Authors accepted manuscript. Article published Vol. 5, Issue 1, 2020
The final published version available at https://doi.org/10.1109/TSUSC.2018.2886164

Time Series-Based GHG Emissions Prediction for Smart Homes

Ana Carolina Riekstin, *Member, IEEE,* Antoine Langevin, Thomas Dandres, *Member, IEEE,* Ghyslain Gagnon, *Member, IEEE,* and Mohamed Cheriet, *Senior Member, IEEE*

Abstract—Smart homes play a crucial role in reducing the residential sector electricity consumption and Greenhouse Gases (GHG) emissions. In this work, we present a time series approach to predict GHG emissions to be integrated into smart home management systems. More specifically, we used Long Short-Term Memory (LSTM), a variant of Recurrent Neural Networks. The prediction results get mean absolute percentage error (MAPE) close to 2 % when the region under study has an energy matrix mostly based on fossil fuels, less intermittent. For regions in which more renewable sources are present, the MAPE is around 12 %. However, in either case, LSTM can predict the hours well with smaller emissions among the next 24 hours. Such day-ahead information brings awareness to the users and allows the scheduling of appliances to work in the hours in which the emissions are minimal, reducing them without significantly affecting the consumers' behavior.

Index Terms—Smart Homes, Emission Factors, GHG Emissions, Prediction, Time Series, LSTM

1 Introduction

The residential sector is among the major contributors to the increasing electricity demand and Greenhouse Gases (GHG) emissions [1]. Only in Canada, the residential sector was responsible for 17% of the total electricity consumption in 2015 and around 13% of the total GHG emissions (including electricity) in this same year [2]. In this context, smart homes play a critical role with its main objectives regarding home automation, energy management, and environmental emissions reduction [3]. As an example, [4] cites that smart home automation can reduce 30% of heating energy and carbon emissions. In a broader context, the building sector is responsible for 30% of global CO_2 emissions [5]. GeSI reported that smart building solutions could cut $2.0Gt$ of CO_2e from the housing sector [6], being one of the eight sectors that will profit the most from the different ICT-enabled possibilities to reduce GHG emissions.

The different sources used during the day to produce electricity and meet the region's demand can make the emission factor of a region (in gCO_2e/Wh) vary a lot, which represents an opportunity to decrease emissions by managing the demand accordingly [7]. The possible adoption and consequent potential to reduce emissions is more limited if the changes largely impact consumers' behavior, but there are flexible loads that allow scheduling with reduced impacts for the users. For instance, charging an electric vehicle during the hours it is parked by selecting the moment in which the emissions (and/or price) are minimal [7]. Home Energy Management Systems (HEMS) may be used to schedule the electrical loads in a smart home to reduce emissions, which requires information about emission factors in advance [7].

Information about forecasted hourly emissions is not very commonly found yet, but some initiatives are starting to appear, like the Carbon Intensity API [8] which provides 2-day forecast for the UK. In other places, this data is not yet explicitly disclosed, but other information available allows the calculation. For instance, the IESO (Independent Electricity System Operator) in Ontario (Canada) is now providing day-ahead energy generation forecast per source[1], which, in conjunction with the forecasted energy imports from neighbors, enable the estimation of the GHG emissions for the next day. However, not all regions are already providing such data publicly. In France, for example, the emission factor of the electricity produced in the country is provided, near real-time, calculated from the energy sources used in the generation [10], but not yet for the next day.

In order to have this day-ahead information, we need to predict the emissions due to electricity generation in the region where the smart home is located. Hourly emission factors (or even in smaller granularities) have gotten increased interest as a more accurate alternative to the traditional annual average emissions. In this paper, we present a time series approach to predict the emission factors for the next day using Long Short-Term Memory (LSTM), a variant of Recurrent Neural Networks (RNN). We then apply this information to reduce the emissions of loads which allow flexible scheduling.

We validate the use of LSTM for emission factors prediction using data available online for regions with different energy profiles: one profile which uses mostly fossil fuels and nuclear sources to generate electricity, and another which also relies partially on nuclear, but has a bigger share of renewables to produce electricity. The second profile is more

- *A. C. Riekstin, A. Langevin, and M. Cheriet are with Synchromedia Laboratory - École de technologie supérieure, Université du Québec (Canada). E-mail: ana-carolina.riekstin.1@etsmtl.net*
- *G. Gagnon is with LaCIME Laboratory - École de technologie supérieure, Université du Québec (Canada).*
- *T. Dandres is with CIRAIG - Polytechnique Montréal, Université de Montréal (Canada).*

1. "Adequacy" report at [9]

challenging due to the intermittency of electricity generation using renewable sources of energy. We demonstrate the prediction applicability by presenting a use case in which an appliance is scheduled to run on the best time of the day from the GHG emissions point of view. We also validate the approach inside an optimization scenario, minimizing emissions while taking into account other aspects.

Despite the extensive literature on predicting load and pricing, to the best of the authors' knowledge, this is the first work to employ time series mechanisms to predict average GHG emissions due to electricity generation and use. In a prior study aiming at predicting emissions [11], we have not considered the dataset as a time series, therefore we did not capture information from prior days that may affect the prediction for the next day. Besides, we used an internal database, which we changed now for open, downloaded and organized data for replicability purposes. The contribution of this work is therefore twofold: we propose a time series approach to predict GHG emission factors for the next 24 hours in order to allow the smart scheduling of flexible electricity loads, and we propose the use of LSTM for such prediction, combining past data and other associated day-ahead information.

The remainder of this paper is organized as follows. Section 2 presents the related work, starting with smart homes energy and emissions management, concluding with related work on predicting electricity data. Section 3 brings the problem formulation and Section 4 details the data collection process. Section 5 presents the prediction model, and Section 6 shows the prediction results, as well as a comparison with a traditional time series approach. Section 7 describes two use cases for the emissions predicted, and Section 8 brings the final remarks.

2 Related Work

The prior publications related to this work are related to smart home energy and emissions management and emissions due to electricity usage prediction.

2.1 Smart Homes Energy and Emissions Management

Smart home energy management with optimization objectives and tools to increase user awareness have been studied for some years. As examples, there are the Energy Aware Smart Home [12] and the SESAME-S project (SEmantic SmArt Metering Services for Energy Efficient Houses) [13]. Such systems can also incorporate the control of generation and storage of energy locally produced from renewables, such as wind turbines and, most commonly, solar panels [14]. GreenCharge [15], for instance, is an optimization approach to minimize electricity costs and peak hours demand which does that, also considering the home energy demand, local batteries, and electricity pricing varying throughout the day.

The pricing information might be available through the utilities websites or by specific protocols. The local energy production can be predicted based on historical solar irradiance in the geographic region under study. Such information is used by [16] in conjunction with a genetic algorithm and linear programming to schedule home appliances, while [17] and [18] do that for electric vehicles charging. Similar predictions can be developed for wind energy production.

The dynamic pricing schemes ("TOU," Time of Use period or "real-time") as used in GreenCharge [15] for smart home energy management has also been studied, such as in the optimization approach in [19] or in conjunction with a list of appliances priority (preferences) in [20] to not significantly degrade consumer comfort.

However, fewer works to date have talked about variable GHG emissions. One example is [21], which differentiated electricity generation emission factors from consumption factors (including imported energy from neighbors), and evaluated the environmental performances of demand-side management programs. Specifically for smart homes, the Swedish Active House [22] proposes an architecture in which the house communicates with the Utility company via a specialized interface to receive a 24-hour forecast of CO_2 emissions. The hourly information was not available at the time of their publication, so the authors defined a model using historical information, assuming that the production sources were dispatched according to their pricing (cheapest first), and also comprising imports and exports between adjacent regions.

Kopsakangas-Savolainen et al. [7] presented theoretical use cases in which the daily variation of emission factors enable GHG emissions reduction from 3 to 8% due to optimization of time of use without significantly affecting the users' behaviors. Only historical information was necessary for the study. The day-ahead GHG emissions, based on the sources used to produce energy, while not provided directly by the Utility to the home energy management systems, need to be predicted in order to be used in such smart systems.

2.2 Emissions Due to Electricity Usage Prediction

Electricity load and pricing prediction are the subjects of diverse prior work [23], using different techniques. Such forecasts can be classified as short-term (1 hour to 1 week), medium-term (one week to one year), and long-term (more than one year) [24]. Time series models like ARMA (Autoregressive Moving Average) and ARIMA (Autoregressive Integrated Moving Average) are popular approaches for predicting load and price and can serve as benchmark models for comparing with new proposals [23]. Load and price are commonly non-stationary, presenting high volatility and nonlinearity [25]. Such characteristics must be tackled before applying time series models, for instance, by differencing the data. There is also a consensus that electricity data suffers from multiple seasonality - per day, week, and year [24] [26], which may be dealt with by using Fourier decomposition, or through the inclusion of dummy variables.

As an example, Contreras et al. [27] used ARIMA for next-day electricity prices, treating daily and weekly seasonalities, assuming residuals as white noise, and accepting the series as stationary after a log transformation for stability. An option to detect and adjust possible unusual observations (outliers) is also selected. Even with all the adjustments, when the data is highly unstable, the model hypotheses are not met. Hinman and Hickey [24] used an ARMAX (Autoregressive Moving Average model with exogenous weather variables) approach to forecast electricity

load 24 hours ahead. In this work, each hour's load is treated separately, as an individual time series, thus, according to the authors, "avoiding modeling complicated intraday patterns" which vary weekly and through the seasons.

Another possibility is to use a multivariate model. VAR models, a generalization of the univariate autoregressive model for forecasting a collection of variables are suitable when all variables affect each other [28]. Raviv et al. [29] used the VAR method to predict the daily average price for the 24 hours of the next day, which are defined all together in the previous day. Even being useful in several contexts, [28] points out that VARs suffer criticism for being "atheoretical", meaning such model is not built on top of an economic theory. Besides, the assumption that every variable affects every other in the system makes the interpretation of estimated coefficients difficult.

According to Dudek [30], one may use two approaches for solving multi-output regression problems: (1) by decomposing the problem into multiple single-output problems or (2) by adapting a model so that it directly handles multi-output data. In (1), the relationship among the variables is ignored, probably reducing the accuracy. But (2) is not very popular due to the inherent complexity, despite having better predictions. For load time series forecasting, the author points out two approaches: conventional (comprised of regression methods like ARIMA and exponential smoothing) and unconventional (neural networks, SVM, fuzzy).

Regarding the ones classified as "unconventional," [24] highlights AI-based techniques, which are flexible methods that may not demand prior experience for getting good forecasts results. Such methods are usually said to be "black-boxes", which may also over fit and take a long time for training, but there is evidence from practical applications that suggests they may perform well. Singhal and Swarup [26] used Artificial Neural Networks (ANNs) for pricing forecasting, using time (day of the week, hour of the day), historical data, and also load forecasting as inputs.

Panapakidis and Dagoumas [23] developed an electricity pricing forecaster based on ANNs and hybrid approaches. They argue that ANNs are suitable for problems without constant mean and which display high volatility, like pricing data, very normally a non-stationary time series. In their work, they evaluated abnormal data to exclude metering failures and also accounted for exogenous variables like the load demand (assuming the day-ahead load is available), the available generation, the market prices of interconnected countries, generation from renewables, and gas prices. Different models were tested, including one with two serial ANNs and a hybrid model, with clustering and one ANN for each cluster, including neighbors' information. The authors concluded that the clustering stage did not improve the operation, although "clustering improves the forecasting accuracy in some subsets of the main test set." In their experiments, weekends presented the higher fluctuations, with poor results on Sundays. Fridays were also difficult to model. Clustering techniques were also used for pricing prediction in [31] and [25].

Specifically for GHG emissions prediction, [32] used Support Vector Machines (SVM) to predict marginal emission factors - more specifically, 24 SVM models corresponding to each of the 24 hours. Marginal emission factors are related to the extra generation required to supply an additional demand [33] [34]. On our case, instead of the marginal value, the average emission factor is a more appropriate selection, since it is assumed that the load under study is part of existing demand. The same type of emission factor was considered by [35] when studying the impacts of electric vehicle charging in the electric grid. Even with the extra demand to charge the vehicles, the authors assumed that the additional demand would not change the hourly characteristic emission level. In other words, the work assumed that the electricity markets would have enough time to react and adapt to this new demand from EVs.

As it may be seen, there is an interesting set of prior work on predicting electricity load and price, but proposals to predict average GHG emission factors are still lacking. The ActiveHouse assumed the emissions forecast information was available and provided by the Utility, which may not be the case yet in many regions. To make the necessary prediction in order to feed smart home management systems, we may leverage on time series approaches, historical sources, and available forecasts to predict emission factors which may then be used to schedule smart home devices.

3 Problem Formulation

Information to compute emission factors of electricity purchased from the grid can be collected in near real time as we described in our prior work [11]. The evaluation of emission factors from the electricity grid should take into account the generation and the demand significant variations over time (during the day and between the seasons of the year) [33] [36], as well as imports and exports from neighbor interconnected regions. The required information can be obtained from some Electricity System Operators websites, like [9], [10], [37]. When such information is not disclosed, depending on the energy sources used, it might be possible to estimate based on historical reports and neighbors' imports and exports, with the side cost of increased uncertainty.

We consider that imported/exported electricity is the average mix from the region where it comes from and it can be obtained from the neighbors of the neighbor as well. For instance, being r_2 a neighbor region of r_1, and r_3 a neighbor region of r_2; r_1 imports from r_2 which, in turn, imports from r_3 and so on. For the sake of simplicity, we consider only one "level" of imports, assuming they are locally consumed and not exported to another place. Going further in the chain would increase the calculation accuracy.

In this work, we based our problem formulation on our prior work [11], being $R = \{r_i, i = 1, ..., I\}$ a set of I interconnected regions, $S = \{s_k, k = 1, ..., K\}$ a set of K sources of energy. $\mathbf{F}$ represents the *emission factor* of the sources in each region, i.e., $\mathbf{F}_{ik}$ is the emission factor of the energy source s_k in the region r_i. $\mathbf{A}$ represents the energy imported between regions, i.e., $\mathbf{A}_{jik}$ is the percentage of the total energy produced from the source s_k that is imported from the region r_j to the region r_i. In case where $j = i$, $\mathbf{A}_{iik}$ is the percentage of the local energy from the source s_k produced and used in region r_i. "Produced and used" in the region means it is necessary to exclude the electricity exported to be used elsewhere, which is not an issue if we

are using percentages. From the notation above, we define the emission factor of region r_i as in:

$$E_i = \sum_{k=1}^{K} \left(F_{ik} \times A_{iik} + \sum_{j=1, j \neq i}^{I} F_{jk} \times A_{jik} \right) \quad (1)$$

Multiplying an emission factor by a load is the simplest and most common way of calculating emissions due to electricity consumption [33]. One sample list of sources' emission factors $\mathbf{F}_{ik,\forall i}$ is listed in Table 1. The values were defined following the method used by [34], having as basis the Ecoinvent database (version 3.1) using Simapro (version 8.1) and IMPACT2002+ (2.20) for 1 kWh of electricity. To simplify, we have been using the same list for regions in the middle and east coast of North America. The listed emission factors consider the power plants construction and other life-cycle steps, such as fuel extraction, following an LCA (Life Cycle Assessment) approach. Refuse is considered as zero until this moment. Similar lists can be generated for other regions, as we did for France. An alternative to LCA emission factors are the *operational* or *direct* emission factors. We selected the LCA approach because it implies that energy generated from renewables is not emissions-free. The other approach can be used by simply changing the sources' emission factors.

TABLE 1
Sources' emission factors used in this work for the middle and east coast of North America (LCA approach)

Source	gCO2/Wh
Biomass	0.166
Coal	1.157
Gas	0.634
Hydro	0.017
Nuclear	0.023
Oil	1.164
Refuse	0
Solar	0.04
Wind	0.031

4 Data Collection

We have been studying near real-time emission factors information in different regions, as well as ways of standardizing such collection and report with different geographic and time granularities [38]. In this work, for replication purposes, we used information downloaded from the utilities' websites. On the one hand, the drawbacks of this approach are the following:

- Only hourly or half-hourly averages are usually available, which was enough for our study, but may not be for those that demand smaller time granularities (e.g. 5 minutes);
- When downloading the data from interconnected neighbors, the imports/exports information may not be available in the desired temporal or regional granularity either. In this study, for instance, we used the neighbors' monthly or annual averages, losing some accuracy in the results. Having a script capturing information near real-time may give more details.

On the other hand, anyone interested in replicating this study or applying it to their own problem may simply obtain the same public data online, subject to fewer errors and for as many years as the utility makes the data available. The dataset may be obtained by following these steps:

1) Download the local energy generation per source, obtaining A_{iik}, or, if available, download the local emission factor already calculated, obtaining directly $F_{ik} \times A_{iik}$;
2) Download the flows (imports and exports) between the region under study and its interconnected neighbors using the best granularity possible;
3) Make sure the local energy percentage A_{iik} excludes the exports to neighbors, since this energy was used elsewhere;
4) Check on the neighbors' websites or annual reports the fuel mix composition, obtaining s_k for each r_j to calculate A_{jik}. If possible, download their production with the same time granularity as (1); if not available, 2 solutions are possible: (a) use a similar region in conjunction with an LCA database/tool to estimate the emission factors, or (b) use the annual average for the region;
5) Calculate the local emission factor E_i, using a table similar to Table 1 or local sources' emission factors.

In this work, we consider hourly average emission factors, from the entire generation mix, which assumes that the load is distributed among all operating power plants [33]. This is a simplification since the imports are not coming from the whole neighbor, but probably from a set of closer plants. The same applies for the energy being generated and consumed in the region. The average mix approach is still commonly used [39], and we are assuming the temporally differentiated case in order to account for hourly and seasonal variations (therefore, the average is about the energy sources, not about time). As mentioned before, the emission factors can also be marginal, but we are assuming the load under study is part of an existing demand and, therefore, the average emission factor is an appropriate selection.

For this work we used data from the following regions: PJM (USA) [37], Ontario (Canada) [9], and France [10]. These regions have different profiles: PJM relies on gas, coal, and nuclear sources, while Ontario and France, despite also relying on nuclear sources, have a bigger share of renewable sources and, therefore, present higher variability of emission factor. For Ontario and France, we obtained complete data for the years of 2015 to 2017. For PJM, we obtained complete data starting in 20th May 2015 until the end of 2017 - older data is available, but all sources were reported as "Others" at the time of this study. Whenever missing data issues arose, which did not frequently happen in the data sets obtained, in order to account for all the relationships of a day, we copied all information from the same day from the prior week.

Another issue, which may happen as well for other regions, is that for France we were not able to find imports information divided by neighbor. The utility does report the contractual imports, but not the physical flows split. To simplify, we used the European Union average emission factor reported at [40]. To conclude, whenever we encountered

"Multiple Fuels" or "Others" among the electricity sources, we searched the utility reports to check how to "split" the amount, aiming at gaining more accuracy. Similarly, if the Utility provides a detailed list (e.g., breaking down "Hydro" in "run-of-river" and "reservoir") we profit on the details and match the items with the list of fuels provided by Ecoinvent.

Table 2 summarizes some important information that characterizes the datasets. The emission factors are rounded to two decimal points - despite using the complete figures for the prediction and analysis, the final value to be reported to the user has to be rounded due to the inherent uncertainties of the LCA process. Figure 1 presents the average emission factors for all regions divided by seasons for the time frames downloaded. Note that the Y axes of the graphs have different scales, which means the emission factor of PJM is much higher than in Ontario and France.

We made a preliminary analysis of each dataset to check the correlation with past data and found that up to four weeks before we still have significant correlation. Using the past weeks as the input for the prediction also helps to account for cold or warm weather that might be happening. Figure 2 illustrates the analysis done for Ontario as an example. It is also possible to note the daily and weekly seasonalities.

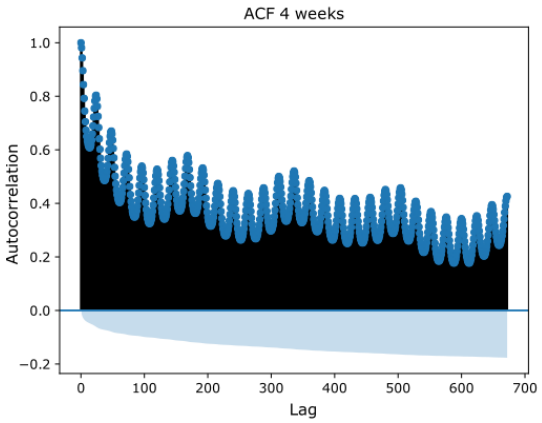


Fig. 2. Autocorrelation Function (ACF) over Ontario hourly data for the period 2015-2017

5 Prediction Model

The problem of emission factors prediction was addressed using LASSO (Least Absolute Shrinkage and Selection Operator) regression analysis in our prior work [11]. In this first study, we have divided the emission factors by hour and season, and have not considered the dataset as a time series. In this paper, we evolved our prediction tool by using a time series approach through Long Short-Term Memory (LSTM).

LSTM is a variant of Recurrent Neural Networks (RNN) introduced by [41]. The main principle of LSTM is its ability to manage its own memory. In contrast with RNN simple neuron, the cell state is not necessarily updated at each time t. In the LSTM cell, there are three *gates*. These gates decide if the cell needs to forget the past information learned, to hold the new incoming data, and to control the cell output. This particularity helps LSTM networks to perform well for long-term dependency contexts, reducing the vanishing and exploding problems encountered with RNN back-propagation algorithm. With gates in cells, important data are kept as long as they are not cleared by them, and irrelevant data are ignored by the cells. This persistence in memory cells make LSTM a good option for the emission factors prediction which takes a time series of 672 steps.

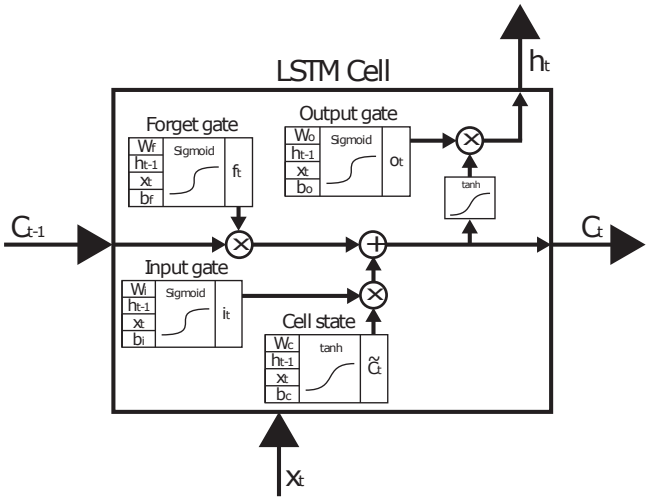


Fig. 3. LSTM Cell.

The LSTM cell shown in Figure 3 represents three layer gates to control flows of information in the network. The *input gate*, represented by Equation 2[2], controls the access of input data. The *forget gate*, considered the most important one and represented by Equation 3, is used to decide if the memory cells need to forget the past information learned in order to keep only the new input information. Finally, the *output gate*, represented by Equation 4 decides if the information inside cells need to be output or not to the next network layer.

$$\mathbf{i}_t = \sigma(\mathbf{W}_i \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i) \quad (2)$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_f \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f) \quad (3)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_o \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_o) \quad (4)$$

$\mathbf{W}_{i,f,o}$ are the weight matrices and $\mathbf{b}_{i,f,o}$ the bias vectors for the input, forget and output gates respectively. $\mathbf{x}_t$ is the input vector of the cell and $\mathbf{h}_{t-1}$ the cell hidden state vector at $t-1$. The brackets in equations 2, 3, 4 and 7 represent the concatenation of $\mathbf{h}_{t-1}$ and $\mathbf{x}_t$. σ is the sigmoid function represented by Equation 5.

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (5)$$

$\mathbf{c}_t$ is the current cells state updated in function of their own previous values $\mathbf{c}_{t-1}$ and the new input information $\tilde{\mathbf{c}}_t$ represented by the Equations 6 and 7. The $\circ$ denotes the element-wise product.

$$\mathbf{c}_t = \mathbf{f}_t \circ \mathbf{c}_{t-1} + \mathbf{i}_t \circ \tilde{\mathbf{c}}_t \quad (6)$$

$$\tilde{\mathbf{c}}_t = tanh(\mathbf{W}_c \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_c) \quad (7)$$

2. LSTM equations have been inspired from http://colah.github.io/posts/2015-08-Understanding-LSTMs/

TABLE 2
Emission factors (gCO_2e/Wh) metrics for the different regions

	ON			PJM			France		
season	**Avg.**	**Min.**	**Max.**	**Avg.**	**Min.**	**Max.**	**Avg.**	**Min.**	**Max.**
spring	0.05	0.03	0.18	0.54	0.40	0.67	0.06	0.02	0.16
summer	0.08	0.02	0.21	0.60	0.43	0.78	0.06	0.03	0.14
autumn	0.06	0.03	0.21	0.54	0.39	0.68	0.12	0.03	0.25
winter	0.07	0.03	0.19	0.55	0.34	0.70	0.10	0.04	0.21

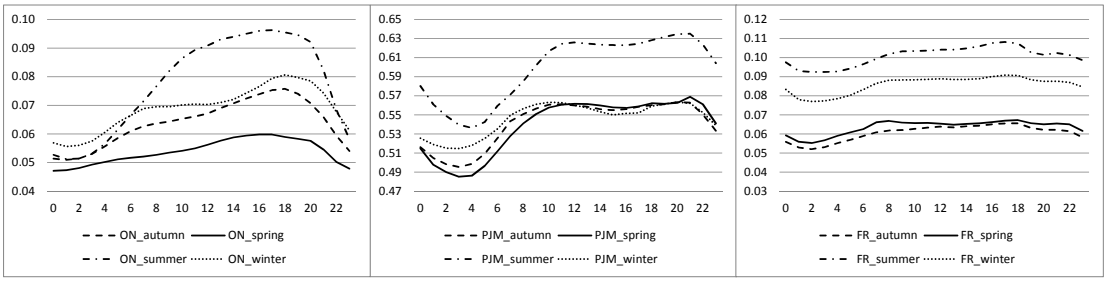


Fig. 1. Average emission factors (gCO_2e/Wh) during the day divided by season

Finally, the output of the cells $\mathbf{h}_t$, Equation 8, is controlled by the *output gate* vector $\mathbf{o}_t$ to determine if the value $\mathbf{c}_t$ needs to be transmitted to the next layer. The $tanh$ is used to limit the range value between -1 and 1.

$$\mathbf{h}_t = \mathbf{o}_t \circ tanh(\mathbf{c}_t) \tag{8}$$

All equations listed before represent one layer of multiple LSTM cells. All of these equations are differentiable allowing to use any gradient descent algorithm to train the neural network model using LSTM cells as neurons. LSTM systems have been achieving good performance in different time series problems, e.g., speech recognition, language modeling, visual recognition, and description [42], [43], [44], [45].

5.1 Dataset Organization

To create the dataset for the LSTM network experiment, we choose to use the last 28 days before the day to predict with a period of 1 hour, $\Delta t = 1$ hour. This 4-week range was defined based on a preliminary analysis of the datasets, as well as based on some tests with different time ranges during the definition of the prediction model. The RNN architecture used is many-to-many, i.e., for each input x_t we get an output y_t.

The time series data is then defined $\mathbf{X} = [x_{-672}, ..., x_t, ..., x_{24}]$ where $t = 0$ is the last known hour. x_t is the input vector given to the network at each time t. x_{-672} is then the input vector 28 days before and x_0 is the input vector just before the 24 hours to predict. The predicted values model output $\mathbf{Y} = [y_{-672}, ..., y_t, ..., y_{24}]$ represents the emission factor predicted for all time t. In our case, we only need the future part of the output is $\mathbf{Y}_{pred} = [y_1, ..., y_{24}]$.

The result when using only the past emission factors as input feature in the model was not at the level of what we were expecting for the datasets with a bigger share of renewables. This might be explained due to the intermittency of such energy sources. In Section 6 we detail more this observation. To improve the emission factor prediction, we decided to leverage on other forecast data provided by the utilities. This day-ahead information can comprise the forecasted demand in the region, generation from renewables (divided by source or aggregated), and generation from non-renewable sources.

Figure 4 shows the actual structure of the dataset used for the experiment for Ontario as an example. For each time t, the network receives an input vector $x_t = [e, d, s, w, m]$ of dimension 5 corresponding to: emission factor, demand, solar production, wind production and the mask value, respectively. The mask m helps the model to differentiate the known past from the unknown future information. When $m = 1$, all features are past information considered known. When $m = 0$ the model understands that the value $e = 0$ should not be taken into consideration and the model needs to predict it from memory and forecast data at this step.

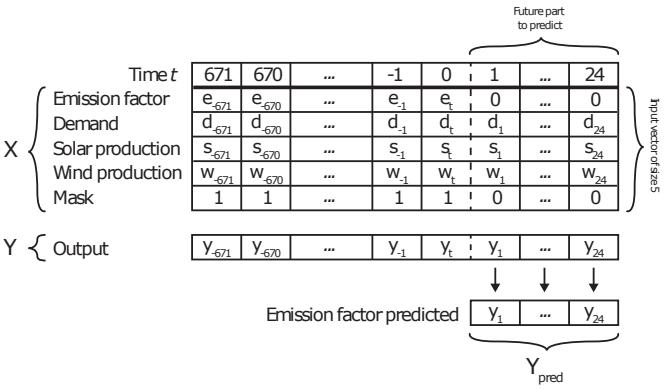


Fig. 4. Dataset structure for the experiment with demand, wind and solar source from IESO.

The process to predict the next 24 hours is to feed the model with the input vector from time $t = -671$ to $t = 24$. The outputs y_1 to y_{24} are the prediction of the emission factor for the next 24 hours.

The training and the experiment were performed using Keras version 2.0.7 with Tensorflow backend engine version

1.2.1 in Python 2.7. The computer OS is Ubuntu 16.04 with a processor i7-4790 and 16 GB of RAM.

The model used is a three-layer RNN: the first layer consists of 22 LSTM cells, the second layer consists of 16 LSTM cells and the output layer consists of 1 linear neuron as shown in Figure 5. Different model architectures have been tested during the research e.g. number of layers and numbers of cells per layer. To create more features, we tried to add a convolution layer before LSTM. We also tested the model with dropout and random noise in data to reduce overfitting, but the results did not improve for any of these configurations. The final model hyperparameters have been found using grid search technique.

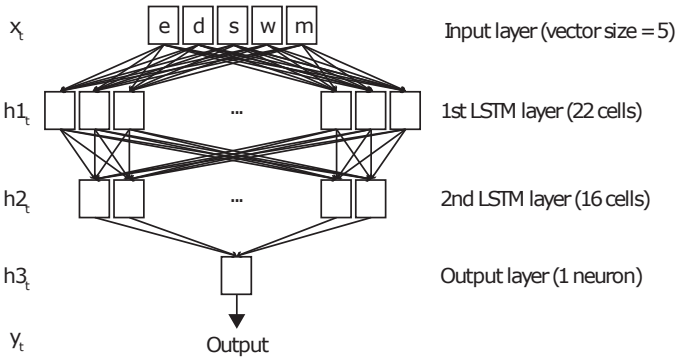


Fig. 5. Dataset structure for the experiment with demand, wind and solar forecasts from Ontario.

The gradient descent optimizer used during the training is the Root Mean Square Propagation (RMSProp). The learning rate adjustment follows the Equation 9 proposed by [46]. The partial warm restart, consisting of restarting the learning rate to the initial value, improves the performance and the convergence rate during the training in our experiment. By doing that, in some circumstances, the gradient descent algorithm will reach a new lower local minimum faster if available in the range.

$$\eta_t = \eta_{min} + \frac{1}{2}(\eta^i_{max} - \eta^i_{min})(1 + cos(\frac{T_{cur}}{T_i}\pi)) \quad (9)$$

η^i_{min} and η^i_{max} are the minimum and the maximum learning rates for the i_{th} run. T_{cur} represents the current epoch since the last restart. T_i is the number of epochs of the next restart. For this paper, we tested multiple sets of parameters and selected $\eta_{min} = 0.0002$ and $\eta_{max} = 0.01$. The $\eta_{min,max}$ are constant for all runs, but T_i doubles each run with $T_0 = 10$. The parameters of Equation 9 have been found by grid search in preliminary experiments.

The experiment follows the Algorithm 1. For the cross-validation, we used Scikit's $cross\_val\_score$ over the entire dataset, using $KFold$ with 10 folds. For each run, we trained the model over 639 epochs. This value comes from the Equation 9, where the new sequence $i = 7$ starts at the 640th epoch. Furthermore, after the 6th sequence, the validation performance stops increasing.

6 Results and Discussion

For the evaluation of different datasets, we used the Mean Absolute Percentage Error (MAPE), a commonly used metric for electricity load forecasting [29] as well in time series prediction in general. This metric evaluates the error between the real and predicted values and is scale-independent. Equation 10 describes MAPE.

Data: $(\mathbf{X}, y)$

1. Split the data $(\mathbf{X}, \mathbf{Y})$ into 10 equal sized folds.
2. Select the first fold as test dataset and all others to train the RNN model
   - 9 folds to train RNN model
   - 1 fold to test the trained model
3. Mix the 9 folds and select 20% for the validation dataset and 80% for the training dataset
4. Train and validate the RNN network over 639 epochs
5. Predict the output values (emission factors) of the test data
6. Perform **cross-validation** step 2 to 5 by changing the test fold and evaluate the average score over all data
7. Compare the results with other algorithms

Algorithm 1: Predicting the emission factor for the 24 next hours in a given region.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y^a_t - y^f_t}{y^a_t}\right| \quad (10)$$

n is the total number of forecasted values, y^a_t and y^f_t are the actual and predicted emission factors of the t-th hour, respectively.

We also used Pearson's Correlation in order to check how much the real values for the day are correlated to the predicted values. To reduce GHG emissions by scheduling smart appliances, the predicted minimum hour has to correspond to the real minimum hour. We are then interested in positive correlations - when the real values increase, so do the predicted values. Equation 11 describes how the Pearson's correlation coefficient is calculated [47].

$$r_{pearson} = \frac{\sum(y^a_t - m_{y^a})(y^f_t - m_{y^f})}{\sqrt{\sum(y^a_t - m_{y^a})^2(y^f_t - m_{y^f})^2}} \quad (11)$$

y^a_t and y^f_t are the actual and predicted emission factor values of the t-th hour, respectively. m_{y^a} and m_{y^f} are the mean of actual and predicted emission factor values.

As mentioned in Section 5, the results using just the historical emission factors as inputs were improved using additional day-ahead information provided by the utility (forecasted demand, generation from renewable, or generation from non-renewables). The additional information can be whichever is available for download, which varies depending on the region under study. The results improved significantly, with the MAPE reducing around 50% for Ontario and 30% for France. For PJM, the results also improved, but the original numbers were already good, since the region mostly used non-renewable energy, and, therefore, presents less intermittency in the dataset. Table 3 lists the metrics for all studied regions, as well as divided by season. For MAPE, a lower result indicates a better result. For

Pearson's Correlation, 1 indicates a perfect correlation between real and predicted values and -1 indicates an inverse correlation.

6.1 Comparison with a time series approach

For validation, we compared the results of LSTM with TBATS [48] (Trigonometric, Box-Cox transform, ARMA errors, Trend, and Seasonal components), a state space model, convenient for multiple seasonal time series [28]. Despite supporting only univariate time series, it is an automated algorithm which allows for automatic box-cox transformation and ARMA errors [28].

The TBATS parameters are ω (Box-Cox parameter, with the value 1 meaning no Box-Cox transformation), ϕ (damping parameter for de-trending), (p, q) (ARMA parameters), and the pairs (m_x, k_x) for the seasonal periods and the corresponding Fourier terms to fit this cycle. [28] gives the following example $TBATS(0.999, \{2, 2\}, 1, \{< 52.18, 8 >\})$, meaning a 0.999 box-cox transformation (essentially doing nothing), ARMA (2,2) errors, a damping parameter of 1 (doing nothing) and a set of 8 Fourier pairs with period $m = 52.18$.

Following the usual out-of-sample approach and using the function $tbats$ of R-programming language, we observed the results described on Table 4. For PJM, with mostly programmable energy sources, the results are close to the real values, but when more renewables are present, the prediction gets worse. For some days we also noticed some seasonality changes, and in other cases, the prediction either did not converge or got constant results for all hours of the day. LSTM did not present such issues for the studied cases, being more generalizable and easier to use for our purposes.

6.2 Comparison with a supervised learning approach

To confirm the effectiveness of the proposed solution, we also compared the results of LSTM with a supervised learning model, SVR (Support Vector Regression) first identified by [49] following a time series to supervised approach. We selected SVR based on the related work [32], which dealt with marginal emissions. For that, we performed feature engineering by converting the univariate dataset of emission factor values so that the lagged observations up to 28 days (times 24 hours) in the past are features (X) and the current observation is the output (y). We then run 24 models corresponding to the next 24 hours to be predicted. In order to benefit from other forecasted values provided by the operators, as we did for LSTM, we also tried adding to the features vector the predicted values for the demand, wind, and solar production for the next hour, but the results did not improve significantly.

For the experiments, we used the $python$ language package $scikit\ learn$ [50]. A polynomial kernel with degree 2 was used and, by performing grid search, we selected C (the penalty parameter of the error term) as 0.001 and ϵ (a margin of tolerance) as 0.1. For the other parameters, the default values were used. We also differentiated the series for stationarity.

The results are described on Table 5. As for the other experiments, PJM has the results closest to the real values, however when more renewables are present, as is the case for Ontario and France regions, the predictions got worse. Despite running faster during the training phase, using this approach demanded some further evaluation on the time series and did not obtain better results than LSTM.

TABLE 3
LSTM results for the different regions

	ON		PJM		France	
season	**MAPE**	$r_{pearson}$	**MAPE**	$r_{pearson}$	**MAPE**	$r_{pearson}$
ALL	12.44	0.52	2.10	0.90	11.30	0.50
spring	11.83	0.32	2.19	0.91	13.42	0.53
summer	12.39	0.71	1.87	0.96	12.49	0.55
autumn	13.00	0.54	2.14	0.88	9.76	0.50
winter	12.61	0.51	2.28	0.83	9.14	0.49

TABLE 4
TBATS results for the different regions

	ON		PJM		France	
season	**MAPE**	$r_{pearson}$	**MAPE**	$r_{pearson}$	**MAPE**	$r_{pearson}$
ALL	19.07	0.40	3.15	0.85	14.70	0.46
spring	14.70	0.23	3.41	0.86	17.68	0.51
summer	20.70	0.63	2.60	0.95	16.77	0.44
autumn	22.48	0.41	3.34	0.83	11.69	0.43
winter	18.45	0.32	3.47	0.72	12.11	0.46

TABLE 5
SVR results for the different regions

	ON		PJM		France	
season	**MAPE**	$r_{pearson}$	**MAPE**	$r_{pearson}$	**MAPE**	$r_{pearson}$
ALL	20.24	0.38	3.09	0.85	15.00	0.49
spring	15.46	0.25	3.28	0.86	17.22	0.54
summer	22.95	0.58	2.74	0.95	16.71	0.51
autumn	22.90	0.36	3.06	0.84	13.79	0.39
winter	19.65	0.30	3.46	0.72	11.71	0.50

6.3 Discussions

As expected, some seasons or days are better than others. Figure 6 brings some examples for the different datasets. For the PJM dataset, a region mostly powered by fossil fuels, therefore more controllable, the emission factors are more dependent on the demand itself. The results are good both through using LSTM and the other approaches, with LSTM being more robust, without the convergence problems that sometimes happen to TBATS and without any additional required analysis on stationarity. We also tested for internal purposes this fossil fuels assumption on a dataset from the province of Alberta (Canada). The results obtained were similar to the PJM region, but not finally retained in our works, due to the small amount of data available at the time of this study.

As expected, more intermittency meant more difficult in predicting. Figure 7 shows the emission factor curve with a margin of 12.44% representing the average MAPE for the region of Ontario. For this graph, July 17 of 2015 is depicted. For this region, we achieved better results comparing LSTM with the other approaches, but, recently, the province started releasing their own 24-hour forecast per source. This dataset, called "Accuracy Report", in conjunction with the predicted imports, can give a more accurate result than using our tool. This report is updated several times per day, so, for the applications discussed in the next Section, we suggest the user to get the information which is closest to midnight for the next day.

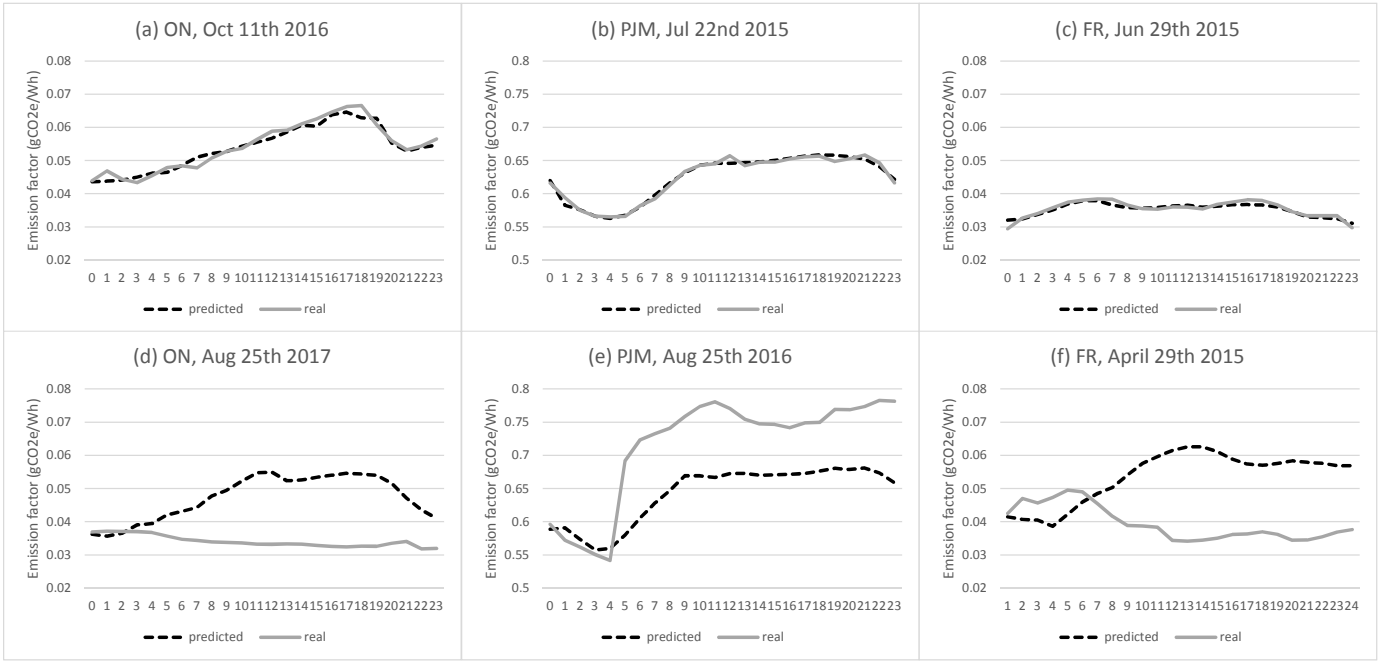


Fig. 6. Graphs comparing the LSTM predicted and true values: (a) (b) (c) represent good approximations, while (d) (e) (f) represent the opposite

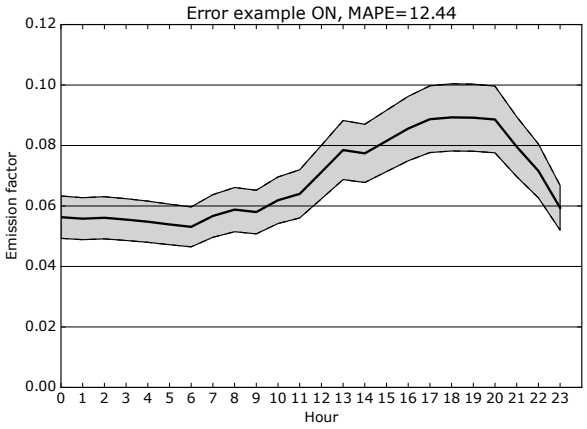


Fig. 7. Graph showing emission factor with a MAPE of 12.44% on July 17, 2015

However, even with this nice dataset now disclosed, other regions do not have such information available yet, which proves the utility of our tool. France, for instance, already reports carbon emissions (not using LCA emission factors for the sources, but this is another discussion), but there is no day-ahead data. One difficulty encountered in the process for this region was that the information about imports is reported as a total, not being divided by neighbor (divided by neighbor are just the commercial agreements, not the actual, physical flows). Therefore, we needed to assume an average emission factor for the imports. Since France is, most of the time, a net exporter, that may be not a big issue for this use case. However, it might be for other regions or even in the moments when France is importing from a region with much higher emission factor, like Germany [21]. Regarding the results, LSTM was also better than the other approaches and, as mentioned, SVR demanded an additional analysis on stationarity and TBATS was not always able to predict the emission factors. For this reason, we argue that LSTM might be more generalizable. Another aspect that could explain the better results obtained by LSTM neural network is its robustness when modeling non-linear time series data when compared to the other linear predictor since emission factor generated by electricity production is generally a non-linear function.

In general, the task of predicting emission factors is not easy. The data is, of course, related to the demand and pricing dynamics in the regions, but it is not totally correlated, demanding its own predictors. Emission factor forecasts usually present worse results on Fridays and during weekends, probably due to reduced users' routines during these days. Another aspect to point out is the difficulty the algorithm might have when the weather drastically changes from one day to the other. In the example of Ontario on August 25, 2017, and France on April 27, 2015, shown in the Figure 7 d) and f), the temperature dropped by 6 °C compared with the prior day. The demand also decreased following the temperature. We suppose air conditioning (AC) might be playing a role in the demand decrease. For the case of Ontario, the province might be using natural gas as a source to meet this energy peaks during these afternoons. Since August 25, 2017, was a cooler day, the demand for the afternoon was lower than previous days, and probably the extra energy generation using natural gas was not necessary. The same happened for the French example, but with coal. The emission factor prediction based on the previous days was, therefore, higher than the real values. The emission factor prediction might be improved with additional sources, like weather forecasts, which are somehow already inside the day-ahead information we used, but maybe using more "explicitly" might yield better

results. It would be interesting to figure out a way to account for such outliers days.

7 Scheduling the Smart Home Devices

The proposed day-ahead GHG emissions predicted was applied to two case studies: one in which the best time of the day is used to decide when to start a dishwasher, and other inside an optimization problem, in conjunction with other variables to decide the optimal time to charge an electric vehicle (EV). Note that both loads are flexible and can be programmed to start when convenient, without significantly affecting users' behavior. Both use cases are currently running in our smart home prototype, controlled by a Raspberry Pi 3 running OpenHab, an open source home automation software. In a real environment, our approach could be readily applied using, for instance, a smart plug. In the near future, smart appliances can also incorporate such functionality.

7.1 Using the predicted best time of the day

In this first use case, we assume the user programs the dishwasher to work during the hours he/she is out of home on the next day, for instance, from 8 a.m. to 5 p.m. Without this, the user could just turn on the dishwasher as soon as it is loaded, after dinner for example. In order to illustrate the advantage of programming to work on the hour with minimum emissions, we bring the example of Ontario on September 23rd, 2017. The predicted best hour for the period between 8 a.m. and 5 p.m. was 8 a.m. The emission factor for 8 a.m. during this day was $0.11gCO_2/Wh$, and the emission factor right after dinner time $0.14gCO_2/Wh$.

The difference between these values vary during the year, and can be even more significant. It is also important to note that, for the locations studied, the hours with minimum emissions are usually during the first morning hours, but that might not always be the case and/or the user might want to schedule the appliance to work during the day hours, and the minimum can take place at another moment, like exemplified in Figure 8. Note also that the total electricity consumed does not change.

Considering a usual power profile for a dishwasher[3], as presented in Figure 9, the difference between running the dishwasher in the minimum GHG emissions time compared to the peak emissions time on the date and location used as reference can be close to $30g$ of CO_2e per day.

3. We obtained a power profile from the Tracebase https://www.tracebase.org/. The used file is the following: "$dev\_B7E6FA\_2011.11.27.csv$"

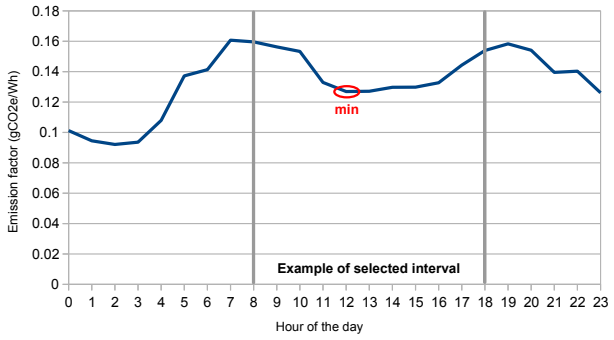


Fig. 8. Emission factors during February 10th, 2015 in Ontario

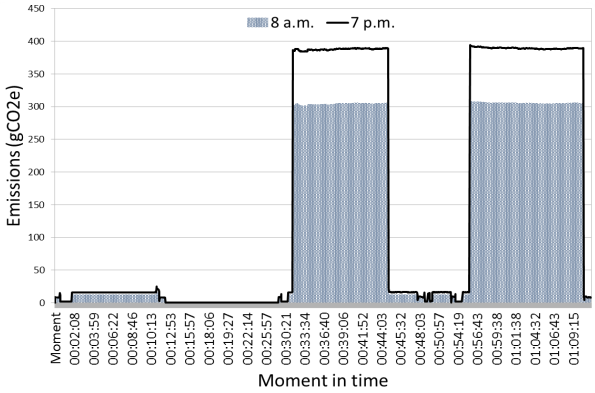


Fig. 9. Dishwasher cycle emissions at 8 a.m. (min. emission factor) and at 7 p.m. (assuming the user would turn on the dishwasher right after dinner time) in Sep. 23rd, 2017

7.2 Using the best time of the day information in an optimization problem

In this second use case, we replicated and extended the GreenCharge proposal from [51] [15] in order to determine the best time to charge an EV using an optimization approach. In this case, we are considering not only (i) the best time of the day from the GHG emissions point of view, but also (ii) the predicted home energy demand during the day; (iii) the local energy generation from a solar panel and (iv) the storage of the local energy production in a home battery. Our objective is to minimize the GHG emissions over the day by deciding when the EV will be charged. The original objective function [51] is represented in Equation 12.

$$Minimize \sum_{i=1}^{T}(p_i + s_i - d_i) * I * c_i \quad (12)$$

where i represents the periods of the day of length I up to T periods (24); p_i is the power consumed from the grid in interval i, s_i is the power charged to the home battery from the grid; $d_i$is the power discharged from the home battery; and c_i is the cost of buying electricity from the grid. On our example, we consider c_i represents the GHG emissions, not price. We also incorporated the EV charging extra demand and the determination of the hours to perform the charging using a binary variable $x_i \in \{0, 1\} \forall i \in H$, where H represents the allowed hours to charge given the time the EV is parked and when it will be used again. While GreenCharge [51] has a linear programming formulation, this extension became a Mixed Integer Linear Programming (MILP) due to the variable used to determine the hour the charging will occur.

For the item (i) above, we selected the province of Ontario and its predicted emission factors for September 23rd, 2017. To represent the predicted home energy demand (item ii), we used data from the Smart* Data Set, also from UMass Amherst, as the GreenCharge [52] [53]. For our use case, we selected the Home B, Meter 1. We took the data from September 23rd, 2016 and calculated the average use for each hour. We assume the same demand curve happened

on this same day of 2017. The EV charging extra demand was calculated by using Tesla's website information and assuming 70 km a day of driving demand, which means 2 hours charging using the NEMA 14-50 (max. 7.7 kWh) [54].

For the energy locally produced by renewables (item iii), we used data from NREL's PVWatts Calculator [55] [56], which estimates the energy production of grid-connected photovoltaic (PV) energy systems. To consider a city in the province of Ontario and, at the same time, consider a place with a close insulation factor as Amherst (for which we have a sample home energy demand), we selected the city of London, ON with the default parameters in this calculator. For the home battery (item iv), since we are replicating [51], and the selected dataset has the same annual average energy demand of 1 kWh, we used the same values as used in [51].

Figure 10 shows the results of this experiment. In this figure, we present as lines the electricity consumed from the grid to charge the home battery and the electricity generated from the renewable source (in this case, the solar panels) charged to the battery. We also present as bars the electricity the house draws from the grid and from the home battery to meet the usual and the additional demand from the EV charging. Note that the EV is being charged when the emission factors are the smallest, between 2 a.m. and 3 a.m. In this same period, the home battery is also being charged, in order to minimize the total emissions. The battery is also charged using solar energy production during the day and is supposed to start and finish the day discharged. Charging at the selected time of the day, at 7.7kWh would mean $770gCO_2e$ emitted against $1232gCO_2e$ if the charging starts as soon as the vehicle arrives home in the evening.

8 Conclusion and Future Work Directions

In this work, we presented a method to predict day-ahead GHG emissions using LSTM. Compared with other traditional approaches, the method achieved better results and proved to be easier to use and more generalizable. We applied the information predicted in two use cases: one simple, in which we used the predicted emission factors to run a dishwasher in the time of the day with minimum emissions, and one more complex scenario which extended the work from [15], to charge an EV. In both use cases we were able to demonstrate GHG emissions reduction in examples that would not significantly change consumers' behavior.

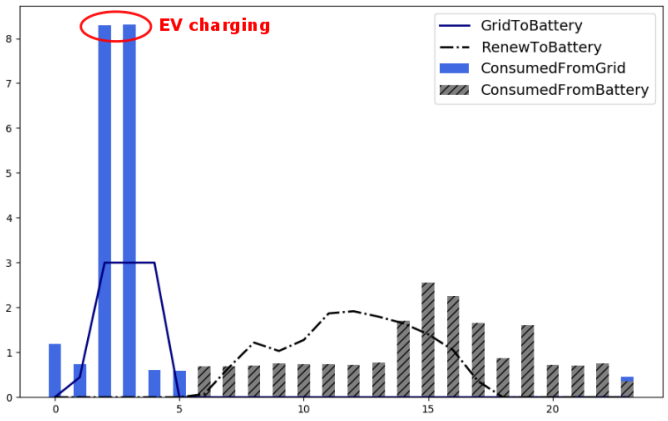


Fig. 10. Optimization result example: the best time to charge an EV

Many newer appliances already have a timer as functionality, allowing the end user to choose the best moment to start [7]. With the upcoming technologies to control appliances remotely, as well as smart grids, more advanced techniques and integration scenarios will be possible. The optimization example showed can also help to handle, at least for some time, the extra load expected from the increasing demand, like for charging EVs. The prediction of GHG emissions can also be used to provide information for different applications in buildings or cloud computing infrastructures. Pricing information may also be associated in the optimization formulation, bringing more value to the customers.

Due to the complexity of the electricity market, as pointed out by [39] for pricing data, the forecast over GHG emissions data might not be able to always produce accurate forecasts. Another limitation is related to the uncertainties and the precision of the emission factors, inherent to LCA analysis. Another important point to consider is the future effect of using such information on the best hours to schedule loads: if the usage spreads, the demand may flatten, and the variable energy generation will flatten as well. Which might be good from the point of view of the utility, with the reduction of peaks, may also reduce the incentives for performing this kind of prediction and scheduling. Approaches considering whole buildings or communities might help with this issue.

During the experiments, we also noticed that the days had different profiles - some had a very low profile, with small emission factors throughout the day, while others have higher emissions factors and/or present high variability during the day. It is possible that a clustering technique could give hints to the predictor and improve the results.

Moreover, one issue noticed during our studies is about finding data. Some regions disclose all information, some regions have just real-time data available, others do not have data about the exchanges with neighbors and others do not disclose information at all. We believe that, considering the increasing demand, applications, and standardization efforts, more datasets will be made available.

Acknowledgments

The authors also thank NSERC and Ericsson for funding the project CRDPJ 469977. This research also receives support from the Canada Research Chair, Tier 1, held by Mohamed Cheriet. The authors thank Rachid Hedjam for the insights in the beginning of the studies. We also thank dr. Mishra for the technical insights during the replication of his work.

References

[1] Y. Geng, W. Chen, Z. Liu, A. S. Chiu, W. Han, Z. Liu, S. Zhong, Y. Qian, W. You, and X. Cui, "A bibliometric review: Energy consumption and greenhouse gas emissions in the residential sector," *Journal of Cleaner Production*, vol. 159, pp. 301 – 316, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0959652617310259

[2] Natural Resources Canada, "Energy Use Data Handbook Tables - Total End-Use Sector," accessed: 28 Feb. 2018. [Online]. Available: http://oee.nrcan.gc.ca/corporate/statistics/neud/dpa/menus/trends/handbook/handbook_aaa_ca.cfm

[3] A. S. Al-Sumaiti, M. H. Ahmed, and M. M. A. Salama, "Smart Home Activities: A Literature Review," *Electric Power Components and Systems*, vol. 42, no. 3-4, pp. 294–305, 2014. [Online]. Available: http://dx.doi.org/10.1080/15325008.2013.832439

[4] G. Mehdi and M. Roshchin, "Electricity consumption constraints for smart-home automation: An overview of models and applications," *Energy Procedia*, vol. 83, pp. 60 – 68, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1876610215028611

[5] B. Dean, J. Dulac, K. Petrichenko, and P. Graham, "Towards zero-emission efficient and resilient buildings.: Global status report," 2016.

[6] G. e Sustainability Initiative *et al.*, "SMARTer2030, ICT Solutions for 21st Century Challenges," *Belgium, GeSI, Accenture Strategy*, 2015.

[7] M. Kopsakangas-Savolainen, M. K. Mattinen, K. Manninen, and A. Nissinen, "Hourly-based greenhouse gas emissions of electricity cases demonstrating possibilities for households and companies to decrease their emissions," *Journal of Cleaner Production*, vol. 153, pp. 384 – 396, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S095965261501656X

[8] "Carbon Intensity API," accessed: 02 Feb. 2018. [Online]. Available: http://carbonintensity.org.uk/

[9] "IESO Website," accessed: 06 Jan. 2017. [Online]. Available: http://www.ieso.ca

[10] "RTE France - Eco2mix tool," accessed: 28 Fev. 2018. [Online]. Available: http://www.rte-france.com/fr/eco2mix/eco2mix-co2

[11] A. C. Riekstin, R. Hedjamt, T. Dandrest, and M. Cheriet, "Statistical-based method to determine the best hour of the day regarding GHG emissions for a smart home appliance," in *2017 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computed, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, Aug 2017, pp. 1–8.

[12] M. Jahn, M. Jentsch, C. R. Prause, F. Pramudianto, A. Al-Akkad, and R. Reiners, "The energy aware smart home," in *Future Information Technology (FutureTech), 2010 5th International Conference on*. IEEE, 2010, pp. 1–8.

[13] A. Fensel, S. Tomic, V. Kumar, M. Stefanovic, S. V. Aleshin, and D. O. Novikov, "SESAME-S: Semantic Smart Home System for Energy Efficiency," *Informatik Spektrum*, vol. 36, no. 1, pp. 46–57, 2013.

[14] B. Zhou, W. Li, K. W. Chan, Y. Cao, Y. Kuang, X. Liu, and X. Wang, "Smart home energy management systems: Concept, configurations, and scheduling strategies," *Renewable and Sustainable Energy Reviews*, vol. 61, no. Supplement C, pp. 30 – 40, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364032116002823

[15] A. K. Mishra, "Energy Optimizations for Smart Buildings and Smart Grids," Ph.D. dissertation, UMass Amherst, May 2014.

[16] M. Rahmani-andebili and H. Shen, "Energy Scheduling for a Smart Home Applying Stochastic Model Predictive Control," in *2016 25th International Conference on Computer Communication and Networks (ICCCN)*, Aug 2016, pp. 1–6.

[17] Y.-M. Wi, J.-U. Lee, and S.-K. Joo, "Electric vehicle charging method for smart homes/buildings with a photovoltaic system," *IEEE Transactions on Consumer Electronics*, vol. 59, no. 2, pp. 323–328, 2013.

[18] L. Liu, F. Kong, X. Liu, Y. Peng, and Q. Wang, "A review on electric vehicles interacting with renewable energy in smart grid," *Renewable and Sustainable Energy Reviews*, vol. 51, pp. 648–661, 2015.

[19] T. Hubert and S. Grijalva, "Modeling for residential electricity optimization in dynamic pricing environments," *IEEE Transactions on Smart Grid*, vol. 3, no. 4, pp. 2224–2231, 2012.

[20] J. Abushnaf, A. Rassau, and W. Grnisiewicz, "Impact of dynamic energy pricing schemes on a novel multi-user home energy management system," *Electric Power Systems Research*, vol. 125, no. Supplement C, pp. 124 – 132, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S037877961500108X

[21] A. Milovanoff, T. Dandres, C. Gaudreault, M. Cheriet, and R. Samson, "Real-time environmental assessment of electricity use: a tool for sustainable demand-side management programs," *The International Journal of Life Cycle Assessment*, Dec 2017. [Online]. Available: https://doi.org/10.1007/s11367-017-1428-2

[22] P. Stoll, G. Bag, J. E. Y. Rosseb, L. Rizvanovic, and M. kerholm, "Scheduling residential electric loads for green house gas reductions," in *2011 2nd IEEE PES International Conference and Exhibition on Innovative Smart Grid Technologies*, Dec 2011, pp. 1–8.

[23] I. P. Panapakidis and A. S. Dagoumas, "Day-ahead electricity price forecasting via the application of artificial neural network based models," *Applied Energy*, vol. 172, pp. 132 – 151, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306261916304160

[24] J. Hinman and E. Hickey, "Modeling and forecasting short-term electricity load using regression analysis," *Journal of IInstitute for Regulatory Policy Studies [ ]*, 2009.

[25] C. H. Jin, G. Pok, I. Paik, and K. H. Ryu, "Short-term electricity load and price forecasting based on clustering and next symbol prediction," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 10, no. 2, pp. 175–180, 2015. [Online]. Available: http://dx.doi.org/10.1002/tee.22050

[26] D. Singhal and K. Swarup, "Electricity price forecasting using artificial neural networks," *International Journal of Electrical Power and Energy Systems*, vol. 33, no. 3, pp. 550 – 555, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0142061510002231

[27] J. Contreras, R. Espinola, F. J. Nogales, and A. J. Conejo, "ARIMA models to predict next-day electricity prices," *IEEE Transactions on Power Systems*, vol. 18, no. 3, pp. 1014–1020, Aug 2003.

[28] R. J. Hyndman and G. Athanasopoulos, "Vector autoregressions," in *Forecasting: principles and practice*. Melbourne, Australia: OTexts, 2013, ch. Section 9.2.

[29] E. Raviv, K. E. Bouwman, and D. van Dijk, "Forecasting day-ahead electricity prices: Utilizing hourly prices," *Energy Economics*, vol. 50, pp. 227–239, 2015.

[30] G. Dudek, "Multivariate Regression Tree for Pattern-Based Forecasting Time Series with Multiple Seasonal Cycles," in *International Conference on Information Systems Architecture and Technology*. Springer, 2017, pp. 85–94.

[31] N. Mahmoudi-Kohan, M. P. Moghaddam, and M. Sheikh-El-Eslami, "An annual framework for clustering-based pricing for an electricity retailer," *Electric Power Systems Research*, vol. 80, no. 9, pp. 1042–1048, 2010.

[32] C. Wang, Y. Wang, C. J. Miller, and J. Lin, "Estimating hourly marginal emission in real time for PJM market area using a machine learning approach," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, July 2016, pp. 1–5.

[33] N. A. Ryan, J. X. Johnson, and G. A. Keoleian, "Comparative Assessment of Models and Methods To Calculate Grid Electricity Emissions," *Environmental Science & Technology*, vol. 50, no. 17, pp. 8937–8953, 2016, pMID: 27499211. [Online]. Available: http://dx.doi.org/10.1021/acs.est.5b05216

[34] T. Dandres, R. F. Moghaddam, K. K. Nguyen, Y. Lemieux, R. Samson, and M. Cheriet, "Consideration of marginal electricity in real-time minimization of distributed data centre emissions," *Journal of Cleaner Production*, vol. 143, pp. 116 – 124, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0959652616322065

[35] V. Tikka, J. Lassila, J. Haakana, and J. Partanen, "Electric vehicle smart charging aims for CO_2 emission reduction?" in *2016 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, Oct 2016, pp. 1–6.

[36] E. Maurice, T. Dandres, R. Farrahi Moghaddam, K. Nguyen, Y. Lemieux, M. Cheriet, and R. Samson, "Modelling of Electricity Mix in Temporal Differentiated Life-Cycle-Assessment to Minimize Carbon Footprint of a Cloud Computing Service," in *ICT for Sustainability 2014 (ICT4S-14)*. Atlantis Press, 2014.

[37] "PJM Data Miner 2," accessed: 02 Feb. 2018. [Online]. Available: http://www.pjm.com/markets-and-operations/etools/data-miner.aspx

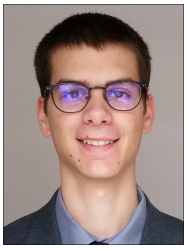
[38] H. Freeman, J. M. H. Elmirghani, and C. Despins, "The President's Page," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 4–6, 2017.

[39] M. B. Amor, E. B. de Villemeur, M. Pellat, and P.-O. Pineau, "Influence of wind power on hourly electricity prices and ghg (greenhouse gas) emissions: Evidence that congestion matters from ontario zonal data," *Energy*, vol. 66, pp. 458 – 469, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360544214000814

[40] B. Koffi, A. K. Cerutti, M. Duerr, A. Iancu, A. Kona, and G. Janssens-Maenhout, "Covenant of Mayors for Climate and Energy: Default emission factors for local emission inventories," *Joint Research Centre (JRC)*, 2017, accessed: 28 Mar. 2018. [Online]. Available: http://publications.jrc.ec.europa.eu/

repository/bitstream/JRC107518/jrc_technical_reports_-_com_default_emission_factors-2017.pdf
[41] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997. [Online]. Available: https://doi.org/10.1162/neco.1997.9.8.1735
[42] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional lstm and other neural network architectures," *Neural Networks*, vol. 18, no. 5, pp. 602 – 610, 2005, iJCNN 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0893608005001206
[43] A. Graves, A. r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 6645–6649.
[44] M. Sundermeyer, R. Schlüter, and H. Ney, "LSTM neural networks for language modeling," in *Thirteenth Annual Conference of the International Speech Communication Association*, 2012.
[45] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 2625–2634.
[46] I. Loshchilov and F. Hutter, "SGDR: stochastic gradient descent with restarts," *CoRR*, vol. abs/1608.03983, 2016. [Online]. Available: http://arxiv.org/abs/1608.03983
[47] E. Jones, T. Oliphant, P. Peterson *et al.*, "SciPy: Open source scientific tools for Python," 2001–, [Online; 16 Mar. 2018]. [Online]. Available: http://www.scipy.org/
[48] A. M. De Livera, R. J. Hyndman, and R. D. Snyder, "Forecasting time series with complex seasonal patterns using exponential smoothing," *Journal of the American Statistical Association*, vol. 106, no. 496, pp. 1513–1527, 2011.
[49] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Berlin, Heidelberg: Springer-Verlag, 1995.
[50] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011.
[51] A. Mishra, D. Irwin, P. Shenoy, J. Kurose, and T. Zhu, "GreenCharge: Managing RenewableEnergy in Smart Buildings," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 7, pp. 1281–1293, July 2013.
[52] S. Barker, A. Mishra, D. Irwin, E. Cecchet, P. Shenoy, and J. Albrecht, "Smart*: An open data set and tools for enabling research in sustainable homes," *SustKDD, August*, vol. 111, p. 112, 2012.
[53] U. T. Repository, "Smart* Data Set for Sustainability," accessed: 05 Oct. 2017. [Online]. Available: http://traces.cs.umass.edu/index.php/Smart/Smart
[54] Tesla, "Home Charging Calculator," accessed: 05 Oct. 2017. [Online]. Available: https://www.tesla.com/en_CA/support/home-charging-calculator
[55] B. Marion and M. Anderberg, "PVWATTS-an online performance calculator for grid-connected PV systems," in *Proceedings of the solar conference*. AMERICAN SOLAR ENERGY SOCIETY; AMERICAN INSTITUTE OF ARCHITECTS, 2000, pp. 119–124.
[56] NREL, "PVWatts Calculator," accessed: 05 Oct. 2017. [Online]. Available: http://pvwatts.nrel.gov/

Ana Carolina Riekstin Post-doctoral Fellow at Synchromedia Laboratory, École de Technologie Supérieure, Montreal. Received her PhD (2015) and her MSc (2012) in Computer Engineering from the Polytechnic School of University of São Paulo. Got her BSc in Computer Science (2007) from the Institute of Mathematical and Computer Sciences of the University of São Paulo. Worked previously at the Laboratory of Sustainability in ICT (LASSU) at USP, Univesp, Microsoft Research (internship), Volkswagen do Brasil, and PromonLogicalis. Her research interests are computer networks and network management, smart homes and cities, and sustainability.



Antoine Langevin Received his BSc in Electrical Engineering from the École de Technologie Supérieure, Montreal in 2016. He then started the graduate studies by joining the Synchromedia laboratory of the École de Technologie Supérieure. His current research project is on energy disaggregation using machine learning techniques. He also participates in the Smart ÉTS Residence project. His areas of interest in research are artificial intelligence, home/smart city, prototyping and embedded systems.

Thomas Dandres Research Officer at CIRAIG (Polytechnique Montréal), during the last 10 years he has been developing the life cycle assessment methodology in the context of energy systems to improve the modelling of the temporal dynamicity of power generation and the economic consequences of energy policies. He has a Ph.D. in life cycle assessment (2012, Polytechnique Montréal) and a Master in water treatment (2003, Polytechnique Montréal). Since 2015 he is also a member of the IEEE Green ICT Initiative where he participates in the development of new standards on the calculation of ICT emissions.

Ghyslain Gagnon Ghyslain Gagnon received the Ph.D. degree in electrical engineering from Carleton University, Canada in 2008. He is now an Associate Professor at cole de technologie supérieure, Montreal, Canada. He is a board member of ReSMiQ and Director of research laboratory LACIME, a group of 13 Professors and nearly 100 highlydedicated students and researchers in microelectronics, digital signal processing and wireless communications. Highly inclined towards research partnerships with industry, his research aims at digital signal processing and machine learning with various applications, from media art to building energy management.

Mohamed Cheriet Received M.Sc. and Ph.D. degrees in Computer Science from the University of Pierre & Marie Curie (Paris VI) in 1985 and 1988 respectively. Since 1992, he has been a professor in the Automation Engineering department at the École de Technologie Supérieure (University of Quebec), Montreal, and was appointed full Professor there in 1998. Prof. Cheriet is the founder and director of Synchromedia which targets multimedia communication in telepresence applications. Dr. Cheriet research has extensive experience in cloud computing and network virtualization and softwarisation. In addition, Dr. Cheriet is an expert in Computational Intelligence, Pattern Recognition, Machine Learning, and Perception. Dr. Cheriet has published more than 350 technical papers in the field. He serves on the editorial boards of several renowned journals and international conferences. As Tier 1 Canada Research Chair on Sustainable and Smart Eco-Cloud, he leads the establishment of the first smart university campus in Canada, created as a hub for innovation and productivity at Montreal. Dr. Cheriet is a Fellow of the International Association of Pattern Recognition (IAPR), a Fellow of Canadian Academy of Engineering (CAE), and the recipient of the 2016 IEEE J.M. Ham Outstanding Engineering Educator Award and of the 2012 Queen Elizabeth II Diamond Jubilee Medal. He is a senior member of the IEEE, the founder and former Chair of the IEEE Montreal Chapter of Computational Intelligent Systems (CIS), Steering Committee Member of the IEEE Green ICT Initiative, and Chair of IEEE ICT Emissions Working Group.