

Article

An Intelligent Time-Series Model for Forecasting Bus Passengers Based on Smartcard Data

Ching-Hsue Cheng ¹, Ming-Chi Tsai ^{2,*} and Yi-Chen Cheng ¹

¹ Department of Information Management, National Yunlin University of Science & Technology, Touliou, Yunlin 640, Taiwan; chcheng@yuntech.edu.tw (C.-H.C.); hsue6771@gmail.com (Y.-C.C.)

² Department of Business Administration, I-Shou University, Kaohsiung City 84001, Taiwan

* Correspondence: mct@isu.edu.tw; Tel.: +886-7-6577711 (ext. 5923)

Abstract: Public transportation systems are an effective way to reduce traffic congestion, air pollution, and energy consumption. Today, smartcard technology is used to shorten the time spent boarding/exiting buses and other types of public transportation; however, this does not alleviate all traffic congestion problems. Accurate forecasting of passenger flow can prevent serious bus congestion and improve the service quality of the transportation system. To the best of the current authors' knowledge, fewer studies have used smartcard data to forecast bus passenger flow than on other types of public transportation, and few studies have used time-series lag periods as forecast variables. Therefore, this study used smartcard data from the bus system to identify important variables that affect passenger flow. These data were combined with other influential variables to establish an integrated-weight time-series forecast model. For different time data, we applied four intelligent forecast methods and different lag periods to analyze the forecasting ability of different daily data series. To enhance the forecast ability, we used the forecast data from the top three of the 80 combined forecast models and adapted their weights to improve the forecast results. After experiments and comparisons, the results show that the proposed model can improve passenger flow forecasting based on three bus routes with three different series of time data in terms of root-mean-square error (RMSE) and mean absolute percentage error (MAPE). In addition, the lag period was found to significantly affect the forecast results, and our results show that the proposed model is more effective than other individual intelligent forecast models.

Keywords: passenger flow; integrated-weight time-series model; public transportation systems; long short-term memory network



Citation: Cheng, C.-H.; Tsai, M.-C.; Cheng, Y.-C. An Intelligent Time-Series Model for Forecasting Bus Passengers Based on Smartcard Data. *Appl. Sci.* **2022**, *12*, 4763. <https://doi.org/10.3390/app12094763>

Academic Editor: Feng Guo

Received: 12 March 2022

Accepted: 4 May 2022

Published: 9 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

Public transportation is considered to be an effective solution to traffic congestion and environmental pollution. The Federal Transit Administration (FTA) also believes that public transportation is an effective way to reduce traffic congestion, air pollution, energy consumption, and private vehicle use [1]. The use rate of buses accounted for 46% of all public transportation use in 2016 by people aged over 15 years according to the Taiwan Ministry of Transportation survey [2].

Taiwan's EasyCard Company promoted the smartcard system in 2002 based on the idea of "one card in hand, unimpeded travel". It was the first card to be issued for Taipei mass rapid transit and was then expanded to the Taiwan railway, Taiwan high-speed railway, and various other types of public transportation. Smartcards can collect information about vehicle routes, schedules, and real-time driving conditions through the automatic fare collection (AFC) system for vehicle monitoring, which can greatly improve public transportation efficiency and safety. The AFC system, when referring to the transportation system [3], is also called the smartcard system. The smartcard system is regarded as a dynamic and real-time data source for the public transportation system. It

has attracted a significant amount of attention from researchers, and many studies have used smartcard data [3,4].

Although the smartcard system helps passengers greatly reduce their transaction time and shortens the time taken to board/exit buses, it also helps the bus industry collect large amounts of data to improve their service quality. Despite the utilization of the bus system, serious traffic congestion still occurs. Traffic flow describes the number of vehicles passing through a specific road section within a predetermined time interval [5]. It is different from traffic congestion, which is caused by excessive travel demand by people, resulting in abnormal traffic flow. There have been many studies on passenger flow predictions. In addition, smartcard systems can increase the convenience of users and help bus operators formulate practical route planning and reform timetables and related policies; however, this all depends on the use of smartcard system data to accurately forecast bus passenger flow. Accurately forecasting passenger flow can help cities implement transportation policies, strengthen local construction, reduce excessive energy consumption and carbon emissions, and improve urban ecosystems to achieve sustainable development.

The accessibility of the urban bus system is greater than for other modes of public transport, as this system utilizes the road network; however, passenger demands are affected by a number of factors such as crowding and different weather conditions. Tang et al. [6] confirmed prediction models would be better if the weather conditions were considered. The number of bus rides varies depending on the time of day, but there are still expected peak periods. For example, there will be many passengers during peak hours on weekdays and working days and at times when leisure activities are taking place during the holidays. We must consider passengers' needs, but external factors are also important.

In the past ten years, many successful traffic flow forecast methods have been proposed, especially deep learning methods. Li et al. [7] proposed a dynamic radial basis function neural network to predict short-term passenger flow through the Beijing subway. Ke et al. [8] proposed a fusion convolution long-term short-term memory network to forecast short-term passenger demand for ride services. Xu et al. [9] used a combined seasonal autoregressive integrated moving average with a support vector regression model to forecast the demand for the aviation industry. Deep learning methods have led to great progress in transportation research, but there have been few studies on forecasting buses' passenger flow compared with other types of public transportation. In our study, we collected data from the smartcard system for the bus industry and considered other external factors that affected the ride. In this article, we propose an integrated-weight time-series model to forecast passenger flow and detail our comparison with the listing methods. In summary, the goals of this study are as follows:

- (1) To identify the important attributes that affect passenger flow from a total of 42 attributes in the smartcard system;
- (2) To add other variables that affect passenger flow, such as climate, time, space, and lag period, to establish a prediction model;
- (3) To apply multilayer perceptron (MLP), support vector regression (SVR), radial basis function (RBF) neural network, and long short-term memory network (LSTM) methods to forecast passenger flow with different types of time data series (weeks, weekdays, and holidays);
- (4) To propose an integrated-weight time-series forecast model that uses forecast data from the top three of the 80 intelligent forecast models as the adaptive factors;
- (5) To provide results that can be used as a reference by the government, industry, and related personnel.

The remaining sections are organized as follows: Section 2 is a literature review. In Section 3, we describe the research model and discuss the research design and methodology. Section 4 shows the results and findings. Finally, Section 5 presents the implications, limitations, and future work.

2. Literature Review

This section introduces related work on forecasting passenger flow using smartcard data, time series forecasting, and intelligent forecast methods.

2.1. Forecasting Passenger Flow by Smartcard Data

The smartcard is popular and convenient and can store a large amount of transaction data. Therefore, in the past decade, researchers have paid more and more attention to smartcard data. Ma et al. [10] used one-month data from smart bus cards to analyze the patterns of commuters in the area and the spatial distribution of movement. Eom et al. [11] applied the smartcard data from a five-day working week to learn about various social roles, such as the distribution of students and office workers in Seoul. Tao et al. [12] used smartcard data to visually compare the spatial-temporal trajectories of bus rapid transit trips and other bus trips.

To investigate factors relevant to forecasting passenger flow, Briand et al. [13] applied a Gaussian mixture model based on weather, time, and space to regroup passengers according to their public transportation habits in terms of time. Arana et al. [14] analyzed the impact of weather conditions on the number of public bus trips taken for shopping and personal business. Tang and Thakuria [15] used the unemployment rate, gasoline prices, weather conditions, transportation services, and socioeconomic factors to implement a quasi-experimental design to examine changes in the monthly average number of bus passengers on weekdays.

The literature on passenger flow forecasting in bus services can be divided into long-term and short-term forecasts. Traditional long-term passenger flow forecasting usually involves the use of regression techniques to estimate future travel demand [16]. The regression model is used to establish the relationship between the number of passengers and influencing factors, which includes demographic, economic, and land use information [17,18]. For short-term passenger flow forecasting, models based on statistics and computational intelligence have been studied extensively [19,20].

There has been much research on passenger flow forecasting, but most has not included bus passenger flow forecasting. We present some of the research techniques and methods that have been used in previous studies. Sun et al. [21] proposed a hybrid model based on wavelet analysis and the support vector machine to evaluate the historical passenger flow through the Beijing subway. Xie et al. [22] applied seasonal decomposition and a least squares support vector to find the best hybrid method for the short-term prediction of airline passengers. Liu and Chen [23] proposed a passenger flow prediction model using deep learning where an autoencoder deeply and abstractly extracts the nonlinear features in many hidden layers and a back-propagation algorithm is applied to train the model.

2.2. Time Series Forecasting

Forecasting passenger flow is a time-series research field because bus data points (including smartcard and meteorology data) are indexed in time order and are therefore time-series data. A time series is a sequence of discrete time data, and the use of a time series model can help organizations understand the underlying causes of trends and systemic patterns over time. As such, the following section briefly introduces relevant knowledge about time series. Time series are data arranged in time order, and the first time series model to be developed was the linear autoregressive moving average model (ARIMA), which was proposed by Box and Jenkins in 1970. The ARIMA model [24] consists of three components, and each component helps model different types of patterns. The autoregressive (AR) component attempts to explain the patterns between any time period and previous lag periods; the moving average (MA) component can adapt the new forecasts to previous forecast errors (error feedback term); and the integrated (I) component indicates trends or other integrative processes in the data.

Traffic flow data are time series of periodic and irregular fluctuations, and many studies have used time-series methods to predict traffic flow. Hou et al. [25] combined

ARIMA with a wavelet neural network to overcome the limitations of using ARIMA for short-term forecasting of traffic flow. Xu et al. [9] used seasonal differences to eliminate nonstationary seasonal ARIMA and combined these data with support vector regression to predict the demand of the aviation industry. Wang et al. [26] used wavelet analysis to detect abnormal passenger flow to estimate the sudden traffic peak and then used a multiple regression model to estimate the peak time. Finally, they used seasonal ARIMA to estimate passenger flow.

In terms of an intelligent time-series model, AR neural network (ARNN) is a classic intelligent time-series model that uses a neural network to learn AR coefficients [27]. Further, we can collect many influential variables and lag periods of the dependent variable, and then we use MLP, SVR, RBF neural network, and LSTM network to train their parameters for building intelligent time series models. From the practical viewpoint, it is critical to properly handle weights in time series, and weighted time-series models include weighting on recent observations, important variables, and the better forecasting methods. For example, Hajirahimi and Khashei [28] proposed a weighted sequential hybrid model to calculate each model weight to construct a final hybrid output for time series forecasting; Tsai et al. [29] proposed a multifactor fuzzy time-series fitting model to weight the three significant variables; Jiang et al. [30] presented a weighted time-series forecasting model to weight recent observations.

2.3. Intelligent Forecast Methods

This section introduces four intelligent forecast methods, and they are applied to forecast the collected data in this study: multilayer perceptron, support vector regression, radial basis function neural network, and long short-term memory network.

(1) Support vector regression (SVR)

The support vector machine (SVM) is a supervised learning algorithm for data classification and regression analysis that was developed by Vapnik and colleagues [31]. The SVM is used for classification problems, known as support vector classification (SVC) problems, and regression problems, known as support vector regression (SVR) problems. The main purpose of SVR is to find the best separation hyperplane to separate clustered data to solve nonlinear problems. SVR is quite good when dealing with small samples and can handle high-dimensional attributes without relying on all available data for classification, but its disadvantage is that its efficiency is very low for a large number of forecast samples. In addition, the SVM needs to find a suitable kernel function, such as linear, polynomial, sigmoid, or radial basis functions, and it is sensitive to missing data. The SVM can be used to solve problems in many fields, such as text classification, image classification, and time-series prediction.

In forecasting traffic passenger flow, SVR is suitable for nonlinear and complex models. Castro-Neto et al. [32] considered that SVR cannot be fully trained with real-time data. To address this, they developed online SVR models. To reduce the computational complexity of the SVM, Xie et al. [22] proposed the combined seasonal decomposition and least squares support vector to get the best hybrid method for short-term forecasting of airline passengers.

(2) Multilayer perceptron neural network (MLP)

Perceptron is a type of artificial neural network invented by Rosenblatt [33]. It can be regarded as the simplest form of feedforward neural network, and it is a binary linear classifier. An MLP consists of at least three layers of nodes (input layer, hidden layer, and output layer). Except for the input nodes, each node is a neuron that uses a nonlinear activation function. The MLP uses supervised backpropagation learning, and its multilayer structure and nonlinear activation function distinguish it from linear perception. The MLP is a nonlinear learning model that can be processed in parallel and has good fault tolerance. It can be used as a real-time online learning model with associative memory, adaptivity, and self-learning ability. To make the output of the MLP as close to the actual target value

as possible, a set of optimal weight values must be found in the training model, and one would need to determine the number of neurons used in each hidden layer.

Past studies have used MLP models to forecast multifactor problems. Ma et al. [34] used the MLP to forecast the network-wide co-movement patterns of all traffic flows, and they used ARIMA to postprocess the residual of the MLP. Tsai et al. [35] proposed a multiple temporal unit MLP to forecast short-term passenger demand.

(3) Radial basis function (RBF) network

The radial basis function (RBF) network proposed by Broomhead and Lowe has an input layer, a hidden layer, and an output layer [36]. In an RBF network, the nonlinear transformation is from the input layer to the hidden layer, and then the linear transformation is from the hidden layer to the output layer. This can achieve mapping from the input layer space to the output layer space, approximate any nonlinear function, and deal with difficult problems. The RBF network is conceptually similar to the k-nearest neighbor (k-NN) algorithm. In the self-organizing learning stage, basis function centers can be obtained; in the supervised learning stage, the weight between the hidden layer and the output layer is obtained, and each parameter can be learned quickly, thus overcoming the local minima problem.

To solve the central problem of the RBF function and the number of neurons in the hidden layer, Li et al. [7] proposed a new dynamic radial basis function (RBF) network to predict outbound passenger traffic. Li et al. [37] proposed a multiscale radial basis function (MSRBF) network to address the issue of when the number of input vectors is large, there may be a large number of candidates in the initial model. The MSRBF network can be applied to forecast irregular fluctuations in subway passenger flow.

(4) Long Short-Term Memory (LSTM)

LSTM is a time recurrent neural network (RNN) that was first proposed by Hochreiter and Schmidhuber [38]. Due to its unique design structure, an LSTM is suitable for processing and predicting important events with exceedingly long intervals and delays in time series. An LSTM network is a special type of regression neural network that uses a forget gate, an input gate, and an output gate to control the storage units. LSTM overcomes the problems of a RNN through gradient disappearance and gradient explosion. LSTM applications include time-series forecasting, language modeling, machine translation, image captions, and handwriting recognition.

In forecasting passenger flow, Ke et al. [8] proposed a fusion convolutional long short-term memory network (FCL-Net) for forecasting short-term passenger demand. Xu et al. [39] developed a long short-term memory network to forecast bike-sharing trip production and attractions at different time intervals.

3. Proposed Model

Passenger flow forecasting is a nonlinear, nonstationary time series problem, and a good forecast result mainly depends on having a large amount of high-quality data and a large number of methods. Nowadays, there are many passenger flow forecast models; however, some issues can be improved to enhance performance, such as:

1. Passenger flow forecasting is a periodic pattern, and many forecast models have been proposed to address this pattern. Previous studies have shown that datasets of no more than one month can be used to predict passenger flow at intervals of 5 or 15 min, and some studies use longer time datasets to predict passenger flow, such as daily or weekly intervals. To avoid the impact of extreme passenger flow, some studies do not consider the data collected on national holidays or weekends, and some studies treat special data as another forecast model for separate training. There is some room for improvement to obtain satisfactory results.
2. To solve the shortcomings of the model, more and more studies are taking advantage of different methods that complement each other and proposing hybrid models to forecast passenger flow. These hybrid models mainly combine traditional algorithms

and neural networks, but their nature still has limitations. Hence, hybrid models can be further strengthened to obtain the dynamics and forecasts of passenger flow.

3. Most research in this area has focused on passenger flow forecasting for railways, high-speed railways, and subways. Compared with other public transportation systems, there have been fewer studies forecasting bus passenger flow based on smartcard data, and few studies have considered time-series lag periods as forecast variables.
4. Previous research on the spatio-temporal nature of smartcard data has been widely conducted, and different attributes have been used in these studies; however, these studies have rarely discussed why these attributes have been selected and which attributes should be used in these methods. The selection and combination of input attributes is an important bridge between methods and forecast results.

Based on the discussion above, current forecast models of bus passenger flow still have limitations in terms of attribute selection, methods, and public transportation models. To process these limitations, we propose an integrated-weight time-series model for forecasting bus passengers using smartcards. First, the proposed model considers the attributes of time, space, and the lag period and uses four intelligent forecast models (multilayer perceptron, support vector regression, RBF network, and LSTM network) to forecast passenger flow for different time series (weeks, weekdays, and holidays). Second, the forecast data from the top three of the 80 combined forecast models (8 lag periods \times 10 algorithms) were used as adaptive factors in the proposed model to enhance the forecast results.

The proposed time series forecasting model was revised from adaptive expectations theory [40,41]. Adaptive expectations theory is an economic theory that gives importance to past events when predicting future outcomes—a hypothetical process through which people can form expectations of what will happen in the future based on what has happened in the past. In a more complex and adaptive expectation model, different weights can be assigned to past values, and we can look at how different the fluctuations are from the predicted fluctuations.

To quickly understand the proposed model, Figure 1 shows a detailed explanation of the procedure to clarify the research process and computational steps involved. The proposed procedure, from top to bottom, includes data collection, data preprocessing, lag period testing, building a time-series forecast model, and evaluation and comparison.

Computational steps

The proposed procedure has five steps (see Figure 1). A detailed breakdown of the five steps is provided in the following sections.

Step 1: Data collection

In this step, two types of data were collected:

- (1) One type was smartcard data from a bus industry in Kaohsiung City, Taiwan; the data were collected over a total of 669 days, including 2,865,763 records from January 2018 to October 2019, 17 bus lines (routes), and 137 bus stations. There were 42 attributes in the collected data (see Table 1), covering 15 administrative districts of Kaohsiung City in Taiwan. Regarding data location, the longitude range is 22.58706 to 22.792377, and the latitude range is 120.32016 to 120.29944.
- (2) The other data type was meteorological data because the number of passengers boarding is often affected by many external factors, especially weather, which has always affected the travel behavior of passengers. Many researchers have presented the impact of weather conditions on passenger flow [13–15]. We collected weather data from the Kaohsiung Meteorological Bureau.

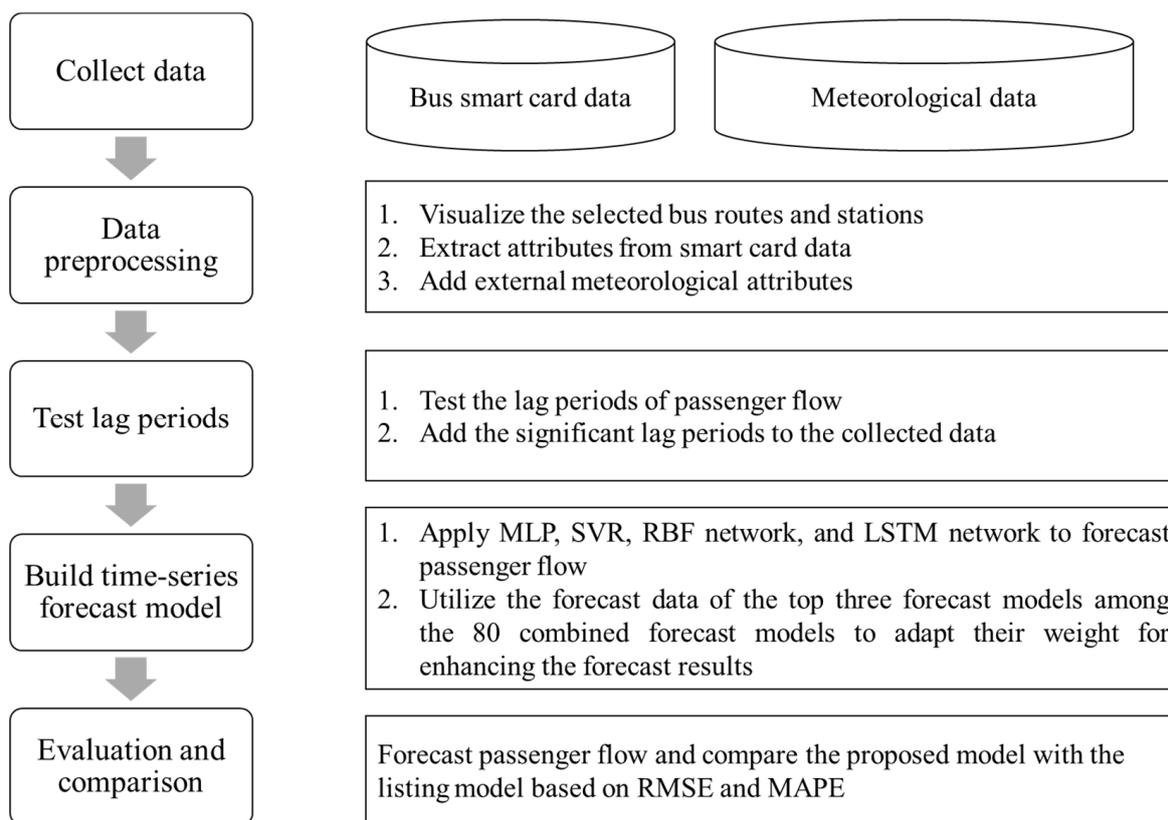


Figure 1. Proposed procedure.

Table 1. Original 42 data attributes.

| | | |
|---------------------------|-------------------------------------|-------------------------------------|
| Bus schedule number | Trading time for boarding | Card payment amount for exiting |
| Station number | Types of trading | Benefit points discount for exiting |
| Station name | Voice code for boarding | Free |
| Driver number | Boarding station code | Cash |
| Driver name | Boarding station name | Penalty fine |
| Bus number | Transferring discount amount | Making up the fare difference |
| Route number | Onboard card payment amount | Company subsidy amount |
| Route name | Boarding by benefit points discount | Transaction file name for boarding |
| Card number | Trading date for exiting | Transaction file name for exiting |
| Service type | Trading time for exiting | Outbound/return |
| Trade tickets | Types of trading for exiting | Counting status |
| Fare | Voice code for exiting | Counting date |
| Smartcard payment amount | Station code for exiting | Transferring group code |
| Trading date for boarding | Station name for exiting | Smartcard company |

Step 2: Data preprocessing

We calculated the total number of bus passengers (22 months) for each route based on smartcard data. Figure 2 shows the total number of passengers for each route. Among the 17 routes, Route 1, Route 7, and Route 52 had the top three numbers of passengers: 508,997, 545,915, and 272,968, respectively.

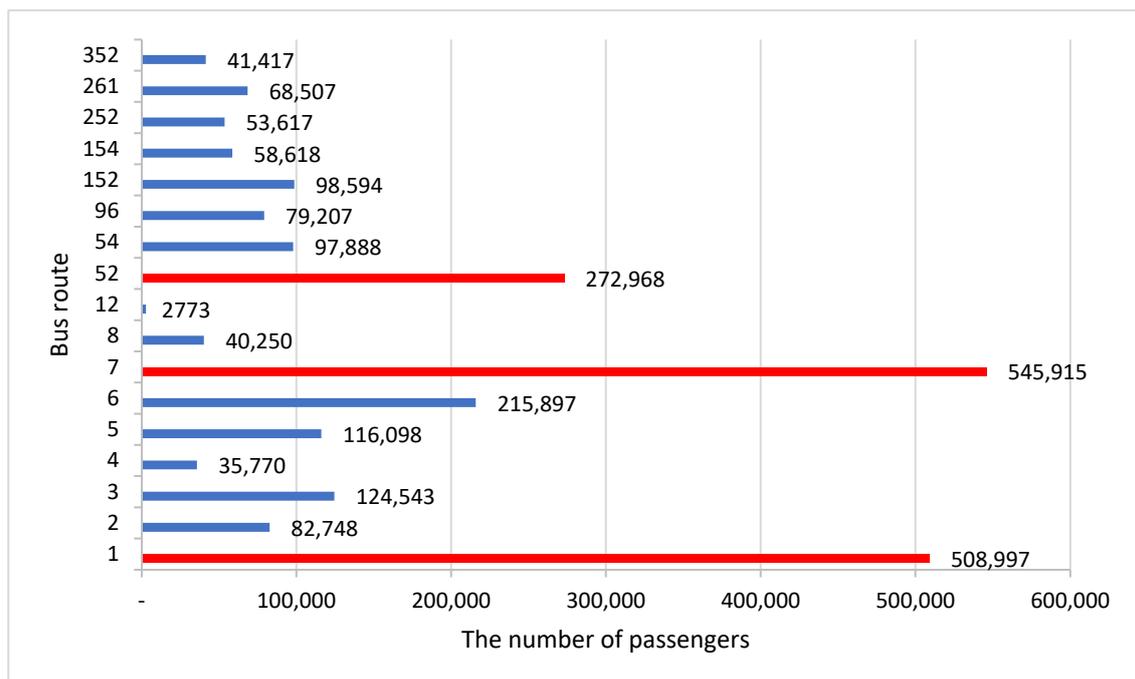


Figure 2. The number of passengers for each route.

Step 2.1: Extraction of attributes from smartcard data

This step involved the extraction of different time attributes from smartcard data as follows:

From the selected top three routes, the data from the three popular routes were divided into seven days (weeks), weekdays, and holidays. Weekdays were Monday to Friday (455 days), and holidays were Saturdays and Sundays (214 days). The total number of days was 669 (455 weekdays + 214 holidays), and a different day was used as an additional attribute. We extracted seven attributes from the smartcard data, including months, days, weeks, bus lines, bus stations, station passengers, and the number of passengers. Table 2 lists all the attributes used in this study in detail.

To check whether the number of rides was periodic, we plotted three figures to show the changes in the number of passengers for the top three routes based on the number of passengers per day on different days, as shown in Figures 3–5.

Figure 3 shows the number of passengers on three routes per day by week, which shows that the number of rides was periodic. Only Route 1 showed peak ride times, January 1 (New Year Day) and December 31 (New Year Eve), which are both national holidays. These long holidays are suitable times for going home or traveling, and as such, there is large-scale passenger flow. Route 7 has many bus stations, and the first half of the route is the same as Route 1; therefore, the number of passengers was found to have a similar periodicity to Route 1. For Route 52, because the bus station is different from the other two routes, the number of rides was found to be less similar to the other two routes, but it still showed periodicity.

Figure 4 shows the changes in the number of passengers per day for the three routes on weekdays. For a few days in the weekday period, peak passenger numbers occur, such as after 1 January (New Year's Day) and before the national holiday on 28 February, as people go home early before the holidays. In addition, 25 December (Christmas) is a religious holiday and a time when various industries launch marketing activities. As a result, people celebrating the festive season go out to purchase discounted goods.

Figure 5 shows the changes in the number of passengers on the three routes during the holidays. Compared with Figure 3, the weekly data show that 1 January (New Year's Day) and 31 December (New Year's Eve) are days with many passengers, and those two days

are also holidays; therefore, the trend for the numbers of passengers on the three routes showed a similar periodicity for holiday data.

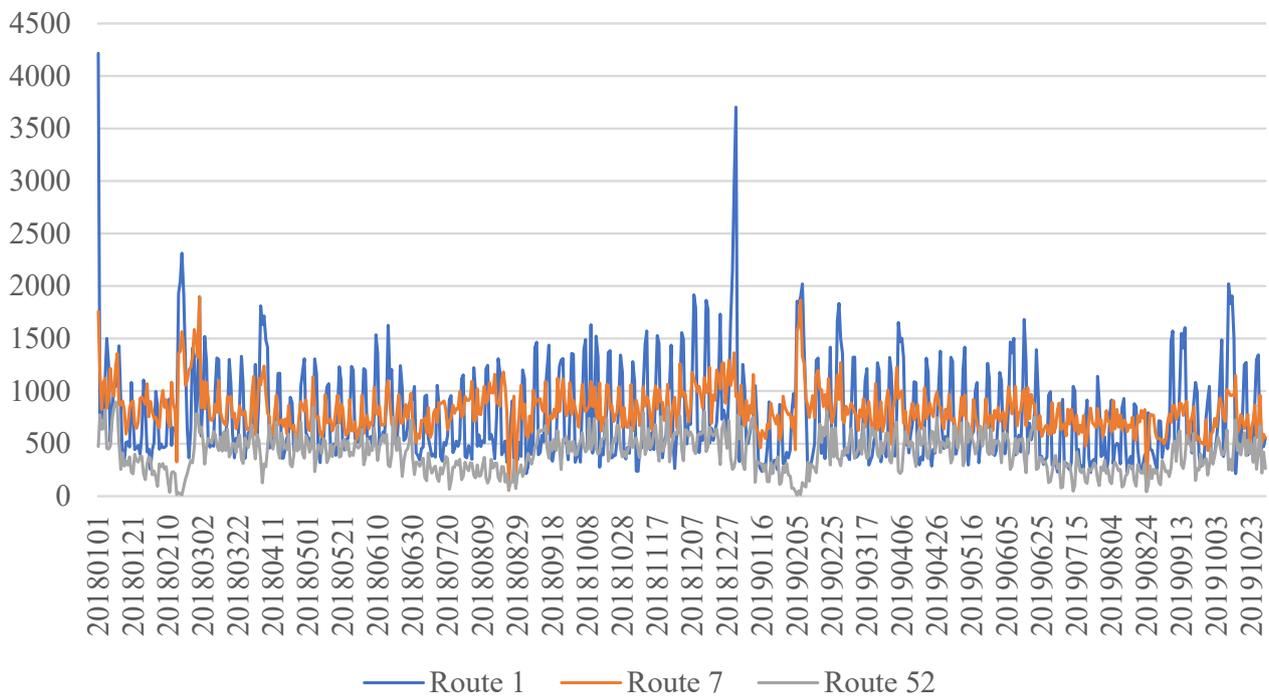


Figure 3. The number of passengers per day for the top three routes by week.

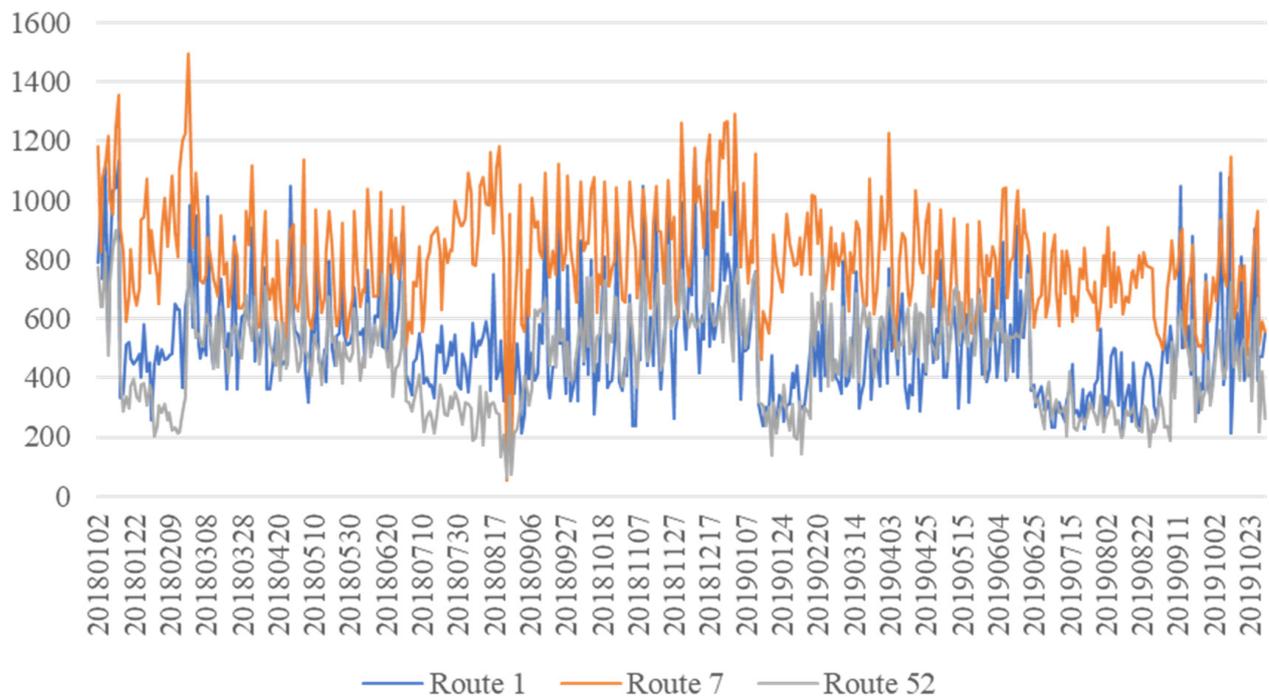


Figure 4. The number of passengers per day for the top three routes on weekdays.

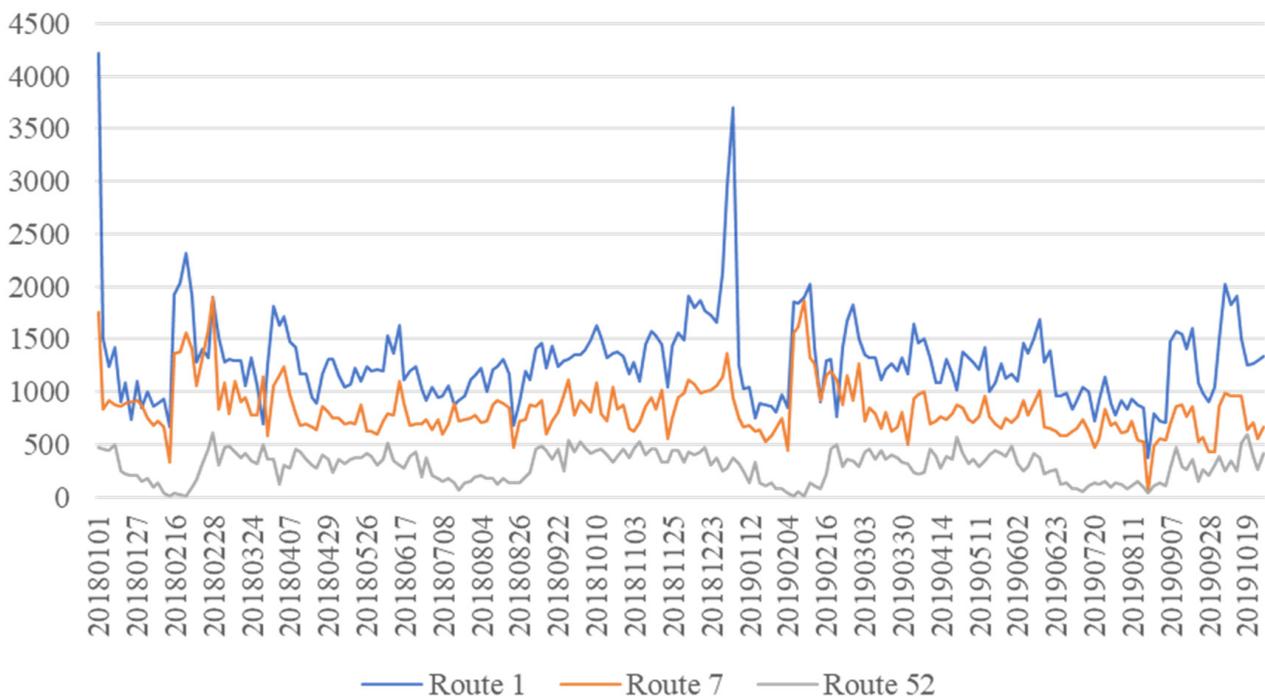


Figure 5. The number of passengers per day for the top three routes on holidays.

Step 2.2: Addition of external meteorological attributes

Previous studies [13–15] presented the impacts of weather factors on passenger flow. This study collected six types of meteorological data, including temperature, humidity, wind speed, rainfall, sunshine, and ultraviolet radiation. From the practical application of bus passenger flow forecasting, we collected data on 15 meteorological attributes, described in the meteorological type of Table 2.

Step 3: Test lag periods

In a time series, the autoregressive model is the output variable that is linearly dependent on its own previous value and random term. In order to test whether passenger flow has a time lag, this study used a partial autocorrelation function (PACF) to test how many passenger flows are significant at the 0.05 significance level, as the PACF is most useful for identifying the amount of lag in an autoregressive model. Further, lag- n is defined as follows: from the original data, the series values are moved forward n periods [42]. For example, lag 1 is moved forward 1 period; lag 10 is moved forward 10 periods. The test results are listed in Table 3 to show the lag periods for the number of passengers travelling on the top three routes based on the number of passengers per day for different daily data series (seven days, weekdays, and holidays). Table 3 shows that the largest lag period was seven for seven-day data; the largest lag period for the weekend data was six, and the holiday data had two lag periods.

Table 2. Attributes used in this study.

| Type | Attribute | Description |
|-------------|---|--|
| Smartcard | Month | "1" denotes Jan, "2" means Feb, . . . , "12" represents Dec |
| | Day | "1" denotes the first day for each month, "2" means the second day for each month, . . . , "31" denotes the last day for each month. |
| | Week | "1" denotes Monday, "2" means Tuesday, . . . , "7" represents Sunday. |
| | Bus line | The attribute was applied to visualize the heat map. |
| | Bus station | The attribute was applied to visualize the heat map. |
| | Station passengers | The attribute represents the passengers boarding at each bus station, which was applied to calculate the number of passengers at all stations for each day |
| | Passengers | Number of passengers on the bus line for each day |
| Meteorology | Temp | Average temperature, degrees Celsius, °C |
| | Tmax | Maximum temperature, degrees Celsius, °C |
| | Tmin | Minimum temperature, degrees Celsius, °C |
| | RH | Relative humidity, percent % |
| | RH_min | Minimum relative humidity, percent % |
| | WS | The wind speed was taken as the average value 10 min before the observation point, meters per second (m/s). |
| | WS_max | The maximum wind speed was taken as the maximum instantaneous wind speed within 1 h before the observation point, meters per second (m/s). |
| | Precp | The precipitation was taken as the total rainfall in a day, milliliters per day. |
| | Precp_hr | Total number of rainy hours in a day, number of hours |
| | Precp_10max | Maximum precipitation within ten minutes of the day, milliliters per ten minutes. |
| | Precp_hrmax | Maximum precipitation within an hour of the day, milliliters per hour. |
| | SunS | Sunshine hours, number of hours |
| | SunS_rate | The sunshine rate is a percentage ratio of the recorded bright sunshine duration and daylight duration in a day, percent %. |
| GloblRad | Global radiation refers to a value used to measure the solar radiation energy for a given time and area, megajoules per square meter and per day, MJ/m ² . | |
| UVImax | The maximum ultraviolet index refers to the international measurement standard for the solar ultraviolet (UV) radiation intensity at a certain place on a certain day; the index value from 0 to 11+ is divided into five levels. | |
| Lag period | Lag 1 | A Lag 1 autocorrelation is the correlation between values that are one time period apart. |
| | Lag 2 | a Lag k autocorrelation is a correlation between values that are k time periods apart, where $k = 2, 3, 4, 5, 6, 7$. |
| | Lag 3 | |
| | Lag 4 | |
| | Lag 5 | |
| | Lag 6 | |
| | Lag 7 | |

Table 3. Number of lag periods for each route in different time data.

| | Week (Seven Days) | Weekday | Holiday |
|----------|-----------------------|-------------------|----------|
| Route 1 | Lag: 1, 2, 4, 6, 7 | Lag:1, 2, 4, 5 | Lag:1 |
| Route 7 | Lag: 1, 2, 3, 4, 7 | Lag:1, 3, 4, 5, 6 | Lag:1, 2 |
| Route 52 | Lag: 1, 2, 3, 4, 6, 7 | Lag:1, 2, 3, 4, 5 | Lag:1, 2 |

Step 4: Establishment of an integrated-weight time-series forecasting model

After extracting/adding the attributes of Steps 2 and 3, all attributes are listed in Table 2. This study proposed an integrated-weight time-series forecasting model to improve forecasting performance. That is, this step applied the four intelligent forecast methods to train these parameters of multi-variables and lag periods based on different data series and forecast passenger numbers. Further, we input the collected data with time order to train their parameters by SVR, MLP, RBF network, and LSTM; hence, these models are called intelligent time-series models. In addition, the four intelligent forecast methods (MLP, SVR, RBF network, and LSTM) were separated into 10 models according to the hidden layer, activation function, and kernel functions, as shown in Table 4.

Table 4. Abbreviation of ten intelligent forecast models.

| Model Abbreviation | Full Name |
|--------------------|--|
| MLP_1_lin | MLP with 1 hidden layer and linear activation function |
| MLP_1_log | MLP with 1 hidden layer and logistic activation function |
| MLP_2_lin | MLP with 2 hidden layer and linear activation function |
| MLP_2_log | MLP with 2 hidden layer and logistic activation function |
| SVM_lin | SVR with linear kernel function |
| SVM_pol | SVR with polynomial kernel function |
| SVM_rbf | SVR with RBF kernel function |
| SVM_sig | SVR with sigmoid kernel function |
| RBF net | Radial basis function network |
| LSTM | Long short-term memory |

The proposed model is based on the concept of adaptive expectation theory [40,41] to adapt the forecast data of the top three of the 80 combined forecast models (10 intelligent forecast models with 8 different lag periods, as shown in Table 5). The equation used in the proposed model is as follows.

$$F(t) = \alpha \times first(t) + \beta \times second(t) + \gamma \times third(t) + T(t - 1) \quad (1)$$

where

$F(t)$ is the forecast of the number of passengers at time t ,

$T(t - 1)$ denotes the actual number of passengers at time $(t - 1)$,

$first(t)$ represents the forecast of the best model for the number of passengers at time t ,

$second(t)$ is the forecast of the second-best model for the number of passengers at time t ,

$third(t)$ denotes the forecast of the third-best model for the number of passengers at time t ,

α is the parameter of $first(t)$,

β represents the parameter of $second(t)$,

γ denotes the parameter of $third(t)$, and the range of α , β , and γ is from -1 to 1 (-1 means a negative correlation, and 1 represents a positive correlation).

Table 5. Results of 80 combined models for Route 1 on week data.

| Algorithm | | No lag | Lag 1 | Lag 2 | Lag 3 | Lag 4 | Lag 5 | Lag 6 | Lag 7 |
|-----------|------|---------|---------|---------|---------|---------------|----------------|----------------|----------------|
| LSTM | RMSE | 346.983 | 323.136 | 314.640 | 313.599 | 311.582 | 270.137 | 246.327 | 235.838 |
| | MAPE | 55.297 | 51.271 | 54.108 | 53.805 | 46.663 | 42.201 | 30.713 | 29.091 |
| MLP_1_lin | RMSE | 419.087 | 457.096 | 440.914 | 464.523 | 487.629 | 465.230 | 444.713 | 441.920 |
| | MAPE | 50.414 | 40.028 | 42.689 | 39.118 | 38.048 | 39.447 | 42.019 | 42.427 |
| MLP_1_log | RMSE | 413.770 | 422.451 | 437.535 | 442.538 | 454.443 | 413.121 | 425.137 | 473.599 |
| | MAPE | 53.621 | 48.826 | 43.573 | 42.355 | 40.548 | 54.963 | 46.756 | 38.588 |
| MLP_2_lin | RMSE | 417.011 | 436.597 | 606.395 | 496.521 | 520.526 | 451.678 | 455.457 | 481.922 |
| | MAPE | 51.544 | 43.844 | 58.925 | 38.235 | 39.951 | 41.111 | 40.371 | 38.435 |
| MLP_2_log | RMSE | 423.232 | 456.767 | 463.386 | 427.124 | 433.571 | 448.466 | 435.913 | 466.868 |
| | MAPE | 48.477 | 40.148 | 39.307 | 46.883 | 44.817 | 41.475 | 43.736 | 38.926 |
| RBF net | RMSE | 407.477 | 406.968 | 436.204 | 416.700 | 413.988 | 406.394 | 405.928 | 400.344 |
| | MAPE | 67.738 | 65.773 | 44.045 | 51.765 | 54.012 | 65.769 | 64.782 | 65.439 |
| SVR_lin | RMSE | 415.328 | 438.604 | 444.973 | 460.935 | 440.719 | 434.932 | 435.827 | 435.218 |
| | MAPE | 52.564 | 43.338 | 41.875 | 39.496 | 43.074 | 45.093 | 43.985 | 43.888 |
| SVR_pol | RMSE | 428.524 | 413.265 | 424.142 | 434.607 | 419.408 | 417.863 | 422.436 | 423.037 |
| | MAPE | 46.397 | 54.061 | 41.223 | 44.322 | 50.989 | 52.713 | 49.107 | 48.805 |
| SVR_rbf | RMSE | 437.040 | 414.256 | 422.609 | 415.376 | 419.644 | 421.224 | 420.811 | 423.770 |
| | MAPE | 43.773 | 53.330 | 48.703 | 52.607 | 50.366 | 50.886 | 50.222 | 48.335 |
| SVR_sig | RMSE | 416.775 | 445.232 | 448.546 | 446.987 | 500.864 | 441.484 | 406.373 | 410.386 |
| | MAPE | 51.681 | 41.895 | 41.223 | 41.478 | 38.510 | 43.311 | 62.067 | 54.371 |

Model abbreviations are shown in Table 4, and the bold print denotes the top three of the 80 combined models in terms of the RMSE and MAPE.

To optimize the parameters for α , β , and γ , the top three forecasted data points were used to adapt these parameters based on the minimum root mean square error (RMSE) and average absolute percentage error (MAPE) by using Equation (1). First, we set a feasible step iteration (step = 0.001) to produce the best parameters for α , β , and γ . Because the right-hand side of equation (1) has the actual number of passengers at the previous time $t - 1$, the parameters of the first three forecast data points would fall between plus and minus one.

Step 5: Evaluation and comparison

In order to evaluate the proposed model and compare it with the listed models, in this step, the minimum RMSE and MAPE criteria were used for the evaluation and comparison. In terms of data, the data from three routes with three different types of days were compared experimentally based on an 8:2 ratio of training and testing data with time order. The data collected for each route were as follows: weekly data 669 = 535 training data + 134 testing data; weekday data 455 = 364 training data + 91 testing data; weekend data 214 = 171 training data + 23 testing data. The performance indicators used were the root mean square error (RMSE) and average absolute percentage error (MAPE). The equations are shown as Equations (2) and (3).

$$RMSE = \sqrt{\frac{\sum_{t=1}^n |F(t) - T(t)|^2}{n}} \tag{2}$$

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{F(t) - T(t)}{T(t)} \right| \tag{3}$$

where $T(t)$ is the actual number of passengers at time t , $F(t)$ is the forecasted number of passengers at time t , and n is the number of data points.

4. Experimental Comparison

From the procedure proposed in Section 3, the initial data analysis and acquisition of necessary attributes were selected. This section describes the experiments implemented to

compare the proposed model with the listed models based on the minimum RMSE and MAPE criteria and then gives some findings.

4.1. Experimental Results

The experiments involved nine data series based on three routes with three different day types; each data series was partitioned into 80% training data and 20% testing data by time sequence. Based on Steps 4 and 5, which were presented in Section 3, this section presents the experimental results.

(1) Route 1

The results of 80 combined models for Route 1 for weekly data are shown in Table 5. In terms of RMSE, the top three best results were LSTM with lag 7, LSTM with lag 6, and LSTM with lag 5. For MAPE, the top three best results were LSTM with lag 7, LSTM with lag 6, and MLP_1_lin with lag 4 for the MAPE. We used Equation (1) to adapt the forecast data of the top three best forecast models to the minimal RMSE and MAPE. Similarly, the other two data series for Route 1 were experimented with based on Steps 4 and 5. Finally, the forecast data for the top three forecast models under the minimal RMSE were collected, and the results for Route 1 are shown in Table 6. The results show that the proposed model performed better than the listed models for Route 1, and the top three models were LSTM intelligent forecast methods.

(2) Route 7

As with the Route 1 experiment, we applied the proposed model to adapt the forecast data from the top three best forecast models for three data series of Route 7 using the minimal RMSE. The results are shown in Table 7. Table 7 shows that the proposed model was better than the listed models for Route 7 in terms of RMSE and MAPE.

Table 6. Results of proposed model for Route 1 based on minimal RMSE.

| Week | | | Weekday | | | Holiday | | |
|----------|------|----------------|-----------------|------|----------------|-----------------|------|----------------|
| LSTM | RMSE | 235.838 | LSTM | RMSE | 175.697 | LSTM | RMSE | 255.503 |
| lag 7 | MAPE | 29.091 | lag 4 | MAPE | 34.973 | lag 4 | MAPE | 21.913 |
| LSTM | RMSE | 246.327 | LSTM | RMSE | 177.303 | LSTM | RMSE | 255.907 |
| lag 6 | MAPE | 30.713 | lag 1 | MAPE | 34.003 | lag 5 | MAPE | 21.469 |
| LSTM | RMSE | 270.137 | LSTM | RMSE | 177.751 | LSTM | RMSE | 259.302 |
| lag 5 | MAPE | 42.201 | lag 2 | MAPE | 36.563 | lag 7 | MAPE | 21.085 |
| Proposed | RMSE | 199.882 | Proposed method | RMSE | 115.963 | Proposed method | RMSE | 171.627 |
| | MAPE | 54.534 | | MAPE | 33.068 | | MAPE | 20.426 |

Model abbreviations are shown in Table 4, and the bold digits denote the optimal results among the four models for RMSE and MAPE.

Table 7. Results of proposed model for Route 7 based on minimal RMSE.

| Week | | | Weekday | | | Holiday | | |
|----------|------|---------------|-----------------|------|---------------|-----------------|------|----------------|
| LSTM | RMSE | 142.964 | RBF net | RMSE | 131.499 | LSTM | RMSE | 148.165 |
| (no lag) | MAPE | 27.182 | lag 4 | MAPE | 15.400 | (no lag) | MAPE | 36.102 |
| LSTM | RMSE | 143.360 | MLP_2_log | RMSE | 131.794 | LSTM | RMSE | 157.255 |
| lag 5 | MAPE | 27.397 | lag 1 | MAPE | 15.607 | lag 2 | MAPE | 37.800 |
| LSTM | RMSE | 144.51 | MLP_2_log | RMSE | 133.274 | LSTM | RMSE | 164.564 |
| lag 4 | MAPE | 27.964 | lag 5 | MAPE | 15.657 | lag 1 | MAPE | 40.155 |
| Proposed | RMSE | 93.682 | Proposed method | RMSE | 82.124 | Proposed method | RMSE | 110.650 |
| | MAPE | 26.728 | | MAPE | 15.295 | | MAPE | 35.097 |

Model abbreviation is shown in Table 4, and the bold digits denote the optimal results among the four models for RMSE and MAPE.

(3) Route 52

Similarly, we only list the results for three data series of Route 52 in terms of the minimal RMSE, as shown in Table 8. The results show that the proposed model performs better than the listed models for Route 7 in terms of RMSE and MAPE.

Table 8. Results of proposed model for Route 52 based on minimal RMSE.

| Week | | | Weekday | | | Holiday | | |
|----------|------|---------------|----------|------|---------------|----------|------|---------------|
| LSTM | RMSE | 117.833 | LSTM | RMSE | 107.196 | LSTM | RMSE | 121.741 |
| lag 7 | MAPE | 37.901 | lag 1 | MAPE | 29.631 | lag 4 | MAPE | 52.935 |
| LSTM | RMSE | 119.873 | LSTM | RMSE | 109.904 | LSTM | RMSE | 121.767 |
| lag 6 | MAPE | 38.999 | lag 7 | MAPE | 28.371 | lag 6 | MAPE | 53.773 |
| LSTM | RMSE | 131.381 | LSTM | RMSE | 110.222 | LSTM | RMSE | 121.943 |
| lag 5 | MAPE | 48.909 | lag 2 | MAPE | 29.768 | lag 5 | MAPE | 52.555 |
| Proposed | RMSE | 79.963 | Proposed | RMSE | 78.179 | Proposed | RMSE | 60.968 |
| | MAPE | 38.126 | | MAPE | 26.025 | | MAPE | 41.958 |

Model abbreviation is shown in Table 4, and the bold digits denote the optimal results among the four models for RMSE and MAPE.

4.2. Findings and Discussion

The experimental results show that the proposed model is better than the listed models based on the minimum RMSE and MAPE criteria. However, there are some other findings to be discussed, as follows.

(1) Key attributes

In the forecast experiments, this study used the forecast data from the top three forecast models to adapt the optimal forecast. Simultaneously, we obtained the attributes of the top three forecast models. Additionally, we used the top three forecast models to rank the smartcard, lag periods, and meteorological attributes based on their impacts on passenger numbers. Then, we took the common attributes (at least two of the same attributes of the top three models) as the key attributes. The ordering of the key attributes of bus routes for different time series are shown in Table 9. Based on the ordering of key attributes, we identified the following features:

Routes 1 and 7: The three different time series for Route 1 have the same top three key attributes: Precp_10max, Precp, and Precp_hrmax. The top three key attributes in the weekly data for Route 7 are the same as for Route 1. This means that rainfall is an important factor for passenger flow on Route 1 and in the weekly data for Route 7. The top two key attributes in the weekday and holiday data for Route 7 are lag 1 and week, which shows that the passenger flow through Route 7 on weekdays and holidays is dependent on the week (Monday, Tuesday, . . . , Sunday) and the passenger flow in the previous period.

Because the bus stations on Routes 1 and 7 are close to the university, high-speed rail, theme park, and tourist attractions, the collected data reveal that most passengers on these routes are students and tourists. Passengers want to go to schools and theme parks, and most of them will take the two routes. In addition, Route 7 has a highway transit station and Buddha memorial station. These two stations are important transportation and tourist attractions; hence, climate attributes (such as SunS, SunS_rate, and Tmin) influence the activity of tourists on Route 7 for the holiday data series.

Route 52: The bus stations on Route 52 are close to the university, high-speed rail, hospital, and detention center. These bus stations are used by students, patients, government employees, and their families, and the students, patients, and government employees are off duty on holiday; hence, the passenger flow is influenced by climate attributes (such as Precp_10max, Precp_hrmax, and Precp_hr) on the weekdays.

Three Routes: From Table 9, it can be seen that the three attributes (smartcard, meteorology, and lag period) affect the passenger flow, but, in different data series, the different attributes have different degrees of influence. Overall, the passenger flow through Route 1 and the weekly data for Route 7 have more attributes in common than other data.

Table 9. Key attributes of bus routes for different time data.

| Route | Dataset | Ordering of Attribute Importance |
|----------|---------|--|
| Route 1 | week | Precp_10max > Precp > Precp_hrmax > lag 7 > lag 1 > Precp_hr > week > lag 5 > lag 2 > RH > month > lag 6 > Tmax > UVImax |
| | weekday | Precp_10max > Precp > Precp_hrmax > Precp_hr > lag 1 > week > Temp > WS > Tmax |
| | holiday | Precp_hrmax > Precp > Precp_10max > lag 1 > Precp_hr > week > month > SunS_rate > Tmax |
| Route 7 | week | Precp > Precp_hrmax > Precp_10max > Precp_hr > lag 1 > GloblRad > WS > RH_min > Temp > Tmin > RH > SunS_rate > UVImax |
| | weekday | lag 1 > week > month > Temp > Tmin > GloblRad > Precp_hr > RH_min > Precp > Precp_10max |
| | holiday | lag 1 > week > SunS > SunS_rate > Tmin > month > GloblRad > Precp_hr > RH_min > WS > lag 2 |
| Route 52 | week | lag 1 > week > lag 2 > SunS_rate > Tmax > GloblRad > WS_max > UVImax > Month > Precp > SunS |
| | weekday | Precp > lag 1 > Precp_10max > Precp_hrmax > Precp_hr > RH > Tmin > WS_max > Day > lag 5 |
| | holiday | lag 1 > SunS_rate > lag 2 > SunS > Precp > Precp_10max > GloblRad > Tmax > Precp_hr |

(2) Lag period

From the lag period test results shown in Table 3, it can be seen that the number of lag periods is consistent with the different time-series data. We can see that the lag period is seven for the weekly data (a week has seven days); the holiday data are organized by Saturdays and Sundays and have a lag period of two; the lag period for the weekday data is five. We checked whether the lag period is consistent with the number of lags for the top three models; if the same lag periods exist, then the data have seasonal variation (seasonality). Seasonality means that the time series data have periodic, repetitive, and predictable patterns [43]. We summarized the data from Tables 6–8, and the lag period results for the top three models for each route for different time series are shown in Table 10.

Based on Table 10, it can be seen that the weekly data for Routes 1 and 52 are consistent with the number of days in a week, as the number of lags is seven. This means that the weekly time-series data for Routes 1 and 52 have a weekly seasonal variation (seasonality), as shown in 6. Thus, it is necessary to add the passenger flow lag periods to forecast the number of passengers.

Table 10. Lag period of top three models for each route in different time data.

| | Week (Seven Days) | Weekday | Holiday |
|----------|-------------------|-----------------|-------------------|
| Criteria | RMSE | RMSE | RMSE |
| Route 1 | LSTM lag 7 | LSTM lag 4 | LSTM lag 4 |
| | LSTM lag 6 | LSTM lag 1 | LSTM lag 5 |
| | LSTM lag 5 | LSTM lag 2 | LSTM lag 7 |
| | LSTM (no lag) | RBF net lag 4 | LSTM (no lag) |
| Route 7 | LSTM lag 5 | MLP_2_log lag 1 | LSTM lag 2 |
| | LSTM lag 4 | MLP_2_log lag 5 | LSTM lag 1 |
| | LSTM lag 7 | LSTM lag 1 | LSTM lag 4 |
| Route 52 | LSTM lag 6 | LSTM lag 7 | LSTM lag 6 |
| | LSTM lag 5 | LSTM lag 2 | LSTM lag 5 |

Bold text denotes that the lag period of the top three models is consistent with the number of days in a week.

(3) Model performance

From Table 10, we can see that most of the top three models used LSTM, because LSTM is suitable for processing and predicting important events with very long intervals and delays in the time series [44]. Furthermore, the proposed model was found to be better than the listed models for each route for different time series, as shown in Tables 6–8. Therefore, the proposed model has some advantages: (1) The proposed model incorporates the smartcard, meteorology, and lag period attributes; (2) To enhance the forecast performance, this study proposed an integrated-weight time-series model to adapt the data from the

top three of the 80 combined forecast models; (3) Bus data (including smartcard and meteorology data) are time series data, and the time series model helps organizations to understand the underlying causes of trends and systemic patterns over time. Therefore, we propose the use of an intelligent time series model to forecast passenger numbers.

(4) Sensitivity analysis

A sensitivity analysis can determine the associations between attributes. It facilitates more accurate forecasting and is the process of adjusting only one input and studying how it affects the overall model [45]. Based on sensitivity analysis involving the removal of attributes, we removed the meteorology attributes, the first key attribute, and the second key attribute to show the forecasting ability and robustness of the proposed model. This study was based on the ordering of key attributes for the three routes, as shown in Table 9, and we used the weekly datasets of the three routes to conduct the sensitivity analysis. The results show that the proposed model without meteorological data has a larger RMSE than the model with data from all attributes for the three routes, as shown in Table 11. Further, removing the first key attribute generates a larger error, and removing the second key attribute also increases the error. Based on the sensitivity analysis, we can confirm that the meteorological data are important when building the proposed model, and the first and second key attributes affect the proposed model.

Table 11. Results of sensitivity analysis based on RMSE.

| | Route 1 | Route 7 | Route 52 |
|-----------------------------------|---------|---------|----------|
| Full attributes | 199.882 | 93.682 | 79.963 |
| Removal of meteorology attributes | 231.714 | 144.949 | 113.959 |
| Removal of first key attribute | 228.594 | 136.121 | 121.006 |
| Removal of second key attribute | 228.066 | 140.402 | 116.717 |

5. Conclusions

Since 2000, Taiwan has been implementing the AFC system, also called the smartcard system, in the transportation system. The widespread use of smartcards helps passengers greatly reduce their transaction time and helps companies collect a large amount of information. Although there is a smartcard system, serious traffic jams still occur. Therefore, a good passenger flow forecast could be used to reduce traffic congestion, increase passenger convenience, and assist enterprises with formulating route planning, resetting timetables, and constructing other policies. In addition, a good passenger flow forecast can help cities reduce excessive energy consumption and carbon emissions and improve urban ecosystems to achieve sustainable development.

In order to achieve better prediction performance, we carried out the following steps:

- (1) We proposed an integrated-weight time-series forecast model to forecast passenger flow. We used real smartcard data to verify that the proposed model has good predictive capabilities, rather than using simulated data to show the research results. The experiments showed that the proposed model performed better than the listed models for each route for different time series, as shown in Tables 6–8.
- (2) In terms of the verification data, we focused on the top three routes with the most passengers out of the 17 routes—Route 1, which showed the largest fluctuations; Route 7, which has the largest number of passengers; and Route 52, which has the least number of passengers of the top three routes—as shown in Figure 2.
- (3) In terms of attribute screening, this study used smartcard data and time attributes as well as 15 external weather attributes. In addition, as the number of passengers varies with time, this is a time-series forecasting problem; hence, seasonal trends had to be considered. Therefore, we added the number of lags to the forecast of passenger flow.
- (4) As shown in Tables 6–8, we found that the data for each route could be partitioned by time (weeks, weekdays, and holidays) to improve the forecast result. Based on the key attributes shown in Table 9 and the lag periods of the top three models shown in

Table 10, the number of lags affected forecast results. Furthermore, most of the top three models are in the LSTM family, which presents a better forecast.

In terms of future work, we have two suggestions to further enhance this topic by making the results less conservative and improving the forecasting performance: (1) other attributes could be used in these forecast models, and (2) other methods (such as deep learning algorithms) could be applied to this topic.

Author Contributions: Conceptualization, C.-H.C. and M.-C.T.; methodology, C.-H.C.; software, Y.-C.C.; validation, C.-H.C. and M.-C.T.; formal analysis, M.-C.T.; investigation, Y.-C.C.; resources, C.-H.C.; data curation, M.-C.T.; writing—original draft preparation, Y.-C.C.; writing—review and editing, C.-H.C.; visualization, M.-C.T.; supervision, C.-H.C.; project administration, M.-C.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors acknowledge partial support from Ministry of Science and Technology, Taiwan (Grant No. MOST-2221-E-224-037-).

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

References

1. FTA. Transit's Role in Environmental Sustainability. 2015. Available online: <https://www.transit.dot.gov/regulations-and-guidance/environmental-programs/transit-environmental-sustainability/transit-role> (accessed on 10 August 2020).
2. TMTC. Public Transport Market Share from Taiwan's Ministry of Transportation and Communications. 2016. Available online: <https://www.motc.gov.tw/uploaddownloaddoc?file=public/201707031545021.pdf&filedisplay=201707031545021.pdf&flag=doc> (accessed on 10 August 2020).
3. Trépanier, M.; Tranchant, N.; Chapleau, R. Individual trip destination estimation in a transit smart card automated fare collection system. *J. Intell. Transport. Syst.* **2007**, *11*, 1–14. [\[CrossRef\]](#)
4. Cheon, S.H.; Lee, C.; Shin, S. Data-driven stochastic transit assignment modeling using an automatic fare collection system. *Transp. Res. Part C Emerg. Technol.* **2019**, *98*, 239–254. [\[CrossRef\]](#)
5. Transportation Research Board. *HCM 2010. Highway Capacity Manual*; Transportation Research Board: Washington, DC, USA, 2010.
6. Tang, T.; Liu, R.; Choudhury, C. Incorporating weather conditions and travel history in estimating the alighting bus stops from smart card data. *Sustain. Cities Soc.* **2020**, *53*, 101927. [\[CrossRef\]](#)
7. Li, H.; Wang, Y.; Xu, X.; Qin, L.; Zhang, H. Short-term passenger flow prediction under passenger flow control using a dynamic radial basis function network. *Appl. Soft Comput.* **2019**, *83*, 105620. [\[CrossRef\]](#)
8. Ke, J.; Zheng, H.; Yang, H.; Chen, X. Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transp. Res. Part C Emerg. Technol.* **2017**, *85*, 591–608. [\[CrossRef\]](#)
9. Xu, S.; Chan, H.; Zhang, T. Forecasting the demand of the aviation industry using hybrid time series SARIMA-SVR approach. *Transp. Res. Part E Logist. Transp. Rev.* **2019**, *122*, 169–180. [\[CrossRef\]](#)
10. Ma, X.; Liu, C.; Wen, H.; Wang, Y.; Wu, Y. Understanding commuting patterns using transit smart card data. *J. Transp. Geogr.* **2017**, *58*, 135–145. [\[CrossRef\]](#)
11. Eom, J.K.; Choi, J.; Park, M.S.; Heo, T.Y. Exploring the catchment area of an urban railway station by using transit card data: Case study in Seoul. *Cities* **2019**, *95*, 102364. [\[CrossRef\]](#)
12. Tao, S.; Corcoran, J.; Mateo-Babiano, I.; Rohde, D. Exploring Bus Rapid Transit passenger travel behaviour using big data. *Appl. Geogr.* **2014**, *53*, 90–104. [\[CrossRef\]](#)
13. Briand, A.; Côme, E.; Trépanier, M.; Oukhellou, L. Analyzing year-to-year changes in public transport passenger behaviour using smart card data. *Transp. Res. Part C Emerg. Technol.* **2017**, *79*, 274–289. [\[CrossRef\]](#)
14. Arana, P.; Cabezudo, S.; Peñalba, M. Influence of weather conditions on transit ridership: A statistical study using data from Smartcards. *Transp. Res. Part A Policy Pract.* **2014**, *59*, 1–12. [\[CrossRef\]](#)
15. Tang, L.; Thakuriah, P. Ridership effects of real-time bus information system: A case study in the City of Chicago. *Transp. Res. Part C Emerg. Technol.* **2012**, *22*, 146–161. [\[CrossRef\]](#)
16. Horowitz, R. Legal notes. *J. Futures Mark.* **1984**, *4*, 229–230. [\[CrossRef\]](#)
17. Taylor, B.; Miller, D.; Iseki, H.; Fink, C. Nature and/or nurture? Analyzing the determinants of transit ridership across US urbanized areas. *Transp. Res. Part A Policy Pract.* **2009**, *43*, 60–77. [\[CrossRef\]](#)

18. Chan, S.; Miranda-Moreno, L. A station-level ridership model for the metro network in Montreal, Quebec. *Can. J. Civ. Eng.* **2013**, *40*, 254–262. [CrossRef]
19. Karlaftis, M.; Vlahogianni, E. Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transp. Res. Part C Emerg. Technol.* **2011**, *19*, 387–399. [CrossRef]
20. Ma, Z.; Xing, J.; Mesbah, M.; Ferreira, L. Predicting short-term bus passenger demand using a pattern hybrid approach. *Transp. Res. Part C Emerg. Technol.* **2014**, *39*, 148–163. [CrossRef]
21. Sun, Y.; Leng, B.; Guan, W. A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system. *Neurocomputing* **2015**, *166*, 109–121. [CrossRef]
22. Xie, G.; Wang, S.; Lai, K. Short-term forecasting of air passenger by using hybrid seasonal decomposition and least squares support vector regression approaches. *J. Air Transp. Manag.* **2014**, *37*, 20–26. [CrossRef]
23. Liu, L.; Chen, R.C. A novel passenger flow prediction model using deep learning methods. *Transp. Res. Part C Emerg. Technol.* **2017**, *84*, 74–91. [CrossRef]
24. Box, G.E.P.; Jenkins, G.M. *Time Series Analysis: Forecasting and Control*, revised ed.; Holden Day: San Francisco, CA, USA, 1976.
25. Hou, Q.; Leng, J.; Ma, G.; Liu, W.; Cheng, Y. An adaptive hybrid model for short-term urban traffic flow prediction. *Phys. A Stat. Mech. Its Appl.* **2019**, *527*, 121065. [CrossRef]
26. Wang, H.; Li, L.; Pan, P.; Wang, Y.; Jin, Y. Early warning of burst passenger flow in public transportation system. *Transp. Res. Part C Emerg. Technol.* **2019**, *105*, 580–598. [CrossRef]
27. Triebe, O.; Laptev, N.P.; Rajagopal, R. AR-Net: A simple Auto-Regressive Neural Network for time-series. *arXiv* **2019**, arXiv:1911.12436.
28. Hajirahimi, Z.; Khashei, M. Weighted sequential hybrid approaches for time series forecasting. *Phys. A Stat. Mech. Its Appl.* **2019**, *531*, 121717. [CrossRef]
29. Tsai, M.-C.; Cheng, C.-H.; Tsai, M.-I. A Multifactor Fuzzy Time-Series Fitting Model for Forecasting the Stock Index. *Symmetry* **2019**, *11*, 1474. [CrossRef]
30. Jiang, Y.; Ye, Y.; Wang, Q. Study on Weighting Function of Weighted Time Series Forecasting Model in the Safety System. In Proceedings of the 2011 Asia-Pacific Power and Energy Engineering Conference, Wuhan, China, 25–28 March 2011; pp. 1–4. [CrossRef]
31. Cortes, C.; Vapnik, V. Support-vector networks. *Mach Learn* **1995**, *20*, 273–297. [CrossRef]
32. Castro-Neto, M.; Jeong, Y.; Jeong, M.; Han, L.D. Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions. *Expert Syst. Appl.* **2009**, *36*, 6164–6173. [CrossRef]
33. Rosenblatt, F. *The Perceptron—A Perceiving and Recognizing Automaton*; Report 85-460-1; Cornell Aeronautical Laboratory: Buffalo, NY, USA, 1957.
34. Ma, T.; Antoniou, C.; Toledo, T. Hybrid machine learning algorithm and statistical time series model for network-wide traffic forecast. *Transp. Res. Part C Emerg. Technol.* **2020**, *111*, 352–372. [CrossRef]
35. Tsai, T.; Lee, C.; Wei, C. Neural network based temporal feature models for short-term railway passenger demand forecasting. *Expert Syst. Appl.* **2009**, *36*, 3728–3736. [CrossRef]
36. Broomhead, D.S.; Lowe, D. Multivariable functional interpolation and adaptive networks. *Complex Syst.* **1988**, *2*, 321–355.
37. Li, Y.; Wang, X.; Sun, S.; Ma, X.; Lu, G. Forecasting short-term subway passenger flow under special events scenarios using multiscale radial basis function networks. *Transp. Res. Part C Emerg. Technol.* **2017**, *77*, 306–328. [CrossRef]
38. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
39. Xu, C.; Ji, J.; Liu, P. The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets. *Transp. Res. Part C Emerg. Technol.* **2018**, *95*, 47–60. [CrossRef]
40. Cagan, P. The monetary dynamics of hyper-inflation. In *Studies in the Quantity Theory of Money*; Friedman, M., Ed.; University of Chicago Press: Chicago, IL, USA, 1956.
41. Kmenta, J. *Elements of Econometrics*, 2nd ed.; Macmillan: New York, NY, USA, 1986.
42. Brockwell, P.J.; Davies, R.A. *Time Series: Theory and Methods*, 2nd ed.; Springer: New York, NY, USA, 1991.
43. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*, 2nd ed.; OTexts: Melbourne, Australia, 2018; Available online: <https://otexts.org/fpp2/> (accessed on 15 January 2021).
44. Zhang, Y.; Xiong, R.; He, H.; Pecht, M.G. Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries. *IEEE Trans. Veh. Technol.* **2018**, *67*, 5695–5705. [CrossRef]
45. Yang, J.; Kim, J.; Jiménez, P.A.; Sengupta, M.; Dudhia, J.; Xie, Y.; Golnas, A.; Giering, R. An efficient method to identify uncertainties of WRF-Solar variables in forecasting solar irradiance using a tangent linear sensitivity analysis. *Sol. Energy* **2021**, *220*, 509–522. [CrossRef]