University of Nevada, Reno

Methane concentration forward prediction using machine learning from measurements in underground mines

A thesis submitted in partial fulfillment of the

requirements for the degree of Master of Science in

Mining Engineering

by

Tulio Dias de Almeida

George Danko/Thesis Advisor

December, 2021

THE GRADUATE SCHOOL

We recommend that the thesis
prepared under our supervision by

Tulio Dias de Almeida

entitled

Methane concentration forward prediction using machine learning from measurements in underground mines

be accepted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

George L. Danko, Ph.D
Advisor

Javad Sattarvand, Ph.D
Committee Member

Behrooz Abbasi, Ph.D
Committee Member

Simon R. Poulson, Ph.D
Graduate School Representative

David W. Zeh, Ph.D., Dean
Graduate School

December, 2021

ABSTRACT

Unmanaged gases inside the mine airways are hazards to health and explosions, mainly methane (CH_4) in coal mines. Temperature rise caused by heat release from the strata and machinery is another factor that may harm the health and safety of workers in underground mines. Control of methane and other gas components and the high temperature near a working face require overall localized ventilation management and adequate mine cooling systems. Continuously monitoring the in-situ atmospheric conditions and the number of contaminant gases, especially methane, are important factors for predicting the necessary actions for keeping the mine a safe and healthy place for workers. Studies are reported for predicting methane concentration variations inside underground mines using a long-short-term memory (LSTM) artificial recurrent neural network. Results will be compared to a simple time-series regression predictor (time-series filter). Different combinations of the variables and techniques are tested in the LSTM model to find the best results for accuracy and applicability. Forward time step variations are tested to explore the best prediction outcome. The results show that the LSTM model is limited to one-step-ahead prediction for reasonable accuracy. Furthermore, increasing the number of variables or the training window size does not seem to increase the accuracy of the LSTM predictions. Comparing the results using artificial data and the measured data from the mine, it is observed that the LSTM performs better if the data has a specific pattern and is as smooth as possible.

DEDICATION

I dedicate this work to my mother, **Edna Dias de Almeida**, and my father, **Tertulino Mascarenhas de Almeida**. They have always supported and motivated me to accomplish my goals. This work proves that their unconditional support, mentoring throughout my life, and energy spent trying to give me the best education possible were totally worth it. Additionally, I dedicate this work to my sisters, **Cleia** and **Clesia**, for their love and support in all moments of my life. ***Amo vocês!!!***

Moreover, I would like to dedicate this work to **Fernando Mai** and **Breno Mai**. They have always believed in me, and without their support and advice, I could not have possibly accomplished this one more goal in my life.

Finally, I would like to dedicate this work to my dearest friends. They have always motivated me to pursue my dreams. They are like family to me, and I am very blessed to have the best friends I could possibly ask for!

ACKNOWLEDGMENTS

Firstly, I would like to thank God for giving me the strength to finish my masters' program. Secondly, I would like to express my sincere gratitude to my advisor, **Dr. George Danko**, for his support, patience, and motivation to achieve the best results. Additionally, I would like to thank the whole committee for being patient, understanding and for always giving me valuable suggestions and comments. Furthermore, I want to acknowledge the **National Institute for Occupational Safety and Health (NIOSH)** for funding this research. Additionally, I would like to thank the whole department of Mining Engineering at the University of Nevada, Reno. As an international student, their support was vital to make me feel at home.

Finally, I would like to thank my dearest friends and family for their support throughout the program. They always believed in me and kept me motivated to achieve my goals

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.

LIST OF EQUATIONS

1. INTRODUCTION

1.1 Underground Mines Atmospheric Conditions

The intake air of underground mines has a composition of 78% nitrogen, 21% oxygen, and 1% of other gases (Mcpherson et al., 2009). The composition changes as they are mixed with gases from strata (e.g., methane and carbon dioxide) and chemical reactions, oxidation, burning of fuels, and explosives. Common toxic gases encountered in underground mines are reviewed by (Osunmakinde, 2013), emphasizing the problem that mine's gases are not easily detected by human sense. The presence of these gases, mainly methane in mine openings (Eltschlager, 2001), may cause serious hazards (Ulery,2008), especially due to gas outbursts. Another harmful factor for health and safety in underground mines is heat release, accompanied by a rise in temperature (Nie et al., 2018).

Altogether, the atmospheric conditions in the mine workings need careful ventilation design and operation. Therefore, continuous monitoring is required to check compliance with safety and health limitations in gas component concentrations and temperature and humidity. For example, section 75.320 of the (Federal Register, 2021) refers to air quality detectors and measurement devices. It deals with standards for the devices that measure methane or other gases that can accumulate inside a mine. According to this section, these devices must be maintained in permissible and operating conditions. Still,

section 75.321 deals with air quality according to the same source. This section states that air where people work or travel has to contain at least 19.5 percent oxygen and no more than 0.5 percent carbon dioxide.

Furthermore, the volume and air velocity in those areas should be sufficient to dilute and carry away flammable, explosive, noxious, harmful gases, dust, smoke and fumes. Continuous AMS (Atmospheric Monitoring Systems) are commonly used in metal and coal mines, assisting the mines in maintaining operations within safety and health limitations. A modern mine may employ an AMS network of thousands of gas, pressure, temperature, humidity, and airflow rate sensors, collecting samples every minute during the mining operation.

The amount of atmospheric data from the AMS sensors is usually evaluated against dangerous threshold crossings to assure compliance with target values for compliance. For instance, (Federal Register, 2021), section 75.323 deals with action for excessive methane. This section states that all electrical, diesel, and battery-powered equipment to be de-energized or shut off in the case where methane concentration of one percent or more is present in a working place, an intake air course, or an area where mechanized mining equipment is being installed or removed. If methane reaches 1.5 percent or more, personnel should be withdrawn from the affected area. The data collected by the sensors is far too rich in information to be used only for spot-like threshold or set-value evaluation. Trend analysis and deep learning from the variations of the AMS signals may allow using the investment value for AMS operation for early-warning

safety and health assurance not only in the present but in future time, and not only in the monitored locations but everywhere in the mine airways. An AMS data evaluation system (G. L. Danko et al., 2019) is under development. The present thesis details a study in AMS signal analysis for trend detection using artificial neural networks.

1.2 Neural Networks

As described by (Haykin, 1999); (Gurney, 1997), an artificial neural network is a machine that performs pattern recognition and perception, enhanced through "experience." All of this is done by artificial "neurons" or "processing units" that, when connected, are called synaptic weights and are used to store knowledge acquired from the surroundings. The "processing units," also called nodes by (Gurney, 1997) are connected, and when there is a large connection of nodes, it is referred to as a net. The goal of a neural network is to find the desired relationship between input and output by changing its weights (Nguyen & Widrow, 1990) . According to (Rojas, 1996) each neuron receives a certain number of signals (x_i) which depends on the number of inputs. These signals are then multiplied by their respective weights (w_i). Then all the different signals are added and integrated into the neuron. The primitive function (f) is evaluated at these points, as shown in Figure 1.

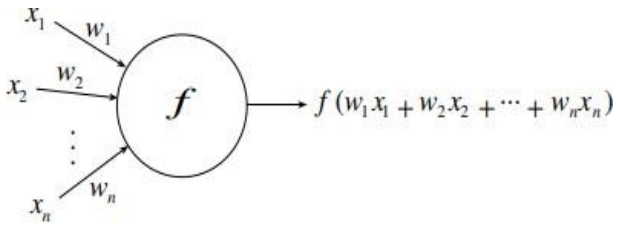

Figure 1-1 A general neuron at work. After (Rojas, 1996)

Network architecture is defined depending on how the neurons are disposed. There are three kinds of architecture: recurrent networks, single-layer networks, and multilayer networks (Mas & Flores, 2008). A layer is a group of neurons working together in a network (Kriesel, 2014). The single-layer network, sometimes also called single-layer perceptron, is the simplest form of an artificial neural network. This kind of network is mainly used in classification patterns. It consists only of a certain number of input neurons and some output neurons (output layer).

Multilayer feedforward networks usually have more than one hidden layer. When combined with the backpropagation algorithm, they are the most used and useful type of neural network (Svozil et al., 1997); (Blum & Socha, 2005). One problem with this type of network is that it results in many calculations; therefore, gradient loss and overfitting may occur (Li et al., 2019). If the error is not small enough and the predictions from the network do not match the target, an interactive process of gradient descent is applied. This process adjusts the weights and, therefore, minimizes the error.

There are many techniques for training a neural network, but supervised learning and unsupervised learning are the two most used. Supervised learning is the case where the correct response (target) is already known. The input and target data are entered into the network. It tries to find some relationship between these data, comparing input and output. On the other hand, Unsupervised learning only uses input data, and the network tries to classify it. Target information is not provided (Gurney, 1997).

Prior to using a multilayer feedforward network, it needs to be trained and tested. As described by (Svozil et al., 1997), the training process requires several steps: First, some random numbers are chosen for the weights. After that, the iteration process starts and passes all the information through all the nodes in the layers, and the result is compared with the desired target. Each complete iteration is called an epoch. Suppose the predictions from the network are not satisfactory. In that case, the network uses backpropagation and tries to adjust the weights and minimize the error. This error can measure the performance of the network's predictions. When a network is well trained, it is said to generalize. Generalization is when the trained network gives perfect or nearly perfect predictions for a new set of data that has never been used in the training set. However, overfitting may occur if the network is trained too many times. Overfitting is when the network "memorizes" the training set and consequently gives perfect predictions for the data that has been trained but fails to give good predictions for a new dataset. To overcome this problem, a sufficiently large training set must be used.

However, it is better to use a time series neural network for doing multi-step ahead predictions. They do not have the problem of long-time dependencies that feedforward networks have, caused by vanishing or exploded gradients (Diaconescu, 2008); (Lin et al., 1996). Some authors, such as (Lin et al., 1996); (Connor Les Atlas FT-lO, n.d.), have used an architectural approach to deal with long-term dependencies called Nonlinear Autoregressive models with exogenous inputs (NARX models). This model is recognized for being well suited for nonlinear modeling systems such as biological wastewater treatment and catalytic reformer in petroleum refineries (Su et al., 1992) and time series (Connor Les Atlas FT-lO, n.d.). In this type of network, outputs from the model are used as new inputs. Figure 1-2 shows a schematic where output at time instant y(*k*+1) is fed back into the network along with the input *u(k).*

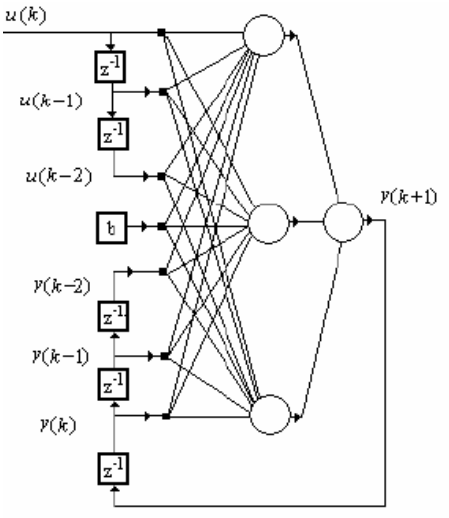

Figure 1-2Schematic of a NARX neural network. After (Diaconescu, 2008)

It was observed for this type of neural network that the more inputs it has, the better the prediction is (Diaconescu, 2008). One variant of the common NN is the

recurrent neural network (RNN) (Liang & Cai, 2020). The difference from RNN to the conventional NN is that there are connections between layers and connections between neurons within layers. . Recurrent neural networks are the most studied types of networks. According to (H. Liu et al., 2020), they possess many remarkable capabilities, such as fault tolerance, self-organizing, nonlinear function approximation, and self-learning. Furthermore, this neural network has many applications in pattern recognition, image processing, and combinatorial optimization. One type of RNN is the long-short term memory (LSTM). It is a powerful version of the conventional NN and RNN because it can deal with the explosion gradient in training (Sangiorgio & Dercole, 2020). It is also used for regression and classification problems of time series data (Liang & Cai, 2020; Wang et al., 2020)(Sangiorgio & Dercole, 2020;). It has four main components: input gate (i_g), forget gate (f_g), output gate(o_t), and a cell (c_t). These gates help the LSTM remember just helpful information across the time series data. These four parameters can be mathematically described as:

$$i_t = \sigma(\boldsymbol{W}^{i})H + b_i$$ *Equation 1*

$$f_t = \sigma(W^f)H_f + b_f$$ *Equation 2*

$$o_t = \sigma(W^o)H_o + b_o$$ *Equation 3*

$$C_t' = tanh(W^{(c)})H_c + b_c$$ *Equation 4*

$$C_t = f_t * C_{t-1} + i_t * C' \qquad \textit{Equation 5}$$

$$h_t = o_t * \tanh(C_t) \qquad \textit{Equation 6}$$

In this definition, σ is the sigmoid function, W^i,W^f, W^o, and W^c are the weights of the input gate, forget gate, output gate, and cell state, respectively. b_i , b_f , b_o, and b_c are the biases of the input gate, forget gate, output gate, and cell state, respectively. H, H_f , H_o and H_c represent the superposition of the current input vector x_t and the output h_{t-1} at the previous moment. The output is represented by h_t. Tanh is the tangent activation function. According to (Miao et al., 2020), the input gate and forget gate are responsible for updating the cell state. These are the gates that control which information the LSTM keeps or discards. The LSTM can be trained in two ways, as stated by (Sangiorgio & Dercole, 2020), with teacher forcing (TF) or without TF. Training with teacher forcing is the case where the predictions from the LSTM are not fed back into it, so new predictions cannot be made based on past predictions. The schematic is shown in Figure 1.3.

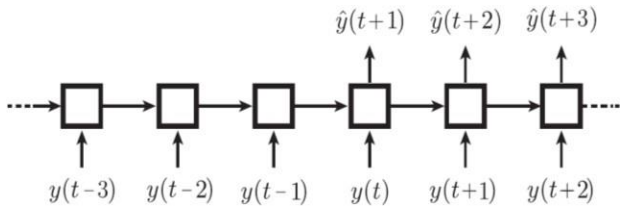

Figure 1-3 Training with TF. After (Sangiorgio & Dercole, 2020)

It is observed from Figure 1-3 that even if more than one step ahead is to be predicted, the data provided for predicting is always actual data, not the own predictions from the LSTM. This type of predicting is suitable because it does not propagate error in time.

On the other hand, training without TF is where the LSTM uses its predictions as new inputs and makes further predictions based on previous ones, as shown in Figure 1-4.

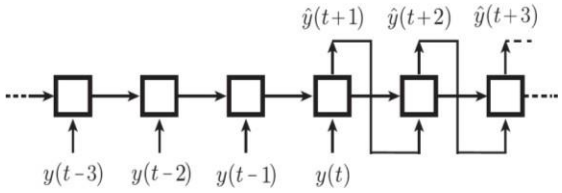

Figure 1-4 Training without TF. After
(Sangiorgio & Dercole, 2020)

This training technique allows for multi-step ahead predictions using past predicted values from the LSTM. In this case, the propagation of error in time is more significant. The value predicted at a specific time step affects future predicted values. Besides the teaching method, the parameters used for tuning the LSTM are fundamental. According to (Sarkar & de Bruyn, 2021), the so-called hyperparameters can increase or decrease an LSTM's performance. Some of the parameters are the number of hidden units, the learning rate, and the batch size.

The hidden units are related to the ability of the LSTM to recognize relationships in the data. If there are too few, the LSTM may not capture the relationship. Conversely, too many increase the chance of overfitting. It happens when the neural network tries to model (fit) every minor variation in the input. The problem

is that this variation is most likely to be noise than the actual signal (Murphy, 2021).The learning rate is responsible for the algorithm to "learn," It is considered the most important parameter by (Greff et al., 2017) (Bengio, 2012). If it is too low, the training process is prolonged and may never converge. If it is too high, instability in the training part may occur. According to (Bengio, 2012), the mini-batch size can vary between 1 and a few hundred, impacting the computation speed. Besides hyperparameters, (Bengio, 2012) emphasizes that preprocessing the data before passing it to the LSTM is also a good practice.

As stated before by (Murphy, 2021), overfitting may happen in the training of a neural network. According to the referred author, regularization is the most common way to avoid overfitting. According to (Nagarajan & L.D Babu, 2021) , the technique called regularization is used to avoid overfitting in machine learning problems. It helps prevent overfitting by introducing a penalty term to the error function. The L1 and L2 regularization are the two most common types used. The L1 regularization adds a regularization term that is the sum of the absolute values of weights. The L2 regularization adds a regularization term that is the sum of the squared value of weights. According to the source mentioned above, the L2 regularization is more computationally efficient and stable than the L1 regularization. According to (Srivastava et al., 2014), another common way to avoid overfitting is using the dropout technique. This method randomly drops out neurons (make them inactive) during training so that they contribute neither to the forward pass nor to the backpropagation (Krizhevsky et al., 2017). The idea is to prevent the units from co-adapting too much. This way, the neuron

is forced to learn more robust features. Both authors highlight that this technique significantly reduces overfitting and improves the neural network's performance. This is done in MATLAB using the function “dropoutLayer”. The description of this function is “ A dropout layer randomly sets input elements to zero with a given probability.”

According to (Siblini et al., 2020), the most common error metrics used when dealing with classification problems are precision, recall, and F1 score. The predictive performance is measured using the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives(FN). According to (Powers, 2007), true and false positives (TP/FP) is the number of predicted positives that were correct/incorrect and similarly for true and false negatives (TN/FN). The recall is defined as the true positive rate, and its notation is:

$$Rec = \frac{TP}{TP + FN}$$ Equation 7

The precision is defined as :

$$Prec = \frac{TP}{TP + FP}$$ Equation 8

According to (Siblini et al., 2020), these error metrics are biased towards a specific type of error. Therefore, more complex error metrics have been proposed, such as the F1 score.

$$F1 = \frac{2 * Prec * Rec}{Prec + Rec}$$ Equation 9

According to (Anderson, Sweeney, Williams, 2011), the error metrics used for

time-series data forecasting are the mean absolute error (MAE). It measures the average of the absolute value of forecast errors.

$$\mathrm{MAE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{j=1}}^{\mathrm{n}} |\mathrm{A_j} - \mathrm{P_j}|$$ Equation 10

Where n is the size of the dataset, A_j is the actual value, P_j is the predicted value, and j is the time step. Another common error metric is the mean squared error (MSE). It computes the average of the squared forecast error.

$$\mathrm{MSE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{j=1}}^{\mathrm{n}} (\mathrm{A_j} - \mathrm{P_j})^2$$ Equation 11

According to (Bratu, 2013), the root mean squared error (RMSE) is also used as an error metric for forecast accuracy. It is defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{n}}\sum_{\mathrm{j=1}}^{\mathrm{n}} (\mathrm{A_j} - \mathrm{P_j})^2}$$ Equation 12

2. LITERATURE REVIEW

2.1 Factors Affecting the Health and Safety of Underground Workers

2.1.1 Temperature And Heat Release

Heat release in underground mines may come from several sources. If the appropriate cooling is not provided, it can rise to uncomfortable levels. According

to (Maurya et al., 2015), the most common source of heat release in underground mines is: Geothermal gradient- refers to the rate of increase in temperature with respect to the increase in depth. In underground airways, the gradient is affected by the virgin rock temperature and airflow (dry or wet). Auto compression- When air goes downwards an airway, it gets compressed, which, in turn, generates heat. Mechanized equipment-Vehicles and devices such as rock-breaking machines, lights, and fans generate heat and thermal pollution that needs to be removed from the mine; Diesel equipment- Diesel is used in almost all the machines in an underground operation; hence, the output produced, exhaust gases, and heat play a major role in contributing to heat release and control; Explosives and blasting- When blasting is used for rock fragmentation, the heat released may stay trapped in the broken rock and released later when the rock is removed. Heat release leads to increased temperature, as shown by the case studied by (Nie et al., 2018). The author studied heat treatment and ventilation optimization in a deep mine where heat could not be discharged from the well in time, mainly because of poor air circulation. The authors used cold mine water to treat the heat damage inside the Jinqu gold mine in China.

Throughout the research is emphasized that without proper cooling, heat release leads to an increase in temperature that may affect mineworkers' physical and mental health, which decreases productivity. Another factor that leads to an increase in temperature in underground mines is the seasonal temperature variation of the surface. (Yi et al., 2019) studied the effect of seasonal air temperature variation on airflow and surrounding rocks in deep coal mines. Data

were collected during one year. At the end of this period, the authors concluded that the further from the surface, the lower the impact on increasing air temperature inside the mine caused by surface air. The surrounding rock would absorb heat during summer, which caused a decrease in underground air temperature. Studying cooling pathways in deep Australian longwall coal mines (Belle & Biffi, 2018) states that wet bulb temperature (WBT) and dry bulb temperature (DBT) are valuable indexes to evaluate risk management in underground mines. The author reports that a WBT below 28 ^{o}C can be considered cool, and productivity increases. However, at 33 ^{o}C WBT, it was reported that productivity decreases significantly, accidents happen more frequently, and the chances of a heart stroke increase dramatically. In hot conditions, miners may sweat a lot and suffer significant weight loss in their working shifts (Brake, 2001). Given the importance of keeping the temperature under adequate levels, so health and safety coupled with productivity increases, some researchers have tried to develop models to forecast temperature inside underground mines.

2.1.2 Models for Temperature forecast in Underground Mines

For instance, (Bascompta et al., 2020) developed a linear model with four variables interacting with each other, namely temperature at the beginning of the drift section, the temperature at the end of the drift section, the airflow changes along the drift section, and the length of the drift section. The authors concluded that the model presented a good correlation between measured and estimated temperature values (R^2=0.933). Furthermore, according to the research, this

model can be used to find good workplace conditions inside the mine in terms of temperature; it is simple to use, adaptable to further modifications, and allows for analyzing different mining planning scenarios. Another research that describes a predictive model for temperature in an underground mine was carried out by (W. Lyu et al., 2017). The authors developed a mathematical model based on energy and mass conservation law which considers the airflow inside the mine, rock temperature, hygrothermal exchange, temperature, and moisture. The authors concluded that the model had a good agreement with the measured data, and the airflow inside the mine was adjusted based on the predictions from the model.

Furthermore, the workforce was able to meet standard requirements. Similar research was carried out by (Zhu et al., 2015), who developed a predictive model based on heat balance for ultra-deep underground coal mines. The authors used several underground heat sources such as heat from the mine shaft, heat from the wall rock, heat from underground machinery, and other parameters to increase the model performance.

Furthermore, a computer program was developed in visual basics to speed up the calculation process. The highest standard deviation from the model prediction compared to the measured value was 1.08 ^{o}C. The authors concluded that more underground heat sources should be considered to increase the model's accuracy further. Moreover, that moisture exchange in the roadway should be analyzed carefully because it can significantly affect the temperature of the air. Some terminologies defined by (Mine Safety & Administration, 2012) define a hot place.

According to this terminology, a hot place is characterized by high air temperatures, high humidity, relatively low air movement.

Furthermore, according to this source, ventilation is the best method of combating the effect of rock temperature. In a survey made by (Donoghue, 2004), it was found that most of the heat-related illnesses occur at the face in underground metal mines, primarily because of low ventilation velocities. A sound ventilation system is required to decrease the rise in temperature caused by heat release and provide a safer and productive environment (Sasmito et al., 2013). Not only that, but a sound ventilation system also allows for bringing hazardous gases concentration, such as methane, to within acceptable limits as required by local statutes (as those described in the introduction section by (Federal Register, 2021) (Kurnia et al., 2014) (Sasmito et al., 2013). If the amount of air provided is insufficient, the potential accumulation of pollutants increases (Parra et al., 2006). However, for a ventilation system to be effective, many factors have to be considered: the ventilation route, method, the operating fan, the ventilation structure as a whole (Tang & Ding, 2011). A study made by (Donoghue et al., 2000) evaluated the heat exhaustion occurrence in deep underground mines. It was reported that to prevent heat exhaustion and decrease in productivity, it is required to provide air cooling power greater than 250 W/m^2. The best way to do this is by refrigeration and ventilation. A sound ventilation system is paramount for keeping the temperature within acceptable working levels in an underground mine. Furthermore, predicting the temperature inside an underground mine would have clear advantages, as aforementioned.

Among the software available for climate simulation underground, it is worth mentioning MULTIFLUX (Danko, 2008). According to (Danko, Bahrami & Jones., 2011), this software couples time-dependent thermal hydrologic and ventilation calculations in an underground environment. The author states that the advantage of using this software lies in the great accuracy it provides. The referred software is tested for different scenarios, such as mine drifts and development ends in coal mines. MULTIFLUX showed very good agreement between the simulations results and the measured field values in all these tests. Several papers have been published about the MULTIFLUX software, such as (G. Danko, 2010; G. Danko et al., 2020; G. L. Danko, 2013; G. L. Danko & Asante, 2017; George et al., 2011; Lu, 2019)

2.1.3 Gas Release- Methane Release In Coal Mines

Methane is one of the most common strata gases. It is produced by bacterial and chemical activity (Mcpherson et al., 2009). It is retained within fractures, voids, pores, and it is liberated when pierced by boreholes or mined openings. Despite not being toxic, its dangerous behavior lies in the fact that it can form explosive mixtures with air. In underground coal mines, methane explosion poses a deadly hazard for miners (Shi et al., 2017). The amount of coal mine methane released during mining activities poses concern about the sufficient ventilation required to ensure that work safety complies with statutory limits (C. Ö. Karacan et al., 2011); (C. Ö. Karacan, 2008)). According to the author, several geological features contribute to the release of this gas. When coal is being extracted, there is a

variation in the amount of gas released. It is not easy to distinguish where the emission is coming from. Furthermore, the author lists the parameters that most influence methane release in coal mines, among them: the level of maturation reached (rank) in a coal seam; the depth of the coal seam; face conveyor and stage loader speed, and presence of degasification. (C. Ö. Karacan, 2008) reports that due to geological, geographical, and operational factors, prediction of methane emission is difficult. For example, methane emissions into a longwall operation can be controlled, or at least mitigated, using gob gas vent holes (GGV). Therefore, understanding the circumstances under which methane accumulations can occur in a mine is crucial (NIOSH, 2006). The referred handbook emphasizes that methane entering a mine through a crack, for example, gets mixed with air and undergoes dilution. As it gets diluted, from 100% to permissible working limits, it might pass through three concentration ranges: the upper explosive limit (UEL), above 15%, where the methane-air mixture is not explosive; from 15% to 5%, known as explosive range; and the lower explosive limit (LEL), below 5%, where the methane-air mixture cannot ignite. During the explosive range, the CH_4 air mixture might be ignited. To make a methane-air mixture non-explosive, the addition of inert gas, such as nitrogen, can be used. Pressure and temperature also influence methane explosibility. The dependency of methane explosibility with pressure varies slightly under reduced pressure. At elevated pressures, the LEL decreases, and the UEL increases strongly. As for temperature, the effect is modest. The LEL at -110 ^oC is 5.6% CH_4 and at +100 ^oC is 4.8% CH_4. (Mcpherson et al., 2009). If the concentration exceeds 1% or 1.25 %, the power must be

switched off. If the concentration exceeds 2% or 2.5 %, all personnel except those concerned with the ventilation should leave the area. Chapter four of the (NIOSH, 2006) lists ways of preventing methane ignitions at longwall faces. It states that the methane released at the longwall represents only 10%-20% of the total methane emitted from the longwall panel. Despite not saying how this measurement is done, the chapter points out four ways to prevent methane ignition at longwall faces. The first one is to avoid methane accumulation around the drum of the shearer, which can be accomplished by providing better ventilation. Another action that can be taken is installing water spray behind each cutter bit to cool down the metal when it strikes the rock, preventing frictional ignitions. The third action is to ensure no eddy zones due to bad installation of the water spays around the shearer. Lastly, ensure that the methane detector in the shearer is at the best location possible.

2.1.4 Methane Release in Metal/Nonmetal Mines

Chapter thirteen of (NIOSH, 2006) lists ways to control methane in metal/nonmetal mines. It emphasizes the difference between methane release in coal mines, where the releasing process is well understood,

and the release in metal mines, where it can occur unexpectedly. Because CH_4

emissions in metal/nonmetal mines are not consistent; its anticipation is hard to be

done. One significant difference between coal mines and metal mines is that methane in coal mines is always a concern. Detection can be done anywhere, and

various instruments are used to monitor and mitigate CH_4 emissions. Therefore, monitoring is a regular activity that creates awareness of methane's potential hazards. This awareness leads to great efforts to mitigate the risk of ignition or explosion. On the other hand, workers may never encounter methane in a metal mine. They may become complacent, and CH_4 testing might be cursory, leading to methane build-up in recirculation areas.

2.1.5 Models for Methane Emission and Forecast in Underground Mines

Due to the importance of monitoring methane concentration in underground mines, several scholars have developed models for methane monitoring systems (Agioutantis et al., 2014); (Jo & Khan, 2018)). Studying the parameters that influence methane release in coal mines, like those highlighted in section 2.1.3, is essential to develop accurate forecasting models (Diaz et al., 2021). These models can be divided into three categories depending on the approach used: the first is the empirical approach, based on data collected by observing a process or phenomenon. The second category is the numerical approach, which creates physical models of the system or process and uses numerical approximation or mathematical tools to solve it. Lastly, the statistical approach, consisting of collecting and analyzing raw data and using mathematical techniques to find a pattern to create a forecast model.

2.1.6 Empirical Models for Methane Emission and Forecast

Studying the methane released from coal mines in Britain (Creedy, 1993) proposed a model based on three primary sources: methane flow from drainage systems,

methane flow in mine ventilation air, and in-situ gas content of coal seams. Data used for the study was from annual historical values (collected from 1966 to 1988). This data consists of methane emissions from deep mines in the United Kingdom. The estimate E_D is as follow:

$$E_D = LfP_w + ((1.857 * D) - (D - U) + RfP_t \quad \textit{Equation 13}$$

Where P_w refers to the annual coal production from mines without methane drainage; P_t is the total annual deep mine coal production; D is the total mass of methane drained from all mines; U is the quantity of methane utilized; L is the specific emission for mines without drainage of 6 m^3/t; R is the residual gas content of coal arriving at the surface of 2 m^3/t; f is a factor for converting volume flow to mass flow. A methane emission model was proposed by (Kirchgessner et al., 2000). It is based on a regression equation and considers the relationship between mine emissions, coalbed methane content, and coal production rate.

$$ME = 1.08x10^{-7}(C\ P\ x\ MC) + 31.44 - 26.76(DV) \quad \textit{Equation 14}$$

Where CP is the annual production of coals (tons/year); MC is the total CH_4 content of the unmined coal(m^3 CH_4/tons coal); DV is a step function: DV =1 if ($CP \times MC$) is less than 7.6×10^5, and DV = 0 if ($CP \times MC$) is greater than or equal to 7.6×10^5. Another methane emission model was proposed by (C. O. Karacan et al., n.d.). The author studied longwall methane emission where the longwall was extended from 229 to 305 meters. The equation generated was through regression analysis.

The longwall face was divided into equal lengths segments to generate the model. Data from the previous two segments were averaged and used as historical data. The trend was then extrapolated to forecast the methane emission for the segment ahead.

2.1.7 Numerical Models for Methane Emission and Forecast

According to (Diaz et al., 2021), the numerical models for methane forecast are mainly based on Darcy's Law of the porous medium. The study carried out by (Owili-Eger et al.,1970) simulated quantity and quality control in mine ventilation. Based on Darcy's Law, a mathematical model was developed to approximate the methane flow pattern and quantity coming from the coal seams and going into the mine atmosphere. The author considered that the methane emitted into the mine atmosphere depended on gas emissivity, boundary conditions, initial gas distribution pressures, and the combination of mining factors. It was emphasized that the functional relationship between these factors is unknown, making it very difficult to develop rigorous mathematical equations to simulate methane flow into the mine air. The model was tested and proved to be reasonable. However, further research showed that the model was only valid for shallow depths mines (Dixon, 1992). A numerical simulation of time-dependent methane flow was carried out by (Ediz & Edwards, 1991). The model considered a medium having variable anisotropic permeability. The solution for the equations was obtained using finite element analysis for time-dependent gas pressure evaluation. The model was based on Darcy's Law.

The results showed that the model reported good agreement with those anticipated from physical considerations. However, the author highlighted that the model's accuracy could

be improved if more field data were available. A dynamic model for methane emission in longwalls mines was developed by (Tauziède et al., 1993). The model is a function of stratigraphy and gas content of the strata surrounding the mined coal seam. The author emphasizes that due to the complex structure of a mine, it is complicated to equate the problem as a whole. The difficulty lies in the number of parameters involved and the challenge of apprehending the values for these parameters. A study carried out by (C. O. Karacan et al., n.d.) proposed a numerical analysis of the impact of longwall panel width on methane emissions. The study was divided into two stages: firstly, the author used a FLAC2D (Fast Lagrangian Analysis of Continua 2D) finite difference method to simulate the rock's geomechanical responses. The output of this first analysis was used to calculate permeability changes based on empirical relationships. The second stage used the calculated permeability field combined with Computer Modeling Groups GEM software to simulate methane flow and emissions in the longwall mining. Studying the effects of porosity and permeability changes on methane emissions in an underground coal mine (Luxbacher & Erdogan, 2009), developed a dynamic model to simulate methane emissions and concentration. The author concluded that the induced stress activities produce the highest amount of methane emission. Another alternative for forecasting methane in underground mines is using a numerical model such as the NTCF (numerical transport code functionalization) proposed by (G. Danko, 2006). Initially proposed for heat and moisture transport problems, this model can be used in combination with TOUGH2 (Transport of unsaturated groundwater and heat) or NUFT (Non-equilibrium, Unsaturated-saturated Flows and Transport) to get important information from these models and then be used for methane forecasting.

2.1.8 Time Series Analysis for Methane Forecast

According to (Diaz et al., 2021), a time series is a series of observations recorded at regular times. When applied to forecasting, past observations can be applied to predict future values of the target variables. The forecasting technique can be a regression analysis, a machine learning model, filtering algorithms, etc. For instance (Mathatho., 2020) proposed an early warning system for methane prediction in an underground coal mine. The approach consisted of a principal component analysis (PCA) to determine what factors most influence the methane levels; and an artificial neural network for methane forecast. The parameters used in the dataset were methane, humidity, pressure, wind(speed and current), and temperature. The authors obtained the dataset publicly from the “knowledge pit mining competition containing sensor data recorded in 2014 from March to June.” The PCA results showed that the most significant variables were humidity, temperature, and pressure. The authors use the Levenberg-Marquardt (LM) algorithm for training.

Furthermore, the data was normalized, and the best number of neurons for the hidden layer was chosen by increasing its initial value by one until it converged to a minimum averaged squared error. The best number of neurons using the original dataset variables was fifteen, and six using the PCA components. The authors obtained a mean squared error of 0.0019 using the LM algorithm, and a root mean squared error of 0.0426. The authors concluded that the results from the neural had a high prediction accuracy. Additionally, they stated that the PCA-based neural

network was a good alternative to find influencing factors for improving the performance of methane predictions models. The paper does not give any other information regarding the actual comparison between predicted and forecasted values. It also does not give any information about the training data size or the number of steps ahead forecasted.

(Geng, 2016) studied short-term prediction of coal mine methane using chaos partial swarms optimized radial basis function neural network model (PSO-RBFNN). The author used a three-layered feed-forward neural network with a Gaussian radial basis function to activate the hidden layers. The data used by the author comes from an underground coal mine monitoring system of the Xuan Dong Cola Mine. The first season of 2013 was used as a dataset, and the sample rate was one minute. Due to the complex environment, human factors may have interfered in the raw data, causing it to be noisy. Therefore, the data needed to be preprocessed. The author used 250 iterations and a hidden layer with five neurons. The output layer was the coal mine methane concentration. The RMSE for the proposed model was 0.0265. The author concluded that the proposed model could accurately forecast methane concentration. The author does not give information about the number of steps ahead predicted.

In an attempt to forecast methane emissions in a coal mine, (Yang et al., 2020) used a modified grey radial basis function neural network to make the forecasting. Seven factors were considered in the study, namely: coal seam depth, seam thickness, coal seam methane content, seam spacing, daily progress, daily output,

and methane emission. The data consisted of 300 sets of methane data from a coal mine in Shanxi. The authors reported that a sensitivity analysis was made to determine the number of training samples. The best value from the sensitivity analysis was 275. Data points from 276 to 300 were used as prediction samples. The relative error comparing predicted and actual values was used to evaluate the model's performance. The author used the six variables mentioned earlier as input variables to train the model. The methane concentration was the target. The data used for training went through logarithmic processing and other transformations to make it smoother and improve the predictions. For this improved model, the results showed that the relative error stayed below 15%, with an average of 5.63%. The authors attribute the network's performance to the fact that combining the grey theory with the neural network, the model weakens the randomness of the sample. Conclusions drawn suggested that more factors affecting methane should be used to improve the reliability of the predictions.

2.1.9 Attempts to Use LSTMs in Underground Mines to Improve Worker's Health and Safety

As seen in sections 2.1.6 and 2.1.7, empirical and numerical models for methane emission and forecast in underground mines have been proposed. However, in most cases, these models consider CH_4 emission per year or month, making it difficult to prevent an accident if the methane concentration exceeds permissible limits during a working shift. Furthermore, as highlighted by some authors, the number of variables interacting with each other makes it difficult to create an

accurate forecasting model. Therefore, the present study aims to develop a predictive model for methane forecast in underground mines. However, this task is evaluated using neural networks (NN), specifically a long short-term memory (LSTM). Because of its powerful capabilities, some researchers have been using LSTM models applied to underground mines to improve workers' health, safety, and productivity. Trying to forecast a sealed-off area's fire status in underground coal mines (Kumari et al., 2021) used an LSTM model to forecast the concentrations of O_2, CO, CH_4, CO2, H_2, N_2, and C_2H_4 inside a mine. The authors used a uniform manifold approximation and projection (UMAP) to reduce the dimensionality of nonlinear high dimensional values (a procedure on how this method works can be found in the author's article). Therefore, it was created a UMAP-LSTM model. This model is used in three steps: dimension reduction of the data, gas concentration prediction, and forecasting of fire status and explosibility of the sealed-off area. The author used around 80% of the data for training the LSTM and 20% to test the prediction result. The model's accuracy was calculated based on the root mean squared error. The LSTM configuration used was a hidden layer with 50 neurons, batch size of 64, a learning rate of 0.001, Adam optimizer, a dropout rate of 0.01, and 200 epochs. The data used by the author comes from sensors installed at the Lakhimata Underground Coal Mine under Eastern Coalfields Limited in India's Jharkhand state. The sampling interval was one day. 143 days were used as training dataset and 31 days as the test dataset. Several gases ratios, such as Graham's ratio, Young's ratio, CO/CO_2 ratio, Jones and Trickett's ratio, C/H ratio, and Ellicott's extension graph of Coward's explosibility

diagram were determined based on the forecasting (detailed explanation about these gases ratios can be found in the author's paper). The UMAP_LSTM model was compared to a Supporting Vector Machine and an ARIMA model. Results showed that the LSTM model was the best. Among the gases forecasted by the model, the root mean squared error for forecasting CH_4 was 0.0995.

The research done by (P. Lyu et al., 2020) presents an LSTM-based encoder-decoder for short-term prediction of CH_4 gas concentration. The data used comes from the Sijiazhuang coal mine. Before inputting the data into the LSTM, the author performed a feature extraction in the data. The correlation coefficient was used to calculate the correlation of the values of each gas sensor. This is done because, in the data used, the peak concentration for methane has a delay time if taken at different locations around the mine. The MSE was used as the loss function for the model. To prevent overfitting, an L1 regularization term was used. The data sampling interval was two minutes. Nineteen days of data were used as a training set, and ten hours of data were used as a testing set. The proposed model was used to predict methane concentration two minutes, four minutes, six minutes, eight minutes, and ten minutes ahead. The proposed encoder-decoder LSTM was compared with an ARMA model and CHAOS model. Results showed that the LSTM model was superior to the other models. The author states that despite achieving better performance, the accuracy of the LSTM model decreased as the number of time steps ahead increased. The study compares the results when data from a single sensor is used to train the model versus when it is trained with multiple sensors. For instance, the first step ahead prediction using a single

sensor presented a mean absolute error (MAE) of 0.01. Using the multi-sensor method, the error was 0.007. As the number of steps ahead increased, the error using a single sensor and multi-sensor also increased. The author highlights that using the multi-sensor dataset for training, the MAE for predicting five steps ahead was 0.0165, lower than predicting just one step ahead, 0.0169 using the single-sensor data. The author states that combining data from different sensors to train the model increases its accuracy. The paper does not give information about the architecture of the LSTM model.

3. THESIS PROBLEM

This study aims to forecast the methane concentration at the tail gate of an undergroundcoalmine. The data used for performing the tasks presented in this st udy comes from four different locations inside the mine around the tail gate. Minute-averaged variations of methane concentration, air flow rate, barometric pressure, and three other gas concentration components (CO2, CO, O2) for about a year time (460,880 sampled values for each variable) were made available for the research project. Only the methane concentrations, barometric pressures, and air flow rates were used in this thesis study. The daily averages of the measured variables were calculated by arithmetic averaging the minute-sampled measurement data for model studies involving daily average variations. The study is part of an ongoing, funded research project on an early warning system (EWS) proposed by (G. L. Danko et al., 2019). The original idea and in-depth details can be found in the author's article. This section gives a brief description of how the proposed model works and lists the goals of this study.

The early warning system proposed by the referenced author in this section considers that methane concentration traveling through an airway might accumulate (due to a distributed source). This build-up in methane concentration may cross a hazard threshold for safe working conditions. Therefore, an early warning system is required to prevent a possible disaster. Figure 3-1 below shows a schematic proposed by (G. L. Danko et al., 2019), considering the interaction between sensors and EWS evaluation for accident

prevention. This thesis focus on complementing part (D) of the schematic below. However, instead of using the dynamic mine ventilation model (DMVM) to forecast, it relies on an LSTM model. As a first approach, the LSTM model will use a combination of the available variables (airflow, barometric pressure, and

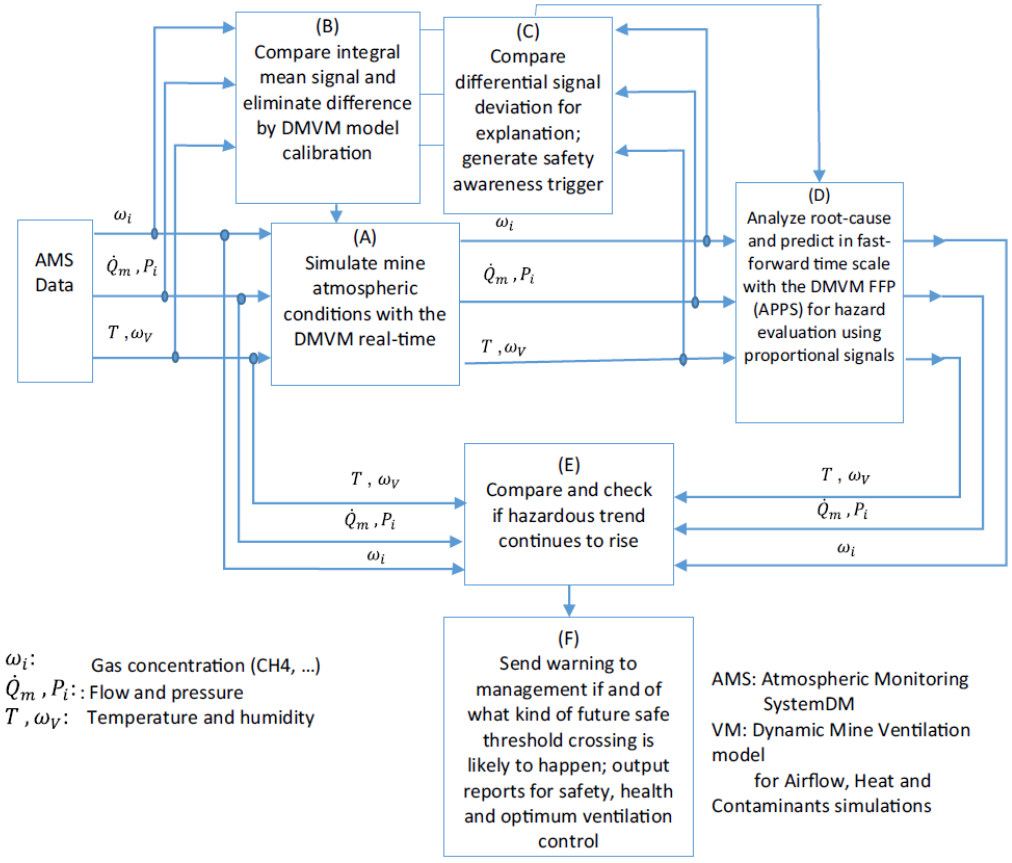

Figure 3-1 Schematic showing interaction between sensors and EWS evaluation for accident prevention. After G. L. Danko et al., 2019

methane concentration from different locations inside the mine) for training. Different combinations of the variables will be tested. The one that gives the best result will be chosen to verify its accuracy and applicability in predicting data that can match the monitored data. The accuracy of the LSTM model will be further

compared to a time series filter model. Conclusions will be drawn regarding the strengths, weaknesses, and usefulness of the LSTM compared to a simpler model.

4. METHODOLOGY

The configuration of the computer used in the proposed methodology is as follow:

Processor: Intel® Xeon(R0 CPU E3-1270 V2 @ 3.5 GHz;

Installed RAM: 16 GB

System type: 64-bit operating system, x64-based processor.

4.1 Testing the Influence of LSTM Tuning Parameters

The LSTM model in use is the one available in (MATLAB R2019a-academic use). It was adapted to reflect the needs for the simulations presented in this study. Before using the LSTM, the best tuning parameters has to be chosen. Therefore, In this section, the influence of the LSTM tuning parameters is tested. The first parameter to be tested is the number of LSTM layers. Firstly, the LSTM will use just one layer. This parameter will be increased to three with increments of one while the other parameters remain constant. Next, the number of hidden units will be tested using the best layer option from the previous simulation. The number of hidden units will start at three; It will be increased to five, ten, and twenty. Lastly, the initial learning rate will be varied using the best layer option and the best number of hidden units from the previous simulations.

Table 4-1 Initial LSTM configuration

Max Epochs	Number of Layers	Initial Learning Rate	Number of Hidden Units
500	1	0.005	3

4.2 Testing the LSTM with Artificial Data (Sine Functions)

Before using the LSTM with actual data from the mine, this step aims at validating the LSTM performance using artificial data (sine functions). This validation compares how well the LSTM can predict with actual data compared to simple data with a specified pattern. This validation allows verifying how powerful the LSTM is and seeing what data can generate good predictions.

4.2.1 Validating the LSTM with one Input

The first elementary function that is going to be tested is:

$$x = 2 * \sin\left(\frac{ti}{10 * \pi}\right) + 3 \qquad \text{input} \qquad \textit{Equation 15}$$

$$Y = \sqrt{2 * \sin\left(\frac{ti}{10 * \pi}\right) + 3} \qquad \text{target} \qquad \textit{Equation 16}$$

Where ti is a vector containing data from one to 327 points;The x vector will be used as a training input into the LSTM, and the target is the Y vector. The figure is shown below:

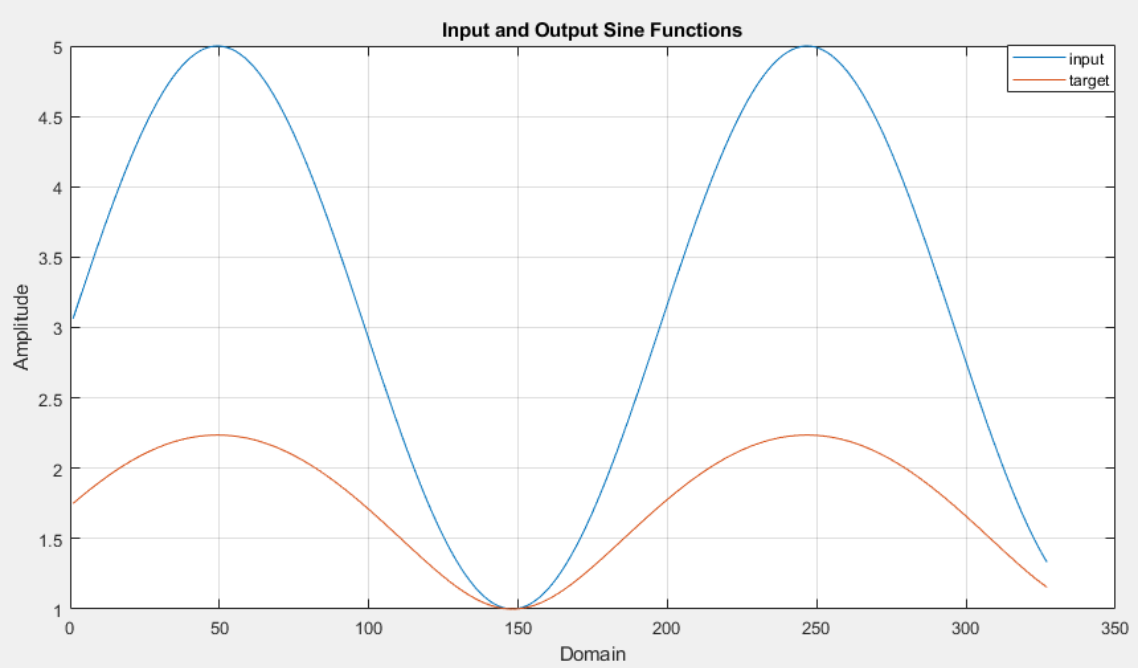

Figure 4-1 Sine functions used as training and target

The LSTM will be trained with one input, and it has to predict one step ahead. A five-step-ahead prediction will also be performed to check the robustness of the model.

4.2.2 Validating the LSTM with Artificial Data and Multiple Inputs

To test if the LSTM performs better if multiple input variables are used, it will also be trained with three inputs and predict one output. The data used in this test comes from a sinusoidal function. The functions used are defined as:

$$input\_1 = \sin\left(\frac{ti}{40 * \pi}\right) + 2$$ Equation 17

$$input\_2 = 2 * \sin\left(\frac{ti}{30 * \pi}\right) + 3$$ Equation 18

$$input\_3 = 3 * \sin\left(\frac{ti}{20 * \pi}\right) + 4$$ Equation 19

$$Output = input_1 + \frac{input_3}{input2}$$ Equation 20

The inputs and output are displayed in Figure 4-2:

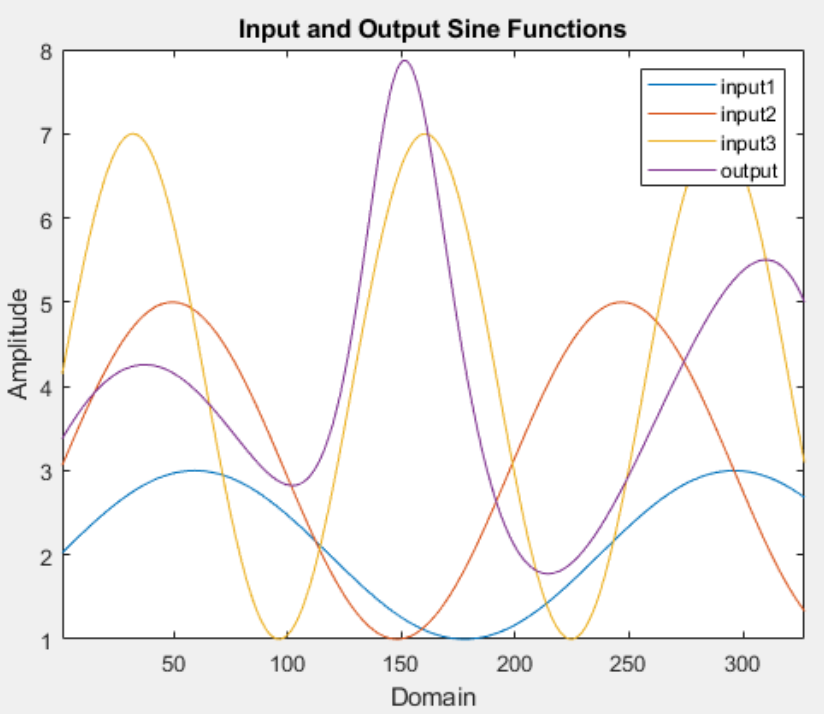

Figure 4-2 Inputs and output used for training the LSTM

According to Figure 4-2, Inputs 1, 2, and 3 will be used as training variables into the LSTM, and it has to predict the output function. Following the previous methodology, one-step-ahead and five-step-ahead will be performed to check the robustness of the model.

4.3 Data Presentation

The data used for performing the tasks presented in this study comes from four different locations inside an underground coal mine. The sensors collect the insitu data from an operating mine. Figure 4-3 shows the schematic of these locations within the mine.

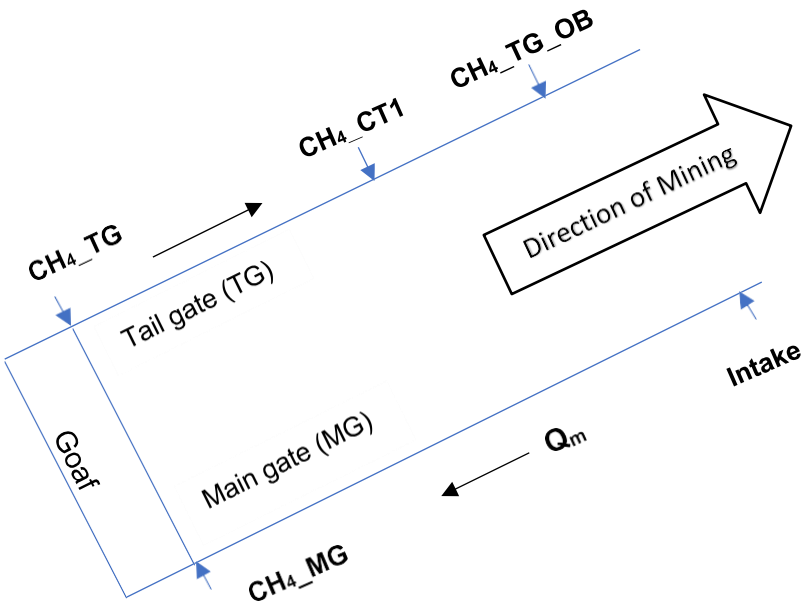

Figure 4-3 Schematic layout of AMS sensor locations inside the mine

The MG location is the intake airway path, and the TG (also called headgate) is the return airway. Qm is the airflow rate, CH4_CT1 is the methane at cross-cut, and CH4_TG_OB is the methane at the outby panel.

As discussed in section 1, the LSTM model and neural networks, in general, perform better if the data is as smooth as possible. Therefore, the data presented in the schematic in Figure 4-3 was filtered so that a comparison can be made later regarding the effect of data filtering on the results.

4.4 Data Filtering

Filtering is an important part of signal processing. According to (Marlin, 2015), the so-called "noise," a high-frequency component, can be generated by factors such

as electrical interference and mechanical vibration. It can damage the quality of the actual signal (Horowitz & Hill, 2021). Therefore, filtering aims to smooth the signal and remove incorporated noise by reducing its amplitude to zero (Marlin.,2000; Koderstani,Xiang, & Ye., 2018). In order to deal with noisy data, a digital or an analog filter can be used (Smith,1999). These types of filters are used for signal separation and signal restoration. Digital filters can be separated into finite impulse response (FIR) and infinite impulse response (IIR). One of the most common and straightforward types of filter used in the industry is the moving average filter (MAF), because of its simplicity and capability to attenuate the noise (Kordestani et al., 2018); (Smith,1999 The drawback of this filter is that it is not suitable for frequency-domain signals because it has a poor capability of separating bands of frequencies. It is an FIR, and its mathematical formulation is:

$$Y_i = \frac{1}{M}\sum_{j=0}^{M-1} X(i+j) \qquad \text{Equation 21}$$

Where Y_i Is the signal output at the ith step, $X(i+j)$ is the signal input from the ith step to the jth step, and M is the number of points in the average. Several other IIR and FIR filters are investigated by (Smith,1999).

In this section, a novel FIR low pass filter is used. In summary, what this particular filter does is: for a set of data, for example:

$$S = \{p_1, p_2, p_3, p_4, p_n\} \qquad \text{Equation 22}$$

It applies the least square fitting technique to a subset T of the data set S. The length l of this subset is user-defined, and it consists of a moving window. For instance, if l is defined as three, then the first subset would be:

$$T_1 = \{p_1, p_2, p_3\}$$ Equation 23

The second subset would be:

$$T_2 = \{p_2, p_3, p_4\}$$ Equation 24

And so on, until the last element of S. For each subset, the least square fitting is applied in the form of a linear model using the 'polyfit' function from the software MATLAB R2019a-academic use.

$$y = a * x + b$$ Equation 25

After the 'polyfit' function finds the best coefficients for the equation, the 'polyval' function is used to evaluate the polynomial at the points going from zero to l. The result of y is stored in a matrix M. For instance, if S is equal to six and l is equal to two, matrix M would have the following format:

$$M = \begin{pmatrix} y_{1,1} & 0 & 0 & 0 & 0 \\ y_{2,1} & y_{2,2} & 0 & 0 & 0 \\ 0 & y_{3,2} & y_{3,3} & 0 & 0 \\ 0 & 0 & y_{4,3} & y_{4,4} & 0 \\ 0 & 0 & 0 & y_{5,4} & y_{5,6} \\ 0 & 0 & 0 & 0 & y_{6,6} \end{pmatrix}$$ Equation 26

Where each column of M contains l points, zeros are placed where the following conditions are true:

$$M_{ij} = 0, if\ j > i\ or\ if\ j + l > i$$ Equation 27

Once matrix M is completed, each row is averaged by l generating the filter output with the reduced frequency component, that is, the daily variation. Besides filtering the data, this same filter will also be used to make predictions. The difference is that for predicting one point ahead, the best-fitted points (for example $y_{3,1}$), the 'polyval' function evaluates the generated polynomial curve at the point $l + 1$.

The example below shows the lines generated by the filter with length l equal three along a section of the original curve for the methane concentration at location MG.

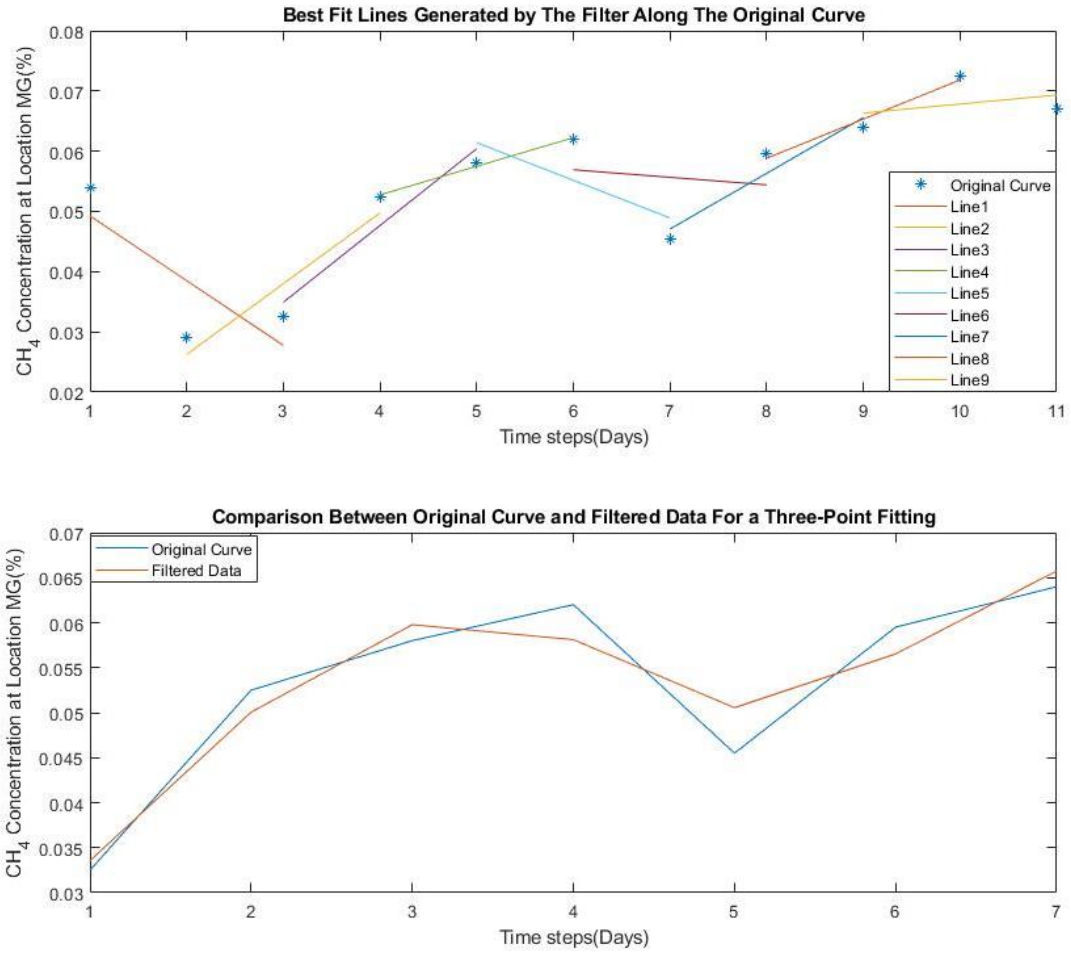

Figure 4-4 Section of a curve showing the lines generated by the filter for a three-point fitting (top figure); smoothed curve generated by the filter (bottom figure).

The smoother curve is generated after averaging each line by the filter length, as shown at the bottom of Figure 4-4. The smaller the "filter length " size," the better the filter's fit values. This is true because the high-frequency component is not being filtered so strongly. If l is equal to two, the best fit lines generated by the filter will match exactly each point of the original curve. However, this filter configuration with l equal to two cannot be used because it would not smooth the original curve. The high-frequency components would not be eliminated because the averaged

values generated by the 'polyval' function would coincide exactly with the original values of the curve.

Figure 4-5 (top) shows a complete example comparing the filtered (filter length equal three) and original daily average CH_4 concentration at locations TG. The two subsequent figures (middle and bottom) show the absolute and relative errors between the original and filtered data. The plot shows that using a filter length equal to three makes the filtered and original curves almost identical. The filter was applied to smooth the data in the different locations in fig Figure 4-5. The goal of the filter is to make the data smoother. However, depending on the number of points used as the "filter length" (points used to find the best fit), the outcome may be a distortion of the original data, causing an increase in relative and absolute errors. That is why a filter length of size three was used, even though it seems that it has no effect. In some cases, a comparison will be made between the original data and the filtered data. The goal is to verify if filtering the data improves the predictions or destroys the daily variation component (even when a small filter length is used).

.

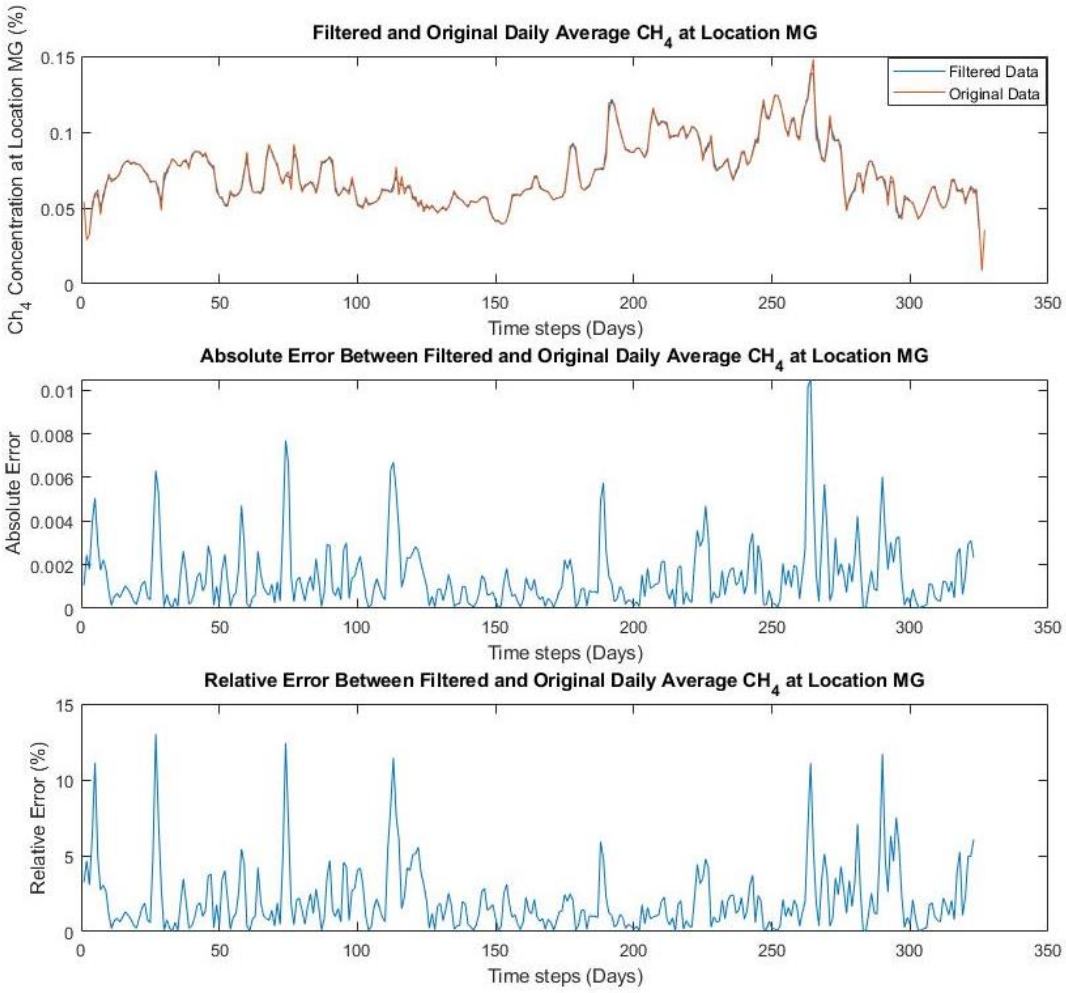

Figure 4-5 Comparison between filtered and original curve at location MG.

The error metrics used in this study are the absolute error and the relative error. The values are normalized.

$$E_{ai} = \frac{(v_{fi} - v_{oi})}{mean\ (CH_4^{TG})} \qquad \text{Equation 28}$$

Where E_{ai} is the absolute error for each time step, v_{oi} is the original measured value and v_{fi} is the forecasted value. The normalization occurs by dividing the absolute error by the mean value of the methane concentration at location TG.

The relative error is defined as:

$$E_{ri} = \frac{E_{ai}}{v_{oi}} * 100 \qquad \text{Equation 29}$$

Where E_{ri} is the relative error for each time step. The mean squared error (MSE) and the root mean squared error (RMSE) will also be used to evaluate the LSTM performance.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(v_{oi} - v_{pi})^2 \qquad \text{Equation 30}$$

Where n is the number of data points and v_{pi} is the predicted value from the LSTM.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(v_{oi} - v_{pi})^2} \qquad \text{Equation 31}$$

Besides the error metrics aforementioned, the performance of the LSTM will also be evaluated using a probability histogram plot. The histogram was generated in MATLAB using the function " histogram" with the property "normalization" set to "probability."

4.5 Tasks

The data coming from the mine will be used as training variables and as a target for training an LSTM and a time-series filter model. To better understand the behavior of the LSTM and its usefulness, a set of tasks are tested:

4.5.1 Testing the Importance of Each Training Variable

This section will test if a specific training variable (e.g., airflow rate) has a more positive effect on the predictions than the methane concentration at location MG, for example. There are four variables available for training. Each one will be individually tested to see how well the LSTM performs. The two variables that give the best result will be used as training variables to see if the accuracy of the predictions is improved. The variables available for training and prediction can be seen below:

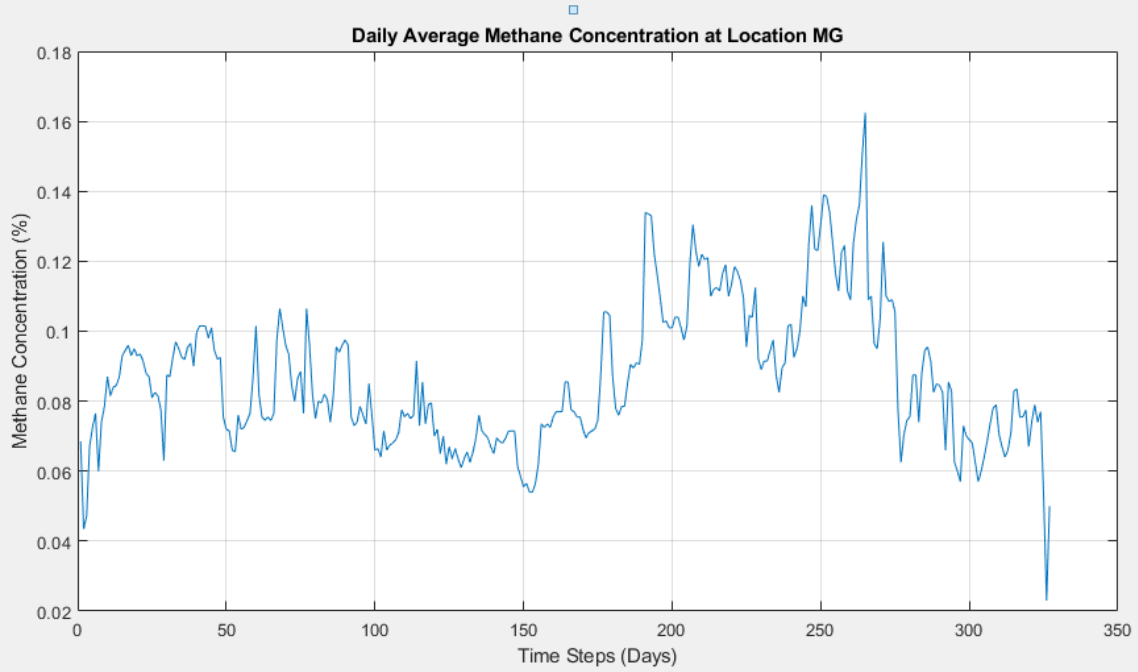

Figure 4-6 Methane Concentration at Location MG

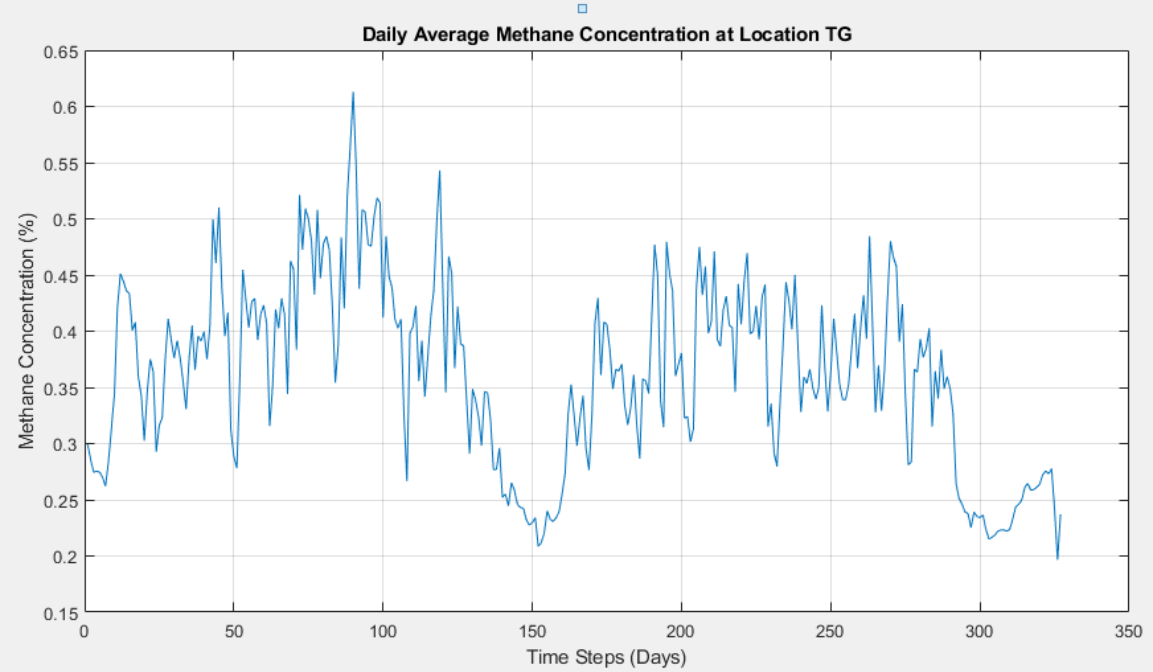

Figure 4-7 Methane Concentration at Location TG

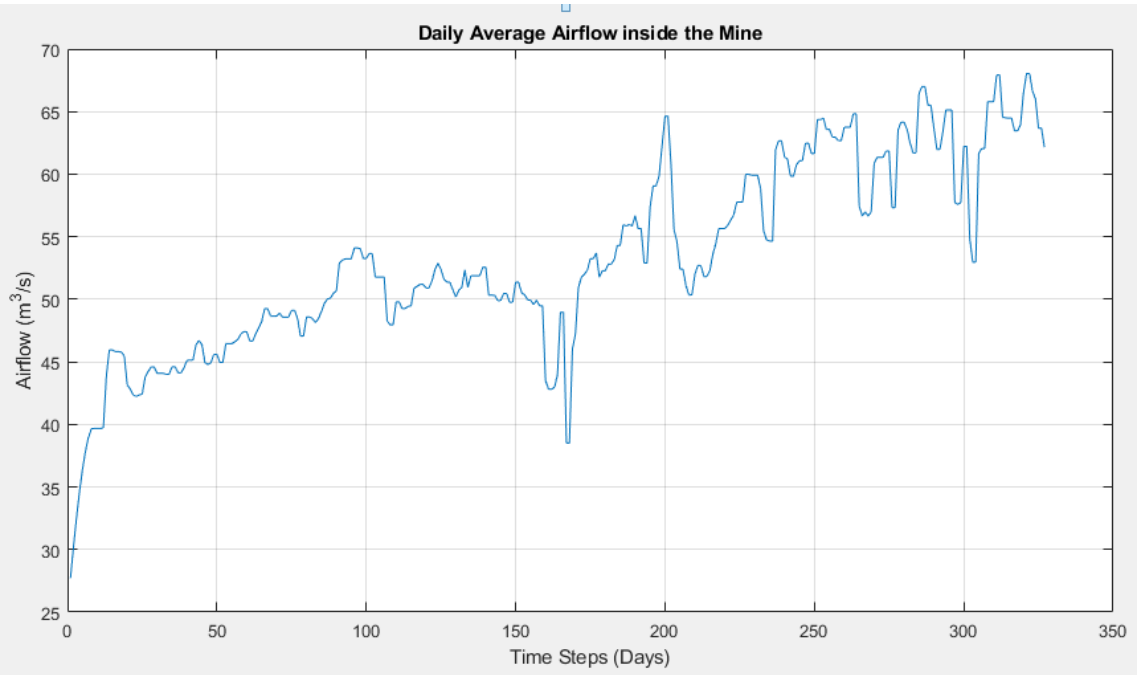

Figure 4-8 Airflow inside the mine in the intake location

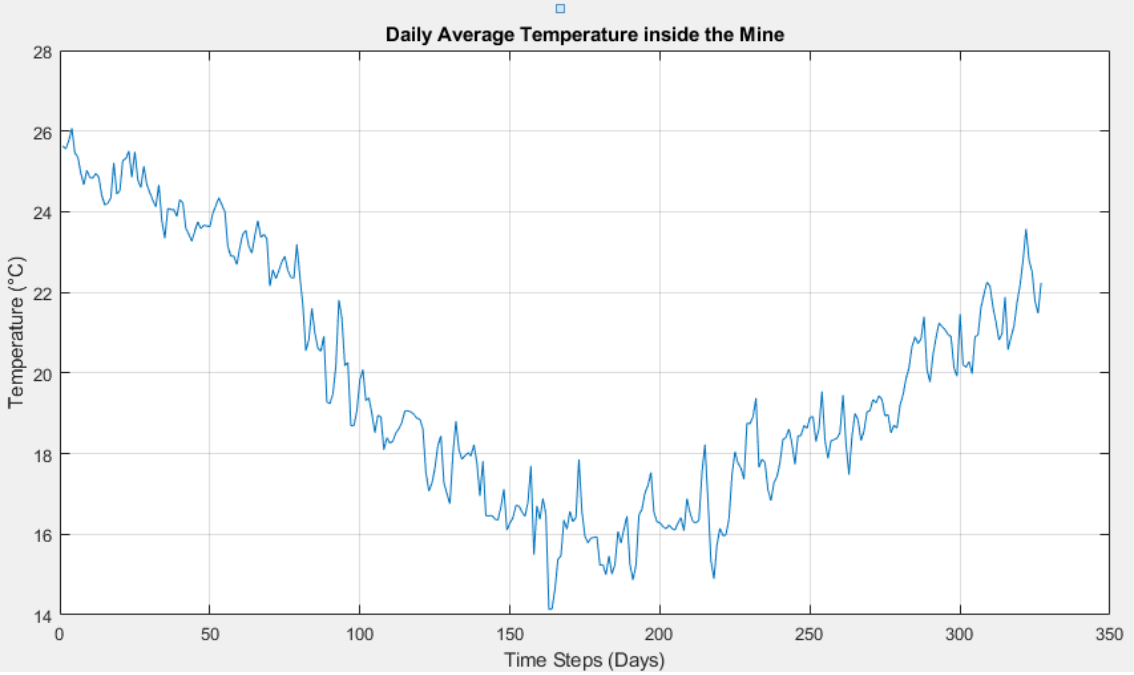

Figure 4-9 Temperature inside the mine at the main gate

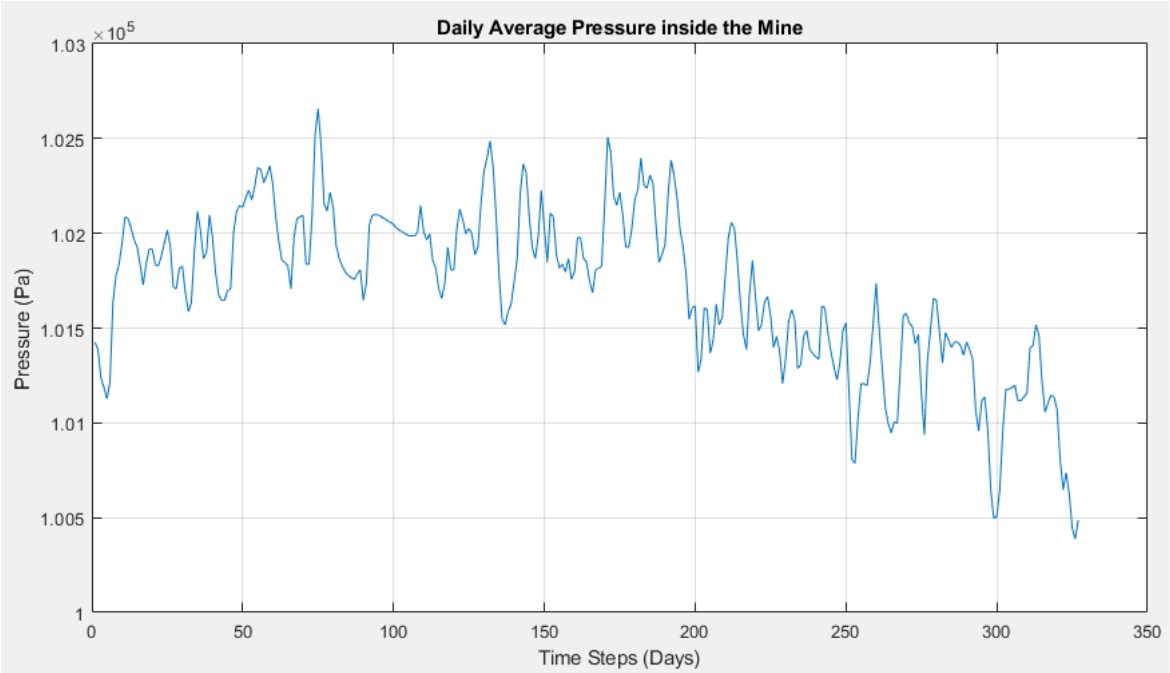

Figure 4-10 Pressure inside the mine at the main gate

The data presented from Figure 4-6 to Figure 4-10 is a daily average. The data in each figure has 327 data points. The four variables available for training (methane concentration at the main gate, airflow rate, temperature, and pressure) were chosen because they are related to the variation of methane concentration at the tailgate.

4.5.2 Direct Forward Prediction of Methane Concentrations With (A) LSTM Model And (B) Time-Series Filter Model

Using the best combination of variables from the previous section, this section aims to see how well the daily average data can be predicted for the next day. Task (b) is based on a 'polyfit' function from MATLAB, and it will be used for the last two (filtered or original) data points. The goal is to fit a 1st-order polynomial function and forward predict the next new point(s) with this 1st-order predictor. This scheme will be used from day 2 through day 324 and forward predict consecutive days (3,4,5), (4,5,6), and so on, through days (325,326, 327). Next, the predicted curves will be compared with the measured curves. As an additional step, it will be checked if the forward prediction is better when using the original or the filtered curves.

The LSTM will be trained in a sliding window fashion. This method is used by (Danko, 2021) and (Hota et al., 2017). Although the latter author uses it for another purpose and in a different style (it gets the average at the end of the window), the concept of “sliding the window” is the same.

A simple example of how this method works (for ten data points) is presented below. Consider that the data set has ten data points.

Table 4-2 hypothetical data points for a sliding window scheme

1	2	3	4	5	6	7	8	9	10

Furthermore, consider that this data set will be used as a training variable (green blocks) for the LSTM. Additionally, consider that the user has defined that the size of the sliding window is five.

Table 4-3 Schematic showing training data in the sliding window

1	2	3	4	5	6	7	8	9	10

In this case, The LSTM will use the first five data points for training, and it has to predict the next point (red block). It is worth mentioning that despite having ten data points, only five data points are passed to the LSTM, and the remaining data points are "hidden" from it. This process of "hiding" the remaining points is necessary; otherwise, the LSTM would "see" the whole data set, and that would be "cheating" since it has the chance of memorizing the values. Continuing with the proposed training, the window slides to the right, and a new set of five data points is created.

Table 4-4 Schematic showing new set of data after the window slides

1	2	3	4	5	6	7	8	9	10

As the window slides to the right, it carries historical values from the previous data points, which might help predict the next point without overloading the LSTM with unuseful information. This sliding window process is repeated until the last value of the data set is forecasted.

Table 4-5 Schematic showing the process of the sliding window for ten data points

1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10

After the LSTM is trained with five points and forecasts the next point, the state of the LSTM is reset. Its values for weights, bias, and information about that training are forgotten. This is done using the “resetState(net)” function in MATLAB. The window slid five times for the simple example shown here. Therefore, it is as if five LSTMs were used in this process. It is important to mention that despite using the “resetState(net)” function, the configuration (hyperparameters) of the LSTM remains the same. The number of layers, number of hidden units, learning rate, and training algorithm are the same regardless of how many times the window slides. What is erased from the LSTM is just the constants used to model the five data points. It is worth mentioning that the size of the window is user-defined.

The goal is to see how well the LSTM can predict at least one step ahead and verify its sensitivity to the training window size.

The 1st null hypothesis compares a simple 'polyfit' forward prediction scheme (explanation of how this scheme works is found in section 4.4) and sees if it is as good or better than the LSTM prediction (for at least the 1st next day).

4.5.3 Forward-Prediction of Root-Cause CH_4 Sources With (a) LSTM Model And (B) Time-Series Filter Model

The transport model used in this section has a complex structure. A full description can be found in the book 'Model Elements and Network Solutions of Heat, Mass and Momentum Transport Processes' written by George Danko (Danko and Springer-Verlag Gmbh, 2017).

In this step, the CH_4 line source responsible for the CH_4 increase in methane concentration will be calculated (MG to TG).

$$QCH_4 = (CH_4^{TG} - CH_4^{MG}) * Q_m \qquad \text{Equation 32}$$

After the predictions are made, the CH_4^{TG} will be back-calculated:

$$CH_4^{TG} = \frac{QCH_4}{Q_m} + CH_4^{MG} \qquad \text{Equation 33}$$

This exercise will include two influencing factors: the CH_4 differences between CH_4^{MG} and CH_4^{TG} And the Q_m flow rate of the air. The same exercises as in (1) will be done, showing the QCH_4 sources back-calculated from the measurements and the predicted values from methods (a) and (b).

4.6 Testing the Influence of the number of Variables Used for Training.

This section will increase the number of training variables to three, based on section 4.5.1. The goal is to verify if increasing the number of training variables there is an increase in the performance of the LSTM. This assumption is based on the fact that maybe the LSTM can extract more relevant information between these variables and the target.

4.7 Testing the Influence of Training Data Size

This section aims at testing the influence of training data size. The goal is to verify if, after increasing the training window size, the LSTM can capture more easily a longer-term dependency between the data and perform better when making the predictions. So far, the tests considered that the LSTM would be trained with five data points. In this section, the proposed number of time steps trained will be ten, fifty, and one hundred data points.

4.8 Testing the Influence of the Number of Steps Ahead Predicted

The data used in this simulation is a methane concentration daily average. However, many things can happen throughout the day that will influence the methane concentration for the next day. Therefore, making the predictions using the minute data or a five-minute average data would be more advantageous. This section aims at verifying how well the LSTM can make the predictions if more than one step ahead is considered. The five minutes average data will be used, and five

steps ahead will be predicted. In this case, every prediction will correspond to 5 minutes ahead. Thus, a total of 25 minutes will be predicted. This test will be done considering all the best-case scenarios from the previous sections. It will consider the best number of training variables, the best training data size, and the best tuning parameters for the LSTM. The test will be done using the unfiltered data.

4.9 Considering that the Next Day Prediction is the Previous Day Value

In this section, the assumption is that the prediction for the next day is the previous day value. This simple case will be compared against the LSTM best case from the previous sections.

4.10 Testing the LSTM Against Overfitting

In this section, the chance of overfitting is taken into account. As discussed in the introduction, L2 regularization and dropout layers avoid overfitting. Therefore, the best case scenario from the previous sections will be re-run in this section. Considering that overfit may have taken place in this best-case scenario, the training is done in this section will use an L2 regularization equal to 0.005 and a dropout layer equal to 0.33.

5. RESULTS

After all results are obtained considering the proposed methodology, the figures presented in this section will be discussed in deep in the “Discussion” section.

5.1 Influence of LSTM Tuning Parameters

The tests carried out in this section used the methane concentration at location MG and the airflow rate as training variables. The predictions were made using the unfiltered values.

Table 5-1 Testing LSTM layers

New Values Tested					
Number of Layers	Learning Rate	Number of Hidden Units	Time spent for predicting (minutes)	MSE	RMSE
1	0.005	3	20	0.004307	0.0656
2	0.005	3	26	0.003057	0.0553
3	0.005	3	30	0.003117	0.0558

From table 5-1, it is observed that two layers give better performance than one layer. However, there is the penalty of spending more time to make the predictions. Based on the results from table 5-1, the LSTM will now be trained with two layers and keep the learning rate at 0.005, and the number of hidden units will be increased to five, ten, and twenty.

Table 5-2 Testing LSTM number of hidden units

New Values Tested

Number of Layers	Initial Learning Rate	Number of Hidden Units	Time spent for predicting (minutes)	MSE	RMSE
2	0.005	5	29	0.004298	0.0655
2	0.005	10	29	0.005318	0.0729
2	0.005	20	30	0.005750	0.0758

Based on tables 5-1 and 5-2, the best configuration so far is having the LSTM with two layers and three hidden units. Now, using this configuration, the LSTM will be further tested, varying the learning rate.

Table 5-3 Testing LSTM number of hidden units

New Values Tested					
Number of Layers	Learning Rate	Number of Hidden Units	Time spent for predicting (minutes)	MSE	RMSE
2	0.004	3	27	0.003686	0.0607
2	0.0001	3	29	0.004136	0.0643

2	0.001	3	29	0.002472	0.0497
2	0.01	3	29	0.003862	0.0622
2	0.0005	3	28	0.002277	0.0477
2	0.0009	3	29	0.002333	0.0483
2	0.0003	3	28	0.002737	0.0523
2	0.0007	3	29	0.002195	0.0469
2	0.0006	3	28	0.002306	0.0480
2	0.00075	3	29	0.002389	0.0489
2	0.00065	3	28	0.002240	0.0473
2	0.00068	3	29	0.002125	0.0461
2	0.00069	3	29	0.002308	0.0480

Combining the results from tables 5-1,5-2 and 5-3, the best configuration is having the LSTM with two layers, three hidden units, and a learning rate of 0.00068.

5.2 Training the LSTM with One Input Using Artificial Data

This section presents the results of the methodology described in section 4.2.1. The proposed method considers that the LSTM will be trained with simple sine waves curves to check how robust the model is. The assumption is that if the LSTM

cannot forecast trivial sine waves with a specified pattern and no noise, it is most likely that it will not be able to accurately forecast real data from the mine, which has no specific pattern and has a lot of noise. For further comparison, the LSTM forecasts the sine waves using the best tunning parameters achieved using measured data from the mine. The results can be seen in the figures below.

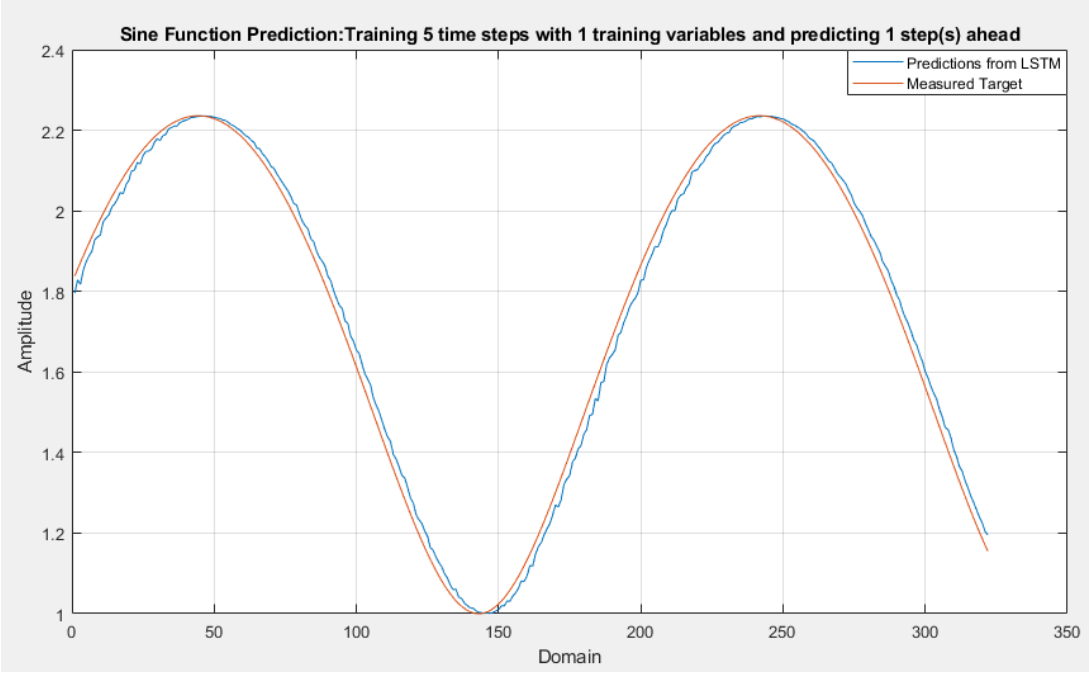

Figure 5-1 LSTM one-step-ahead forecast for the sine function

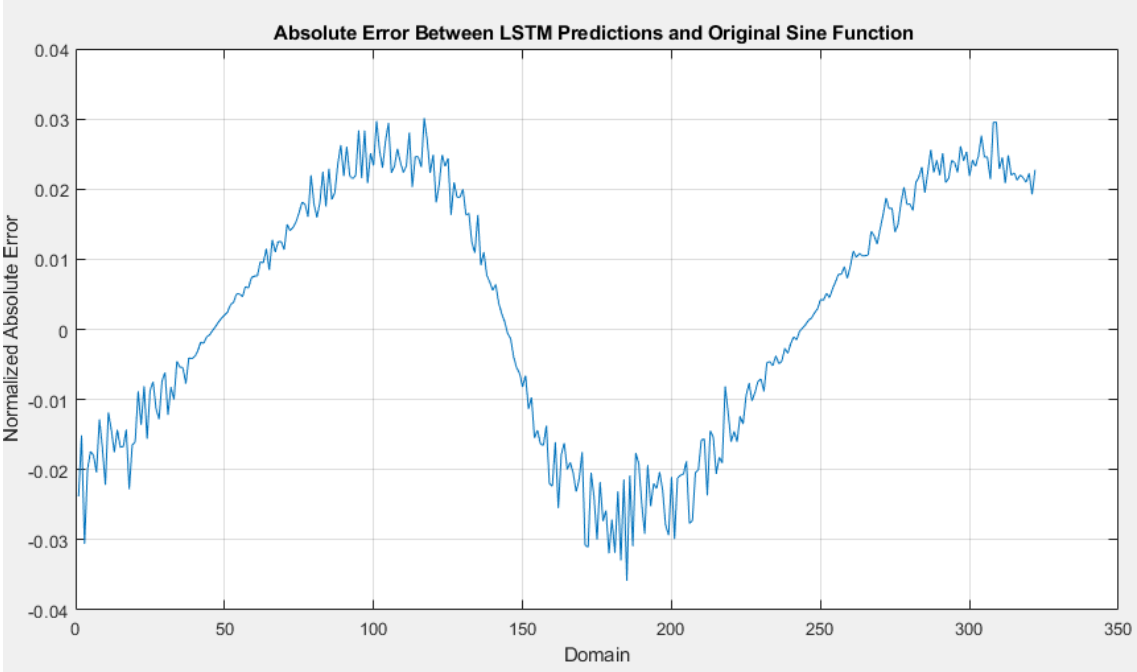

Figure 5-2 Absolute error from the LSTM one-step-ahead forecasting the sine function

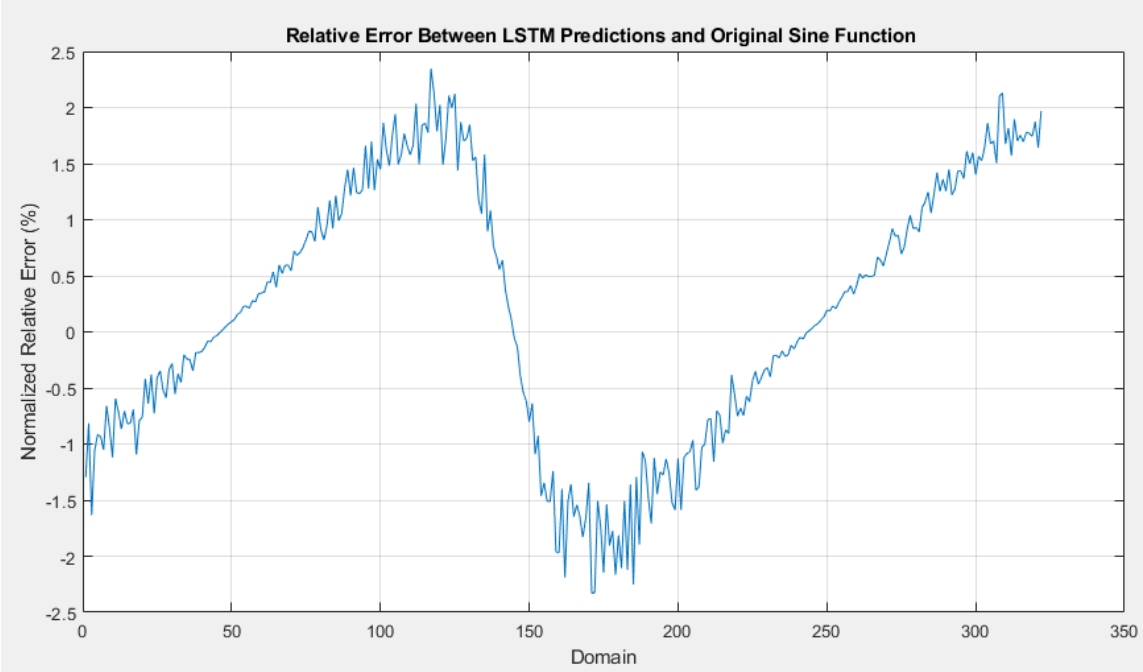

Figure 5-3 Relative error from the LSTM one-step-ahead forecasting the sine function

The MSE and RMSE for this prediction are 0.001024 and 0.0320, respectively. The time spent for training and predicting was 29 minutes. The same data was used; however, this time, five predictions ahead were made with a training window size of five.

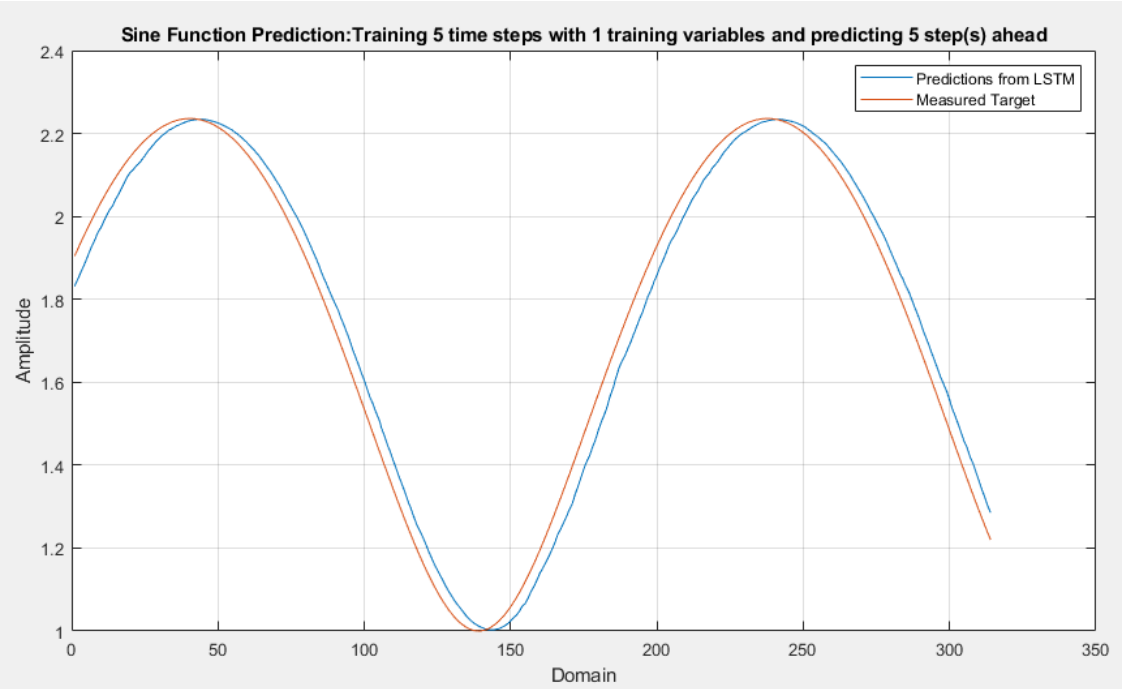

Figure 5-4 LSTM five-step-ahead forecast for the sine function

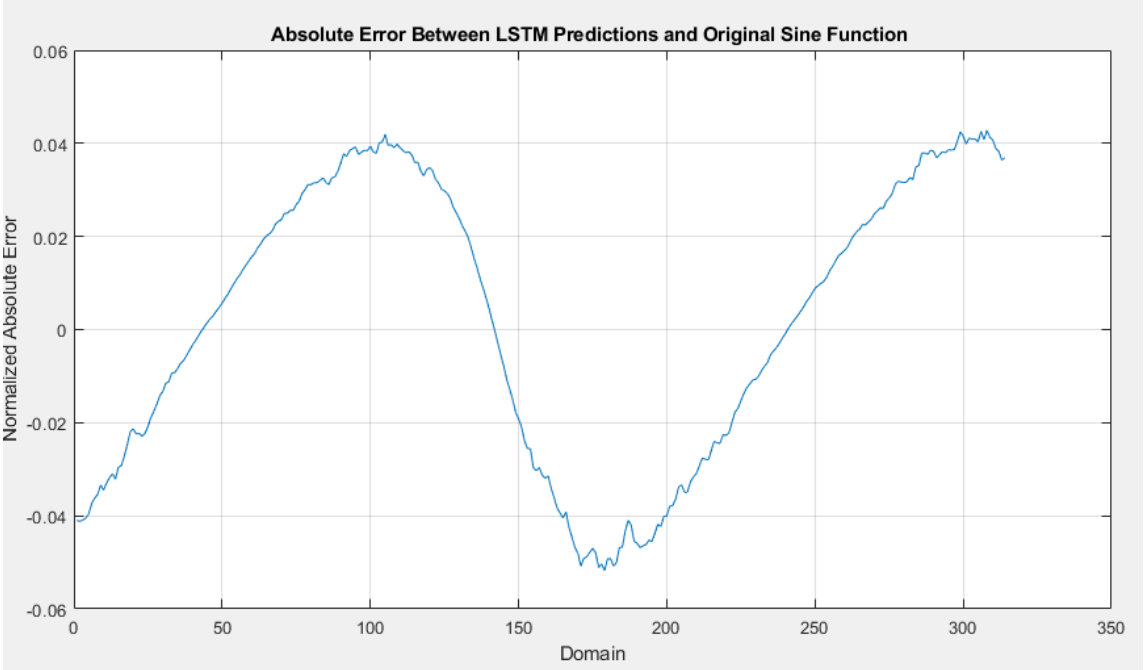

Figure 5-5 Absolute error from the LSTM five-step-ahead forecasting the sine function

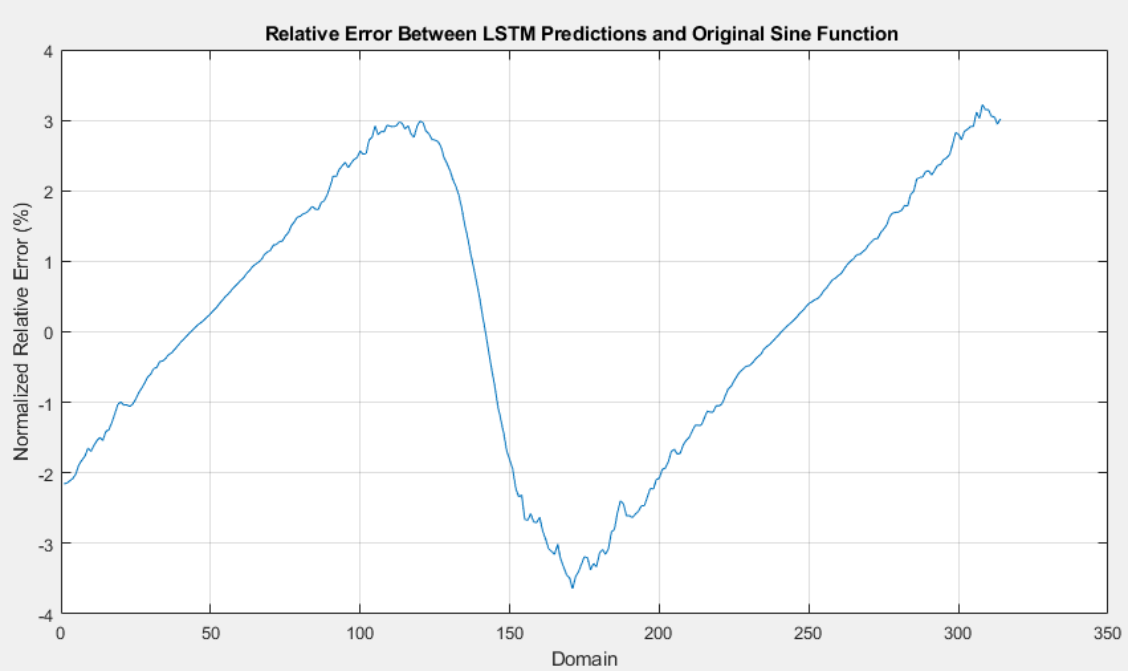

Figure 5-6 Relative error from the LSTM five-step-ahead forecasting the sine function

The MSE and RMSE for this forecast are 0.002831 and 0.0532, respectively. The time spent for training and predicting was 29 minutes.

5.3 Training the LSTM with Artificial Data and Multiple Inputs

The tests shown in this section follow the methodology described in section 4.2.2. The predictions are shown below:

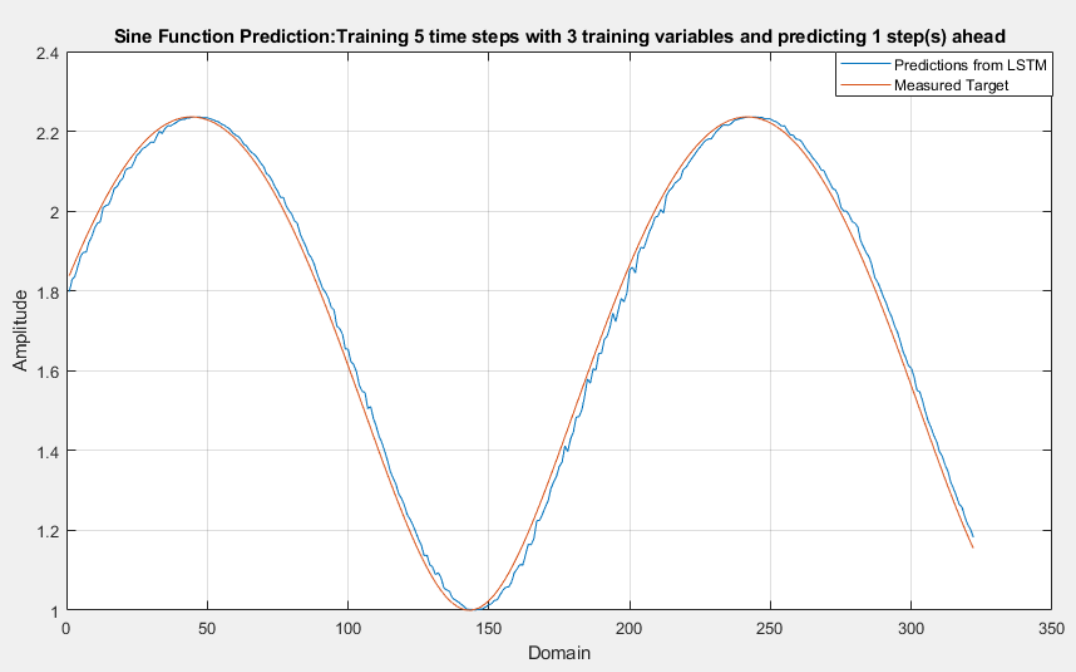

Figure 5-7 LSTM one-step-ahead forecast for the sine function with three training variables

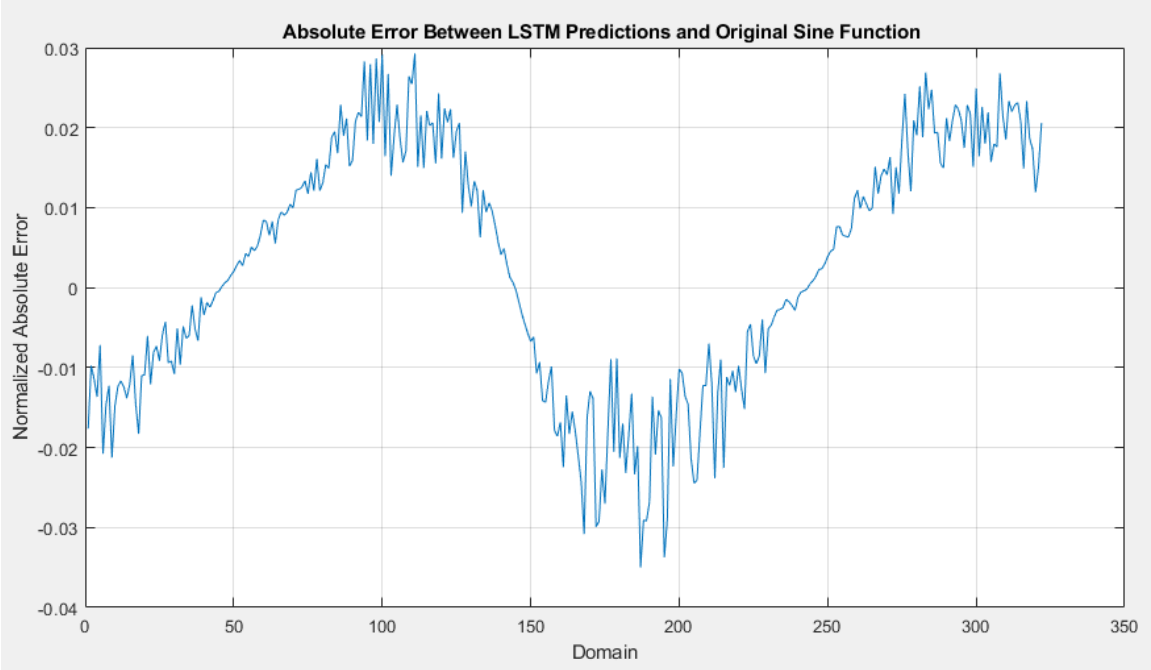

Figure 5-8 Absolute error from the LSTM one-step-ahead forecasting the sine function with three training variables

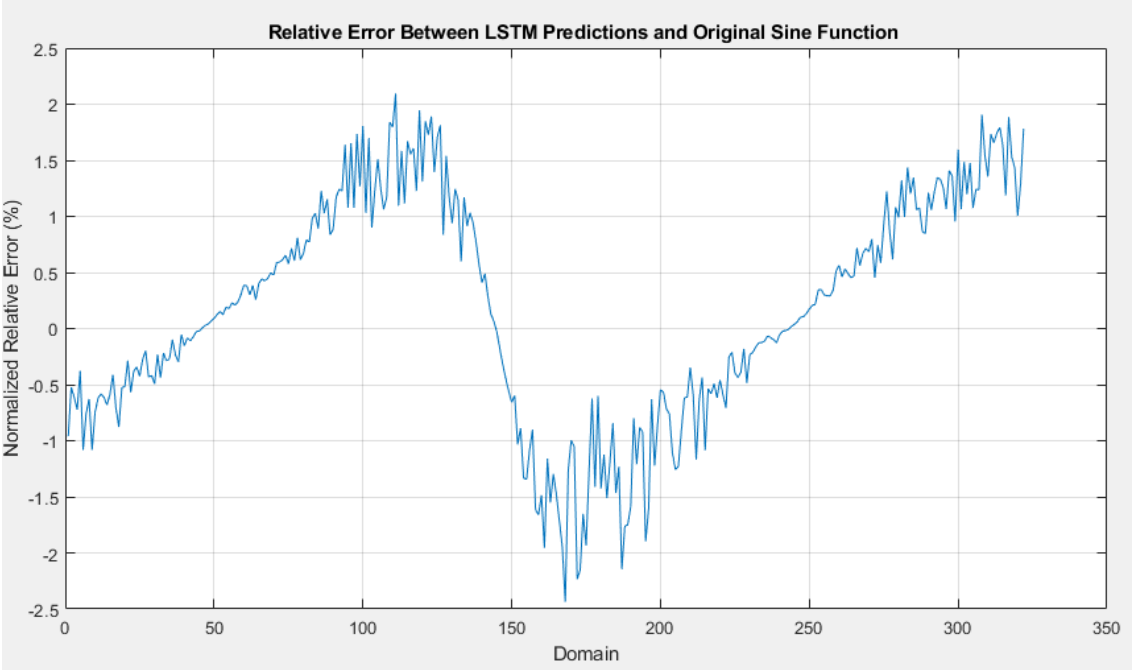

Figure 5-9 Relative error from the LSTM one-step-ahead forecasting the sine function with three training variables

The MSE and RMS for this prediction are 0.0007604 and 0.0276, respectively. The time spent for training and predicting was 28 minutes.

The same data set was used; however, this time, five predictions ahead were made.

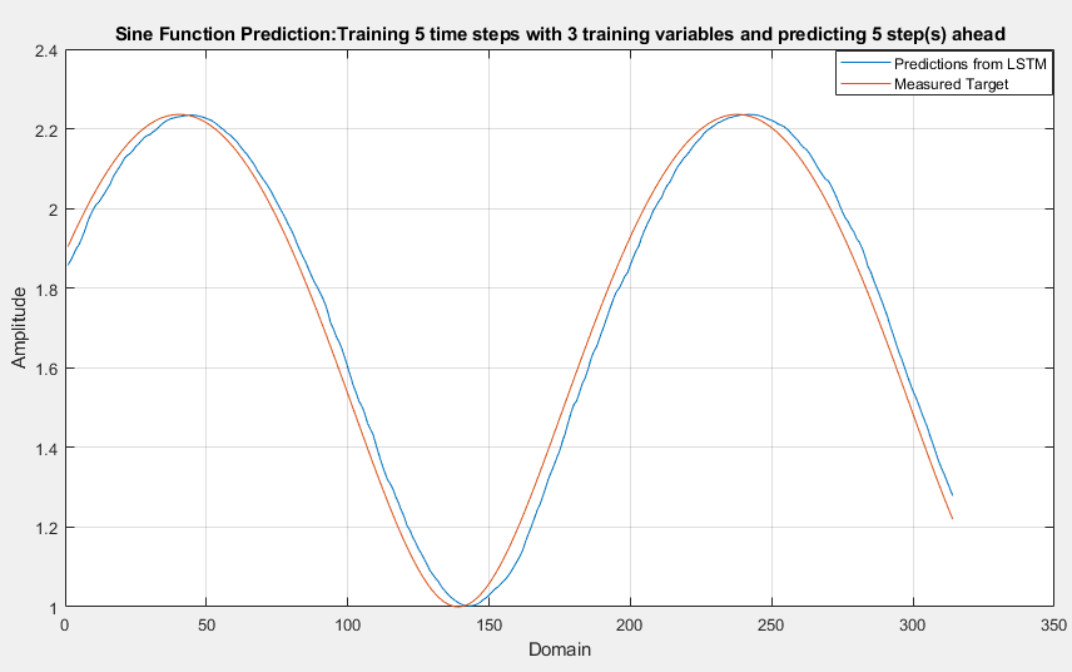

Figure 5-10 LSTM five-steps-ahead forecast for the sine function with three training variables

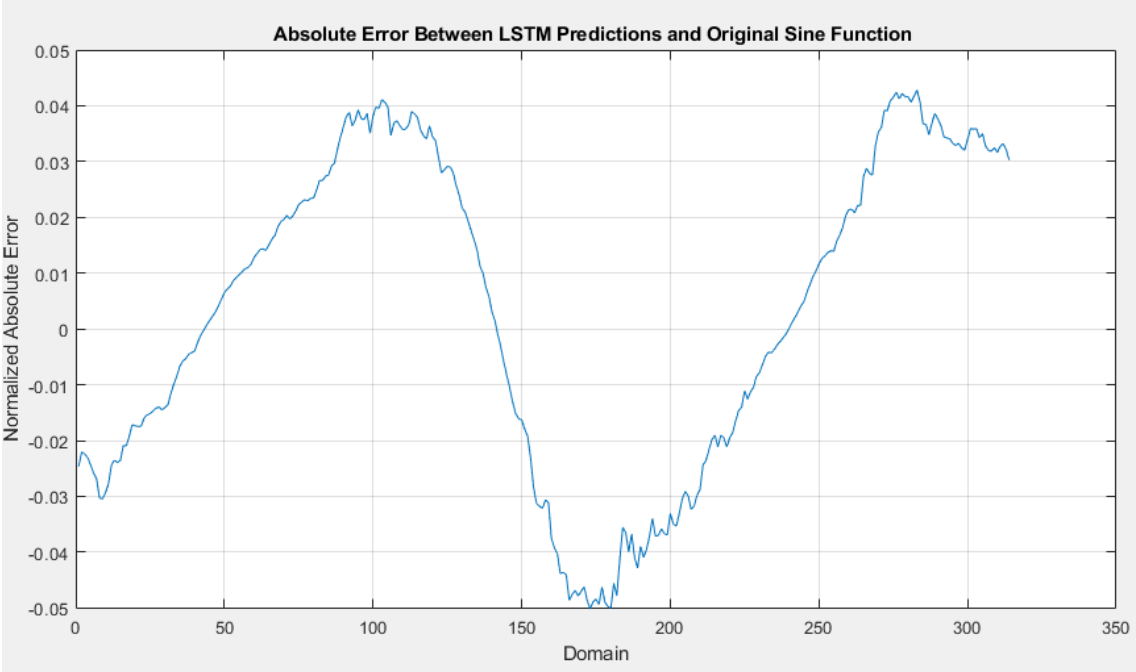

Figure 5-11 Absolute error from the LSTM five-steps-ahead forecasting the sine function with three training variables

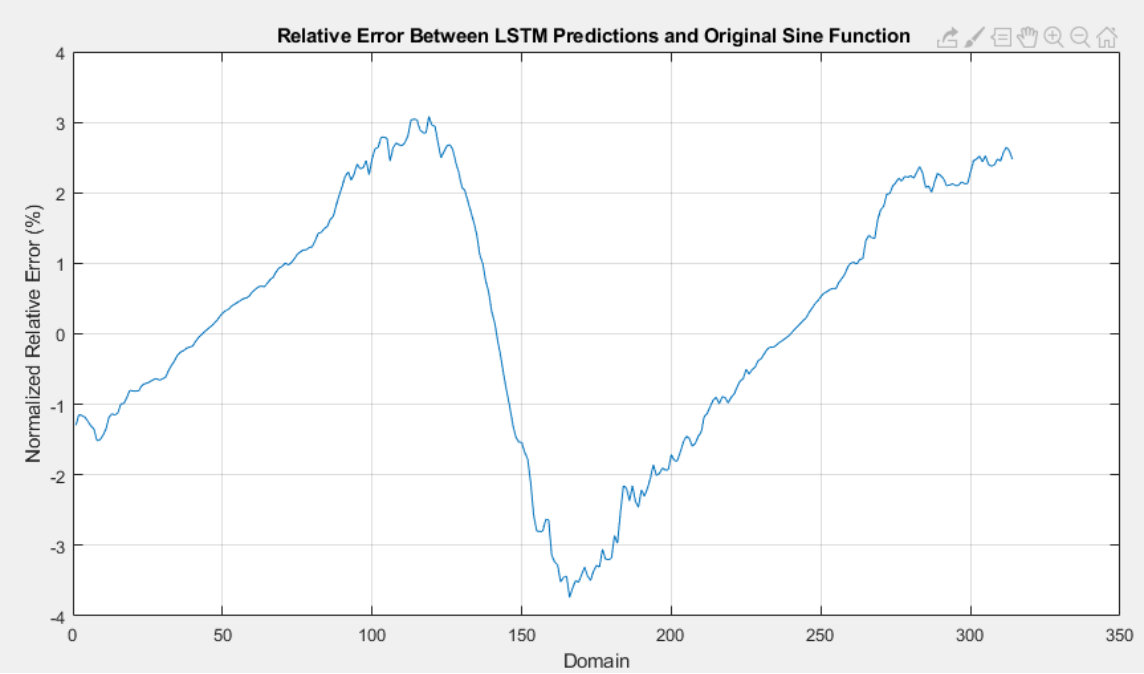

Figure 5-12 Relative error from the LSTM fie-steps-ahead forecasting the sine function with three training variables

The MSE and RMSE for this simulation are 0.002521 and 0.0502, respectively. The time spent for training and predicting the first set of data was six seconds, and for training and predicting, the whole set of data was 28 minutes.

5.4 Choosing the Best Combination of Variables

This section presents the results of the best combination of variables that most influence the LSTM predictions using the measured data from the mine. The target is the methane concentration at location TG.

Table 5-4 Error calculation for choosing the variables that most influence the predictions

Error	MSE	RMSE
CH_4 at location MG	0.002374	0.0487
Airflow rate	0.002364	0.0486
Temperature	0.002731	0.0523
Pressure	0.002671	0.0517

According to the MSE and RMSE values from table 5-4, the variables that give the best predictions in combination with the target are the CH_4 concentration at location MG and the airflow rate. Therefore, the LSTM will be trained using these two variables.

5.5 Direct Forward Prediction of Methane Concentrations with LSTM Model Using Measured Data from the Mine

The tests carried out in this section follow part (a) of the methodology presented in section 4.5.2. The LSTM is trained with the airflow rate and the CH_4 concentration

at location MG. It has to predict one step ahead for the target (CH_4 concentration at location TG). The predictions shown in the figure below were made using the unfiltered data.

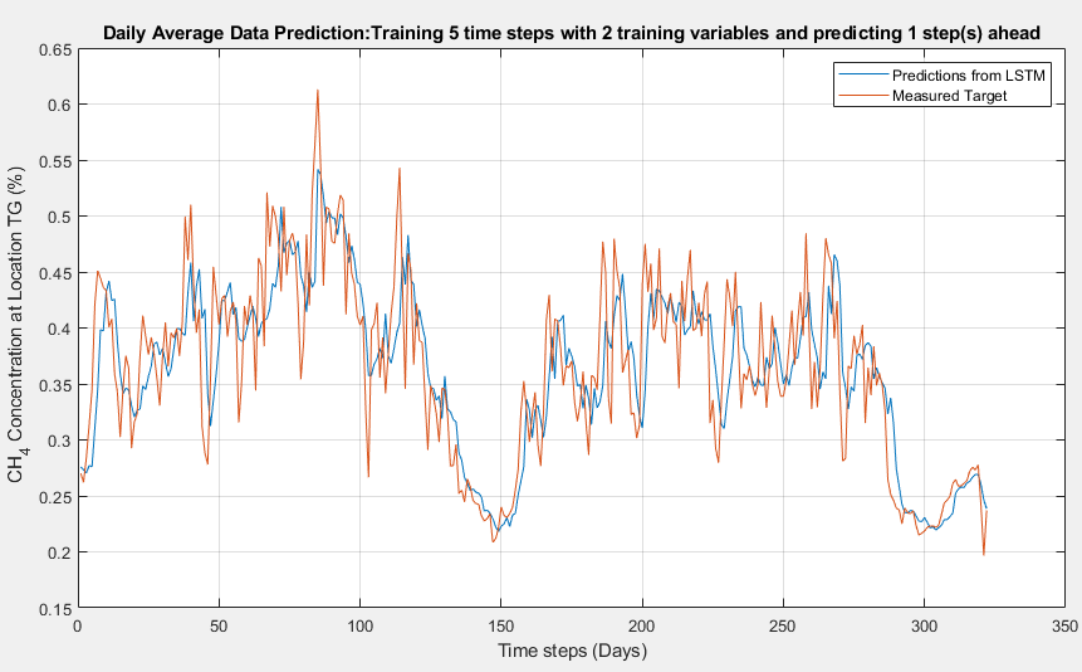

Figure 5-13 LSTM one-step-ahead prediction for the methane concentration at location TG using the unfiltered data.

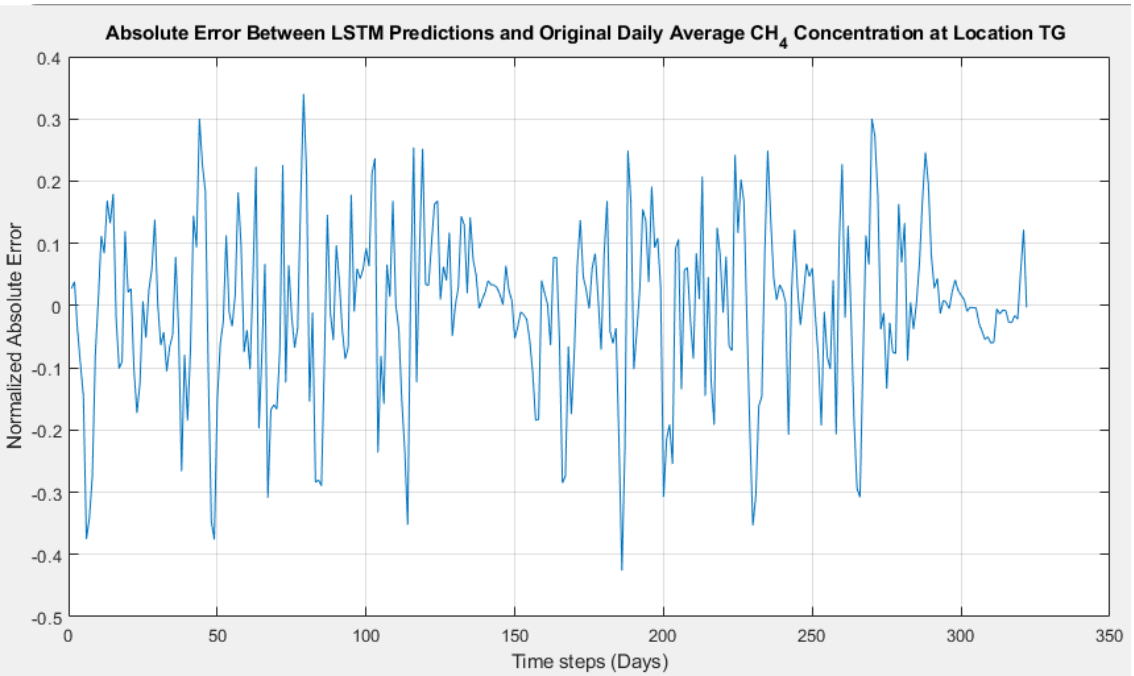

Figure 5-14 Absolute error for the LSTM one-step-ahead prediction for the methane concentration at location TG using the unfiltered data

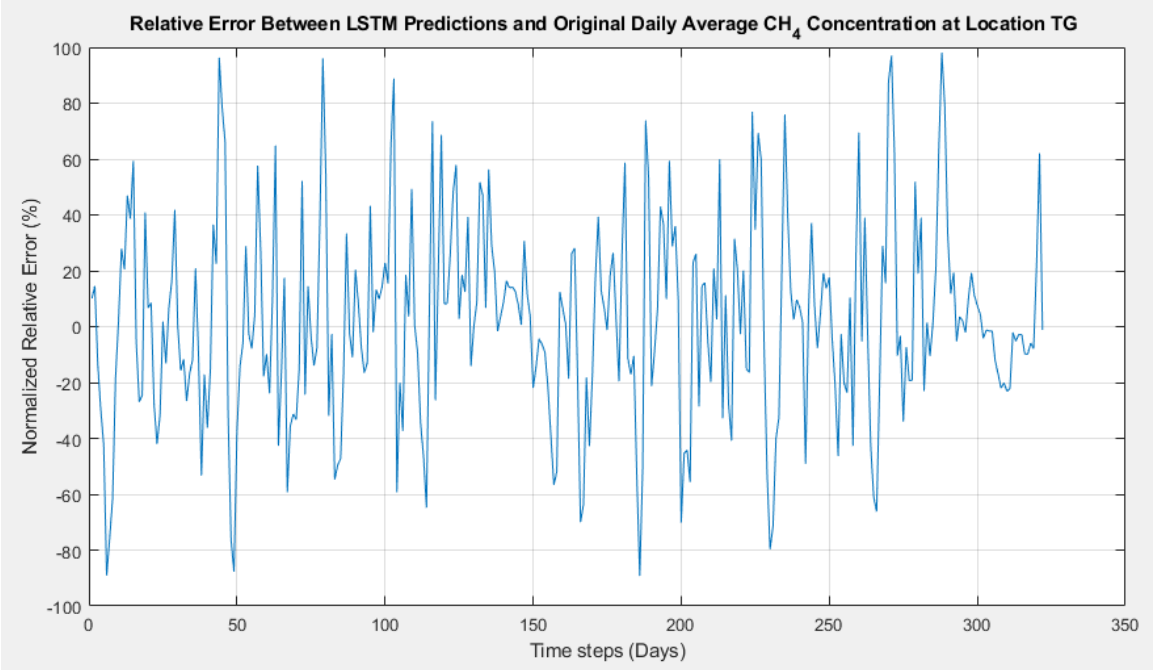

Figure 5-15 Relative error for the LSTM one-step-ahead prediction for the methane concentration at location TG using the unfiltered data

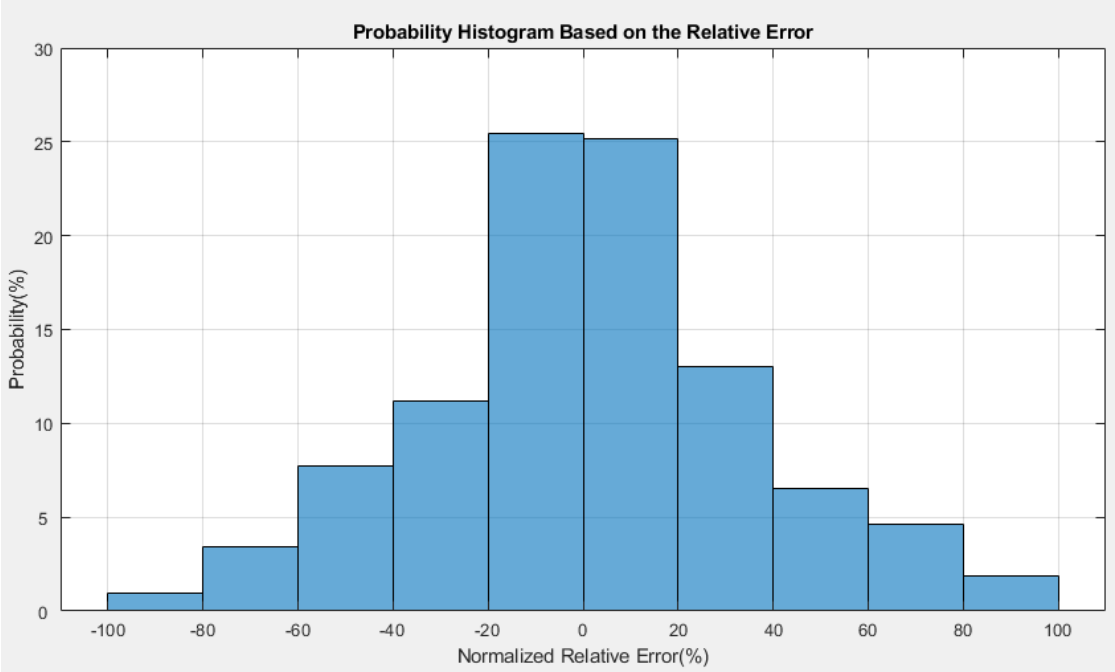

Figure 5-16 Probability histogram based on the relative error for the LSTM one-step-ahead prediction for the methane concentration at location TG using the unfiltered data

The MSE and RMSE for this forecast are 0.002125 and 0.0461, respectively. The time spent training the first five steps and predicting one-step-ahead was six seconds, and the whole simulation lasted 29 minutes.

The same data was used; however, this time, it was filtered. The results are shown below:

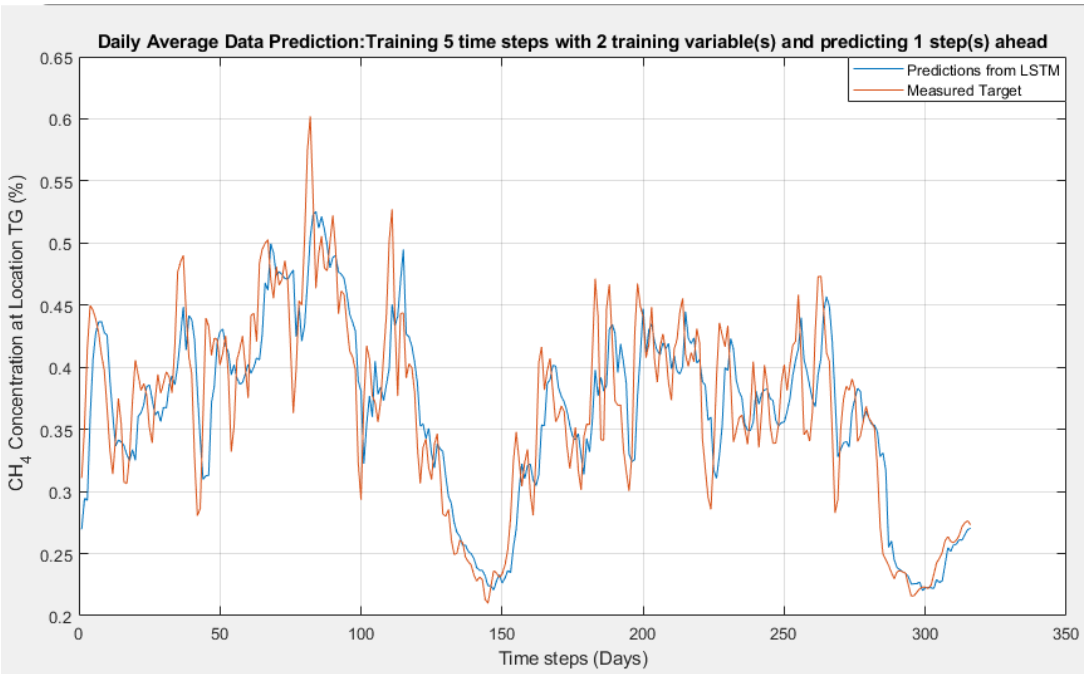

Figure 5-17 LSTM one-step-ahead prediction for the methane concentration at location TG using the filtered data.

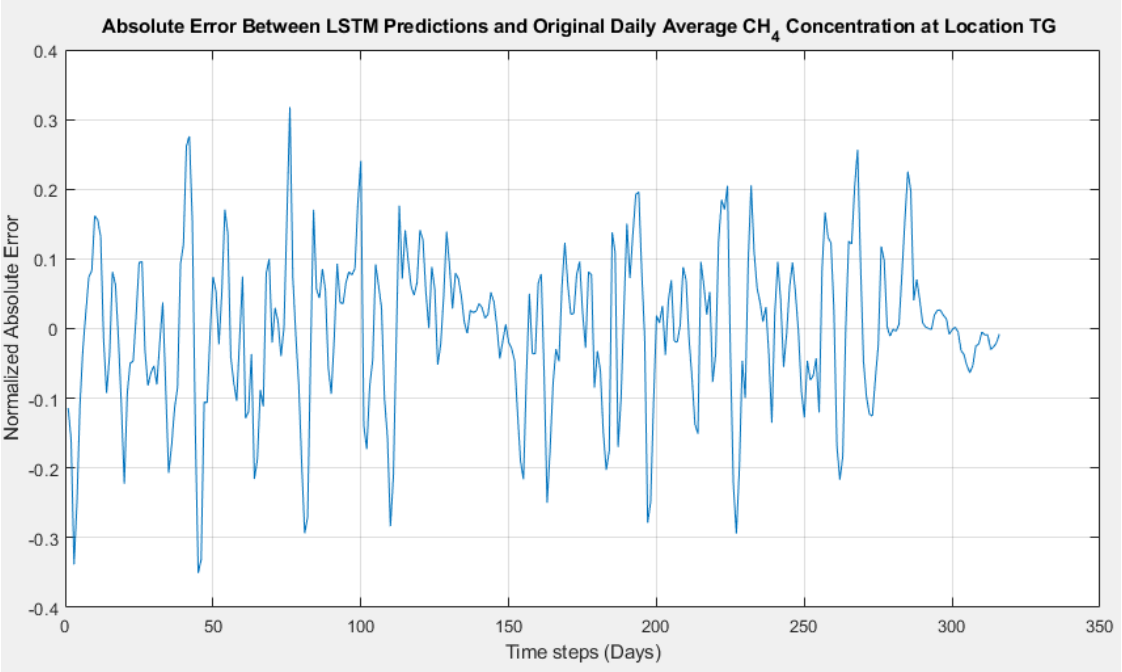

Figure 5-18 Absolute error for the LSTM one-step-ahead prediction for the methane concentration at location TG using the filtered data

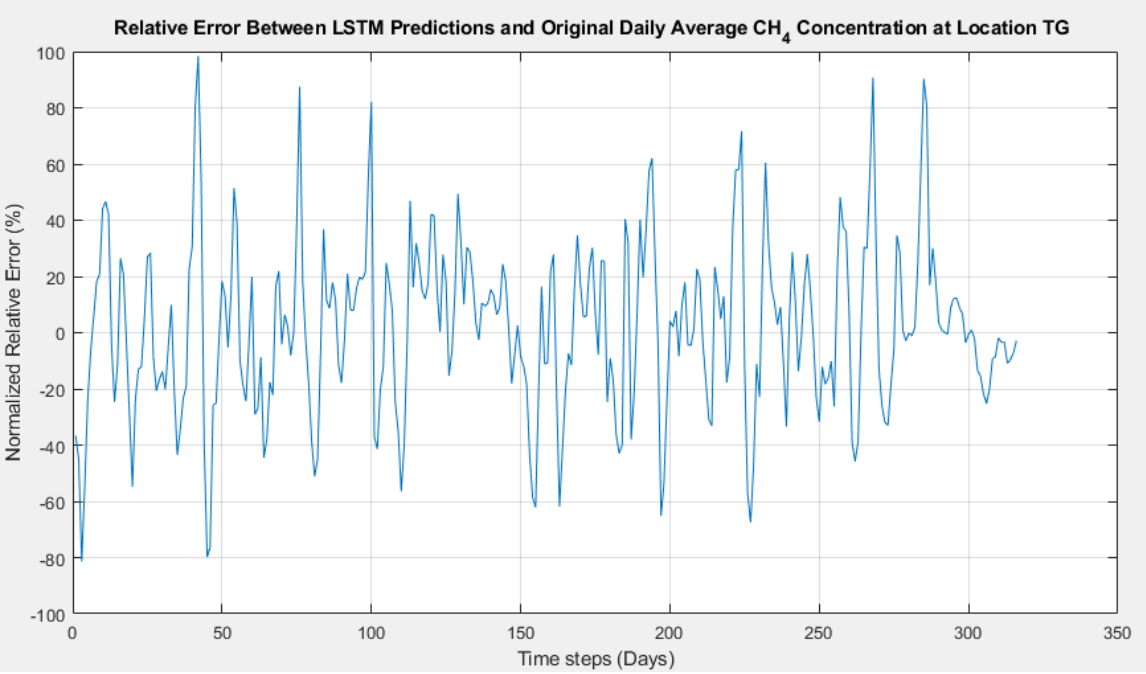

Figure 5-19 Relative error for the LSTM one-step-ahead prediction for the methane concentration at location TG using the filtered data

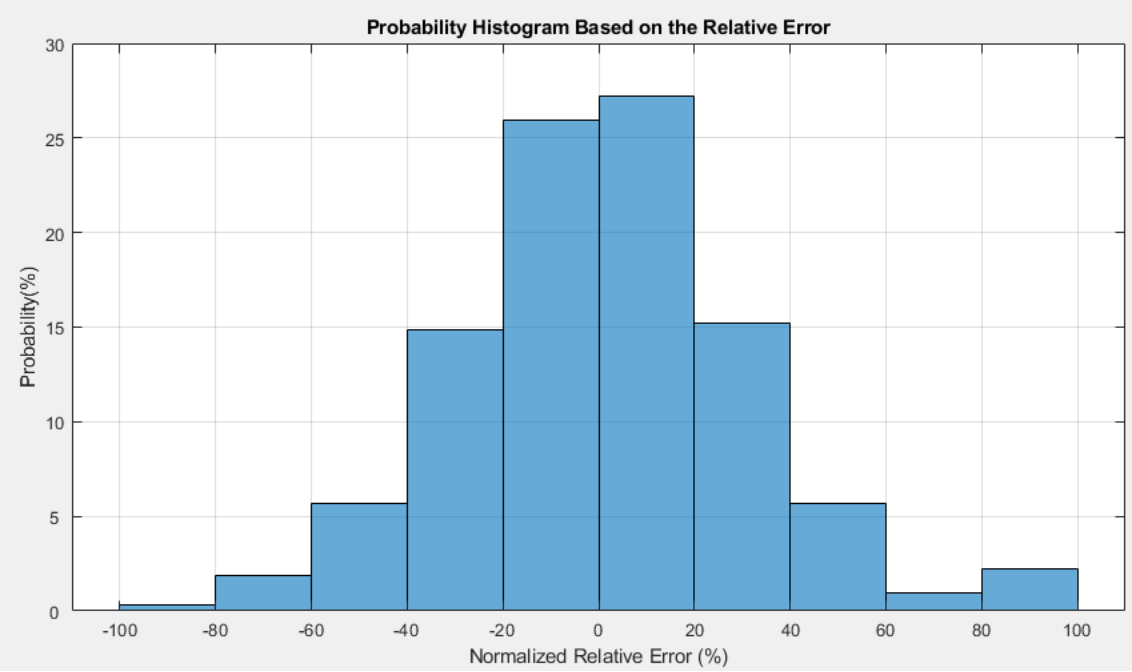

Figure 5-20 Probability histogram based on the relative error for the LSTM one-step-ahead prediction for the methane concentration at location TG using the filtered data

The MSE and RMSE for this simulation are 0.001679 and 0.0410, respectively. The time spent for training the first five data points was 5.2 seconds, and for training and predicting, the whole simulation was 28 minutes.

5.6 Direct Forward Prediction of Methane Concentrations with the Time Series Filter Using Measured Data from the Mine

The tests carried out in this section follow part (b) of the methodology presented in section 4.5.2. The time-series filter (TSF) is trained with the target and has to forecast one-step-ahead. This simulation was done using unfiltered data.

Figure 5-21 TSF one-step-ahead prediction for the methane concentration at location TG using the unfiltered data.

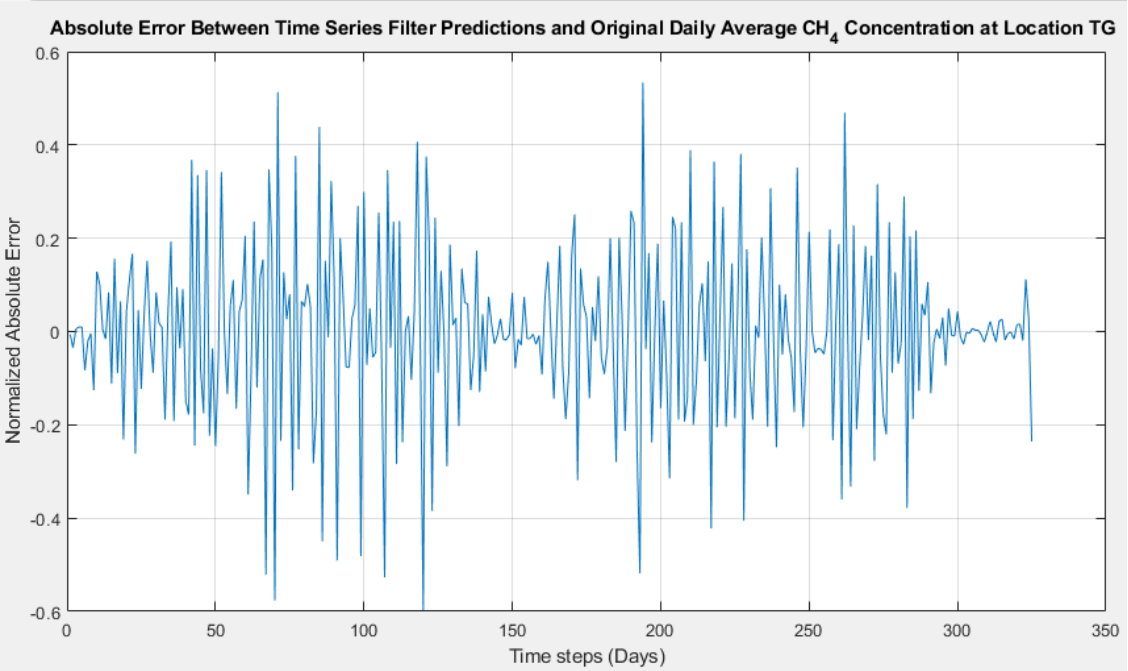

Figure 5-22 Absolute error for the TSF one-step-ahead prediction for the methane concentration at location TG using the unfiltered data

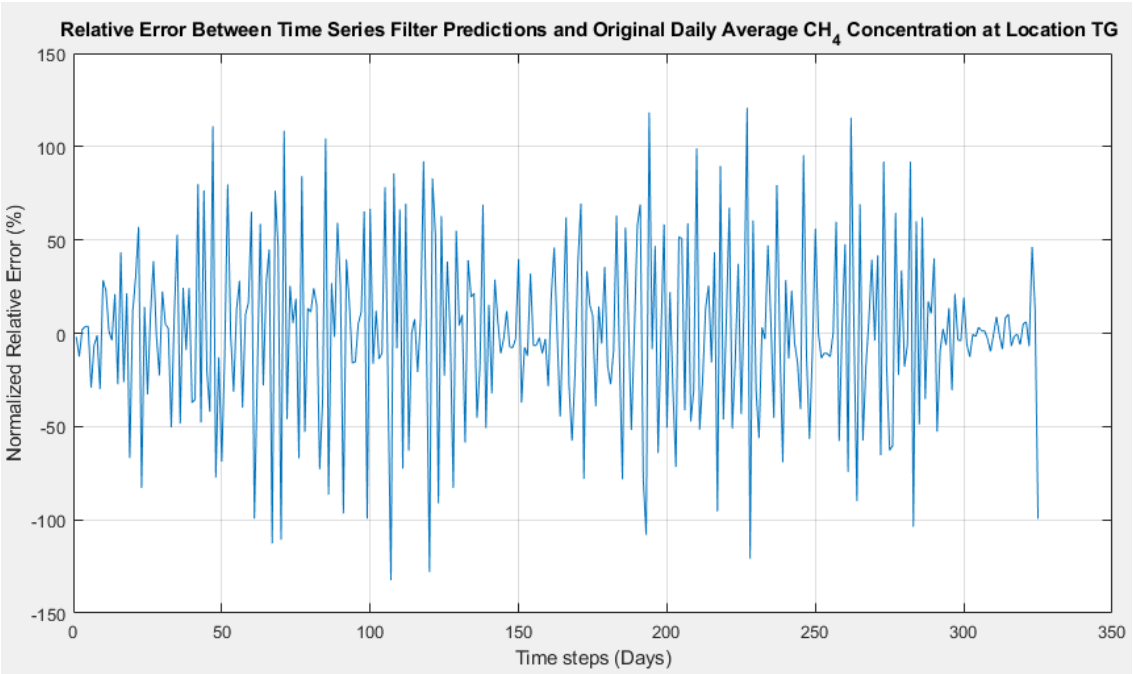

Figure 5-23 Relative error for the TSF one-step-ahead prediction for the methane concentration at location TG using the unfiltered data

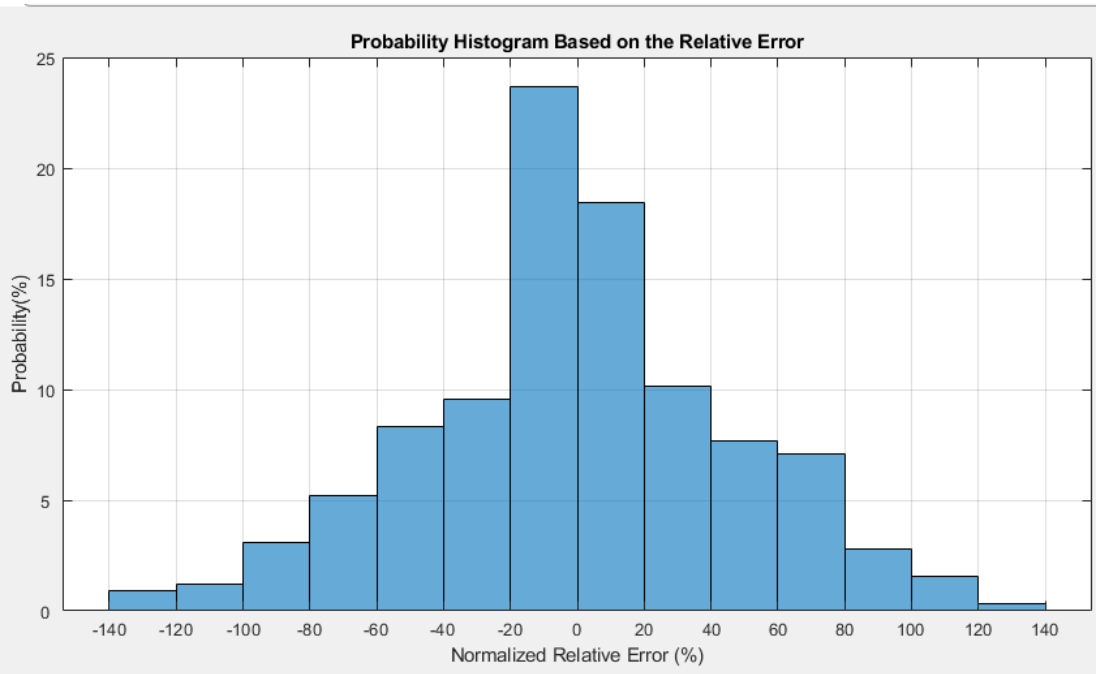

Figure 5-24 Probability histogram based on the relative error for the TSF one-step-ahead prediction for the methane concentration at location TG using the unfiltered data

The MSE and RMSE for this simulation are 0.004619 and 0.0680, respectively. The time spent for the whole simulation was 0.52 seconds. The simulation was made again, however, with the filtered data.

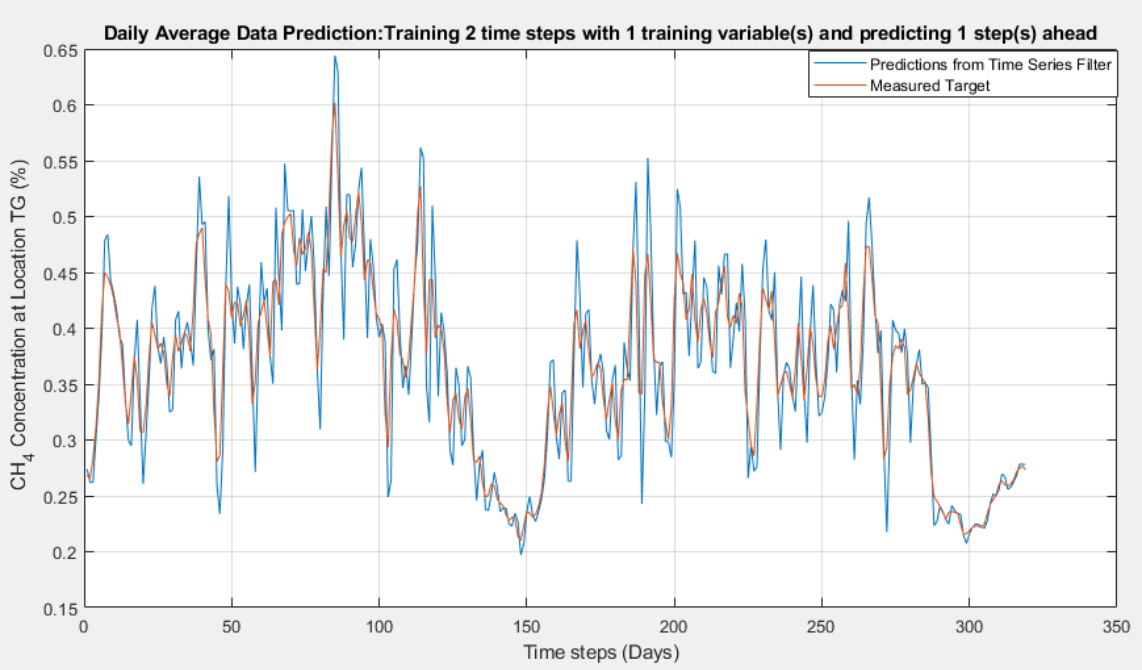

Figure 5-25 TSF one-step-ahead prediction for the methane concentration at location TG using the filtered data.

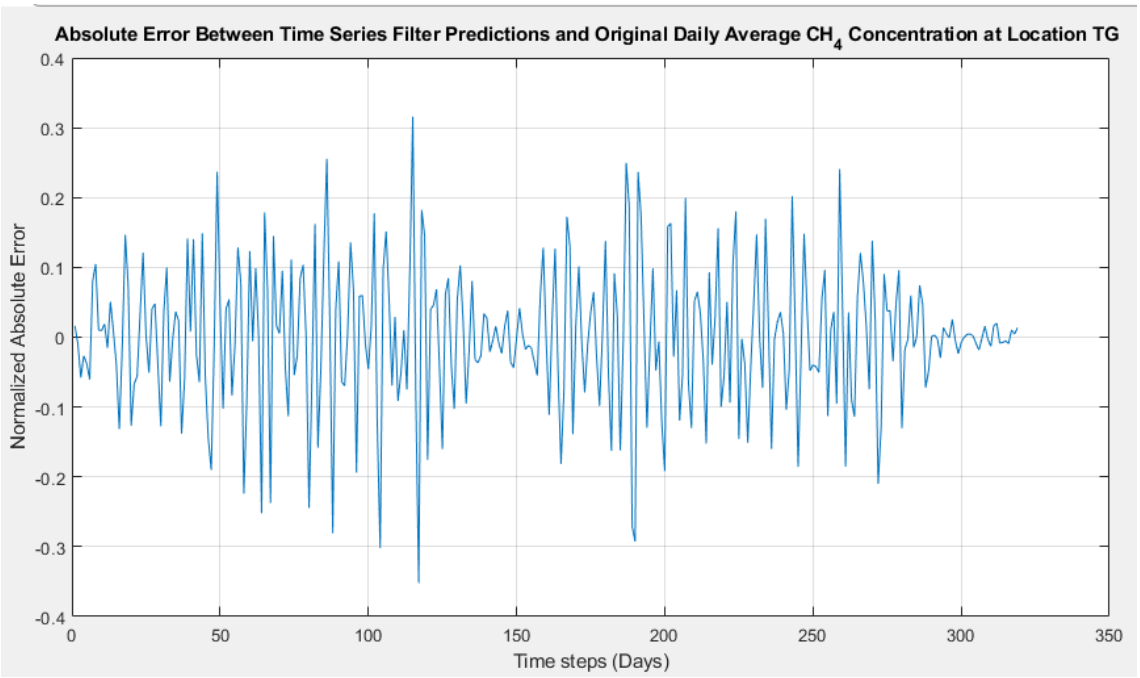

Figure 5-26 Absolute error for the TSF one-step-ahead prediction for the methane concentration at location TG using the filtered data

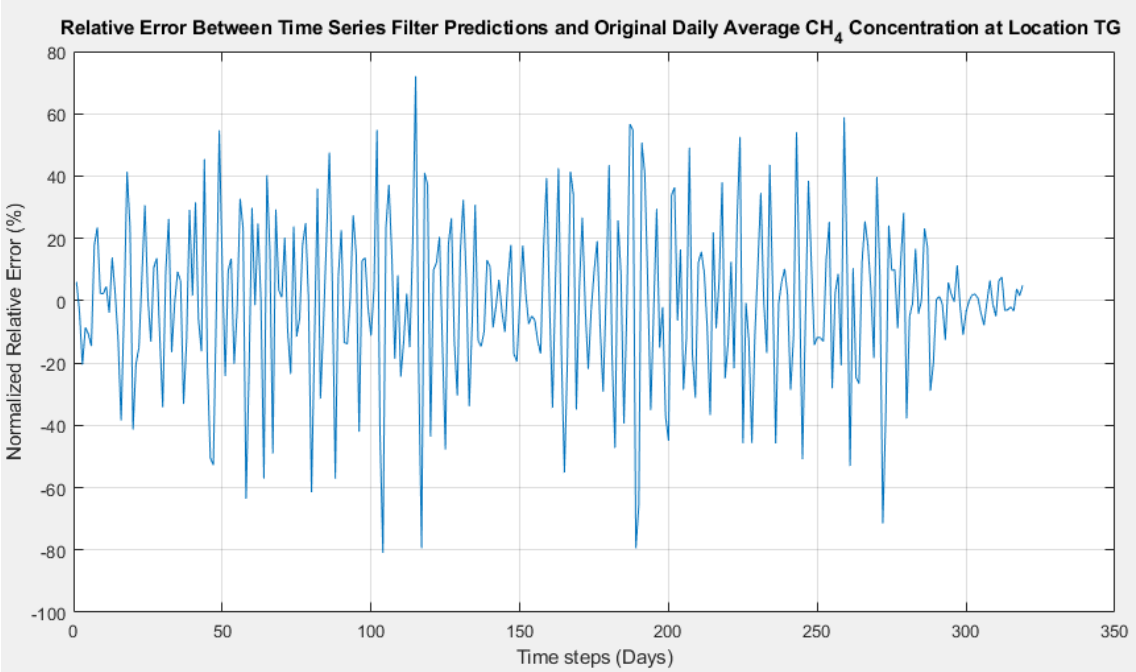

Figure 5-27 Relative error for the TSF one-step-ahead prediction for the methane concentration at location TG using the filtered data

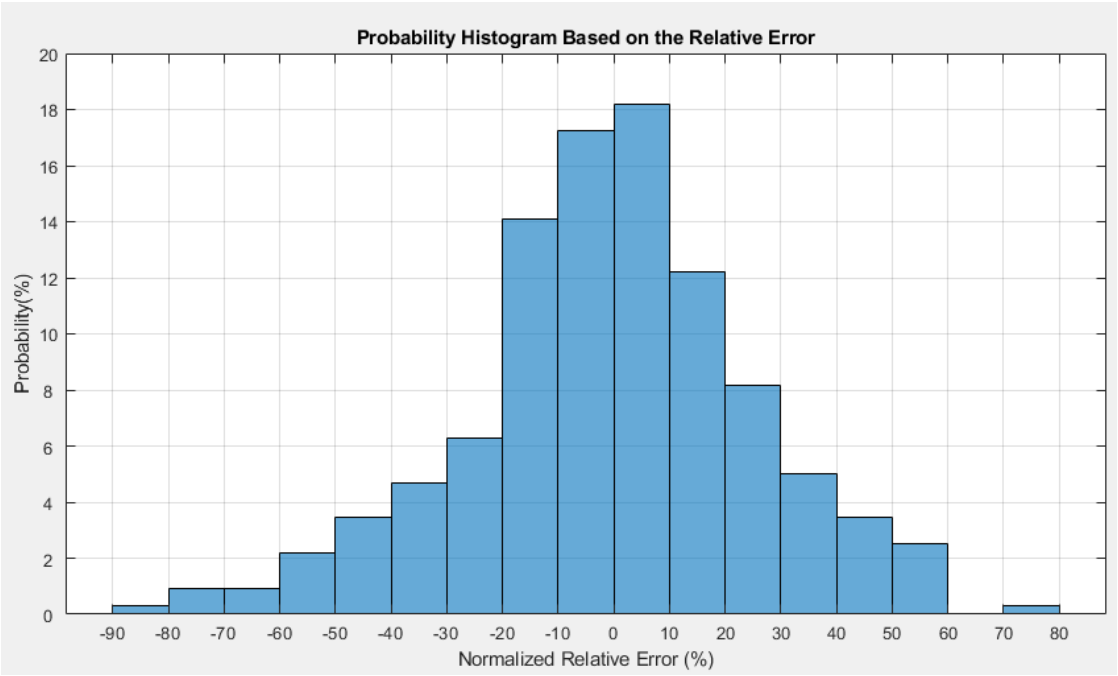

Figure 5-28 Probability histogram based on the relative error for the TSF one-step-ahead prediction for the methane concentration at location TG using the filtered data

The MSE and RMSE for this simulation are 0.001345 and 0.00367, respectively. The time spent for the whole simulation was 0.19 seconds.

5.7 Forward-Prediction of Root-Cause CH_4 Sources With the LSTM Model

The results shown in this section are based on part (a) of the methodology in section 4.5.3. The figure below shows the predictions for the back-calculated line source using the airflow rate as a training variable.

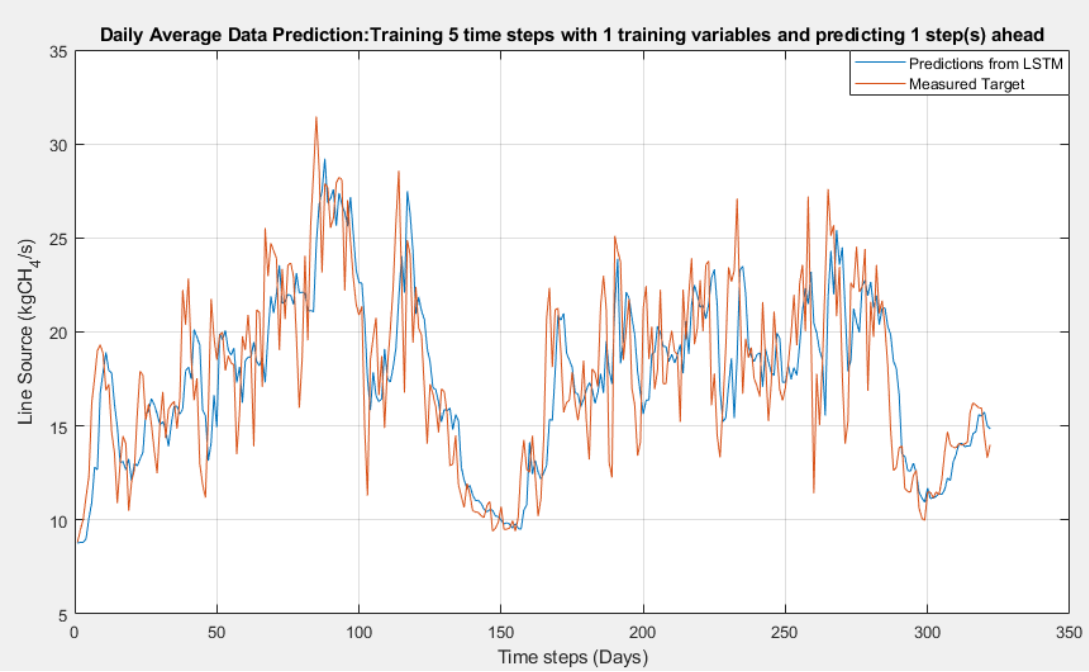

Figure 5-29 LSTM one-step-ahead prediction for the line source using the unfiltered data

The back-calculated methane concentration at location TG is presented below:

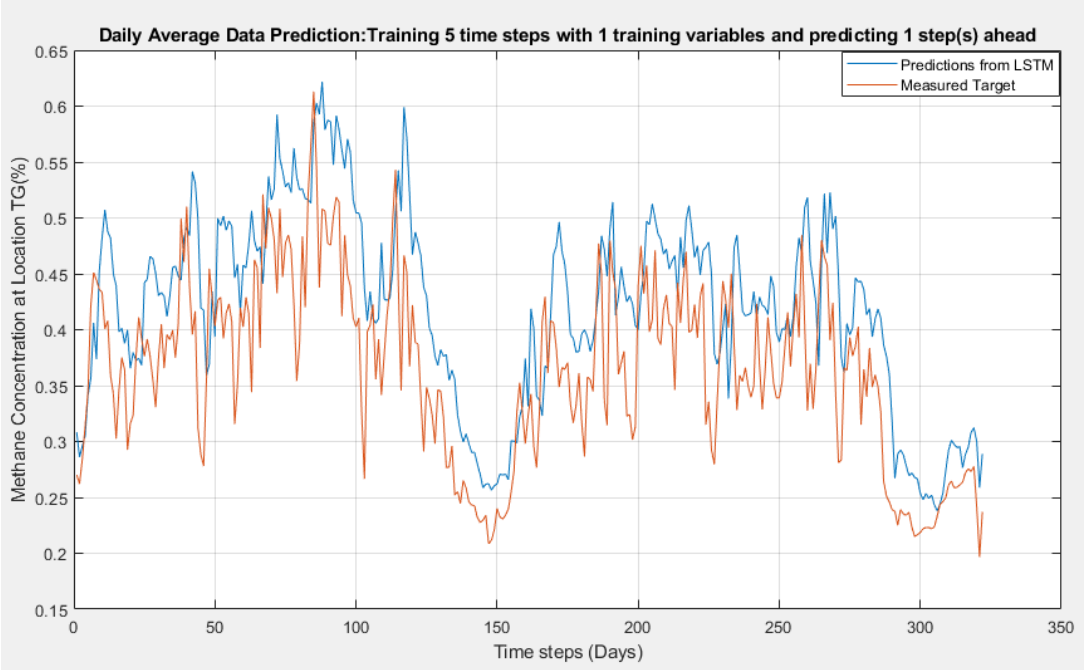

Figure 5-30 LSTM back-calculated one-step-ahead prediction for the methane concentration at location TG using the unfiltered data

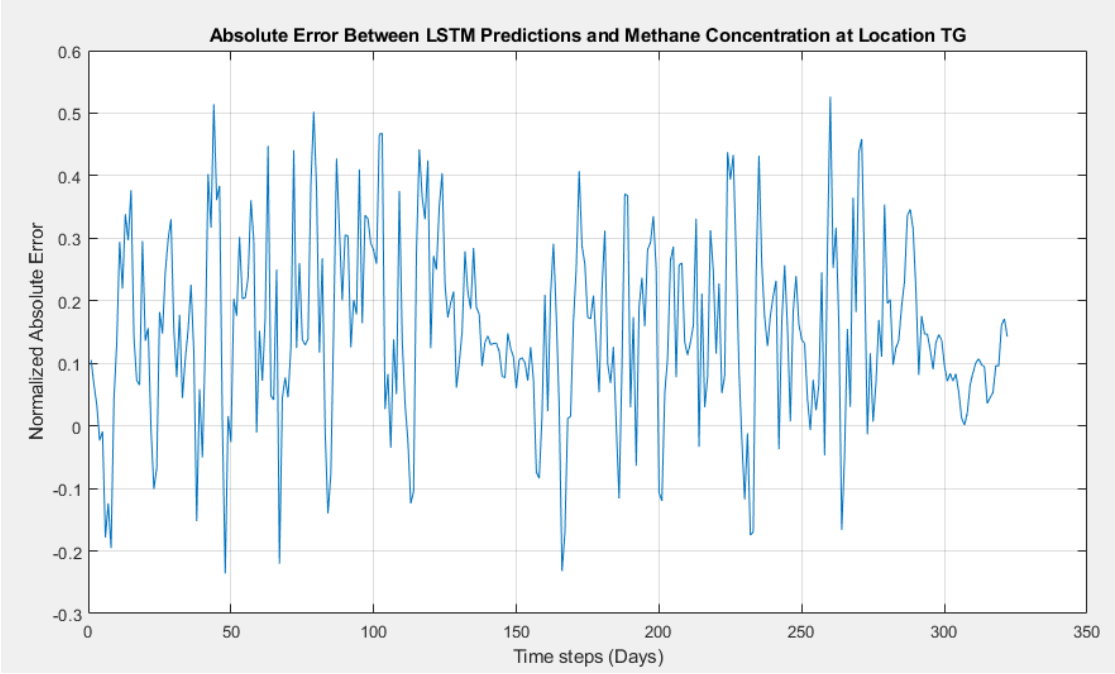

Figure 5-31 Absolute error for the LSTM one-step-ahead prediction for the back-calculated methane concentration at location TG using the unfiltered data

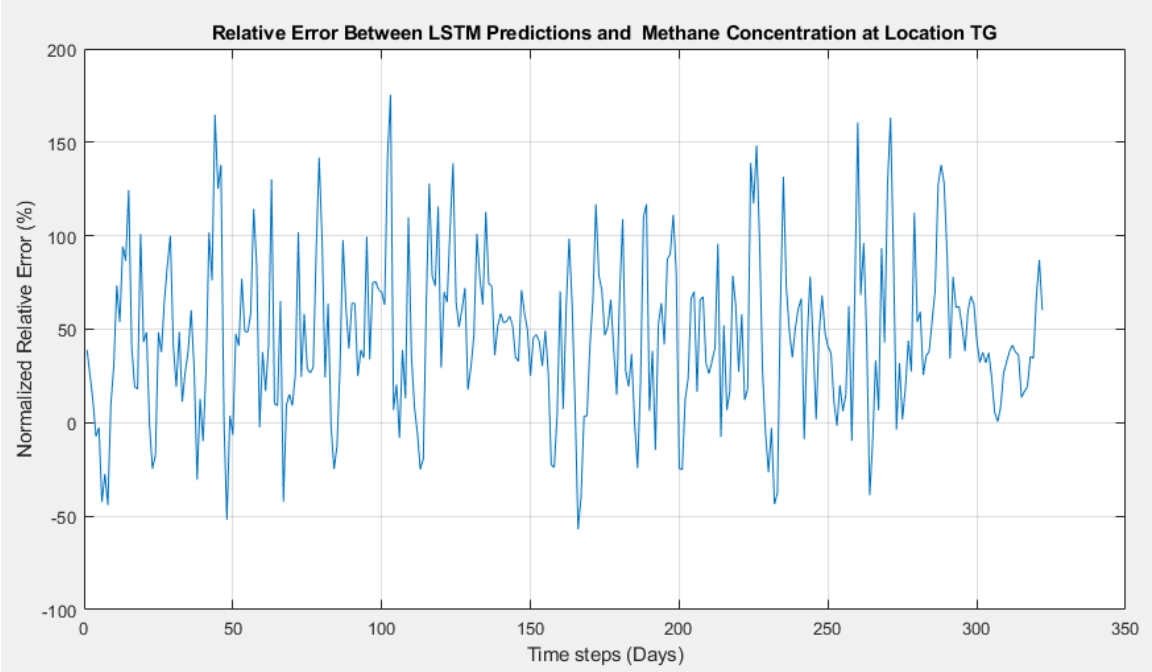

Figure 5-32 Relative error for the LSTM one-step-ahead prediction for the back-calculated methane concentration at location TG using the unfiltered data

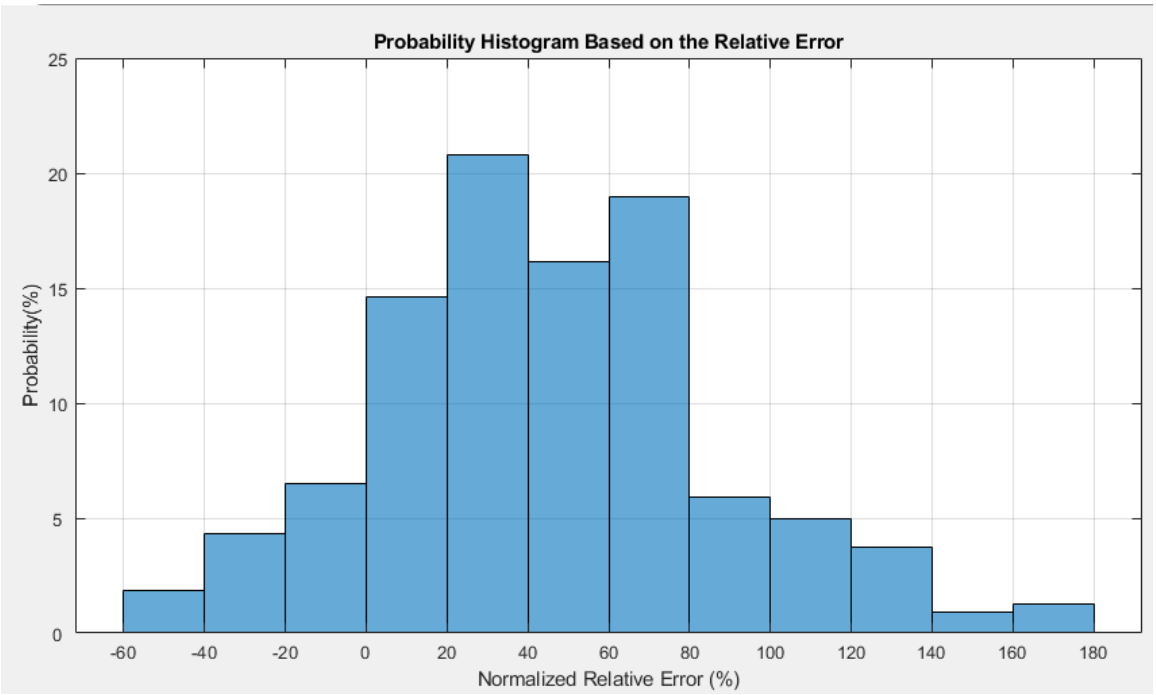

Figure 5-33 Probability histogram based on the relative error for the LSTM one-step-ahead prediction for the back-calculated methane concentration at location TG using the unfiltered data

The MSE and RMSE for this simulation are 0.005861 and 0.0766, respectively. The time spent training and predicting the first set of data was 16 seconds and the time spent on the whole prediction was 28 minutes.

5.8 Forward Prediction of Root-Cause CH_4 With the Time Series Filter Model

The methodology followed in this section is described in section 4.5.3 part b. The figure below shows the back-calculated methane concentration using the Time Series model.

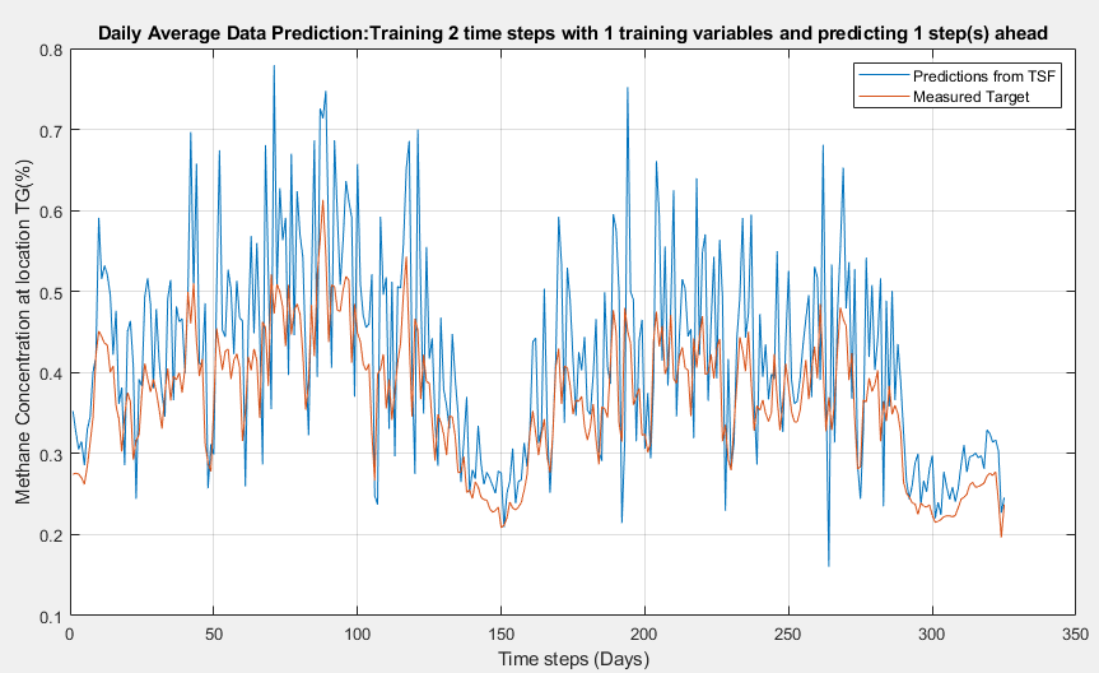

Figure 5-34 TSF back-calculated one-step-ahead prediction for the methane concentration at location TG using the unfiltered data

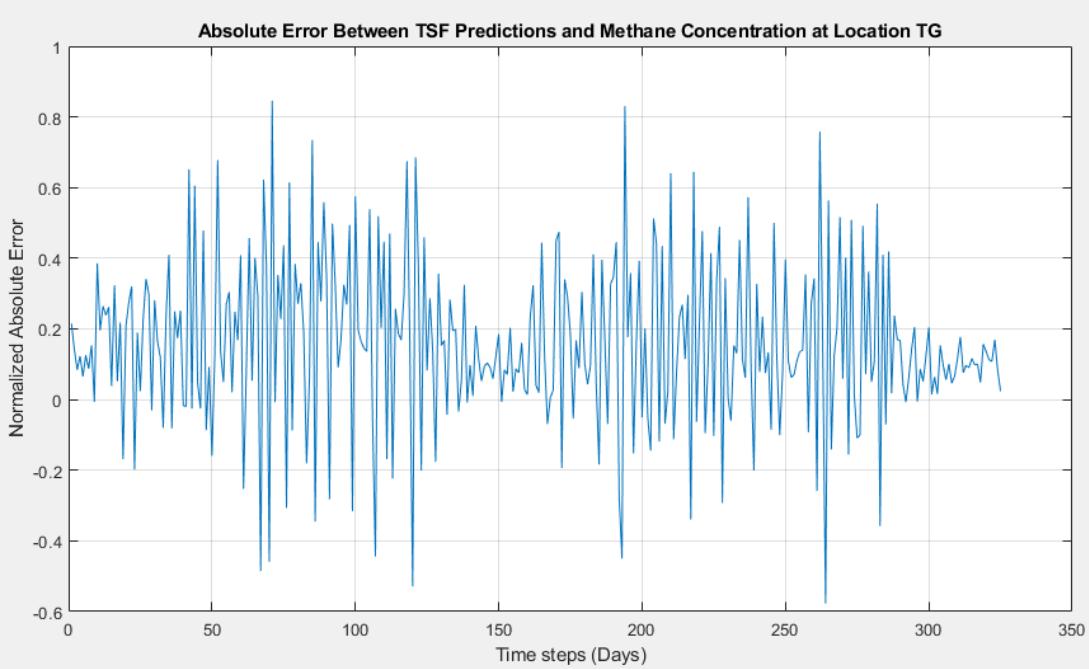

Figure 5-35 Absolute error for the TSF one-step-ahead prediction for the back-calculated methane concentration at location TG using the unfiltered data

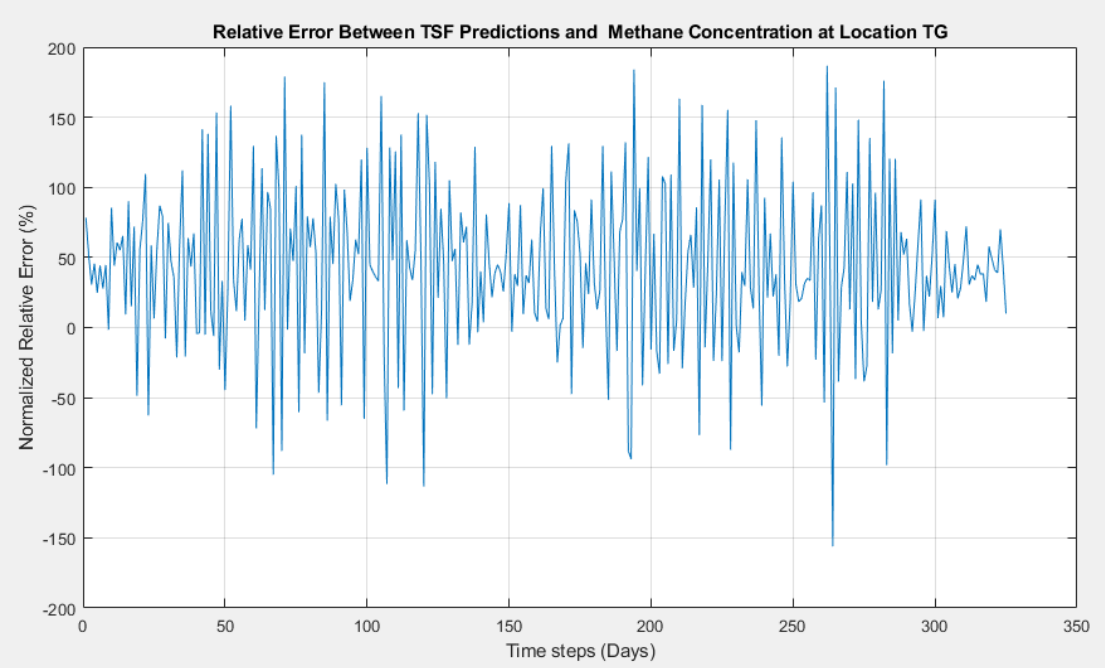

Figure 5-36 Relative error for the TSF one-step-ahead prediction for the back-calculated methane concentration at location TG using the unfiltered data

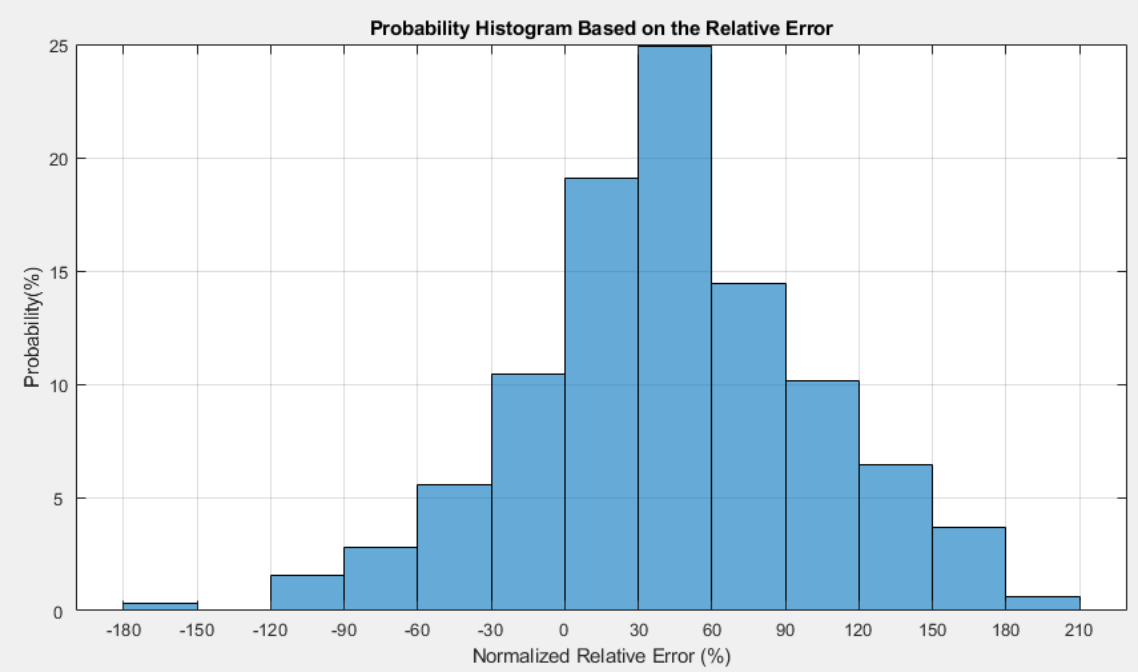

Figure 5-37 Probability histogram based on the relative error for the TSF one-step-ahead prediction for the back-calculated methane concentration at location TG using the unfiltered data

The MSE and RMSE for this simulation are 0.01044 and 0.1022, respectively. The time spent making the predictions was 0.25 seconds.

5.9 Testing the Influence of the Number of Training Variables

The methodology followed in this section is described in section 4.6. The LSTM was trained with the airflow rate, the methane concentration at location MG, and the pressure. The figure below shows the results using the unfiltered data.

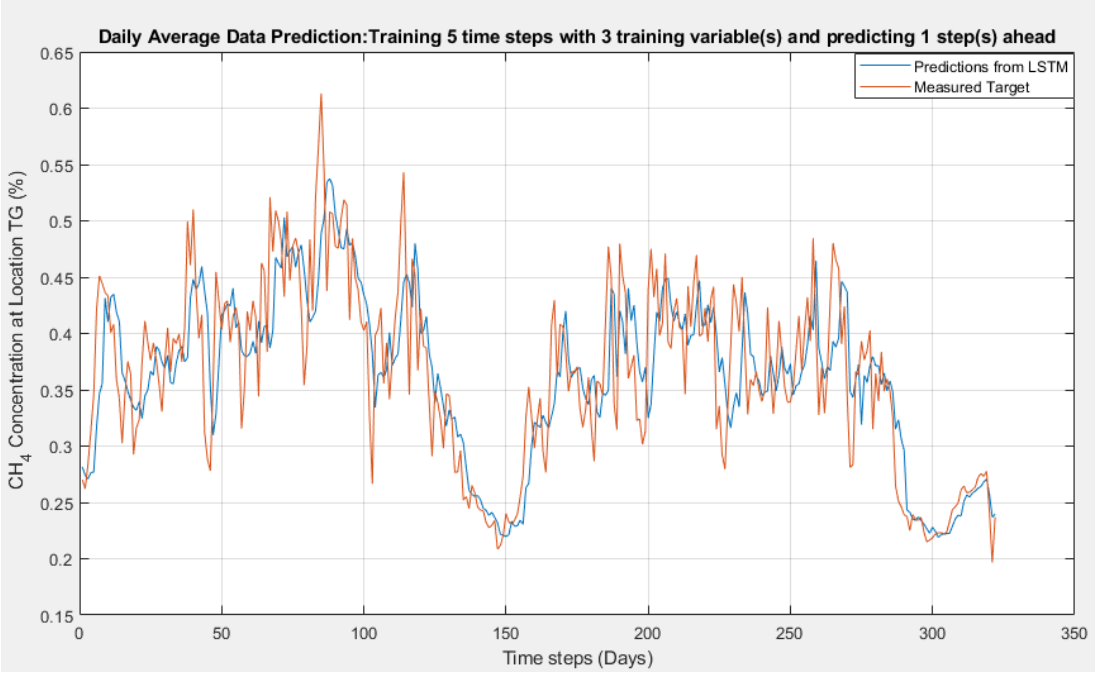

Figure 5-38 LSTM one-step-ahead prediction for the methane concentration at location TG using the unfiltered data and three training variables

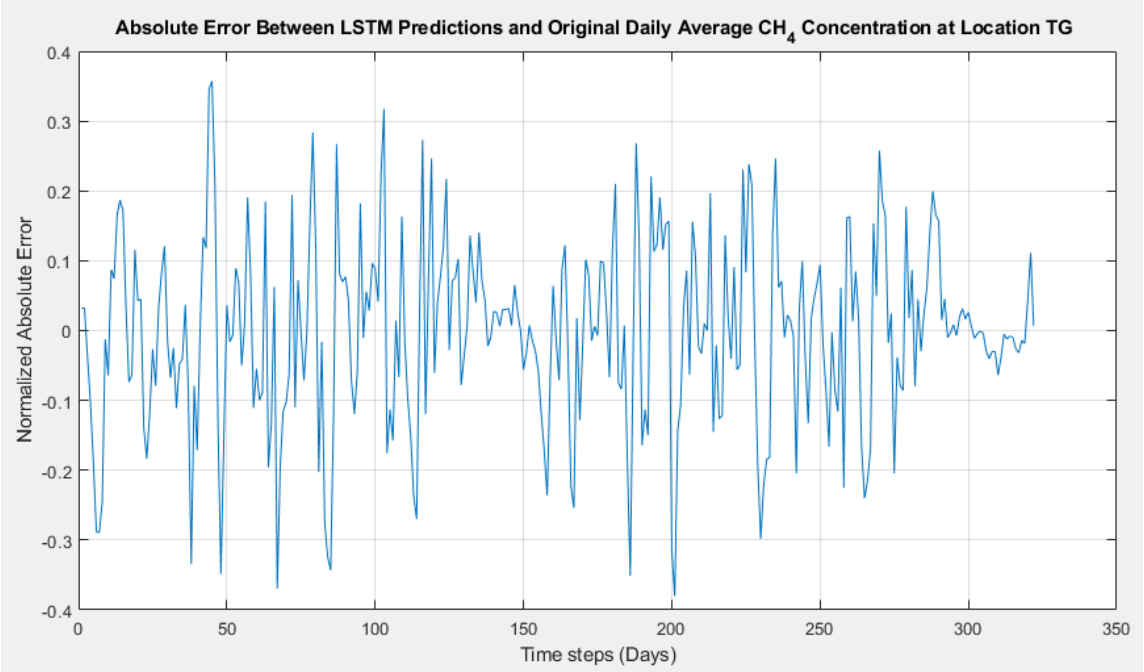

Figure 5-39 Absolute error for the LSTM one-step-ahead prediction for the methane concentration at location TG using the unfiltered data and three training variables

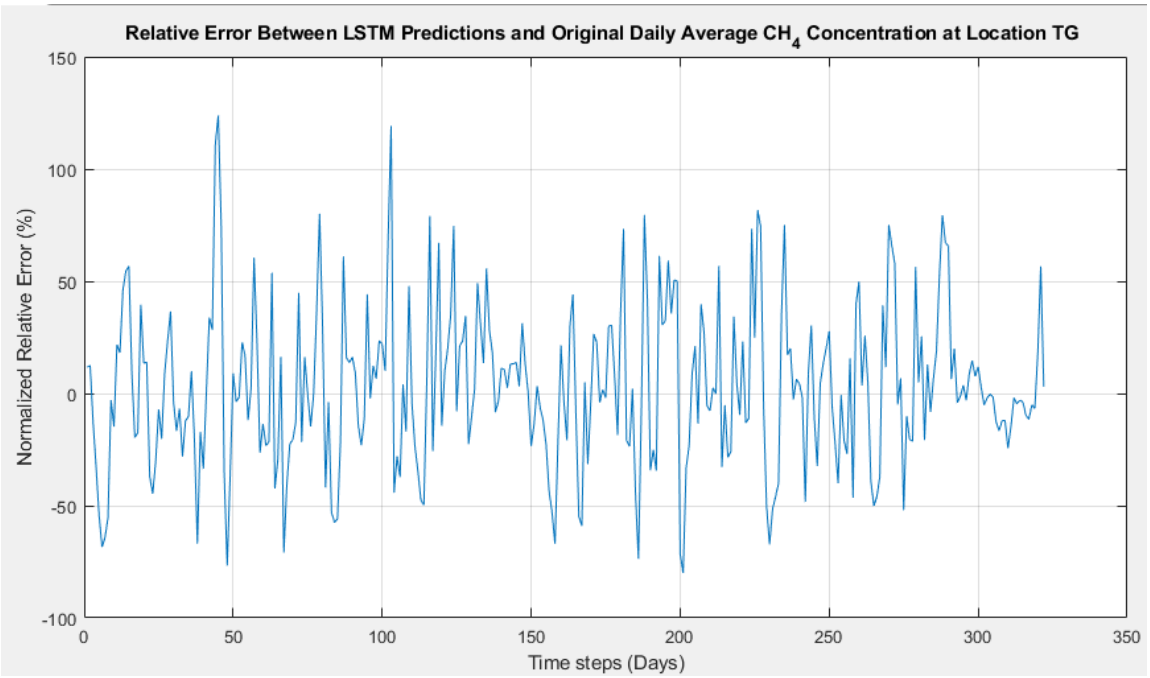

Figure 5-40 Relative error for the LSTM one-step-ahead prediction for the methane concentration at location TG using the unfiltered data and three training variables

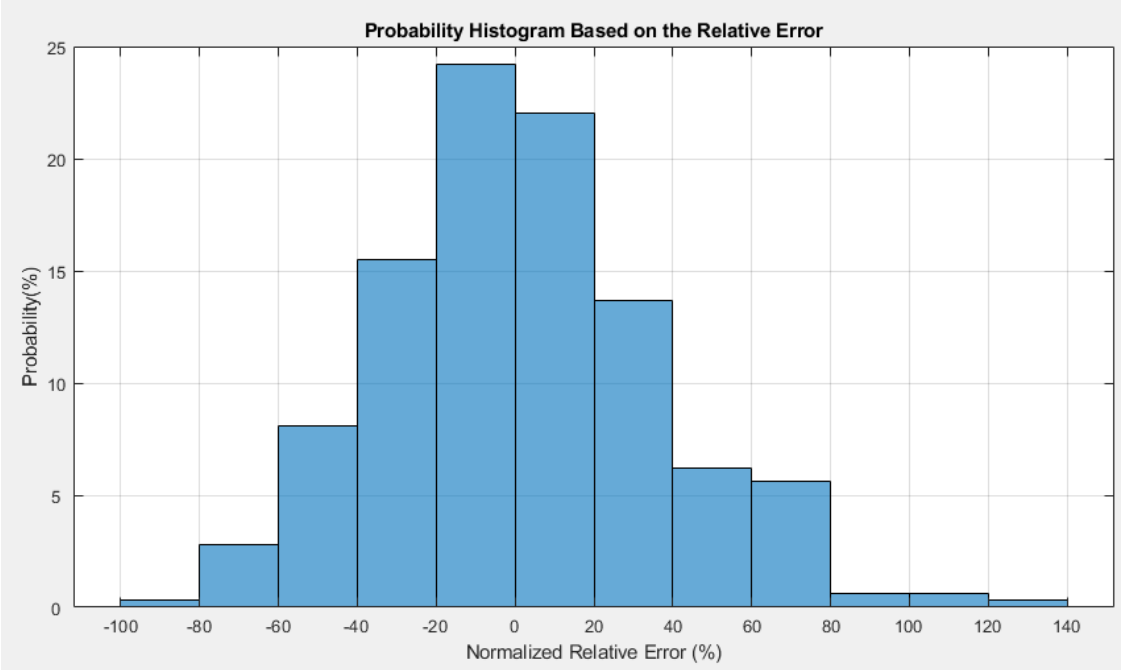

Figure 5-41 Probability histogram plot based on the relative error for the LSTM one-step-ahead prediction for the methane concentration at location TG using the unfiltered data and three training variables

The MSE and RMSE for this prediction are 0.002301 and 0.0480. The time spent making one prediction was 5 seconds and for making the predictions for the whole data set was 29 minutes. The same data was used; however, this time, it was filtered.

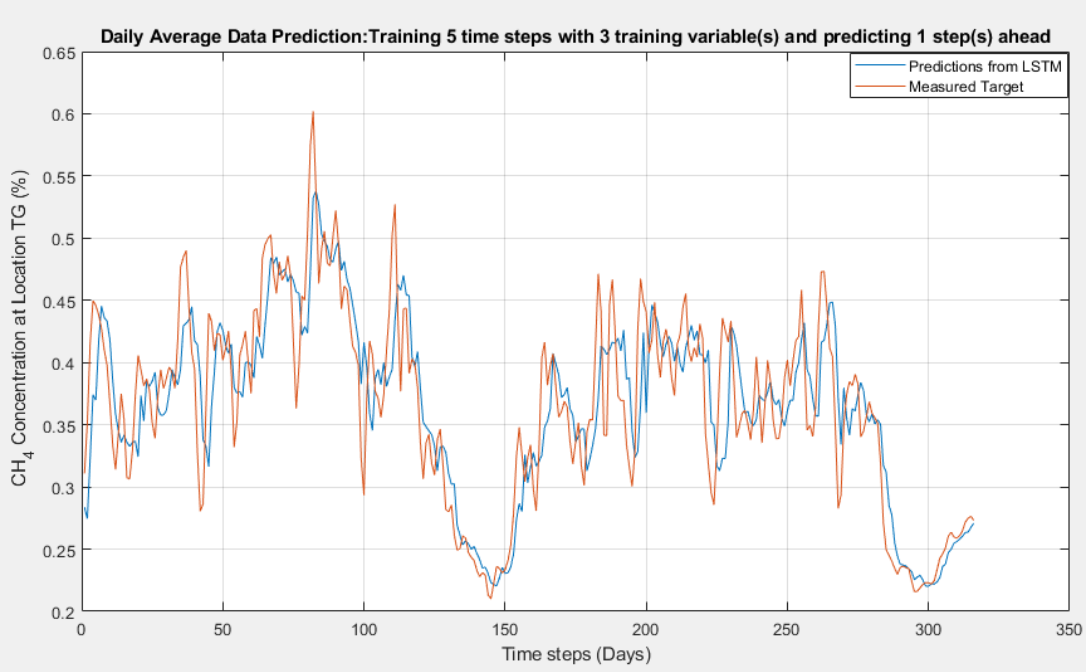

Figure 5-42 LSTM one-step-ahead prediction for the methane concentration at location TG using the filtered data and three training variables

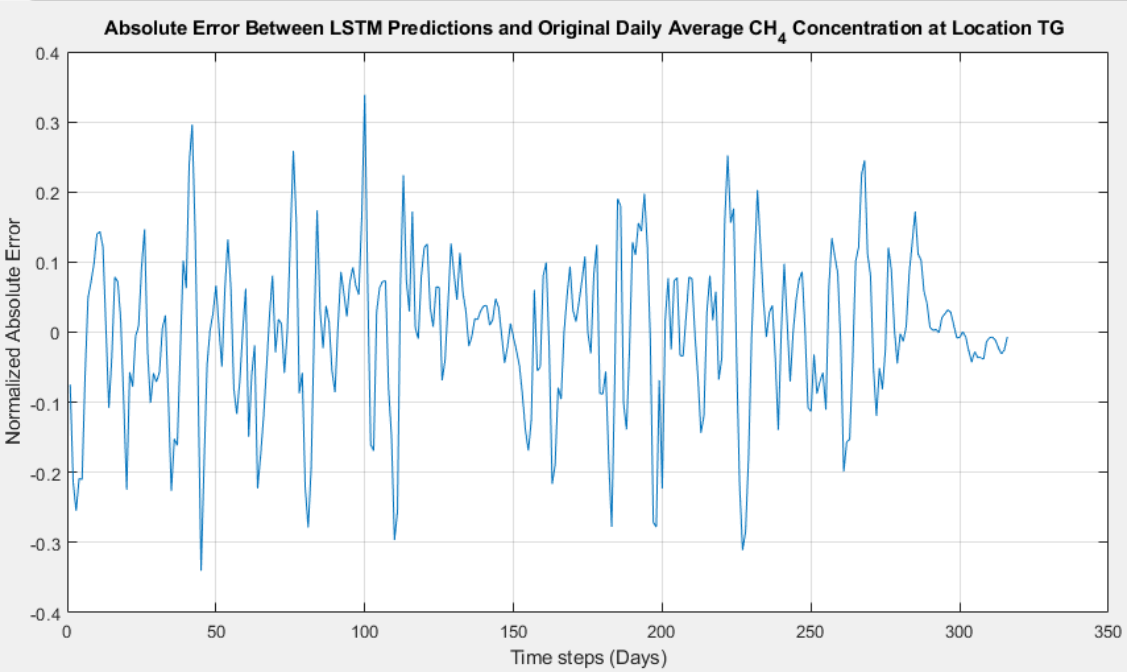

Figure 5-43 Absolute error for the LSTM one-step-ahead prediction for the methane concentration at location TG using the filtered data and three training variables

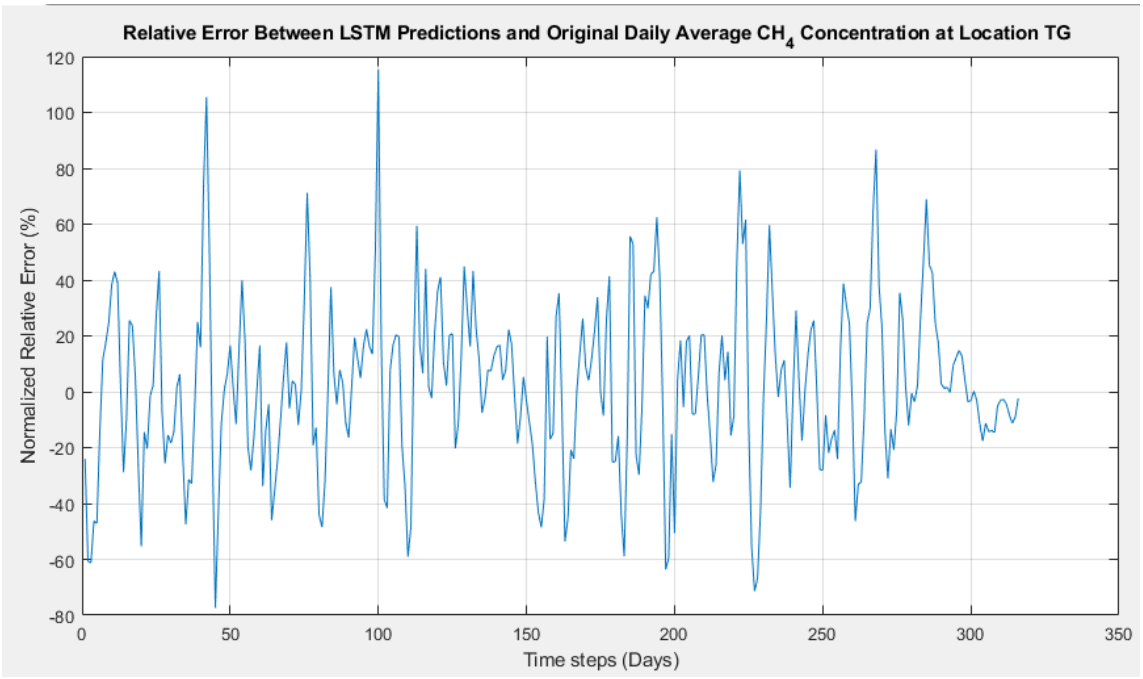

Figure 5-44 Relative error for the LSTM one-step-ahead prediction for the methane concentration at location TG using the filtered data and three training variables

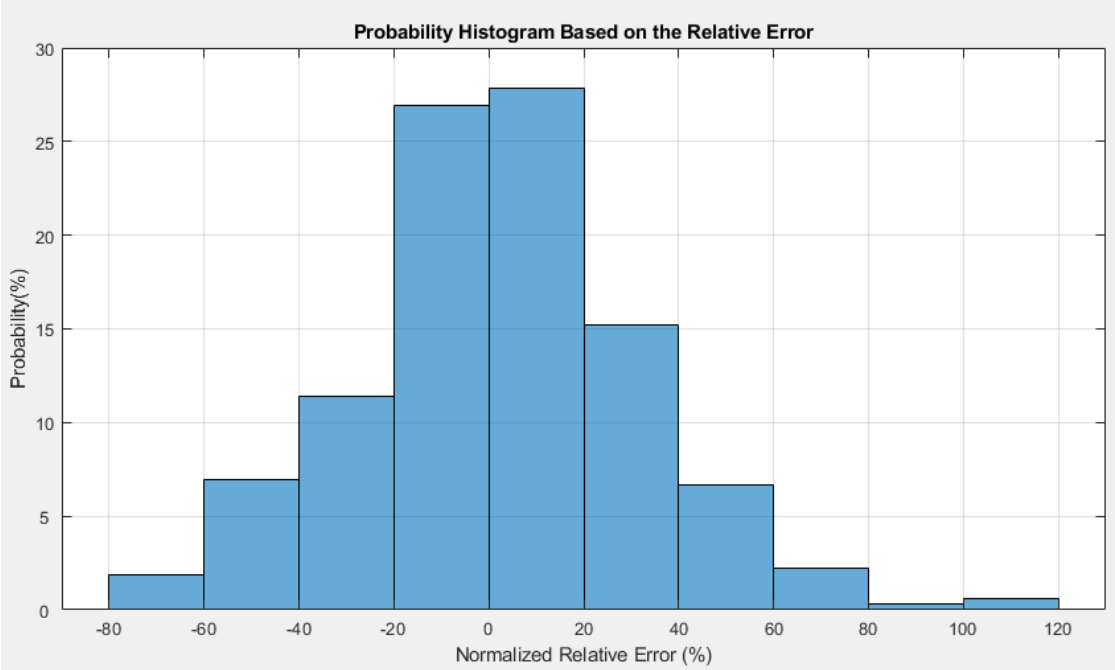

Figure 5-45 Probability histogram based on the relative error for the LSTM one-step-ahead prediction for the methane concentration at location TG using the filtered data and three training variables

The MSE and RME for this simulation are 0.001641 and 0.0405, respectively. The time spent for training the first five data points was 9 seconds, and for training and predicting, the whole data set was 31 minutes.

5.10 Testing the Influence of Training Data Size

The results presented in this section follow the methodology presented in section 4.7. The variables used for training were the methane concentration at location MG and the airflow rate.

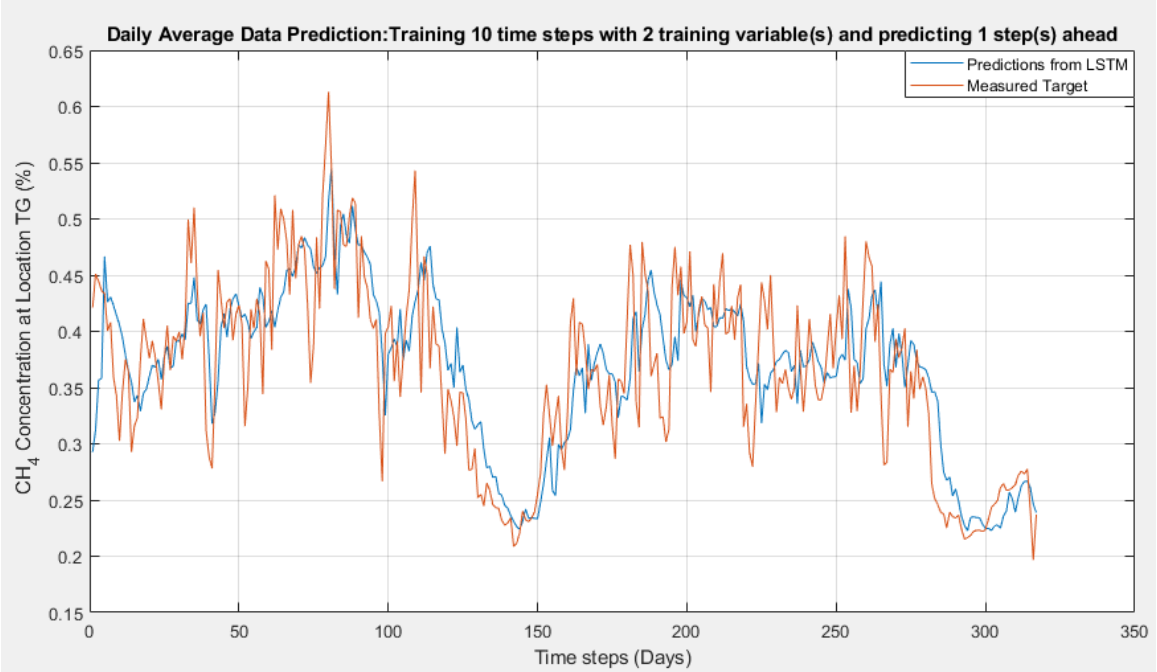

Figure 5-46 LSTM one-step-ahead prediction for the methane concentration at location TG using the unfiltered data and a training window size of ten

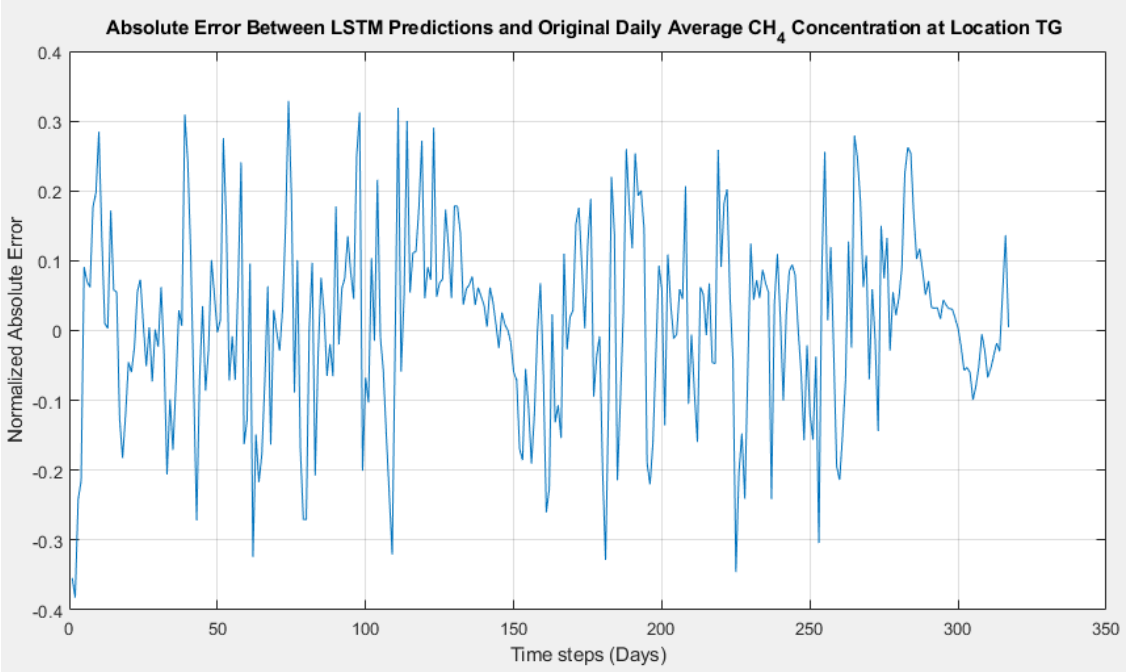

Figure 5-47 Absolute error for the LSTM one-step-ahead prediction for the methane concentration at location TG using the unfiltered data and a training window size of ten

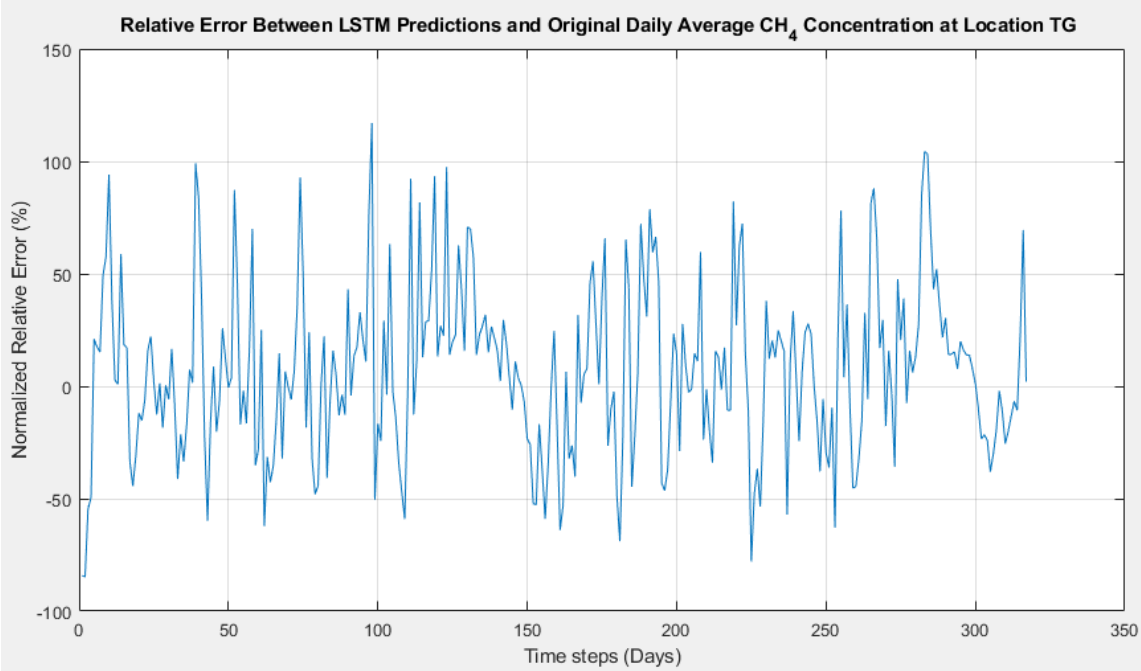

Figure 5-48 Relative error for the LSTM one-step-ahead prediction for the methane concentration at location TG using the unfiltered data and a training window size of ten

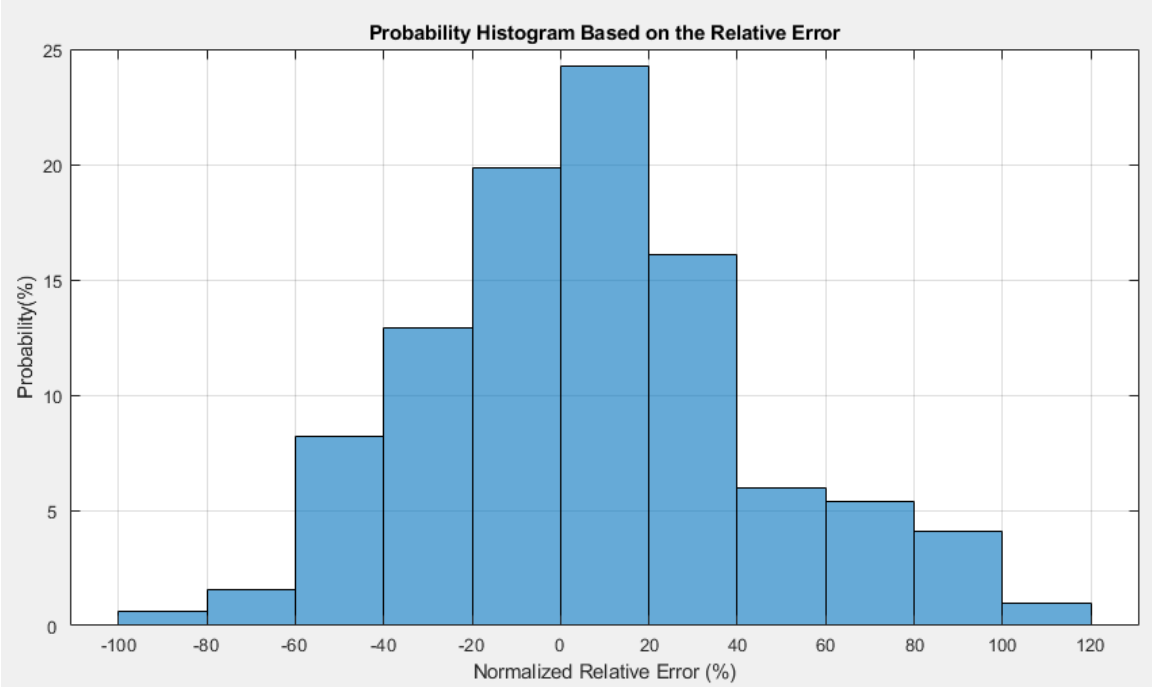

Figure 5-49 Probability histogram based on the relative error for the LSTM one-step-ahead prediction for the methane concentration at location TG using the unfiltered data and a training window size of ten

The MSE and RMSE for this simulation are 0.002437 and 0.0494, respectively. The time spent predicting the first training window was 16 seconds, and the whole data set was 64 minutes.

The same data set was used to train the LSTM with a window size of fifty-time steps. The predictions can be seen below:

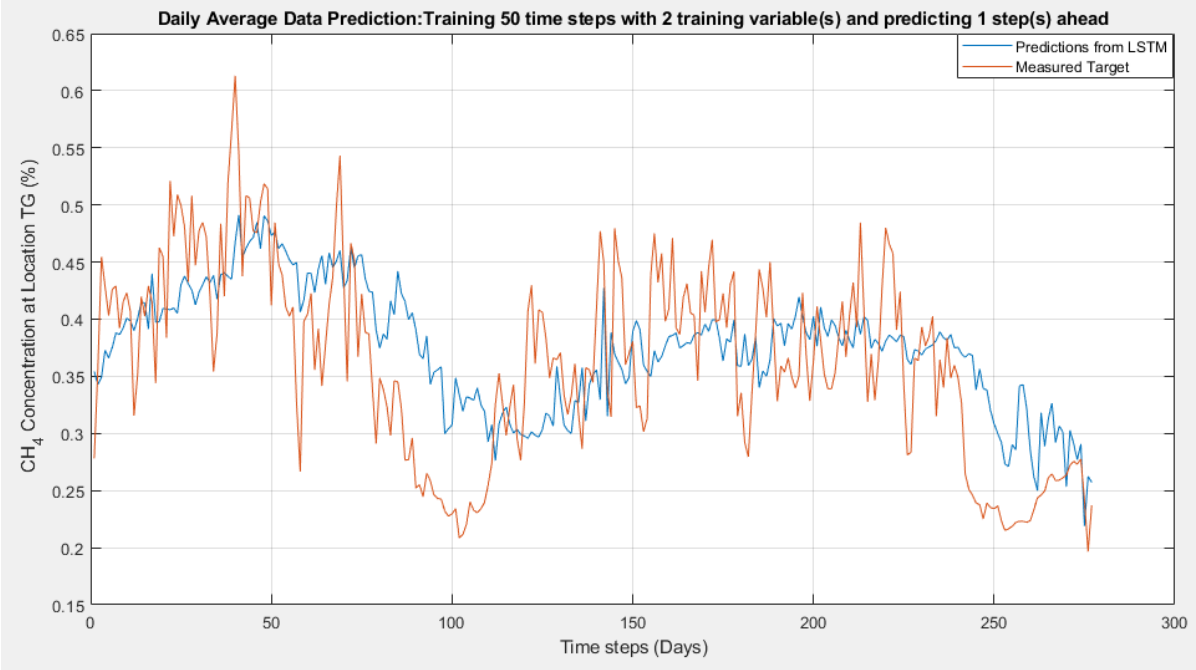

Figure 5-50 LSTM one-step-ahead prediction for the methane concentration at location TG using the unfiltered data and a training window size of fifty time-step

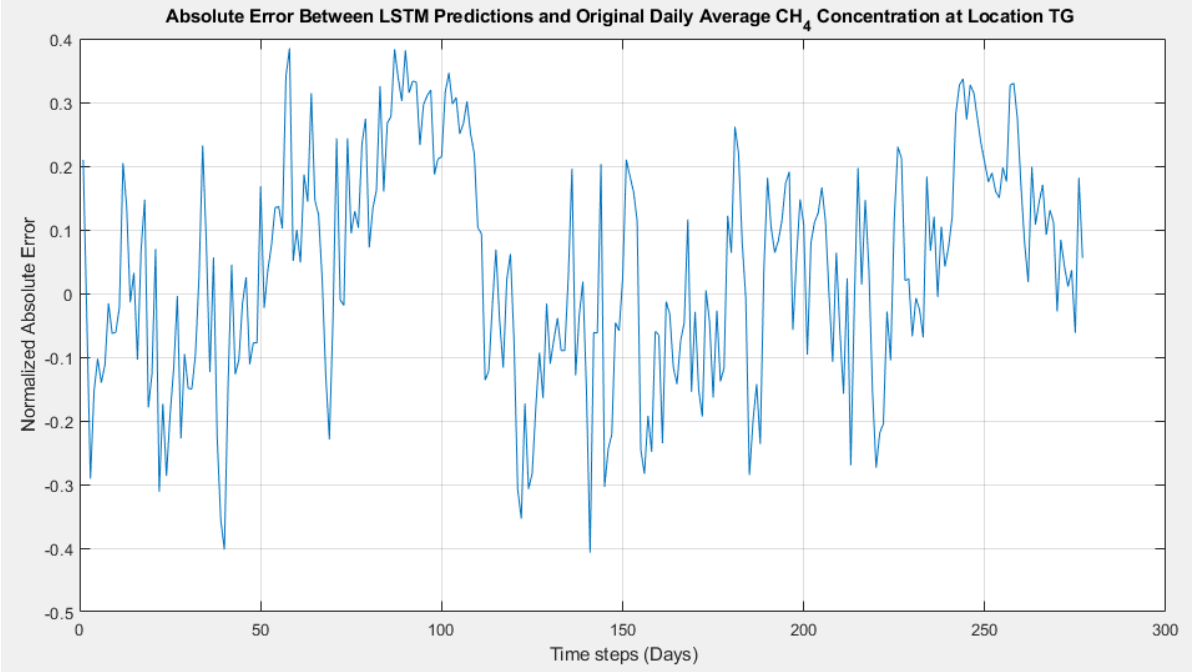

Figure 5-51 Absolute error for the LSTM one-step-ahead prediction for the methane concentration at location TG using the unfiltered data and a training window size of fifty time-steps

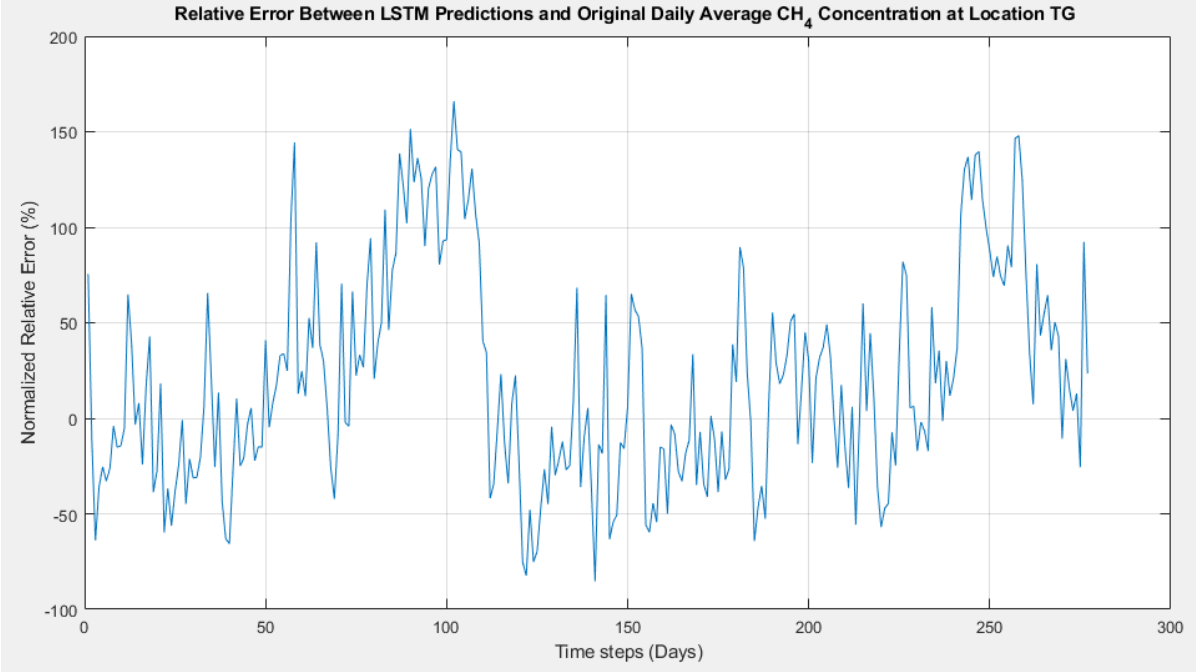

Figure 5-52 Relative error for the LSTM one-step-ahead prediction for the methane concentration at location TG using the unfiltered data and a training window size of fifty time-steps

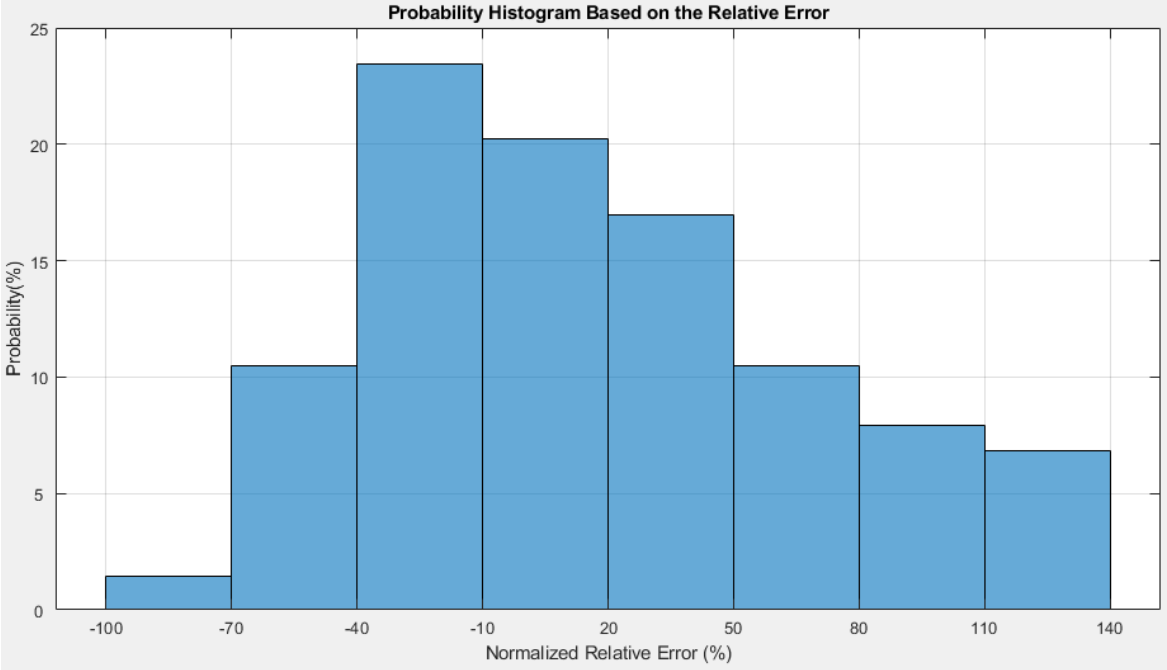

Figure 5-53 Probability histogram based on the relative error for the LSTM one-step-ahead prediction for the methane concentration at location TG using the unfiltered data and a training window size of fifty time-steps

The MSE and RMSE for this simulation were 0.004208 and 0.0649, respectively. The time spent training the first fifty steps was 23 seconds and for the whole simulation was 83 minutes.

The same data set was used; however, the sliding window size was one hundred data points this time. The LSTM has to predict one step ahead.

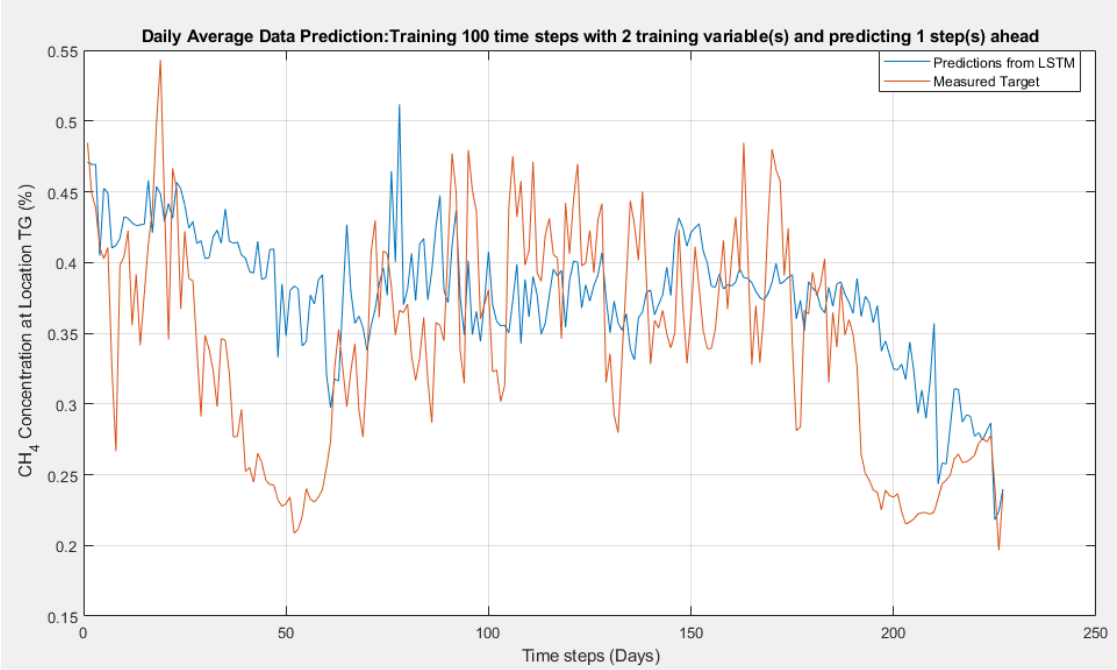

Figure 5-54 LSTM one-step-ahead prediction for the methane concentration at location TG using the unfiltered data and a training window size of one hundred time steps.

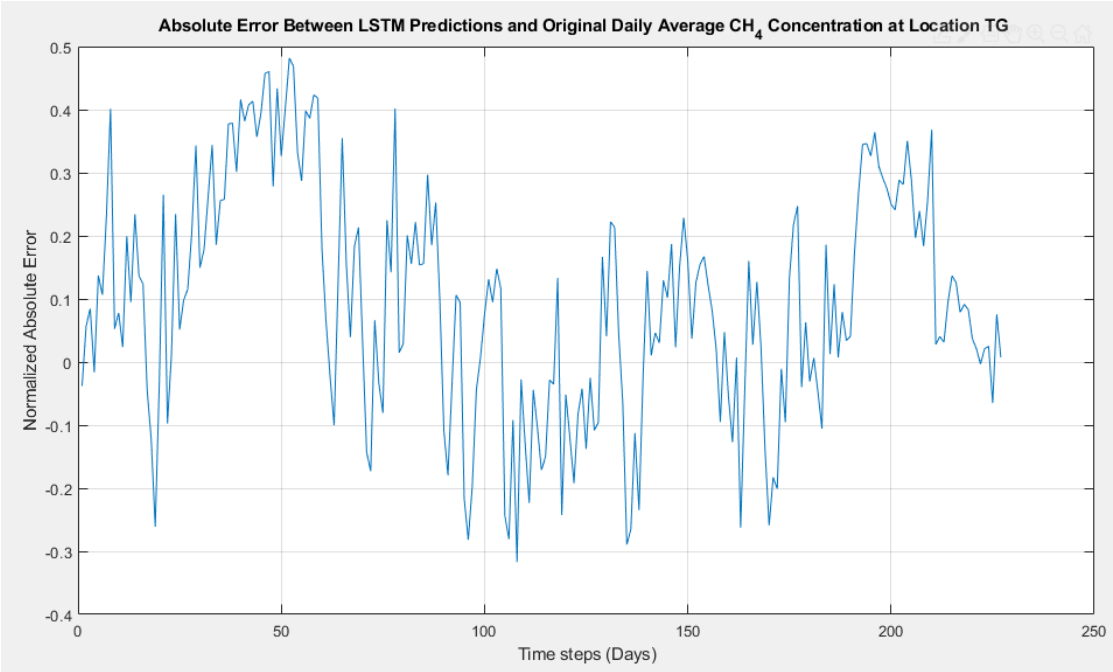

Figure 5-55 Absolute error for the LSTM one-step-ahead prediction for the methane concentration at location TG using the unfiltered data and a training window size of one hundred time steps

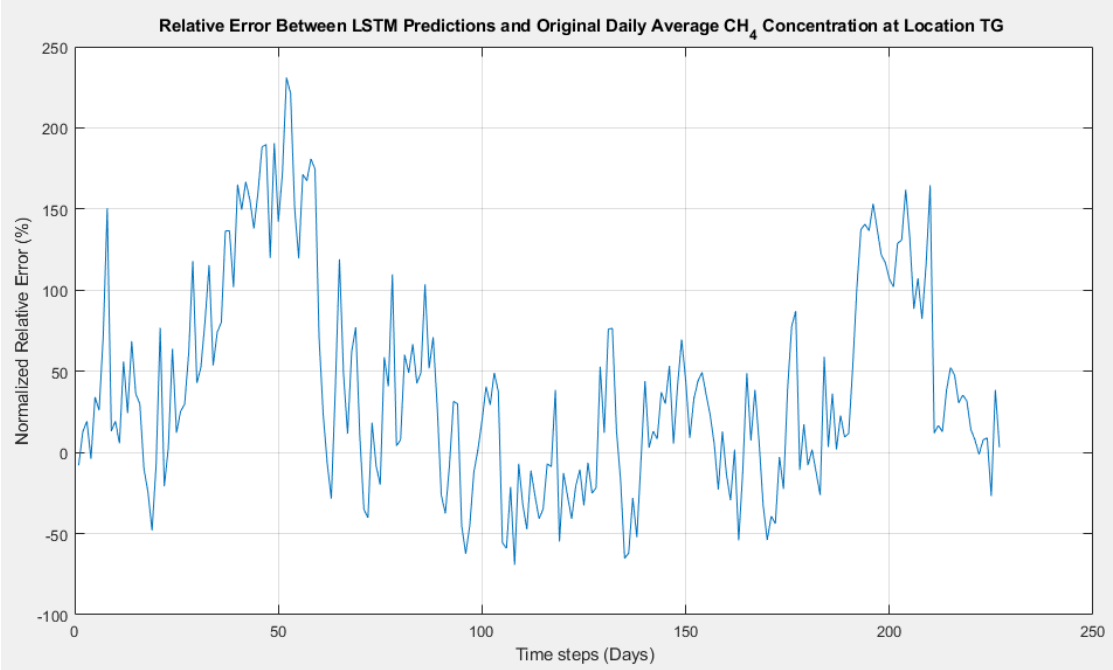

Figure 5-56 Relative error for the LSTM one-step-ahead prediction for the methane concentration at location TG using the unfiltered data and a training window size of one hundred time steps

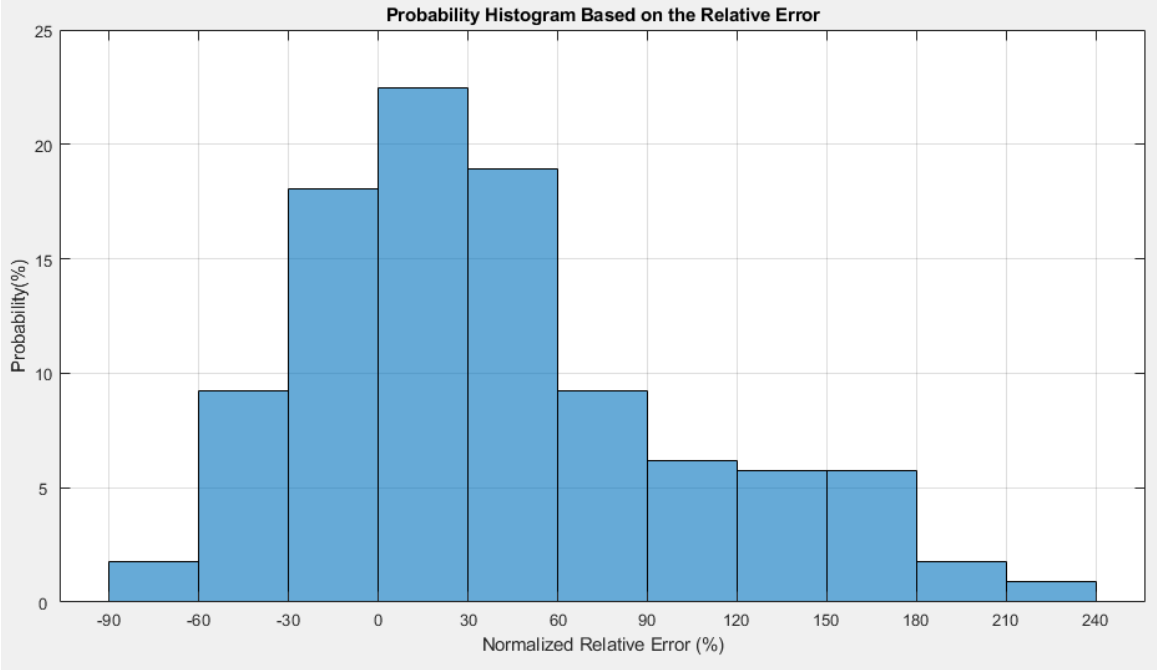

Figure 5-57 Probability histogram based on the relative error for the LSTM one-step-ahead prediction for the methane concentration at location TG using the unfiltered data and a training window size of one hundred time steps

The time spent training the first 100 points was 37 seconds and for the whole simulation was 84 minutes. The MSE and RMSE are 0.0055 and 0.0741, respectively.

5.11 Testing the Case Where the Next Day Prediction is the Previous Day Value

The results from the methodology presented in section 4.9 are displayed below.

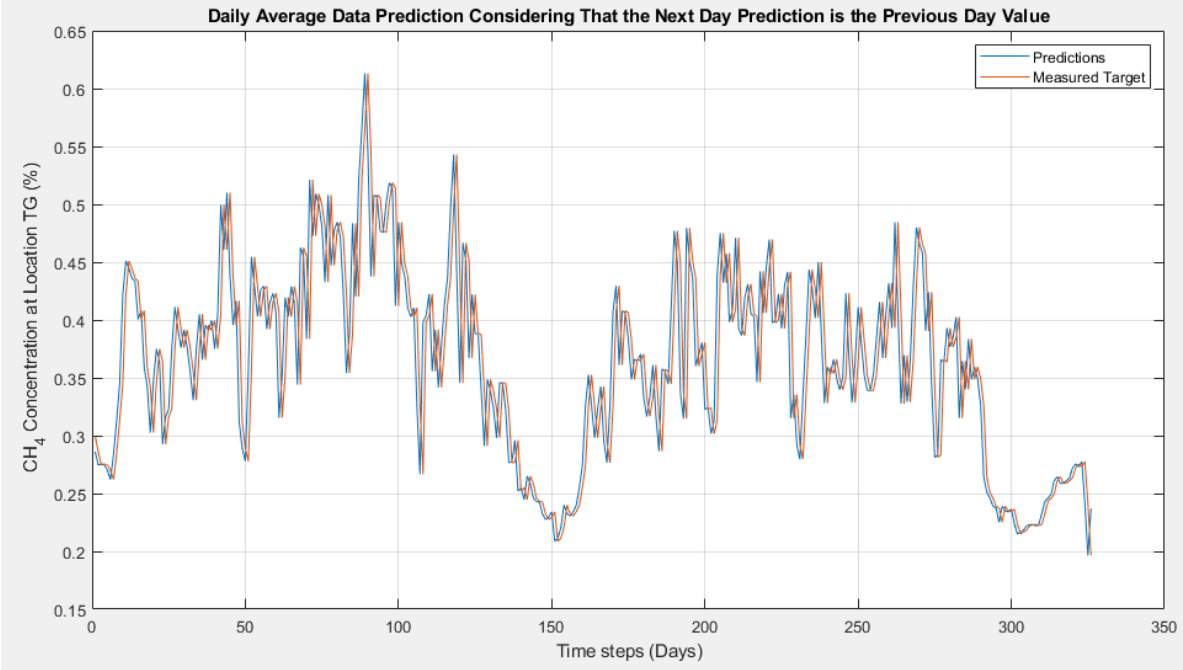

Figure 5-58 one-step-ahead prediction for the methane concentration at location TG using the unfiltered data and considering that the next day prediction is equal to the previous day value

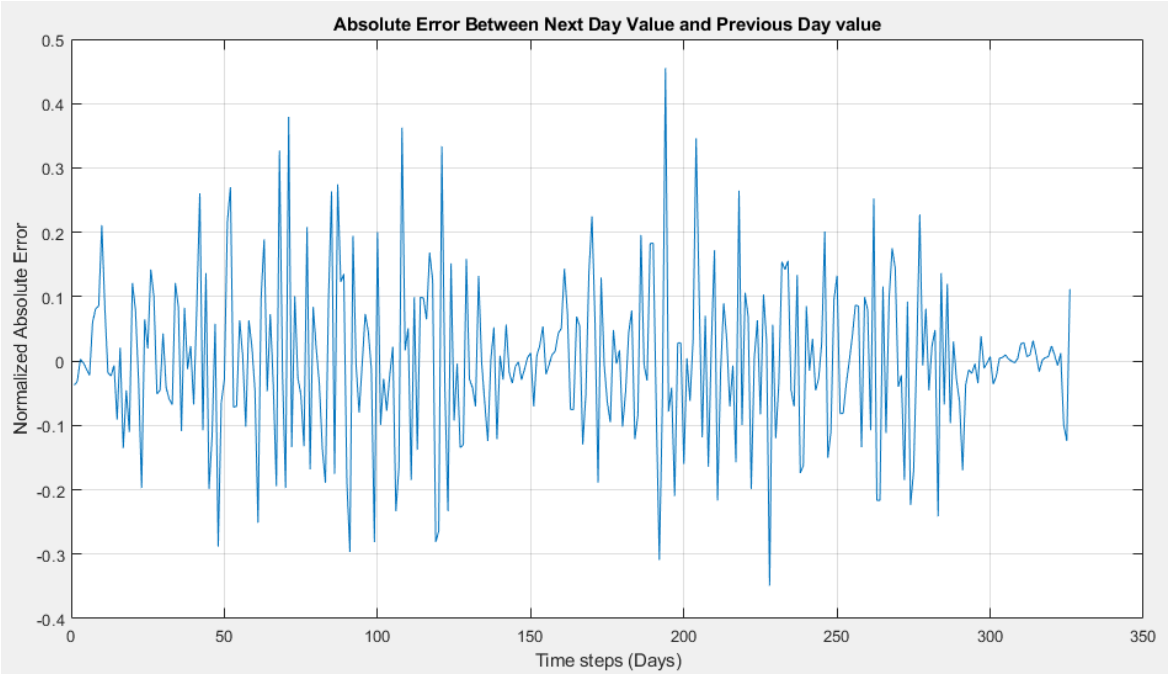

Figure 5-59 Absolute error for one-step-ahead prediction for the methane concentration at location TG using the unfiltered data and considering that the next day prediction is equal to the previous day value

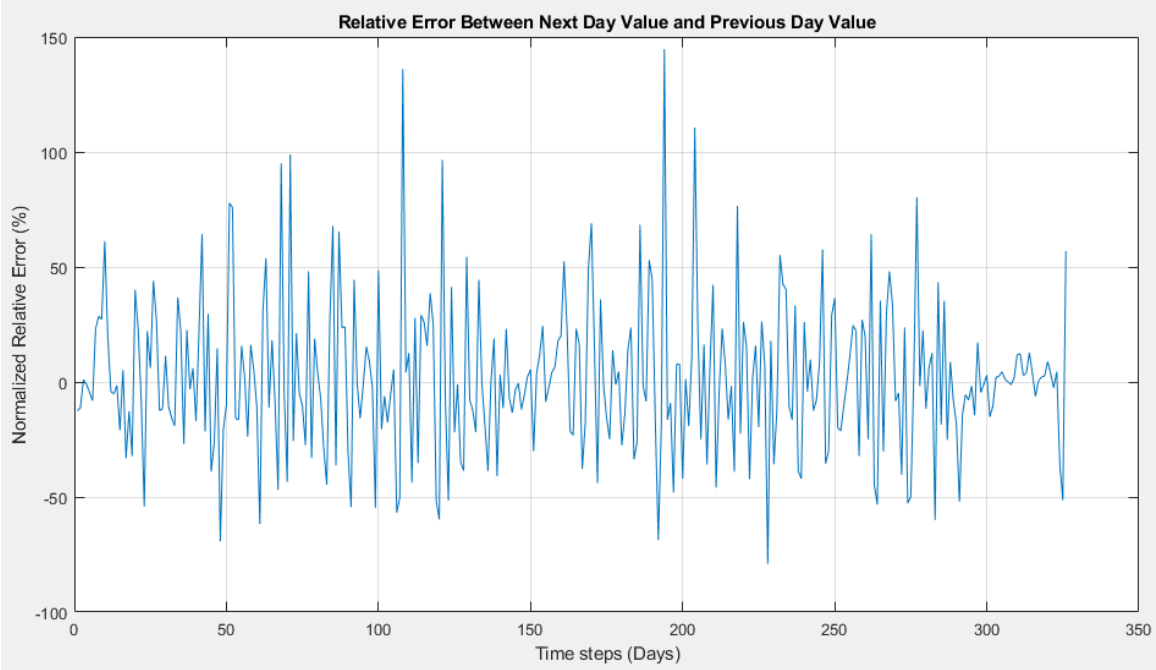

Figure 5-60 Relative error for one-step-ahead prediction for the methane concentration at location TG using the unfiltered data and considering that the next day prediction is equal to the previous day value

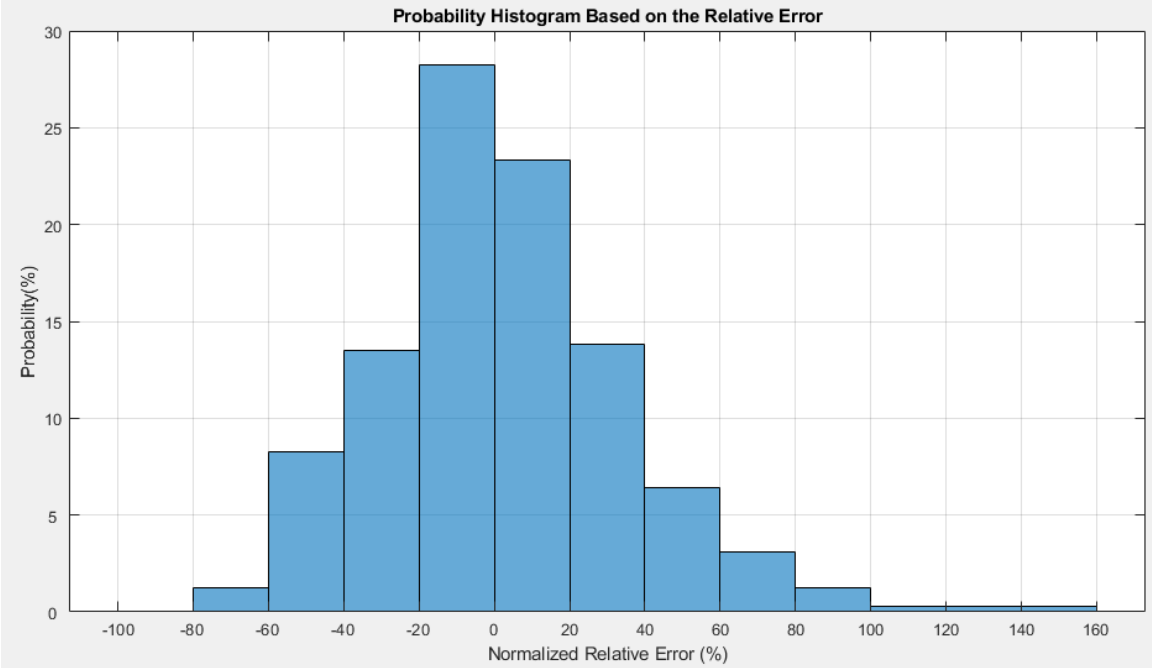

Figure 5-61 Probability histogram plot for one-step-ahead prediction for the methane concentration at location TG using the unfiltered data and considering that the next day prediction is equal to the previous day value

The MSE and RMSE for this simulation are 0.002003 and 0.0448, respectively.

5.12 Testing the Influence of Steps Ahead Predicted

This section follows the methodology described in section 4.8. The LSTM will be trained using two training variables, the airflow and the methane concentration at location MG. It will use a training window size of five data points and the unfiltered data.

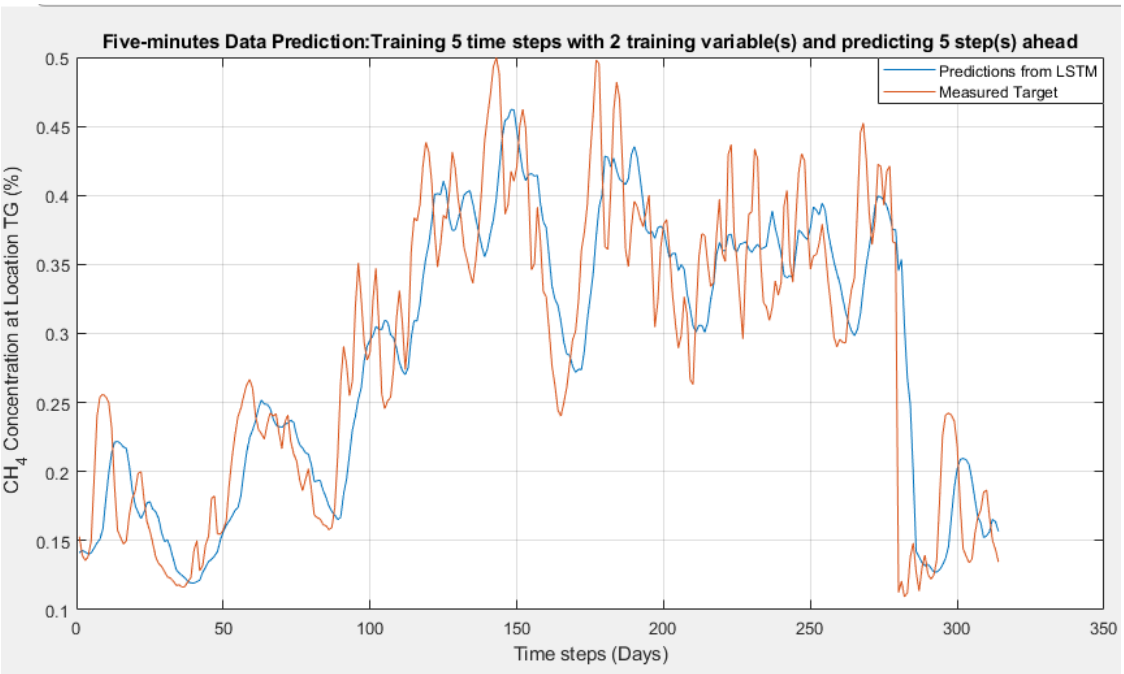

Figure 5-62 LSTM five-steps-ahead prediction for the methane concentration at location TG using the unfiltered data

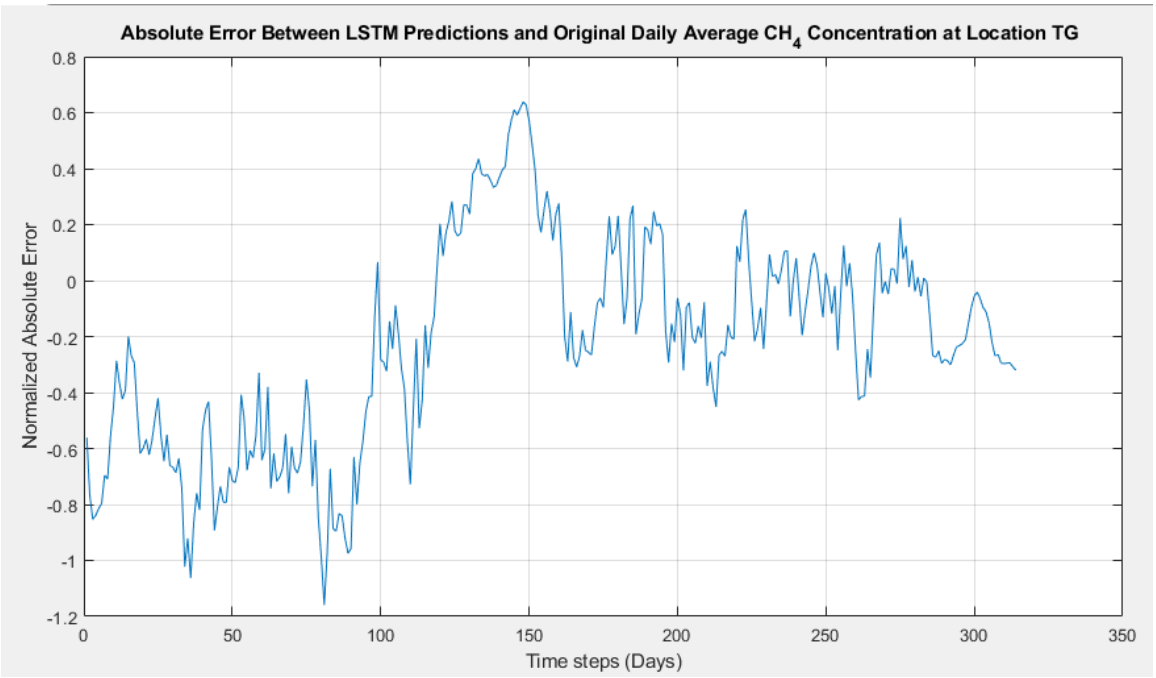

Figure 5-63 Absolute error for the LSTM five-steps-ahead prediction for the methane concentration at location TG using the unfiltered data

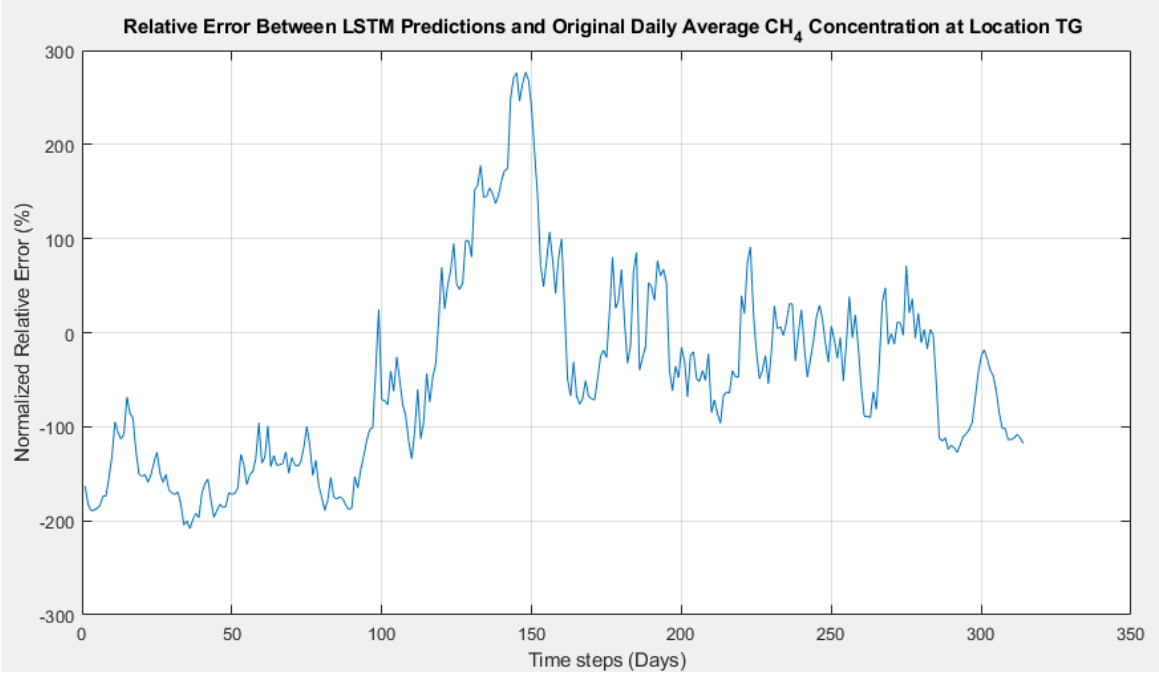

Figure 5-64 Relative error for the LSTM five-steps-ahead prediction for the methane concentration at location TG using the unfiltered data

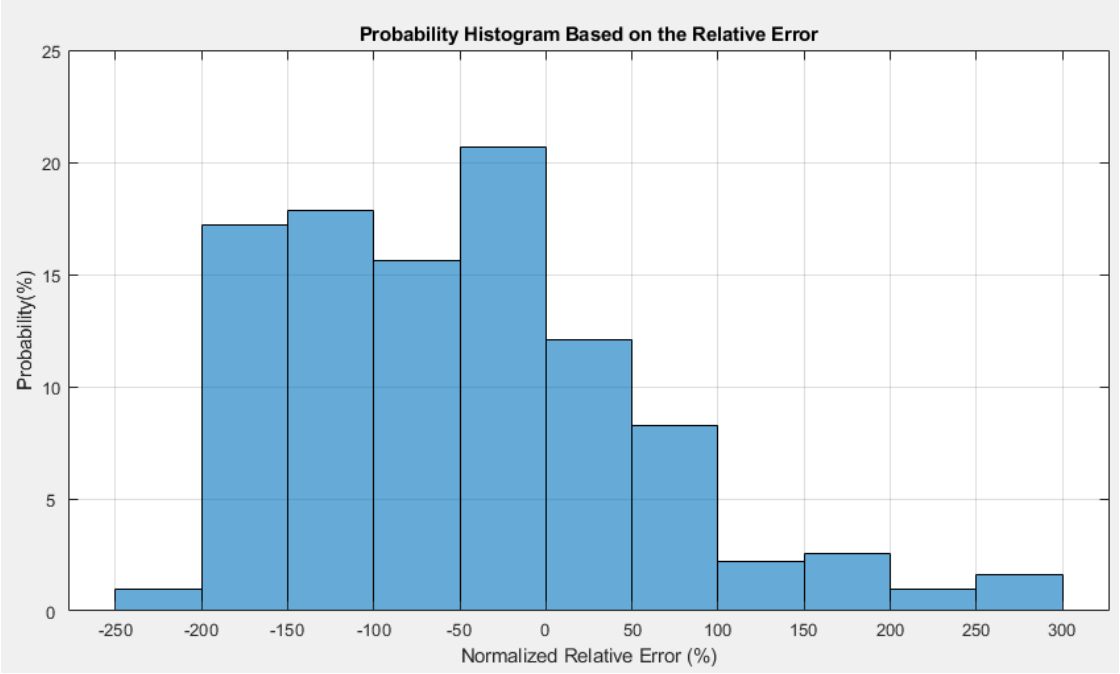

Figure 5-65 Probability histogram based on the relative error for the LSTM five-steps-ahead prediction for the methane concentration at location TG using the unfiltered data

The MSE and RMSE for this simulation are 0.002783 and 0.0528, respectively. The time spent training and predicting the first five data points was 5 seconds and for the whole data set was 27 minutes.

5.13 Testing overfitting

This section presents the results of section 4.10. Here, the LSTM uses regularization techniques to avoid overfitting. The results are shown below:

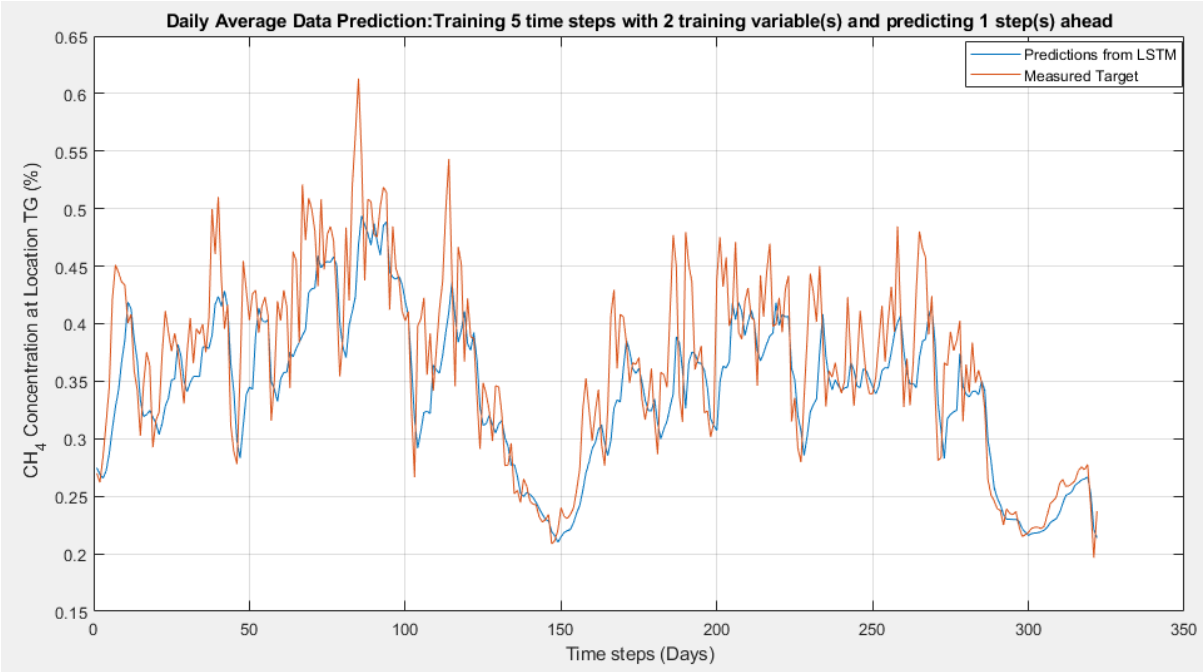

Figure 5-66 LSTM one-step-ahead prediction for the methane concentration at location TG using the unfiltered data and regularization methods to avoid overfitting.

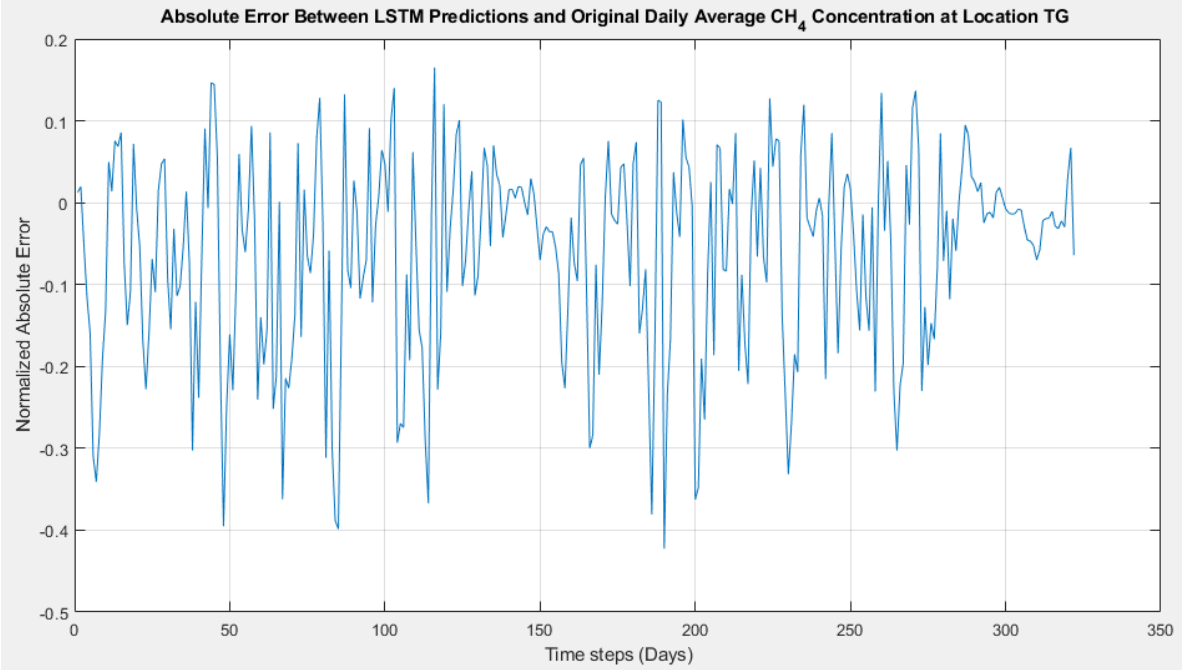

Figure 5-67Absolute error for the LSTM one-step-ahead prediction for the methane concentration at location TG using the unfiltered data and regularization methods to avoid overfitting.

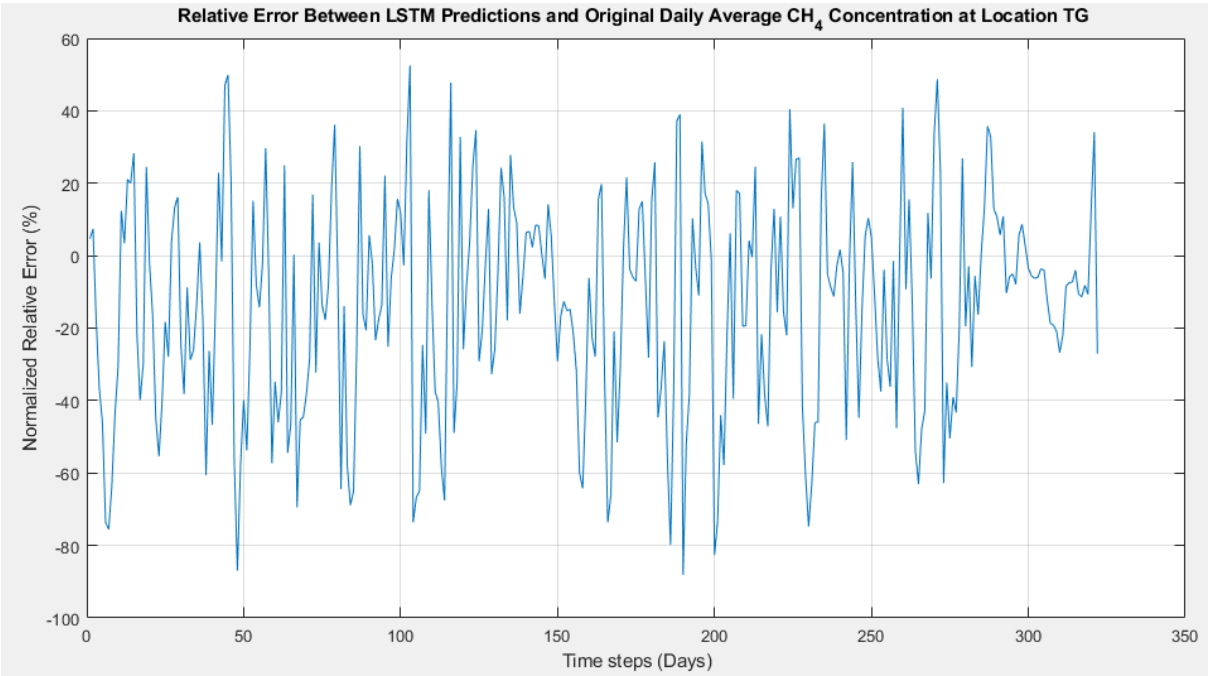

Figure 5-68 Relative error for the LSTM one-step-ahead prediction for the methane concentration at location TG using the unfiltered data and regularization methods to avoid overfitting.

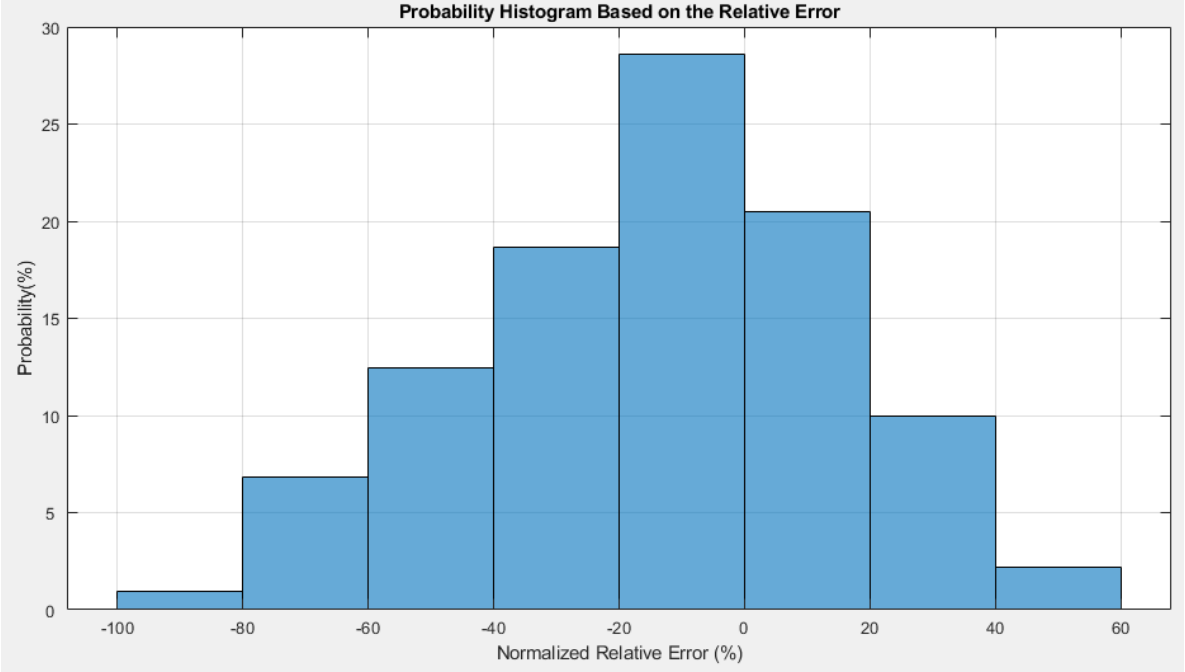

Figure 5-69 Histogram plot based on the relative error for the LSTM one-step-ahead prediction for the methane concentration at location TG using the unfiltered data and regularization methods to avoid overfitting.

The MSE and RMSE are 0.002449 and 0.0495, respectively. The time spent for training just the first five points was 4 seconds, and the whole dataset was 20 minutes.

6. DISCUSSION

6.1 LSTM Configuration

The tests results shown in section 5.1 show that the tuning parameter that most affects the LSTM performance is the learning rate. The fine-tuning presented in Table 5-3 shows that even a small change in the learning rate can significantly impact the error metrics. However, changing the learning rate does not affect the time spent on training and predicting. Despite not having a significant impact on

the LSTM performance, the number of LSTM layers seems to have a high impact on training and predicting time. Increasing the number of hidden units decreased the LSTM performance; however, it did not impact the time spent for training and predicting.

6.2 LSTM Results Using the Artificial Data

The table below shows a summary of the results from the LSTM using the artificial data (sine waves)

Table 6-1 Summary of the results from the simulation using artificial data

Number of variables used for training	Number of steps ahead predicted	MSE	RMSE	Time (minutes)
1	1	0.001024	0.0320	29
1	5	0.002831	0.0532	29
3	1	0.0007604	0.0276	28
3	5	0.002521	0.0502	28

The artificial data results, either using one variable for training or three variables, show that the predictions are accurate. When just one training variable is used, there is no difference in the time spent for making the predictions for the whole data set when comparing one-step-ahead predictions or five-step-ahead predictions. However, It is observed that the error increases when the number of time-steps-ahead predicted increases. Furthermore, it is seen, in both cases, that the predictions seem to be shifted when compared with the target. When three variables are used for training, it is observed that it significantly impacts the LSTM predictions. The MSE is very low when compared to just one training variable. It is worth noticing that the target that the LSTM has to match when it is training with three variables is more complex than when it is trained with just one variable.

Maybe, one explanation why the LSTM performs better with more training variables is that it can obtain more relevant information between the training variables and the target.

6.3 LSTM and Time Series Filter Results Using the Measured Data from the Mine

6.3.1 LSTM Results

The table below summarizes the results obtained from the simulations shown in section 5.5. The LSTM uses the airflow and methane concentration at location MG for training and predicts the methane concentration at location TG. The LSTM uses a training window size of five-time steps.

Table 6-2 Summary of results from the simulation using measured data from the mine

		Number of variables used for training	Number of steps ahead predicted	MSE	RMSE	Time for training first set of data(seconds)	Time for training the whole set of data(minutes)
Unfiltered data	LSTM	2	1	0.002125	0.0461	6	29
	TSF	1	1	0.004619	0.0680	-	0.00866
Filtered data	LSTM	2	1	0.001679	0.0410	5.2	28
	TSF	1	1	0.001345	0.0367	-	0.003166

Table 6-2 shows that filtering the data positively impacts the predictions. Comparing figures 5-13 and 5-17, it is seen that the shifting-looking behavior is still present in the predictions. Regarding the time spent making the predictions, filtering does not have a high impact. If the time is compared with the time spent using the artificial data, it is observed that they are about the same. The relative error plot using the unfiltered data shows that there are some high error picks. However, the probability histogram in Figure 5-16 shows that the probability of occurrence of these high picks (±100 to ±60) is below 5%.

Furthermore, the probability histogram shows that the lowest error picks (-20% to 0% to 20%) have the highest chance of occurrence, around 25% of the time. The probability histogram for the unfiltered data also shows that based on the error picks from (-60% to -40%), the LSTM is undershooting the predictions. Despite being in a low probability range, this might become a severe issue since predicting a point below the original value can lead to a failure in the early warning system, and a disaster might happen. It is better to be on the safe side, where the LSTM predicts a point higher than the original value, so in this case, the early warning system is activated, and due precautions are taken.

Analyzing the results using the filtered data in the same section, table 6-2 shows that the accuracy of the predictions increased based on the lower values of MSE and RMSE errors. The relative error plot in Figure 5-19 coupled with the results from the probability histogram in Figure 5-20 shows that the occurrence of high picks decreased. These high picks (±100% to ±40%) may reach around 5.5% of

the time. Furthermore, the probability of the lower error picks (-20% to 0% to 20%) increased a little compared with the unfiltered data. The simulation with the filtered data confirms that filtering positively affects the predictions. The issue of undershooting the predictions also happens in this simulation. Now, the range is from (-80% to -60%) relative error. As mentioned before, the chances of occurrence are below 5%, but this might become a problem.

6.3.2 Time Series Filter Results

Table 6-2 shows that the accuracy of the TSF is also affected when the filtered data is used. Analyzing the results using the unfiltered data, the relative error plot in Figure 5-23 and the probability histogram in Figure 5-24 show very high error picks; however, these picks (±140 to ±80) are below a 5% chance of occurrence. As discussed in the previous section, having predicted values lower than the original values may cause a failure in the early warning system. The probability histogram for the unfiltered data shows that in the range of (-20% to 0%) relative error, the TSF undershoots the predictions compared with the original values. This range presents a probability of almost 25%. Suppose the predictions coming from the TSF are not reliable enough. In that case, it might predict a point that is below the original value. Following preceding discussions, this might become a problem if the original value is above permissible limits.

The results from the filtered data show that the very high error picks decreased. However, there is still error picks reaching 90% relative error. The lowest error picks (-10% to 0% to 10%) have the highest probability, around 18%. In this

simulation, the probability histogram plot is less symmetric (in the range of -30% to 0% to 30% relative error) than the one using the unfiltered data. In the range of (-20% to -10%) relative error, the histograms show that the TSF will undershoot the predictions, with a probability of around 14%.

6.3.3 Comparing the Predictions from the LSTM with the TSF

Table 6-2 shows that if the unfiltered data is used, the LSTM gives better predictions than the TSF. However, the TSF will give better predictions if the filtered data is used. Comparing the unfiltered data results, the LSTM probability histogram plot (Figure 5-16) shows that the range of relative error picks for LSTM predictions reach a maximum of ±100%. On the other hand, the TSF probability histogram plot (Figure 5-24) shows that the errors reach ±140%. Moreover, the low relative error range (-20% to 0% to 20%) in the LSTM predictions presents almost the same probability of occurrence (around 25%). Whereas in the TSF histogram plot, the probability of occurrence of this range is below 25%. Comparing the filtered results, Figure 5-20 (LSTM) and Figure 5-28 (TSF), the histogram plots show that despite having less high error pick than the raw data, the LSTM has more high error picks than the TSF. Furthermore, the LSTM presents a higher probability of occurrence for the lower relative error range (-20% to 0% to 20%) than the TSF. Moreover, the histogram plot for the TSF shows that in the relative error range of (±40% to ± 20%), the probability of occurrence reaches a maximum of 8%. Whereas in the LSTM histogram plot, the probability is around

15% for the same range. Comparing the time spent for training and predicting, the TSF is much faster than the LSTM, regardless of the data used (filtered or not).

6.4 LSTM and TSF Results Using the Root-Cause CH_4 Source

The table below shows a summary of the results obtained from section 5.7. The LSTM uses a training window size of five-time steps.

Table 6-3 Summary of the LSTM and TSF results using the line source

		Number of variables used for training	Number of steps ahead predicted	MSE	RMSE	Time for training first set of data(seconds)	Time for training the whole set of data(minutes)
Unfiltered data	LSTM	1	1	0.005861	0.0766	16	28
	TSF	1	1	0.01044	0.1022	-	0.004166

Table 6-3 shows that the predictions from the LSTM are better than the TSF. Comparing the probability histogram plots Figure 5-33 (LSTM) and Figure 5-37 (TSF), it is observed that the LSTM presents a lower probability of having high error picks. Moreover, the LSTM histogram plot shows that in the range where the predictions are undershot (-60% to 0%), the probability stays around 7%. Whereas for the TSF model, in the range of (-180% to 0%), the probability goes above 10%. Both histogram plots show that the LSTM and TSF predict points above the original values most of the time. This effect can be verified in Figure 5-30 and Figure 5-34.

Comparing the time for training and predicting, the TSF is much faster than the LSTM.

6.4.1 Comparing the Results from the LSTM and TSF using the Direct Methane Concentration Versus the Results Using the Root-Cause

The table below summarizes the LSTM results when the methane concentration is used directly versus the line source. The LSTM uses a training window size of five-time steps.

Table 6-4 Summary showing the results when the CH_4 is directly used versus the line source(results for the unfiltered data)

		Number of variables used for training	Number of steps ahead predicted	MSE	RMSE	Time for training first set of data(seconds)	Time for training the whole set of data(minutes)
Direct CH_4	LSTM	2	1	0.002125	0.0461	6	29
	TSF	1	1	0.004619	0.0680	-	0.00866
Line Source	LSTM	1	1	0.005861	0.0766	16	28
	TSF	1	1	0.01044	0.1022	-	0.004166

Comparing the time spent for training and predicting, there is just a one-minute difference in the LSTM results. The MSE and RMSE results from Table 6-4 show

that the LSTM performs better if the direct CH_4 concentration is used. Comparing the probability histogram plots, Figure 5-16 (direct CH_4) and Figure 5-33 (line source), it is observed that the direct CH_4 gives a more symmetric distribution. Furthermore, the low relative error range (±20% to ± 20%) using direct CH_4 gives a probability of around 25%. Using the line source, the probability varies around 7% to almost 15% for the same range of relative error values.

Moreover, the line source histogram plot shows high error picks reaching 180%, while the direct CH_4 limits to 100%. Additionally, the histogram for the line source shows that high error picks from 60% to 80% relative error happens almost 20% of the time. Whereas for the direct CH_4, the probability for the same range is below 5%.

The TSF results based on the MSE and RMSE of Table 6-4 show that using the direct CH_4 concentration gives better predictions. Comparing the histogram plots Figure 5-24 (direct CH_4) and Figure 5-37 (line source), it is possible to see that the predictions coming from the line source have more high relative error picks. Comparing the interval of (±140% to ± 20%) on the direct CH4 predictions, it is observed that the probability of occurrence stays below 10%. Whereas for the line source, the probability varies reaches 10% and 25% for the negative and positive sides of the plot, respectively.

6.5 Influence of the Number of Training Variables in the LSTM Predictions

The table below summarizes the results when LSTM is trained with the airflow rate, the methane concentration at location MG and the barometric pressure; versus

when it uses just the airflow rate and the methane concentration at location MG. The LSTM uses a training window size of five-time steps.

Table 6-5 Summary for the results when the LSTM is trained with three variables versus when it is trained with two variables

		Number of variables used for training	Number of steps ahead predicted	MSE	RMSE	Time for training first set of data(seconds)	Time for training the whole set of data(minutes)
Unfiltered data	LSTM	3	1	0.002301	0.0480	5	29
Unfiltered data	LSTM	2	1	0.002125	0.0461	6	29
Filtered data	LSTM	3	1	0.001641	0.0405	9	31
Filtered data	LSTM	2	1	0.001679	0.0410	5.2	28

The results show that increasing the number of training variables causes a decrease in the LSTM performance when the unfiltered data is used. One possible explanation is that when the number of variables is increased, the LSTM experience more noise during training since the data is not filtered. When the trained model tries to forecast, the predicted values reflect the noise in the training set. Another possible explanation is that the third variable included in the training

set has little or no influence on output prediction. However, since the LSTM uses it for training, this variable introduces more noise to the system. It makes it even harder for the LSTM to make the predictions. The time spent on training and predicting is the same. Using the filtered data, the predictions are about the same regardless of the number of variables used for training. Comparing the histogram plots for the unfiltered data, Figure 5-16 for the case with two training variables and Figure 5-41 for the case with three training variables, it is observed that the latter case increases the range of the relative error. Furthermore, training with two variables gives a more symmetric probability distribution. Comparing the low relative error range (±20% to ± 20%) on both histogram plots, it is observed that the probability is higher for the case where two variables are used for training.

Comparing the histogram plots for the filtered data, Figure 5-20 (two variables) and Figure 5-45 (three variables), it is seen that the range of high error picks increased when three variables were used for training. On both histogram plots, the relative error range of (±100% to ± 20%) stays below the probability of 15%. For the low relative error range (±20% to ± 20%), the probability is about the same, around 25%.

6.6 Influence of Training Data Size

The table below summarizes the results when the training window size is increased to ten. The results are presented for the unfiltered data. Variables used for training were the methane concentration at location MG and the airflow rate.

Table 6-6 Summary of the LSTM results for a training window size of ten-time-steps

		Number of variables used for training	Number of steps ahead predicted	MSE	RMSE	Time for training first set of data(seconds)	Time for training the whole set of data(minutes)
Training window size of 10	LSTM	2	1	0.002437	0.0494	16	64
Training window size of 50	LSTM	2	1	0.004208	0.0649	23	83
Training window size of 100	LSTM	2	1	0.0055	0.0741	37	130
Training window size of 5	LSTM	2	1	0.002125	0.0461	6	29

The results show that increasing the number of time-steps used for training did not positively impact the LSTM performance. Both the error and the time spent predicting increased. Comparing Figure 5-16 (training window size of five) and Figure 5-49 (training window size of ten), it is seen that increasing the training window size makes the histogram less symmetric. The probability of the low relative error range (±20% to ± 20%) is below 25% when the training window size increases. Furthermore, the range of high error picks increases compared to the case where five steps are used for training. Looking at Figure 5-46, it seems that the shifting-looking behavior increased compared to the training window size of

five (Figure 5-13). Analyzing the case where the training window size is increased to fifty-time steps, Figure 5-50 shows that the predictions are very off compared to the target. The histogram plot in Figure 5-53 shows that the LSTM is overshooting the predictions most of the time. The number of high picks increases compared to the case where the window size is five, and the probability of these high picks has increased. When one hundred time steps are used for training the model, Figure 5-54 shows that this case is the worst scenario. The histogram plot in Figure 5-57 shows that the number of high picks has increased and that the LSTM is overshooting the predictions most of the time.

The assumption made in the methodology, section 4.7, does not seem to be true. The LSTM does not seem to capture the relationship between training variables and target more efficiently when the training window size increases. Maybe when the training size increases, the LSTM experiences more fluctuations in the data. Later, when it tries to account for these fluctuations, it predicts less accuracy. Another explanation for the poorer performance of the LSTM when the training window size increases is that it carries information that is not useful for the predictions. In the case where the LSTM is trained with a larger amount of data (training window size of 50 and 100 time-steps), the internal structure of the LSTM generates weights and biases for each slide of the window. The LSTM has to predict the next point based on the weights and biases created in this training. The problem is that, possibly, the methane release on days one and two, for example, does not affect the methane release on day 51 (for the training window size of 50). It also does not affect the methane release on day 101 (for the training window

size of 100). However, as the LSTM was trained with this larger amount of data, it considers the effects of days one and two. Therefore, the LSTM might be experiencing more noise and have outdated constants because of the large number of points used.

Thus, the constants generated (weights and biases) are unsuitable for predicting the next day and subsequent days because they reflect unuseful information from very long past days. On the other hand, when the sliding window of five data points is used, it also carries information from past days; however, just useful information without overloading the model. Therefore, for the case studied in this paper, the methane concentration can be regarded as having a low memory with time.

6.7 Verifying the Influence of Steps Ahead Predicted

The table below compares the LSTM predictions using the five-minutes-average and daily average values. The training variables used were the methane concentration and airflow rate at location MG. The LSTM uses a training window size of five-time steps.

Table 6-7 Summary for the LSTM results using the 5-minutes-average data and predicting three steps ahead

	Number of variables used for training	Number of steps ahead predicted	MSE	RMSE	Time for training first set of data(seconds)	Time for training the whole set of data(minutes)

Unfiltered data(daily average)	LSTM	2	1	0.002125	0.0461	6	29
Unfiltered data(5-min-average)	LSTM	2	5	0.002783	0.0528	5	27

The results show that increasing the number of steps ahead predicted, the error increases. Comparing the probability histogram error plots, Figure 5-16 (two predictions ahead) and Figure 5-65 (five predictions ahead), it is seen that for the latter case, the probability histogram plot is not symmetric anymore. For the case where five predictions are made, the histogram plot shows high error picks (-200% to 0%) relative error that almost reached 20% probability. This might generate an issue with the early warning system, as previously discussed. Besides high error picks, the shift-looking behavior of the predictions also increased. One of the requirements for the early warning system is that it has to give enough time in case the LSTM predicts that a disaster might happen. Following the results shown in this section, the accuracy for predicting five minutes ahead is not enough to use the LSTM as an early warning system. If used as an early warning system, one step ahead must be considered for reasonable accuracy.

6.8 Evaluating the Case Where the Next Day Prediction is the Previous Day Value

The most straightforward case tested in this thesis is considering that the next day prediction is the previous day's value. The results shown in section 5.11, Figure

5-58 show that the number of high picks throughout the predictions is very low compared to the LSTM. However, it is clearly visible that the predictions look shifted. The table below shows a comparison between this simple case and for the best LSTM case scenario.

Table 6-8 Comparison between the best-case scenario for the LSTM and the simple case considering no machine learning

		Number of variables used for training	Number of steps ahead predicted	MSE	RMSE	Time for training first set of data(seconds)	Time for training the whole set of data(minutes)
Unfiltered data(daily average)	LSTM	2	1	0.002125	0.0461	6	29
Filtered data	LSTM	2	1	0.001679	0.0410	5.2	28
Unfiltered data(daily average)	Simple case	1	1	0.002003	0.0448	-	0.001

Table 6-8 shows that if the unfiltered data is used, the predictions from the LSTM are slightly worse than the simple case. If the filtered data is used, the error coming from the LSTM is lower. However, considering that there is no algorithm implementation with the simplest case, it is safe to use it. Furthermore, the time spent generating the predictions using this simple case is extremely faster than the LSTM. Figure 5-61 shows the histogram plot for this simple case. It shows that the

few high picks throughout the predictions happen in a very low probability range. Furthermore, it shows that the predictions for the low probability range (-20% to 0 to +20%) is almost the same for the LSTM best-case scenario. For just one step ahead prediction, using this simple case is more advantageous than using the LSTM or the TSF.

6.9 Testing for Overfitting in the LSTM

This section considers that overfitting may have happened in the LSTM best-case scenario. Therefore, L2 regularization and dropout techniques were used to avoid overfitting. The table below summarizes the results obtained.

Table 6-9 Testing overfitting in the LSTM

		Number of variables used for training	Number of steps ahead predicted	MSE	RMSE	Time for training first set of data(seconds)	Time for training the whole set of data(minutes)
Unfiltered data(daily average)	LSTM	2	1	0.002125	0.0461	6	29
Unfiltered data(L2, dropout)	LSTM	2	1	0.002449	0.0495	4	20

As shown in Table 6-9, adding the regularization terms did not improve the predictions of the LSTM. The errors are more prominent than in the previous case. One possible explanation for the increase in error is the addition of the dropout

layer. As discussed before in the introduction, this technique randomly deactivates neurons during training to avoid overfitting. From the LSTM configuration, the best number of neurons is three. Suppose the dropout layer deactivates one of them during training. In that case, it is expected that the error for that simulation will be bigger than for the optimum number of neurons. This might also explain why the time spent on training is lower than when there is no regularization. Another explanation is that the LSTM randomly initializes the weights and biases during training, so the outcome of the predictions will be slightly different even if the same set of data is used. To reproduce the values every time the LSTM uses the same set of data, the random number generator (RNG) in MATLAB has to be set to "default" and then the seed value equal to one. Unfortunately, the RNG was not fixed for the LSTM best case scenario discussed. Two main conclusions can be taken from the results presented in Table 6-9. The first one is that overfitting did not happen. Thus there was no additional improvement in the accuracy of the LSTM model. The second conclusion is that the values used for the L2 regularization and dropout are not optimum.

6.10 Comparing the LSTM Results Using Measured Data from the Mine Versus Artificial Data and Results from Literature

The LSTM behaves differently when trained with artificial data (sine waves) versus measured data from the mine. From the artificial data predictions, the best predictions were those with three inputs, and even a five-step ahead prediction was accurate. The best predictions are for the training with two inputs with actual data. A smaller number of points must be predicted to give the best predictions. It

is also interesting to see that increasing the number of inputs trained, in this case, did not have a significant impact on the result of the predictions using the measured data from the mine. Based on the results presented in this study using the artificial data, it is safe to say that the LSTM improves the accuracy of the predictions when trained with data with a repetitive pattern (such as the sine functions tested) and is very smooth.

Considering the papers discussed in the literature review, where authors tried to forecast methane using neural networks, for instance (Mathatho.,2020); despite not giving any information about the training data size or the number of time steps forecasted, the authors obtained an MSE of 0.0019 and an RMSE of 0.0426. Comparing these numbers with the ones found in this study, they are very similar to the case where the LSTM was trained with five time-steps, using the filtered data and predicting one step (MSE of 0.001679 and RMSE of 0.0410). The author used some procedures similar to the ones presented in this study. For example, in that paper, the best number of neurons was chosen by increments of one until it converged to a minimum error. Comparing the results from this study with that of (Geng, 2016), the main difference is the type of neural network used in that paper. The author used partial chaos swarms optimized radial basis function neural network (details on how this neural network works can be found in the author's article). The best results were obtained using a three-layered neural network using this neural network. The best number of iterations was 250 (half of that used in this thesis study), and the best number of neurons was five (just two more than the ones used in this thesis study). Because of noise, the author preprocessed the

data before using it (using another method than the filter used in this thesis study). For that paper, the author used 6880 points for training the neural network and around 1700 (this number is not explicit in the text but can be estimated from the figure provided) for predicting. With this configuration, the author obtained an RMSE of 0.0265. Comparing this number with the one for the best-case scenario in this thesis (using the LSTM), the RMSE is 0.0410. It can be seen that the result obtained for (Geng, 2016) is better. This better result may be due to the different types of neural networks used in that paper. The author for that paper does not explicitly say how many steps ahead were predicted at a time. However, it is said in the paper that the data used had a collecting sampling rate of one minute. The author also does not explain how long it took for the model to train or forecast. The work done by (Yang et al., 2020) also presents a similar case as this thesis. Its similarity lies in the fact that the authors from that paper tried to forecast methane emissions from an underground coal mine. The first difference is that the authors used an improved grey radial basis function neural network (details on how this network works can be found in the author's article). Another difference is that six factors (coal seam depth, seam thickness, coal seam methane content, seam spacing, daily progress, daily output) related to methane concentration were used as training variables. In this thesis study, a maximum of three variables were used for training the LSTM. Another difference is that the authors used 275 data points for training the model and 25 data points for forecasting. It also used logarithmic processing to make the data smoother and improve the predictions. The results showed that the relative error stayed below 15%, with an average of 5.63%. This

value is relatively low compared to the relative error presented for the best case scenario using the LSTM in this paper (around 20%). The accuracy of the results from that paper, according to the authors, lies in the fact that the author combined a neural network with another predictor (the grey theory in that case), so the model's performance and accuracy improved as a whole. Despite not saying how many time steps were predicted at a time, the paper suggests that the twenty-five-time steps were predicted at once. The paper also does not give any information about the time taken to train the model and forecast. Another similar research was done by (Kumari et al., 2021). The author forecasted methane concentration in an underground coal mine, among other gases. The first difference from that paper to this study is that the author used a uniform manifold approximation and projection (UMAP) combined with the LSTM for dimension reduction of the data. This UMAP-LSTM model was trained with 143 data points (each data point corresponds to one day), and 31 days were used for forecasting. Using the UMAP-LSTM predictor, the best combination of hyperparameters for the model was 50 neurons (compared to 3 used in this thesis study), a learning rate of 0.001 (0.00068 used in this thesis), and 200 epochs (500 used in the thesis). The paper reports that a RMSE of 0.0995 was obtained from this configuration for the methane forecast. This number is very low compared to the one obtained in this thesis study for the LSTM best-case scenario. The main difference that can explain their low error is that the author used another predictor combined with the LSTM, improving the model's performance as a whole. The research done by (P.Lyu et al.,2020) presents an LSTM-based encoder-decoder for short-term prediction of methane concentration

in a coal mine. As outlined in the paper, the proposed model was used to predict methane concentration two minutes, four minutes, six minutes, eight minutes, and ten minutes ahead. The data sampling used was two minutes. Despite not saying the configuration of the proposed model, the author highlights that the LSTM model loses accuracy as the number of time steps forecasted increases. This conclusion is similar to the one found in this thesis; the higher the number of steps ahead, the worse the model's accuracy. That paper also reported that using data from several sensors is better than using a single sensor. This conclusion is similar to the one presented in this thesis study. Table 5-4 and Table 6-2 show the contrast between using just one variable for training versus using two variables for the LSTM model training. In this case, combining information from two sensors is better than using just data from one sensor.

The main conclusion from comparing results from the literature versus those presented in this thesis is that the predictions coming from the LSTM model are improved when combined with other predictors.

6.11 Recommendations in Case the Methane Concentration is Above Permissible Limits

According to chapter six of (NIOSH, 2006), by law, the methane concentration inside mines is restricted to 1%-1.25%. Therefore, if the threshold limit for methane is set to 1% in this present study, the accuracy of the predictions depends on how far the original point is from the threshold limit. For instance, consider that the

LSTM predicts that the next day's concentration will be 0.6% methane. However, the actual value for the next day's concentration is 0.4% methane. In this case, there is a 50% relative error. However, the prediction would still be acceptable since it is below the threshold limit. Analyzing the results shown in the discussion section, it can be concluded that if the LSTM is used as an early warning system, it has to be used with caution. Suppose the methane concentration measured from the mine is close to the threshold limit. In that case, the LSTM predictions might predict that it will be above permissible limits when in reality, it is not. If that is the case, the early warning system might activate, but it would be a false alarm.

The contrary situation also might be true and even more dangerous. Suppose the LSTM predicts that the next's day concentration is 0.9% methane. However, the actual value for the next-day concentration is 1.1% methane. There is a -22.2% relative error in this case, which is much better than a 50% error. The problem is that, in this case, the LSTM undershot the prediction. Therefore, a disaster might happen, and the early warning system alarm will not be activated because the predicted point is below the threshold limit. Since we deal with human lives and lots of money involved in the operation, an early warning system must be flawless.

The appealing side of the LSTM, or neural networks in general, is that one does not have to worry about the physics of the process under consideration, which can sometimes be very complex. The use of LSTM in this present study, for example, would simplify the need for understanding the relationship between the target and the several possible factors that influence methane release inside the mine.

Another important point to notice is that with a physics-based model, the more parameters it has, the greater the model's accuracy (although there is difficulty in collecting data for these parameters). However, as shown by the training using the LSTM in this study, increasing the number of parameters did not increase the model's accuracy. Therefore, one can conclude that the model is not smart enough to “see” the importance of that variable or consider the variable important. However, the model is so sensitive to the amount of noise in the data that it does not perform well. Looking at the cases reported in the literature review, the goal of forecasting methane using machine learning has been attempted by others, even though they are using combined techniques to achieve this goal.

6.12 Analyzing the Case Where the Methane Concentration Increases Suddenly

The methane concentration used in this study was collected throughout the year. Therefore, the production variation is in the data. It is not possible to separate the cause for CH_4 up or down variation except for the dependence on the air flow rate and pressure variations. These are measured, and the data already deals with the effects of airflow variations. Other possible causes for methane variation are intercepting CH_4 pockets, fissures, CH_4 flow channels, various coal-seam parts that contain and release varying amounts of CH_4, varying amounts of CH_4 leakage from the gob, and production rate. If any of these factors happened throughout the year, these effects are also included in the data.

7. CONCLUSIONS

Working in subsurface mines may become dangerous if the appropriate environmental conditions are not provided. One problem that may arise is the accumulation of contaminants, such as CO, SO_2 for heath, or methane for explosion hazards due to lousy ventilation inside the mine, which may also cause a temperature rise. The present work proposed an LSTM model to create an early warning system for forecasting methane concentration inside an underground coal mine. The model used a combination of variables related to the variation of methane inside the mine, such as airflow, pressure, and methane concentration at the main gate. The target is the methane concentration at the headgate. Several scenarios were tested to increase the performance of the proposed model, such as the influence of the LSTM tuning parameters, the best combination of variables for training the model, filtering the data, training window size, and the number of steps ahead predicted. The results were compared with an in-house developed time-series filter. Results show that the best configuration of the LSTM uses two hidden layers, three neurons per layer, and a learning rate of 0.00068. The results show that the LSTM model is limited to one-step-ahead prediction for reasonable accuracy. It also shows that when the number of time steps forecasted increased, the accuracy of the predictions decreased. This observation is in agreement with some authors from the literature.

Furthermore, increasing the number of variables or the training window size does not seem to increase the accuracy of the LSTM predictions. The main reason

behind this observation is that when the LSTM is trained with more data or the number of variables increases, it experiences more noise. When it tries to make the predictions, the model considers this noise and combines it with outdated information that decreases the performance of the LSTM.

Comparing the results with the time series filter, both present high error picks. However, the TSF is much faster for training and predicting. Filtering positively impacted the predictions, both using the LSTM and the TSF. It is important to notice that despite using a small filter length, the filter data contributes to increasing the accuracy of the predictor models. Comparing the results using artificial data (sine waves) and the measured data from the mine, it is observed that the LSTM performs better if the data has a specific pattern and is as smooth as possible. Possibly, this is the reason why filtering the measured data gives better predictions. Comparing the results from the LSTM with those from the most simple case (considering that the next day prediction is the previous day value), it is possible to verify that the predictions from the simple case are as good as the predictions from the LSTM. Furthermore, it is safe to assume that using the most simple case is more advangeous than using the LSTM or the time series filter for simplicity reasons.

The comparison with the results from the literature shows that using the LSTM in combination with other predictors greatly improves the model's accuracy. The use of regularization techniques to test the model against overfitting shows that it does

not improve the model's accuracy. Therefore, the conclusion is that overfitting did not happen or that the regularization is ineffective for the parameters used.

Due to many fluctuations in the predictions, the proposed model is not accurate enough to be used as the first stage in an early warning system to predict multiple steps ahead. Future works might be considered to investigate what can be done to improve the LSTM performance. Suggestions are:

- Include more parameters in the fine-tuning of the LSTM
- Make sure the data used does not have interference from the sensors
- Use the LSTM in combination with other predictors
- Use a faster computer if time is a concern

8. REFERENCES

Agioutantis, Z. G., Karmis, M., & Schafrik, S. J. (2014). Development of an atmospheric data management system for underground coal mines Automation of Shuttle Car for Underground Coal Mines View project. In *Article in Journal-South African Institute of Mining and Metallurgy*. Vol. 114, pages 1059-1063.

Anderson, Sweeney, Williams (2001). *Statistics for Business and Economics- Ed11*. SBN 13: 978-0-324-78325-4

Bascompta, M., Rossell, J. M., Sanmiquel, L., & Anticoi, H. (2020). Temperature prediction model in the main ventilation system of an underground mine. *Applied Sciences (Switzerland)*, *10*(20), 1–11. https://doi.org/10.3390/app10207238

Belle, B., & Biffi, M. (2018). Cooling pathways for deep Australian longwall coal mines of the future. *International Journal of Mining Science and Technology*, *28*(6), 865–875. https://doi.org/10.1016/j.ijmst.2018.02.001

Bengio, Y. (2012). Practical recommendations for gradient-based training of Deep Architectures. *Lecture Notes in Computer Science*, 437–478. https://doi.org/10.1007/978-3-642-35289-8_26

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning Long-Term Dependencies with Gradient Descent is Difficult. *IEEE Transactions on Neural Networks*, *5*(2), 157–166. https://doi.org/10.1109/72.279181

Blum, C., & Socha, K. (2005). Training feed-forward neural networks with ant colony optimization: An application to pattern classification. *Fifth International Conference on Hybrid Intelligent Systems (HIS'05)*. https://doi.org/10.1109/ichis.2005.104

Bougrain, L. (2004). Practical introduction to artificial neural networks. *IFAC Proceedings Volumes (IFAC-PapersOnline)*, *37*(15), 347–352. https://doi.org/10.1016/s1474-6670(17)31048-0

Brake, D. (2001) . Fluid consumption, sweat rates and hydration status of thermally-stressed underground miners and the implications for heat illness and shortened shifts. *Proc 2001 Qld Mining Ind Occ Health and Safety Conf. Townsville, 2001. Qld Mining Council, Brisbane (2001).*

Bratu, M. (2013). New accuracy measures for point and interval forecasts. A case study for Romania's forecasts of inflation and unemployment rate, Economic Analysis Working Papers (2002-2010). Atlantic Review of Economics (2011-2016), Colexio de Economistas de A Coruña, Spain and Fundación Una Galicia Moderna, vol. 1, pages 1-1, June.

Connor, J., Atlas, L. E., & Martin, D. R. (n.d.). *Recurrent networks and Narma Modeling - NeurIPS*. Retrieved December 7, 2021, from https://proceedings.neurips.cc/paper/1991/file/5ef0b4eba35ab2d6180b0bca7e46b6f9-Paper.pdf.

Creedy, D. P. (1993). Methane emissions from coal related sources in Britain: Development of a methodology. *Chemosphere*, *26*(1-4), 419–439. https://doi.org/10.1016/0045-6535(93)90435-8

Danko, G. (2006). Functional or operator representation of numerical heat and mass transport models. *Journal of Heat Transfer*, *128*(2), 162–175. https://doi.org/10.1115/1.2136919

Danko, G. L. (2013). Subsurface flow and transport process model for time dependent mine ventilation simulations. *Transactions of the Institutions of Mining and Metallurgy, Section A: Mining Technology*, *122*(3), 134–144. https://doi.org/10.1179/147490013X13639459465691

Danko, G. L., & Asante, W. (2017). Time-Dependent Contaminant Transport in Ventilating Air from a Moving Source. *Applied Mathematics*, *08*(05), 671–692. https://doi.org/10.4236/am.2017.85054

Danko, G. L., Asante, W. K., Bahrami, D., & Stewart, C. (2019). Dynamic Models in Atmospheric Monitoring Signal Evaluation for Safety, Health and Cost Benefits. *Mining, Metallurgy and Exploration*, *36*(6), 1235–1252. https://doi.org/10.1007/s42461-019-0099-x

Danko, G., Bahrami, D., & Stewart, C. (2020). Applications and verification of a computational energy dynamics model for mine climate simulations. *International Journal of Mining Science and Technology*, *30*(4), 483–493. https://doi.org/10.1016/j.ijmst.2020.03.019

Danko, G., Birkholzer, J., Bahrami, D., & Halecky, N. (2010). Temperature, humidity, and airflow in the emplacement drifts using convection and dispersion transport models. *Nuclear Technology*, *171*(1), 74–87. https://doi.org/10.13182/nt10-a10773

Diaconescu, E. (2008). *The use of NARX neural networks to predict chaotic time series. - WSEAS Transactions on computer research*. *ISSN:1991-8755. Issue3, volume3*. Pages 182-191

Diaz, J. C., Agioutantis, Z., Schafrik, S., & Hristopulos, D. T. (2021). Towards atmospheric monitoring data analysis in underground coal mines. In *Mine Ventilation* (pp. 498–506). CRC Press. https://doi.org/10.1201/9781003188476-51

Dixon, D. W. (1992). *A statistical analysis of monitored data for methane prediction* (thesis). University of Nottingham.

Donoghue, A. M. (2000). Heat exhaustion in a deep underground metalliferous mine. *Occupational and Environmental Medicine*, *57*(3), 165–174. https://doi.org/10.1136/oem.57.3.165

Donoghue, A. M. (2004). Heat Illness in the U.S. Mining Industry. *American Journal of Industrial Medicine*, *45*(4), 351–356. https://doi.org/10.1002/ajim.10345

Ediz, I. G., & Edwards, J. S. (1991). Numerical simulation of time-dependent methane flow. *Mining Science and Technology*, *12*(1), 1–15.

Eltschlager, K. K. H. (2001). *Technical measures for the investigation and mitigation of fugitive methane hazards in areas of coal mining.* NTIS. Retrieved December 7, 2021, from https://ntrl.ntis.gov/NTRL/dashboard/searchResults/titleDetail/PB2002106133.xhtml.

Geng, Y. (2016). Short term prediction of coal mine methane concentration with Chaos PSO-RBFNN model in underground coal mines. *Proceedings of the 2016 International Conference on Mechatronics Engineering and Information Technology*. https://doi.org/10.2991/icmeit-16.2016.7

George, D., Davood, B., & Pierre, M. J. (2011). Ventilation and climate simulation with the Multiflux code. *Journal of Coal Science and Engineering*, *17*(3), 243–250. https://doi.org/10.1007/s12404-011-0304-0

Greff, K., Srivastava, R. K., Koutnik, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM: A Search Space Odyssey. *IEEE Transactions on*

Neural Networks and Learning Systems, *28*(10), 2222–2232. https://doi.org/10.1109/TNNLS.2016.2582924

Gurney, K. (Kevin N.). (1997). *An introduction to neural networks*. UCL Press. ISBN 0-203-45622-X

Horowitz, P., & Hill, W. (2021). *The Art of Electronics*. Cambridge University Press. ISBN 978-0-521-80926-9

Hota, H. S., Handa, R., & Shrivas, A. K. (2017). Time Series Data Prediction Using Sliding Window Based RBF Neural Network. In *International Journal of Computational Intelligence Research* (Vol. 13, Issue 5). http://www.ripublication.com. ISBN:1-85728-503-4

Jo, B. W., & Khan, R. M. A. (2018). An internet of things system for underground mine air quality pollutant prediction based on azure machine learning. *Sensors (Switzerland)*, *18*(4). https://doi.org/10.3390/s18040930

Karacan, C. Ö. (2008). Modeling and prediction of ventilation methane emissions of U.S. longwall mines using supervised artificial neural networks. *International Journal of Coal Geology*, *73*(3–4), 371–387. https://doi.org/10.1016/j.coal.2007.09.003

Karacan, C. O., Diamond, W. P., Esterhuizen, G. S., & Schatzel, S. J. (n.d.). *Numerical Analysis of the Impact of Longwall Panel Width on Methane Emissions and Performance of Gob Gas Ventholes* . Retrieved

December 7, 2021, from https://www.cdc.gov/niosh/mining/UserFiles/works/pdfs/naotio.pdf.

Karacan, C. Ö., Ruiz, F. A., Cotè, M., & Phipps, S. (2011). Coal mine methane: A review of capture and utilization practices with benefits to mining safety and to greenhouse gas reduction. In *International Journal of Coal Geology* (Vol. 86, Issues 2–3, pp. 121–156). https://doi.org/10.1016/j.coal.2011.02.009

Kirchgessner, D. A., Piccot, S. D., & Masemore, S. S. (2000). An improved inventory of methane emissions from coal mining in the United States. *Journal of the Air & Waste Management Association*, *50*(11), 1904–1919. https://doi.org/10.1080/10473289.2000.10464227

Kordestani, H., Xiang, Y. Q., & Ye, X. W. (2018). Output-Only Damage Detection of Steel Beam Using Moving Average Filter. *Shock and Vibration*, *2018*. https://doi.org/10.1155/2018/2067680

Kriesel, D. (2014, December 18). *A brief introduction to neural networks* Retrieved December 7, 2021, from http://www.dkriesel.com/en/science/neural_networks.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional Neural Networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Kumari, K., Dey, P., Kumar, C., Pandit, D., Mishra, S. S., Kisku, V., Chaulya, S. K., Ray, S. K., & Prasad, G. M. (2021). UMAP and LSTM based fire status and explosibility prediction for sealed-off area in underground coal mine. *Process Safety and Environmental Protection*, *146*, 837–852. https://doi.org/10.1016/j.psep.2020.12.019

Kurnia, J. C., Sasmito, A. P., & Mujumdar, A. S. (2014). Simulation of a novel intermittent ventilation system for underground mines. *Tunnelling and Underground Space Technology*, *42*, 206–215. https://doi.org/10.1016/j.tust.2014.03.009

Li, T., Zhang, Z., & Chen, H. (2019). Predicting the combustion state of rotary kilns using a Convolutional Recurrent Neural Network. *Journal of Process Control*, *84*, 207–214. https://doi.org/10.1016/j.jprocont.2019.10.009

Liang, L., & Cai, X. (2020). Forecasting peer-to-peer platform default rate with LSTM neural network. *Electronic Commerce Research and Applications*, *43*. https://doi.org/10.1016/j.elerap.2020.100997

Lillicrap, T. P., & Santoro, A. (2019). Backpropagation through time and the brain. In *Current Opinion in Neurobiology* (Vol. 55, pp. 82–89). Elsevier Ltd. https://doi.org/10.1016/j.conb.2019.01.011

Lin, T., Horne, B. G., Tiiio, P., Lee Giles, C., & Member, S. (1996). Learning Long-Term Dependencies in NARX Recurrent Neural Networks. In *JEEE TRANSACTIONS ON NEIJRAL NETWORKS: Vol. I* (Issue 6).

Liu, H., Ma, L., Wang, Z., Liu, Y., & Alsaadi, F. E. (2020). An overview of stability analysis and state estimation for memristive neural networks. *Neurocomputing*, *391*, 1–12. https://doi.org/10.1016/j.neucom.2020.01.066

Liu, Y., Hou, D., Bao, J., & Qi, Y. (2018). Multi-step ahead time series forecasting for different data patterns based on LSTM recurrent neural network. *Proceedings - 2017 14th Web Information Systems and Applications Conference, WISA 2017*, *2018-January*, 305–310. https://doi.org/10.1109/WISA.2017.25

Liu, Z., Li, L., Fang, X., Qi, W., Shen, J., Zhou, H., & Zhang, Y. (2021). Hard-rock tunnel lithology prediction with TBM construction big data using a global-attention-mechanism-based LSTM network. *Automation in Construction*, *125*. https://doi.org/10.1016/j.autcon.2021.103647

Lu, C. (2019). *Analysis of The Effects of Thermal History of The Strata upon Temperature Variations in Underground Airways* (thesis). *University of Nevada, Reno.* Scholar Works, Reno.

Luxbacher, K. D., & Erdogan, S. S. (2009). *Modeling Methane Emissions and Ventilation Needs by Examination of Mining Induced Permeability*

Changes and Related Damage to Ventilation Controls. 43rd U.S. Rock Mechanics Symposium/4th U.S.-Canada Rock Mechanics Symposium, June 28 - July 1, 2009, Asheville, North Carolina, paper no. ARMA 09-146. Alexandria, VA: American Rock Mechanics Association, 2009 Jun; :1-8

Lyu, P., Chen, N., Mao, S., & Li, M. (2020). LSTM based encoder-decoder for short-term predictions of gas concentration using multi-sensor fusion. *Process Safety and Environmental Protection*, *137*, 93–105. https://doi.org/10.1016/j.psep.2020.02.021

Lyu, W., Cai, S., Yang, P., & Zhang, Y. (2017). Underground environment parameter prediction in a deep mine. *Proceedings of the Eighth International Conference on Deep and High Stress Mining*, 949–960. https://doi.org/10.36487/acg_rep/1704_65_lyu

Marlin, T. E. (2015). *Process Control Designing Processes and Control Systems for Dynamic Performance 2 nd Edition*. ISBN 978-0070393622

Mas, J. F., & Flores, J. J. (2008). The application of artificial neural networks to the analysis of remotely sensed data. In *International Journal of Remote Sensing* (Vol. 29, Issue 3, pp. 617–663). Taylor and Francis Ltd.

Mathatho, S., Owolawi, P. A., & Tu, C. (2019). Prediction of methane levels in underground coal mines using Artificial Neural Networks. *2019 International Conference on Advances in Big Data, Computing and Data Communication Systems (IcABCD)*. https://doi.org/10.1109/icabcd.2019.8851041

Maurya, T., Karena, K., Vardhan, H., Aruna, M., & Raj, M. G. (2015). Potential Sources of Heat in Underground Mines – A Review. *Procedia Earth and Planetary Science*, *11*, 463–468. https://doi.org/10.1016/j.proeps.2015.06.046

Mcpherson, M. J., Eng, C., Aime, M., & Ashrae, M. (2009). *Subsurface Ventilation Engineering*. https://wipp.energy.gov/Library/Information_Repository_A/Supplemental_Information/2019/References/McPherson.%202009.pdf

Miao, K. chao, Han, T. ting, Yao, Y. qing, Lu, H., Chen, P., Wang, B., & Zhang, J. (2020). Application of LSTM for short term fog forecasting based on meteorological elements. *Neurocomputing*. https://doi.org/10.1016/j.neucom.2019.12.129

Mine Safety, D., & Administration, H. (2012). *Heat Stress in Mining Safety Manual Series SM 6*. U.S Department of Labor. National Mine Health and Safety Academy.

https://www.msha.gov/sites/default/files/Alerts%20and%20Hazards/Heat%20Stress.pdf

Murphy, K. P. (2021). *Machine learning: A probabilistic perspective*. MIT Press. ISBN 978-0-262-01802-9

Nagarajan, G., & Dhinesh Babu, L. D. (2021). Missing data imputation on biomedical data using deeply learned clustering and L2 regularized regression based on symmetric uncertainty. *Artificial Intelligence in Medicine*, 102214. https://doi.org/10.1016/j.artmed.2021.102214

Nguyen, D. H., & Widrow, B. (1990). Neural networks for self-learning control systems. *IEEE Control Systems Magazine*, *10*(3), 18–23. https://doi.org/10.1109/37.55119

Nie, X., Wei, X., Li, X., & Lu, C. (2018). Heat Treatment and Ventilation Optimization in a Deep Mine. *Advances in Civil Engineering*, *2018*. https://doi.org/10.1155/2018/1529490

NIOSH. (2006). *Handbook for Methane Control in Mining Centers for Disease Control and Prevention National Institute for Occupational Safety and Health TM*. www.cdc.gov/niosh

Osunmakinde, I. O. (2013). Towards safety from toxic gases in underground mines using wireless sensor networks and ambient intelligence. *International Journal of Distributed Sensor Networks*, *2013*. https://doi.org/10.1155/2013/159273

Parra, M. T., Villafruela, J. M., Castro, F., & Méndez, C. (2006). Numerical and experimental analysis of different ventilation systems in deep mines. *Building and Environment*, *41*(2), 87–93. https://doi.org/10.1016/j.buildenv.2005.01.002

Pascanu, R., Mikolov, T., & Bengio, Y. (2013). *On the difficulty of training recurrent neural networks*. *Proceedings of the 30thInternational Conference on Ma-chine Learning, Atlanta, Georgia, USA*, 2013.JMLR:W&CP volume 28. Copyright 2013 by the author(s)

Rojas ; R (1996). *Neural Networks*, *A Systematic Introduction*. Springer-Verlag.

Sangiorgio, M., & Dercole, F. (2020). Robustness of LSTM neural networks for multi-step forecasting of chaotic time series. *Chaos, Solitons and Fractals*, *139*. https://doi.org/10.1016/j.chaos.2020.110045

Sarkar, M., & de Bruyn, A. (2021). LSTM Response Models for Direct Marketing Analytics: Replacing Feature Engineering with Deep Learning. *Journal of Interactive Marketing*, *53*, 80–95. https://doi.org/10.1016/j.intmar.2020.07.002

Sasmito, A. P., Birgersson, E., Ly, H. C., & Mujumdar, A. S. (2013). Some approaches to improve ventilation system in underground coal mines environment - A computational fluid dynamic study. *Tunnelling and*

Underground Space Technology, *34*, 82–95. https://doi.org/10.1016/j.tust.2012.09.006

Shi, L., Wang, J., Zhang, G., Cheng, X., & Zhao, X. (2017). A risk assessment method to quantitatively investigate the methane explosion in underground coal mine. *Process Safety and Environmental Protection*, *107*, 317–333.

Siblini, W., Fréry, J., He-Guelton, L., Oblé, F., & Wang, Y.-Q. (2020). Master your metrics with calibration. *Lecture Notes in Computer Science*, 457–469. https://doi.org/10.1007/978-3-030-44584-3_36

Smith, S. W. (1999). *The scientist and engineer's Guide to Digital Signal Processing (2nd ed)*. California Technical Pub. https://www.analog.com/en/education/education-library/scientist_engineers_guide.html

Srivastava, N., Hinton, G., Krizhevsky, A., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. In *Journal of Machine Learning Research* (Vol. 15). Pages 1929-1958

Su, H.-T., Mcavoy, T. J., & Werbos, P. (1992). Long-Term Predictions of Chemical Processes Using Recurrent Neural Networks: A Parallel Training Approach. In *Ind. Eng. Chem. Res* (Vol. 31). https://pubs.acs.org/sharingguidelines

Svozil, D., Kvasnieka, V., & Pospichal, J. (1997). Chemometrics and intelligent laboratory systems Introduction to multi-layer feed-forward neural networks. In *Chemometrics and Intelligent Laboratory Systems* (Vol. 39). Pages 43-62

Tang, M., & Ding, Y. (2011). The reliability of ergonomics in the ventilation system of an underground metal mine. *Procedia Engineering*, *26*, 1705–1711. https://doi.org/10.1016/j.proeng.2011.11.2357

Tauziède, C., Pokryszka, Z., & TAUZffiDE, C. (1993). *Dynamic prediction of CH4 emission in longwalls* (Vol. 25). *Conférence Internationale des Instituts de Recherches sur la Sécurité dans les Mines*, Sep 1993, Pretoria, South Africa. pp.41-50. ⟨ineris-00971873⟩ https://hal-ineris.archives-ouvertes.fr/ineris-00971873

Ulery, J. P. (2008). Explosion hazards from methane emissions related to geologic features in coal mines. *DEPARTMENT OF HEALTH AND HUMAN SERVICES*. https://doi.org/10.26616/nioshpub2008123

Wang, H., Li, J., Yu, Q., Hong, T., Yan, J., & Zhao, W. (2020). Integrating recurrent neural networks and reinforcement learning for dynamic service composition. *Future Generation Computer Systems*, *107*, 551–563. https://doi.org/10.1016/j.future.2020.02.030

Yang, Y., Du, Q., Wang, C., & Bai, Y. (2020). Research on the method of methane emission prediction using improved grey radial basis function

neural network model. *Energies*, *13*(22). https://doi.org/10.3390/en13226112

Yi, X., Ren, L., Ma, L., Wei, G., Yu, W., Deng, J., & Shu, C. (2019). Effects of seasonal air temperature variation on airflow and surrounding rock temperature of mines. *International Journal of Coal Science and Technology*, *6*(3), 388–398. https://doi.org/10.1007/s40789-019-00268-1

Zhu, S., Wu, S., Cheng, J., Li, S., & Li, M. (2015). An underground air-route temperature prediction model for ultra-deep coal mines. *Minerals*, *5*(3), 527–545. https://doi.org/10.3390/min5030508