Multivariate Time Series Forecasting with Deep Learning Proceedings in Energy Consumption

Nédra Mellouli[1][a], Mahdjouba Akerma[2], Minh Hoang[2], Denis Leducq[2] and Anthony Delahaye[2]

[1]*LIASD EA4383, IUT de Montreuil, Université Paris 8, Vincennes Saint-Denis, France*

[2]*Irstea, UR GPAN, Anthony, France*

Keywords: Demand Response, Deep Learning, Time Series Forecasting.

Abstract: We propose to study the dynamic behavior of indoor temperature and energy consumption in a cold room during demand response periods. Demand response is a method that consists of smoothing demand over time, seeking to reduce or even stop consumption during periods of high demand in order to shift it to periods of lower demand. Such a system can therefore be tackled as the study of a time-series, where each behavioral parameter is a time-varying parameter. Different network topologies are considered, as well as existing approaches for solving multi-step ahead prediction problems. The predictive performance of short-term predictors is also examined with regard to prediction horizon. The performance of the predictors are evaluated using measured data from real scale buildings, showing promising results for the development of accurate prediction tools.

1 INTRODUCTION

In France as elsewhere in the world, the balance between electricity production and consumption is a necessity that must be maintained and for which electricity suppliers are required to index electricity production to the demand of grid users. To do this, suppliers must regularly deploy more production resources with fast response times. In addition, the European Union has set itself the target of increasing the share of renewable energy in energy consumption to 27% by 2030, from 17% in 2016. However, these new means of production, whose productivity may vary according to time of day, season or climate, require a rethinking of the use of the global electricity grid to allow more flexibility. Due to this high complexity and when the main objective is the final result obtained at the output of the system independently of internal operation, it may be interesting to examine the use of Black Box models, the purpose is to predict the output parameters according to the inputs. The study of the dynamic behavior of such a system can therefore be approached as the study of a time series, where each behavioral parameter is a time-varying parameter. A time series is a sequence of real-valued signals that are measured at successive time intervals. Long short-term memory (LSTM)(Hochreiter and Schmidhuber, 1997), a class of recurrent neural networks (RNNs)(andY. Bengio and Hinton, 2015), is particularly designed for sequential data. For time series prediction task LSTM has particularly shown promising results. Four deep neural network architectures derived from the LSTM architecture were studied, adapted and compared. Their validation was carried out using experimental data collected in a cold room in order to evaluate their performance in predicting demand response. In this paper, we present our methodology allows us to effectively answer the following questions:1) Which deep learning models are best suited to represent our specific data? 2) From the selected models, we have sought to define and characterize the most efficient predictive models and to highlight all the parameters that induce them. 3) In addition to their relevance, we are looking to evaluate the robustness of the selected models at the last iteration against the data quality (noise), data temporal window scaling and large scaling data. The paper is structered as following. In the next section we present the related work, then we detail our approach. Finally we describe the datasets used to compare four deep leraning architectures and we discuss their results.

[a] https://orcid.org/0000-0001-8858-9902

Mellouli, N., Akerma, M., Hoang, M., Leducq, D. and Delahaye, A.
Multivariate Time Series Forecasting with Deep Learning Proceedings in Energy Consumption.
DOI: 10.5220/0008168203840391
In *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2019)*, pages 384-391
ISBN: 978-989-758-382-7

2 RELATED WORK

2.1 Time Series Models in Electrical Demand Response Prediction

Time series is defined as a sequence of discrete time data. It consists of indexed data points, measured typically at successive times, spaced at (often uniform) time intervals. Time series analysis comprises the different methods for analyzing such time series in order to understand the theory behind the data points, i.e. its characteristics and the statistical meaning (Nataraja et al., 2012). A time series forecasting model predicts future values based on known past events (recent observations). Conventional time series prediction methods commonly use a moving average model that can be autoregressive (ARMA)(Rojo-Alvarez et al., 2004), integrated autoregressive (ARIMA) (Hamilton, 1994) or vector autoregressive (VARMA)(Rios-Moreno et al., 2007), in order to reduce data. Such methods must process all available data in order to extract the model parameters that best match the new data. These methods are useless in the face of massive data and real-time series forecasting. To address this problem, online time learning methods have emerged to sequentially extract representations of underlying models from time series data. Unlike traditional batch learning methods, online learning methods avoid unnecessary cost retraining when processing new data. Due to their effectiveness and scalability, online learning methods, including linear model-based methods, ensemble learning and kernels, have been successfully applied to time series forecasting. Each time series forecasting model could have many forms and could be applied to many applications. For more detailes we can see (Amjady, 2001) (Aman et al., 2015).

In our application context, (Hagan and Behr, 1987) have been reviewed time series based models for load forecasting. Then in 2001 (Amjady, 2001) has studied time series modeling for short to medium term load forecasting. To predict energy consumption some authors have used time series data. For example, (J.W et al., 2006) have concentrated their study on the comparison of the performance of the methods for short-term electricity demand forecasting using a time series. (Simmhan et al., 2013) have focused their study on prediction of energy consumption using incremental time series clustering, (Sheng and Duc-Sonand, 2018) have also forecasted the energy consumption time series using machine learning techniques. In (Aman et al., 2015) the work is focused on increasing the accuracy of prediction models for dynamic demand response, this prediction is based on a very small data granularity (15 min intervals). The focus on demand response has been on large industrial and commercial consumers (Ziekow et al., 2013) which are expected by their high contribution and adopted for the smart meters (Simmhan et al., 2013).

2.2 Nonlinear Models

Due to the very high complexity and need for accuracy that the use of linear modeling, which is very time-consuming, can imply, the transition to another more applied type of modeling can simplify the study. Indeed, when the main objective is the final result obtained at the output of a system, regardless of internal operation, it may be interesting to look at the use of non-linear models, whose purpose is solely to predict the output parameters from the inputs. This presents in addition the advantage of being much more easily generalizable, at least in the presence of data of sufficient good quality and by finding a model corresponding to our problematic, without requiring a reshaping of the problem and adaptation of the different parameters when studying a new system. In particular, Artificial Neural Networks (ANN) could provide an alternative approach, as they are widely accepted as a very promising technology offering a new way to solve complex problems. ANNs ability in mapping complex non-linear relationships, have succeeded in several problems such as planning, control, analysis and design. The literature has demonstrated their superior capability over conventional methods, their main advantage being the high potential to model non-linear processes, such as utility loads or energy consumption in individual buildings . At present, although studies (Hu et al., 2017)(Xue et al., 2014) have been carried out within the wide framework of demand response, no such method does appear to have been applied to demand response in the field of refrigeration. In consideration of the energy importance of this field which provides a panel of significant opportunities, the use of a pertinent modelling approach can demonstrate (or invalidate) the use of demand response in cold rooms and cold stores, allowing (or not) a significant increase in the application of electrical cut-off. An LSTM network, or "Long Short Term Memory", is a model for retaining short-term information (recent variations and current trends in data) and long-term ones (periodicity, recurring or non-recurring events). It is a matter of a deep learning model widely used for time series processing. It is popular due to the ability of learning hidden long-term sequential dependencies, which actually helps in learning the underlying representations of time series (Kuo and Huang, 2018).A convolutional LSTM model was proposed based on the Fully Connected LSTM model (Shi et al., 2015), particularly for its application to predicting changes in spatial images. This model has allowed them to obtain better results than using a simple LSTM or Fully-Connected LSTM network. The architecture of an LSTM network therefore consists of a sequence of LSTM layers for both past and future data, which are then processed together to predict current data. To conclude, the modeling of a refrigeration system being characterized by the non linearity and the coupling of several parameters, classical physical models encounter difficulties in predicting the dynamic behaviour of such systems, in particular during disturbances such as electrical cut-off periods. Neural network methods, due to their ability to adjust and self learning, can therefore be very promising in responding to this type of issues offering a new way to solve complex problems. The LSTM ability in mapping complex non-linear relationships, have succeeded in several problems such as planning, control, analysis and design of energy systems. The literature has demonstrated their superior capability over conventional methods, their main advantage being the high potential to model non-linear processes.

3 OUR APPROACH

3.1 Experimental Setup of Cold Room

A cold room or cold store ensures that the products are kept in satisfactory conditions. This therefore requires the use of a refrigeration system, which can take the form of a regular supply of cold air, in order to keep the air and products below a setpoint temperature, generally below $-18°C$. This phenomenon follows the refrigeration cycle, passing through its four main stages, namely compression, condensation, expansion and evaporation. In order to avoid significant heat leakage, it is also necessary to reduce external inputs, through optimal thermal insulation and to minimize door openings as well as human or mechanical activities 1. The products are stored in the form of distributed pallets to reduce the phenomenon of natural warming while facilitating access to the products during loading and unloading. The cold room used in this study is about $2.4m$ long x $2.4m$ width x $2m$ high. For wall insulation, a $10cm$ layer of polyurethane with $\lambda = 0.023W/m.K$ is used. A global heat transfer coefficient $Uc = 0.29W/m2K$ is obtained by measurement. The room temperature is controlled using an on/off strategy and is kept at $-16.3°C$ (within a limited range (set-point) between $-15.5°C$ and $-19°C$). The refrigerated unit consists of a single evaporator with single speed fan. The measured coefficient of performance (COP) is 1.4 at -16°C. The cold room is installed inside an external cell equipped with an air conditioning unit to simulate the meteorological conditions (summer, spring…). To limit the infiltration load, the doorway is kept closed during measurements. In order to measure the air and product temperature, thermocouple sensors are used (Fig. 2). These thermocouples are distributed as follows (*cf.* Figure 1):

- 3 for cold blowing air (Ts) and 3 for return air (Tr)
- 14 for air temperature inside the cold room: 6 in the middle of each wall, 1 near the door, 1 at one wall surface (sample inside cold room), 6 at the corners (2 other corners are not accessible by thermocouples)
- 5 for air temperature in the external cell (Text) on the surface of each wall
- Two wattmeters are used to measure the instantaneous electricity consumption of the refrigeration system (compressor and auxiliary).

The main features of a cold room are of two categories. As a first category, we find fixed features which take into account building geometry (as building dimensions, wall thickness), building composition (as material conductivity, density and overall heat exchange coefficient), outdoor contributions (as outdoor temperature, solar flux, air renewal), cold production (as setpoint temperature, blowing temperature, blowing rate, operation of the refrigeration machine, cooling capacity) and operations on building (as Loading, product conductivity, product density, human presence, lighting, ventilation, defrosting, etc). The second category includes mainly temporal features like the demand response periods ($T_{erasure}$) including both the demand response phase itself (switching off the cooling system) and the recovery phase of the cooling system (restoring the set temperature); The defrost periods, ($T_{defrost}$) which occur several times a day without any decision-making power on the time of appearance. These periods correspond to a specific temperature increasing related to the defrosting of the cold room fan; The compressor on/off periods ($compressor$), which occur very regularly and on an ad hoc basis; The time elapsed since the last demand response period ($\delta T_{erasure}$), allowing the model to better predict the behavior of the cold room in the moments following the demand response period, while its condition is not yet restored; And the time elapsed since the last defrost ($\delta T_{defrost}$), also allowing better prediction of the behaviour of the cold room in the moments following defrosting,

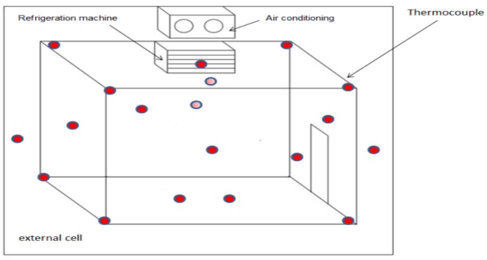

Figure 1: Example of a used cold room.

when the condition of the cold room is not yet stabilized. According to the time parameters defined below we obtain a system with three modes behavior. The first mode (Fig.2(1))represents the steady state where regular temperature variations according to set points (high/low) are measured. The second mode ((Fig.2(2))is the critical state of the system refers to electrical demand response. During this mode, the temperature increases over a long period of time. This period depends mainly on the demand response time interval, typically varies between 30 minutes and 3 hours.The third mode ((Fig.2(3))is the defrost state in which the temperature suddenly increases over a short period of time.

3.2 Multivariate Electrical Demand Response Data Time Series

In this section we define the list of time series data we aim to predict. Indeed, we consider four data classes. The first class is dedicated to measure the Indoor Temperature (Tp_{In}) using 5 sensors thus we have five Tp_{In} time series. Sensors are located in five strategic positions in the cold room. The second class is the Recovery Temperature of the Air (Tp_{Air}) in the cold room and we have three sensors to measure three Tp_{Air} time series. The third class is focused to measure the Product Temperature (Tp_{Prod}) with 8 sensors and thus we have eight Tp_{Prod}) time series. The last class is to measure the Outdoor Temperature (Tp_{Out} using five sensors and we obtained five outdoor temperature time series. Compressor consumption ($Comp_Energy$) which is a particular data time series correlated to the previous variable classes is also considered. To solve the problems related to the simultaneous occurrence of defrosting and demand response phenomena, i.e. the confusion of their measurement periods, we have chosen to recover the defrosting and demand response programs directly by observing the temperature variations that occur during demand response period and defrosting. All time series data are measured each five seconds.

3.3 LSTM Models to Forcasting Electrical Demand Response

We note $\{tp_1, tp_2, ..., tp_n\}$ a time series representing any parameter Tp_x in the output set where $x \in \{Air, In, Product\}$. Our predictive model learns useful features from a set of time series parameters to give a prediction $t\hat{p}_t$ that compares with tp_t to update itself, where tp_t is the real value measured at time t and $t\hat{p}_t$ is the time series data point forecasted at the same time. As it is proved in the literature, Long Short Term Memory (LSTM) is required for discovering a dependence relationships between the time series data by using specialized gating and memory mechanisms. For this purpose, we are aimed to compare four LSTM models : LSTM, Convolutionnal LSTM, Stacked LSTM, Bidirectional LSTM. As a first model we have used **LSTM**.

After a huge number of experiments following several evaluations of the model, the parameters were selected for this model:

- 1024 units are able to store enough information. This choice was made balancing the learning time and the quality of obtained prediction.
- A linear or SeLu activation function, giving us better results than the other functions
- A cost function using the Root Mean Square Error (RMSE),

The second model is **stacked LSTM**. The developed Stacked LSTM model is a modification of the LSTM network described above, with the addition of several layers of LSTM.

For this purpose, and after experimental tests, the selected model consists of a stack of three layers of LSTM, namely: A first layer of 1024 memory units, allowing to store a large amount of information in the short and long term; A second layer of 512 memory units; A third layer, of 256 memory units;

Each layer of the network is separated from the next layer by a dropout layer, allowing less overtraining and robust generalization results.

For the **Bidirectional LSTM** network, it consists of a stack of two layers of LSTMs each with 512 memory units. This model thus created makes it possible to keep information related to both past and future data.

The **convolutional LSTM** network was chosen with the following parameters: 40 filters, corresponding to the outputs of the convolutional part of the model; A kernel size of 2x10, corresponding to the dimensions of the convolution window; A normalization layer, allowing to normalize the activations of the convolutional LSTM layer.

3.4 Performance Metric for Evaluation

In order to compare our different models and to select the appropriate model(s) for our study, different criteria were implemented and computed. In the following, the reference values will be indicated by Y, the predicted values by $\widehat{Y}$, the average of the reference values by $\bar{Y}$ and the number of observations by N. The Fit criterion is needed for measuring the proximity between the reference values and the predicted values. The closer its value is to 100%, the more it indicates a correctly predicted variable. It therefore corresponds to a percentage and is defined by: $Fit(Y) = 100.(1 - \frac{|\widehat{Y}-Y|}{|Y-\bar{Y}|})$.

The Mean-Squared-Error (MSE), or mean square error, is the arithmetic mean of the squares of the differences between the forecasts and the actual observations. The objective of a good prediction is therefore to obtain the lowest possible mean square error. The advantage of squaring is to highlight high errors, and therefore to minimize low prediction errors. This value is therefore defined by: $MSE(Y) = \frac{1}{N}.\sum_{i=1}^{N}(\widehat{Y}_i - Y_i)^2$. The root of this value, or Root-Mean-Squared-Error (RMSE), is also often used, which is simply calculated by: $RMSE(Y) = \sqrt{MSE(Y)}$.

The Mean Absolute Error (MAE), or absolute mean error, is the arithmetic mean of the differences between forecasts and actual observations. Since there is no squaring, this measure treats each difference with equal importance. The objective is of course to minimize this value, and it is defined over prediction horizon $[t+1, t+H]$ by: $MAE(Y) = \frac{1}{N}.\sum_{i=1}^{N}|\widehat{Y}_i - Y_i|$.

The coefficient of variation (CV) is a little-known measure that has been proposed by[Karatasou et al., 2006][Amasyali et al., 2016] to evaluate the prediction of models for building energy consumption. This value is defined as a percentage, by the formula : $CV(Y) = 100.\frac{RMSE(Y)}{\bar{Y}}$.

As this value is only used in the field of energy consumption, it will therefore only be evaluated on consumption values, and not on temperatures (the latter may also be negative).

4 EXPERIMENTAL RESULTS

As mentioned above, we implemented the four LSTM architectures described in the section 3.3. To evaluate their performance in predicting demand response, a set of use cases were developed based on experimental data collected in cold rooms. For the cold room, we had access to large periods of measurement time with acceptable accuracy where measurements were made every five seconds. These measurements were made in a cold room that was replicated in a controlled environment to obtain data to form the models and prove their predictive power. For these purposes, we have proposed five use cases allowing us to answer efficiently to the following questions :1) Which deep learning models are more adapted to represent our specific data? 2) Which model(s) is less sensitive to the stochastic occurrence of electricity demand response? 3) Which model(s) is more robust to the data quality, i.e. signal noise (electrical demand response) and horizon window? We assume that the reference behaviour of a cold room is characterized by an ideal "undisturbed" operation, with no phenomenon of demand response or door opening and this over a long period of time, in order to stabilize the internal temperature. These data measurements are used only to initialize the four architectures. Then we have established five time series datasets to evaluate each LSTM architecture. These datasets are described as follows. We note by E_i a dataset time series where $i = 1..5$ elaborated for each use case i:

- The use case 1 (train: 127975, test: 60000)hypothesis is to consider a set of measurements with three electrical demand response periods, uniformly distributed over 3 days. Here we simulate the stochastic disturbance of the system on considering a uniform distribution of the noise signal. Hence $\delta T_{erasure}$ is randomly decreased and $T_{erasure}$ is fixed ;
- The use case 2 (train: 223545, test: 149030) hypothesis is to consider a set of measurements with two electrical demand response periods per day, uniformly distributed over 5 days. In this case we increase the frequency of occurrence of the noise in use case 1 and increase the total number of measurements. Indeed, this case allows us to study the bias of the frequency of noise occurrence as well as the amount of data;
- The case 3 (train: 490985, test: 294590) hypothesis is to consider a set of measurements over 5 days with one electrical demand response period randomly occured per day and with a random period. It means both $\delta T_{erasure}$ and $T_{erasure}$ are random.
- The case 4 (train: 630920, test: 420610) hypothesis is to consider the union of the three previous hypothesis. It corresponds to a generalized model of the electrical demand response problem, i.e. We have a large amount of data, more noise and more randomness.
- The case 5 (train: 214080, test: 142700) hypoth-

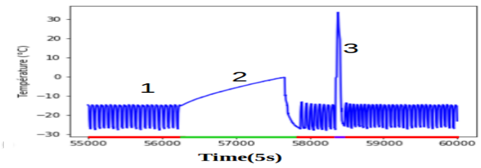

Figure 2: Three modes: (1)Steady state, regular temperature variations according to set points (high/low) ; (2)Electrical demand response, temperature increases over a long period of time (30min-3h) ; (3)Defrosting, sudden temperature increase for a very short time (5-10min).

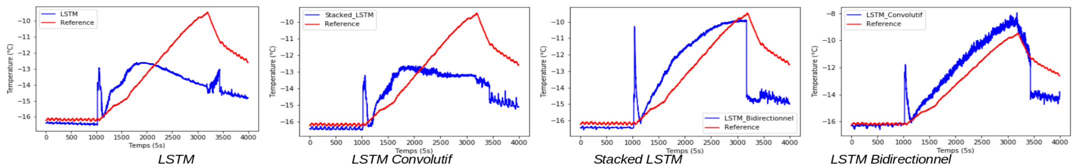

Figure 3: LSTM models prediction of $Tp_{product}$ in use case 1.

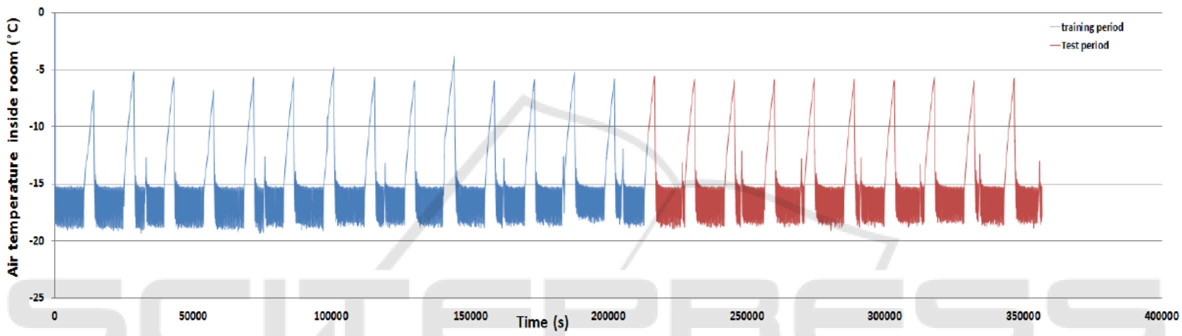

Figure 4: LSTM data time series prediction with $E1$ datasets.

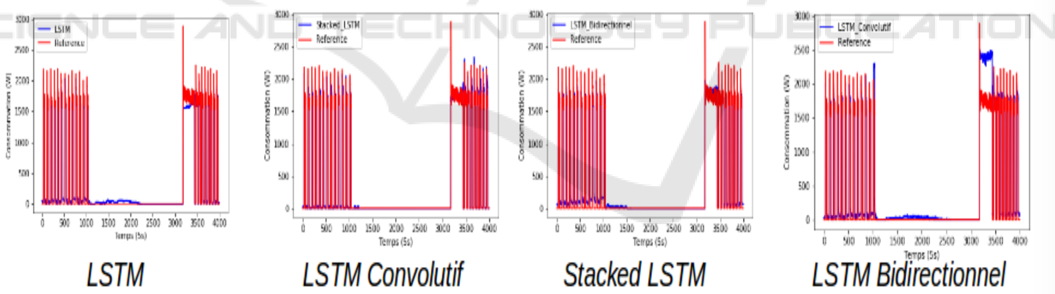

Figure 5: LSTM models prediction of $Compressor_{Energy}$ in use case 1.

esis is to consider a fixed $\delta T_{erasure}$, a fixed $T_{erasure}$ varing between 1 and 3 hours and we increase prediction horizon H.

Each dataset E_i is respectivly splited into 60% of data for training (Train) and 40 for validation (Test) sets, except the E_5 used as Test sample.

4.1 Results Analysis

As observed graphically (Figures: 4,3,5), we can therefore see much more significant results thanks to the use of the convolutional LSTM network using the E_1 dataset, on all features except $Compressor_{Energy}$, which could be verified graphically. The other models developed and derived from the LSTM model give us interesting but much less significant results than those found with the convolutional LSTM model. However, it can be noted that the Stacked LSTM and Bidirectional LSTM models obtain fairly high performance in terms of air temperature. However, these results are still very modest, which can easily be explained by the small amount of data. In addition, it should be noted that some features, such as $Tp_{product}$, are insufficiently predicted by all models. These problems

Table 1: E_i Temperature T_p prediction with the four derived LSTM models.

T_p	LSTM	ConLSTM	StackedLSTM	BidirectionalLSTM
E_1	(Mae =0.38,Fit=35.8)	(Mae=0.38, Fit=60.5)	(Mae=0.44, Fit=42.8)	(Mae =0.5, Fit=41.7)
E_2	(Mae =0.16,Fit=29.7)	(Mae=0,33, Fit=14.8)	(Mae=0.14, Fit=30.4)	(Mae =0.31, Fit=24.4)
E_3	(Mae =0,27,Fit=48.0)	(Mae=0.26, Fit=46.9)	(Mae=0.23, Fit=50.9)	(Mae =0.31, Fit=52.1)
E_4	(Mae =0.27,Fit=58.2)	(Mae=0.33, Fit=55.5)	(Mae=0.27, Fit=61.1)	(Mae =0.37, Fit=48.4)
E_5	-	-	(Mae=0.19,Fit=69.64)	-

Table 2: E_i $Compressor_{Energy}$ prediction with the four derived LSTM models.

$CV(Compressor_{Energy})$	LSTM	ConLSTM	StackedLSTM	BidirectionalLSTM
E_1	18.9	23.5	18.9	20.3
E_2	15.7	13.7	47.1	20.08
E_3	15.1	13.8	54.42	22.6
E_4	16.2	15.9	16.1	16.1
E_5	-	-	28.5	-

seem to be partly related to the sensors. A more detailed analysis will be carried out later to check the proper functioning of these different sensors. The results obtained with the dataset E_2 are less efficient for predicting temperature-related features. In particular, we note a sudden increase in the predicted temperature as soon as the electrical cut-off is triggered, followed by an equally sudden decrease at the end of the demand response period. As far as the energy consumption is concerned, the behaviour is always reproduced as faithfully as ever. With this dataset Bidirectional and Stacked LSTM outperform the other models. We can note that Stacked and bidirectional LSTM are less sensible the amount of data. With the E_3 dataset, simple and Convolutional LSTM outperform the other models. They seem to be more efficient in the presence of random noises. Bidirectional and Stacked LSTM are able to predict the dynamics of time series but are sensitive to noise, especially during the starting times of both the demand response and the retakes. This is manifested by peaks of values predicted by the last two models. With the E_4 dataset, we recall that is the union of the three previous datasets, we obtained comparable results with LSTM and Stacked LSTM. Their predictive accuracy outperforms the other two models. With the dataset E_4 we obtained comparable results with LSTM and Stacked LSTM, their predictive accuracy outperforms the other two models. It should be noted, however, that four models are trained on the union of E_1 and E_2 and they have to predict on the E_3 dataset according to the percentages of the Train and Test samples. What is interesting is that we expected the results in this use case to be similar to the previous use case (use case 3). However, learning more noise allows stacked systems to better predict noise and random data. Also they seem to be less sensible to the amount data. Finally the dataset E_5 is used as Test sample for the pre-trained Stacked LSTM. Since the horizon size of E_5 is 20 minutes, four times bigger than the previous ones, time series data are less. As long as the size of the E_5 horizon is four times larger than the previous ones, the data are more smoothed and less noisy. As a result, the model was able to better predict with a Gain of approximately 10%. The $Compressor_{Energy}$ is efficiently predicted in all use cases where the fitting is arround 90% and the MAE is arround 0.1. This is due to its independent state from the defrosting and electrical demand response and it has a stationary state.

5 CONCLUSIONS AND FUTURE WORK

Although the results obtained in the study of series E_1 and E_4 were quite satisfactory, the results of series E_2 and E_3 remain rather moderate, as can be seen from the graphs and values given above (Table 1, 2). Indeed, the different models, although trained and then tested on larger data sets, seem to encounter difficulties in generalizing prediction on test values. However, some models and their predictions are encouraging, suggesting that the use of deep learning methods could lead to better results through improvements and the use of more data. In particular, Stacked LSTM seems to be the efficient deep learning architecture providing acceptable predictions in the context of our specific data. Indeed, the modeling of a refrigeration system being characterized by the non linearity and the coupling of several parameters, classical physical models encounter difficulties in predicting the dynamic behaviour of such systems, in particular during disturbances such as electrical demand response periods. Stacked LSTM, due to its ability to adjust and self learning, can therefore be very promising in responding to this type of issues. To increase its efficiency, it could be a possible perspective to use weight masks during training. Therefore, transfer learning could be able to favour the adjustment of weights during demand response periods, and would obtain predictions closer to the reference values.

REFERENCES

Aman, S., Frincu, M., Chelmis, C., Noor, M., Simmhan, Y., and Prasanna, V. K. (2015). Prediction models for dynamic demand response. In *IEEE International Conference on Smart Grid Communications (SmartGridComm)*.

Amjady, N. (2001). Short-term hourly load forecasting using time series modeling with peak load estimation capability. In *IEEE Transactions on Power Systems*.

andY. Bengio, Y. L. and Hinton, G. (2015). Deep learning. In *Nature*, volume 521, pages 436–444. Google Scholar.

Hagan, M. T. and Behr, S. M. (1987). The time series approach to short term load forecasting. In *IEEE Transaction on Power System*, pages 785–791.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, New Jersey, NJ, USA.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. In *Neural Computation*, pages 1735–1780. Google Scholar.

Hu, M., Xiao, F., and Wang, L. (2017). *Investigation of demand response potentials of re-sidential air conditioners in smart grids using grey-box room thermal model*.

J.W, T., de Menezes L.M, and P.E, M. (2006). A comparison of univariate methods for forecasting electricity demand up to a day ahead. In *International Journal of Forecasting*, volume 22, pages 1–16.

Kuo, P.-H. and Huang, C. (2018). *A High Precision Artificial Neural Networks Modelfor Short-Term Energy Load Forecasting*.

Nataraja, C., G.M.B, Shilpa, G., and Harsha, J. S. (2012). Short term load forecasting using time series analysis: A case study for karnataka. In *International Journal of Engineering Science and Innovative Technology*, volume 1, India.

Rios-Moreno, G., Trejo-Perea, M., Castañeda-Miranda, R., Hernández-Guzmán, V., and Herrera-Ruiz, G. (2007). Modelling temperature in intelligent buildings by means of autoregressive models. In *Automation in Construction*, volume 16, pages 713–722.

Rojo-Alvarez, J. L., Martınez-Ramon, M., and de Prado-Cumplido, M. (2004). Support vector method for robust arma system identification. In *IEEE Transactions on Signal Processing, vol. 52, no. 1*, pages 155–164.

Sheng, C. J. and Duc-Sonand, T. (2018). Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. In *Energy*, volume 165, pages 709–726.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., and Woo, W. (2015). *Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting*.

Simmhan, Y., Aman, S., Kumbhare, A., Liu, R., and Stevens, S. (2013). Cloud-based software platform for data-driven smart grid management. In *IEEE/AIP Computing in Science and Engineering*.

Xue, X., Wang, S., Yan, C., and Cui, B. (2014). *A fast chiller power demand response control strategy for buildings connected to smart grid*.

Ziekow, H., Goebel, C., Struker, J., and Jacobsen, H. (2013). The potential of smart home sensors in forcasting household electricity demand. In *IEEE International Conference on Smart Grid Communications (SmartGridComm)*.