Empirical Approaches to Improve the Predictive Performance of Spatial Distribution Neural Network Models

Gusri YALDI[a], Michael A P TAYLOR[b], Wen Long YUE[c]

[a] *Civil Engineering Department, Padang State Polytechnic, Padang, Indonesia 25163*
E:mail: gusri.yaldi@yahoo.com
[b] *School of Natural and Built Environments, University of South Australia, Adelaide*
E:mail: michael.taylor@unisa.edu.au
[c] *School of Natural and Built Environments, University of South Australia, Adelaide*
E:mail: wen.yue@unisa.edu.au

Abstract: Neural Network (NN) approach has been used for both people and commodity spatial distribution modeling since about two decades ago. However, this artificial intelligent based approach seemed to have the capability in calibrating the trip and commodity distribution only. It tends to have poor predictive capability when new datasets are used, especially for doubly constrained distribution models, where the information of the trip and commodity production and attraction is available. This paper reports empirical approaches integrated in a modeling framework used in order to improve not only the predictive capability of neural models for spatial distribution estimation, but also its calibration performance. A case study using US Commodity Flow Survey conducted in 2007 is used to implement those empirical approaches. Findings from this study suggest a promising improvement in the predictive capability of neural models.

Keywords: Neural model, Modeling framework, Predictive capability, Commodity flow distribution

1. INTRODUCTION

Due to its potential capabilities, NN has been adopted as a transportation modeling tool including as the traffic forecasting tool since 1990s (Dougherty, 1995). According to Cantarella and de Luca (2005), the NN has the capability to address three main demand simulation issues, namely (1) Trip generation, (2) Trip distribution, and (3) Modal split.

Shmueli (1998) suggested that NN has some advantages that can overcome the problems faced by the behavioral or disaggregate models. Those advantages are (1) NN discovers the relationship between variables automatically (without using any function representing the relationship between those variables) and the fitting takes place naturally which is found to be a complicated task for a disaggregate model to specify, and (2) NN does not face the inaccuracy problem in the relationship of the models to the real choice of the travellers as faced by the disaggregate model, where the model outcome differs from the reality. In addition, NN directly works on the data without the aid of additional models. Further, Shmueli et al. (1996) suggested that NN is relevant for addressing the problems requiring large scale, highly dimensional data analysis.

2. PREDICTIVE CAPABILITY

The initial utilization of neural network spatial distribution model for strategic forecasting was reported by Chin et al. (1992) as described by Dougherty (1995). The next one was reported by Black (1995). Both studies suggested promising results. The latter study compared the gravity and neural models, where neural models are developed based on the traditional form of the gravity model. Although both gravity and neural models produce good results, the neural models are found as a better calibration tool and recommended for future flow forecasting.

Mozolin et al. (2000) reported another strategic forecasting by neural models. The difference between the studies by Black (1995) and Mozolin et al. (2000) is that Black (1995) used neural models to calibrate the commodity and migration flows, while Mozolin et al. (2000) extended the analysis to the predictive/testing level. It was reported the neural model has poor generalization performance. Both studies developed neural models based on the structure of the traditional doubly constrained gravity model.

The common things from previous neural model studies generally suggested the neural model is good for calibration, however, it is not recommended for predicting trip distribution numbers for new datasets as its predictive capability is claimed as poor. However, a recent study by Yaldi et al. (2011b) suggested that the neural model predictive capability can be improved to the same level as the doubly constrained gravity model.

3. MODELING PROCEDURE ISSUES

Basically, the former neural models were developed in relatively the same modeling procedures. It starts from specifying the model architecture, number of layers, number of nodes for each layer, training algorithm (TA), and the activation function (AF). It is followed by training the model based on the specified input and target data. Prior to the training, the data must be normalized. These two major steps (model specification and training) are the steps to calibrate the doubly constrained trip distribution neural models.

The properties of the neural models such as the number of layers, TA, and AF could be different. The calibration results tend to outperform other modeling techniques such as the gravity (Black, 1995) and Box-Cox (Celik, 2004b) models. This is likely because the model was trained with an excessively number of epochs like 150000 and 100000 iterations. Training the neural models with these high epoch numbers could cause over fitting.

Different properties of neural models for trip distribution estimation may lead to different performance of its testing outputs as indicated in the study reported by Mozolin et al. and Yaldi et al. (2011b). The first study claimed neural models had a poor predictive capability, while the second one reported an opposite results. It used Quickrop TA, an ad hoc and heuristic modified backpropagation (BP) version, and Levenberg-Marquardt (LM) respectively. Details on LM can be found in (Hagan and Menhaj, 1994a).

The main reason behind this inconsistent testing performance of neural models is because there is no guideline or standardization for using NN for people and commodity distribution modeling. Therefore, Yaldi (2012) proposed a neural network modeling framework for a trip distribution model as illustrated in Figure 1. It contains the procedures in developing the neural models and its application for new datasets. This modeling framework is developed based on empirical works and comprehensive literature reviews on the NN and its application.

Generally, the previous studies constructed the neural models and calibrated them based on this framework. For examples are the network architecture, number of layers, and number of nodes for each layer. Those three sub aspects have a common feature in trip distribution modeling as depicted by Figure 2.

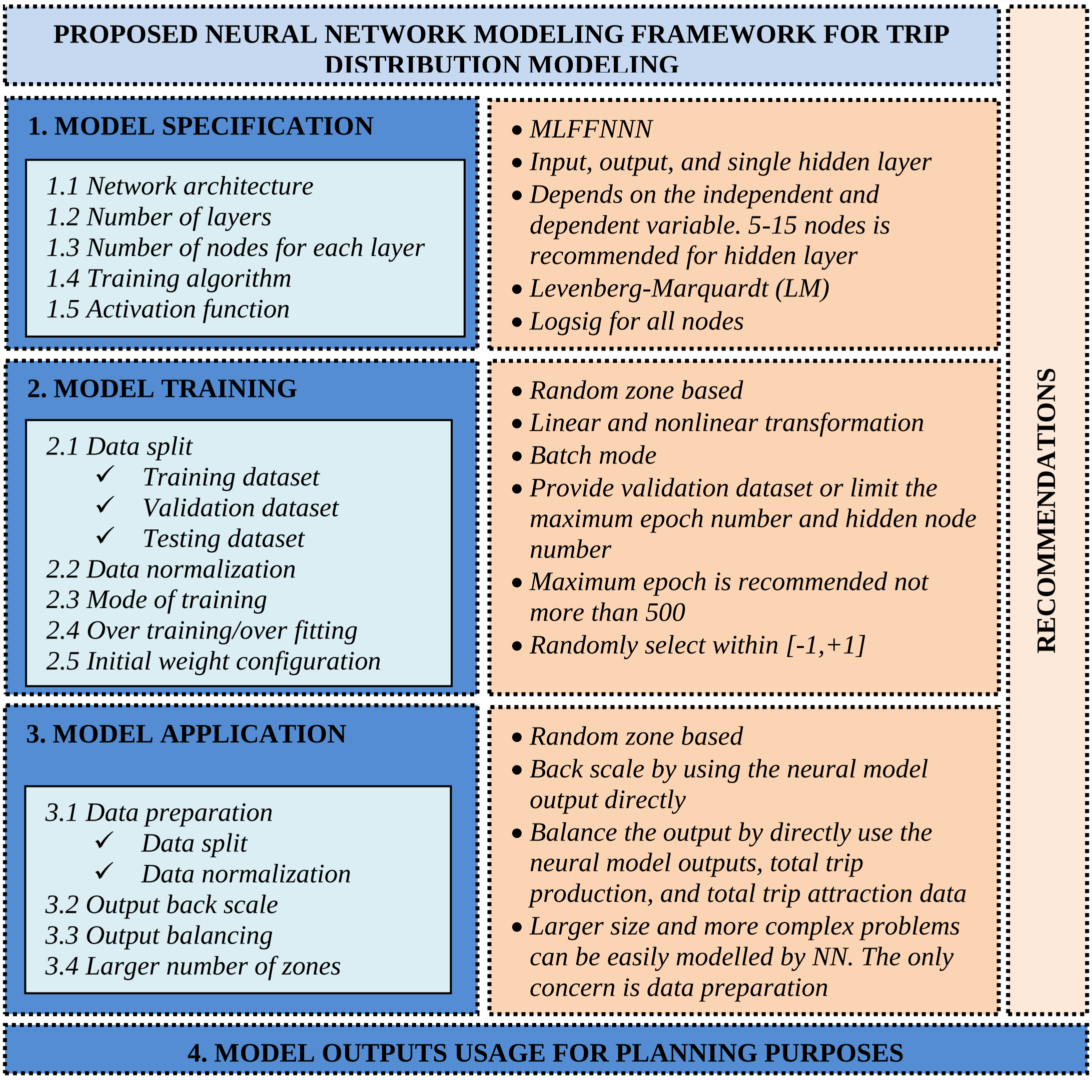

Figure 1. Modeling procedures and recommendations

Figure 2 illustrates the common multi layer feed forward neural network (MLFFNN) used in trip distribution neural models. Three input nodes are attached to this model to store the input vector, namely total trip/commodity produced in zone ‘i' and attracted to zone ‘j’ (O_i and D_j), and the distance between them (d_{ij}). One node is used in the output layer to store the target vector (t_{ij}) and the estimated trip numbers or commodity tonnage (T_{ij}). There is one hidden layer between input and output layer, where the number of nodes in this layer (h) is arbitrarily defined. This layer is connected with randomly defined connection weights, w_{ji} and w_{kj}.

To train the model, LM is recommended as the training algorithm. It has been proven to be a fast and efficient TA (Wilamowski et al., 2001). It was also found that the LM algorithm can converge in various cases where the Variable Learning Rate (VLR) and also other forms of second order approach like the conjugate gradient TAs are failed to converge (Hagan and Menhaj, 1994b). Yaldi et al. (2010a) found that LM calibrates the trip distribution matrix with a statistically significant higher accuracy than VLR and BP. In addition, this algorithm would be suitable for modeling spatial interaction as the network structure requires fewer connection weights, based on the number of nodes in the input, hidden and output layers. Meanwhile, logsig is recommended for the AF in all nodes as it transforms all of the numbers into positive value within [0, 1], and this is suitable for trip number data (Yaldi et al., 2009a). All data is normalized nonlinearly by using the logsig function, the same function as the AF.

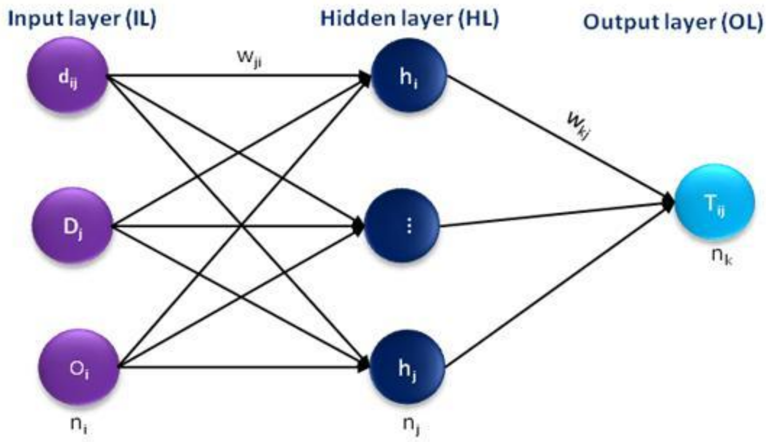

Figure 2. Common neural model architecture

3.1 Transforming Input Data Nonlinearly

To estimate trip number distribution by using a neural model, an iterative procedure is used to minimize error (difference/diff) between estimated and real trip numbers. The difference is computed as:

diff = Network output – Observed trip number

$$\Delta = T_{ij} - t_{ij} \tag{1}$$

When logsig function is used to transform the neural model outputs in both hidden and output layers, it means the results are nonlinearly normalized. Thus, the difference is computed as the gap between the nonlinearly transformed trip numbers (T_{ij}) and real trip numbers (t_{ij}), which is linearly normalized. Thus the difference becomes the gap between nonlinear model output and the linear target pattern, or:

$$\Delta = \text{(Non-Linear) } T_{ij} - \text{(Linear) } t_{ij} \tag{2}$$

$\rightarrow$ Unmatched! (Systematic error)

This is incorrect as the comparison should be based on the same basis, i.e. nonlinear output data against nonlinear target pattern. Wilamowski et al. (2001) and Hagan & Menhaj (1994b) described that the gradient of the neural models trained with LM is a function of the Jacobian transpose and the error. This will affect the computation of the gradient. It can stop the training process earlier and so influence its performance. There is a systematic mismatch in the difference computation in the above equation. Therefore, it needs to be corrected so that:

Corrected diff = (Non-Linear) T_{ij} - (Non-Linear) t_{ij} (3)
$\rightarrow$Matched!

Zhang (1998) investigated the properties of neural models used in different research projects. Some research was reported as using logistic functions in data normalization due to the type of activation function used in the output node. The normalization is undertaken prior to the training process, so that the outputs of the neural models will be in the same state as the target vectors. This requires transforming the data to either logsig or double logsig function, which can be expected to benefit the neural model performance. However, this research applied this correction for the target vector or observed trip numbers (t_{ij}) only.

3.2 Neural Model Output Balancing

There is another important issue related to the neural model for trip distribution. Mozolin et al. (2000) and Yaldi et al. (2009b) both reported that the neural model outputs for row and column totals are different from the observed values. Mozolin et al. (2000) actually balanced the neural model output before comparing the results with the doubly constrained gravity model. It indicated that the importance of the constraints to be satisfied in neural model testing performance. However, lack of effort was devoted in this essential issue. In addition, balancing output itself is inadequate. It must be preceded by the data which has to be in the square form as the doubly constrained O-D matrix.

This study proposes a balancing method for the neural model output which is applicable for both calibration and testing levels. It is somewhat similar to the Furness method (Ortuzar and Willumsen, 1994). The difference is the balancing is undertaken directly with the output generated by the neural model.

These balancing procedures are also different from the balancing procedures undertaken by Mozolin et al. (2000). The balancing is undertaken by directly using the neural model outputs. Therefore, it does not require multiplication with any normalization factors. The matrix form for calibration, validation, and testing datasets is preserved. The data is randomly selected based on the random zone number. This is different from the procedures in Mozolin et al. (2000), Yaldi et al. (2009b), and Shir-mohammadli et al. (2010) which randomly selected the data from the whole samples, so that the matrix form is unpreserved (termed as random vector basis). Thus, the random zone number method will enhance the balancing procedures for the neural model outputs (see Yaldi (2011b) for more detail on random zone method).

The disadvantage of the method used in the previous research is that it is difficult to measure whether the model outputs are able to satisfy the O_i and D_j constraints. Both studies by Mozolin et al. (2000) and Yaldi et al. (2009b) reported that the neural model's ability to satisfy both O_i and D_j constraints was inadequate.

3.3 Random Zone Data Split Method

The data for calibration and testing is divided on the basis of random zone number in order to maintain the square form of the O-D matrix and hence improve the neural model testing performance. It is unlikely the model output can automatically fulfill the O_i and D_j constraints. Thus, it will negatively influence the performance of the neural models. Therefore, an alternative data split method is proposed, that is compatible with the characteristics of the spatial interaction data and model. The data split is undertaken by randomly selecting a zone number for three blocks instead of randomly selecting the data pattern directly, and then forming the data vectors for each block. The details for this method can be found in Yaldi et al. (2011b).

The advantage of this technique is the performance of the neural model toward O_i and D_j constraints can be assessed and should contribute positively to neural model performance in forecasting the trip numbers for an unseen dataset (generalization ability).

All these three aspects (the nonlinear transformation, output balancing, and random zone based split methods) are expected to refine the performance of neural models, not only at the calibration level, but also at the testing level. Please note that the balancing commences after all of the trials are completed.

Training is conducted in batch mode, where the connection weight is updated after all of the input patterns are presented to the network. To prevent over fitting, training is recommended not more than 500 epochs.

After the training, the model is ready for forecasting trip or commodity distribution of new datasets. Given the data of the total trip or commodity produced and attracted in zone 'i ' and 'j', the neural model can predict the doubly constrained trip number or commodity tonnage (Tij) distribution. The same as doubly constrained gravity model, the neural model outputs require to be balanced. The Furness technique can be adopted (Ortuzar and Willumsen, 1994). Finally, these procedures can be easily used for larger and more complex problems of spatial distributions.

4. CASE STUDY WITH US 2007 CFS DATA

The procedural steps as illustrated by Figure 1 are rigorously followed in developing the neural model for commodity distribution in this paper. This model uses the US Commodity Flow Survey (CFS) data, collected in 2007, and extends the neural model evaluation to the testing level. For model benchmarking purposes, the neural model is compared with the gravity model, as in the previous studies. The calculation of the gravity model is conducted rigorously, especially during the calibration and balancing process. The Hyman (1969) algorithm is used to calibrate the gravity model.

4.1 The Neural Model Properties

The information regarding the properties of the neural model used to forecast commodity distribution based on the 2007 US CFS is reported in Table 1.

Table 1. Commodity neural model properties

Property	Remark
Model architecture	Multi layer feed forward neural network/MLFFNN
Number of layers	3 layers (Input, Hidden, and output layers)
Number of input nodes	3 nodes (Commodity total production, attraction, and estimated length)
Number of hidden nodes	10 nodes
Number of output nodes	1 node (estimated commodity tonnage)
Training algorithm	Levenberg-Marquardt (LM)
Activation function	Sigmoid (Logsig)
Data split	'Random zone based'
Data normalization	'Simple mix' (simple linear and logsig transformation)
Mode of training	Batch mode
Initial weight method	Closed random [-1, +1]
Maximum epoch number	500 epochs

4.2 Model Data

The data is based on 2007 US CFS. The data is available online thanks to the US Census Bureau which can be accessed on the http://www.census.gov/. There are 42 different commodity groups. Initially ten commodities were selected; however, only four commodities are used to test the applicability of the proposed framework. This is because the information on commodity tonnage is only available for a few zones and hence was considered ineffective to be modeled. Some of the data shows that the commodity flows were reported for intra zonal flows only. For the distance between each zone, it is roughly estimated and obtained from http://www.bing.com/maps/.

The commodity groups which are used in this research are reported in Tables 2 and 3. These groups have more data than other groups. The data is sparse with the number of zones reported without commodity flows (zero trip zones) reaching up to 84 per cent of the total flows. Table 2 shows the commodity code, name, and the terms used for each commodity. Meanwhile, Table 3 shows the total tonnages for each commodity in ton thousand/year.

Table 2. Commodity code and name

Commodity code	Commodity name	Remark
05	Meat, fish, seafood, and their preparations	Termed as C05
06	Milled grain products, and preparations and bakery products	Termed as C06
07	Other prepared foodstuffs and fats and oils	Termed as C07
08	Alcoholic beverage	Termed as C08

Table 3. Commodity tonnage (Ton thousand/year)

Commodity code	Calibration	Testing (Z16)	Testing (Z15)
05	15374	25273	12225
06	22574	19327	16470
07	100298	121923	70547
08	27979	37094	16516

There are 51 origin and destination geographies reported in 2007 US CFS representing the number of states in the USA, however, only 49 states or zones are used in this study. Alaska and Hawaii are excluded. This is due to the type of mode used to transport the commodities considered in this model. The model only considers those commodities which are transported by trucks. The truck is one kind of mode used to transport the commodity as listed in the website. Then, the number of zones for training is set to be 18 zones.

The commodity neural model in this application does not allocate data for validation. Instead, it conducts the testing twice. The number of zones for the first and the second tests are 16 and 15 respectively. They are termed as Z16 and Z15 consecutively. The member of training and testing group is randomly selected from the 49 states in USA. Each model is trained for 30 trials and hence there are 2x4x30 = 240 trials.

5. MODEL OUTPUT AND DISCUSSION

5.1 Calibration Performance: RMSE and Goodness-of-Fit

The RMSE and correlation coefficient (r) for calibration are reported in Tables 4 and 5. Each table reports the performance for gravity model (GM) and neural model (NM) respectively. Table 5 also shows the difference of RMSE between neural and gravity models in terms of percentage (see the percentage inside the bracket).

Table 4. Gravity model performance

Commodity	**Calibration**		**Testing_1 (Z16)**		**Testing_2 (Z15)**	
	RMSE	r	RMSE	R	RMSE	r
C05	50	0.992	115	0.969	51	0.965
C06	40	0.998	79	0.978	47	0.985
C07	202	0.997	419	0.982	138	0.993
C08	73	0.997	44	0.998	28	0.997

Table 5. Neural model performance

Commodity	**Calibration**		**Testing_1 (Z16)**		**Testing_2 (Z15)**	
	RMSE	**r**	**RMSE**	**R**	**RMSE**	**r**
C05	13 (-74)	0.999	82 (-29)	0.989	33 (-35)	0.985
C06	18 (-55)	0.999	76 (-4)	0.980	59 (25)	0.980
C07	97 (-52)	0.999	256 (-39)	0.993	135 (-2)	0.995
C08	18 (-75)	0.999	38 (-14)	0.999	24 (-14)	0.997

Both models have all of the correlation coefficients above 0.9 indicating that both models have considerably high mapping ability. However, there are huge differences in the RMSE. It can be seen that all of the neural model RMSE at calibration level are considerably lower than the gravity model. The differences vary from 52 to 75 per cent. This percentage supports the finding reported by Black (1995).

Black (1995) reported that the neural model can have the RMSE for calibration up to 50 per cent lower than the gravity model. However, in his case the neural model was trained with BP and the maximum epoch was 150000. This number is significantly higher than the one used to train the model in this research, which is not more than 500 epochs. Therefore, the advantages of the proposed framework depicted by Figure 1 is not only improved neural model performance at calibration level, but also reduced number of epoch, hence shortening the training time, and avoiding over-fitting.

5.1.1 Calibration performance: t-test

In order to test the significance of the difference between average RMSE of neural and gravity models, a t-test was conducted. There was no significance test study reported prior to 2009. The initial use of the t-test was reported at calibration level only by Yaldi et al. (2009a). Thus, the two-tailed t-test in this research was conducted for degrees of freedom 29, and level of confidence 95 per cent. The results are reported in Table 6. Actually, the results reported in Table 5 strongly demonstrate that neural model outperforms the gravity model. The test results as reported in Table 6 only confirm these results – but indicate their strong statistical significance. All of the neural models have significantly lower RMSE than the gravity model.

Table 6. T-test for average RMSE (Calibration, 30 trials)

Commodity	**t-calculation**	**Remark**	
C05	-41.915	Critical t	2.04
C06	-84.565	Degree of freedom	29
C07	-77.492	Level of confidence	95%
C08	-75.217	Two-tailed	

5.1.2 Calibration performance: Linear regression relationship

The linear regression relationship between the observed and estimated commodity tonnage is illustrated by Figures 3-6. These figures suggest that the gravity model tends to overestimate the commodity flows, except for commodity CO7. The neural model is shown to have better goodness-of-fit than the gravity model in all cases. The regression equation on the left top side of the figures represents the neural model, and the other one represents the gravity model.

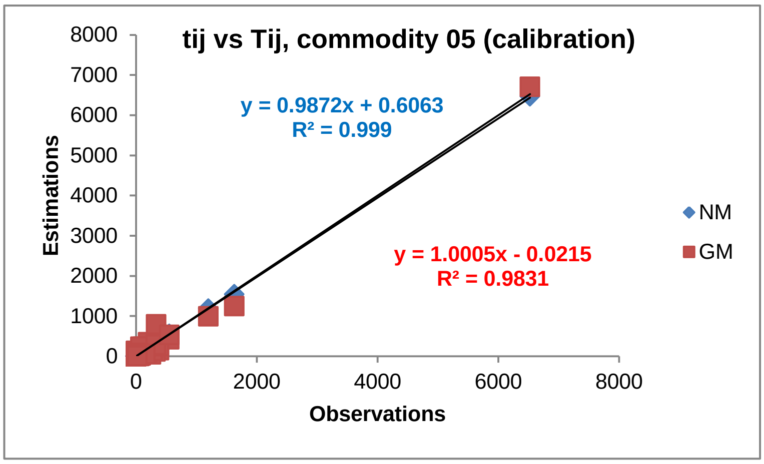

Figure 3. Observed and estimated commodity tonnage at calibration level (C05)

This is one of the main advantages of the NN, namely its ability to learn from the data pattern without having prior information of relationship between the independent and dependent variables. The gravity model forecasts the movements of either passengers or commodities with specific decay functions. It assumes that zone pairs with short distances or trip lengths have more movements or flows than those with long distance or trip length separations. This is not always true as the distribution of people and commodities can also be determined by other factors such as socioeconomic and demographic factors. Investigating those factors and analyzing them are time consuming and costly, and in the case of the gravity model have led to the occasional use of empirical ‘K-factors’, which are difficult to justify in theory.

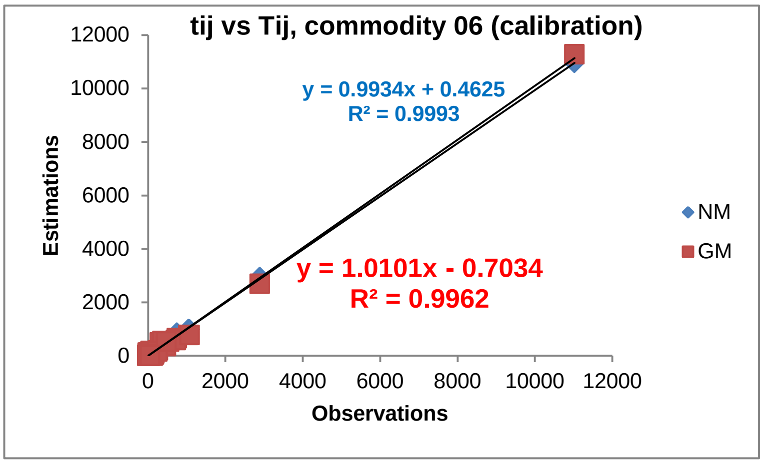

Figure 4. Observed and estimated commodity tonnage at calibration level (C06)

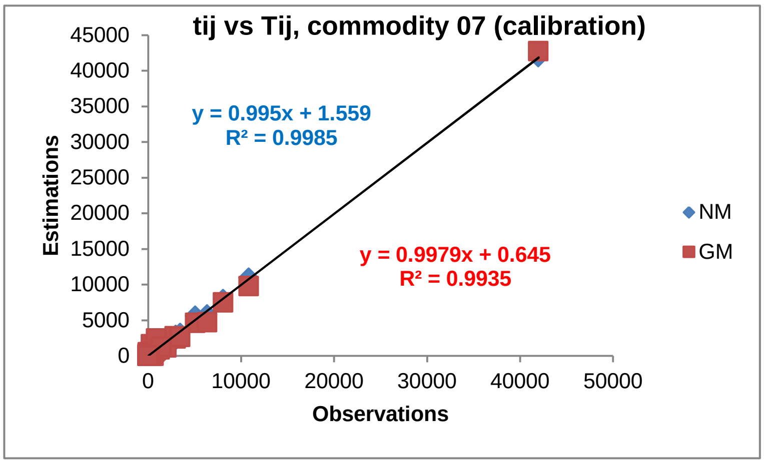

Figure 5. Observed and estimated commodity tonnage at calibration level (C07)

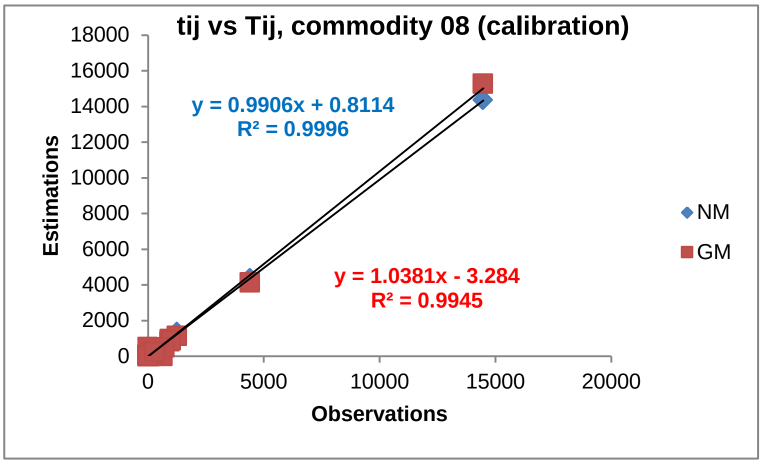

Figure 6. Observed and estimated commodity tonnage at calibration level (C08)

On the other hand, a neural model has the freedom to forecast person movement or commodity flow. It is not bounded by any specific functions, except the functions used in the training algorithm and the activation function. Those functions are used so that the training can be conducted. They do not represent any relationship between the independent and dependent variables. The neural model distributes the flow based on the pattern captured during the training and then uses it to generalize trip patterns for new sets of independent variables. Then, the results are improved by applying external processes such as the balancing procedures as described before. This is neither time consuming nor costly. This is then another advantage of the NN, especially for use in countries and regions where limited budgets are available for transport data collection and modeling.

5.2 Testing Performance: RMSE and Goodness-of-Fit

The neural model performance at the testing level is also reported in Table 5. The results suggest that almost all of neural models have a lower RMSE than the gravity model, ranging from 2 to 39 per cent. It also can be seen in the same table that only one gravity model has a lower RMSE than the neural model (by 25 per cent, for commodity CO6 in Testing 2 (Z15)). According to Tables 3 and 4, a majority of the neural model have higher generalization ability than the gravity model as indicated by the observation that more than 60 per cent of the neural models have higher correlation coefficients than gravity models. These results indicate that NN can be used not only for passenger trip distribution, but also to forecast commodity flows. This applies for both calibration and testing or generalization levels. The usage of the proposed framework enhances the reliability of this ‘black box’ approach for travel demand modeling as indicated by the results in this research.

5.2.1 Testing performance: t-test

Like the calibration, t-tests are also conducted to test the significance of differences between neural and gravity models average RMSE. The results are reported in Tables 7 and 8 for Z16 and Z15 consecutively. The entire commodity neural models for Z16 have a significantly lower RMSE than the gravity model, except for the commodity C06. Although the neural model has a lower RMSE compared to the gravity model, the difference is statistically insignificant.

Table 7. T-test for average RMSE for Z16 (Testing, 30 trials)

Commodity	t-calculation	Remark		
C05	-16.856	**Significant**	Critical t	2.04
C06	-1.482	Insignificant	Degree of freedom	29
C07	-23.987	**Significant**	Level of confidence	95%
C08	-3.156	**Significant**	Two-tailed	

Table 8. T-test for average RMSE for Z15 (Testing, 30 trials)

Commodity	t-calculation	Remark		
C05	-24.198	**Significant**	Critical t	2.04
C06	21.223	**Significant**	Degree of freedom	29
C07	-1.121	Insignificant	Level of confidence	95%
C08	-2.519	**Significant**	Two-tailed	

Meanwhile, the commodity neural model for Z15 has slightly different results to those for Z16. According to Table 8, the gravity model has a significantly lower RMSE than the neural model for commodity C06 which is also indicated by the results reported in Table 5. Both neural and gravity models have insignificant difference in average RMSE for commodity C07 although the neural model has a lower RMSE. Then, the remaining commodity neural model has significantly lower average RMSE than the gravity model. Again, this result confirms the finding in Table 5. It confirms that the neural model not only significantly outperforms the gravity model at calibration level, but also at the testing level. It demonstrates that a neural model which is properly prepared and trained will be able to generalize with significantly lower deviation between the model outputs and the observed values.

5.2.2 Testing performance: Linear regression relationship

The linear regression between observed and estimated commodity flow tonnages are represented by Figures 7&8 for Z16, while for Z15 represented by Figures 9&10. These figures illustrates the regression for some commodity types only. As described before, a majority of the neural models have better goodness-of-fit than the gravity model (for 63 per cent). When the gravity model outperforms the neural model, the gaps are relatively low compared to when the neural model outperforms the gravity model where the gaps can reach as much as 39 per cent.

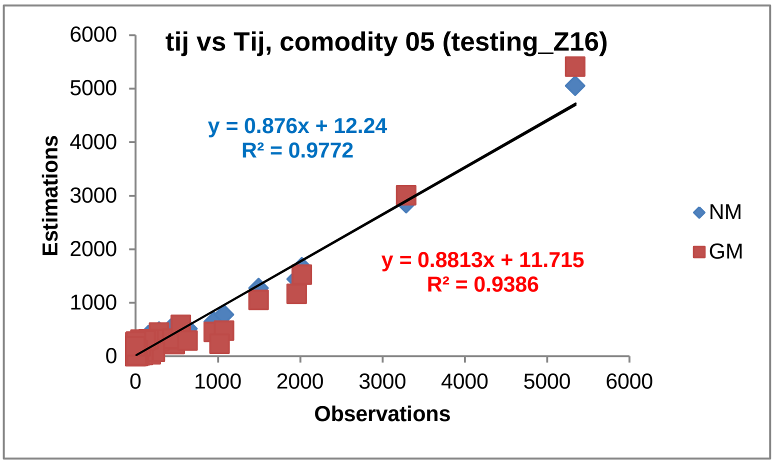

Figure 7. Observed and estimated commodity tonnage at testing level (C05_Z16)

It can be seen from all of those figures that the neural and gravity models have almost identical relationship between observation and estimation, except that some points belong to the gravity model are located further away than the neural model. Thus, the neural model has relatively higher goodness-of-fit than the gravity model.

Both models have good estimations, fitting the observed values well, with correlation coefficients above 0.9. The intercept is relatively low, where the highest one belongs to commodity code C07. This commodity has the highest total cumulative tonnage. However, the gravity model has the highest intercept for commodity C07. According to Table 5, the gravity model for this commodity has a RMSE 39 per cent higher than the neural model.

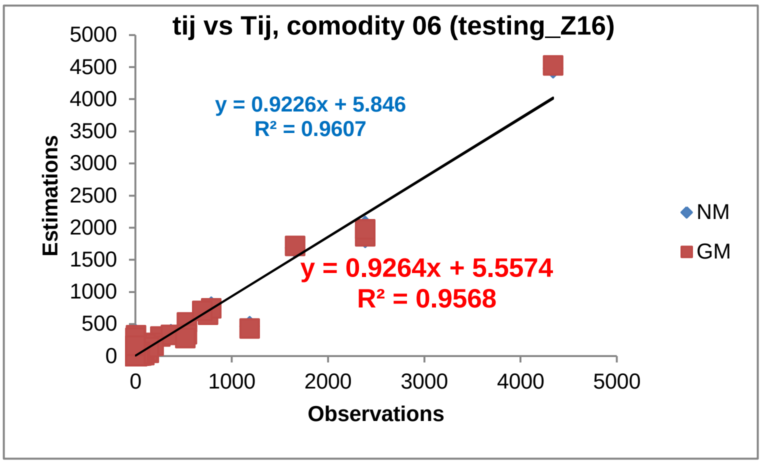

Figure 8. Observed and estimated commodity tonnage at testing level (C06_Z16)

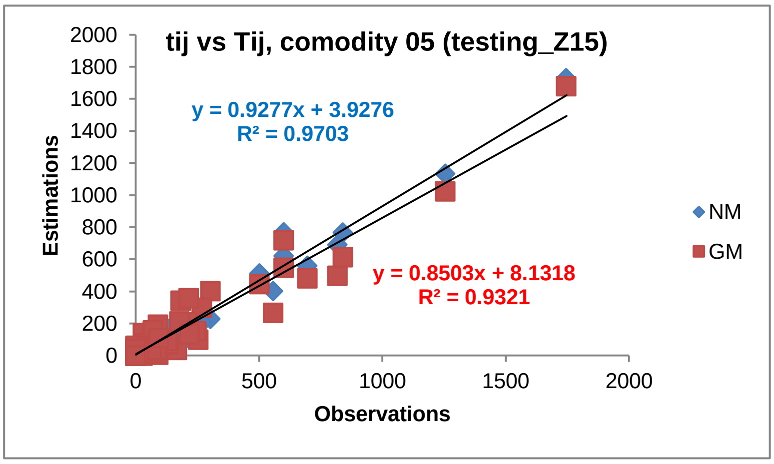

Figure 9. Observed and estimated commodity tonnage at testing level (C05_Z15)

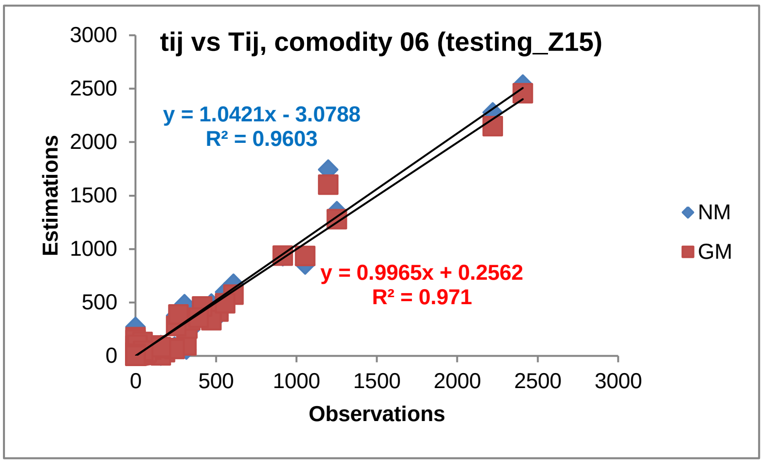

Figure 10. Observed and estimated commodity tonnage at testing level (C06_Z15)

5.2.3 Testing performance: Zero flow estimation

There are zones reported to have no commodity flows in each commodity O-D matrix used in this research. Thus, the evaluation of both neural and gravity models also covers the estimation for those zones which have no commodity flows according to the observed data. The results are reported in Tables 9-12, and also illustrated by Figures 11-12.

Table 9. Percentage of underestimate and overestimate (Z16)

Commodity code	Zero flow percentage (%)	Underestimate percentage (%)		Overestimate percentage (%)	
		NM	GM	NM	GM
C05	52	24	23	25	25
C06	66	23	24	11	9
C07	51	29	27	20	22
C08	77	15	18	8	6

Table 10. Percentage of underestimate and overestimate (Z15)

Commodity code	Zero flow percentage (%)	Underestimate percentage (%)		Overestimate percentage (%)	
		NM	GM	NM	GM
C05	58	21	20	21	21
C06	67	24	22	9	12
C07	48	28	35	24	17
C08	84	12	11	4	4

Table 11. Average statistic for Z16

Commodity code	Average zero flow estimation (T)		Average underestimate (%)		Average overestimate (%)	
	NM	GM	NM	GM	NM	GM
C05	13	17	-43	-49	599	771
C06	13	12	-48	-57	272	297
C07	52	66	-40	-67	336	505
C08	8	6	-42	-74	414	69

Table 12. Average statistic for Z15

Commodity code	Average zero flow estimation (T)		Average underestimate (%)		Average overestimate (%)	
	NM	GM	NM	GM	NM	GM
C05	6	7	-39	-41	177	199
C06	6	7	-59	-51	122	150
C07	23	17	-42	-55	155	177
C08	3	1	-43	-61	49	46

It can be seen in Tables 9 and 10 that the percentages of zones which have no commodity flows range from 48 to 84 per cent. This percentage is crucial, as both neural and gravity models forecast these zones to have some positive commodity flow distribution. In general, the neural and gravity models forecast the zero flow zones along the same trend. Both neural and gravity models estimate the zero trip zones as having positive commodity flows, ranging from 1 to 66 tons (see Tables 11 and 12). This wide range is because each commodity has a different total tonnage. The highest total tonnage belongs to commodity C07, which is about ten times higher than the lowest one (which is commodity C05). See Table 3 for the tonnage information for each commodity.

5.2.4 Testing performance: Underestimate and overestimate distribution

The evaluation of neural and gravity models is expanded by analyzing the distribution of estimated commodity tonnages for two different classifications, something that has not been done in previous studies. This is important as the trend of both neural and gravity models can be investigated, for either underestimation or overestimation. Then, the modeled distribution may be classified as either underestimate or overestimate.

The percentage of zones having underestimated tonnage tends to be the same as the overestimated percentage for both neural and gravity models (see Tables 9 and 10). However, when the zero flow is included, there tends to be more overestimation than underestimation (because the models will always generate a positive value for zero flow zones). See Tables 9 and 10 for the percentages of O-D pairs with zero commodity flows. The average tonnage for underestimation ranges from 39 to 59 and from 41 to 74 per cent below the real tonnage for neural and gravity models respectively as can be seen in Tables 11 and 12. This means the gravity model has higher gaps between observed and estimated commodity tonnages for the zones forecast as underestimate, as can be seen in Figures 11 and 12.

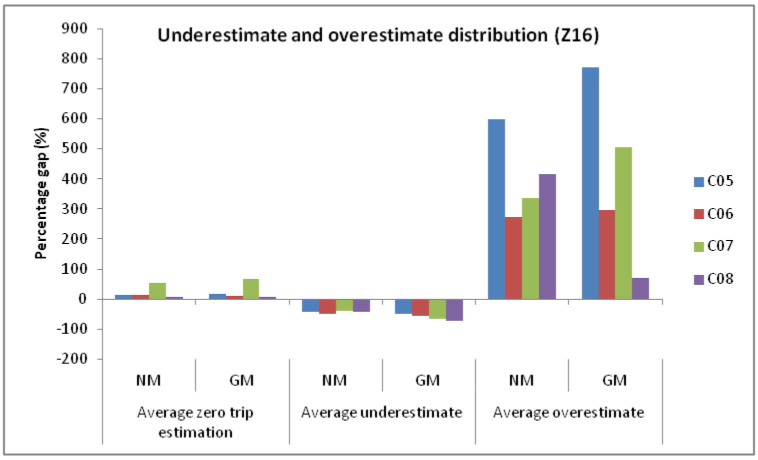

Figure 11. Underestimate and overestimate distributions (Z16)

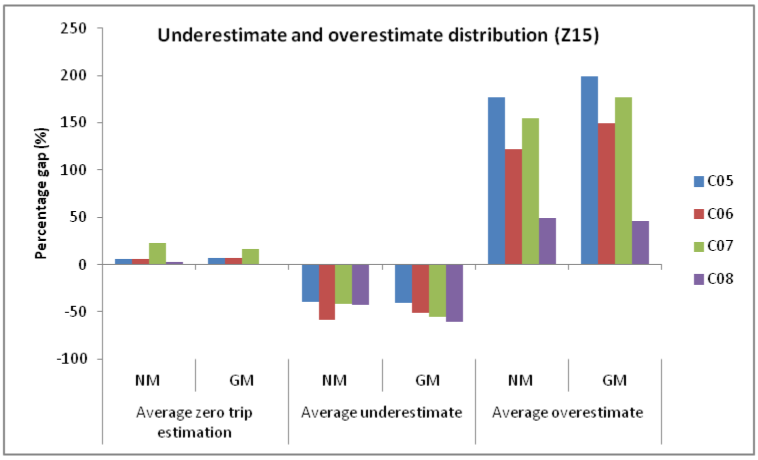

Figure 12. Underestimate and overestimate distributions (Z15)

Meanwhile, the average percentage for overestimation ranges from 49 to 599 per cent and from 46 to 771 per cent above the real tonnage for neural and gravity models respectively. Thus the gravity model has higher gaps between observed and estimated commodity tonnage for the zones forecast as overestimate. In this regard, the neural model forecasts the commodity tonnages closer to the observed values, for both categories of underestimate and overestimate, and neural model estimates theses distributions without using any decay function as in the gravity model.

5.2.5 Calibration and testing performance: Outlier removal

Figures 3-10 show the linear regression among the neural and gravity models' outputs and the observed values. All of them show both models' outputs fit the observed values. However, it can be seen that some commodity flows are predicted much bigger or smaller than the observed ones as indicated by their locations which are far from majority of the points in those figures (outliers). These points will have a strong influence on the regression analysis and the computed R^2 values. Thus, the points which are considered as 'outliers' are removed from the figures, to remove this influence. The results are represented in Figures 13-15 for calibration, testing (Z16), and testing (Z15) respectively. There are only three figures representing the 'outlier removal', however, the details of linear regression for all of the models are reported in Tables 13-15.

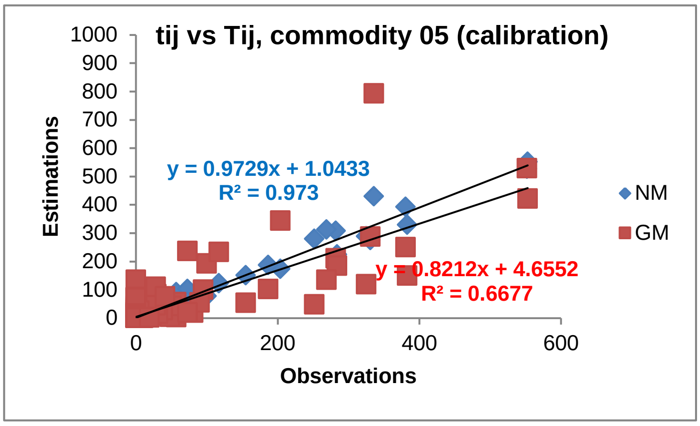

Figure 13. Observed and estimated commodity tonnage without 'outlier' (Calibration)

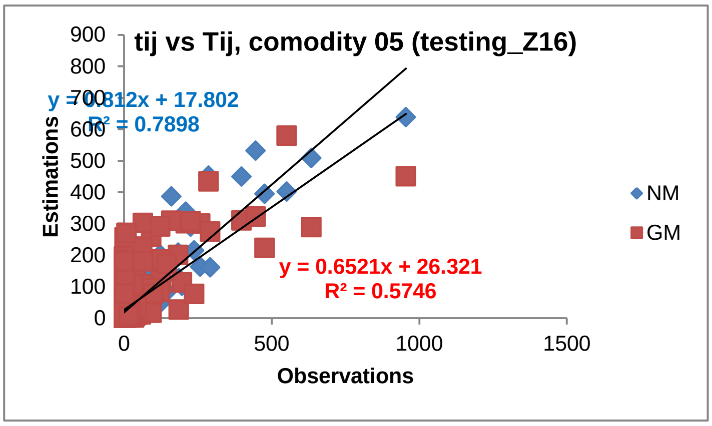

Figure 14. Observed and estimated commodity tonnage without 'outlier' (Testing_Z16)

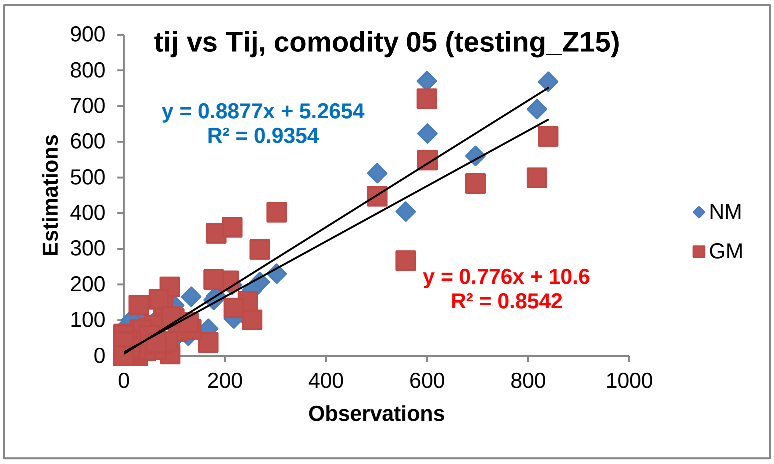

Figure 15. Observed and estimated commodity tonnage without 'outlier' (Testing_Z15)

It can be seen in the figures that there seems to be more points compared to before the outliers are removed – this results from the change of scale for the scatter plots. These figures also show different functions of linear regression and coefficient of determinations as a result of the outlier removal. In general, the outlier removal reduces the regression slope and increases its intercept constant.

There is no slope above value one. All of Z16 commodities (C05-C08) have a higher slope than the Z15 commodities. It is due to the Z16 commodities have a higher tonnage than for Z15 as seen in Table 3. The highest intercept constant belongs to commodity C07 which has the highest tonnage of commodity among others. It also can be seen that C07 for Z16 has a higher slope than the C07 for Z15. This is because the C07 data for Z16 has a higher tonnage. Then, the coefficient of determination decreases.

Tables 13-15 show the decrease for each coefficient of determination. A majority of the coefficient of determinations are nearly or above 0.8 for both neural and gravity models. All of coefficients of determination were above 0.9 before the outliers were removed. The highest decrease belongs to commodity C07 for Z16. It also has the poorest goodness-of-fit. Its coefficient of determination is 0.406 and 0.360 for neural and gravity models respectively. However, the neural model continues to have higher goodness-of-fit than the gravity model for both calibration and testing levels, as seen in Figures 13-15 and Tables 13-14.

Table 13. Linear regression slope, intercept, and R^2 (Calibration-no outliers)

Commodity code	**Neural model**			**Gravity model**		
	Slope	**Intercept**	**R^2**	**Slope**	**Intercept**	**R^2**
C05	0.973	1.043	0.973 (-2.6)	0.821	4.655	0.668 (-31.8)
C06	0.991	0.792	0.967 (-3.2)	0.906	3.573	0.898 (-9.9)
C07	0.899	6.829	0.830 (-16.9)	0.797	19.288	0.432 (-56.5)
C08	0.914	1.754	0.982 (-1.7)	0.750	4.971	0.766 (-23.0)

Table 14. Linear regression slope, intercept, and R^2 (Testing Z16-no outliers)

Commodity code	**Neural model**			**Gravity model**		
	Slope	**Intercept**	**R^2**	**Slope**	**Intercept**	**R^2**
C05	0.812	17.802	0.790 (-19.2)	0.652	26.321	0.575 (-38.8)
C06	0.852	9.657	0.836 (-13.0)	0.845	9.756	0.823 (-14.0)
C07	0.742	41.777	0.406 (-58.9)	0.967	52.726	0.360 (-67.2)
C08	0.893	5.880	0.936 (-6.2)	0.997	2.070	0.906 (-9.1)

Table 15. Linear regression slope, intercept, and R^2 (Testing Z15-no outliers)

Commodity code	**Neural model**			**Gravity model**		
	Slope	**Intercept**	R^2	**Slope**	**Intercept**	R^2
C05	0.888	5.265	0.935 (-3.6)	0.776	10.600	0.854 (-8.4)
C06	0.888	0.934	0.816 (-15.1)	0.828	4.517	0.860 (-11.4)
C07	0.913	-0.151	0.944 (-4.7)	0.908	13.868	0.924 (-6.3)
C08	0.953	1.603	0.980 (-1.5)	0.968	-1.176	0.941 (-2.0)

6. SUMMARY

The important findings from the discussion in this paper may be summarized as below:

1. The usage of the empirical approaches integrated in the modeling framework proposed by Figure 1 has resulted in all of the neural models having lower RMSE and higher goodness-of-fit than the gravity model for calibration level. The neural model has a significantly lower average RMSE than the gravity model.
2. About 63 per cent of the neural model has significantly higher generalization performance than the gravity model in term of RMSE and correlation coefficient, when applied to an independent dataset (the 2007 US commodity flow data).
3. Both neural and gravity models have the same trend for percentage of underestimated and overestimated outputs.
4. When the outputs are underestimated, the neural model has lower average gaps between its outputs and the real ones compared to the gravity model.
5. When the outputs are overestimated, the neural model has lower average gaps between its outputs and the real ones compared to the gravity model.
6. The neural model does not require any function representing the relationship between independent and dependent variables. Therefore, it forecasts the distribution of person movement or commodity flow based on the pattern in the calibration data. It does not require any additional data. The usage of additional data is believed capable of further enhancing the neural model generalization performance. Thus, neural models become an effective and efficient reliable modeling tool for forecasting person travel and commodity flow distribution.
7. The results in this research also support the finding by Tillema et al. (2006) which claimed that neural models outperform the gravity model for scarce data. This research extends it to the testing level.
8. The gravity model relies on its decay function in order to calibrate and forecast the person movement and commodity flow distribution. There are likely to be other socio-economic and socio-demographic variables which determine the movement of people and commodities. However, identifying the relevant O-D pairs and collecting relevant information about them is time consuming and costly

7. REFERENCE

BLACK, W. R. 1995. Spatial interaction modeling using artificial neural networks. *Journal of Transport Geography,* 3**,** 159-166.

CANTARELLA, G. E. & DE LUCA, S. 2005. Multilayer feedforward networks for transportation mode choice analysis: An analysis and a comparison with random utility models. *Transportation Research Part C: Emerging Technologies,* 13**,** 121-155.

CELIK, H. M. 2004a. Forecasting interregional commodity flows using artificial neural networks: an evaluation. *Transportation Planning and Technology,* 27**,** 449-467.
CELIK, H. M. 2004b. Modeling freight distribution using artificial neural networks. *Journal of Transport Geography,* 12**,** 141-148.
CHIN, S. M., HWANG, H. L. & MIAOU, S. P. Year. Transportation demand forecasting with a computer-simulated neural network model. *In:* International Conference on Artificial Intelligence Applications in Transportation Engineering, 1992 San Buenaventure, CA.
DANTAS, A., YAMAMOTO, K., LAMAR, M. V. & YAMASHITA, Y. Year. Neural network for travel demand forecast using GIS and remote sensing. *In:* Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on, 2000 2000. 435-440 vol.4.
DOUGHERTY, M. 1995. A review of neural networks applied to transport. *Transportation Research Part C: Emerging Technologies,* 3**,** 247-260.
HAGAN, M. T. & MENHAJ, M. B. 1994a. Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks,* 5**,** 989-993.
HAGAN, M. T. & MENHAJ, M. B. 1994b. Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks* 5**,** 989-993.
HYMAN, G. M. 1969. The Calibration of Trip Distribution Models. *Environment and Planning,* 1**,** 105-112.
MOZOLIN, M., THILL, J. C. & LYNN, U. E. 2000. Trip distribution forecasting with multilayer perceptron neural networks: A critical evaluation. *Transportation Research Part B: Methodological,* 34**,** 53-73.
ORTUZAR, J. D. D. & WILLUMSEN, L. G. 1994. *Modelling transport,* West Sussex, England, John Wiley & Sons Ltd.
SHIR-MOHAMMADLI, M., SHETAB-BUSHEHRI, S. N., POORZAHEDY, H. & HEJAZI, S. R. 2010. A comparative study of a hybrid Logit–Fratar and neural: network models for trip distribution: case of the city of Isfahan. *Journal of Advanced Transportation,* 45**,** 80-93.
SHMUELI, D. 1998. Applications of neural networks in transportation planning. *Progress in Planning,* 50**,** 141-204.
SHMUELI, D., SALOMON, I. & SHEFER, D. 1996. Neural network analysis of travel behavior: Evaluating tools for prediction. *Transportation Research Part C: Emerging Technologies,* 4**,** 151-166.
TILLEMA, F., VAN ZUILEKOM, K. M. & VAN MAARSEVEEN, M. F. A. M. 2006. Comparison of Neural Networks and Gravity Models in Trip Distribution. *Computer-Aided Civil and Infrastructure Engineering,* 21**,** 104-119.
WILAMOWSKI, B. M., IPLIKCI, S., KAYNAK, O. & EFE, M. Ö. 2001. An Algorithm for Fast Convergence in Training Neural Networks. *IEEE,* 3**,** 1778-1782.
YALDI, G. 2012. *Developing a neural network modeling framework for a trip distribution model.* Doctor of Philosophy, University of South Australia.
YALDI, G., TAYLOR, M. A. P. & YUE, W. L. 2009a. Improving Artificial Neural Network Performance in Calibrating Doubly-Constrained Work Trip Distribution by Using a Simple Data Normalization and Linear Activation Function. *The 32nd Australasian Transportation Research Forum.* Auckland, New Zealand. Available at www.patrec.org/atrf.aspx.

YALDI, G., TAYLOR, M. A. P. & YUE, W. L. 2009b. Using Artificial Neural Network in Passenger Trip Distribution Modelling (A Case Study in Padang, Indonesia). *Journal of Eastern Asia Society for Transportation Studies,* 8**,** 682-693.

YALDI, G., TAYLOR, M. A. P. & YUE, W. L. 2010a. Refining the Performance Neural Network Approach in Modelling Work Trip Distribution by Using Lavenberg-Marquardt Algorithm. *Journal of the Society for Transportation and Traffic Studies (JSTS),* 1**,** 18-29.

YALDI, G., TAYLOR, M. A. P. & YUE, W. L. 2010b. A Robust Neural Network Approach for Commodity Flow Distribution Modelling Tool. *The 3rd International Conference in Transportation and Logistics.* Fukuoka, Japan.

YALDI, G., TAYLOR, M. A. P. & YUE, W. L. 2011a. A Consistent Neural Network Model for Doubly Constrained Spatial Movement Estimation. *Eastern Asia Society for Transportation Studies,* 9**,** 512-525.

YALDI, G., TAYLOR, M. A. P. & YUE, W. L. 2011b. Forecasting origin-destination matrices by using neural network approach: A comparison of testing performance between back propagation, variable learning rate and levenberg-marquardt algorithms. *The 34th Australasian Transportation Research Forum.* Adelaide, South Australia.

ZHANG, G., PATUWO, B. E. & HU, M. Y. 1998. Forecasting with artificial neural networks:: The state of the art. *International Journal of Forecasting,* 14**,** 35-62.