engineering proceedings

MDPI

Proceeding Paper

Improving Effectiveness of Energy Baseline Using Deep Learning †

Chun-Wei Chen [1,*], Chen-Yu Lin [1], Jung-Hsing Wang [1] and Hao-Kai Tu [2]

[1] Taiwan Instrument Research Institute, National Applied Research Laboratories, Hsinchu 300092, Taiwan; chenyulin@narlabs.org.tw (C.-Y.L.); kevin@narlabs.org.tw (J.-H.W.)
[2] Department of IE&EM, NTHU, Hsinchu 300092, Taiwan; iamtonydu@gmail.com
* Correspondence: alexchen52088@yahoo.com.tw
† Presented at the 2025 IEEE 5th International Conference on Electronic Communications, Internet of Things and Big Data, New Taipei, Taiwan, 25–27 April 2025.

Abstract

Energy conservation and carbon reduction are critical in energy policies. Therefore, numerous energy-saving methods, such as the introduction of new technologies and the replacement of outdated equipment, have been proposed. To determine whether these methods are effective in energy conservation and carbon reduction, scientific validation is required. The most common validation method is energy baseline. An energy baseline refers to the use of data measured before energy-saving improvements. It is used to construct a mathematical model that describes energy consumption. Using the baseline, the energy consumption during the baseline period after improvements is calculated. By subtracting the measured consumption from the value, the amount of energy saved is estimated. Traditionally, linear regression is used to establish energy baseline prediction. However, linear regression has limitations with complex energy data. Therefore, we used deep learning models to handle nonlinear data in the air compression system for comparative analysis. The developed long-short-term memory (LSTM) model showed superior capabilities for processing nonlinear data, aligning with the actual data distribution, and reducing errors. Compared with linear regression models, the LSTM model reduced uncertainty, risk, and cost by 40.3%.

Keywords: energy conservation and carbon reduction; energy baseline; linear regression model; deep learning; LSTM

check for updates

Academic Editors: Teen-Hang Meen, Shu-Han Liao and Cheng-Fu Yang

Published: 1 September 2025

Citation: Chen, C.-W.; Lin, C.-Y.; Wang, J.-H.; Tu, H.-K. Improving Effectiveness of Energy Baseline Using Deep Learning. *Eng. Proc.* **2025**, *108*, 12. https://doi.org/10.3390/engproc2025108012

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Energy conservation and carbon reduction are crucial in energy policies. Various energy-saving methods, such as those involving the introduction of new technologies and the replacement of outdated equipment, have been proposed. However, a scientific method is required to verify the effectiveness of these methods in achieving energy conservation and carbon reduction with the establishment of energy baselines.

An energy baseline in a mathematical model is used to describe energy consumption using data measured before energy-saving improvements. Energy baselines are also used to estimate baseline energy consumption after improvements. The differences between the estimated baseline consumption and the measured consumption represent the energy saved using the improvement methods [1]. Currently, the most commonly used mathematical model for establishing energy baselines is the linear regression model. This model is simple and easy to use, and it facilitates rapid computation. However, it often fails to

accurately fit when processing complex data. Therefore, a mathematical model with higher energy baseline fit performance must be developed to enable precise comparisons of energy consumption before and after the implementation of energy-saving measures, facilitate accurate assessments of the effectiveness of energy-saving methods, and provide a more convincing basis for energy-saving decisions. In this study, we focused on the air compressor system at a specific site.

The rest of this paper is arranged as follows. Section 2 introduces the linear regression model and the operational principles of the long-short-term memory (LSTM) network. Section 3 describes the examined air compressor system, and Section 4 details the simulations and comparative analyses conducted in this study. Finally, Section 5 provides the conclusions of this study. According to the simulation results of this study, deep learning is a more suitable method for establishing energy baselines.

2. Energy Baseline Model

Linear regression and deep learning methods were used to establish energy baselines in this study. An energy baseline model and optimal model parameters were constructed under the assumption that a dataset [i.e., (X_n, Y_n), where $X \in R^m$ and $Y \in R$] contains n pieces of data and m features.

2.1. Linear Regression

The linear regression equation for the aforementioned dataset is expressed as follows [2]:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{1}$$

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta m \end{bmatrix}, \boldsymbol{X} = \begin{bmatrix} 1 & X_{11} & \dots & X_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \dots & X_{nm} \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{2}$$

where $\boldsymbol{\beta}$ represents the parameters of the linear regression model, and $\boldsymbol{\varepsilon}$ is the error term [$\varepsilon_n \sim N(0, \sigma^2)$].

The loss function of the model is defined using the least squares method as follows:

$$Loss(\boldsymbol{\beta}) = (\boldsymbol{Y} - \boldsymbol{Y})^{\boldsymbol{T}}(\boldsymbol{Y} - \boldsymbol{Y}) \tag{3}$$

This loss function is rearranged and expressed as $\boldsymbol{Y}^T\boldsymbol{Y} + \boldsymbol{X}^T\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}\boldsymbol{X} - 2\boldsymbol{X}^T\hat{\boldsymbol{\beta}}\boldsymbol{Y}$. To find the optimal parameter $\boldsymbol{\beta}$, the partial derivative of the loss function is taken to determine its extreme value.

$$\frac{\partial\, Loss(\boldsymbol{\beta})}{\partial\, \boldsymbol{\beta}} = 0 \Rightarrow \hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y} \tag{4}$$

2.2. LSTM Network

An LSTM network is an extension of a recurrent neural network (RNN) that addresses the limitations of RNNs in retaining memory and obtaining the optimal parameters [3,4]. An LSTM network consists of four main components: a cell state, a forget gate, an input gate, and an output gate. The structure of this network is shown in Figure 1.

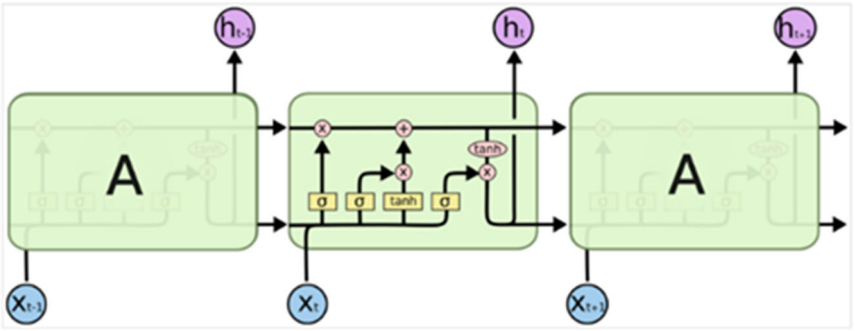

Figure 1. Structure of LSTM model.

In Figure 1, each rectangle with rounded edges represents an LSTM cell. The horizontal line running through all cells is the LSTM state, which is the main location where memory is stored. The memory state at time step t is denoted as C_t, which is updated from C_{t-1} by discarding and adding information within the cell. The first σ rectangle inside the cell is the forget gate, which is denoted as f_t and determines the extent to which the information in the cell state C_{t-1} should be discarded. The term f_t is computed as follows:

$$f_t = \sigma\left(W_f \times \boldsymbol{X}_t + U_f \times h_{t-1}\right) \tag{5}$$

where W represents the weights of the input data, h_{t-1} is the output at time $t-1$ (corresponding to the arrow below the cell in Figure 1), and U is the weight of the previous time step's output at time $t-1$. Moreover, σ is a sigmoid function, which has values between 0 and 1. The second σ rectangle inside the cell represents the input gate, which is denoted as i_t and determines the extent to which new input data $\widetilde{C}_t$ must be added. The values of the input gate also range from 0 to 1.

$$i_t = \sigma(W_i \times \boldsymbol{X}_t + U_i \times h_{t-1}) \tag{6}$$

The third rectangle inside the cell represents the input content, which is denoted as $\widetilde{C}_t$ and expressed as follows:

$$\widetilde{C}_t = tanh(W_c \times \boldsymbol{X}_t + U_c \times h_{t-1}\Big) \tag{7}$$

The activation function of $\widetilde{C}_t$ is the hyperbolic tangent function (tanh).

As displayed in Figure 1, the memory state is updated through the memory gate and input gate. This update is expressed as follows:

$$C_t = f_t \times C_{t-1} + i_t \times \widetilde{C}_t \tag{8}$$

This equation represents the sum of the previous memory state multiplied by the forget gate and the new input content multiplied by the input gate. It describes the process of discarding a part of the old information while incorporating new information.

After the memory state is updated, C_t is output through two pathways. The first pathway passes along the straight line that runs through all cells, with C_t passing to the next cell at time step $t+1$. The second pathway generates the output of the estimated value $\hat{Y}_t$ at time step t, which is controlled by the output gate (denoted as o_t).

$$o_t = \sigma(W_o \times \boldsymbol{X}_t + U_o \times h_{t-1}) \tag{9}$$

The output information is represented as h_t, which is expressed as follows:

$$h_t = o_t \times tanh(C_t) \tag{10}$$

After h_t is calculated, it must be passed to the output layer and the next time step $t + 1$. The output layer produces the desired label, and $\hat{Y}_t$ is used to denote the estimated value at time step t.

$$\hat{Y}_t = \sigma(W_h \times h_t) \tag{11}$$

Because the execution of an LSTM network involves time step computation, parameter optimization is conducted using the backpropagation through time (BPTT) algorithm, which combines traditional neural network backpropagation with the chain rule from calculus.

The loss function is defined using entropy. The loss function at time step t is as follows:

$$E_t(Y_t, \hat{Y}_t) = -Y_t log(\hat{Y}_t) \tag{12}$$

where Y_t is the actual value and $\hat{Y}_t$ is the predicted value. Although equations contain numerous parameters, an LSTM network benefits from weight sharing, meaning that the parameters f_t, i_t, $\widetilde{C}_t$, and o_t essentially have the same value. Therefore, only three main parameters (i.e., W_h, W_o, and U_o) must be calculated. BPTT derivation is a complex and lengthy process; a more detailed explanation of this method can be found in [5].

3. Air Compressor System

An air compressor system was examined in this study. A schematic of this system is shown in Figure 2.

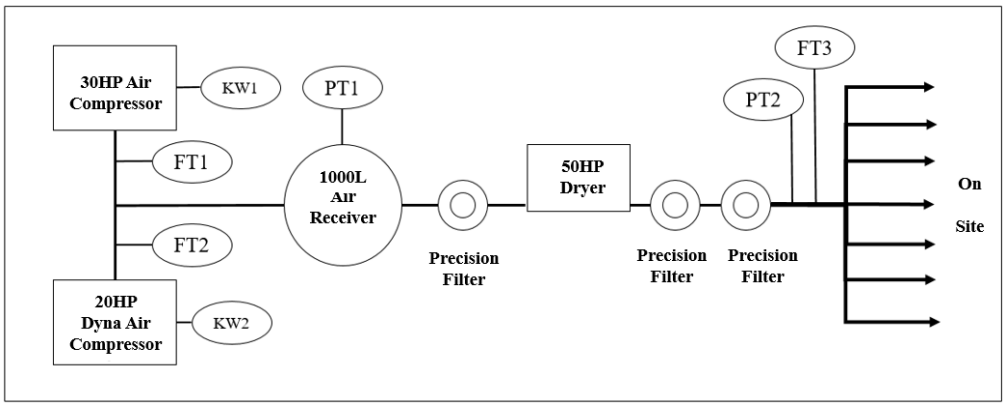

Figure 2. Schematic of air compressor system.

KW1 and KW2 are the power meters used to record the power consumption data of the air compressor system. These data include three-phase current, three-phase line voltage, power consumption, power factor, and total system energy consumption. Moreover, PT1 and PT2 are pressure gauges that measure pressure-related variables, including air tank pressure, system pressure, and differential pressure. Finally, FT1, FT2, and FT3 are flow meters that record gas-flow-related variables, such as fluid temperature, flow pressure, flow velocity, and flow rate.

To collect relevant energy consumption data from the air compressor system's flow meters, pressure gauges, and power meters, the Pearson correlation coefficient was calculated for each variable with power consumption (kW) and energy consumption (kW/CMM). The variables are ranked by the descending order of their correlations in Tables 1 and 2.

Table 1. Power-consumption-related correlation coefficients.

Correlation coefficient	Flow rate (CMM)	Differential pressure (mBar)	Air tank pressure(Bar)	Output frequency (Hz)	System pressure (Bar)
	0.8594	0.8454	0.4049	0.3972	0.2365

Table 2. Energy-consumption-related correlation coefficients.

Correlation coefficient	Flow rate (CMM)	Differential pressure (mBar)	Air tank pressure(Bar)	Output frequency (Hz)	System pressure (Bar)
	0.4764	0.4763	0.3854	0.2982	0.2273

After the correlation coefficients were calculated and variables directly related to power consumption (i.e., current and voltage) were excluded, the results indicated that flow rate and differential pressure were highly correlated with power consumption, air tank pressure and output frequency were moderately correlated with power consumption, and system pressure had a low correlation with power consumption. The other variables were not correlated with power consumption.

Flow rate (CMM) represents the rate of gas flow in an air compressor system. The differential pressure (mbar) refers to the pressure difference across filters, as recorded by pressure gauge PT2. Air tank pressure (bar) is measured by pressure gauge PT1, and output frequency (Hz) represents the variable frequency data of the air compressor system. Finally, system pressure (bar) is the pressure value recorded by pressure gauge PT2.

4. Simulations and Results

Energy baselines established using deep learning methods enabled a model to better fit the data distribution characteristics compared with those derived from linear regression models. Power consumption (kW) and energy consumption (kW/CMM) were used as the prediction targets. The variables were highly correlated with the prediction targets selected as explanatory variables to construct the energy baseline model.

Data from October to December 2024 for an air compressor system at a specific site were selected as training data. Moreover, data from January 2025 for this system were selected as the test data. Model performance was evaluated using mean squared error (MSE) as follows:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2 \tag{13}$$

The descriptive statistics of the examined air compressor system were obtained for flow rate, differential pressure, air tank pressure, and output frequency, denoted as $\left\{\left(\boldsymbol{X}_n^{train1}, Y_n^{train1}\right)\middle| \mathrm{X}_{\in}\mathbb{R}^4,\ \mathrm{Y}_{\in}\mathbb{R},\ n = 1,477,470\right\}$, to predict power consumption. Figure 3 illustrates the power consumption scatterplot, in which the x-axis represents days, and the y-axis represents daily average power consumption values. Because the dataset was large, the scatterplot was created using daily average values to ensure readability. Table 3 presents the descriptive statistics for power consumption.

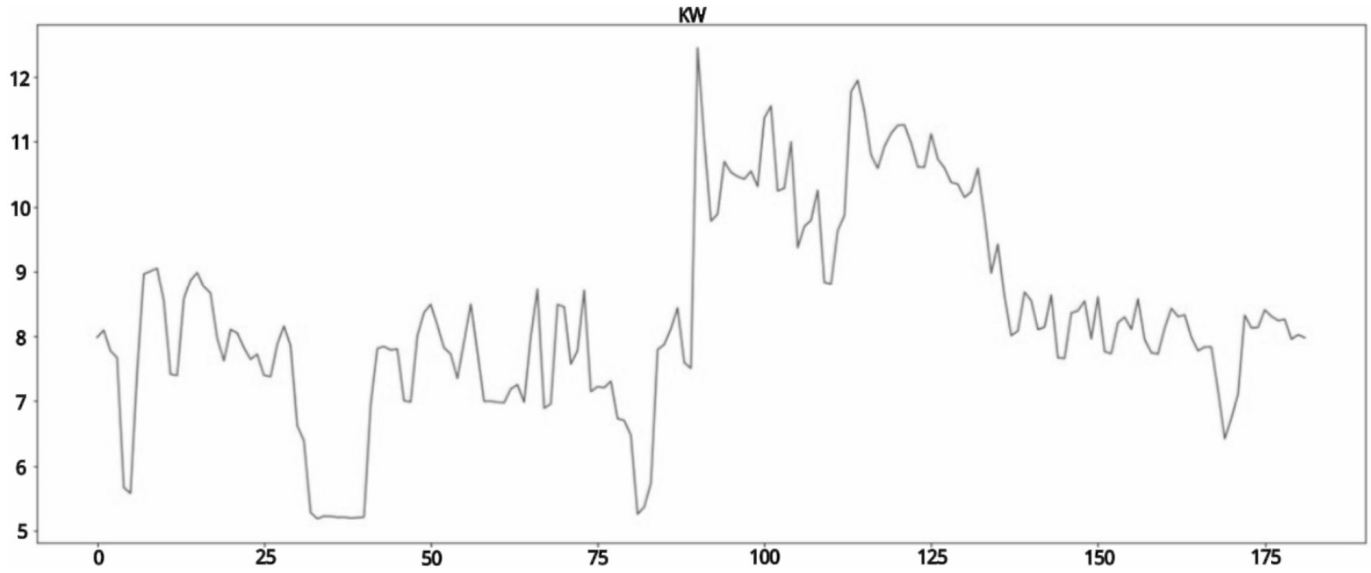

Figure 3. Scatterplot of daily average power consumption.

Table 3. Descriptive statistics for power consumption.

Statistical Measure	Mean	Standard Deviation	Maximum Value	Minimum Value
Value	8.4029	4.5918	28.3	1.4

Flow rate, differential pressure, and air tank pressure, denoted as $\left\{\left(X_n^{train2},\ Y_n^{train2}\right)\middle|X_{\in}\mathbb{R}^3,\ Y_{\in}\mathbb{R}, n = 1,477,470\right\}$, were selected to predict energy consumption. Figure 4 illustrates the energy consumption scatterplot, and Table 4 presents the descriptive statistics for energy consumption.

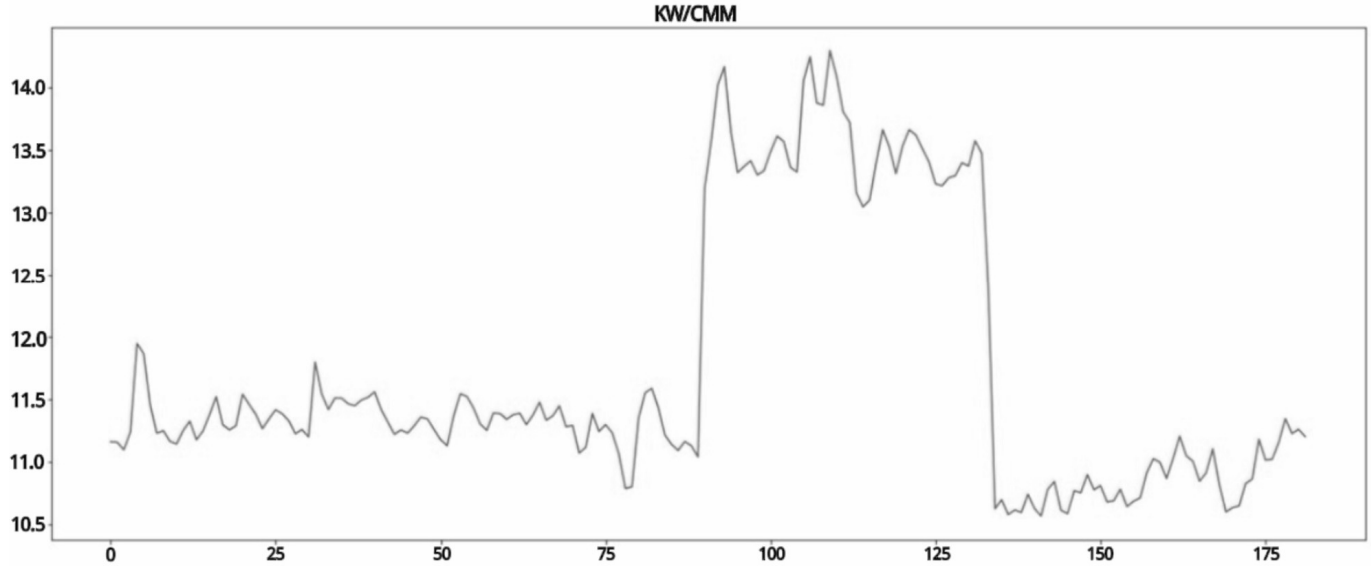

Figure 4. Scatterplot of daily average energy consumption.

Table 4. Descriptive statistics for energy consumption.

Statistical Measure	Mean	Standard Deviation	Maximum Value	Minimum Value
Value	11.748	3.9118	53	1.56

Next, $\{(X_n^{test1}, Y_n^{test1}) | X_{\in}\mathbb{R}^4, Y_{\in}\mathbb{R}, n = 163,869\}$ was used to represent the test data for power consumption, whereas $\{(X_n^{test2}, Y_n^{test2}) | X_{\in}\mathbb{R}^3$ was used to represent the test data for energy consumption. Figures 5 and 6 show the scatterplots for the test data on daily average power consumption and daily average energy consumption, respectively.

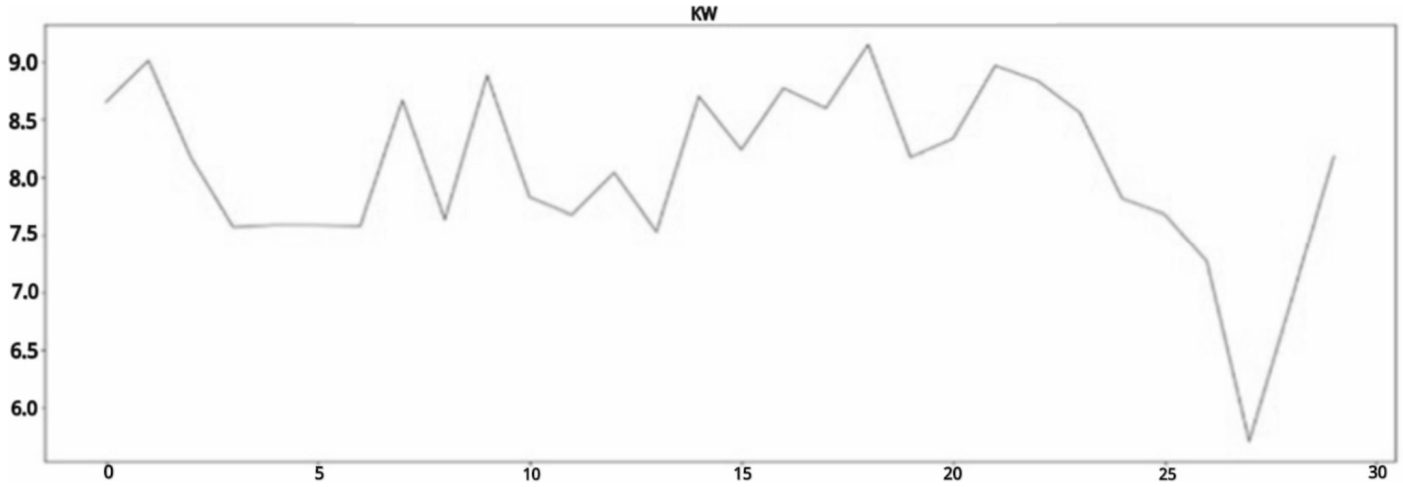

Figure 5. Scatterplot of the test data for daily average power consumption.

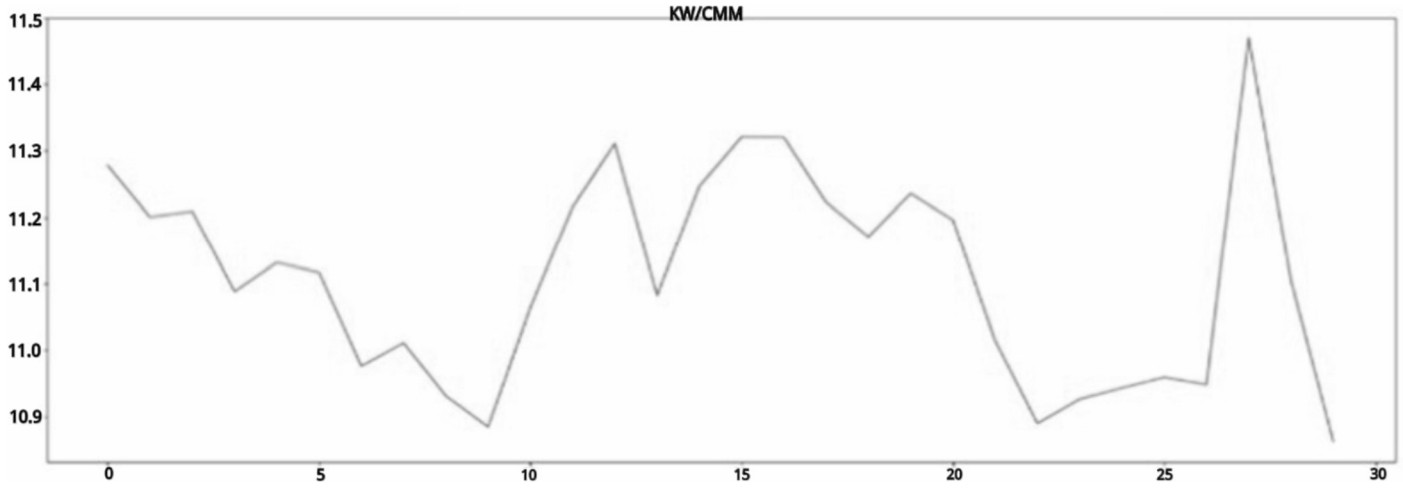

Figure 6. Scatterplot of the test data for daily average energy consumption.

Linear regression and LSTM models were used to establish baselines for the power consumption of the examined air compression system. The simulation results are described as follows.

Table 5 presents the numerical simulation results obtained with the linear regression and LSTM models for power consumption. The performance of the linear regression model was inferior to that of the LSTM model, making the linear regression model unsuitable for evaluating the effectiveness of energy-saving methods. Although the LSTM model produced a smaller error than did the linear regression model, a notable drawback of deep learning models is that they must conduct numerous computations, leading to a long simulation time. Therefore, we also calculated the computation times of both models in the power consumption simulations (Table 6).

Table 5. Numerical simulation results obtained for power consumption by using the linear regression and LSTM models.

Model	Linear Regression	LSTM
MSE	4.8533	2.8047

Table 6. Computation times of the linear regression and LSTM models in power consumption simulations.

Model	Linear Regression	LSTM
Training time	0.25(s)	4383(s)
Test time	0.007(s)	3.399(s)

The training and testing times for both models were calculated. The linear regression model was considerably faster than the LSTM model in the training process. However, both models exhibited short testing times. Therefore, the LSTM model exhibited a more favorable overall performance than the linear regression model in the establishment of energy baselines for evaluating the effectiveness of energy-saving methods.

The linear regression and LSTM models were also used to establish baselines for the energy consumed by the air compression system. Table 7 presents the numerical simulation results obtained with these models for energy consumption. The linear regression model exhibited inferior performance, likely because of the weak linear relationships of the explanatory variables with energy consumption, as indicated by the low correlation coefficients. The LSTM model considerably outperformed the linear regression model because deep learning models are well-suited for processing nonlinear data. Table 8 presents the calculation times of both models in the energy consumption simulations. The results in Table 8 are consistent with those in Table 6. The linear regression model again exhibited a considerably shorter training time than the LSTM model. However, both models exhibited short testing times. Therefore, the LSTM model is more suitable to the linear regression model for establishing energy baselines to evaluate the effectiveness of energy-saving methods.

Table 7. Numerical simulation results obtained for energy consumption by using the linear regression and LSTM models.

Model	Linear Regression	LSTM
MSE	10.6136	6.3364

Table 8. Computation times of the linear regression and LSTM models in energy consumption simulations.

Model	Linear Regression	LSTM
Training time	0.25(s)	2315(s)
Test time	0.007(s)	2.149(s)

5. Conclusions

We conducted empirical analyses to compare the performance of linear regression and LSTM models in establishing energy baselines. The results indicated that the LSTM model provided more accurate energy baselines than the linear regression model, which is commonly used for energy-saving evaluations. The performance of the linear regression model was hindered when it processed nonlinear data, resulting in poor model fitting.

Consequently, decision-makers using this model might struggle to make accurate energy-related decisions, and they fail to accurately determine the effectiveness of energy-saving methods. Conversely, the LSTM model exhibited a strong capability to process nonlinear data and generated models that accurately aligned with the actual data distribution. Consequently, the LSTM model showed lower error rates in energy consumption prediction, being more appropriate for establishing energy baselines than the linear regression model. The LSTM model reduced uncertainty, risk, and cost by 40.3% compared with traditional regression models.

Author Contributions: Conceptualization: C.-W.C.; Methodology: C.-W.C.; Software: C.-W.C.; Data curation: C.-Y.L., J.-H.W. and H.-K.T.; Writing—original draft preparation: C.-W.C.; Writing—review and editing: C.-Y.L., J.-H.W. and H.-K.T.; Visualization: C.-W.C.; Project administration: C.-W.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Science and Technology Council of Taiwan, grant number NSTC 113-2224-E-492-001. The APC was funded by National Science and Technology Council of Taiwan.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Due to commercial confidentiality, all data from this study will not be publicly available.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Efficiency Valuation Organization. *International Performance Measurement and Verification Protocol*; Volume I: 2007, Volume II: 2002, Volume III: 2006. Available online: http://www.evo-world.org (accessed on 13 May 2025).
2. Montgomery, D.C.; Peck, E.A.; Vining, G.G. *Introduction to Linear Regression Analysis*, 5th ed.; Wiley: Hoboken, NJ, USA, 2012; Volume 821.
3. Chniti, G.; Bakir, H.; Zaher, H. E-commerce time series forecasting using LSTM neural network and support vector regression. In Proceedings of the International Conference on Big Data and Internet of Thing (BDIOT 2017), London, UK, 20–22 December 2017; pp. 80–84. [CrossRef]
4. Olah, C. Understanding LSTM Networks. 2015. Available online: http://colah.github.io/posts/2015-08-Understanding-LSTMs/ (accessed on 13 May 2025).
5. Guo, J. *Backpropagation Through Time*, Harbin Institute of Technology: Harbin, China, 2013; (*unpublished manuscript*).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.