Transportation Research Part D 88 (2020) 102593

Contents lists available at ScienceDirect

Transportation Research Part D

journal homepage: www.elsevier.com/locate/trd

Greenhouse gas emission prediction on road network using deep sequence learning

Lama Alfaseeh [a], Ran Tu [b], Bilal Farooq [a,*], Marianne Hatzopoulou [b]

[a] Laboratory of Innovations in Transportation (LiTrans), Ryerson University, Canada
[b] Transportation and Air Quality Research Group, University of Toronto, Toronto, Canada

ARTICLE INFO

Keywords:
Greenhouse Gas (GHG) emissions
Carbon dioxide equivalent (CO_{2eq})
Machine learning
Clustering
Long-short term memory network (LSTM)
Autoregressive integrated moving average (ARIMA)
Link level GHG prediction

ABSTRACT

Mitigating the substantial undesirable impact of transportation systems on the environment is paramount. Thus, predicting Greenhouse Gas (GHG) emissions is one of the profound topics, especially with the emergence of intelligent transportation systems (ITS). We developed a deep learning framework to predict link-level GHG emission rate (ER) (in CO_{2eq} gram/second) based on the most representative predictors, such as speed, density, and GHG ER of previous time steps. In particular, various specifications of the long short-term memory (LSTM) networks with explanatory variables were examined, and were compared with clustering and the autoregressive integrated moving average (ARIMA) model with explanatory variables. The downtown Toronto road network was used as the study area, and highly detailed data were synthesized using a calibrated traffic microsimulation and MOVES. It was found that LSTM specification with speed, density, GHG ER, and in-links speed from three previous minutes performed the best while adopting two hidden layers, and when the hyper-parameters were systematically tuned. Adopting a 30-second updating interval slightly improved the correlation between true (simulated) and predicted GHG ERs (from predictive models), but contributed negatively to the prediction accuracy as reflected in the increased root mean square error (RMSE) value. Efficiently predicting GHG emissions at a higher frequency with lower data requirements will pave the way for various applications, e.g. anticipatory eco-routing in large-scale road networks to alleviate the adverse impact on global warming.

1. Introduction

Transportation systems have been consistently ranked as the largest source of GHG emissions (mainly carbon dioxide) in the U.S., producing 29% of the total emissions (United States Environmental Protection Agency, 2017). GHG emission is among the main contributors to global warming and climate change (Liu et al., 2020). For dynamic applications, such as anticipatory eco-routing, accurately predicted GHG link cost is urged. This can be achieved by utilizing high resolution data points as predictors, such as speed, density, and flow. It is worth mentioning that the relationship between CO_2 ER and speed is quasi-convex, as found in Djavadian et al. (2020) and Anas et al. (2009). In this case, ARIMA is not a suitable model for prediction when speed is a predictor.

To capture non-linear relationships between predictors and responses, artificial neural networks (ANNs) have been introduced. It has been shown that neural networks (NNs) outperformed other models when they were adopted to predict environmental pollutants (Singh et al., 2012). Clustering is another technique used for prediction in transportation systems; examples include

* Corresponding author.
E-mail addresses: lalfaseeh@ryerson.ca (L. Alfaseeh), ran.tu@mail.utoronto.ca (R. Tu), bilal.farooq@ryerson.ca (B. Farooq), marianne.hatzopoulou@utoronto.ca (M. Hatzopoulou).

https://doi.org/10.1016/j.trd.2020.102593

Available online 22 October 2020
1361-9209/

Poucin et al. (2018) and Tu et al. (2019b) and Gmira et al. (2017). In some cases, clustering was used in combination with other models, such as the MapReduce Framework, as in Zhao et al. (2011).

Unlike the aforementioned studies, which can be classified as top-down approaches, Dong et al. (2019) developed a predictive model for diesel trucks and gasoline passenger cars based on volume to capacity (V/C) ratio as an explanatory variable. Perugu et al. (2017) developed a high-resolution heavy-duty truck emission inventory for an urban case study. The authors estimated link-level total daily emissions by utilizing link truck miles travelled (TMT) and corresponding emissions factors for hourly temperatures and relative humidity (Perugu et al., 2017). For prediction applications, remote sensing, as in Hoogendoorn et al. (2003), has been proposed to collect high-resolution data points. A helicopter was used and a camera of very high standards regarding the resolution of the images as well as the frequency of collecting images was adopted (Hoogendoorn et al., 2003). Their case study included different sets of highways. However, the aforementioned tool can be challenging and expensive in the case of large urban networks. The use of nighttime light is another suggested approach to gather microscopic data points, as illustrated in Zhao et al. (2019).

While the existing studies in the literature forecast GHG emissions at an aggregated level (spatially and temporally) and generally utilize fuel and economical factors, this study aims to predict GHG ER (in CO_{2eq} g/s) at a link level using a deep learning framework, based on LSTM with explanatory variables, while using microscopic data points. When GHG emissions are estimated based on fuel consumption, vehicle kilometres travelled (VKT) is required (United States Environmental Protection Agency, 2020a). There are two conversion factors for this type of GHG estimation. One estimates the amount of consumed fuel based on the VKT. The other estimates the GHG produced based on the amount of consumed fuel (United States Environmental Protection Agency, 2020a). In this case, GHG emissions are associated with under- or overestimation due to the aggregated level of estimation. For estimating GHG emissions with a higher level of spatial (link level) and temporal (one-minute) resolution, it was found that the link GHG marginal cost was the best (Djavadian et al., 2020). The link GHG marginal cost represents the produced GHG by one vehicle traversing a studied link. In other words, GHG cost of a link requires the GHG ER and the time required to travel the link in concern. This explains why GHG ER was predicted in this study. Mainly, GHG gases include carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O). According to Althor et al. (2016), GHG gases combined are more harmful to the environment than CO_2 alone. Thus, GHG was taken into account in this study instead of just CO_2. CH_4 and N_2O contribute to climate change and have an effect that is larger than that of CO_2. Therefore, their effect is converted and GHG is reported in terms of "CO_2 equivalent" (United States Environmental Protection Agency, 2020b). This illustrates why we used CO_{2eq} ER as the modelling variable in our prediction models. The best developed predictive model in this study is suitable for various applications, such as dynamic anticipatory eco-routing.

The main contributions of this work are as follows:

1. Development of a deep learning framework based on LSTM to predict GHG ER (in CO_{2eq} g/s) at link level in a highly congested urban network, while utilizing microscopic data points.
2. Systematic analysis of the importance of various predictors contributing to the developed models.
3. Comparison of the results with the two commonly used emission prediction models, ARIMA and clustering, and demonstration of the strengths and limitations associated with each.
4. Illustration of the impact of utilizing different lengths of time interval for predictors and demonstrating the effect of the systematic tuning while applying LSTM.

This work is organized as follows. The methodology related to the utilized predictive models is in Section 2. The description of the study area is illustrated in Section 3. The traffic simulation and the emission model for collecting data are presented in Section 4, including the details of the simulated scenarios. Results and discussion are in Section 5. Finally, concluding remarks and future outlook are presented in Section 6.

2. Methodology

Fig. 1 illustrates the steps followed in this research work. The diagram shows which predictors (traffic and environmental variables) were adopted and how they were collected. In addition, it presents how the traffic and emission models interacted. As well, the deployed level of resolution (spatial and temporal) is demonstrated in Fig. 1. Before discussing the LSTM methodology in detail, we describe the two dominantly used models, ARIMA and clustering, for comparison purposes. ARIMA is a statistical model, while clustering is a machine learning algorithm.

2.1. ARIMA with explanatory variables

ARIMA (p,d,q) has been traditionally used for time series data that can be made to be "stationary" by differencing of "d" order. In other words, there should be no trends in the data in order to predict using the ARIMA model efficiently. Where: "p" and "q" are small integers reflecting on the "autoregressive" (AR) and "moving-average" (MA), respectively. To include r explanatory variables while using ARIMA, Eq. (1) is applied.

$$\hat{y}_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{k=1}^{r} \beta_k x_{tk} + e_t + \sum_{j=1}^{q} \theta_j e_{t-j} \tag{1}$$

where: $\hat{y}_t$ is the response at time t, μ is the constant, ϕ_i is the AR coefficient at lag i, y_{t-i} is the value of the variable in concern for prediction at time $t-i$, β_k is a coefficient of explanatory variable k, x_{tk} is explanatory variable k at time t, θ_j is the MA coefficient at lag j, and $e_{t-j} = y_{t-j} - \hat{y}_{t-j}$ is the forecast error that was made during period $t-j$ (Box et al., 2015).

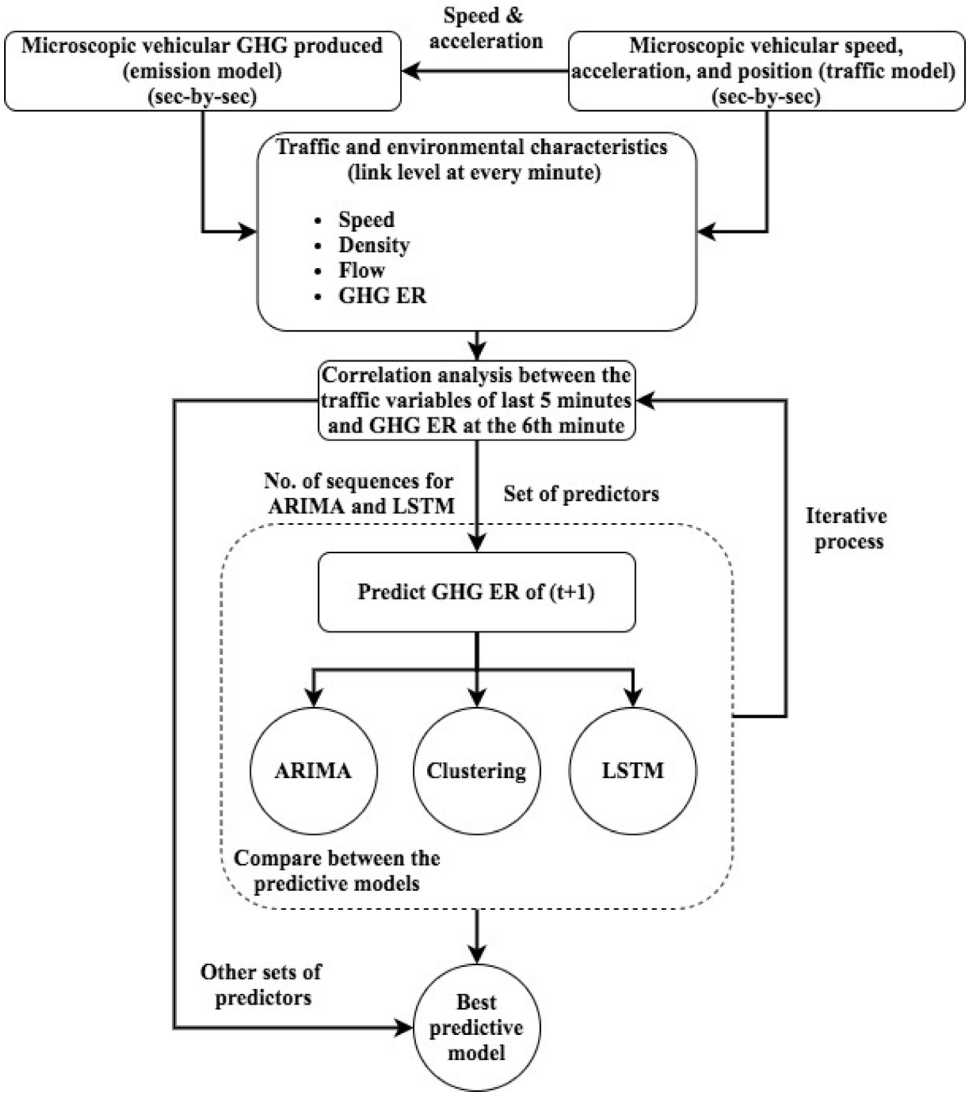

Fig. 1. Methodology followed in this study.

Fig. 2(a) illustrates the steps followed to develop the ARIMA model and predict the GHG ERs. A ratio of 80% to 20%, training to validation, was considered. To avoid overfitting, we have tried several sets of data and made sure there were no trends in the dataset.

To define the optimal parameters (p,d,q), an iterative process took place while considering the auto-correlation, partial auto-correlation plots of the differenced series, and root unit, which is a measure that signals when the time series is under- or overdifferenced. One of the major shortcomings of ARIMA models in our context was that a separate model was required for each link, based on the link data, to assure that the time series was stationary. In other words, ARIMA lacked the spatial dimension.

2.2. Clustering

Clustering is an important tool in data mining applications. It mainly groups objects so that in one group the objects are more related to each other than those in other groups (Mann and Kaur, 2013). To classify the traffic conditions at link level using the most important variables, K-means clustering has been utilized. It aims to classify data points into K clusters. Every data point belongs to a cluster with the minimum distance to the centroids of that cluster. In K-means clustering, the optimal cluster is defined when the total intra-cluster variance, or the squared error is minimized (Poucin et al., 2018), following Eq. (2).

$$M = \sum_{m=1}^{k} \sum_{i=1}^{n} \|x_i^{(m)} - c_m\|^2 \tag{2}$$

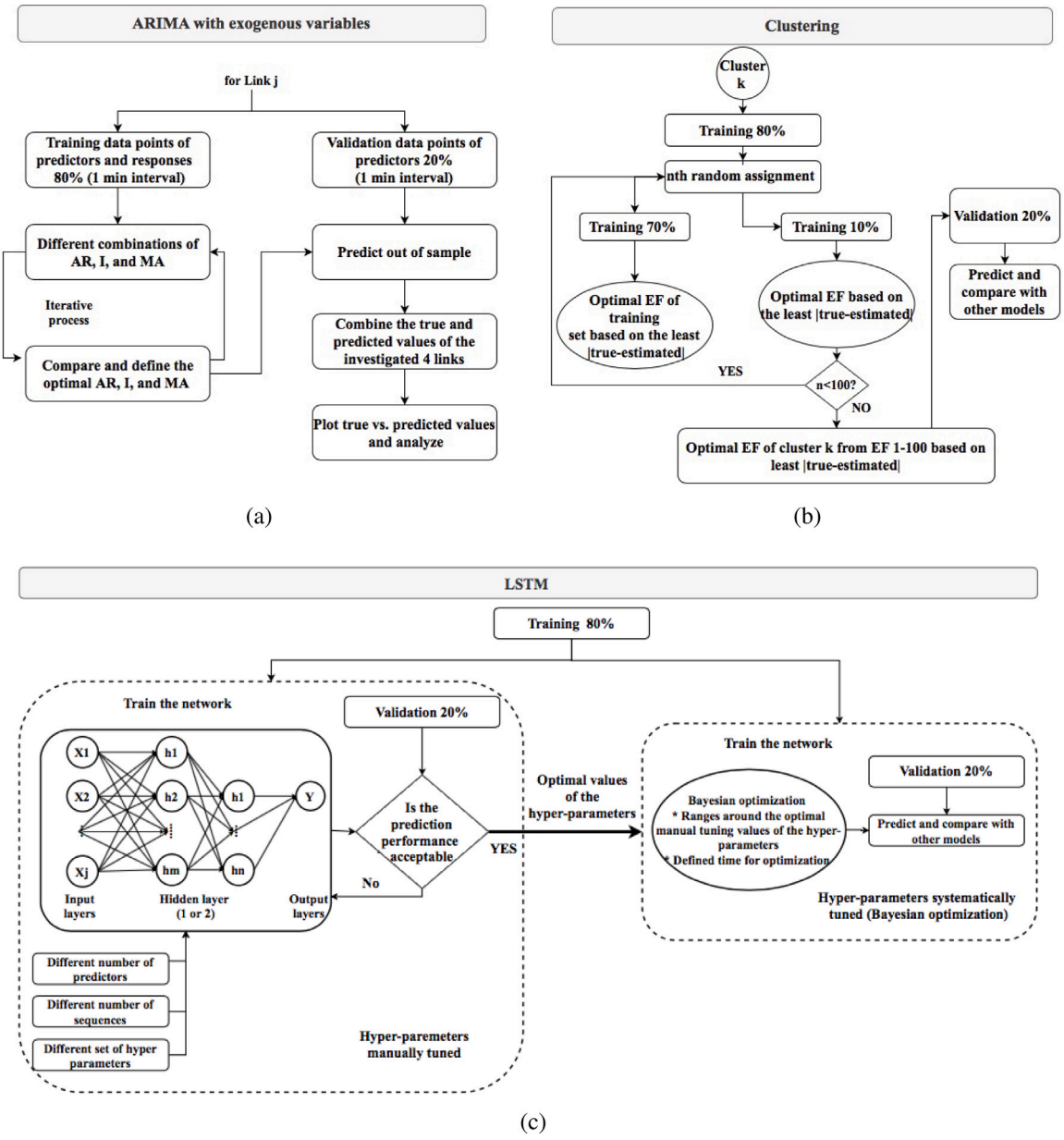

Fig. 2. Methodology for (a) ARIMA with explanatory variables, (b) clustering, and (c) LSTM with explanatory variables.

where M is the objective function, k is the number of clusters, n is the number of data points (observations), $x_i^{(m)}$ is the observation i being tested for cluster m, c_m is the centroid of cluster m. The number of clusters should be defined based on the data and a statistical analysis. To define the optimal number of clusters, the elbow method/sum of squared error (Poucin et al., 2018), which measures the sum of squared distances between the points within a cluster, was the guide. Fig. 2(b) presents the steps followed. For clustering, dataset was divided into 80% and 20% for training and validation, respectively (Tu et al., 2018). The optimal GHG ER (g/s) of each cluster was defined based on the minimum sum of absolute distances between GHG ERs and the centroid of the examined cluster. To avoid overfitting, cross validation was conducted for 100 iterations while training the model. For every cross-validation iteration, the training data of 80% was divided randomly into 70% and 10%. The optimal GHG ER for every cluster of the 70% data points was defined based on the least distance between true and estimated GHG ER. Then the optimal GHG ER was used for the cross validation on the 10% data points.

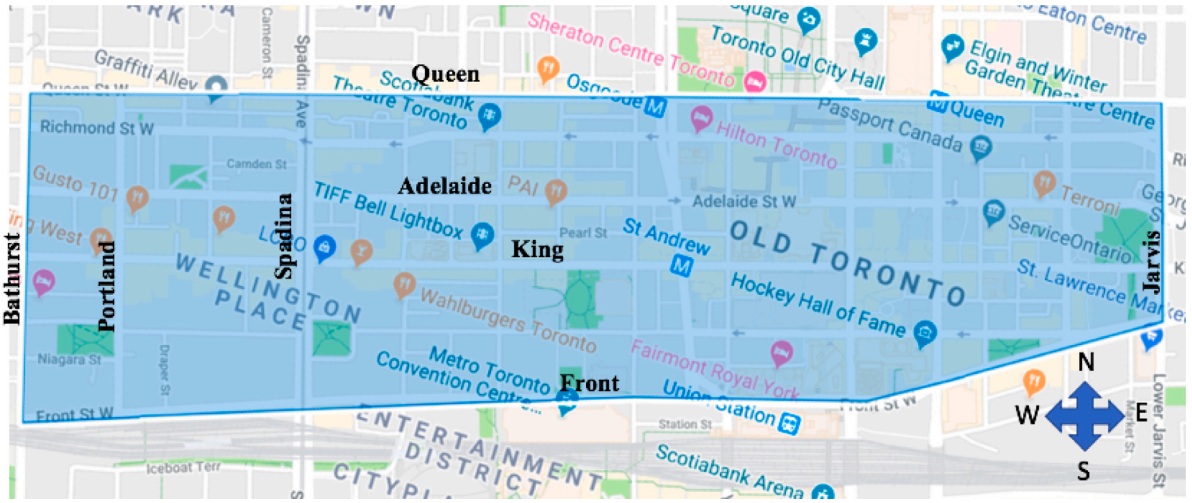

Fig. 3. Study area, downtown Toronto.

2.3. LSTM with explanatory variables

The main focus of this work was to develop an LSTM-based learning framework and compare it to already established models including ARIMA and clustering. LSTM is associated with three gates (Hochreiter and Schmidhuber, 1997)-input, forget, and output-following Eqs. (3), (4), and (5) below, respectively.

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i) \tag{3}$$

$$f_t = \sigma(w_f[h_{t-1}, x_t] + b_f) \tag{4}$$

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \tag{5}$$

Where: i_t represents the input gate, f_t represents the forget gate, o_t represents the output gate, σ represents the sigmoid function, w_x represents the weight for gate x neurons, h_{t-1} represents the output of the previous LSTM block at time $t-1$, x_t represents the input at current time step t, and b_x represents biases for respective gates (x).

Defining the best network is an iterative process and is dependent on the selection of the predictors, number of sequences, and set of hyper-parameters (Reimers and Gurevych, 2017; Hutter et al., 2015). It is worth mentioning that increasing the depth of NNs (Hermans and Schrauwen, 2013; Pascanu et al., 2013) and effectively tuning the network (Snoek et al., 2012) may contribute positively to the prediction performance. Thus, one and two hidden layers were investigated in this work, and Bayesian optimization (Wu et al., 2019) was utilized for the systematic tuning. With reference to the hyper-parameters tuning procedure, two stages took place. The first was the manual, while, the second was the systematic. Fig. 2(c) presents the followed methodology for LSTM application. A ratio of 80% to 20% was considered for training to validation, respectively. Adding drop out layers is an effective and easy way to avoid overfitting in NNs in general (Srivastava et al., 2014). This was the considered technique for the three investigated LSTM networks. To compare the performance of LSTM to ARIMA and clustering, four indicators were employed: the correlation coefficient between true (simulated) and predicted GHG ERs (in CO_{2eq} g/s), the fit to the ideal straight curve, the R^2, and the RMSE.

3. Study area

The road network of downtown Toronto was used as the study area. It was selected due to its high level of recurrent congestion during the morning peak period. It is crucial to note that the best GHG ER predictive model in this study can be deployed to apply anticipatory single- or multi-objective eco-routing for the same study area, downtown Toronto, or any urban network. The network consists of 223 links and 76 nodes (intersections). Links have different characteristics in terms of the number of lanes, free flow speed, and number of directions to assure heterogeneity. With reference to the speed limit, 132 and 18 links in the study area are of 60 and 80 km/h, respectively. Sixty six links in the study area have a speed limit of 40 km/h, while only five and two links have speed limits of 10 and 30 km/h, respectively. Fig. 3 illustrates the study area, including the major roads.

4. Data collection

Tackling the GHG prediction problem from a disaggregated perspective, requires high-resolution data points. To the best of our knowledge, currently there are no datasets available that report GHG at link level for a large and congested road network. An agent-based traffic and emission simulation has been used in this study to synthesize GHG and traffic characteristics for links at a very high

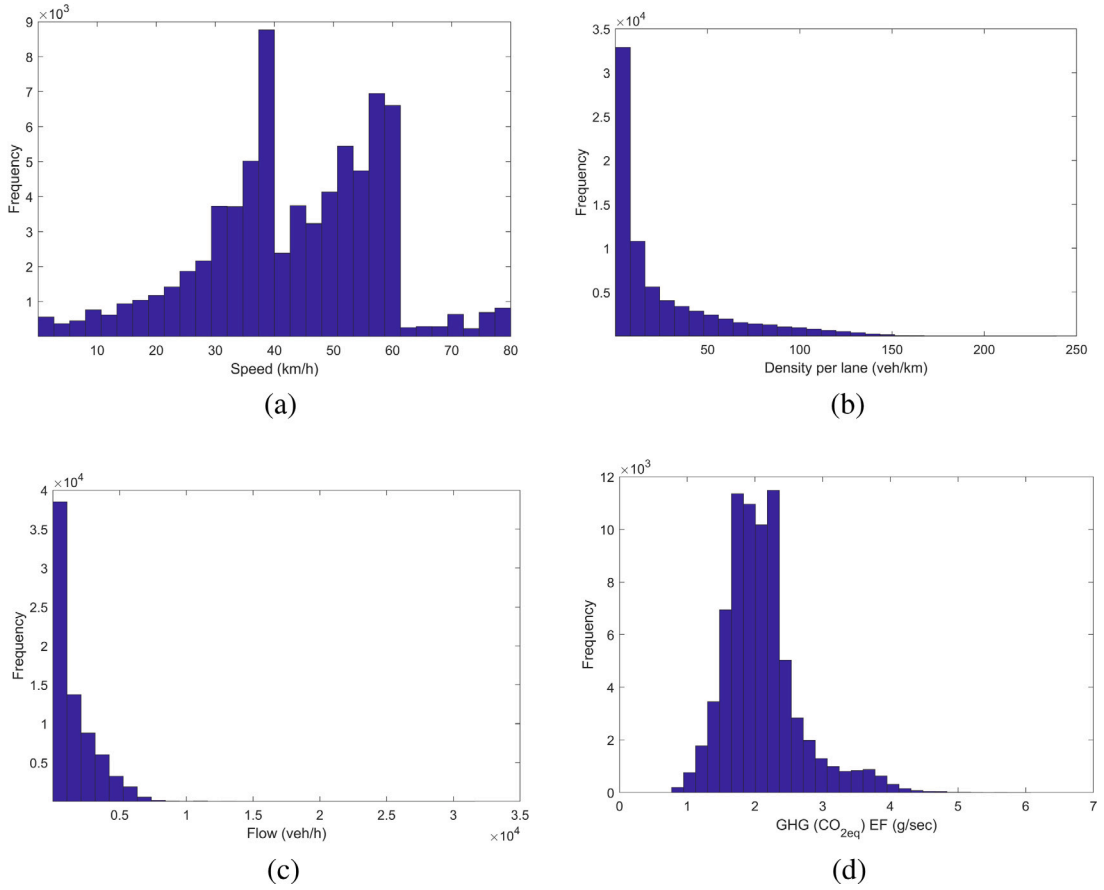

Fig. 4. Histogram of (a) speed, (b) density per lane, (c) flow, and (d) GHG ERs (in CO_{2eq} g/s).

temporal resolution (Djavadian and Farooq, 2018). Furthermore, we are of the view that in the near future, with the adoption of smart cities technologies, high-resolution datasets from real road networks will become available. In that case, the proposed LSTM framework can be retrained on such datasets. The used agent-based simulator in this study implemented a calibrated Intelligent Driver Model (IDM) (Treiber et al., 2000) for vehicular movement. Vehicles were dynamically routed on the network based on real-time traffic information (Djavadian and Farooq, 2018). Simulation ended once all of the vehicles reached their destinations. The link level space mean speed, density, and flow information was recorded every minute. Fig. 4 demonstrates the statistical analysis of the variables reflecting on the traffic conditions and GHG ER (in CO_{2eq} g/s) on links obtained from the simulation. It can be noticed in Fig. 4(a) that the mode for speed is 40 km/h. The speed average in the network based on second-by-second data points is 56.16 km/h. Nevertheless, speed varies from 0 to 80 km/h. Similarly, density (veh/km.lane), as in Fig. 4(b), and flow (veh/h), as in Fig. 4(c), are associated with a wide range reflecting different traffic conditions. Finally, GHG ERs (in CO_{2eq}) start from less than 1 g/s to more than 5 g/s as in Fig. 4(d).

With regards to the emission modelling, Motor Vehicle Emission Simulator (MOVES), which was developed by the USEPA, was adopted to generate GHG ERs (in CO_{2eq} g/s) (United States Environmental Protection Agency, 2014). MOVES estimates the second-by-second emissions by defining the vehicle operating mode, which is based on the vehicle specific power (VSP) as illustrated in Eq. (6). As shown in Fig. 1, the estimated second-by-second vehicular speed and acceleration by IDM model were used as inputs for MOVES to estimate the vehicular second-by-second GHG emissions. Unlike the GHG estimation approaches based on fuel, MOVES reflects on the environmental state more accurately (United States Environmental Protection Agency, 2020a), which makes it suitable for routing application. With regards to vehicle type, light duty vehicles (LDVs) only, including passenger cars (46%) and passenger trucks (54%), were only considered. Vehicle composition data depends on the report of Ministry of Transportation Ontario (2017). The distribution of the age (0 to 30 years) of the vehicles was also based on a 2017 report by the Ministry of Transportation Ontario (2017).

$$P_{V,t} = \frac{Av_t + Bv_t^2 + Cv_t^3 + mv_t a_t}{m} \tag{6}$$

Where:

$P_{V,t}$ is the vehicle-specific power (VSP) at time t

Table 1
Scenarios considered for data generation.

Demand factor	No. of vehicles	Departure time distribution
0.7	2437	Exponential, uniform, and normal
1	3477	Exponential, uniform, and normal
1.3	4520	Exponential, uniform, and normal
1.5	5259	Exponential
2	6988	Exponential

v_t is the speed of a vehicle at time t (m/s)

a_t is the acceleration of a vehicle at time t (m/s^2)

m is the mass of a vehicle, usually referred as "weight" (Mg).

A, B and C are track-road coefficients, representing rolling resistance, rotational resistance and aerodynamic drag, in unit kW-s/m, kW-s^2/m^2, and kW-s^3/m^3.

The second-by-second CO_2 emissions of every vehicle on every link were then used to estimate the space mean GHG ER (in CO_{2eq} g/s) of each link, based on a defined updating interval. Based on the 2012 *Canada–U.S. Air Quality Agreement Progress Report* (International Joint Commission, 2012), Canada switched to using MOVES during the summer of 2012.

In terms of the data points, a variety of scenarios have been simulated to trigger different traffic conditions on links, as shown in Table 1.

Travel demand was obtained from the *Transportation Tomorrow Survey* (TTS). The time dependent exogenous demand Origin–Destination (OD) matrices were based on five-minute intervals from TTS (DMG, 2011). With regards to the demand, it ranged from 2,437 to 6,988, reflecting demand factors from 0.7 to 2. To assure more heterogeneity in the generated traffic conditions at link level, different distributions were considered for the departure time, normal, uniform, and exponential. Two datasets were extracted corresponding to two time intervals, i.e. 30 s and one minute, to examine the impact of the two levels of resolution on the prediction performance. The data has been pre-processed to suit each of the discussed models in Section 2 and to assure a fair comparison between the models. When one minute was the updating interval of predictors for clustering and LSTM, 48,652 and 12,159 data points for training and validation were employed, respectively. For the 30 s analysis, the numbers double. The data has been divided into 80% training and 20% validation for clustering and LSTM. For the ARIMA model, four representative links were considered, which are associated with different characteristics to give an indication of the prediction performance at the network level. The total number of sequences was 83. The dataset has been divided into 80% to 20% training to validation as well. Training and validation datasets consisted of 66 and 17 sequences, respectively.

5. Results and discussion

This section presents the major findings, starting with the detailed correlation analysis outcome. Then we present the results of the ARIMA, clustering, and LSTM models, as well as a comparison between them. It is important to note that the considered explanatory variables for comparing the aforementioned three predictive models were the top three highly correlated variables found in Section 5.1.

5.1. Correlation analysis

The correlation analysis was the judging factor used to define not only the most important predictors for the three models, but also the optimal number of sequences/minutes for the predictors while applying LSTM. Traffic and environmental information of studied links were captured. In order to expand the spatial dimension, the characteristics of in-links (upstream links) were not neglected. Traffic conditions (e.g. speed flow, density, etc.) at time t on the upstream links would give a strong indication of the traffic condition on the downstream studied link at time $t+1$. The considered variables for this analysis were speed, density, flow, delay (difference between free-flow travel time and actual travel time), in-links speed, in-links density, in-links flow, and GHG ER (in CO_{2eq} g/s). The maximum link length in the network is around 450 m. Speed varies from 0 to 80 km/h. Under the free-flow traffic condition, the maximum travel time required to traverse a link is around 0.8 min. Nevertheless, when the network is congested, it takes a longer time to traverse a link. Thus, a five-minute period was the starting point for investigating the best sequence to capture the changes in the GHG ER at the sixth minute. The main outcome of this analysis is the correlation between the aforementioned indicators and the GHG ER (in CO_{2eq} g/s) at the sixth sequence/minute. Linear correlation has been employed for this part of the analysis. With reference to the results, the absolute value of the correlation coefficient between any variable and the GHG ER (in CO_{2eq} g/s) at the sixth minute, except for delay, increases from minute one to minute five (top to down), as shown in Fig. 5.

In terms of the importance order, speed, GHG ER (in CO_{2eq} g/s), density, and in-links speed are the top four highly correlated variables with the GHG ER at the sixth minute. Speed is the variable of the highest correlation coefficient with the GHG ER. This is due to the explicit dependency of GHG estimation on speed (United States Environmental Protection Agency, 2015). Based on the traffic fundamental relationships, the relationship between speed and density is observed to be monotonically decreasing, which justifies the high correlation between density and the GHG ER (Papacostas and Prevedouros, 1993). Among the in-links characteristics, in-links speed is the most correlated variable with the GHG ER, which is due to the fact that speed is the variable GHG depends on for estimation. This analysis triggered the choice of not only the predictors for the three models, but also the number of sequences for LSTM.

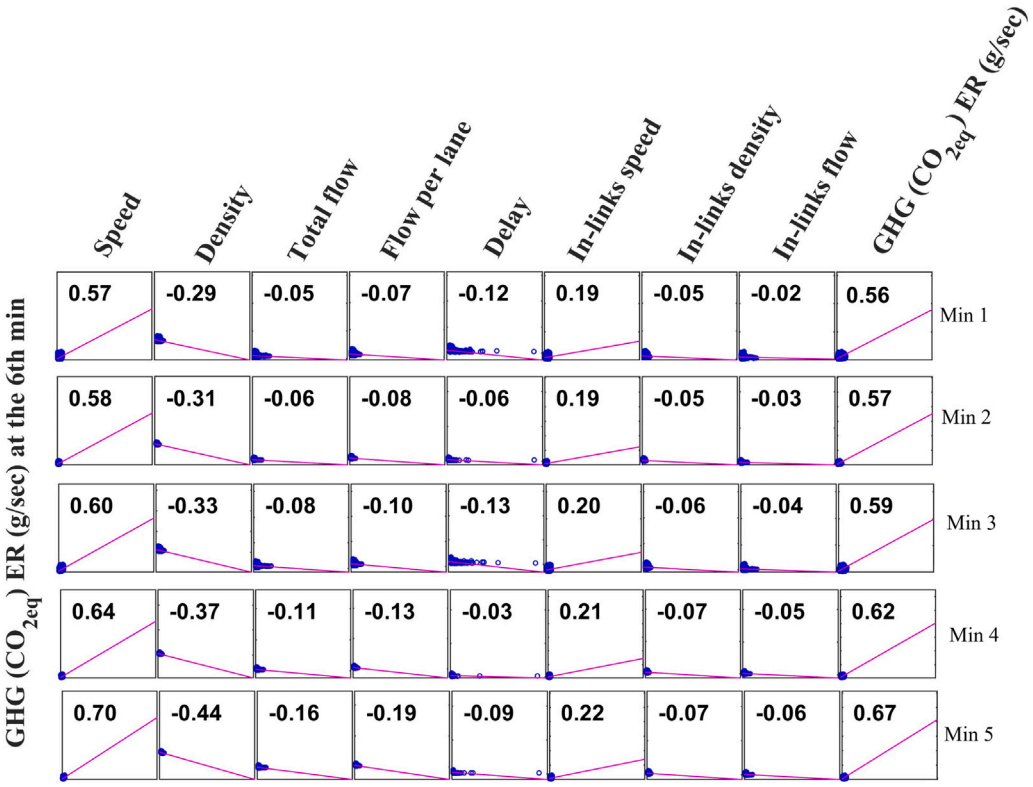

Fig. 5. Correlation between the proposed variables at every minute from one to five (from top to bottom), with GHG ER (in CO_{2eq} g/s) at the sixth minute.

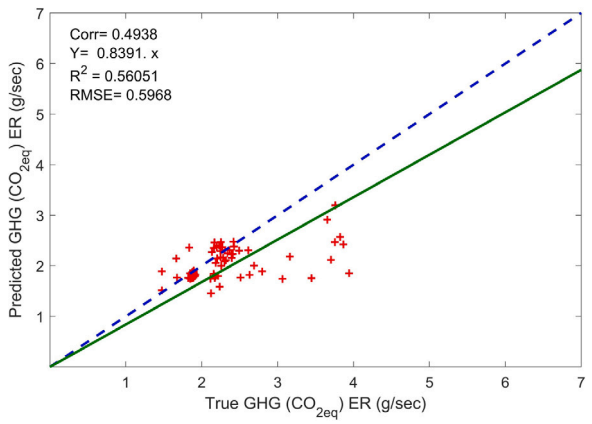

Fig. 6. Sample representation (four sampled links) of network level prediction using ARIMA with explanatory variables.

5.2. ARIMA with explanatory variables

The ARIMA model with explanatory variables does not consider the non-linearity between variables (Zhang, 2003), which is a serious drawback when the data and relationships are complex. Furthermore, ARIMA model is not scalable, in which every link required a model based on its data, which is tedious while dealing with a large number of links. Nevertheless, to compare between the models, a sample of four links, which are associated with different characteristics (number of lanes and free flow speed) and conditions (congested and uncongested), were considered. In terms of the adopted predictors (explanatory variables), the top three variables, speed, density, and GHG ER, were used. As shown in Fig. 6 and Table 3, the correlation coefficient and R^2 of the ARIMA model of the four links were 0.49 and 0.56, respectively. With regards to the fit to the straight curve, there was a slight overestimation and the RMSE was 0.597 g/s.

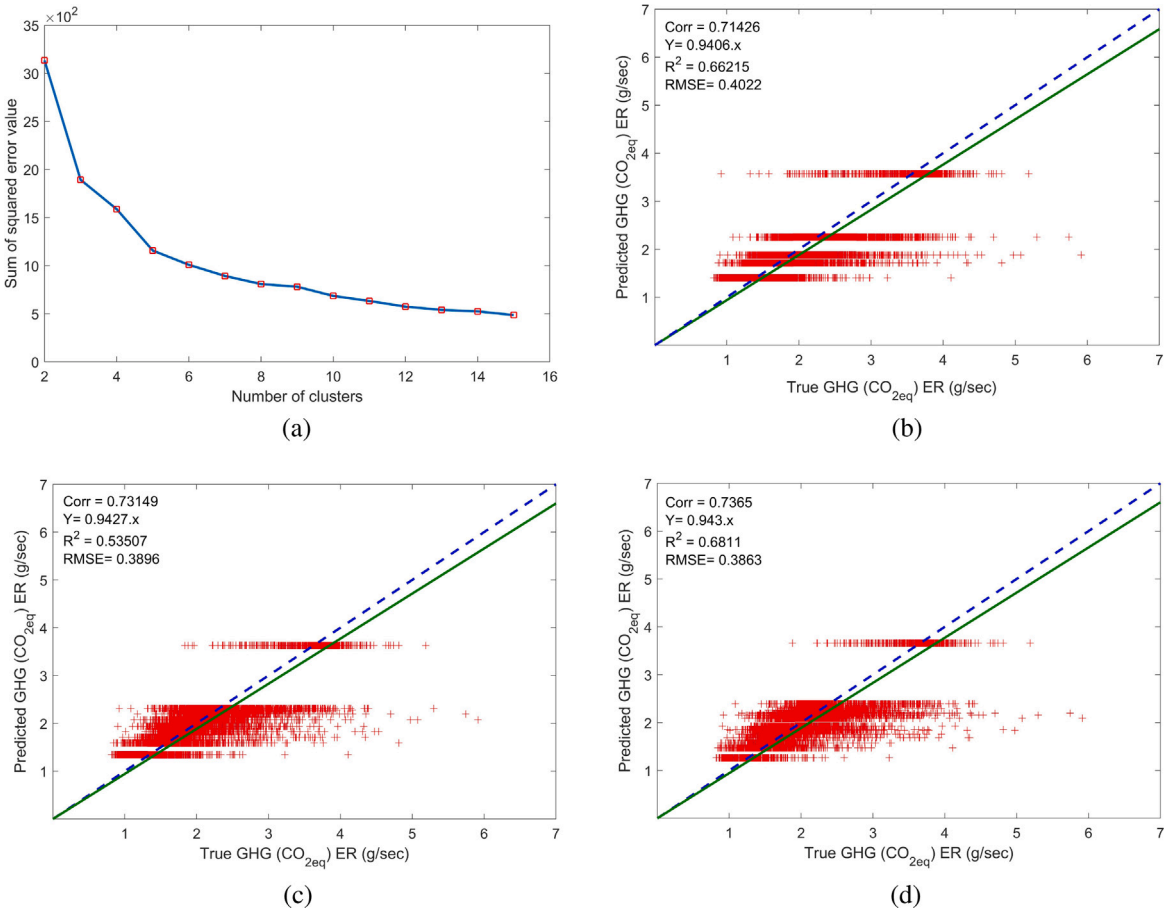

Fig. 7. Sum of squared error (elbow method) of 15 clusters in figure (a) and true versus predicted GHG ERs (in CO_{2eq} g/s) of clustering where: (b) is for 5 clusters, (c) is for 10 clusters, and (d) is for 15 clusters.

5.3. Clustering

To define the optimal number of clusters, the sum of squared error (elbow method) was adopted. From Fig. 7(a), the optimal number of clusters is five. However, 10 and 15 clusters were examined for further enhancements. For a fair comparison, the top three explanatory variables, speed, density, and GHG ER, were used. Table 3 includes the performance indicators of the whole investigated predictive models for comparison purposes. From Fig. 7(b), it is noticed that predicted values of GHG ERs were considered as discrete variables triggering dramatic over- and underestimation. Values of GHG ERs ranging from 1 to 4 were predicted to be 1.5, 1.8, 1.9, or 2.3 g/s. The correlation coefficient, R^2, and RMSE were 0.71, 0.66, and 0.4 g/s, respectively. While comparing true (simulated) versus predicted GHG ERs of 5 to 10 clusters as in 7(c), and to 15 clusters as in 7(d), it is illustrated that even when the number of clusters increased, the performance indicators did not improve substantially, and the GHG ERs between 2.5 and 4 g/s were either underestimated to be 2.5 or overestimated to be 4 g/s.

With reference to the fit to the straight curve, it is noticed in Fig. 7(d) that GHG ERs between 2.7 to around 4 g/s were either underestimated, taking one value of around 2.5 g/s, or over estimated. It can be concluded that since clustering considered the predicted GHG ER to be a discrete variable, the dynamic nature of traffic conditions was not captured efficiently, compared to ARIMA, triggering the aforementioned outcome. As well, there is a proportional relationship between the number of clusters and the amount of data required for training. Thus, even when the number of clusters is increased the data requirement can be hard to fulfil.

5.4. LSTM with explanatory variables results

Table 2 shows the specifications of the trained networks including the predictors, number of sequences (previous minutes), number of LSTM layers, updating interval of data points, and hyper-parameter tuning approach.

Before demonstrating the results, it is important to note that the presented networks in this section were systematically tuned adopting the Bayesian optimization (Gelbart et al., 2014). The search range for each of the hyper-parameters was around the optimal

Table 2
LSTM specifications considered for application.

Model ID	Predictors	No. of sequences/ min	No. of LSTM layers	Sequence interval	Hyper-param. tuning approach
LSTM1	Speed, density, and GHG ER (in CO_{2eq} g/s)	3	1	1 min	Bayesian
LSTM2	Speed, density, GHG ER (in CO_{2eq} g/s), and in-links speed	3	1	1 min	Bayesian
LSTM3	Speed, density, GHG ER (in CO_{2eq} g/s), and in-links speed	3	2	1 min	Bayesian

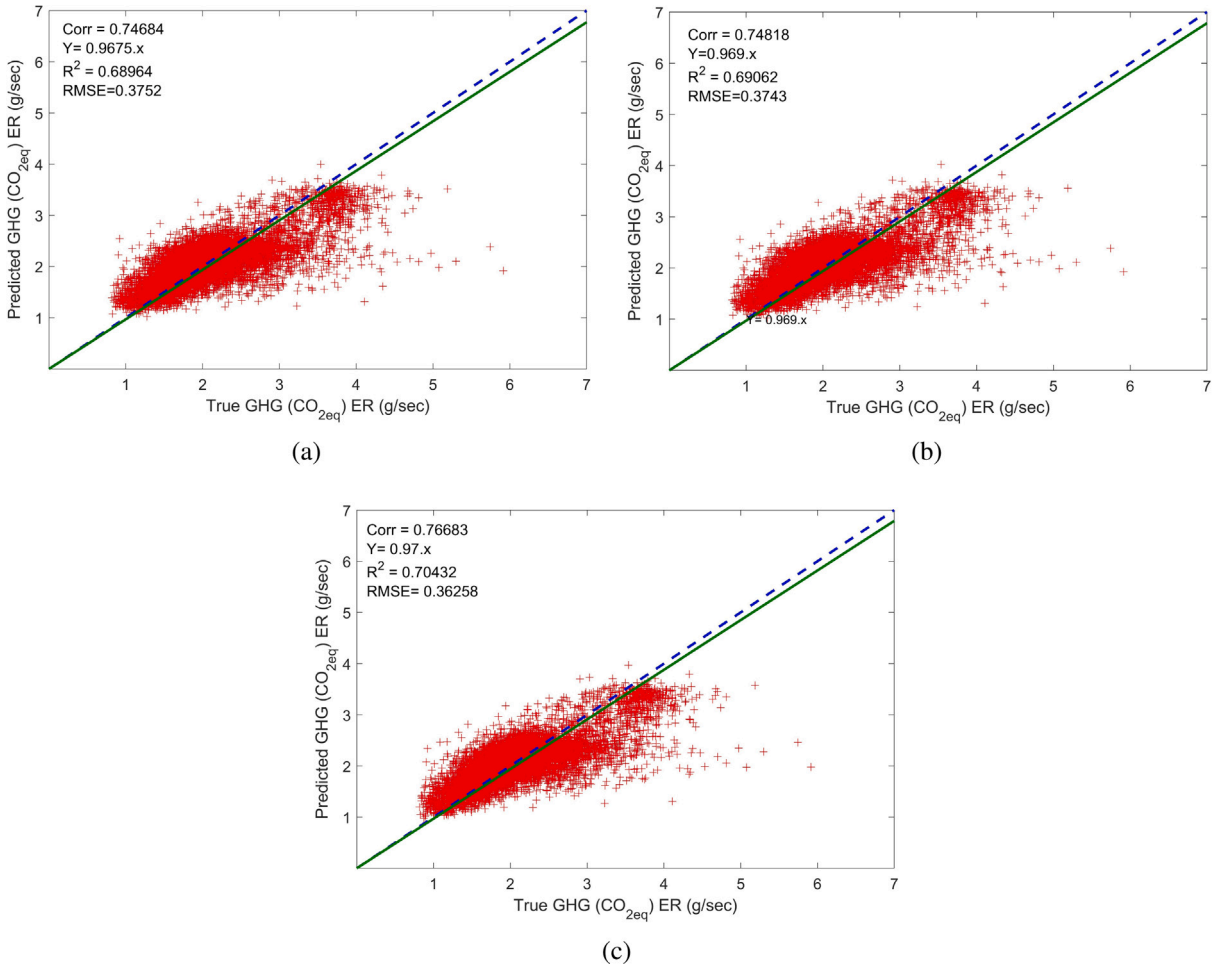

Fig. 8. True versus predicted GHG ERs (in CO_{2eq} g/s) of (a) LSTM1, (b) LSTM2, and (c) LSTM3.

values of the manual tuning for each of the networks. For the manual tuning, a comprehensive list of tuning sets was examined to narrow the search range for a more efficient systematic tuning. Two solvers were considered: the stochastic gradient descent with momentum (sgdm) (Robert, 2014) and adaptive moment estimation (Adam) methods (Kingma and Ba, 2014). The sgdm adds a momentum term to the parameter update to reduce oscillation associated with the standard stochastic gradient descent solver along the path of steepest descent towards the optimal solution (Robert, 2014). Adam is derived from adaptive moment estimation. It adopts different learning rates for the parameters unlike sgdm, but uses a momentum term like the sgdm to further improve network training (Kingma and Ba, 2014).

The hyper-parameters considered include: the initial learning rate, momentum, max epochs, learning rate drop factor, learning rate drop period, number of hidden units of the first LSTM (hidden) layer, and the number of hidden units of the second (LSTM) layer when used. Not to neglect that different numbers of sequences (previous minutes) have been assessed for the best application. Three previous minute sequences contributed to the best performance, as the period was sufficient to capture the changes in traffic

Table 3
Performance indicators of the examined predictive models.

Predictive model	Correlation	R^2	Linear fit	RMSE
ARIMA	0.4938	0.56051	Y = 0.8391.x	0.5968
Clustering (5 clusters)	0.714	0.662	Y = 0.94.x	0.402
Clustering (10 clusters)	0.731	0.535	Y = 0.942.x	0.39
Clustering (15 clusters)	0.736	0.681	Y = 0.943.x	0.3863
LSTM1	0.747	0.689	Y = 0.9675.x	0.3752
LSTM2	0.748	0.691	Y = 0.969.x	0.374
LSTM3	0.767	0.704	Y = 0.97.x	0.362

conditions and according to the linear correlation analysis as in Fig. 5. The performance indicators of the whole examined predictive models are presented in Table 3. Fig. 8(a) demonstrates the performance indicators of LSTM1, which employs the link's three highly correlated predictors of the last three minutes. This was the adopted set of predictors (explanatory variables) and number of sequences in the case of ARIMA and clustering application for a proper comparison. The LSTM1 was associated with a correlation coefficient, R^2, and RMSE of 0.747, 0.69, and 0.375 g/s, respectively, while underestimating the GHG ERs by 3%. Comparing LSTM2 as in Fig. 8(b) to LSTM1 as in Fig. 8(a) illustrates a negligible enhancement in the correlation coefficient and R^2, while the RMSE is similar. LSTM2 underestimated the GHG ER by 3%. This slight improvement stemmed from the fact that LSTM2 did not only employ the three mostly correlated predictors with the response (GHG ER), but also the in-links highly correlated characteristic (in-links speed), as illustrated in Fig. 5. When speed of in-links (upstream) at time t is very low due to congestion contributing to more GHG emissions, this means that at time $t+1$ these vehicles driving at low speed would be on the studied downstream link. In other words, speed on in-links at time t provides an indication of what the speed would be on the downstream link at time $t+1$. The results of LSTM2 demonstrate the potential of including the highly correlated in-links characteristics to better reflect on the spatial dimension while predicting GHG ERs. LSTM3 as in Fig. 8(c), which adopts two hidden layers to account for the higher order of correlation between the variables, gave the best results in terms of the four performance indicators. The LSTM3 network underestimated by only 3% and was associated with a correlation coefficient, R^2, and RMSE of 0.77, 0.7, and 0.36 g/s, respectively.

A mutual feature in Figs. 8(a), 8(b), and 8(c) was that the models were unable to predict GHG ERs that were higher than 4 g/s. This is due to the fact that data points reflecting the associated traffic conditions were not sufficient, as illustrated in Fig. 4(d). That is, more data points representing the traffic conditions associated with GHG ERs higher than 4 g/s may enhance prediction accuracy.

With reference to the required computational power, manually tuning the LSTM networks took a longer time compared to systematically tuning them. This is due to the fact that different number of sequences and sets of predictors were examined in addition to the different sets of hyper-parameters. One, three, and five sequences were assessed for different predictors for LSTM1. Then the optimal hyper-parameters of every network were defined following two stages, manual and systematic, for every LSTM network. Every manual training iteration of LSTM1 and LSTM2 took from 3 to 10 h based on the specifications of the predictors, hyper-parameters, number of sequences, and number of hidden layers. On the other hand, every manual tuning iteration for LSTM3, of two hidden layers, took as much as double the time compared to LSTM1 or LSTM2. Around 40 manual training iterations were conducted for each of LSTM1, LSTM2, and LSTM3 to define the best set of predictors, number of sequences, and hyper-parameters. The systematic tuning depended on the outcome of the manual tuning process. Every two hyper-parameters were optimized for two days (48 h). Then the whole hyper-parameters were optimized for two days as well, while considering the ranges of the partial optimization process of every two hyper-parameters. Increasing the period of time given for the systematic tuning (Bayesian optimization) may contribute to further enhancements in the prediction performance for future applications. Compute Canada's high performance clusters were used to run several instances at once to accelerate the training process.

Investigating the impact of a shorter updating interval is of an added value. Hence, three LSTM models of the same specifications of LSTM1, LSTM2, and LSTM3 have been developed for 30 s intervals. The best model resulted in RSME of 0.48 g/s, R^2 of 0.707, and a correlation of 0.78. It was found that increasing the level of resolution had an adverse impact on the RMSE, while slightly improving the correlation coefficient for the three networks. More specifically, for LSTM3 when the updating interval was 30 s, both the correlation coefficient and the RMSE increased by 1.34% and 35%, respectively. The main justification for this is that the shorter updating interval triggered more noise and local fluctuations, which were not well captured by a general LSTM that was not trained for a specific link. Therefore, a one-minute updating interval was sufficient at this point to account for the high level of heterogeneity in link characteristics.

Comparing the LSTM model outcome to the other models, clustering, and ARIMA, shows that the three aforementioned LSTM networks outperformed the best case of clustering as in Fig. 7(d) and ARIMA as in Fig. 6 in terms of the four performance indicators. More specifically, the RMSE of LSTM3 was less by 39% and 6% compared to ARIMA and the best case of clustering, respectively. Unlike clustering, LSTM considered GHG ER (in CO_{2eq} g/s) to be a continuous variable. Compared to ARIMA with explanatory variables, LSTM is scalable and was applicable to any link in the studied network. It is important to note that the prediction performance of ARIMA was link-based, and was impacted by the level of congestion on a studied link. For two uncongested links, RMSE was 0.0718 and 0.0054 g/s. On the other hand, in the case of two congested links, RMSE of ARIMA was 0.95 and 0.4494 g/s. This means that ARIMA is link-based and cannot be generalized. Unlike ARIMA, LSTM was more stable whatever the traffic condition was on links. RMSE ranged from 0.3577 to 0.4494 regardless of the traffic condition on the studied links. This means that LSTM is a generic model. The accuracy of the predictive model is profound. When single- or multi-objective eco-routing is applied, the low accuracy would contribute to over- or underestimation of GHG cost of links. This would have a negative impact on the efficiency of routing, as the real traffic and environmental states would not be represented reliably.

6. Conclusion and potential directions

It has been shown that the efficiency of routing is a key tool to reduce the undesirable impact of transportation systems on the environment (Tu et al., 2019a; Alam et al., 2018). Thus, reliably predicting GHG ER (in CO_{2eq} g/s) is essential. It paves the way for the switch from myopic to anticipatory eco-routing and can also be adopted in other applications that can efficiently mitigate the produced GHG emissions. Previous studies were predominantly based on macroscopic data points. Even when microscopic data were used, the case studies involved a single or few intersections/links (Alfaseeh and Farooq, 2020). Therefore, there is a need for studies tackling issues related to GHG emissions at a high spatial and temporal resolution. To predict GHG ERs at link level, a deep learning approach, namely LSTM with explanatory variables, has been employed in this study. LSTM has been chosen because it has overcome various limitations of other NNs in the context of time-series data, such as the vanishing gradient problem (Amarpuri et al., 2019). To assure satisfying prediction outcome, a sufficient amount of representative data points were generated from MOVES and traffic microsimulation for downtown Toronto. The optimal set of predictors and number of time sequences was identified based on a comprehensive correlation analysis. A comparison with the commonly used approaches for times series prediction, namely clustering and ARIMA, was also conducted.

In this study, the LSTM3 network of two hidden layers, which was systematically tuned using Bayesian optimization, outperformed the best case of clustering and ARIMA models in terms of the four performance indicators. More specifically, the LSTM3 network reduced the RMSE by 39% and 6% compared to ARIMA and the best case of clustering, respectively. While the LSTM can scale up to a network level and considered the GHG ER as a continuous variable, ARIMA required a model for every link and clustering considered GHG ER as a discrete variable. ARIMA's major drawbacks are related to scaling and non-linearity consideration between variables. In clustering, the main limitation is that the predicted GHG ER (in CO_{2eq} g/s) was considered to be a discrete variable, which has affected the fit quality to the straight curve. In other words, GHG ERs (in CO_{2eq} g/s) between 2.5 and 4 g/s were either underestimated to be 2.5 g/s or overestimated to be 4 g/s, even when the number of clusters was 15. This is due to the fact that clustering lacks the temporal dimension compared to the LSTM model. Utilizing a smaller updating interval of 30 s in the case of LSTM modelling improved the correlation coefficient slightly compared to the case of a one-minute updating interval, but had a negative impact on the RMSE of the three LSTM models. The shorter updating interval triggered more local fluctuations that were not well captured by the network level model.

For future work, one such application that would benefit from the predictive model of this study is anticipatory single- or multi-objective eco-routing. Anticipatory eco-routing takes the GHG prediction (future time-step) of links in order to define the dynamic route for vehicles. This will allow for the switch from reactive (myopic) to proactive (anticipatory) routing for less GHG emissions while routing vehicles to their destinations. Fig. 4 illustrates that the frequency of speed higher than 60 km/h or lower than 30 km/h is very small. Hence, to enhance the current work, generating more data points reflecting on the aforementioned condition may contribute positively to the current outcomes. With reference to the hyper-parameters tuning, dedicating more time for Bayesian optimization may result in further improvement of the model's prediction capacity. Developing models for categorized links based on their speed limit and number of lanes may introduce further enhancements. In addition, utilizing physical constraints, which consider suitable traffic characteristics (flow over capacity, density over jam density, etc.), may contribute to higher prediction accuracy. Since the traffic and environmental information were obtained from simulation, employing sensors to collect real data would further enhance the model's prediction capability. It has been found that the link-level marginal cost, which reflects on the produced GHG by one vehicle traversing a studied link, gave optimal results in terms of system-level GHG reductions (Djavadian et al., 2020). The marginal cost takes into account the estimated GHG ER and time to travel a studied link at a specific time interval. Thus, future policies may consider taking into account the link-level GHG concentrations while developing new regulations. As well, other pollutants, such as nitrogen oxide (NOx) and particulate matter (PM) should be considered in future studies.

Acknowledgements

This research is funded by Dr. Bilal Farooq's Ontario Early Researcher Award. We would also like to thank Dr. Shadi Djavadian for providing the source code for traffic microsimulation used in this research.

References

Alam, M., Perugu, H., McNabola, A., 2018. A comparison of route-choice navigation across air pollution exposure, CO_2 emission and traditional travel cost factors. Transp. Res. D 65, 82–100.

Alfaseeh, L., Farooq, B., 2020. Multifactor taxonomy of ecorouting models and future outlook. J. Sens. 2020.

Althor, G., Watson, J.E., Fuller, R.A., 2016. Global mismatch between greenhouse gas emissions and the burden of climate change. Sci. Rep. 6, 20281.

Amarpuri, L., Yadav, N., Kumar, G., Agrawal, S., 2019. Prediction of CO_2 emissions using deep learning hybrid approach: A Case Study in Indian Context. In: 2019 Twelfth International Conference on Contemporary Computing. IC3, IEEE, pp. 1–6.

Anas, A., Timilsina, G.R., Zheng, S., 2009. An Analysis of Various Policy Instruments to Reduce Congestion, Fuel Consumption and CO2 Emissions in Beijing. The World Bank.

Box, G.E., Jenkins, G.M., Reinsel, G.C., Ljung, G.M., 2015. Time Series Analysis: Forecasting and Control. John Wiley & Sons.

Djavadian, S., Farooq, B., 2018. Distributed dynamic routing using network of intelligent intersections. In: ITS Canada ACGM. 2018.

Djavadian, S., Tu, R., Farooq, B., Hatzopoulou, M., 2020. Multi-objective eco-routing for dynamic control of connected & automated vehicles. arXiv preprint arXiv:2005.00815.

DMG, 2011. Transportation Tomorrow Survey. Department of Civil Engineering, University of Toronto Toronto, Ontario, Canada.

Dong, Y., Xu, J., Liu, X., Gao, C., Ru, H., Duan, Z., 2019. Carbon emissions and expressway traffic flow patterns in China. Sustainability 11 (10), 2824.

Gelbart, M.A., Snoek, J., Adams, R.P., 2014. Bayesian optimization with unknown constraints. arXiv preprint arXiv:1403.5607.

Gmira, M., Gendreau, M., Lodi, A., Potvin, J.-Y., 2017. Travel speed prediction using machine learning techniques. In: Proc ITS World Congr. pp. 1–10.
Hermans, M., Schrauwen, B., 2013. Training and analysing deep recurrent neural networks. In: Advances in Neural Information Processing Systems. pp. 190–198.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.
Hoogendoorn, S., Van Zuylen, H.J., Schreuder, M., Gorte, B., Vosselman, G., 2003. Microscopic traffic data collection by remote sensing. Transp. Res. Rec. 1855 (1), 121–128.
Hutter, F., Lücke, J., Schmidt-Thieme, L., 2015. Beyond manual tuning of hyperparameters. KI-K. Intell. 29 (4), 329–337.
International Joint Commission, 2012. Canada/United States Air Quality Agreement: Progress report 2012. International Joint Commission.
Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
Liu, Z., Wang, F., Tang, Z., Tang, J., 2020. Predictions and driving factors of production-based CO_2 emissions in Beijing, China. Sustainable Cities Soc. 53, 101909.
Mann, A.K., Kaur, N., 2013. Review paper on clustering techniques. Glob. J. Comput. Sci. Technol.
Ministry of Transportation Ontario, 2017. Statistics and management reporting.
Papacostas, C.S., Prevedouros, P.D., 1993. Transportation Engineering and Planning.
Pascanu, R., Gulcehre, C., Cho, K., Bengio, Y., 2013. How to construct deep recurrent neural networks. arXiv preprint arXiv:1312.6026.
Perugu, H., Wei, H., Yao, Z., 2017. Developing high-resolution urban scale heavy-duty truck emission inventory using the data-driven truck activity model output. Atmos. Environ. 155, 210–230.
Poucin, G., Farooq, B., Patterson, Z., 2018. Activity patterns mining in Wi-Fi access point logs. Comput. Environ. Urban Syst. 67, 55–67.
Reimers, N., Gurevych, I., 2017. Optimal hyperparameters for deep lstm-networks for sequence labeling tasks. arXiv preprint arXiv:1707.06799.
Robert, C., 2014. Machine learning, a probabilistic perspective. J. Electron. Sci. Technol. 27 (2).
Singh, K.P., Gupta, S., Kumar, A., Shukla, S.P., 2012. Linear and nonlinear modeling approaches for urban air quality prediction. Sci. Total Environ. 426, 244–255.
Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical bayesian optimization of machine learning algorithms. In: Advances in Neural Information Processing Systems. pp. 2951–2959.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (1), 1929–1958.
Treiber, M., Hennecke, A., Helbing, D., 2000. Congested traffic states in empirical observations and microscopic simulations. Phys. Rev. E 62 (2), 1805–1824. http://dx.doi.org/10.1103/PhysRevE.62.1805, arXiv:0002177.
Tu, R., Alfaseeh, L., Djavadian, S., Farooq, B., Hatzopoulou, M., 2019a. Quantifying the impacts of dynamic control in connected and automated vehicles on greenhouse gas emissions and urban NO_2 concentrations. Transp. Res. D 73, 142–151.
Tu, R., Alfaseeh, L., Djavadian, S., Saleh, M., Farooq, B., Hatzopoulou, M., 2019b. What happens to on-road emissions when travel time on a road network is improved through end-to-end dynamic routing for connected autonomous vehicles?. In: Transportation Research Board TRB (2019).
Tu, R., Kamel, I., Wang, A., Abdulhai, B., Hatzopoulou, M., 2018. Development of a hybrid modelling approach for the generation of an urban on-road transportation emission inventory. Transp. Res. D 62, 604–618.
United States Environmental Protection Agency, 2014. Motor Vehicle Emission Simulator (MOVES): User's Guide For MOVES2014a. United States Environmental Protection Agency.
United States Environmental Protection Agency, 2015. Exhaust Emission Rates for Light-Duty On-road Vehicles in MOVES2014: Final Report. United States Environmental Protection Agency.
United States Environmental Protection Agency, 2017. Sources of greenhouse gas emissions.
United States Environmental Protection Agency, 2020a. Greenhouse gases equivalencies calculator - calculations and references. https://https://www.epa.gov/energy/greenhouse-gases-equivalencies-calculator-calculations-and-references, (Online; Accessed 15 July 2020).
United States Environmental Protection Agency, 2020b. How does MOVES calculate CO_2 and CO_2 equivalent emissions? https://www.epa.gov/moves/how-does-moves-calculate-co2-and-co2-equivalent-emissions, (Online; Accessed 9 July 2020).
Wu, J., Chen, X.-Y., Zhang, H., Xiong, L.-D., Lei, H., Deng, S.-H., 2019. Hyperparameter optimization for machine learning models based on bayesian optimization. J. Electron. Sci. Technol. 17 (1), 26–40.
Zhang, G., 2003. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50, 159–175.
Zhao, J., Zhang, J., Jia, S., Li, Q., Zhu, Y., 2011. A mapreduce framework for on-road mobile fossil fuel combustion CO_2 emission estimation. In: 2011 19th International Conference on Geoinformatics. IEEE, pp. 1–4.
Zhao, M., Zhou, Y., Li, X., Cao, W., He, C., Yu, B., Li, X., Elvidge, C.D., Cheng, W., Zhou, C., 2019. Applications of satellite remote sensing of nighttime light observations: Advances, challenges, and perspectives. Remote Sens. 11 (17), 1971.