calendar is harmful, and predictions cannot reach higher accuracy without it. As stated above, this feature's complexity seems too high for the LSTM to handle multivariate sequences directly.

The physical features that we decided to monitor are intuitively those related to a close physical system, with the HVAC being a function of outdoor weather and temperature and indoor temperatures. Doors are not open often enough to have a 'real' impact on this system.

Thus, forecasting ability encompasses the measurement of the building's physical properties. A pitfall for such a goal is the lack of information about temperature setpoints, which can be changed at any time by building occupants and directly impact the link between HVAC consumption and indoor temperature. It is also relevant as the indoor temperature and HVAC have significant inertia and take time to reach the chosen temperature target. This inertia makes it very difficult to find setpoint changes by discontinuities in the data.

5.2. History Length for Hourly Forecasting

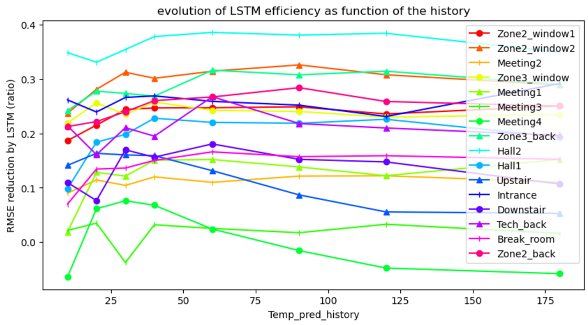
AI is often considered a robust tool for any complexity. However, while processing older data for time series forecasting, a loss of correlation may be inducted, which decreases the performance. This also applies to dubious physically correlated data, like the weather of a different place or wind speed near the building. We wished to give an insight into history relevance in a forecasting framework by deep learning.

In this paper, we considered the time correlation within the history parameter of the LSTM, which is the duration covered to forecast each value. The length of history always refers to the duration of the sequence, which ends 1 h before the predicted value.

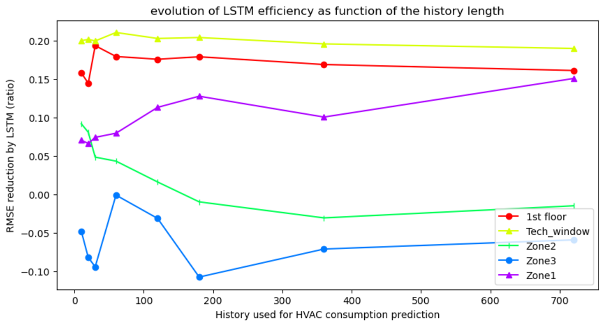
With a longer history, only sequence–target couples with continuous data are kept despite remaining holes, so a longer history provides fewer couples. The high quality of HVAC consumption data allowed us to have no data cut until 720 min. However, the temperature dataset shrank in the training and test sets at 30 min. For experiment consistency on temperature history, we selected 11 weeks for the training and 4 weeks for the test. The moment of evaluated sequences may change, but the amount of data remained the same for each history tried. Furthermore, no 'easy-to-observe' anomalous behavior or events occurred in the building.

Figure 5 shows how the length of the sequences used impacts the performances of the forecasting for each sensor, as we still compared it to a simple delay. Figure 6 presents the average performances over sensors for each history length. However, Figure 5 shows that each forecasting performance evolved differently depending on the sensors, so the standard deviation is also presented. As this metric decreases or stays constant, this shows that the average performances are not led by a single value, but tend to represent most of the sensors. We observed a rapid increase before a maximum of 1 h of data before decreasing, for both consumption history in Figure 6a and temperature history in Figure 6b. The standard deviation of performances was high over the sensors, so we show it on the same scale as the average performances.

We emphasize here, and according to our dataset, that the relevance of the data is paramount rather than the sheer quantity. Here, a long history brings increased complexity to the regression, but older samples do not carry interesting information to balance the complexity.

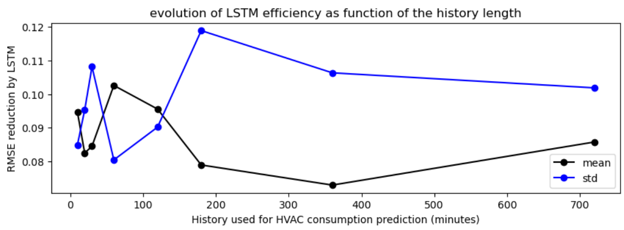


(a)

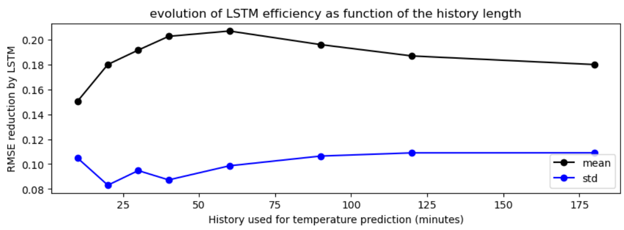


(b)

Figure 5. Study of performances as a function of history length in minutes. (**a**) Temperature forecasting performances; (**b**) HVAC consumption forecasting performance.



(a)



(b)

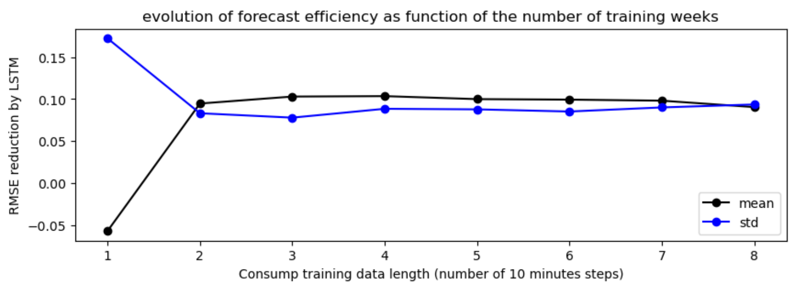
Figure 6. Study of performances as a function of history length. (**a**) Consumption forecasting performances; (**b**) temperature forecasting performances.

5.3. Training Data Length Study

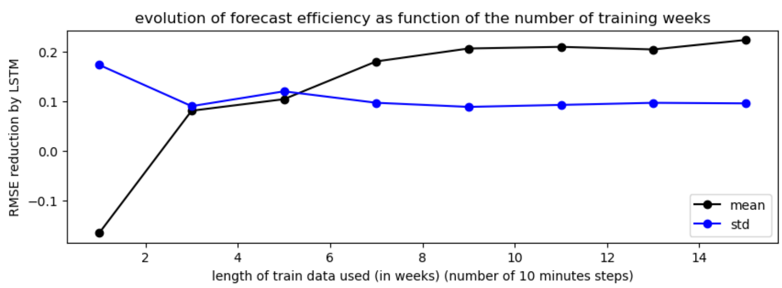
This section focuses on the amount of data necessary to train the LSTM to forecast and its impact on the consistency of the forecasting ability. With a fixed test set, we changed the number of weeks used to train the algorithm. The results are presented in Figure 7. For mean HVAC consumption forecasts, rapid growth leads to a plateau reached in 2 weeks, with barely any increase in further weeks, as seen in Figure 7a. Mean temperature forecasts increase until 9 weeks, with a fast increase, which slows down as more data are added for the training set, as seen in Figure 7b. The standard deviation is also presented. As it decreases or stays constant, it shows that aggregates are not led by a single value, but tend to represent most of the sensors, validating an average trend.

It is unclear how this behavior should be attributed to the nature of data or the presence of anomalies in the training or test data. Unknown anomalies might be expected for simultaneous employees' days off or exceptionally warm or cold days. In the event of exceptional behaviors with no label, we might attribute them to the training data or testing data. In the first case, we expect plateaus in the performance to increase, lasting as much as irrelevant data. In the second case, we expect a hard cap on performance increases with hardly any increase while using further weeks. Notice that both cases may coexist and can be different between sensors.

We consequently highlight the need for high-quality relevant test data, showing the diversity of the common use of the building. We address this topic against the over-representation of the training data, which may not be possible for an on site industrial solution with small datasets. Establishing a monitoring campaign for smart building forecasting solutions would require a lengthy test dataset, depending on the diversity and exceptional events that cannot be quantified regarding building behavior irregularity.



(a)



(b)

Figure 7. Performances of forecasting as a function of training data length. (**a**) Performances of HVAC consumption forecasting; (**b**) performances of temperature forecasting.

6. Findings

6.1. Synthesis on the Methodology

We considered the complexity of the forecasting AI algorithm, resulting from the LSTM memory network and the data-processing policy. We found that higher complexity in the architecture could induce faster training with little impact on accuracy, as highlighted in Figure 3. On average, the reason for the training acceleration could be fewer epochs during the iterative process. LSTM networks are known to be robust to exploding and vanishing gradients by their architecture, providing a stable learning curve, with stability increasing with the size of the LSTM network. Training AI algorithms is not linear, and the training duration should be studied with the training graph and not as a metric.

We furthermore demonstrated the the weakness of LSTM in processing multivariate correlated time series, showing its limits in adaptability. We chose to use an independent forecast combination rather than a simultaneous processing. The former is labeled

`Forecasts` in Figure 4. Our insight lies in the ability to separate temporal correlation from feature correlation. By executing both tasks consecutively, the policy we chose would simplify the forecasting issue. Another explanation could be linked to causality, as the observation of some features could hint at behaviors without being linked by a direct causality relation. Before the analysis by another Neural Network, primary independent forecasts could limit causality induction between features by the LSTM.

6.2. Synthesis on Data Relevance

We evaluated the relevance of the data used to forecast HVAC consumption and temperature by selecting the features, the length of sequences involved in each hourly forecast, and the quantity of training data.

The analysis of the history parameters also showed that 1 h sequence inputs enabled the best accuracy. This observation could be directly linked to the work schedule of occupants, which uses hourly steps. The choice of history results from a trade-off between bringing more data and increasing the complexity of AI processing. Older data may not bring new information and only increase the size of sequences and the number of parameters of AI models, thus decreasing performances. Likewise, data older than 1 h can testify to different meetings, breaks, or technical activities and lack relevance to complete the latest data.

The training data length encompasses a different trade-off between data diversity and common data behavior. We want AI to witness the opposite realistic behavior, and much of the data often offers this opportunity, but dilutes anomalies since anomalies are scarce. In this study, the performances reached their best with two weeks of data for energy consumption forecasting, but kept growing with the data for temperature forecasting. As mentioned in the analysis of the features, the reason might be the scheduling of some HVAC, while the temperature is irregular. We still see a logarithmic shape that allowed us to expect a performance increase with more training data for temperature forecasts. Such an experiment would allow a facility manager to decide when he/she starts operating his/her building or whether he/she needs a few more weeks.

7. Conclusions and Future Works

Experiments with the massive deployment of smart sensors are becoming increasingly numerous, and the technology to utilize them is lagging. As AI holds the promise of automatic optimization, its complexity and uncertainty make it hard to use in an industrial framework. This paper is based on industrial data from one smart building following a real industrial monitoring campaign. The summary of our industrial return of experience is as follows.

Our work proposes experiments on AI architectures and data usefulness to improve HVAC consumption forecasting and indoor temperature forecasting. The issues to address were related to data selection by the size and physical impact on the forecasts. We tackled the issue of AI reliability by focusing on the data used by our algorithm, providing all necessary information about them and sharing our dataset [5] to promote reproducibility.

We have proposed different analyses and observations on 'real' data to illustrate an industrial forecasting framework. To do so, we provided a cross-domain metric to question the use of AI and a real case evaluation, allowing us to cover several aspects of indoor environment modeling. It consequently highlighted a few possibilities for Neural Network architectures and data processing.

In short, the study conducted in this paper provides insight into the design of several aspects of a forecasting framework for smart buildings, as we gave the insight to focus on meaningful architectures, features, and key policies to optimize predictions. We hope to push toward industrial solutions for smart building Predictive Control as we develop trust and generalizability.

Future work could investigate the issues mentioned in this paper, focusing on data and AI uncertainty. Detecting regime changes and discontinuities in data holds a major interest in improving time series processing. A first approach would be subsequences' clustering, where typical sequences are identified to map the behavior of a building. This method includes sequence recognition and expert knowledge to raise trust in the chosen windows. Based on the subsequences found, this approach could be broadened to spatial behavior. The spatial approach alone has been tried and dropped as the behavior varieties over the time series impede the clustering task. It would enable scenario-specific algorithms and wiser choices about the variety of algorithms needed, instead of oversized architectures, which might be ill-designed for such problems.

Training graphs are important for AI engineering to simultaneously monitor the training of the algorithm and its ability to process new data. In this paper, we dropped the issue of generalizability over further datasets, but it appears in 'real' operating systems. They could be worth investigating to enable in-training data selection by the shape of the training graphs and the comparison between the training, validation, and testing steps. Such a study will aim at more-relevant and faster results by deep AI, as the algorithm's generalization would have been strictly monitored during the training step.

Quantifying forecast uncertainty is paramount for industrial solutions. A promising field of research is the use of Bayesian Neural Networks, which allow quantified algorithm flexibility, thus providing a range for prediction uncertainty. The comparison of datasets with the analysis of training graphs could also provide expectations about uncertainty before any Neural Network. This would save time and constitute a step further toward reasonable AI, as we could prioritize projects with higher potential.

Finally, despite the cost and complexity of building management systems, temperature setpoints and their changes must be monitored to enable control issues, especially Predictive Control.

8. Legal Disclaimer

Author Contributions: Conceptualization, L.C.; Methodology, L.C.; Software, L.C.; Validation, C.C. and D.D.; Resources, D.D. and J.-L.B.; Data curation, C.C. and D.D.; Writing – original draft, L.C.; Supervision, C.C. and J.-L.B.; Project administration, J.-L.B.; Funding acquisition, D.D. and J.-L.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been co-funded by a CIFRE grant (reference 2021/1336) and partially supported by the Multi-disciplinary Institute on Artificial Intelligence MIAI at Grenoble Alpes (ANR-19-P3IA-0003). This work was conducted during the Délégation with Centre National de la Recherche Scientifique (CNRS) of Mr Cérin. Thanks to the institutional support of the CNRS, University of Grenoble Alpes, and University Sorbonne Paris Nord.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available at the aforementioned DOI: https://doi.org/10.18709/perscido.2023.08.ds398, accessed on 1 February 2024. A Jupyter notebook is also provided along with the data.

Acknowledgments: Thanks to the institutional support of the CNRS, University of Grenoble Alpes, and University Sorbonne Paris Nord.

Conflicts of Interest: The Authors declare no conflicts of interest.

Appendix A. How to Play with Our Notebook?

To respect the principles of open science, i.e., to produce clear, shared hypotheses, methods, and protocols subject to critical analysis and discussion aimed at their improvement and accessible to as many people as possible, we are making our dataset and the

Jupyter Notebook public. The anonymized data and the Jupyter notebook necessary to build the models and reproduce the results (https://github.com/CampusIoT/datasets/tree/main/BuildPred/notebooks, accessed on 1 January 2020) are published under licenses ODbL-1.0 and GPLv3 in a public Git repository and on the Persido open-data platforms.

In the recent past, independent verification of published research results has not been systematic, nor even typical in computer science, despite innovation in the field relying heavily on software, hardware, and data experimentation. More recently, public authorities have given new impetus to this situation by requiring data management plans to accompany grant applications for sharing experimental results and recognizing that access to the experimental artifacts underlying reported results is the best way to ensure scientific integrity and advance the field.

Consider now the configuration step of the relevant variables of the LSTM as an example. This part of the code can be found in entry [17] in the notebook.

The parameter `depth` corresponds to the history in this paper. It describes the length of sequences, in minutes, used as the input for each value predicted. It was designed to be a multiple of 10, which is the time step of the data.

The parameter `horizon` corresponds to the horizon of prediction, which is the number of minutes between the last sample of the sequence and the value to forecast. It should also be a multiple of 10.

The parameter `nb_samples_train` allows a cutting for data, delimiting the maximum length of the training data in the number of samples. According to a mini-batch learning policy, the training and datasets will be separated into complete batches of size `batchsize`, and incomplete batches are discarded.

The parameter `LSTM_size` refers to the side of each LSTM layer's latent space. By default, they all have the same size. Further modification to the architecture, like the number of layers, can only be performed in entry [6], which defines choices in architecture.

The parameter `param_range_array` is used in a for-loop to cover several configurations during a unique run of the notebook. This one and `param_name` appear in the JSON data, saving experimental results after the trained algorithm's inference on the test data. A JSON file is saved under the name `save_name`.

The unused parameter `nb_segments` allows a k-fold cut in the data portioning between the training and test data. The loop is blocked to the first iteration in the main code, and plotting functions may not be adapted yet to this other complexity in JSON files.

The parameter `restrict_day` allows limiting all data processing to one day of the week, selected by its number between 1 and 7. However, it requires shortening the `batchsize` parameter to 1 day or less, and missing data greatly impact the results' significance. This might not be valuable without more data to dilute anomalous missing values. Time series graphs do not show the discontinuities in the data, and the partitioning details are not provided in the JSON file, which implies cautious use of such parameters.

Finally, users might pay close attention to the part about *Dataset Loading* and *Data Selection*, which specify how data are used to process selected features of interest. It also gives information about the dataset length as it provides an overview of the main dataset to process.

References

1. IEA. Buildings, IEA, Paris. 2022. Available online: https://www.iea.org/reports/buildings (accessed on 1 January 2020).
2. Dos Santos Ferreira, G.; Martins Dos Santos, D.; Luciano Avila, S.; Viana Luiz Albani, V.; Cardoso Orsi, G.; Cesar Cordeiro Vieira, P.; Nilson Rodrigues, R. Short- and long-term forecasting for building energy consumption considering IPMVP recommendations, WEO and COP27 scenarios. *Appl. Energy* **2023**, *339*, 120980. [CrossRef]
3. Schwenzer, M.; Ay, M.; Bergs, T.; Abel, D. Review on Model Predictive Control: An engineering perspective. *Int. J. Adv. Manuf. Technol.* **2021**, *117*, 1327–1349. [CrossRef]
4. Demchenko, Y.; De Laat, C.; Membrey, P. Defining Architecture Components of the Big Data Ecosystem. In Proceedings of the 2014 International Conference on Collaboration Technologies and Systems (CTS), Minneapolis, MN, USA, 19–23 May 2014; Volume 10, p. 1168. [CrossRef]

5. Closson, L.; Donsez, D.; Baudouin, J.L.; Trystram, D.; Cérin, C. A Smart Tertiary Building Dataset. 2023. Available online: https://github.com/CampusIoT/datasets/blob/main/BuildPred/README.md (accessed on 1 January 2020).
6. Kobiela, D.; Krefta, D.; Król, W.; Weichbroth, P. ARIMA vs LSTM on NASDAQ stock exchange data. *Procedia Comput. Sci.* **2022**, *207*, 3836–3845. [CrossRef]
7. Abbas, Z.; Ivarsson, J.R.; Al-Shishtawy, A.; Vlassov, V. Scaling Deep Learning Models for Large Spatial Time-Series Forecasting. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 1587–1594. [CrossRef]
8. Chaturvedi, S.; Rajasekar, E.; Natarajan, S.; McCullen, N. A comparative assessment of SARIMA, LSTM RNN and Fb Prophet models to forecast total and peak monthly energy demand for India. *Energy Policy* **2022**, *168*, 113097. [CrossRef]
9. Makridakis, S.; Spiliotis, E.; Assimakopoulos, V. The M4 Competition: Results, findings, conclusion and way forward. *Int. J. Forecast.* **2018**, *34*, 802–808. [CrossRef]
10. Zheng, H.; Yuan, J.; Chen, L. Short-Term Load Forecasting Using EMD-LSTM Neural Networks with a Xgboost Algorithm for Feature Importance Evaluation. *Energies* **2017**, *10*, 1168. [CrossRef]
11. Smyl, S. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *Int. J. Forecast.* **2020**, *36*, 75–85. [CrossRef]
12. Yan, K.; Wang, X.; Du, Y.; Jin, N.; Huang, H.; Zhou, H. Multi-Step Short-Term Power Consumption Forecasting with a Hybrid Deep Learning Strategy. *Energies* **2018**, *11*, 3089. [CrossRef]
13. Dudek, G.; Pełka, P.; Smyl, S. A Hybrid Residual Dilated LSTM and Exponential Smoothing Model for Midterm Electric Load Forecasting. *Conf. Name IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 2879–2891. [CrossRef]
14. Nguyen, T.A.; Aiello, M. Energy intelligent buildings based on user activity: A survey. *Energy Build.* **2013**, *56*, 244–257. [CrossRef]
15. Li, Y.; Zhu, N.; Hou, Y. Comparison of empirical modal decomposition class techniques applied in noise cancellation for building heating consumption prediction based on time-frequency analysis. *Energy Build.* **2023**, *284*, 112853. [CrossRef]
16. Zhang, L.; Alahmad, M.; Wen, J. Comparison of time-frequency-analysis techniques applied in building energy data noise cancellation for building load forecasting: A real-building case study. *Energy Build.* **2021**, *231*, 110592. [CrossRef]
17. Wood, M.; Ogliari, E.; Nespoli, A.; Simpkins, T.; Leva, S. Day Ahead Electric Load Forecast: A Comprehensive LSTM-EMD Methodology and Several Diverse Case Studies. *Forecasting* **2023**, *5*, 297–314. [CrossRef]
18. Jin, X.B.; Wang, Z.Y.; Gong, W.T.; Kong, J.L.; Bai, Y.T.; Su, T.L.; Ma, H.J.; Chakrabarti, P. Variational Bayesian Network with Information Interpretability Filtering for Air Quality Forecasting. *Mathematics* **2023**, *11*, 837. [CrossRef]
19. Bouktif, S.; Fiaz, A.; Ouni, A.; Serhani, M.A. Optimal Deep Learning LSTM Model for Electric Load Forecasting using Feature Selection and Genetic Algorithm: Comparison with Machine Learning Approaches. *Energies* **2018**, *11*, 1636. [CrossRef]
20. Chen, X.; Li, X. Virtual temperature measurement for smart buildings via Bayesian model fusion. In Proceedings of the 2016 IEEE International Symposium on Circuits and Systems (ISCAS), IEEE, Montreal, QC, Canada, 22–25 May 2016; pp. 950–953. [CrossRef]
21. Kaligambe, A.; Fujita, G.; Keisuke, T. Estimation of Unmeasured Room Temperature, Relative Humidity, and CO2 Concentrations for a Smart Building Using Machine Learning and Exploratory Data Analysis. *Energies* **2022**, *15*, 4213. [CrossRef]
22. Li, Y.; Tong, Z.; Tong, S.; Westerdahl, D. A data-driven interval forecasting model for building energy prediction using attention-based LSTM and fuzzy information granulation. *Sustain. Cities Soc.* **2022**, *76*, 103481. [CrossRef]
23. Zhuang, D.; Gan, V.J.L.; Duygu Tekler, Z.; Chong, A.; Tian, S.; Shi, X. Data-driven Predictive Control for smart HVAC system in IoT-integrated buildings with time-series forecasting and reinforcement learning. *Appl. Energy* **2023**, *338*, 120936. [CrossRef]
24. Alawadi, S.; Mera, D.; Fernández-Delgado, M.; Alkhabbas, F.; Olsson, C.M.; Davidsson, P. A comparison of machine learning algorithms for forecasting indoor temperature in smart buildings. *Energy Syst.* **2020**, *13*, 689–705. [CrossRef]
25. Wei, W.; Ramalho, O.; Malingre, L.; Sivanantham, S.; Little, J.C.; Mandin, C. Machine learning and statistical models for predicting indoor air quality. *Indoor Air* **2019**, *29*, 704–726. [CrossRef] [PubMed]
26. Le, T.; Vo, M.T.; Kieu, T.; Hwang, E.; Rho, S.; Baik, S.W. Multiple Electric Energy Consumption Forecasting Using a Cluster-Based Strategy for Transfer Learning in Smart Building. *Sensors* **2020**, *20*, 2668. [CrossRef]
27. Ferreira, P.M.; Silva, S.M.; Ruano, A.E. Energy savings in HVAC systems using discrete model-based Predictive Control. In Proceedings of the 2012 International Joint Conference on Neural Networks (IJCNN), Brisbane, QLD, Australia, 10–15 June 2012; pp. 1–8. [CrossRef]
28. Coraci, D.; Brandi, S.; Hong, T.; Capozzoli, A. Online transfer learning strategy for enhancing the scalability and deployment of deep reinforcement learning control in smart buildings. *Appl. Energy* **2023**, *333*, 120598. [CrossRef]
29. Afram, A.; Janabi-Sharifi, F.; Fung, A.S.; Raahemifar, K. Artificial Neural Network (ANN) based Model Predictive Control (MPC) and optimization of HVAC systems: A state of the art review and case study of a residential HVAC system. *Energy Build.* **2017**, *141*, 96–113. [CrossRef]
30. Aguilar, J.; Ardila, D.; Avendaño, A.; Macias, F.; White, C.; Gomez-Pulido, J.; Gutierrez de Mesa, J.; Garces-Jimenez, A. An Autonomic Cycle of Data Analysis Tasks for the Supervision of HVAC Systems of Smart Building. *Energies* **2020**, *13*, 3103. [CrossRef]
31. Bakr, M.A.; Lee, S. Distributed Multisensor Data Fusion under Unknown Correlation and Data Inconsistency. *Sensors* **2017**, *17*, 2472. [CrossRef]

32. Kharbouch, A.; Berouine, A.; Elkhoukhi, H.; Berrabah, S.; Bakhouya, M.; El Ouadghiri, D.; Gaber, J. Internet-of-Things Based Hardware-in-the-Loop Framework for Model-Predictive-Control of Smart Building Ventilation. *Sensors* **2022**, *22*, 7978. [CrossRef] [PubMed]
33. Peña, M.; Biscarri, F.; Guerrero, J.I.; Monedero, I.; León, C. Rule-based system to detect energy efficiency anomalies in smart buildings, a data mining approach. *Expert Syst. Appl.* **2016**, *56*, 242–255. [CrossRef]
34. Martin Nascimento, G.F.; Wurtz, F.; Kuo-Peng, P.; Delinchant, B.; Jhoe Batistela, N. Outlier Detection in Buildings' Power Consumption Data Using Forecast Error. *Energies* **2021**, *14*, 8325. [CrossRef]
35. Wang, Z.; Srinivasan, R.S. A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models. *Renew. Sustain. Energy Rev.* **2017**, *75*, 796–808. [CrossRef]