การประมาณตารางจุดต้นทางปลายทางการขนส่งสินค้าโดยใช้ข้อมูลร่วมจากการสำรวจการเคลื่อนย้ายสินค้าและข้อมูลสัมภาษณ์ริมทาง

นายวิรัช หิรัญ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรดุษฎีบัณฑิต

สาขาวิชาวิศวกรรมโยธา ภาควิชาวิศวกรรมโยธา

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2552

ESTIMATING FREIGHT ORIGIN DESTINATION MATRICES USING COMBINED COMMODITY FLOW SURVEY AND ROADSIDE SURVEY DATA

Mr. Wirach Hirun

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Civil Engineering
Department of Civil Engineering
Faculty of Engineering
Chulalongkorn University
Academic year 2009

Thesis Title	ESTIMATING FREIGHT ORIGIN DESTINATION MATRICES USING COMBINED COMMODITY FLOW SURVEY AND ROADSIDE SURVEY DATA
By	Mr. Wirach Hirun
Field of Study	Civil Engineering
Thesis Advisor	Associate Professor Sompong Sirisoponsilp, Ph.D.

Accepted by the Faculty of Engineering, Chulalongkorn University in Partial Fulfillment of the Requirements for the Doctoral Degree

.... Dean of the Faculty of Engineering

(Associate Professor Boonsom Lerdhirunwong, Dr. Ing)

THESIS COMMITTEE

.... Chairman

(Assistant Professor Manoj Lohatepanont, Sc.D.)

.... Thesis Advisor

(Associate Professor Sompong Sirisoponsilp, Ph.D.)

.... Examiner

(Assistant Professor Kasem Choocharukul, Ph.D.)

.... External Examiner

(Associate Professor Pongsak Suriyavanagul, Ph.D.)

.... External Examiner

(Assistant Professor Terdsak Rongwiriyapanich, Ph.D.)

วิรัช หิรัญ : การประมาณตารางจุดต้นทางปลายทางการขนส่งสินค้าโดยใช้ข้อมูลร่วมจากการสำรวจการเคลื่อนย้ายสินค้าและข้อมูลสัมภาษณ์ริมทาง. (ESTIMATING FREIGHT ORIGIN DESTINATION MATRICES USING COMBINED COMMODITY FLOW SURVEY AND ROADSIDE SURVEY DATA) อ.ที่ปรึกษาวิทยานิพนธ์หลัก : รศ.ดร.สมพงษ์ ศิริโสภณศิลป์, 109 หน้า.

จากข้อจำกัดของข้อมูลที่มีอยู่ในปัจจุบัน ส่งผลให้การศึกษาด้านการขนส่งสินค้าในประเทศไทยมีไม่มากนัก เพื่อแก้ปัญหาดังกล่าวรัฐบาลจึงได้ดำเนินโครงการสำรวจการเคลื่อนย้ายสินค้าและสำรวจการขนส่งสินค้าทางถนนด้วยรถบรรทุกโดยการสัมภาษณ์ริมทาง เพื่อรวบรวมข้อมูลการขนส่งสินค้าทางถนนในประเทศไทย แต่เนื่องจากโครงการดังกล่าวเป็นโครงการใหม่ที่ยังไม่เคยริเริ่มในประเทศไทยรวมทั้งข้อจำกัดด้านงบประมาณ จึงส่งผลให้ข้อมูลที่รวบรวมได้ยังไม่สมบูรณ์และจำเป็นต้องได้รับการปรับแก้

สืบเนื่องจากความต้องการพัฒนาฐานข้อมูลต้นทางปลายทางการขนส่งสินค้าจากข้อมูลที่มีอยู่ในประเทศไทย ได้แก่ข้อมูลสำรวจการเคลื่อนย้ายสินค้าและข้อมูลการสัมภาษณ์ริมทาง ดังนั้นการวิจัยนี้จึงได้กำหนดวัตถุประสงค์ไว้ 2 ประเด็นคือ การพัฒนาวิธีการรวมข้อมูลการเคลื่อนย้ายสินค้าและข้อมูลการสัมภาษณ์ริมทาง และการพัฒนาวิธีการเพื่อแก้ปัญหาเซลว่างในตารางต้นทางปลายทางการขนส่งโดยใช้วิธีการอะแด๊ปทีฟนิวโรฟัซซี่อินเฟอร์เร็นซีสเต็ม

งานวิจัยนี้ได้พัฒนาวิธีการรวมข้อมูล 2 วิธีโดยใช้ข้อดีของข้อมูลทั้งสองประเภท โดยวิธีที่ 1 ได้แก่วิธีการแจกแจงการเดินทางตามระยะทาง ซึ่งใช้การปรับการกระจายการเดินทางตามระยะทางของข้อมูลการเคลื่อนย้ายสินค้าเพื่อปรับปริมาณการขนส่งให้เท่ากับปริมาณการขนส่งรวมของข้อมูลการสัมภาษณ์ริมทาง สำหรับวิธีที่ 2ได้แก่วิธีการแบบจำลองแรงโน้มถ่วง ซึ่งใช้ฟังก์ชั่นความยากลำบากในการเดินทางของข้อมูลการเคลื่อนย้ายสินค้า ในการปรับข้อมูลการสัมภาษณ์ริมทาง การทดสอบวิธีการทั้งสองใช้การทดสอบกับสองชุดข้อมูลที่สำรวจต่างเวลาและหน่วยงานที่สำรวจ ผลการทดสอบแสดงให้เห็นว่าวิธีการทั้งสองให้ค่าปรับแก้ที่ใกล้เคียงกันในทั้งสองชุดข้อมูล ดังนั้นจึงสามารถสรุปได้ว่าวิธีการที่พัฒนาขึ้นสามารถนำไปประยุกต์ใช้ในการปรับแก้ข้อมูลได้

สำหรับการพัฒนาวิธีการเพื่อแก้ปัญหาเซลว่างในตารางต้นทางปลายทางการขนส่ง งานวิจัยนี้ได้พัฒนาวิธีการใช้ตัวแปลงบ๊อกค๊อกและอะแด๊ปทีฟนิวโรฟัซซี่อินเฟอร์เร็นซีสเต็ม เพื่อเปรียบเทียบกับวิธีแบบจำลองแรงโน้มถ่วง โดยได้ทำการทดลองโดยใช้ตัวแปรมาตรฐานของวิธีแบบจำลองแรงโน้มถ่วงและตัวแปรข้อมูลเศรษฐกิจและสังคมของพื้นที่ ผลการวิจัยแสดงให้เห็นว่าวิธีการอะแด๊ปทีฟนิวโรฟัซซี่อินเฟอร์เร็นซีสเต็ม ให้ผลการวิเคราะห์ที่ดีกว่าวิธีการใช้ตัวแปลงบ๊อกค๊อกและวิธีแบบจำลองแรงโน้มถ่วง ซึ่งเป็นการยืนยันประสิทธิภาพของวิธีการอะแด๊ปทีฟนิวโรฟัซซี่อินเฟอร์เร็นซีสเต็มในการจำลองระบบที่ซับซ้อนและสามารถนำมาใช้ในการสร้างแบบจำลองการกระจายการเดินทาง

ภาควิชา วิศวกรรมโยธา..............................	ลายมือชื่อนิสิต..
สาขาวิชา วิศวกรรมโยธา..............................	ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก ....
ปีการศึกษา 2552....................................	

487 18449 21: MAJOR CIVIL ENGINEERING

KEYWORDS: FREIGHT / ORIGIN DESTINATION / ANFIS / CFS

WIRACH HIRUN: ESTIMATING FREIGHT ORIGIN DESTINATION MATRICES USING COMBINED COMMODITY FLOW SURVEY AND ROADSIDE SURVEY DATA. THESIS ADVISOR: ASSOCIATE PROFESSOR SOMPONG SIRISOPONSILP, Ph.D., 109 pp.

The paucity of available data was limiting studies of freight flow in Thailand. To overcome this problem, commodity flow survey (CFS) and comprehensive freight transportation by truck using roadside survey (RS) were launched to collect comprehensive freight flow data throughout the kingdom of Thailand. Since these two surveys were pioneering and due to budgetary limitations, the resulting data are still incomplete and must be adjusted.

The need to produce a freight origin destination matrix using available data from CFS and RS led to the objectives of this research. This research has two main objectives. The first is to develop a methodology for combining CFS and RS. The second is to develop a method for filling gaps in the origin destination matrix based on the Adaptive Neuro Fuzzy Inference System (ANFIS) approach.

The methodology to combine these two data sources was developed which uses the strengths of each method, the CFS distribution pattern and the RS marginal total. The first method is Trip Length Distribution Adjusting (TLDA), which uses adjustments to CFS trip length distribution to meet RS marginal total. The second method is Gravity Model Approach (GMA), which uses CFS friction functions to adjust RS data matrix. The method was calibrated using two difference sources of roadside survey. The results indicated that the adjusted volumes of the two data sources agreed despite being collected at different times and by different authors, and that the differences between the total adjusted volumes were quite small. It can therefore be concluded that the developed method can be used to adjust the data.

For the second component, a model using BOX-COX transformation and Adaptive Neuro Fuzzy Inference System (ANFIS) was developed and verified against a convention gravity model. Two types of model, using convention gravity variables and using socio-economic variables, were developed. The results showed that the ANFIS model outperformed both the conventional gravity model and the BOX-COX model. These results proved the performance of the Adaptive Neuro Fuzzy Inference System for modeling complex system and its ability to model freight trip distribution.

Department : ...Civil Engineering........... Student's Signature.....................

Field of study: ..Civil Engineering........... Advisor's Signature.....................

Academic Year : .2009..........................

Acknowledgements

The author would like to express his deepest gratitude to his advisor, Associate Professor Dr. Sompong Sirisoponsilp, for his excellent guidance, inspiration, strong support, and constructive criticisms throughout this study. He provided not only all that was required to accomplish this study, but he also gave valuable advice regarding the author's personal concerns. Without his help both academically and personally, this dissertation work would not have been completed.

The author expresses profound appreciation to his director and his dissertation examination committee, Associate Professor Dr. Pongsak Suriyavanagul, for the opportunity to receive a doctoral scholarship from the Ministry of Higher Education under Kasetsart University Chalermprakiat Sakonnakhon Province Campus, Sakonnakhon, Thailand, and his help, encouragement, suggestions, and constructive comments.

The author also expresses sincere appreciation to the other committee members, Assistant Professor Dr. Terdsak Rongwiriyapanich, Assistant Professor Dr. Manoj Lohatepanont, and Assistant Professor Dr. Kasem Choocharukul, for their help, encouragement, suggestion, and constructive comments, and for serving as members of his dissertation examination committees.

The author wishes to thank his colleagues at Transportation Institute of Chulalongkorn University, Ms. Visarat Sunyalukleachai, Ms. Wannipa Bunnim and Ms. Thitima Wonginta for their help, encouragement, and discussions during his study.

The author would like to acknowledge with deep gratitude and appreciation the National Statistics Office (NSO) for supporting Commodity Flow Survey data and the visiting data survey from field experience.

Thank you to Ms. Yuwadee Hirun for her sacrifice in taking care of the author's lovely daughter, Nong Run, by herself while the author was studying. The lovely and sweet voice of Nong Run gave the author willpower during this trying time in his life. Without them he would not have succeeded.

The author gives the pride for this Ph.D. degree to his mother. Although she does not have a degree, she enlightened the author to understand the important of education. Although she was not able to stand with him on the day of the ceremony, he hopes that she is watching this success. For all of the good things she provided for him, he will remunerate her in the next life.

Contents

Page

List of Tables

Page

List of Figures

Chapter I
Introduction

The aim of this chapter is to explain the background of the research and the problem statement. It also presents the objectives of the research, its scope and limitations, and the expected benefits of this research.

1.1 Background

In land surface transportation links, traffic flow consists of passenger flow and freight flow, the latter of which consists of truck volume on highways or freight trains. Both components are important in transportation planning. Especially in the industrial area, truck volume on the highway is near to or equal to passenger car volume. But while freight flow plays an important role in transportation planning, such as regional, highway and infrastructure planning, limited freight data has led to more extensive studies of passenger flow than freight flow. Since the flow of freight is primarily about economic activities of origin and destination, freight flow data involves many shippers, manufacturers, and receivers, making the data more difficult and costlier to collect than the passenger flow data.

The most important data for modeling freight flow is freight O-D distribution data. The survey method to obtain this data can be divided into two methods. The first is the Commodity Flow Survey (CFS), which collects shipment data such as O-D of shipment, weight of shipment, value of shipment, etc. from sampled shippers. Another type of survey method is the roadside survey which collects shipment data by interviewing drivers along the transportation link. CFS may be preferable to roadside surveys for data accuracy. However, CFS surveys are costlier than roadside surveys, especially national studies. Only two countries (US and Japan) have successfully completed a CFS, with a third (Sweden) finalizing its survey (The Office of National Statistics of Thailand, 2005).

Thailand faces lags in available suitable data for transportation and logistics planning. The developed freight model in the past was calibrated against a small size of roadside interview data. The government is attempting to solve this lack of data by setting up a national logistics data strategy. This strategy propels government agencies to collect important transport and logistics data including Commodity Flow Survey (CFS) and truck O-D survey by the roadside interview method.

The Office of National Statistics of Thailand launched the Commodity Flow Survey (CFS) in 2007. The survey collected data including commodity type, origin and destination, weight, product value, mode of transport, etc. from a large sample of shippers in Thailand. Commodity classification is based on the Harmonized System. The survey's database will be the most complete collection of commodity flow data in Thailand. In 2008, the Department of Land Transport launched the Truck Flow survey. The aim of this project is to collect O-D of trucks by roadside interviews on highways in 10 major provinces of Thailand. Due to budget limitation, this project collected data in the harvest season and out of the harvest season. These two databases complement each other and will be of great use in freight modeling for transportation and logistics planning in Thailand.

According to the US experience, CFS provides a wealth of freight movement data within the US, but CFS still has a number of weak points. A combination of data suppression for confidentiality reasons, limited sample size, and limitations to industrial coverage of the CFS led to empty O-D cells. These empty cells are a gap in O-D matrices which must be explored to determine which cells should contain positive flows and what size of flow should be filled. Data from other sources including USACE's Waterborne Commerce database, Railroad Waybill data, and OAI's air freight database were used as auxiliary data to fill these gaps. Besides auxiliary data, combining two gap-filling methods, Iterative Proportional Fitting and Log-Linear modeling, was employed to solve the gap problem of US CFS (US DOT, 2007). However, the criteria for selecting the method are not based on systematic procedure or statistics evaluation.

As mentioned above, the development of freight O-D distribution from the CFS data must be concerned with two issues, auxiliary data and gap-filling. In the case of Thailand, truck transport has more than 80% of mode share. Thus, the most important auxiliary data should come from the road transport sector. Unfortunately, there is no comprehensive road transport database available in Thailand. Therefore, truck O-D by roadside interviews conducted by the Department of Land Transport may be preferable as auxiliary data for filling the gaps in Thailand's CFS data. There are a number of gap-filling methods besides the two methods proposed by US DOT, such as Gravity model, Regression based model, and soft computing techniques.

The most commonly used method is Gravity model. This method is mainly concerned with replicating the observed flows between every pair of origin and destination with minimum error. The flow is a function of some proxy variable of origin and destination, such as total production of origin zone or total attraction of destination zone. The weaknesses of this method are its inability to explain the relationship of explanatory factors and socio economic variables of the study area, and that it requires a large amount of data for calibration. Moreover, previous works which employ Gravity model focus on best fit with little theoretical foundation (Celik, 2004).

Another method is the regression-based method. Celik and Guldmann (2002) estimated a flexible Box-Cox model with a set of explanatory variables that characterize the economic structure of the origins and destinations. The calibrated model from this technique may give decision makers the ability to control the flow since it may unveil the causative relationship of the flow with the set of policy variables (Celik, 2004). Although this technique can overcome a major inefficient point of Gravity technique, it still suffers from the performance of the model.

Recently, soft computing techniques such as Neural Networks, Fuzzy logic, and Neuro Fuzzy (NFS) have been accepted as efficient alternative tools for modeling complex non-linear systems and widely used for prediction. This new method has proven to be an efficient tool in many disciplines, as well as transportation engineering.

Artificial Neural Network (ANN) is an artificial intelligence technique that mimics the function of the human brain. ANN can approximate a nonlinear relationship between the input and output variables of nonlinear and complex systems. ANN has outperformed many conventional computing methods in many disciplines, as well as in transportation planning (Celik, 2004). However ANN still suffers from the black box phenomenon, since ANN does not provide any mathematic representations between the constituting parts of a system.

The fuzzy logic process is close to human thinking and is easier to use with complex non-linear systems. Fuzzy logic offers the important concepts of fuzzy set theory, fuzzy if-then rules, and approximate reasoning which deals with imprecision and information granularity. Fuzzy models are able to model the uncertain or ambiguous data so often encountered in real life. The fuzzy inference system (FIS) is a popular computing framework based on the concept of fuzzy logic. It has found successful applications in a wide variety of fields such as automatic control, data classification, decision analysis, expert system, time series prediction, and robotics and pattern recognition (Jang, Sun and Mizutani, 1997). Nevertheless, the main problem with fuzzy logic is that there is no systematic procedure for the design of a fuzzy controller (Chang and Chang, 2005).

Neuro-fuzzy (NFS) is an approach where the fusion of neural networks and fuzzy logic find their strengths. These two techniques complement each other. The NFS approach combines the semantic transparency of rule-based fuzzy systems with the learning capability of neural networks. The advantage of the NFS approach is initializing with parameters relating to the problem domain. NFS models have recently gained much popularity for calibrating nonlinear relationships because they offer more advantages over conventional modeling techniques, including the ability to handle large amounts of noisy data from dynamic and nonlinear systems, especially where the underlying physical relationships are not fully understood (Aqil et al., 2006). The NFS approach has recently gained a lot of interest in research and application. The NFS approaches mix ANN with fuzzy inference systems (FIS) in three ways: cooperative, concurrent, and fused. The most common architecture is the fused NFS that uses neural networks ideas just to learn some internal parameters of a fixed structure (Nauck et al., 1997). The most well known of fused NFS architecture was introduced by Jang (1992), the so-called Adaptive Neuro Fuzzy Inference System (ANFIS), which is able to approach any linear or non-linear function (Jang, 1993). ANFIS has been applied as an efficient non-linear approximator in many studies.

1.2 Statement of the Problem

The Thailand CFS is the first comprehensive freight database to be developed in Thailand; thus it has encountered a number of limitations. Therefore, the most important task for a transportation researcher using this powerful technique is to adjust the data in order to provide a suitable dataset for development of a freight O-D distribution database.

As mentioned earlier, soft computing is widely used for modeling complex non-linear systems and making predictions in many disciplines. For trip distribution research, Black (1995) introduced ANN as tool for modeling seven groups of commodity flow between nine census regions of the US. Black concluded that the ANN model has a lower prediction error than a doubly constrained gravity model. A similar result of ANN performance was reported by Celik (2004). Using 1993 Commodity Flow Survey data, the performance of ANN for 15 commodity groups was compared with a regression-based model. The ANN using a conventional gravity model was used as a benchmark model. Celik concluded that the ANN using conventional gravity model variables provides a slight improvement over the regression-based model and that the performance of ANN using theoretically relevant variables from a regression-based model is better than regression-based model. For another technique of soft computing, fuzzy logic, only one study dealing with trip distribution has been introduced, by Kalic and Teodorovic (2003). Using fuzzy logic and genetic algorithm as a tools for modeling passenger distribution, they concluded that these techniques perform well in predicting passenger flows.

For the most efficient technique of soft computing, Adaptive Neuro Fuzzy Inference System (ANFIS) has been applied in many disciplines as well as in civil engineering. ANFIS has been used in the hydrological forecasting context such as reservoir operation (Chang and Chang, 2005; Firat and Gungor, 2006) and modeling hydrological time series (Zounemat-Kermani and Teshnehlab, 2007). Successful applications of ANFIS modeling in water resources forecasting have been widely reported. Nevertheless, ANFIS has not been used in a freight distribution context.

Taking advantage of the soft computing technique and the availability of two types of freight flow data in Thailand (Commodity Flow Data and truck O-D by roadside interview surveys), this research will attempt to focus on how to apply the Adaptive Neuro-Fuzzy Inference System (ANFIS) method to adjust the Thailand commodity flow survey data and to develop freight distribution data in Thailand using the truck O-D by roadside interview method data as auxiliary data. Additionally, the performance of ANFIS method will be evaluated against the traditional Gravity and regression-based methods. Since the study that uses ANFIS approach as a tool for this approach is limited, the development of an ANFIS approach to efficiently fill gaps in the data is a merit of this research. Moreover, combining CFS and roadside interview data for CFS gap filling is the second merit of this research.

1.3 Research Objectives

The goal of this research is to develop a new method to combine CFS and roadside survey data and to develop a novel ANFIS technique for filling gaps in the data. In order to achieve this goal, the objectives of this research are as follows:

1) develop a technique for combining CFS and roadside survey data

2) develop a gap-filling method using an ANFIS approach

3) develop a freight O-D database using combined Commodity Flow Survey and truck O-D by roadside interview method data in Thailand

1.4 Scope and Limitations

Commodity Flow Data from the Office of National Statistics and truck O-D by roadside interview method data conducted by the Department of Land Transport are the primary raw data for this research.

The traffic analysis zone in this research is divided at the provincial level and at special generator zones such as major ports.

The socio-economic data associated with traffic analysis zones are the data published by the Office of National Statistics and related government agencies.

1.5 Expected benefits

The expected benefits of this research can be divided into two contexts:

1) Taking advantage of the commodity flow survey (CFS) data and the roadside survey of truck O-D data in Thailand, this research will create a new method for combining data to develop freight data for transportation planning in Thailand.

2) This study will create an ANFIS gap-filling method which is useful for freight O-D distribution analysis.

Chapter II
Literature review

In order to develop efficient freight O-D matrices for transport and logistics planning in Thailand, the literature relevant to this research was reviewed. This chapter starts with details about commodity flow survey data. Then it describes the method for CFS gap-filling provided by US DOT and the concept of O-D estimation from roadside survey data. Next is a review of modeling freight distribution using commodity flow data research in the past. The last part of this chapter is dedicated to theoretical considerations.

2.1 Commodity flow survey data

United States of America initiated Commodity flow surveys in 1993, 1997, and 2002. As a result of these surveys U.S. now has data on the annual volume of commodity movements taking place into, out of, and within each state, the District of Columbia, and the largest metropolitan areas, broken down by mode of transport. Although this survey has filled a large gap in U.S. freight data, U.S. transportation planning encounters two kinds of problems. First, the surveys do not cover all U.S. freight movements. Second, the surveys only support the representation of origin-to-destination (O-D) movements between quite large geographic regions and are limited in the level of commodity detail (Southworth, 2005).

Southworth (2005) summarizes the strengths and weaknesses of U.S. commodity flow survey data as shown below.

The CFS has a number of unique strengths, in particular:

- the survey's coverage at the national level;
- the survey covers all the major surface transportation modes, including truck, rail, water, petroleum pipelines, and air freight;
- the survey identifies the true geographic origin and destination of each shipment and provides estimates of "door-to-door" shipment distances;
- it collects data on both the weight and dollar value of all in-scope shipments;
- the survey has a time series in the form of the 1993, 1997, and 2002 surveys; and
- it is done in conjunction with the Economic Census, providing concurrency with other datasets.

It also has weaknesses, in particular:

- not all commodities are covered by the CFS;

- imports are out-of-scope of CFS;

- the spatial detail available to its mode specific O-D matrices is limited to a small number of rather large geographic regions;

- the volume of "intermodal" freight reported may be low;

- the shipment length detail available from non-geographically disaggregated products is very limited in its supporting commodity-level detail;

- the surveys have seen some content changes. Reduction in sample size from four times to one between 1993 and 2002 leads to some large coefficients of variation in reported estimates; and

- there are discrepancies in the estimates generated by the CFS and the U.S. Army Corps of Engineers' waterborne commerce data, the latter based on industry-wide carrier reporting that produces larger ton and ton-mileage figures.

A combination of data suppression for confidentiality reasons, limited sample size, and limitations to the scope of the CFS has led to many empty cells which should contain a flow. These empty cells are a gap in O-D matrix which must be explored to determine what size of flow should be filled and which cells ought to contain a positive flow. Southworth (2005) suggests that missing values can be estimated from reported cell values by applying a mathematical equation. Moreover, combining data from the CFS matrix with data from other sources, such as the railcar waybills (suitably modified to match CFS regions and commodity classes), can be used as a second estimate or "data model" of the rail flows in each commodity class. Combining can be done in a number of ways.

- Replacing CFS-based missing cell data with waybill estimated values and then using a gap-filling method to bring the full matrix back into compliance with the original CFS flow margins.

- Making the railcar waybill flows as though they were a separate dimension or set of commodity specific tables in the rail portion of the CFS flow matrix, and filling in the missing cell values using a combination of CFS and waybills data.

Figure 2.1 show the process to modify CFS data, which will be discussed in the next section.

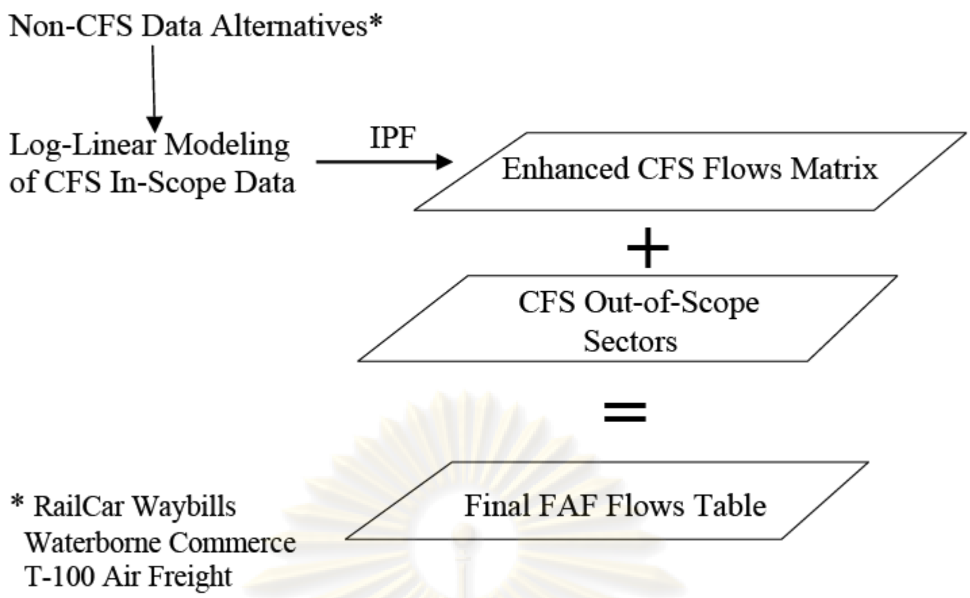

Fig 2.1 Modification of US CFS data process
Source: Southworth (2005)

2.2 US CFS gap-filling methodology

US DOT provided the methodology for filling gaps in US CFS data. This includes several major steps and numerous assumptions. Details are as follows.

Step 1: Preparing CFS data

Because CFS reports shipments at the state level, not the regional level, the aim of the first step is to disaggregate the data from the state level to the regional level. The method for disaggregation is dividing shipments equally across all new regions that comprise each state.

Step 2: Identify "True Zeros" in original CFS data

The purpose of this step is to identify the cell which had no samples. By assumption, those "true zero" cells were constrained to be "0."

Step 3: Auxiliary data and conversions to standard commodity type

Auxiliary data were obtained from other data sources including USACE's Waterborne Commerce database, Waybill data, and OAI's air freight database. These data were converted from the base commodity categorizations used by each data source to standard classification of transported goods (SCTG).

Step 4: Verify "True Zeros" with auxiliary data

The auxiliary data were compared to verify agreement between the two data sources. Waybill, Waterborne Commerce, and air freight data were compared with the dataset from step two for those particular cells to verify that neither of those datasets contradicted the true-zero cells. In cases of contradiction, where auxiliary data showed that cells previously marked as "true zero" contained a flow, the restriction on that cell or margin was lifted.

Step 5: Augment original CFS data with auxiliary data

Original CFS data was augmented with water, rail, and air freight data as constructed from auxiliary data in Step 3. Thus, some cells had two values – one from the CFS and one from an auxiliary source.

Step 6: Log-linear modeling

Log-linear models, specialized cases of general linear models, were employed for estimating the most likely values of those missing cells, based upon statistical relationships extracted from cells with known values.

From the Log-linear model concept, flow between origin (i) and destination (j) by mode of transport (m) can be written as follows:

$$F_{ijm} = \tau_O \tau_i^O \times \tau_j^D \times \tau_m^M \times \tau_{im}^{OM} \times \tau_{jm}^{DM} \times \tau_{ij}^{OD} \times \tau_{ijm}^{ODM} \qquad (2.1)$$

Taking the logs leads to

$$\ln F_{ijm} = \theta + \lambda_i^O + \lambda_j^D + \lambda_m^M + \lambda_{im}^{OM} + \lambda_{jm}^{DM} + \lambda_{ij}^{OD} + \lambda_{ijm}^{ODM} \qquad (2.2)$$

The various λ's are a set of model-estimated parameters that will return the original cell estimates.

Step 7: Iterative proportional fitting

Iterative Proportional Fitting is a well accepted approach to adjust values within cells. The benefit of this method is the ability to maintain the relationships between variables and to ensure that rows and columns are consistent with the appropriate marginals. The concept behind IPF is to seed each of the missing data cells with an initial estimate of some form, then iterate over all the different margins of the matrix until a new balance has been obtained that does the least damage to the estimates in the rest of the matrix, while retaining the values of the statistically more robust (typically) marginal totals that often represent the reported data.

Assume that the simplest two-dimensional case, in which O(i) and D(j) are a set of row (i) and column (j) totals respectively (e.g. annual freight tons produced at each i and consumed at each j), and where T(i,j) is the tons of freight shipped from region i to region j annually. A simple IPF routine applied to this problem can be stated as:

$$T(i,j,r+1) = T(i,j,r) / \sum_j T(i,j,r) * O(i) \qquad (2.3)$$

$$T(i,j,r+1) = T(i,j,r+1) / \sum_i T(i,j,r+1) * D(j) \qquad (2.4)$$

where r, r+1, and r+2 refer to successive iterations, and where equations (2.3) and (2.4) can be applied iteratively until at some iteration r+g is gotten:

$$\sum_j T(i, j, r+g) = O(i) \text{ for all } i \text{ and} \tag{2.5}$$

$$\sum_i T(i, j, r+g) = D(j) \text{ for all } j \tag{2.6}$$

Step 8: Adding Out-of-Scope shipments

CFS does not include traffic flows originating for several "Out-of-Scope" business sectors. Several commodities were totally absent in the 2002 US CFS survey which would be divided into these three contexts:

- One or more shipments in a commodity's supply chain were absent from the CFS survey.
- Whole categories of shipments were omitted from the survey, such as the movement of retail commodities from the point of final purchase to the home, business, etc.
- There was evidence that the 2002 CFS undercounted some commodities and types of shipments – based on significant differences with other reliable data sources.

US DOT reports 15 CFS gaps and undercounts which consist of

- **Farm Based:** These include shipments of farm commodities from the farm to the first point of sale.
- **Fisheries:** These include shipments of fish and seafood from boats on the dock to processors or from fish farms to processors.
- **Crude Petroleum:** Crude petroleum shipments are completely outside the scope of the 2002 CFS.
- **Natural Gas:** Natural gas shipments are completely outside the scope of the 2002 CFS.
- **Municipal Solid Waste (MSW):** MSW shipments are completely outside the scope of the 2002 CFS.
- **Logging:** These include shipments of logs from points of harvest to initial points of processing.
- **Construction:** These include shipments that originate from the construction sector, such as construction companies or establishments engaged in construction of residential and nonresidential buildings; utility systems; highway, street, and bridge construction; and specialty trade contractors.

- **Services:** These include shipments which originate from establishments involved in service industries: finance and insurance; real estate, rental and leasing; professional, scientific, and technical services; administrative and support; waste management and remediation services; education services; health care and social assistance; arts, entertainment, and recreation; accommodation and food services; other services (e.g., repair and maintenance, personal and laundry, religious, etc.); and public administration.

- **Publishing:** The CFS data gap on the publishing industry is primarily due to the adoption of the North American Industry Classification System (NAICS) in the 2002 CFS for selection of business establishments. In the 1997 and 1993 CFS, businesses were selected based on their descriptions in the Standard Industry Classification (SIC).

- **Retail:** Retail trade stores, including motor vehicle and parts dealers, furniture and home furnishings stores, electronics and appliance stores, building materials and garden equipment and supplies dealers, food and beverage stores, health and personal care stores, gasoline stations, clothing and clothing accessories stores, sporting goods stores, book and music stores, general merchandise stores, florists, used merchandise dealers, manufactured home dealers, etc., are not included in sample size.

- **Household and Business Moves:** CFS does not capture freight movements by carriers that transport household and business furniture, equipment, etc.

- **Imports:** Imports are completely outside the scope of the 2002 CFS. However, once import commodities enter the United States and change ownership, further shipments of those "imports" are captured within the CFS.

- **Petroleum Products:** Petroleum products are technically within the scope of the CFS. However, previous research suggested that the 2002 CFS undercounted petroleum products.

- **Exports:** Although CFS included exports from the United States by all freight modes, analysis of the 1993 and 1997 CFS export data suggests that the CFS underestimated U.S. export shipments.

- **In-transits:** The CFS does not include shipments of commodities that originate outside of the United States, enter the United States by whatever mode, and then are shipped to some other country.

In this Step, these 15 categories of shipments must be added to the table from Step 7 to arrive at the final 2002 Commodity Origin-Destination database. In order to generate an expedient and reasonable regionalization of out-of-scope commodity flows, one needs to reflect the relative regional differences in economic activity that generate the truck commodity flows using readily and openly available data on state and local economic activity.

The process for estimating the regionalization of national truck freight flows are as follows:

1. The allocation of the national freight estimates to the county in which the freight generation occurs
2. The estimation of county-to-county freight flows for each commodity shipped in the out-of scope business sectors
3. The aggregation of the county-to-county flows to regional commodity flows used in the CFS matrix

Step 9: Analysis of Results

The 2002 Commodity Origin-Destination Database contains 3 four-dimensional matrices (tons, ton miles, and value) for 43 commodities, 138 origins, 138 destinations, and 11 modes – for a total of more than 27 million cells which will be explored.

Step 10: Validation

Three validations approaches are used for validating the modified CFS database which consists of

- The first approach is Cross validation, in which random cells from the final 2002 Commodity Origin-Destination Database are removed and the 2002 Commodity Origin-Destination Database is re-estimated (Steps 1 through 8). The re-estimated tables are then compared to the tables from Step 8 using standard statistical approaches.
- The second validation approach compares the statistical relationships among the parameters derived from Step 8 with the same statistical relationships derived from the auxiliary data.
- The third validation approach compares the absolute values of each cell from Step 8 with known absolute values for those same cells from the auxiliary data sources.

2.3 O-D Estimation from roadside survey data method

Another type of data for estimating O-D matrices is roadside survey data. The survey collects sampling driver-interview data from survey stations. The O-D sampling from different survey stations can be expanded by the following relationship:

$$\hat{T}_{ij} = \sum_{a \in ST_{ij}}^{m} \frac{t_{ij}^{a'}}{r_a} \tag{2.7}$$

where T_{ij} is true total O-D flow from i to j

$\hat{T}_{ij}$ is the estimated total O-D flow from i to j

$t_{ij}^{a'}$ is the observed O-D flow from i to j at a roadside survey station on link a

r_a is the sampling rate at the roadside survey station on link a

m is the number of roadside stations

ST_{ij} is the set of links where roadside survey stations are located and which have a nonzero probability of being used for travel from i to j

From the relationship shown above, $\hat{T}_{ij}$ is a biased estimator. Because trips which pass by more than one survey station appear in multiple samples, these trips are over-represented in the expanded complete trip table. The situation is the so-called "double counting problem" which must be eliminated while developing the expanded matrices.

Kuwahara and Suliwan (1987) proposed five methods for estimating O-D from roadside interview data that eliminate the double counting problem. Three methods were based on the least square estimator, while the other methods employed the maximum likelihood technique. Details of these methods are as follows.

Method 1: This method is based on the principle that each sample observation is weighted equally, so $\hat{T}_{ij}$ can be written as follows:

$$\hat{T}_{ij} = \frac{\sum_{a \in ST_{ij}} t_{ij}^{a'} / P_{ij}^{a} r_a}{\sum_{a \in ST_{ij}} 1} \tag{2.8}$$

where P_{ij}^{a} is probability of a trip from an origin i to destination j which use link a.

Method 2: This method is based on an assumption that the representative of the data from each station is proportional to the sampling rate, so $\hat{T}_{ij}$ can be written as follows:

$$\hat{T}_{ij} = \frac{\sum_{a \in ST_{ij}} t_{ij}^{a'} / P_{ij}^{a}}{\sum_{a \in ST_{ij}} r_a} \tag{2.9}$$

Method 3: This method is based on the concept that the errors are more important at stations having a high proportion of the O-D trips, so $\hat{T}_{ij}$ can be written as follows:

$$\hat{T}_{ij} = \frac{\sum_{a \in ST_{ij}} t_{ij}^{a'} / r_a}{\sum_{a \in ST_{ij}} P_{ij}^{a}} \tag{2.10}$$

Method 4: This method employs maximum likelihood method with two assumptions:

1. Each $\hat{T}_{ij}$ is an independent random variable
2. The probability distribution of $\hat{T}_{ij}$ is hypergeometric

For a large sampling rate resulting in a large total number of vehicles counted at the survey station, $\hat{T}_{ij}$, can be written as follows:

$$T_{ij} = \frac{\sum_{a \in ST_{ij}} \left(\frac{r_a}{1 - r_a} \right) \frac{t_{ij}^{a'}}{r_a}}{\sum_{a \in ST_{ij}} \left(\frac{r_a}{1 - r_a} \right) P_{ij}^{a}} \tag{2.11}$$

Method 5: This method is based on the maximum likelihood method which is applied to a very small sampling rate, so $\hat{T}_{ij}$ can be written as follow

$$\hat{T}_{ij} = \frac{\sum_{a \in ST_{ij}} t_{ij}^{a'}}{\sum_{a \in ST_{ij}} P_{ij}^{a} r_a} \tag{2.12}$$

Kuwahara and Suliwan (1987) concluded that the most suitable method depends on the network structure, the sampling strategy, and the observed data. Thus, it is necessary to evaluate and choose the expansion method case by case.

2.4 Modeling freight distribution from commodity flow data using soft computing technique

Black (1995) estimated the flow of seven commodity groups between nine census regions of the US. The 1997 commodity flow data were used as input data. Three methods were used in this study. The first is an unconstrained gravity model, the second is a fully constrained gravity model, and the third is an Artificial Neural Network based model, the so-called Gravity Artificial Neural Network (GANN). Input of two gravity models consists of flow production, flow attraction, and distance between all flow regions, while the input of GANN consists of regional flow production, regional flow production, and the interregional distance between the origin and destination region. Black concluded that the GANN model has a lower prediction error than a doubly constrained gravity model.

Celik and Guldmann (2002) determined the flow of 16 commodity groups for 48 continental states of the US using the 1993 Commodity Flow Survey. The Box-Cox functional form was employed as a transformation function. A set of explanatory variables that characterized the economic structure of the origins and destinations was used as input variables to the model. These variables were divided into three types, consisting of

- Origin variables. The origins serve as supply points but also consume part of this supply. Thus, origin variables should be proxies for both supply and demand conditions. These variables include sectoral employment, sectoral value-added, wholesale employment, total population, personal income per-capita, and the average plant size.

- Destination variables. The destinations serve as demand points, with the destination variables serving as proxies for commodity demands, both intermediate and final. These variables include manufacturing employment, personal income per capita, and total population.

- Geographical variables. This variable is a friction variable of flow between origin and destination. Distance is the most conventional friction variable used in a number of models. Two additional variables, competing destinations and intervening opportunities, are employed to capture the effects of the spatial configuration of states.

Celik and Guldmann (2002) conclude that the selected variables and the Box-Cox function form are successful in explaining shipment variables.

Celik (2004) estimated the flow of 15 commodity groups of 1993 US Commodity Flow Survey data. The performance of an ANN based model using variables derived from regression based model was compared with an original regression-based model. ANN using the conventional gravity model was used as a benchmark model. Celik concluded that the ANN using conventional gravity model variables provides a slight improvement with respect to the regression-based model, and that the performance of ANN using theoretically relevant variables from regression-based model are surprisingly superior to the regression-based model.

Celik (2004) suggests that the “black-box’’ phenomenon is a main limitation of ANN, since it is unable to establish a causal relationship between variables to present the constituting parts of a system. For this reason, the ANN model suffers when defining weights for a policy variable in the model, unlike regression models. Moreover, ANN is unable to accommodate a change in network structure.

2.5 Theoretical Considerations

2.5.1 Gravity model

The gravity model is derived from Newton’s gravity model. The attraction between two objects is proportional to their mass and inversely proportional to their respective distance. Consequently, the general formulation of spatial interactions can be adapted to reflect this basic assumption to form the elementary formulation of the gravity model:

$$T_{ij} = k\frac{P_i A_j}{f(t_{ij})} \tag{2.13}$$

where T_{ij} is freight flow between supply node i and demand node j,
Pi is total freight volume at supply node I,
A_j is total consumption volume at demand node j,

$f(t_{ij})$ is impedance function for freight flow between supply node i and demand node j, and
k is a proportionality constant.

The gravity model can be divided into three types as follow.

Trip attraction constrained: The sum of all trips in a row, the trip production, should equal the total interaction flows exiting a particular zone.

$$\sum_j T_{ij} = O_i \qquad \text{for all } i \tag{2.14}$$

Trip production constrained: The total number of all trips in a column, the trip attraction, should equal the total interaction flows entering a particular zone.

$$\sum_i T_{ij} = D_j \qquad \text{for all } j \tag{2.15}$$

A doubly constrained gravity model: The total number of all trips in a row should equal the total interaction flows exiting a particular zone and the total number of all trips in a column should equal the total interaction flows entering a particular zone.

$$\sum_j T_{ij} = O_i \qquad \text{for all } i \tag{2.16}$$

$$\sum_i T_{ij} = D_j \qquad \text{for all } j \tag{2.17}$$

There are a number of methods for calibrating the gravity model. Bergkvist and Westin (1997) compared a performance of four specifications of the gravity model. The four models were the traditional gravity model with ordinary least square (OLS), non-linear least squares (NLS), Poisson distributed model, and semi-parametric neural network. Bergkvist and Westin (1997) concluded that estimations with OLS and NLS are inferior to those with Poisson and Neural network models. Moreover, the neural network model outperformed the other models in terms of Root Mean Square Error (RMSE).

In general, the least square method encountered the zero flow problem, especially in cases of gravity with logarithm transform, since the logarithm is then undefined. Fotheringham and O'Kelly (1989) suggested several solutions to this problem.

- Remove all zero from the analysis. The resulting parameter estimates of this solution would not reflect the low volumes of interaction that occur between certain origins and destination, and thus would be misleading.

- Remove all origins and destinations associated with zero interactions from the analysis. However, a great deal of useful information would be lost in this way, and in particularly sparse matrices, there may be no origin that has a non-zero interaction to every destination.

- Adding a constant to elements of the interaction matrix which would be divided into two ways. The first is to add the constant to every flow in the matrix; the other is to add the constant only to the zero flows. In practice, there seems little difference between the two methods in terms of the resulting parameter estimates. In both cases, some uncertainty exists over the value of the constant to be added. Probably the most frequently encountered method of dealing with zero interaction is to add one to every zero flow; this can be justified on the grounds that recorded flows are generally integers, and one is the closest approximation to zero.

Another type of widely used calibration method for the gravity model is maximum likelihood estimation (MLE). The technique of MLE is to find parameter estimates that maximize the likelihood of observing a sample set of interactions from a theoretical distribution. The steps of MLE calibration include identifying a theoretical distribution for the interaction, maximizing the likelihood function of the distribution with respect to the parameters of the model, and deriving equations that ensure the maximization of the likelihood function.

2.5.2 Regression based model

The regression based model was derived from the concept of spatial price equilibria of interregional trade. At each point which supplies commodities does so in the form of the firm's production, while at the demand point there are firms and households demanding certain quantities from that supply point. From this concept the model was developed from four equations.

The supply function at the supply point is defined as

$$S_i = \sigma_i(p_i, s_i) \qquad (2.18)$$

where

S_i is the supply quantity at supply point i
p_i is the f.o.b. price at i
s_i is a vector of other variables

The relation at demand point j is defined as

$$y_j = \delta_j(q_j, w, d_j) \qquad (2.19)$$

where

y_j is the O-D flow terminating at j
q_j is the c.i.f. price vector
w is a vector of parameters that measure the supply characteristics influencing purchase choices
d_j is a vector measuring demand characteristics

The commodity price at the supply point and the demand point is quite different. The firms at the supply point are faced with f.o.b. (free on board) while the firms and household at the demand point with c.i.f. (cost, insurance, freight). The relation between c.i.f. prices and f.o.b. prices is defined as follows:

$$q_{ij} = p_i + c_{ij} \tag{2.20}$$

where c_{ij} is the transportation cost between i and j At equilibrium condition the relationship between y_{ij} and S_i can be written as follows:

$$\sum_j y_{ij} = S_i \qquad \text{for } \forall i \tag{2.21}$$

Eliminating price leads to a reduced form of the model in which the equilibrium flow is directly assigned to the vector of exogenous variable (s, w, d, c). The function of the flow can be written

$$Y^* = \xi(s, w, d, c) \tag{2.22}$$

An approach to analyze the flow from this equation is to view as empirically functional forms such as gravity.

$$\xi_{ij}(s, w, d, c) = a_i(s, w, d, c) f(c_{ij}) b_j(s, w, d, c) \tag{2.23}$$

where

a_i is the supply point factor
b_j is the demand point factor
c_{ij} is the interaction factor

Celik and Guldmann (2002) suggested the Box-Cox transformation as the transform function for model parameter estimation. The Box and Cox (1964) power transformation is widely used to achieve a normalizing transformation on a positive-valued response variable as shown below.

$$w = \begin{Bmatrix} (Y^{\lambda} - 1) \\ \ln Y \end{Bmatrix} \qquad \begin{matrix} \text{for } \lambda \neq 0 \\ \text{for } \lambda = 0 \end{matrix} \tag{2.24}$$

The λ value was estimated by Maximum Log-likelihood function as shown:

$$LL \quad = \quad -\frac{n}{2}\ln\left[\sum_{i=1}^{n}\frac{(y_i^{(\lambda)} - \bar{y}^{(\lambda)})^2}{n}\right] + (\lambda - 1)\sum_{i=1}^{n}\ln y_i \tag{2.25}$$

where	LL	is	Log-likelihood function
	λ	is	Maximum likelihood of Box-Cox parameter
	n	is	Number of data
	$y_i^{(\lambda)}$	is	Number of transformed data
	$\bar{y}^{(\lambda)}$	is	arithmetic mean) of transformed data calculated by

$$\bar{y}^{(\lambda)} = \frac{1}{n}\sum_{i=1}^{n} y_i^{(\lambda)} \tag{2.26}$$

There are a number of variable transformed methods such as left-hand-side only, right-hand-side only, transform independent and dependent variable with the same transform variable, and transform independent and dependent variable with the different transform variable and no transform of some variable. Details of these transformation methods are shown below.

Table 2.1 Method of BOX COX transformation

Function	**Format**
lhsonly	$y_j^{(\theta)} = \beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \beta_k x_{kj} + \varepsilon_j$
rhsonly	$y_j = \beta_1 x_{1j}^{(\lambda)} + \beta_2 x_{2j}^{(\lambda)} + \cdots + \beta_k x_{kj}^{(\lambda)} + \varepsilon_j$
rhsonly notrans( )	$y_j = \beta_1 x_{1j}^{(\lambda)} + \beta_2 x_{2j}^{(\lambda)} + \cdots + \beta_k x_{kj}^{(\lambda)} + \gamma_1 z_{1j} + \gamma_2 z_{2j} + \cdots + \gamma_l z_{lj} + \varepsilon_j$
lambda	$y_j^{(\lambda)} = \beta_1 x_{1j}^{(\lambda)} + \beta_2 x_{2j}^{(\lambda)} + \cdots + \beta_k x_{kj}^{(\lambda)} + \varepsilon_j$
lambda notrans( )	$y_j^{(\gamma)} = \beta_1 x_{1j}^{(\lambda)} + \beta_2 x_{2j}^{(\lambda)} + \cdots + \beta_k x_{kj}^{(\lambda)} + \gamma_1 z_{1j} + \gamma_2 z_{2j} + \cdots + \gamma_l z_{lj} + \varepsilon_j$
theta	$y_j^{(\theta)} = \beta_1 x_{1j}^{(\lambda)} + \beta_2 x_{2j}^{(\lambda)} + \cdots + \beta_k x_{kj}^{(\lambda)} + \varepsilon_j$
theta notrans( )	$y_j^{(\theta)} = \beta_1 x_{1j}^{(\lambda)} + \beta_2 x_{2j}^{(\lambda)} + \cdots + \beta_k x_{kj}^{(\lambda)} + \gamma_1 z_{1j} + \gamma_2 z_{2j} + \cdots + \gamma_l z_{lj} + \varepsilon_j$

2.5.3 Adaptive Neuro-Fuzzy inference system (ANFIS)

The ANFIS is a multilayer feed-forward network which fuses ANN learning algorithms and fuzzy reasoning to map an input into an output. ANFIS structure consists of nodes and directional links which connect the node. Moreover, parts of all the nodes are adaptive, which means that each output of these nodes depends on the parameters pertaining to the node and that the learning rule specifies how these parameters should be changed to minimize a prescribe error measure. To explain the ANFIS structure, two fuzzy if-then rules based on the first order Sugeno as shown in Fig 2.2 are considered.

Rule 1: IF x is A_1 and y is B_1 THEN $f_1 = p_1^* x + q_1^* y + r_1$ (2.27)

Rule 2: IF x is A_2 and y is B_2 THEN $f_2 = p_2^* x + q_2^* y + r_2$ (2.28)

where x and y are the crisp inputs to the node i, and $\{p_i, q_i, r_i\}$ are the linear parameter set in the consequent part of the first order Sugeno fuzzy model.

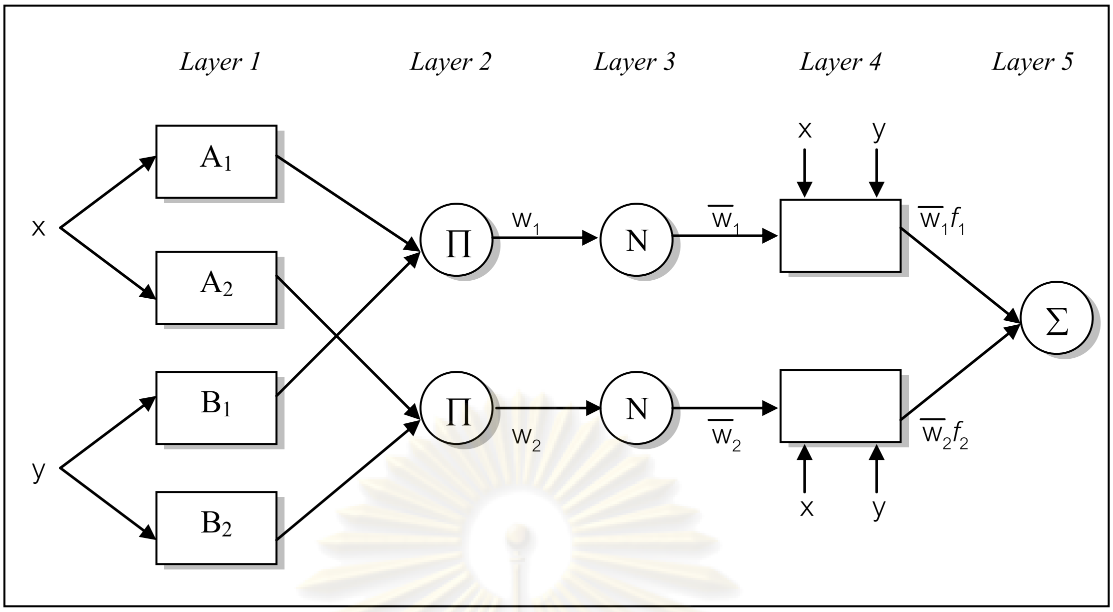

Fig 2.2 ANFIS architecture for two-input Sugeno fuzzy model with two rules

Layer1: *input nodes*. Every node in this layer is an adaptive node. Each node generates membership grades of the crisp input which belong to each of the appropriate fuzzy sets using the membership function. The output of this layer are given by

$$O_{1,i} = \mu_{Ai}(x) \qquad \text{for } i = 1,2 \qquad (2.29)$$

$$O_{1,i} = \mu_{Bi-2}(y) \qquad \text{for } i = 3,4 \qquad (2.30)$$

where x and y are the crisp inputs to the node; *Ai* and *Bi_2* are a linguistic label (such as small or large) by appropriate membership function μ_{Ai} and μ_{Bi-2} respectively. Many various membership functions such as trapezoidal, triangular, Gaussian functions, generalized bell function, or other shapes can be applied to determine the membership grade.

If the Gaussian function is employed, $\mu_{Ai}(x)$ is given by

$$\mu_{Ai}(x) = \exp\left[-\left(\frac{x-c_i}{a_i}\right)^2\right] \qquad (2.31)$$

If the generalized bell function is employed, $\mu_{Ai}(x)$ is given by

$$\mu_{Ai}(x) = \frac{1}{1+\left|\dfrac{x-c_i}{a_i}\right|^{2b}} \qquad (2.32)$$

where $\{a_i,\ b_i,\ c_i\}$is the parameter set that changes the shape of membership function. Parameters in this layer are referred to as premise parameters.

Layer 2: *rules nodes.* Every node in this layer is a fixed node labeled ∏ indicating that it performs as a simple multiplier. The AND operator is applied to obtain one output that represents the result of the antecedent for a fuzzy rule, that is, firing strength. Firing strength means the degree by which the antecedent part of a fuzzy rule is satisfied and it indicates the shape of the output function for the rule. The output of this layer can be represented as

$$O_{2,i} = w_i = \mu_{Ai}(x)x\mu_{Bi}(y) \qquad \text{for } i = 1,2 \qquad (2.33)$$

Layer3: *average nodes.* Every node in this layer is a fixed node labeled N. The node in this layer plays a normalization role to the firing strength from the previous layer. The output of this layer is given by

$$O_{3,i} = \overline{w}_i = \frac{w_i}{w_1 + w_2} \qquad \text{for i} = 1,2 \qquad (2.34)$$

Layer4: *consequence nodes*. Every node in this layer is an adaptive node. The node function of this layer computes the contribution of each *i*th rule toward the model output with the function defined as

$$O_{4,i} = \overline{w}_i f_i = \overline{w}_i(p_i x + q_i y + r_i) \qquad \text{for i} = 1,2 \qquad (2.35)$$

where w_i is normalized firing strength from the previous layer and $\{p_i,\ q_i,\ r_i\}$ is the parameter set. Parameters in this layer are referred to as consequent parameters.

Layer5: *output node.* The single node in this layer is a fixed node labeled ∑. This node computes the overall output as the summation of all incoming signals. This layer is the last step of ANFIS. Hence, the overall output of the model is given by

$$O_{5,i} = \sum_i \overline{w}_i f_i = \frac{\sum_i \overline{w}_i f_i}{\sum_i w_i} \qquad (2.36)$$

2.5.4 ANFIS Learning algorithm

There are many algorithms to identify the parameters in an adaptive network. But with a simple optimization method such as back propagation, for example, the steepest descent takes a long time to reach convergence. Since the output of an adaptive network is linear in some of its network parameters, this study uses the linear least squares method to identify these linear parameters. Note here that the non-linear parameters are fixed; thus the output of the ANFIS model can be written as follows:

$$f = \frac{w_1}{w_1 + w_2} f_1 + \frac{w_2}{w_1 + w_2} f_2 \qquad (2.37)$$

Substituting Eqs (2.32) – (2.35) yields

$$f = \overline{w_1} f_1 + \overline{w_2} f_2 \tag{2.38}$$

Substituting the fuzzy if then rules into Eq (x), it becomes

$$f = \overline{w_1}(p_1 x + q_1 y + r_1) + \overline{w_2}(p_2 x + q_2 y + r_2) \tag{2.39}$$

After rearrangement, the output can be written as follows:

$$f = (\overline{w_1} x) p_1 + (\overline{w_1} y) q_1 + (\overline{w_1}) r_1 + (\overline{w_2} x) p_2 + (\overline{w_2} y) q_2 + (\overline{w_2}) r_2 \tag{2.40}$$

which is a linear combination of the consequent parameters. This approach leads to a hybrid algorithm, which combines a gradient descent method to tune premise non-linear parameters with a least squares method to identify consequent linear parameters for fast identification of those parameters (Jang et al., 1997).

The hybrid algorithm has two process steps, forward pass and backward pass. In the forward pass the premise parameters are held fixed, node outputs go forward until layer 4, and the consequent parameters are identified by the least squares method. Once the optimum consequent parameters are found, the backward pass starts immediately. The consequent parameters are held fixed, the error signals propagate backward, and the premise parameters are updated by gradient descent method. The output of ANFIS is calculated by employing the consequent parameter found in the forward pass. The output error is used to adapt the premise parameters by means of standard back-propagation algorithm. It has been proven that this hybrid algorithm is highly efficient in ANFIS training (H. Esen, et al., 2007).

2.5.5 Constraints of ANFIS

ANFIS supports only Sugeno-type fuzzy inference systems. Moreover, when modeling with ANFIS, the following must be achieved:

- Be first or zeroth order Sugeno-type systems.
- Have a single output.
- All output membership functions must be the same type and be either linear or constant.
- Have no rule sharing. Different rules cannot share the same output membership function; namely the number of output membership functions must be equal to the number of rules.
- Have unity weight for each rule.

An error occurs if FIS structure does not comply with these constraints. Moreover, ANFIS cannot accept all the customization options that basic fuzzy inference allows. That means it cannot customize membership functions and defuzzification functions.

2.5.6 Validation approach

US DOT provided three validations approaches for validating the modified CFS database.

- The first approach is Cross validation, in which random cells from final Commodity Origin-Destination Database are removed and the Commodity Origin-Destination Database is re-estimated. The re-estimated tables are then compared to the final Commodity Origin-Destination Database using standard statistical approaches.

- The second validation approach compares the statistical relationships among our parameters derived from mathematical modeling with the same statistical relationships derived from auxiliary data.

- The third validation approach compares the absolute values of annual tons, annual dollar value, and annual ton-miles from final Commodity Origin-Destination Database with known absolute values for those same cells from auxiliary data sources.

Under the first validation approach as shown above the standard statistic method would be employed. Although a number of statistical approaches for evaluating the model exist, the most popular approach consists of RMSE, MRE, and R-square. Details of theses approach are as follows.

The root mean square error, RMSE, is calculated by:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(e_i - p_i)^2} \qquad \text{For } i = 1 \text{ to } n \qquad (2.41)$$

where e_i is the actual value from experiments

p_i is the predicted value by models

N is the numbers of data points

The mean relative error, MRE, is calculated by

$$MRE(\%) = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{p_i - e_i}{e_i} x100\right| \qquad (2.42)$$

where e_i is the actual value from experiments

p_i is the predicted value by models

N is the numbers of data points

The R-square is given by the following:

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \qquad \text{For } i = 1 \text{ to } n \qquad (2.43)$$

Chapter III
Freight flow data in Thailand

The aim of this chapter is to give information on freight flow data in Thailand. It starts with comprehensive discussions of Commodity Flow Survey data, Roadside survey data from the Department of Land Transport, and roadside survey data from Chonburi road network strategic planning for supporting logistics development project. Next is a comparison of Commodity Flow Survey data, Roadside survey data from the Department of Land Transport.

3.1 Commodity Flow Survey data in Thailand

The office of National Statistics of Thailand (NSO) started a Commodity Flow Data survey on January 2007 which ended on February 2008. The survey collected comprehensive data of freight transportation in Thailand including commodity type, origin and destination, weight, product value, mode of transport, etc. from a large sample of shippers in Thailand. The details of CFS are as follows.

3.1.1 Scope and frame

The sample was selected according to the ISIC (International Standard Industrial Classification of All Economic Activities: ISIC Rev.3.0) classification. Shippers with 11+ workers were the population of the survey. A total of 17,149 shippers were finally included in the survey. Industrial types included in the sample size were

- Mining and quarrying (1.9%)
- Manufacturing (69.6%)
- Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household goods (28.1%)
- Transport, storage and communications (0.4%).

Details are as follows.

Table 3.1 Details of Thailand CFS sample

Industrial Type	Two-digit ISIC
C - Mining and quarrying	10 - Mining of coal and lignite; extraction of peat 11 - Extraction of crude petroleum and natural gas; service activities incidental to oil and gas extraction excluding surveying 13 - Mining of metal ores 14 - Other mining and quarrying
D - Manufacturing	15 - Manufacture of food products and

Industrial Type	Two-digit ISIC
	beverages 20 - Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials 21 - Manufacture of paper and paper products 22 - Publishing, printing, and reproduction of recorded media 24 - Manufacture of chemicals and chemical products 25 - Manufacture of rubber and plastics products 26 - Manufacture of other non-metallic mineral products 27 - Manufacture of basic metals 28 - Manufacture of fabricated metal products, except machinery and equipment 29 - Manufacture of machinery and equipment n.e.c. 31 - Manufacture of electrical machinery and apparatus n.e.c. 32 - Manufacture of radio, television and communication equipment and apparatus 34 - Manufacture of motor vehicles, trailers and semi-trailers 36 - Manufacture of furniture; manufacturing n.e.c.
G - Wholesale and retail trade; repair of motor vehicles, motorcycles, and personal and household goods	50 - Sale, maintenance, and repair of motor vehicles and motorcycles; retail sale of automotive fuel 51 - Wholesale trade and commission trade, except of motor vehicles and motorcycles
I - Transport, storage, and communications	63 - Supporting and auxiliary transport activities; activities of travel agencies

Source: The office of National Statistics of Thailand

Commodities were classified using the Harmonized System. The survey divided the commodities into 20 categories as follows.

Table 3.2 Details of CFS data in the Harmonized System

Category	Commodity type
1	Live animals & Animal products
2	Vegetable products
3	Animal or Vegetable Fats
4	Prepared Foodstuffs
5	Mineral Products
6	Chemical Products or allied industries
7	Plastic & Rubber
8	Hides & Skins
9	Wood & articles of wood
10	Wood & Pulp Products
11	Textile & Textile Articles
12	Footwear, Headgear
13	Articles Of Stone, Plaster, Cement, Asbestos
14	Pearls, Precious Or Semi-Precious Stones, Metals
15	Base Metals & Articles Thereof
16	Machinery & Mechanical Appliances
17	Transportation Equipment
18	Instruments - Measuring, Musical
19	Arms and Ammunition; parts and accessories thereof
20	Miscellaneous

Source: The office of National Statistics of Thailand

3.1.2 Timing

The CFS survey period were divided into quarters. The first quarter was from January to March 2007, the second quarter was from April to June 2007, the third quarter was from July to September 2007, and the fourth quarter was from October to December 2007.

3.1.3 Data items details

The CFS collected comprehensive freight transportation data from establishment to destination. The details collected were

- Number of shipments within one week
- Value and weight of shipment
- Shipment type
- Origin and destination of shipment
- Mode of transportation
- Import and export data

3.1.4 Details of CFS Data

NSO (2008) provided a primary data report of CFS in 2008. The report summarizes important characteristic of regional freight flow in Thailand. Unfortunately, the report presents survey result in percentage units, not in weight of shipment or value of shipment, and some freight flow information still undiscovered. Details of the CFS data are listed below.

3.1.4.1 Shipment origin

More than 50% of freight movement originated from manufacturing establishments, followed by mining and quarrying (30.7%), then wholesale and retail trade (11.3%), then warehouses (0.8%).

Table 3.3 Details of CFS shipment origin

Origin	Q1	Q2	Q3	Q4	Total
Mining and quarrying	10.0	27.4	42.8	40.4	30.7
Manufacture	75.7	59.0	46.9	49.0	57.2
Wholesale and Retail trade	13.1	12.4	9.7	10.4	11.3
Warehouse	1.2	1.2	0.6	0.2	0.8

Source: The office of National Statistics of Thailand

3.1.4.2 Shipment destination

The top three freight destinations are manufacturing establishments (42.0%), mining retail trade (24.8%), and wholesale (19.8%). Two types of transportation hub, marine port and airport, have a 2.9% proportion. The least common freight destination is border crossing (0.2%).

Table 3.4 Details of CFS shipment destination

Destination	Q1	Q2	Q3	Q4	Total
Manufacturer	42.1	43.1	42.0	41.1	42.0
Retail trade	23.9	27.1	25.5	23.0	24.8
Wholesale	21.0	18.5	18.0	21.3	19.8
Agriculture	3.9	2.0	5.0	4.4	3.8
Marine port	1.9	1.7	1.2	1.6	1.6
Airport	1.0	1.0	1.4	1.8	1.3
Mining and quarrying	0.6	1.0	1.3	0.6	0.9
Border crossing	0.1	0.2	0.2	0.2	0.2
Others	5.5	5.4	5.4	6.0	5.6

Source: The office of National Statistics of Thailand

3.1.4.3 CFS outbound shipment weight

Characteristics of shipment movement from origin are shown in the table below. Mineral Products is the highest proportion of movement, at approximately 34.4%, followed by Prepared Foodstuffs (15.2%), then Articles of Stone, Plaster, Cement, Asbestos (12.7%). Some categories of commodity not reported are

- Hides & Skins
- Footwear, Headgear
- Pearls, Precious or Semi-Precious Stones, Metals
- Instruments - Measuring, Musical
- Arms and Ammunition; parts and accessories thereof

Table 3.5 CFS outbound shipment weight characteristics by commodity description

Category	Commodity Description	Percent
1	Live animals & Animal products	0.9
2	Vegetable products	6.2
3	Animal or Vegetable Fats	1.0
4	Prepared Foodstuffs	15.2
5	Mineral Products	34.4
6	Chemical Products or allied industries	7.7
7	Plastic & Rubber	4.8
8	Hides & Skins	-
9	Wood & articles of wood	2.2
10	Wood & Pulp Products	3.1
11	Textile & Textile Articles	0.1
12	Footwear, Headgear	-
13	Articles of Stone, Plaster, Cement, Asbestos	12.7
14	Pearls, Precious or Semi-Precious Stones, Metals	-
15	Base Metals & Articles Thereof	8.6
16	Machinery & Mechanical Appliances	2.0
17	Vehicles, aircraft, vessels and associated transport Equipment	0.6
18	Instruments - Measuring, Musical	-
19	Arms and Ammunition; parts and accessories thereof	-
20	Miscellaneous	0.5

Source: The office of National Statistics of Thailand

3.2 Department of Land Transport roadside survey data (RS1)

The Department of Land Transport launched "the study of freight transportation of the road by truck survey" project in 2008 and delegated it to The Transportation Institute of Chulalongkorn University. The project was completed in 2009. The survey gathered the origin and destination of freight carried by truck on highways in 10 major provinces of Thailand. Data items include

- Weight and value of shipment
- 37 commodity types
- Origin and destination of transport
- Truck types
- Time of shipping
- Type of transport (private or hired)

Details of the survey are as follows.

3.2.1 Survey station

The survey stations were located on major highway links which connect important manufacturing areas, agricultural production areas, and major consumption areas. A number of survey stations were located at adjacent provinces to Bangkok (Ayudhaya, Rajaburi, Nonthaburi, Samutprakarn, and Chacheangchao) to collect flow into and out of Bangkok. Comprehensive flow data was collected at Chonburi province in which the important marine port, Lam Chabang port, is located. Moreover, survey stations were located at major provinces in all parts of Thailand, consisting of Chaingmai, Nakonsawan, Khonkean, Nakonrajchima, Chumpon, and Songkha. Details of the survey stations are shown in Figure 3.1 and Table 3.6.

Table 3.6 Details of survey stations

Num	Province	Region	Number of survey station
1	Bangkok	Central	8
2	Chonburi	Eastern	3
3	Prachinburi	Eastern	1
4	Ratchaburi	Western	1
5	Khonkaen	North Eastern	3
6	Nakhonratchasima	North Eastern	3
7	Chiang Mai	Northern	3
8	Nakhonsawan	Northern	4
9	Chumphon	Southern	2
10	Songkhla	Southern	3

Source: Department of Land Transport

3.2.2 Timing

Due to budget limitations, this project collected data in the harvest season and out of the harvest season. The 24-hour data were collected at all survey stations. Since characteristics of freight movement in Bangkok and Chonburi are different on weekdays than weekends, data for Bangkok and Chonburi were collected on both weekdays and weekends. The survey provides comprehensive regional freight flow data, and reveals the transport carried by the buyers of the goods which cannot be represented in the CFS data.

Fig 3.1 RS1 survey station
Source: Department of Land Transport

3.2.3 Commodities classification

Although the Harmonized System is a freight classification system which is accepted throughout the world, it isn't appropriate for freight transportation and analysis in Thailand, where agricultural products are a major commodity with high shipment volumes per year. Instead, the Transportation Institute of Chulalongkorn University provided new commodities classification for the project. Commodities were classified into 37 categories as shown below.

Table 3.7 Details of RS1 commodities classification

Category	Commodity type
1	Paddy rice
2	Corn
3	Rice
4	Chemical
5	Machinery
6	Food
7	Consumer goods
8	Electronics
9	Flowers and trees
10	Soil stone sand
11	Coal
12	Sugar
13	Fuel
14	Fertilizer
15	Cement
16	Flour
17	Other agricultural products
18	Paper products
19	Wood products
20	Plastic products
21	Rubber products
22	Vegetables and fruits
23	Cassava
24	Wood
25	Para rubber
26	Automobile
27	Mineral
28	Metal and nonmetal
29	Construction material
30	Wood fuel and Agricultural residue
31	Aquatic animals
32	Live animals
33	Textiles
34	Steel
35	Sugarcane
36	Veterinary food
37	Other

Source: Department of Land Transport

3.2.4 Details of RS1 Data

The Department of Land Transport provided its final report in 2009. The report summarized important characteristics of freight movement by truck in Thailand as explained below.

3.2.4.1 Shipment volume per year

The survey shows that the total volume of freight transportation in Thailand was 469,369,365.53 ton per year. The top three items shipped per year by volume were soil stone sand (55,772,446.35 tons), food (36,999,111.19 tons), and cement (35,778,074.34 tons). Rice was the highest proportion in the agricultural product category. A question arose with the volume of paddy rice (6,968,652.90) which is a raw product of rice (21,364,826.85), and also with sugar cane (324,525.77), which is a raw product of sugar (11,068,491.42). The report explained that the survey collected inter-regional and long trip transportation while paddy rice and sugar cane are shipped from crop fields to the vicinity for manufacture, and therefore this types of shipment will be under-reported.

Table 3.8 Shipment volume per year

Category	Commodity type	Volume (Tons per year)
1	Paddy rice	6,968,652.90
2	Corn	6,716,448.62
3	Rice	21,364,826.85
4	Chemical	10,671,260.63
5	Machinery	2,913,628.42
6	Food	36,999,111.19
7	Consumer goods	10,126,992.45
8	Electronics	6,295,500.37
9	Flowers and trees	1,635,736.27
10	Soil stone sand	55,772,446.35
11	Coal	4,331,103.41
12	Sugar	11,068,491.42
13	Fuel	26,531,149.30
14	Fertilizer	10,292,075.37
15	Cement	35,778,074.34
16	Flour	5,412,767.87
17	Other agriculture products	3,661,417.61
18	Paper products	12,234,140.47
19	Wood products	3,846,097.71
20	Plastics products	15,542,165.37
21	Rubber products	2,797,497.17
22	Vegetables and fruits	12,920,419.20
23	Cassava	4,436,797.25
24	Wood	26,141,557.36
25	Para rubber	7,501,794.02
26	Automobile	13,652,349.73
27	Mineral	7,440,633.46
28	Metal and nonmetal	5,449,717.48
29	Construction material	18,937,842.29
30	Wood fuel and Agricultural residue	7,281,974.71
31	Aquatic animals	5,690,403.04
32	Live animals	2,819,054.86

Category	Commodity type	Volume (Tons per year)
33	Textiles	3,827,477.68
34	Steel	24,917,582.07
35	Sugarcane	324,525.77
36	Veterinary food	19,518,266.12
37	Other	17,549,386.39
	Total	**469,369,365.53**

Source: Department of Land Transport

3.2.4.2 Shipment volume of 10 target provinces

The total volume of freight transported in the 10 target provinces is 382,687,966.60 tons per year. Chonburi province has the highest volume in both origin (47,642,069.06) and destination (62,961,790.28), followed by Bangkok with 35,217,439.31 for origin and 58,322,759.03 for destination, then Nakonrajchasima (23,055,461.93 for origin, 21,108,455.78 for destination). The values of volume from origin and volume to destination were used to classify type of province. A province with higher volume from origin than volume to destination was classified as a production province. A province with a lower volume from origin than volume to destination was classified as a consumption province. Where the two volumes were nearly identical, the province was classified as a pass through province. The details are

- Production provinces are Prachinburi, Rajchaburi, and Nakonrajchasima
- Consumption provinces are Bangkok, Chiang Mai, Songkla, and Chonburi
- Pass through provinces are Khonkean, Nakonsawan, and Chumpon

Table 3.9 Shipment volumes of 10 target provinces

Num	Province	Volume (Tons per year)	
		Origin	Destination
1	Bangkok	35,217,439.31	58,322,759.03
2	Chonburi	47,642,069.06	62,961,790.28
3	Prachinburi	5,769,194.86	4,869,823.00
4	Ratchaburi	17,508,193.20	8,656,332.73
5	Khonkaen	11,286,715.19	11,374,984.68
6	Nakhonratchasima	23,055,461.93	21,108,455.78
7	Chiang Mai	8,232,357.75	11,379,247.53
8	Nakhonsawan	9,117,320.36	8,410,687.78
9	Chumphon	3,762,605.38	3,360,044.02
10	Songkhla	11,443,570.97	19,208,913.76
	Total	**173,034,928.01**	**209,653,038.59**

Source: Department of Land Transport

3.2.5 Limitations of the data

The final report of RS1 provided awareness concerning limitations to the data.

"The objective of the survey is to collect data to support carriers of the Department of Land Transport which manages freight transport into and out of 10 target provinces. Thus, the survey station may be located on major highways within or outside the target provinces. For this reason, the survey cannot capture all freight transportation within the whole kingdom, especially some kinds of freight transport including

1. Short distance shipments between districts within a province that do not pass through survey stations which are located central to the province.
2. Shipments in rural areas such as shipment from crop areas to manufacturers in the vicinity.
3. Shipments with origins and destinations outside the 10 target provinces. The survey in this project will capture this type of shipment by chance and cannot capture shipments which do not pass through survey stations."

The report categorizes the reliability of the data into three categories:

1. High confidence**:** origin and destination of shipment located within the 10 provinces
2. Moderate confidence: origin or destination of shipment located within the 10 provinces
3. Low confidence: origin and destination of shipment located in other provinces

Table 3.10 Reliability of the RS1 data

	Destination province	
Origin province	Transportation within the 10 target provinces High confidence data	Transportation from within the 10 target provinces to other provinces Moderate confidence data
	Transportation from other provinces to the 10 target provinces Moderate confidence data	Transportation within other provinces Low confidence data

Source: Department of Land Transport

3.3 Roadside Survey data from Chonburi road network strategic planning for supporting logistics development project (RS2)

The Transportation institute of Chulalongkorn University under contract of Chonburi province government launched the Chonburi road network strategic planning for supporting logistics development project in 2009. An importance part of the project is to collect O-D of truck by roadside interview method on the highway in Chonburi provinces. Due to budget and time limitation, this project collects data at 13 selected sites on major highway in Chonburi province. Data items include

- Weight and value of shipment
- 16 commodity types
- Origin and destination of transport
- Truck type
- Time of shipping
- Type of transport (private or hired)

3.3.1 Survey station

The project located 13 survey stations on major highway in Chonburi province. Since, the objective of the project is to model freight transportation in Chonburi province, the analysis zone was divided into subdistrict area. Survey stations were located between urbanized area of Chonburi such as between Maueng district and Banglamung district.

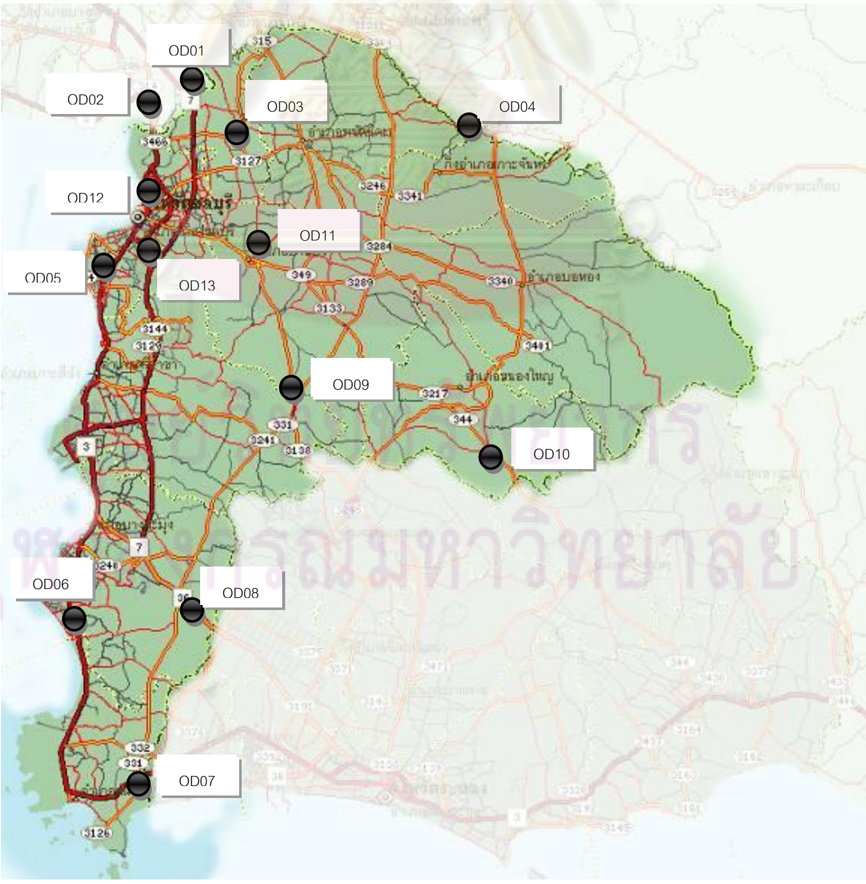

Fig 3.2 RS2 survey station
Source: Chonburi province government

3.3.2 Timing

The data were collected from June 2009 through October 2009. The 24-hour data were collected at all survey stations, and a number of survey stations collected data on both weekdays and weekends.

3.3.3 Commodities classification

Like the RS1 data, the Harmonized System is not appropriate for freight transportation analysis and modeling in Chonburi. Thus, RS1 Classification was not used in this project. This is because transportation behavior in Chonburi has high volumes of freight which is related to chemicals and manufacturing. Moreover, import and export by container is important. New commodities classification for the project was used. Commodities were classified into 15 categories as shown below.

Table 3.11 Details of RS2 commodities classification

Category	Commodity type
1	Gas
2	Chemical
3	Fuel
4	Food
5	Cement
6	Vegetables and fruits
7	Plastics
8	Crop
9	Electronics
10	Wood
11	Automobiles and parts
12	Construction material
13	Stone soil sand
14	Steel and other metals
15	Others
16	Commodities

Source: Chonburi province government

3.3.4 Details of RS2 data

The Transportation Institute of Chulalongkorn University released its final report in late 2009. The report shows the freight transportation data of Chonburi province. The summarized data follows.

3.3.4.1 Shipment volume per year

The total volume of freight transportation through the province is 82,233,689.41 tons per year. The top three items were other commodities (14,952,007.61 tons), chemicals (8,151,087.45 tons), and food (6,960,108.94 tons).

Table 3.12 RS2 Shipment volume per year

Category	Commodity type	Volume (Tons per year)
1	Gas	4,937,030.87
2	Chemical	8,151,087.45
3	Fuel	6,766,208.37
4	Food	6,960,108.94
5	Cement	3,041,106.72
6	Vegetables and fruits	4,046,136.64
7	Plastics	4,287,879.24
8	Crops	2,394,461.83
9	Electronics	1,111,482.66
10	Wood	2,562,413.68
11	Automobiles and parts	4,000,701.30
12	Construction material	3,034,525.62
13	Stone soil sand	7,719,664.18
14	Steel and others metal	6,896,563.75
15	Others	14,952,007.61
16	Consumer goods	1,372,310.49
	Total	**82,233,689.41**

Source: Chonburi province government

Freight flow data was collected from roadside interview data in this project. It consisted of four types of movement:

- Within Chonburi
- Into Chonburi
- Out of Chonburi
- Pass through Chonburi

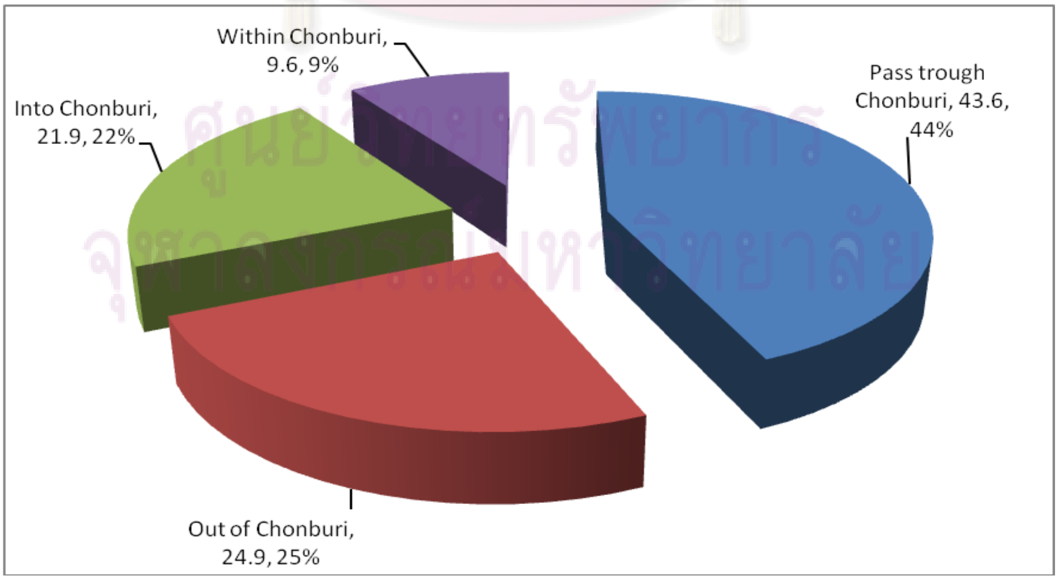

Fig 3.3 Proportion of freight transportation data

The highest proportion of the collected data passed through Chonburi (43.6%), followed by outs of Chonburi (24.9%) and within Chonburi (9.6%). Since Chonburi is located between Bangkok and Rayong, an important industrial province, the survey captured a high volume of trips passing through Chonburi between Bangkok and Rayong.

Table 3.13 Shipment into, out of, and within Chonburi volume per year

Category	Commodity type	Volume (tons per year)		
		Out of	Into	Within
1	Gas	2,024,351.38	878,774.80	367,740.41
2	Chemical	1,304,776.99	1,929,663.67	373,925.18
3	Fuel	2,609,721.22	1,010,091.76	509,606.06
4	Food	2,493,919.09	2,683,456.69	1,321,618.50
5	Cement	852,644.89	930,608.89	726,196.93
6	Vegetables and fruits	891,842.67	1,556,183.67	134,307.55
7	Plastics	239,290.00	741,576.15	44,166.60
8	Crops	927,617.61	1,138,962.60	152,143.64
9	Electronics	383,797.06	520,740.27	239,222.36
10	Wood	552,872.07	1,290,523.18	243,591.00
11	Automobiles and parts	1,704,047.35	1,601,578.54	493,627.20
12	Construction material	1,670,294.51	1,342,123.03	740,238.05
13	Stone soil sand	5,193,265.45	2,280,358.90	1,446,512.38
14	Steel and other metals	1,845,057.20	1,329,980.06	449,339.32
15	Others	5,179,247.66	6,108,094.72	1,606,586.90
16	Consumer goods	742,056.03	485,286.42	169,390.77
	Total	**28,614,801.18**	**25,828,003.37**	**9,018,212.84**

Source: Chonburi province government

3.3.4.2 Average trip length

The average trip length into Chonburi was near the trip length out of Chonburi. Stone soil sand has the shortest average trip lengths while gas has the longest distance for out of Chonburi trips. Average trip lengths show that commodities from Chonburi which were distributed to provinces in the vicinity traveled around 200 kilometer.

Table 3.14 RS2 Average trip length

Category	Commodity type	Average trip length (Km)		
		Out of	Into	Within
1	Gas	281.51	67.31	13.38
2	Chemical	123.17	90.39	17.25
3	Fuel	266.88	134.42	17.47
4	Food	89.71	116.11	17.51
5	Cement	64.46	96.14	20.65
6	Vegetables and fruits	154.54	330.48	21.32
7	Plastics	79.91	80.53	18.65
8	Crops	71.50	130.51	15.74

Category	Commodity type	Average trip length (Km)		
		Out of	Into	Within
9	Electronics	122.09	84.83	13.37
10	Wood	93.26	168.01	17.52
11	Automobiles and parts	124.86	76.36	15.05
12	Construction material	143.97	95.31	19.89
13	Stone soil sand	63.56	69.79	23.23
14	Steel and others metal	109.74	81.23	21.77
15	Others	124.02	119.18	23.95
16	Commodity	132.47	131.81	21.68
	Total	**113.69**	**113.73**	**17.64**

Source: Chonburi province government

3.4 Comparing CFS and Roadside Survey Data (RS1)

The CFS report provided by NSO presented results using percentage of shipment volume, making them difficult to compare with data. Fortunately, under MOU between the Transportation Institute of Chulalongkorn University and NSO, NSO provided raw data to the Transportation Institute of Chulalongkorn University for analysis and research. According to the CFS final reports, shipment volume was reported by percentage, so comparing was undertaken by percentage. The following is a discussion of the comparison.

3.4.1 Volume from origin

The CFS shows that Bangkok is the province that generates the most trips. Next is Chumpon. Meanwhile, RS1 shows that Chonburi generates the most trips. The difference in sampling may have caused this difference. The CFS captures trip data from origin by sampling at the source of the trip, whereas RS1 captures trip data on roads around the province.

Table 3.15 Comparison of volume from origin

Province	Volume (tons per year)	
	CFS	RS1
Bangkok	44.79	20.35
Chonburi	2.89	27.53
Prachinburi	2.49	3.33
Ratchaburi	5.13	13.32
Khonkaen	7.31	6.52
Nakhonratchasima	4.81	4.76
Chiang Mai	10.34	5.27
Nakhonsawan	1.63	10.12
Chumphon	15.54	2.17
Songkhla	5.06	6.61
Total	**100.00**	**100.00**

3.4.2 Volume to destination

The CFS reports that more trips go to Bangkok than any other province; 54.54% of freight was shipped to Bangkok. RS1, meanwhile, shows that more trips go to Chonburi than to any other province. Both CFS and RS1 place both Bangkok and Chonburi in the top three destinations.

Table 3.16 Comparison of volume to destination

Province	Volume (tons per year)	
	CFS	RS1
Bangkok	54.45	27.82
Chonburi	12.10	30.03
Prachinburi	2.78	2.32
Ratchaburi	5.26	10.07
Khonkaen	5.64	5.43
Nakhonratchasima	5.14	5.43
Chiang Mai	4.71	4.01
Nakhonsawan	5.20	4.13
Chumphon	2.66	1.60
Songkhla	2.07	9.16
Total	**100.00**	**100.00**

3.4.3 Volume by commodity type

The CFS reports that construction material has the highest shipping volume, while RS1 reports that soil stone and sand have the highest shipping volume. As mentioned earlier, CFS reports that manufacturing establishments ship approximately 69.6% of all materials, and mining and quarrying approximately 1.9%. These proportions may be causing the high volume shipment of construction material.

Table 3.17 Comparison of volume by freight type

Num	Commodity type	Percent	
		CFS	RS1
1	Paddy rice	0.19	1.48
2	Corn	0.20	1.43
3	Rice	3.36	4.55
4	Chemical	3.27	2.27
5	Machinery	0.82	0.62
6	Food	8.04	7.88
7	Consumer goods	1.06	2.16
8	Electronics	1.33	1.34
9	Flowers and trees	0.03	0.35
10	Soil stone sand	0.21	11.88
11	Coal	0.08	0.92
12	Sugar	4.20	2.36
13	Fuel	0.27	5.65

Num	Commodity type	Percent	
14	Fertilizer	1.29	2.19
15	Cement	1.54	7.62
16	Flour	1.32	1.15
17	Other agricultural products	0.34	0.78
18	Paper products	3.17	2.61
19	Wood products	0.86	0.82
20	Plastics products	3.27	3.31
21	Rubber products	0.51	0.60
22	Vegetables and fruits	0.22	2.75
23	Cassava	0.50	0.95
24	Wood	1.30	5.57
25	Para rubber	1.03	1.60
26	Automobile	0.62	2.91
27	Mineral	0.04	1.59
28	Metal and nonmetal	0.90	1.16
29	Construction material	46.56	4.03
30	Wood fuel and Agricultural residue	0.04	1.55
31	Aquatic animals	0.37	1.21
32	Live animals	0.02	0.60
33	Textiles	0.09	0.82
34	Steel	7.95	5.31
35	Sugarcane	0.06	0.07
36	Veterinary food	4.47	4.16
37	Other	0.49	3.74
Total		**100.00**	**100.00**

3.4.4 Average trip length

CFS shows lower transportation distances than RS1. This is because CFS captures short distance shipment (nearby province transportation) while RS1 captures long distance shipment (region to region transportation). The two sources capture different samples which complement each other. Combining CFS and RS1 reveals complete freight flow matrices.

Table 3.18 Average trip length

Num	Commodity type	Average trip length (Km)	
		CFS	**RS1**
1	Paddy rice	107.50	223.54
2	Corn	226.36	396.50
3	Rice	252.33	371.63
4	Chemical	93.94	267.21
5	Machinery	96.42	239.51
6	Food	121.99	297.09
7	Consumer goo	122.20	361.57
8	Electronics	128.57	325.61
9	Flowers and tree	195.33	357.43

Num	Commodity type	Average trip length (Km)	
10	Soil stone sand	112.20	117.88
11	Coal	62.89	366.20
12	Sugar	236.20	401.36
13	Fuel	115.35	296.01
14	Fertilizer	218.58	384.42
15	Cement	70.99	252.36
16	Flour	259.31	352.86
17	Other agricultural products	102.14	427.09
18	Paper products	86.38	238.73
19	Wood products	164.21	334.69
20	Plastics products	103.91	220.98
21	Rubber products	141.32	263.45
22	Vegetables and fruits	245.51	382.03
23	Cassava	129.80	225.81
24	Wood	234.85	358.99
25	Para rubber	133.26	305.84
26	Automobile	124.45	257.32
27	Mineral	148.28	336.43
28	Metal and nonmetal	100.91	280.18
29	Construction material	90.85	276.36
30	Wood fuel and Agricultural residue	177.07	408.37
31	Aquatic animals	280.10	431.25
32	Live animal	132.21	243.19
33	Textiles	139.44	278.85
34	Steel	96.94	256.81
35	Sugarcane	75.80	121.92
36	Veterinary food	196.08	313.34
37	Other	116.17	333.52
Total		**128.37**	**301.62**

3.4.5 Freight Origin

Most commodity types have higher numbers of origins in CFS data than in RS1 data. This is because CFS capture data spreads across the country; its samples were selected from all provinces, whereas RS1 focused only on the 10 target provinces. The only numbers which are close are food, with 76 origins according to CFS and 74 origins according to RS1.

Table 3.19 Comparison of number of origin

Num	Commodity type	Number of origin	
		CFS	RS1
1	Paddy rice	60	39
2	Corn	42	43
3	Rice	64	64
4	Chemical	50	48
5	Machinery	63	41
6	Food	76	74

Num	Commodity type	Number of origin	
7	Consumer good	64	69
8	Electronics	46	63
9	Flowers and trees	25	64
10	Soil stone sand	41	59
11	Coal	6	27
12	Sugar	39	49
13	Fuel	26	61
14	Fertilizer	48	55
15	Cement	54	58
16	Flour	35	37
17	Other agricultural products	43	53
18	Paper products	67	66
19	Wood products	62	69
20	Plastics products	73	58
21	Rubber products	51	41
22	Vegetables and fruits	53	72
23	Cassava	45	37
24	Wood	74	74
25	Para rubber	27	45
26	Automobile	49	58
27	Mineral	25	43
28	Metal and nonmetal	64	51
29	Construction material	75	63
30	Wood fuel and Agricultural residue	48	60
31	Aquatic animals	40	52
32	Live animals	18	60
33	Textiles	30	55
34	Steel	69	60
35	Sugarcane	39	12
36	Veterinary food	74	68
37	Other	71	73

3.4.6 Freight destinations

CFS reported nine commodities types which were distributed to all the provinces: machinery, food, consumer goods, electronics, paper products, rubber products, construction material, steel, and others. CFS and RS1 reported quite different numbers of destinations, but the numbers were very close for Food and for Consumer Goods.

Table 3.20 Comparison of number of destinations

Num	Commodity type	Number of destinations	
		CFS	RS1
1	Paddy rice	60	39
2	Corn	42	43
3	Rice	75	62
4	Chemical	75	56

Num	Commodity type	Number of destinations	
5	Machinery	76	48
6	Food	76	75
7	Consumer goods	76	76
8	Electronics	76	64
9	Flowers and trees	61	56
10	Soil stone sand	49	58
11	Coal	12	24
12	Sugar	70	48
13	Fuel	59	73
14	Fertilizer	74	72
15	Cement	68	73
16	Flour	52	44
17	Other agricultural products	53	50
18	Paper products	76	61
19	Wood products	75	62
20	Plastics products	76	68
21	Rubber products	76	47
22	Vegetables and fruits	57	72
23	Cassava	41	29
24	Wood	74	67
25	Para rubber	35	42
26	Automobile	75	72
27	Mineral	34	42
28	Metal and nonmetal	74	57
29	Construction material	76	74
30	Wood fuel and Agricultural residue	45	52
31	Aquatic animals	60	64
32	Live animals	41	56
33	Textiles	71	61
34	Steel	76	69
35	Sugarcane	25	12
36	Veterinary food	72	67
37	Other	76	75

3.5 Comparing RS1 and RS2 data

RS2 collected freight flow data from Chonburi province, which was the target province of RS1. For this reason, Chonburi has freight flow data from two sources. However, RS1 and RS2 used different survey stations and had different objectives, so the purpose of this comparison is to verify agreement between the two sources of data.

3.5.1 Volume per year

The two data sources used different commodity classifications. There are 12 commodity types which are reported by the two sources of data. However, RS2 collected data only one time, so seasonal effects may result in misleading data. For this reason, a comparison of the two data sources deals only with freight that is not seasonally affected. Details are shown below.

3.5.1.1 Out of Chonburi province

There are very large differences between RS1 and RS2 for plastic, Electronics, and automobile, while food has a large difference.

Table 3.21 Comparison of volume from Chonburi

Num	Commodity type	Volume (tons per year)		Diff (%)
		RS1	RS2	
1	Food	1,281,534.06	1,172,300.59	9
2	Plastic	549,149.96	195,123.40	64
3	Electronics	469,408.42	144,574.70	69
4	Automobile	1,465,578.03	1,210,420.16	17
5	Construction material	1,047,644.21	930,056.46	11
6	Consumer goods	509,896.18	572,665.26	-12

3.5.1.2 Into Chonburi province

There is quite a large difference between RS1 and RS2 for food, while plastics also and Construction material have large differences.

Table 3.22 Comparison of volume to Chonburi

Num	Commodity type	Volume (tons per year)		Diff (%)
		RS1	RS2	
1	Food	1,281,534.06	1,361,838.19	-6
2	Plastic	549,149.96	697,409.55	-27
3	Electronics	469,408.42	281,517.90	40
4	Automobile	1,465,578.03	1,107,951.35	24
5	Construction material	1,047,644.21	601,884.98	43
6	Consumer goods	509,896.18	315,895.65	38

3.5.1.3 Within Chonburi province

There are large differences between RS1 and RS2, except for electronics and automobile. The results indicate that RS1 captured trips within the province less than RS2 due to the number of survey stations. RS2 had 13 survey stations and some stations located between urbanize area whereas RS1 had three survey stations located on major highways into and out of the province.

Table 3.23 Comparison of volume within Chonburi

Num	Commodity type	Volume (tons per year)		Diff (%)
		RS1	RS2	
1	Food	384,866.53	1,321,618.50	-243
2	Plastic	69,652.58	44,166.60	37
3	Electronics	326,225.97	239,222.36	27
4	Automobile	671,903.80	493,627.20	27
5	Construction material	155,648.20	740,238.05	-376
6	Consumer goods	501,654.43	169,390.77	66

3.5.2 Average distance

3.5.2.1 Out of Chonburi province

Commodities from Chonburi were shipped to nearby provinces less than 200 kilometers. Most commodities in RS1 had average trip lengths more than RS2. These results show that RS1 captured long trips while RS2 captured short, moderate, and long trips.

Table 3.24 Comparison of average trip length from Chonburi

Category	Commodity type	Average trip length (Km)	
		RS1	RS2
1	Food	195.14	116.11
2	Plastics	148.12	80.53
3	Electronics	154.33	84.83
4	Automobiles and parts	128.08	76.36
5	Construction material	193.42	95.31
6	Commodities	196.29	131.81

3.5.2.2 Into Chonburi province

Most commodities of RS1 had average trip lengths more than RS2, except Automobiles and parts. These results show that RS1 captured long trips while RS2 captured short, moderate, and long trips.

Table 3.25 Comparison of average trip length to Chonburi

Category	Commodity type	Average trip length (Km)	
		RS1	RS2
1	Food	193.10	89.71
2	Plastics	121.83	79.91
3	Electronics	208.57	122.09
4	Automobiles and parts	111.50	124.86
5	Construction material	161.98	143.97
6	Consumer goods	162.95	132.47

3.5.2.3 Within Chonburi province

RS1 divided analysis zones whereas RS2 used subdistricts. As a result of these very different methods, RS1 was unable to measure exact trip lengths for trips within the district. Thus, RS1 used 10 kilometers as trip length distance for within district trips.

Table 3.26 Comparison of average trip length within Chonburi

Category	Commodity type	Average trip length (Km)	
		RS1	RS2
1	Food	10.00	17.51
2	Plastics	10.00	18.65
3	Electronics	10.00	13.37
4	Automobiles and parts	10.00	15.05
5	Construction material	10.00	19.89
6	Consumer goods	10.00	21.68

3.6 Summary and discussion

As mentioned earlier, there are three sources of comprehensive freight flow data available in Thailand: CFS, RS1, and RS2. However, the pioneering Thailand CFS has a number of weak points, especially marginal totals, while RS1 captures data at only 10 target provinces out of 76 provinces in Thailand. These two data sources are still incomplete and should be adjusted. Since RS2 used survey stations to capture short, moderate, and long trips within Chonburi province, it captured preferable freight distribution characteristics, as did CFS. For this reason, RS2 is used to verify the developed adjusting data method.

In order to develop a method for adjusting the data, some category data was selected using the following criteria.

1. The commodity should originate from 76 provinces and ship to all provinces.
2. The commodity must originate from Chonburi province.
3. The volume reported by RS1 and RS2 must not be extremely different.

Three freight types meet the first two criteria: food, construction material, and consumer goods. However, RS1 and RS2 showed very different volumes of construction material shipped within Chonburi, so construction material is not appropriate. The food category also has a large volume difference in RS1 and RS2, but a combined consumer goods and food freight category reduce this gap. For this reason, a combined consumer goods and food freight category will be used in the next step.

Chapter IV
Methodology

This chapter presents a comprehensive framework to identify the research methodology, research approach, and detailed methodologies.

4.1 Research methodology

Achieving the proposed main objective involves developing a method to combine the data, and applying ANFIS to enhance the freight O-D distribution database from CFS and truck O-D by roadside surveys in Thailand. The framework of this research will relate to many databases, analysis tools, modeling techniques, and analytical computer program such as MATLAB. The framework of this research is divided into two components, data preparation and modeling.

The first component, data preparation, must be separated into two tasks, compiling commodity flow survey (CFS) data and compiling roadside interview data. The CFS data from NSO will be transferred to an analysis format and re-categorized from the Harmonized System to the Transportation Institute of Chulalongkorn University classification system. Next, data from roadside surveys will be verified against CFS for sample size to ensure that origin and destination data come from the same establishment category. After that, a method to combine the data will be developed, and this method will be applied to develop the combined data. The next step is to verify and augment "true zeros" in combining data using roadside survey data. The objective of this step is to fill any empty cells from CFS which were reported in roadside survey, since those will not be "true zeros" and must be filled with roadside survey data.

The second component is to calibrate three models: the gravity model, the regression based model, and the ANFIS model. A single constraint gravity model using zone total and distance as input data will be used as a branch marks model. The regression based model using socio economic data of traffic analysis zones such as population, employment will be calibrated to explore the influence of these input data. After that, the influential input data will be used as input data for the ANFIS model. The performance of the ANFIS model will be evaluated against the gravity model and the regression based model. With the best developed ANFIS model, freight O-D matrices will be developed.

The final step of this research is the conclusion and recommendations. The overall research approach including processes and activities are shown in Figure 4.1.

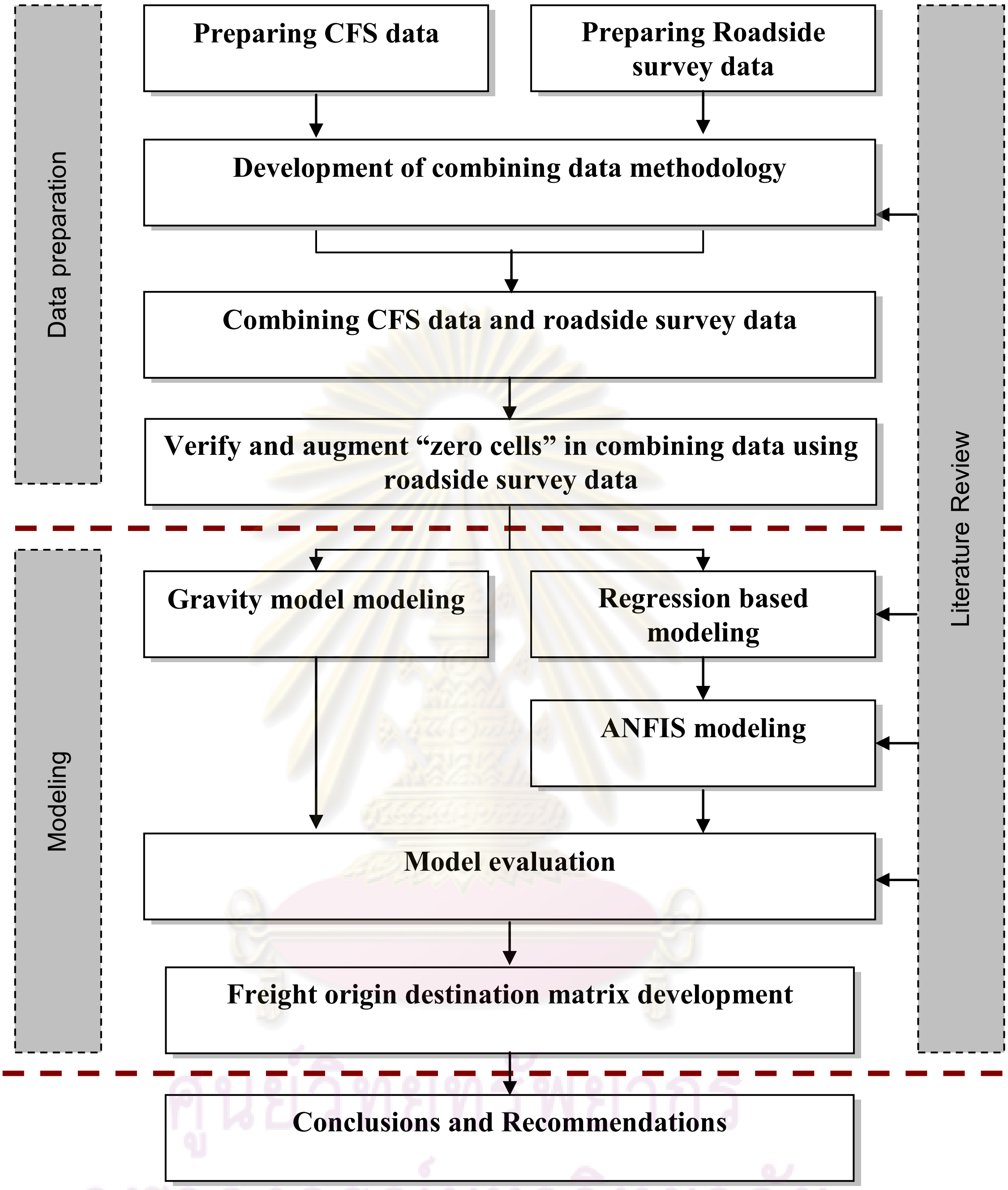

Figure 4.1 Analytical framework

4.2 Detailed methodology

4.2.1 Preparing CFS data

CFS collected Harmonized-based commodity types, which are not appropriate for freight transportation analysis and modeling in Thailand. Thus this research re-categorized the raw data using a commodity classification system provided by the Transportation Institute of Chulalongkorn University and which classified the commodities into 37 categories. After that, a combined food and consumer goods category was extracted from the main database for the next step.

4.2.2 Preparing roadside survey data

Due to the collection method of the roadside survey, the trips that passed the survey stations were randomly selected for interview. Thus, trips outside the scope of the CFS sample may have been selected. For this reason, freight data outside the scope of CFS origin will be excluded from analysis.

4.2.3 Development of combining data methodology

The combining method uses the strengths of its two data sources, the marginal total of RS1 and the distribution pattern of CFS, to produce an adjusted matrix. Two methods will be developed, Trip Length Distribution Adjusting method (TLDA) and Gravity Model Approach method (GMA). TLDA will adjust CFS trip length distribution to meet RS1 or RS2 marginal total, while GMA will adjust RS1 origin-destination data using friction impedances from CFS. The details of these two methods will be discussed the next chapter.

4.2.4 Combining CFS data and roadside survey data

The developed method will be applied to the combined food and consumer goods category to produce an adjusted origin destination matrix.

4.2.5 Verify and augment "Zero cells" in combining data using Roadside survey data

Since combined data uses the CFS distribution pattern, empty cells still remain. To verify the empty cells, the adjusted data will be compared to the roadside survey data. These two datasets will be compared for those particular cells to verify that neither of those datasets contradicts the true-zero state. In cases of contradiction, i.e. where roadside observations are found for CFS empty cells, the cell will be filled with roadside survey data.

4.2.6 Gravity model modeling

The gravity models will be used as a benchmark for evaluating the performances of the ANFIS model. A single constraint gravity model using zone total and distance as input variables will be employed. Since a zero cell is a major problem for calibrating a model with the regression technique, and the zero cell problem of the database cannot be avoided, maximum likelihood techniques will be employed.

4.2.7 Regression based modeling

The regression based models will be used to screen input variables which be used as input variables for the ANFIS model and for benchmarking as gravity model. Many groups of input variables such as population, employment, crop area, industrial production power, etc. will be evaluated. Box-Cox transformation will be used as the transformation function and the maximum likelihood technique will be used to calibrate the model. The output of the model will reveal the relationship of explanatory factors and socio economic variables of the study area which are important for setting up the rules for the ANFIS model.

4.2.8 ANFIS modeling

The ANFIS model will be developed using the procedures of the MATLAB Fuzzy logic toolbox (The MathWorks Inc., 2004). The structure of the ANFIS model consists of a Sugeno type fuzzy system with generalized bell input membership functions and a linear output membership function. A number of significant input variables from the regression base model will be used as input variables for the ANFIS model to ensure that the input variables have strong correlation with the output. The hybrid training algorithm consisting of a gradient descent and least squares estimation for the adjustment of premise and consequent parameters of the ANFIS will be used in this study.

4.2.9 Model evaluation

Three performance indicator are selected for validating developed model:

The root mean square error, RMSE, is calculated by:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(e_i - p_i)^2} \quad \text{For } i = 1 \text{ to } n \qquad (4.1)$$

where e_i is the actual value from experiments

p_i is the predicted value by models

N is the number of data points

The mean relative error, MRE, is calculated by

$$MRE(\%) = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{p_i - e_i}{e_i} x100\right| \qquad (4.2)$$

where e_i is actual value from experiments

p_i is the predicted value by models

N is the number of data points

Correlation coefficient, R, is

The R-square is given by the following

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad \text{For } i = 1 \text{ to } n \qquad (4.3)$$

4.2.10 Freight O-D matrices development

The best model from previous step will be employed to develop the origin and destination matrix of combined food and consumer goods in Thailand. The model will be applied to empty cells data for filling gap in the whole matrix.

4.2.11 Conclusion and Recommendations

The last section of the research methodology is the conclusion and recommendations. The developed data combining method will be summarized and the reliability of developed freight O-D matrices will be reported. The description of the process of ANFIS for freight distribution modeling will be summarized. Moreover, the performance of ANFIS compared with conventional gravity and regression-based models will be summarized, and recommendations will be made for future study.

Chapter V
Combining CFS and Roadside Survey Data

This chapter presents a comprehensive framework to combine commodity flow data and roadside survey data, and detailed methodologies. It ends with a discussion of the methodology.

5.1 Introduction

The objective of the Thailand CFS survey is to capture freight transportation data throughout the kingdom. Since it is a pioneering survey, it encountered several difficulties. One major problem was in respondent cooperation: some small establishments didn't keep records of their shipments, some larger businesses wished to keep shipping destinations confidential, or sometimes data was simply mislaid. The report of total volume shipped in Thailand astonished the well-informed. However, the scope and frame of CFS covered all shippers in Thailand, so the CFS may have captured all shipment transportation characteristics.

Against the roadside interview method, CFS uses the most efficiency survey method, household interview survey, in the context of passenger flow. CFS is able to capture movement from the place origin, making CFS more statistically reliable than the roadside interview but costlier. Moreover, CFS can capture all movements, which include short trips, moderate trips, and long trips, while short trips always vanish from roadside interview surveys when the survey stations located far from zone centers.

Although the RS1 collected data from only 10 target provinces, these were important provinces in the kingdom which generated a great deal of freight movement. These data represent approximately 90% of total freight movement in entire kingdom, making RS1 marginal total data preferable to CFS data and other data sources.

A method to combine these two data sources uses the strengths of each, the marginal total of RS1 and the distribution patterns of CFS, to produce an adjusted matrix. The marginal total of RS1 is present in its database while the CFS distribution pattern was determined by producing trip length distribution or calibrating friction impedances of the gravity model.

Two combining method are presented in the next section. The first method is Trip Length Distribution Adjustment method (TLDA) and the other is Gravity Model Approach method (GMA). Details of the two methods follow.

5.2 Trip length distribution method (TLDA)

Trip length distribution is calculated by accumulating the flow between each pair of zones according to the distance or travel impedance between zones. Trip length distribution reveals characteristics of freight distribution across distance or travel impedance. CFS freight distribution is acceptable for describing freight distribution in Thailand and RS1 marginal total is preferable than other available data. Trip length distribution method uses the strengths of these two data sources. CFS trip length distribution is used as a distribution pattern of adjusted data and is adjusted to meet RS1 marginal total. To verify this method, RS1 and RS2 data are used to produce adjusted data. Then, adjusted data from these two sources, RS1 and RS2, will be compared.

5.2.1 Assumptions

To create this combined performance method, we begin with the following assumptions.

1. The CFS trip length distribution is accepted as representative of freight distribution in Thailand.
2. The CFS shipment volume is under-reported by the same proportion at all distances.
3. RS1 and RS2 have shortages in reporting short trips.
4. There are no differences in long trips between CFS, RS1, and RS2.
5. Adjustments to CFS distribution data must be meet RS1 or RS2 long trips.

5.2.2 Methodology

Using the aforementioned assumptions, the methodology to combine the data to produce freight OD adjusted data is as shown in Fig 5.1.

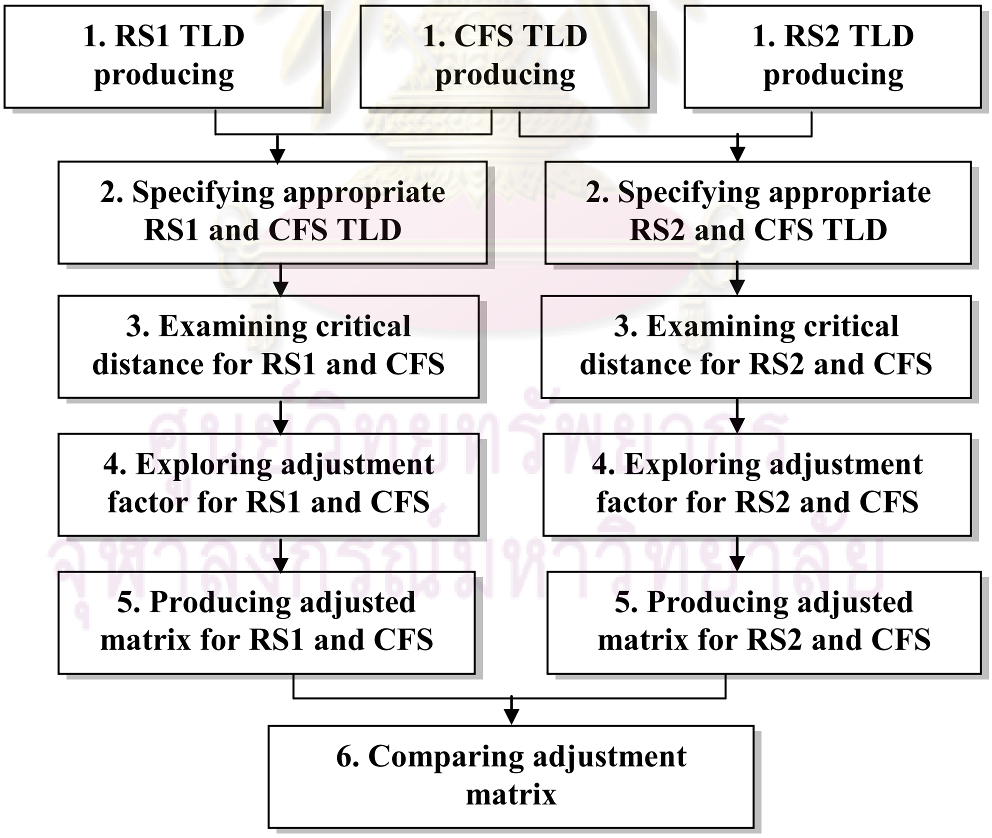

Fig 5.1 TLDA Combining method

5.2.3 Trip Length Distribution

RS1 data was extracted to produce Chonburi freight trip length distribution while RS2 raw data was used to produce trip length distribution. In order to produce trip length distribution, the shipping distance was divided into 25 kilometer intervals. However, statistics testing shows that there were no appropriate distributions for CFS due to data robustness. Thus, the distance was instead divided into 50 kilometer intervals to produce trip length distribution. Standard distributions were used to find the most appropriate distribution model for the data. Gunyoung (2003) employed four standard distributions (Gamma, Lognormal, Weibull, and Log logistics) to analyze trip length distribution between the commodity-based model and the truck trip based model in the Seoul metropolitan area, and used the K-S test value to verify the most appropriate distribution. Moreover, this research employs six standard distributions to verify appropriate distribution for the data:

- Exponential distribution
- Power distribution
- Gamma distribution
- Lognormal distribution
- Weibull distribution
- Log logistic distribution

The results of trip length distribution are shown in the appendix.

5.2.4 Specifying appropriate trip length distribution

For representing trip length distribution of the data, the most appropriate distribution was selected. Elect distribution must accord with these criteria.

1. CFS TLD and RS1 TLD or RS2 TLD must be super-imposed in long distance.
2. CFS TLD must represent short trips more than RS1 TLD and RS2 TLD.
3. After adjusting process, CFS TLD must equal or exceed RS1 TLD or RS2 TLD for all distance intervals.

5.2.4.1 CFS trip length distribution

Among six distributions, Gamma distribution is the most appropriate for the CFS trip length distribution with K-S statistic of 0.34415 while the critical value at $\alpha = 0.01$ is 0.34427. The results of K-S show that it shall not be able to reject the null hypothesis, thus there is no significant difference between the observed frequency and expected frequency at $\alpha = 0.01$. The shape of the CFS consumer goods trip length distribution is depicted below. The Gamma distribution of the data is shown below.

$$f(x) = \frac{(x)^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)} e^{\left(\frac{-x}{\beta}\right)}$$

$$f(x) = \frac{(x)^{-0.091}}{109.5^{0.909}\Gamma(0.909)} e^{\left(\frac{-x}{109.5}\right)}$$

Where x means the trip length in kilometers.

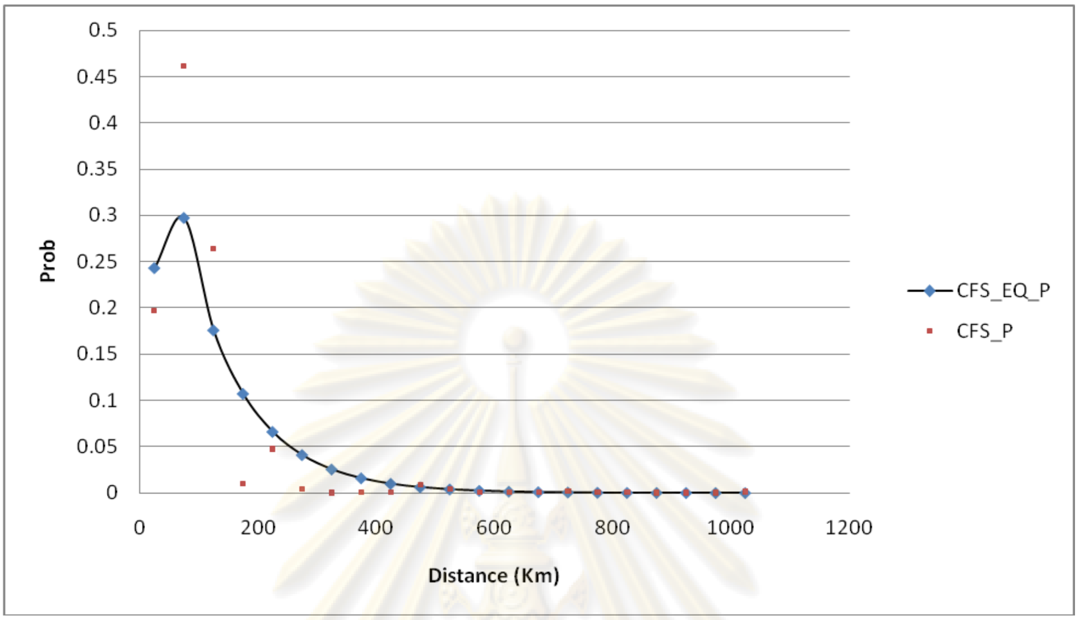

Fig 5.2 Trip length distribution of CFS combined food and consumer goods data

5.2.4.2 RS1 trip length distribution

The shape of the RS1 consumer goods trip length distribution is depicted below. The shape of distribution is according to the distribution of CFS consumer goods trip length distribution. Although there are few long trips, a proportion of short trips is slightly lower than in the CFS data. Moreover, RS1 shows a dramatic decrease of trips at 500 kilometers while the CFS shows this decrease at 300 kilometers. This difference shows that the RS1 data has more long distance trips due to the location of the survey stations. In other words, there is a shortage of short trips in the RS1 data.

Among six distributions, the results of K-S show that Weibull distribution is most appropriate for trip length distribution of the data with K-S statistic of 0.20349 while the critical value at $\alpha = 0.01$ is 0.34427. The results of K-S show that it shall not be able to reject the null hypothesis, thus there is no significant difference between the observed frequency and expected frequency at $\alpha = 0.01$. The Weibull distribution of the data is shown below.

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left(-\left(\frac{x}{\beta}\right)^{\alpha}\right)$$

$$f(x) = \frac{1.260}{139.480}\left(\frac{x}{139.480}\right)^{(1.260)-1} \exp\left(-\left(\frac{x}{139.480}\right)^{1.260}\right)$$

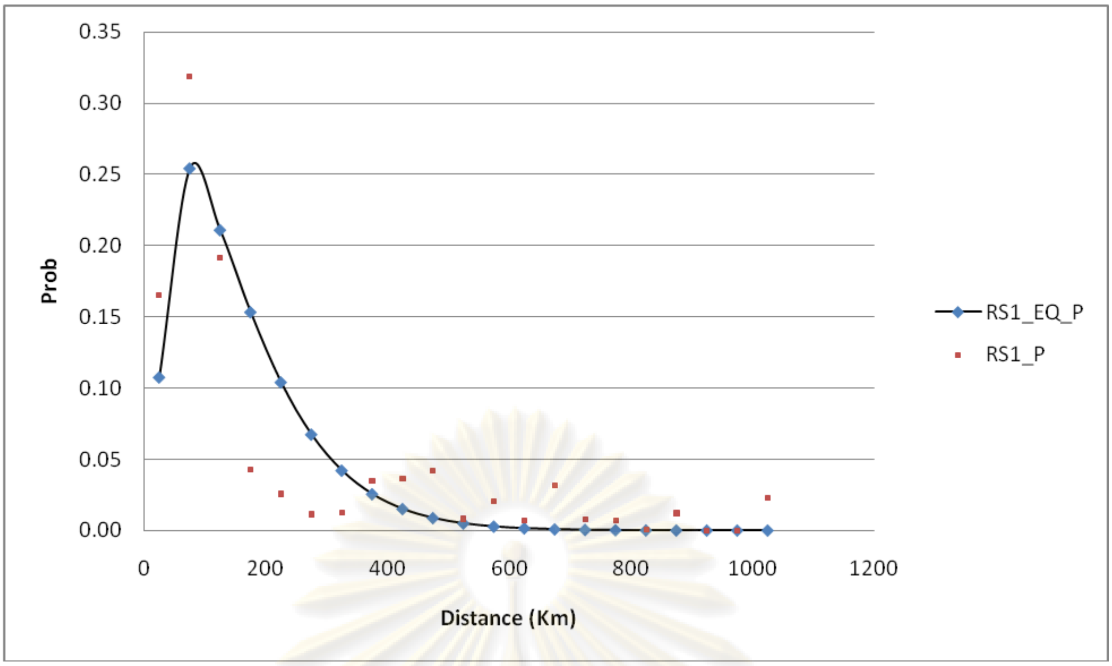

Fig 5.3 Trip length distribution of RS1 combined food and consumer goods data

5.2.4.3 RS2 trip length distribution

The shape of the RS2 consumer goods trip length distribution is depicted below. The shape of distribution is according to the distribution of CFS consumer goods trip length distribution. Although there are few long trips, a proportion of short trips is slightly lower than CFS data. Moreover, there is a dramatic decrease of trips at 500 kilometers while this decrease begins at 300 kilometers in the CFS data. The difference shows that the RS2 data has more long distance trips due to the locations of its survey stations. In other words, there is a shortage of short trips in the RS2 data.

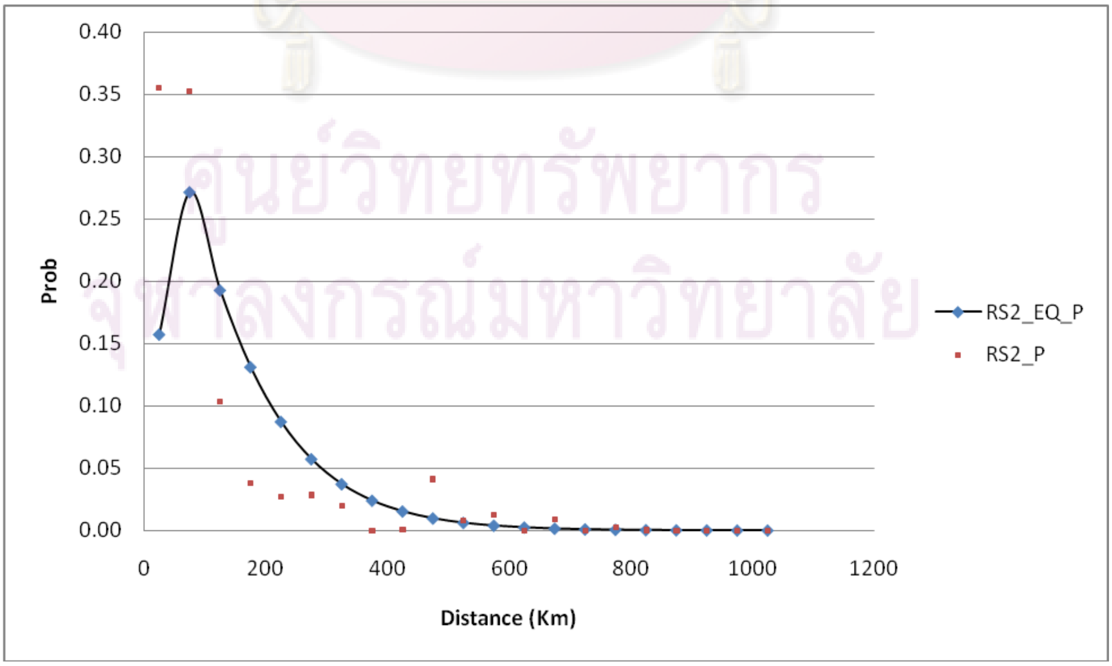

Fig 5.4 Trip length distribution of RS2 combined food and consumer goods data

Among six distributions, the result of K-S show that Weibull distribution is most appropriate for trip length distribution of the data with K-S statistic of 0.27927 while the critical value at α = 0.01 is 0.34427. The results of K-S show that it shall not be able to reject the null hypothesis, thus there is no significant difference between the observed frequency and expected frequency at α = 0.01. The Weibull distribution of the data is shown below.

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}\exp\left(-\left(\frac{x}{\beta}\right)^{\alpha}\right)$$

$$f(x) = \frac{1.079}{128.350}\left(\frac{x}{128.350}\right)^{(1.079)-1}\exp\left(-\left(\frac{x}{128.350}\right)^{1.079}\right)$$

5.2.5 Examining critical distance

As pictured below, CFS data acquired more short trips than RS1 and RS2, and RS2 captured more short trips than RS1. The results of testing of independence show that CFS and RS1 are different at 0-325 kilometers (α = 0.05) while CFS and RS2 are different at 0-275 kilometers (α = 0.05). The results of the test reveal that the CFS and RS2 long trips are not different at distances more than 275 kilometers, and CFS and RS1 long trips are not different at distances more than 325 kilometers.

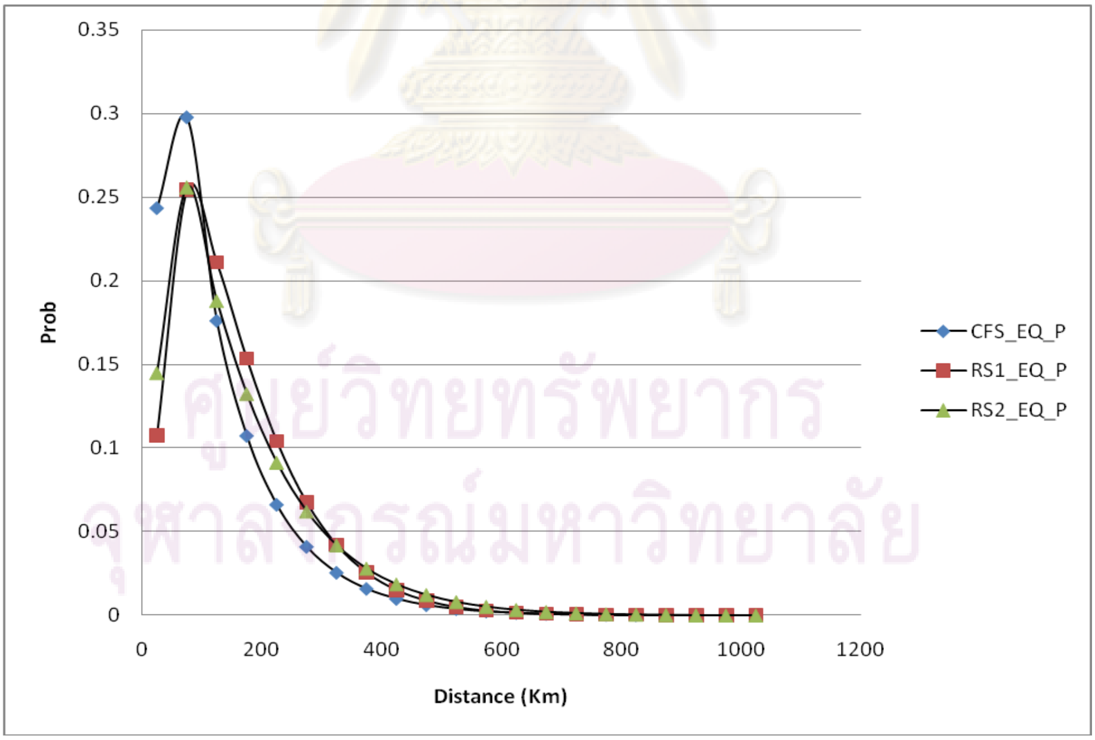

Fig 5.5 Comparing CFS RS1 and RS2 trip length distribution of combined food and consumer goods data

5.2.6 Exploring adjustment factor

The calculation of the adjusting factor to reconcile CFS distribution with RS1 or RS2 data is shown below.

$$Min \sum_{i=1}^{n} (P_{CFS}\,(\mathrm{x_i} > \mathrm{dist}) \times CFSVolume \times ADJ - Ton_{RSx}(x_i > dist))^2$$

Subject to

$$P_{CFS}\,(\mathrm{x_i} > \mathrm{dist}) \times CFSVolume \;\times ADJ > TON_{RSx}(x_i > dist) \qquad \forall x_i$$

Where ADJ is adjustment factor
$dist$ is critical distance
P_{CFS} ($\mathrm{x_i}$ > dist) is probability of CFS in distance interval X_i
CFSvolume is marginal volume of original CFS
Ton_{RSx}($\mathrm{x_i}$ > dist) is shipment volume of $\mathrm{RS_x}$ data in distance interval x_i

Combining CFS and RS1 indicates that short trips of RS1 are adjusted by 61%. The largest adjustment is at 0-50 kilometer interval by 181% which accords with the assumption that RS1 under-reports short trips. Moreover, the adjustment magnitude decreases when distance increases, which accords with the assumption that the adjusted data and RS1 will be super-imposed in long distance. The adjusted data reveal significant evidence that although roadside interview data is able to capture freight movement data, short trips may be under-reported in cases where the survey stations are inappropriately located for collecting short trip data.

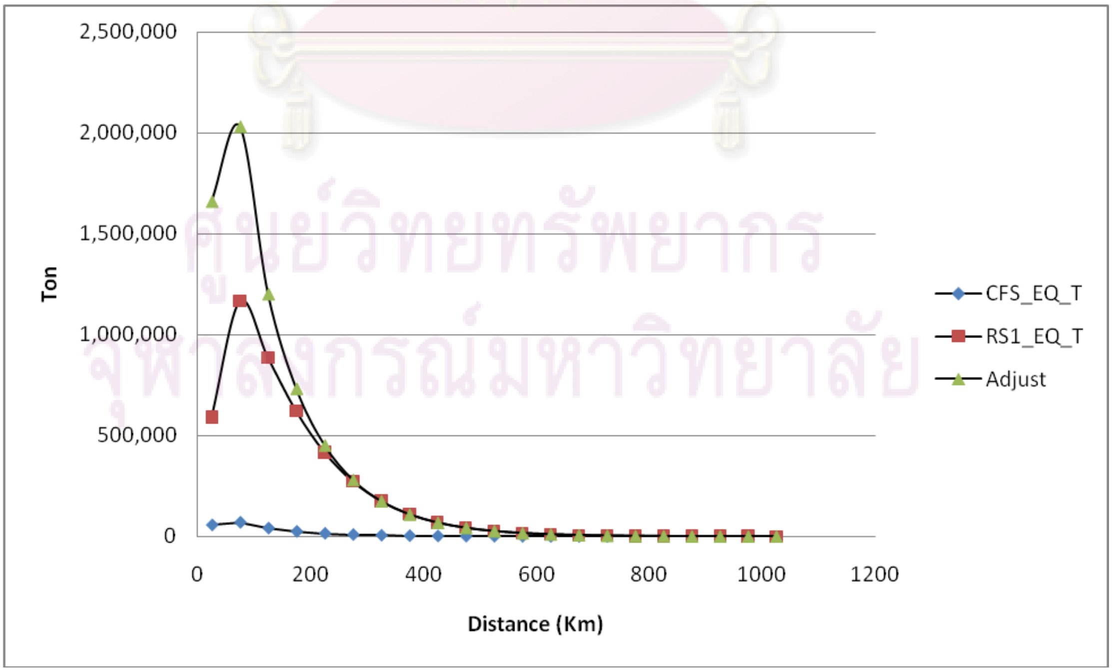

Fig 5.6 Adjusted RS1 combined food and consumer goods data

Table 5.1 Adjusted RS1 consumer goods data

Distance	RS1	Adjusted Data	Percent
0-50	590,570	1,662,329	181
50-100	1,167,804	2,033,195	74
100-150	886,490	1,202,650	36
150-200	620,610	733,318	18
200-250	417,089	452,271	8
250-300	272,781	280,644	3
Total	3,955,344	6,364,407	61

According to the combination of CFS and RS1, combining CFS and RS2 indicates that short trips of RS2 are adjusted by 48%. The largest adjustment is at 0-50 kilometer interval by 121% which is lower than the case of RS1. However, the result still supports the assumption that RS2 under-reports short trips. Moreover, the adjustment magnitude decreases when distance increases, according with the assumption that the adjusted data and RS2 will be super-imposed in long distance.

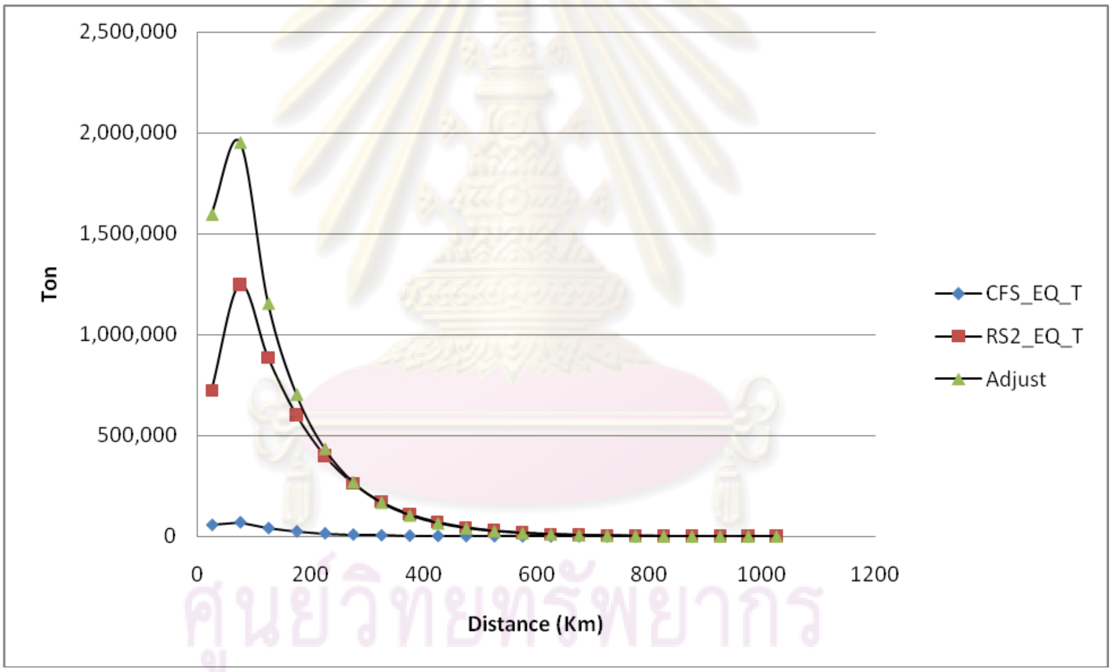

Fig 5.7 Adjusted RS2 combined food and consumer goods data

Table 5.2 Adjusted RS2 consumer goods data

Distance	RS2	Adjusted data	Percent
0-50	723,458	1,596,646	121
50-100	1,248,564	1,952,858	56
100-150	886,629	1,155,131	30
150-200	602,905	704,342	17
200-250	401,283	434,401	8
Total	4,126,198	6,112,933	48

Comparing adjusted RS1 and RS2 indicates that RS1 has more adjustment than RS2. This could be due to the number and location of the survey stations. RS1 located three survey stations on major highways into and out of Chonburi province while RS2 used 13 survey stations. Moreover, a number of survey stations of RS2 were located in urbanize areas. For this reason, it is not surprising that RS2 captured more short trips than RS1, although not as well as CFS. The adjustment factors of RS1 and RS2 are 29.44 and 28.28 respectively. The higher adjustment factor of RS1 reveals that RS1 captured fewer short trips than RS2, which is to be expected given is lower number of survey stations.

5.2.7 Producing adjusted matrix

The adjustment factor is used to produce the adjusted matrix by multiplying each cell in the CFS matrix by that adjustment factor, which will not disturb trip length distribution. However, the CFS matrix has a number of empty cells which should instead contain a positive flow. Thus the adjusted matrix still does not correct the empty cell problem. For this reason, adjusted total volume will be presented here without empty cell recovery. The total volumes of combined food and consumer goods adjusted matrix of RS1 and RS2 are 6,831,812.84 tons and 6,561,870.33 tons respectively while the original volumes of RS1 and RS2 are 4,415,747.60 and 4,598,681.78 respectively.

5.2.8 Comparing adjusted matrix

Although RS1 and RS2 are different data sources which collect data from different times and different authors, the difference of total volume is 269,942.51 tons or 3.95%, which is rather close together. This result strongly advocates the preferability of the CFS distribution pattern and the adjusting method's performance.

5.2.9 Summary and Discussion

Trip length distribution method uses the strengths of its two data collection methods, CFS trip length distribution and roadside interview marginal total. CFS trip length distribution is used as a distribution pattern for the adjusted data and will be adjusted to meet RS1 marginal total. The method was developed using five assumptions involving trip length distribution and characteristics of these data. To verify the method, RS1 and RS2 data are used to produce adjusted data. Then, adjusted data from these two sources, RS1 and RS2, were compared.

Six standard distributions were employed to produce trip length distribution for CFS, RS1, and RS2 including Exponential, Power, Gamma, Lognormal, Weibull, and Log logistic distribution. The most appropriate distribution was selected statistically using three specific criteria to ensure that these trip length distributions agree with developed methodology.

Gamma distribution is the most appropriate distribution for CFS while Weibull distribution is the most appropriate distribution for RS1 and RS2. Although RS1 and RS2 have the same appropriate distribution, their distribution parameters are different. Then, critical distance is examined to discover where CFS trip length distribution and roadside trip length distribution are not super-imposed. This examination revealed that CFS and RS2 long trips are not different at a distance of more than 275 kilometers and that CFS and RS1 long trips are not different at a distance of more than 325 kilometers.

With trip length distributions and critical distances, adjustment factors were explored. Adjustment factors of RS1 and RS2 were 29.44 and 28.28 respectively. The short trips of RS1 were adjusted by 61% with the largest adjustment at 0-50 kilometer interval while short trips of RS2 were adjusted by 48% with the largest adjustment at 0-50 kilometer interval by 121% which is lower than the case of RS1. These differences in adjustments may be due to the numbers and locations of survey stations. RS1 used three survey stations on major highways into and out of Chonburi province, while RS2 used 13 survey stations, some of which were located in urbanized areas of the province.

The adjusted matrix shows total volumes of combined food and consumer goods of 6,831,812.84 tons for RS1 and 6,561,870.33 tons for RS2. The difference in these figures is only 269,942.51 tons or 3.95%, which is quite close, despite RS1 and RS2 coming from different data sources collected at different times by different authors. This result strongly supports the use of the CFS distribution pattern and the method of adjustment.

The developed method uses the strengths of CFS, RS1, and RS2. It is tested by comparing the results of applying it to RS1 and RS2. The results indicate agreement of adjusted volumes of RS1 and RS2 which despite different collection times and authors. For this reason, the developed combining method is a satisfactory adjustment tool for this purpose under the constraint of available data limitation.

5.3 Gravity Model Approach method (GMA)

The gravity model is well-known as a trip distribution tool and is widely used in passenger and freight transportation. The gravity model allocates trip from origin to destination zones in proportion to the total number of trips produced at each origin zone and which are attracted to each destination zone, and in inverse proportion to a measure of the separation, the so-called friction, of origin and destination. The general form of the gravity model is shown below.

$$T_{ij} = k \frac{P_i A_j f(t_{ij})}{\sum_j A_j f(t_{ij})}$$

where T_{ij} is the freight flow between supply node *i* and demand node *j*
Pi is the total freight volume at supply node *i*
A_j is the total consumption volume at demand node *j*
$f(t_{ij})$ is the impedance function for freight flow between supply node *i* and demand node *j*
k is a proportionality constant

Ogden (1978) argued that it is not appropriate to distribute trip generated from origin to destination as passenger flow because freight movement is governed by demand. Therefore, produced freight is attracted to a destination from a range of origins. Ogden (1978) suggests general form of the gravity model for freight distribution as shown below.

$$T_{ij} = k \frac{A_j P_i f(t_{ij})}{\sum_i P_i f(t_{ij})}$$

where T_{ij} is the freight flow between supply node i and demand node j
Pi is the total freight volume at supply node i
A_j is the total consumption volume at demand node j
$f(t_{ij})$ is the impedance function for freight flow between supply node i and demand node j
k is a proportionality constant

5.3.1 Assumptions

To create this combined performance method, we begin with the following assumptions.

1. The CFS distribution pattern is accepted as representative of freight distribution in Thailand.
2. The CFS shipment volume is under-reported by the same proportion at all distances.
3. RS1 under-reports short trips.
4. There are no differences in long trips between CFS and RS1.
5. Total production and total attraction of RS1 are underreported.

5.3.2 Methodology

Based on the aforementioned assumptions, the GMA methodology to combine the data to produce freight OD adjusted data is as shown in Fig 5.8.

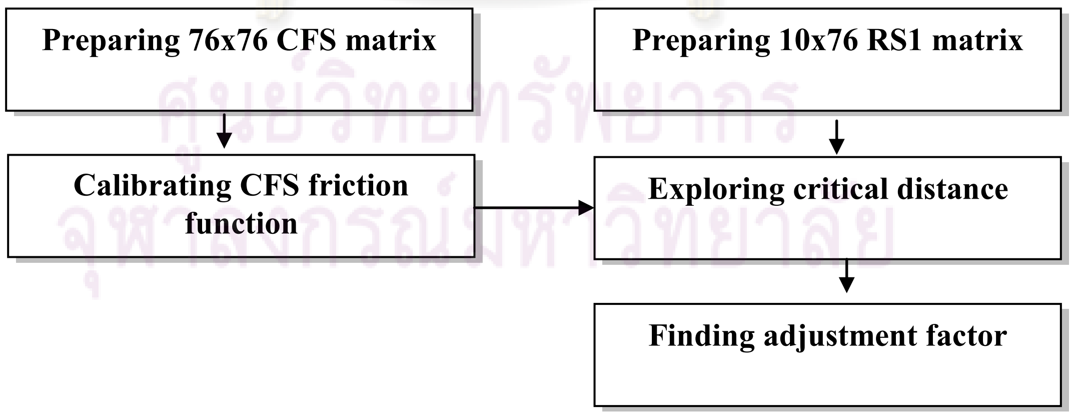

Fig 5.8 GMA combining method

5.3.3 Calibrating impedance function

Obtaining the impedance factor distribution is the main part of the gravity model calibration process. Popular fiction functions were employed for calibrating the gravity model including

- exponential function $f(t_{ij}) = \exp(-\beta t_{ij})$
- power function $f(t_{ij}) = t_{ij}^{-n}$
- gamma function $f(t_{ij}) = t_{ij}^{-n} \exp(-\beta t_{ij})$

Fortunately, the O-D matrices can be derived in a spreadsheet database. This means that the calibration can be performed by spreadsheet software. Microsoft Excel Solver analysis tool is employed to calibrate gravity model. The convergence criterion is using The Root Mean Square Error (RMSE) of predicted trip and observed trip. Since there is no recommended value for freight flow distribution available, the initial values of friction factors are assumed to be equal to 1.

The results show that power function is outstanding against the other functions. The calibrated impedance functions of the CFS data are as follows.

$$f(t_{ij}) = t_{ij}^{-1.578}$$

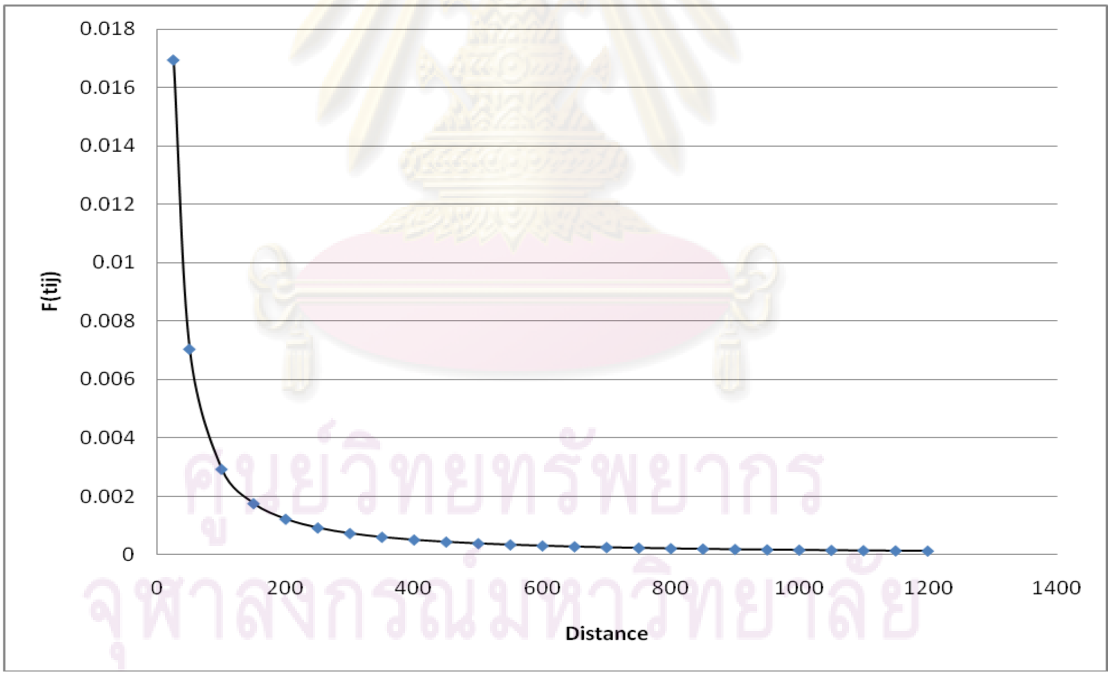

Fig 5.9 CFS impedance function

5.3.4 Exploring critical distance

Unfortunately, there is no appropriate friction function for RS1, so trip length distribution was employed to explore critical distance. Appropriate distribution of CFS is Gamma distribution (α = 0.757 β = 228.58) with K-S statistic of 0.22994 while the critical value at α = 0.01 is 0.2618. For RS1, the most appropriate distribution is Weibull (α = 1.167 β = 347.16) with K-S statistic of 0.11025 while the critical value at $\alpha = 0.01$ is 0.2618. The results of testing of independence show that CFS and RS1 are different at 0-350 kilometers ($\alpha = 0.05$), thus the CFS and RS1 long trip are not different at distances above 350 kilometers.

5.3.5 Finding adjustment factor

The CFS Impedance function was applied to a 10x76 RS1 matrix to find the adjustment factor. The critical distance was used to verify that original and adjusted RS1 cell values were consistent. The convergence criterion used was the Root Mean Square Error (RMSE) of original and adjusted trips, particularly in cells where the distance between the origin and destination exceeded critical distance. Microsoft Excel Solver analysis tool was employed to calibrate the adjustment factor.

As mentioned earlier, the gravity model for passenger and freight are different. While the passenger model focuses on the distribution of production trips from origins to competing destinations using the destinations' attractive potential and origin-destination impedance, freight flow focuses on destinations that are attractive from a range of origins. For this reason, the adjustment factor was calibrated for two cases of the gravity model, passenger concept (PCGM) and freight concept (FCGM), to compare the difference.

The results indicate that PCGM has an adjustment factor equal to 1.698 and adjusts 10x76 freight transport volume from RS1 reported 24,178,825.14 tons to 37,873,530.90 or 69.8%, while FCGM has 2.129 for adjustment and its adjusted volume is 51,469,072.57 tons or 112.9%. FCGM shows a higher adjustment factor than PCGM. Focusing on Chonburi data, the adjusted Chonburi freight transport volume from RS1 using PCGM reported 4,415,747.60 tons to 7,106,023.63 or 60.9%, while the adjusted volume of FCGM is 8,838,150.48 tons or 100.2%. It should be noted that the adjusted volume as presented includes empty cell filling while the excluded empty cell filling adjusted volumes for Chonburi of PCGM and FCGM are 6,888,709.67 and 8,571,624.32 respectively.

5.3.6 Discussion of the Results

Two methods for adjusting data were developed, the Trip Length Distribution Approach (TLDA) and the Gravity Model Approach (GMA). As mentioned earlier, only Chonburi province had data collected from all three sources (CFS, RS1, and RS2), making it the most appropriate for data comparison. The adjusted Chonburi freight transportation data is shown below. The total Chonburi freight volume reported by RS1 is 4,415,747.60 tons while the adjusted volumes reported by TLDA-RS1, TLDA-RS2, GMA-PCGM, and GMA-FCGM are 6,831,812.84, 6,561,870.33 6,888,709.67, and 8,571,624.32 respectively. These results reveal that the adjusted volume of TLDA is close to both GMA-PCGM and GMA-FCGM. However GMA-FCGM is different from the others.

Comparing the results in percentages reveals that GMA-FCGM has the highest adjusted percentage, near to 100%, which is far from the others. However, the range of adjustment of other three adjusting methods, TLDA-RS1, TLDA-RS2, and GMA-PCGM, is between 50-60%. These three methods have similar adjustment percentages.

Although the concept of GMA-FCGM is appropriate for freight distribution analysis as suggested by Ogden (1978), this research focuses on how to adjust available data, CFS and roadside interview data, not studying the behavior of distribution patterns. Moreover, the data collection methods of CFS focus mainly on the origin of freight since CFS collects data at establishments. For this reason, GMA-PCGM is more appropriate than GMA-FCGM.

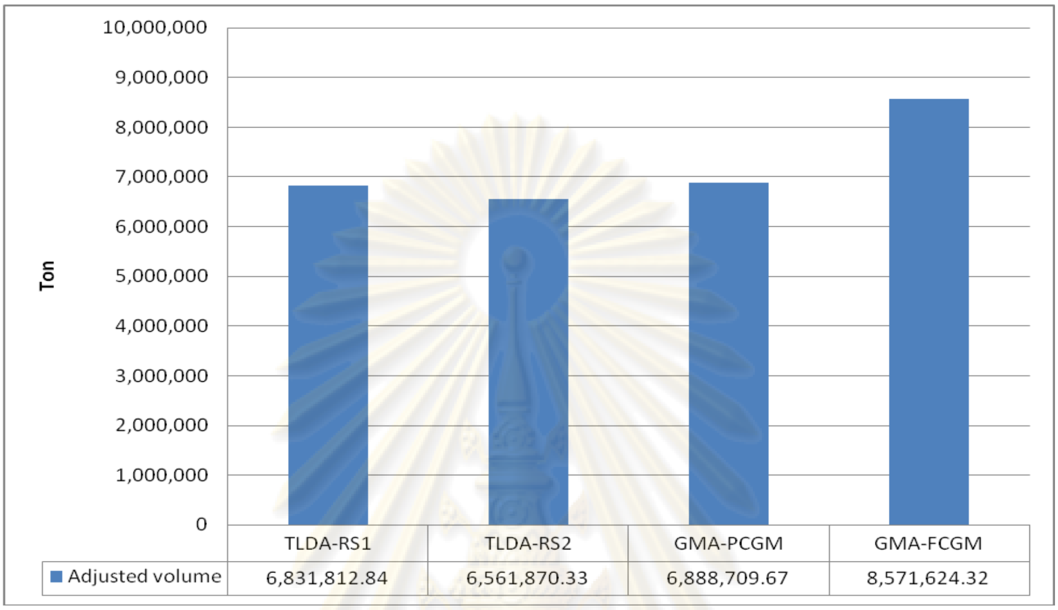

Fig 5.10 Comparing adjusted volume

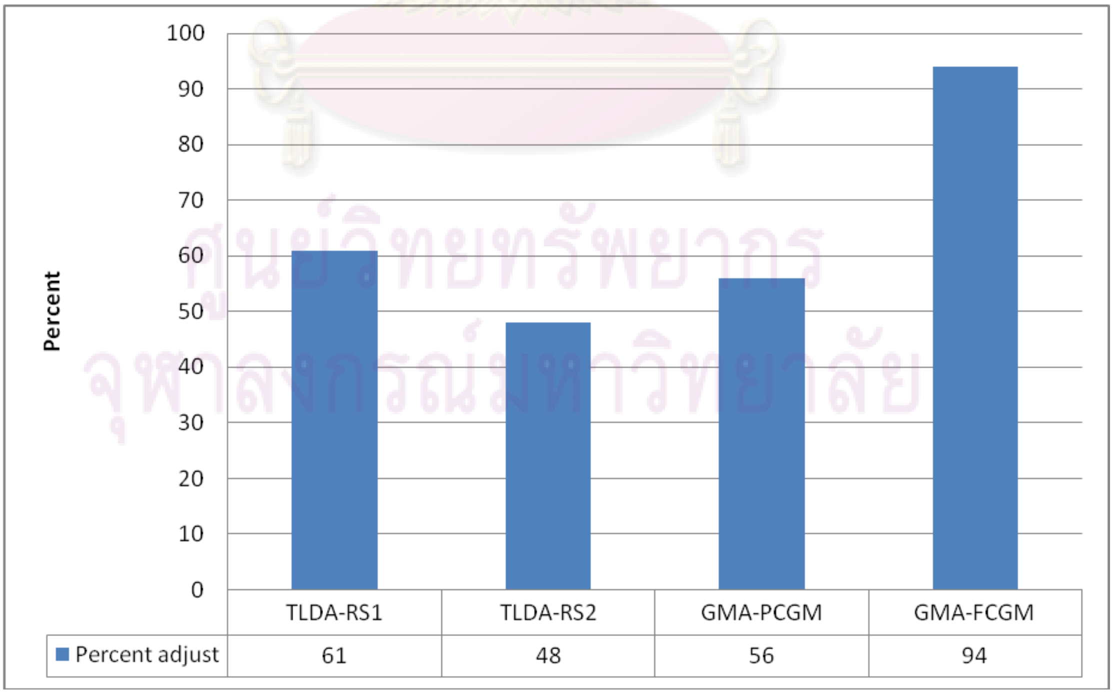

Fig 5.11 Comparing percentage adjustment

Since the three developed method have closely adjusted volumes and since GMA-PCGM agrees with the CFS collection data method, it can be concluded that the developed method can be used to adjust the data.

5.4 Summary and concluding remarks

Unfortunately, this pioneering compressive freight flow data in Thailand has a number of weak points. However, the available data does have strengths which are well suited for the method of adjustment, and the strengths of CFS and roadside data complement each other. Using the strengths of the two types of data, two methods were developed to adjust the data.

The first method is Trip Length Distribution Adjustment (TLDA), which aims to adjust bare CFS original destination matrix, which under-reported marginal total, to meet the more reliable marginal total of roadside interview data. Two data sets, RS1 and RS2, were used to calibrate the method and test its performance for adjusting data from difference sources. Using the trip length distribution of CFS and roadside interview data, the method was developed. Standard distribution was employed to depict trip length distribution, consisting of Exponential, Power, Gamma, Lognormal, Weibull, and Log logistic distribution. Gamma distribution was performed for CFS while RS1 and RS2 accepted Weibull distribution. Using the appropriate distribution, the adjustment factor was explored. The results indicate that adjusted volumes of RS1 and RS2, which collected data at different times and from different authors, agree. The difference of total adjusted volumes between RS1 and RS2 is only 269,942.51 tons or 3.95%, which is quite close.

The second method is Gravity Model Approach (GMA), used to adjust a 10x76 matrix of RS1 using the CFS distribution pattern because the CFS distribution pattern was preferable to available data. Since the main assumption of this method is that RS1 under-reports short trips, this method tries to adjust the volume of short trips using the CFS distribution pattern while keeping long trip volume constant. The 76x76 matrix of CFS is used to calibrate impedance function. The power function is outstanding among conventional Exponential and Gamma functions. Then the calibrated impedance function is applied to a 10x76 matrix of RS1. Microsoft Excel Solver analysis tool is employed to calibrate adjustment factor and Root Mean Square Error (RMSE) of RS1 trips and adjusted trips, particularly in cells where the distance between origin and destination exceeded critical distance (convergence criterion). Since Ogden (1978) suggests a modified form of the gravity model which emphasizes the destination’s attractiveness from a range of origins, this research develops the adjusting method using conventional and modified forms of the gravity model. The method developed employs PCGM using the conventional form and FCGM using a modified form of the gravity model. The results indicate that PCGM adjusts Chonburi freight transport volume from RS1 reported 4,415,747.60 tons to 7,106,023.63 or 60.9%, while adjusted volume of FCGM is 8,838,150.48 tons or 100.2%. It should be noted that the adjusted volume as presented includes empty cell filling while the excluded empty cell filling adjusted volumes for Chonburi of PCGM and FCGM are 6,888,709.67 and 8,571,624.32 respectively.

Comparing TDLA and GMA method indicates that adjusted Chonburi freight volumes of TLDA-RS1, TLDA-RS2, GMA-PCGM, and GMA-FCGM are

6,831,812.84 (61%) 6,561,870.33 (48%) 6,888,709.67 (56%), and 8,571,624.32 (94%) respectively. These results reveal that the adjusted volume of TLDA is close to both TLDA-RS2 and GMA-PCGM. However GMA-FCGM is different from the others.

Although the adjusted volume for Chonburi freight transportation reported by GMA-FCGM is different from the others, reported volumes from the other three methods are quite close. Moreover, the data collection methods of CFS mainly focus on the origin of freight since CFS collected data at establishments. For this reason, GMA-PCGM is more appropriate than GMA-FCGM.

From the discussion above, it can be concluded that the developed method can be used to adjust the data. Next, the two matrices will be developed. The first is CFS based using TLDA adjustment factor to produce a 76x76 combined food and consumer goods matrix. The other is an RS1 based matrix using GMA to produce a 10x76 matrix which will be used as an auxiliary matrix for empty cell verification and augmentation.

Chapter VI
Origin destination matrix gap-filling

This chapter presents a comprehensive framework to develop a gap-filling method. This chapter starts with describe of data preparation. Then comes verifying "true zeros" in the prepared matrix. Next are modeling using gravity model Box-Cox transformation and Adaptive Neuro Fuzzy Inference System (ANFIS). Then comes evaluation of the developed model. The last section is dedicated to concluding remarks.

6.1 Data preparation

In order to combine methodology as mentioned earlier, CFS combined food and consumer goods were placed within a 76x76 dimension matrix for modeling. The adjustment factor of 10 target provinces is show in Table 6.1. It is applied to adjust shipment volume of the 10 target provinces. For the other provinces, an average value is used as an adjustment factor.

Table 6.1 Adjustment factor

Province	Adjustment Factor
Bangkok	12.6416
Chonburi	29.4449
Pracinburi	22.3106
Nakhon Ratchasima	14.8890
Khonkean	15.8003
Chiang Mai	10.7357
Nakhon Sawan	6.5984
Rajchaburi	27.2403
Chumpon	2.0832
Songkla	15.3841
Average	15.7128

By applying the adjustment factor into the original CFS matrix, a 76x76 Adjusted CFS matrix was produced. The matrix had a total of 5,776 cells in the matrix, consisting of 4,666 zero cells and 1,170 non-empty cells. A total of 91,912,662 tons of shipments were available in the matrix. Moreover, adjusted RS1 combined food and consumer goods in a 10x76 matrix were prepared to verify zero cells. Adjusted RS1 had 37,873,531 tons of shipments available in the matrix while the 10x76 matrix subset of the 76x76 CFS adjusted matrix had 41,025,249 tons of shipments, which is close to the value in the adjusted RS1 matrix.

6.2 Verify and augment "Zero cells" in the prepared matrix

The prepared matrix was compared to verify agreement between prepared data and adjusted RS1 data. These two datasets were compared for those particular cells to verify that neither of them contradicted the true zero. In cases of contradiction, i.e.

where roadside observations were found for cells previously marked as "true zero," the restriction on that cell was lifted.

According to this combining methodology, trip length distribution of the CFS was acceptable. Thus, contradictions between these two data sources leads to a problem, namely that using data from non-empty roadside surveys into the prepared matrix will destroy the prepared matrix's trip length distribution. However, non-empty cells must be filled with roadside survey data while keeping TLD. Thus, roadside data was used to aggregate non-empty cells into the matrix while the margin of each distance interval was held constant. A total of 868 empty cells were thus augmented. A total of 5,828,682.62 tons of freight shipment augmented the 868 empty cells while aggregation within trip length intervals were used to keep marginal totals and thus maintain trip length distribution Then, the non-empty cells were used to develop the model. Since the purpose was to fill gaps which occurred in inter-province trips, intra-province cells were excluded.

6.3 Gravity Model

Since the gravity model is well-known and has been used in trip distribution including freight transportation, it was used as a benchmark to evaluate the performances of the other models. This research uses the single constraint conventional gravity model (CGM) using zone total and distance as input variables. The general formulation of the gravity model is as follows.

$$T_{ij} = k \frac{P_i A_j}{f(t_{ij})} \tag{6.1}$$

where T_{ij} is freight flow between supply node i and demand node j
Pi is total freight volume at supply node i
A_j is total consumption volume at demand node j
$f(t_{ij})$ is impedance function for freight flow between supply node i and demand node j
k is a proportionality constant

Obtaining the friction factor distribution is the main part of the gravity model calibration process. Three popular fiction functions were employed for calibrating gravity model:

- exponential function $f(t_{ij}) = \exp(-\beta t_{ij})$
- power function $f(t_{ij}) = t_{ij}^{-n}$
- gamma function $f(t_{ij}) = t_{ij}^{-n} \exp(-\beta t_{ij})$

Microsoft Excel Solver analysis tool was employed to calibrate the gravity model. The convergence criteria used the Root Mean Square Error (RMSE) of predicted trip and observed trip. The results show that the power function is outstanding against other functions. The calibrated friction functions of the CFS data are as follows:

$$f(t_{ij}) = t_{ij}^{-0.657}$$

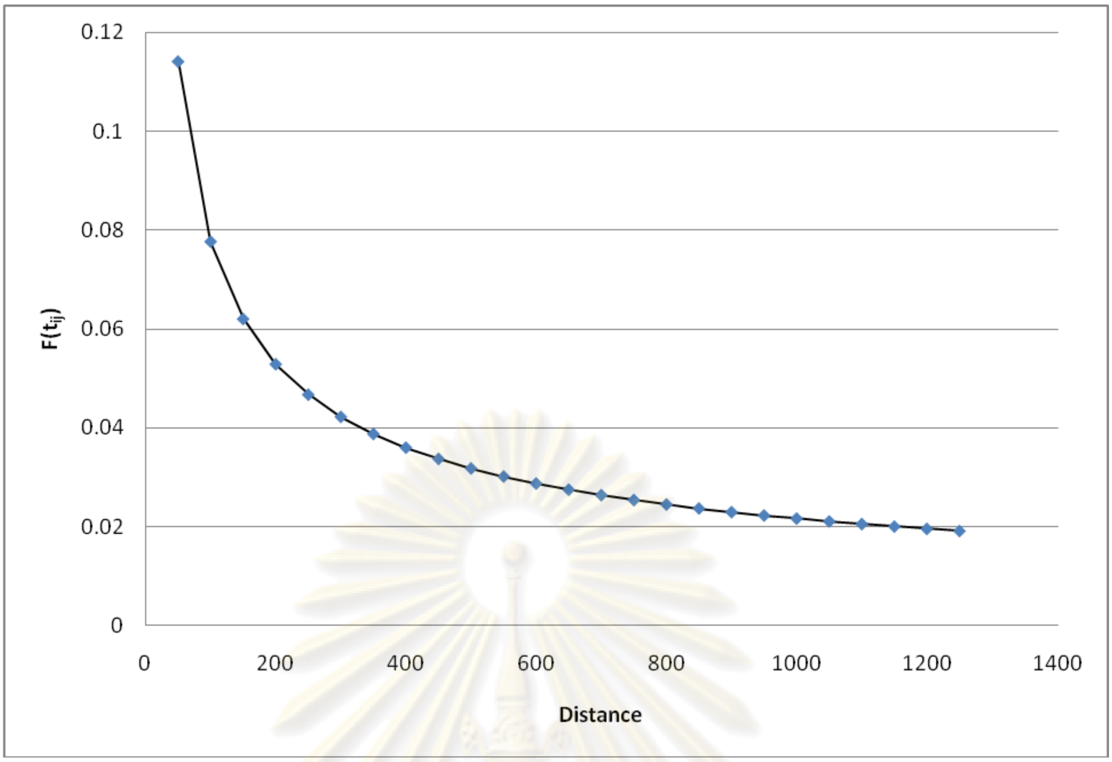

Fig 6.1 Friction function of calibrated gravity model

The calibration for the gravity model shows that CGM has an RMSE value of 101,632.03, MRE is 215.61, and R^2 is 0.589. Predicted and observed data are depicted in the picture below.

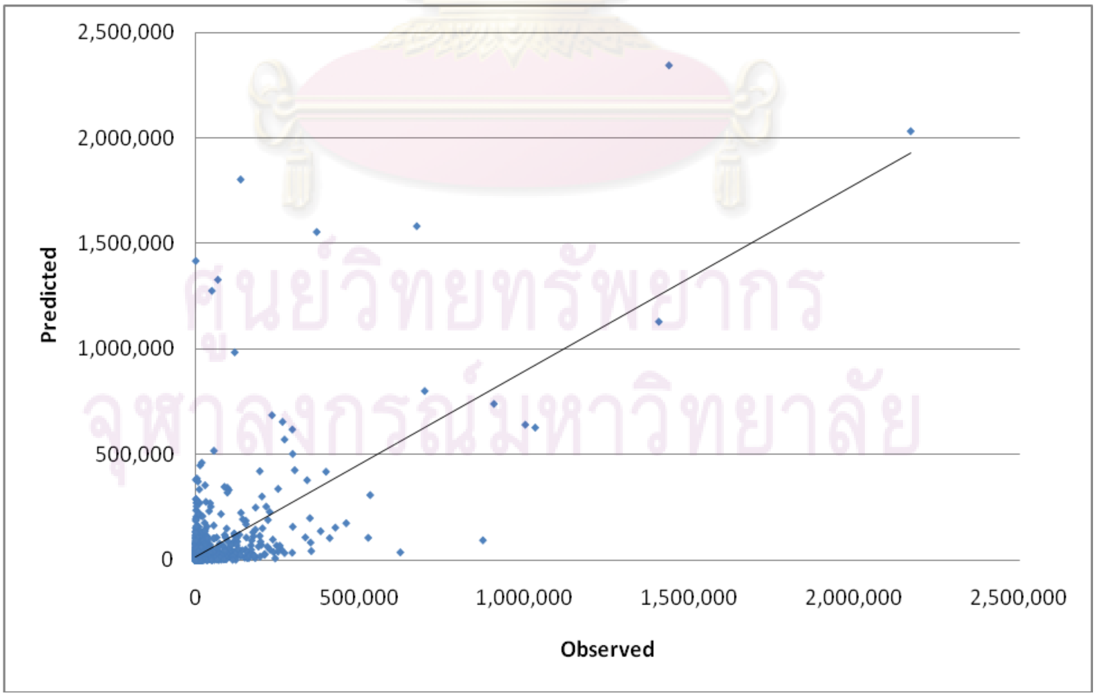

Fig 6.2 Comparing predicted and observed data of CGM

6.4 Box-Cox Transformation Model

The gravity model uses zonal total and distance as proxy variables to explain the system. These variables are not able to link the system to socio economic data of the origin and destination areas where the trip is generated from or attracted to. In contrast, the regression based model attempts to use area socio-economic data as input variables to link the model to characteristics of the origins and destinations.

Celik and Guldmann (2002) used a regression based model to determine freight distribution in US by using a Box-Cox functional form as a transformation function. The flow of 16 commodity groups for 48 continental states of the US using the 1993 Commodity Flow Survey was explored. A set of explanatory variables that characterized the economic structure of the origins and destinations was used as input variables to the model. Celik and Guldmann (2002) used the model to evaluate the performance of a developed artificial neural network model. However, they did not develop a regression model using zonal total and distance as input variables for verifying the performance of the regression base model with the same input variables against the gravity model.

For this reason, this research develops two regression based model. The first model (CBCM) uses three conventional gravity model variables: origin production, destination attraction, and distance. The second model (SBCM) attempts to include socioeconomic variables into the model to mimic constituting parts of the trip that generate and are attracted to the zone.

6.4.1 Variables

The (CBCM) model uses a conventional gravity model with three input variables:

- Zone production, which expects the impact to be positive
- Zone attraction, which expects the impact to be positive
- Distance between zones, which expects the impact to be negative

For the second model (SBCM), Celik (2005) suggests input variables to the model, which can be divided into three groups: origin variables, destination variables, and geographic variables.

Origin variables explain the behavior of the trip as generated at the place of origin. The origin acts as the supply point where freight is manufactured for transport to customers in either the zone of origin or in other zones. Therefore activity at the origin consists of both supply and demand, thus the variables should be proxies for both. Origin variables consist of

- Employment data presents potential productivity. Since labor is an input for production, origins with greater employment should generate more freight. For this reason, the impacts are expected to be positive. This research uses wholesale & retail sartorial employment datasets as input variable.

- Population and personal income per capita are proxy variables which reflect consumption at the origin, so the impacts are expected to be negative.

- The average plant size presents scale or diversification effects in the industry. It is estimated by dividing employment by the number of establishments in a sector. The assumption of the variable is that large establishments should produce and export more freight than small establishments.

Destination variables serve as proxies for commodity demands, both intermediate and final. Intermediate demand is the demand of manufacturing which uses freight as raw material in its production processes, while final demand represents the demand of customers. Since this research focuses on food and consumer goods, final demand is more appropriate than intermediate demand. Personal income per capita and total populations are proxies for final demand and the impacts are expected to be positive.

Geographical variables are variables that represent impedance of transportation between origin and destination. Distance is the most conventional friction variable used in all spatial interaction models, and is measured by highway distance. The impact of distance is expected to be negative.

Table 6.2 Input variable signs initial expectations

Variable	**Description**	**Expected sign**
Pro	Zone production	+
Att	Zone attraction	+
Dist	Distance	-
Oemp	Wholesale & retail sectoral employment at origin	+
Opop	Total population at origin	-
Oinc	Personal income per capita at origin	-
Opz	Average plant size at origin	+/-
Dpop	Total population at destination	+
Dinc	Personal income per capita at destination	+

6.4.2 Model structure

The commodity flow T_{ij} between two points (i,j) can be formulated as a function of the variables described:

CBCM model

$$T_{ij} = a_i(Pro) \,.\, b_j(Att) \,.\, g_{ij}(Dist)$$

SBCM model

$$T_{ij} = a_i(Oemp\ Opop\ Oinc\ Opz) \,.\, b_j(Dpop\ Dinc) \,.\, g_{ij}(dist)$$

Where

a_i is the supply point factor
b_j is the demand point factor
g_{ij} is the interaction factor

6.4.3 Discussion of the results

The results show that all variables have high chi square values, especially distance. The signs of the variables all agree with expected signs and the behavior of freight transportation. Each zone which has high production (*Pro*) or high attraction (*Att*) must generate high freight transportation while the distance between each pair of zones reduces freight transport volume between that pair.

Table 6.3 Parameters of CBCM model

Variable	Parameter	Chi square
Pro	0.417	289.168
Att	0.419	203.617
Dist	-1.248	308.454
Const	1.483	-

Table 6.4 Parameters of SBCM model

Variable	Parameter	Chi square
Oemp	0.599	9.212
Opop	-0.014	0.005
Oinc	0.455	22.924
Opz	0.097	0.525
Dpop	0.572	98.412
Dinc	0.492	68.913
Dist	-1.451	370.715
Const	-14.917	-

Almost all variables agree with the expected sign, except origin income (*Oinc*). Variable origin average plant size (*Opz*) and origin population (*Opop*) perform even more poorly. The unsatisfactory variables were excluded and a new model was calibrated. The results of the final model are shown in Table 6.3. All variables have high chi square values, especially distance.

Table 6.5 Parameters of SBCM final model

Variable	Parameter	Chi square
Oemp	0.668	124.317
Dpop	0.523	77.437
Dinc	0.509	67.517
Dist	-1.483	382.651
Const	-7.511	-

CBCM has an RMSE value of 89,413.20, MRE of 36.92, and R^2 of 0.4581 while SBCM has an RMSE value of 94,216.32, MRE of 30.56, and R^2 of 0.2910. Although CBCM has higher performance than SBCM, SBCM mimics trip distribution behavior using socio economic data which is useful for the planner to explore the constructive part of the trip distribution model.

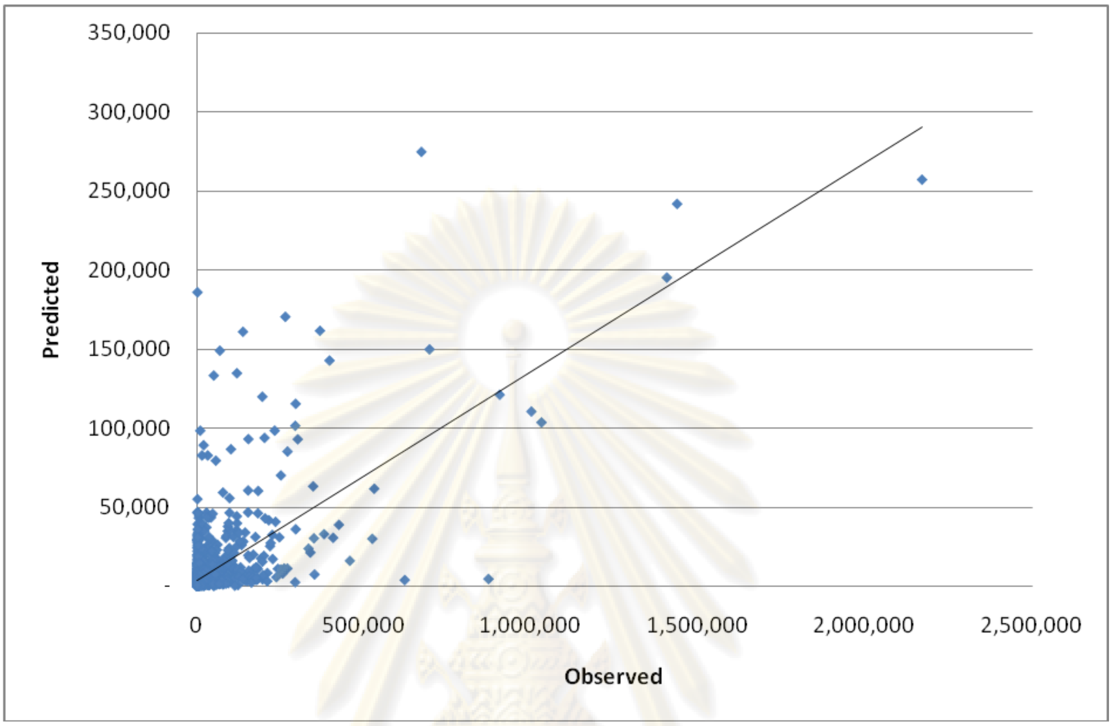

Fig 6.3 Comparing predicted and observed data of CBCM

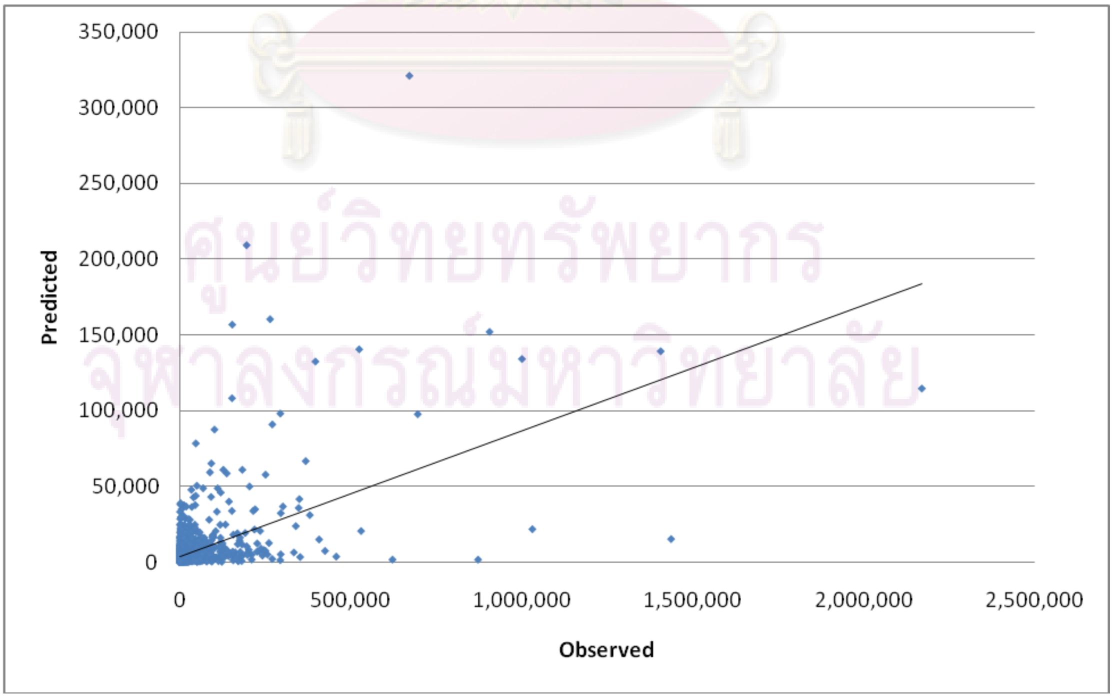

Fig 6.4 Comparing predicted and observed data of SBCM

6.5 ANFIS Model

To verify the performance of ANFIS against the Box-Cox model and the gravity model, two ANFIS models were developed. The first model (CAFM) used the same input variables as the gravity model, zonal total and zonal impedance. Another enhanced model (SAFM) used efficiency to explain the constructive part of the system derived from Box-Cox regression based model. The purpose of the first model is to compare with the gravity model in the context of performance of prediction. The latter model was developed to verify the performance of ANFIS for explaining the system using socio economic variables as inputs, which the gravity model is unable to do. Since ANFIS is not able to present a parameter of relation between input and output variables, a significant set of variables derived by a statistical procedure is used to ensure that input variables correlate with output.

6.5.1 Input Variable

The CAFM model uses a conventional gravity model with three input variables:

- Zone production, which expects the impact to be positive
- Zone attraction, which expects the impact to be positive
- Distance between zones, which expects the impact to be negative

The SAFM model uses only the statistically significant variables, at the 95% confidence level, from the set of variables identified by the Box-Cox regression based model:

- Origin employment in wholesale and retail sector (Oemp)
- Destination population (Dpop)
- Destination income (Dinc)
- Distance between province (Dist)

6.5.2 Development ANFIS model

The commodity flow T_{ij} between two points (i,j) can be formulated as a function of the variables described earlier:

CAFM model

$$T_{ij} = a_i(Pro) \,.\, b_j(Att) \,.\, g_{ij}(Dist)$$

SAFM model

$$T_{ij} = a_i(Oemp) \,.\, b_j(Dpop\ Dinc) \,.\, g_{ij}(dist)$$

Where

a_i is the supply point factor
b_j is the demand point factor
g_{ij} is the interaction factor

The model structure is developed using the fuzzy logic toolbox of the MATLAB software package. Three generalized bell-shaped membership functions are used for each of the inputs to build the ANFIS in this study. $\mu_{Ai}(x)$ is given by

$$\mu_{Ai}(x) = \frac{1}{1 + \left| \dfrac{x - c_i}{a_i} \right|^{2b}}$$

where {ai, bi, ci} is the parameter set of the membership functions in the premise part of the fuzzy if-then rules that change the shapes of the membership function.

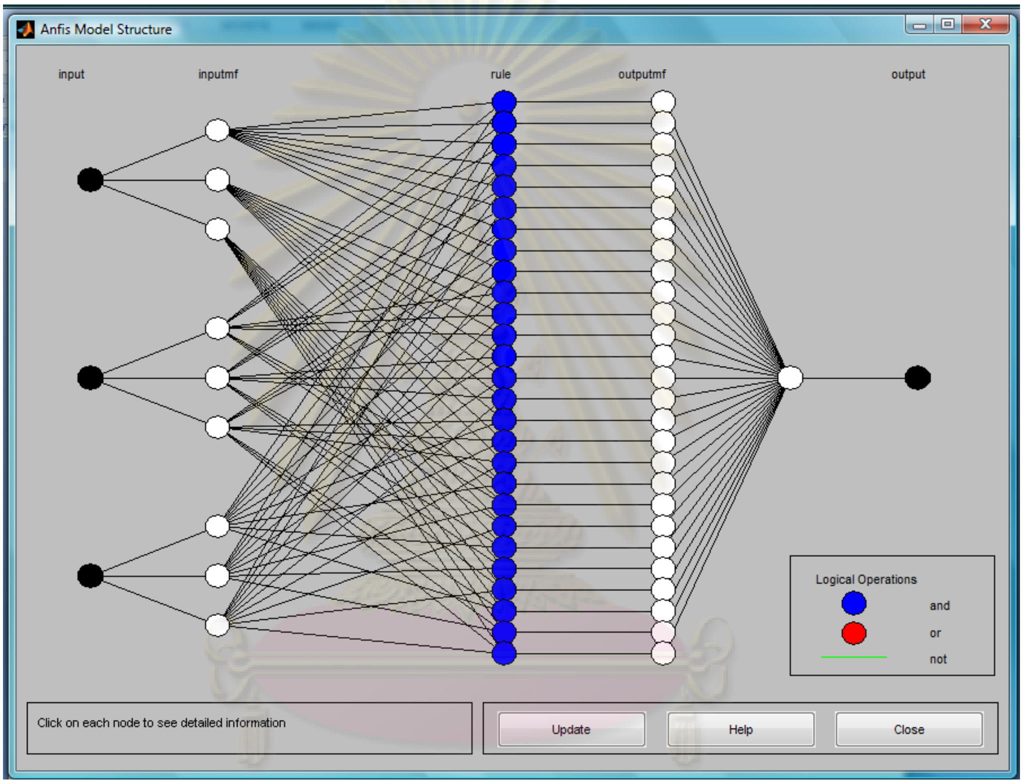

Fig 6.5 ANFIS Model structure for CAFM

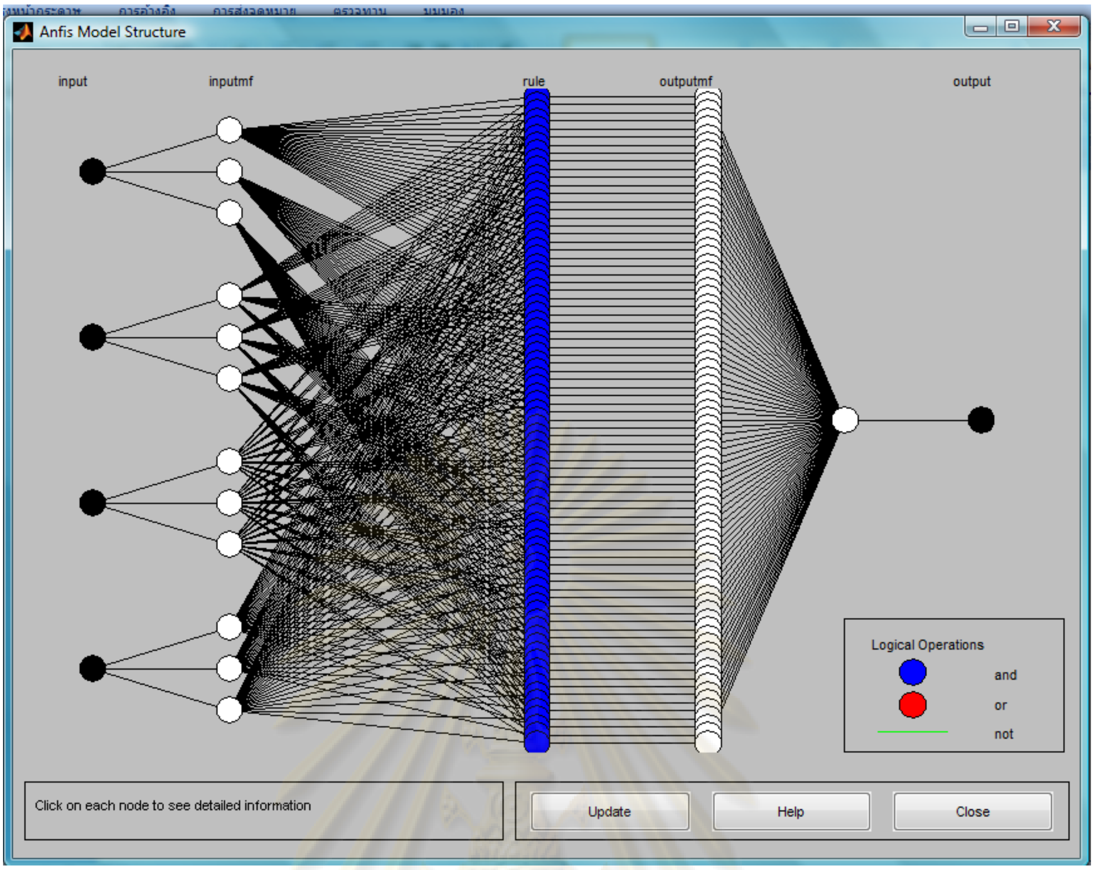

Fig 6.6 ANFIS Model structure for SAFM

This research applies the hybrid-learning algorithm that combines the gradient descent method and the least-squares method. The gradient descent method is used to assign the nonlinear input parameters while the least-squares method is used to identify the linear output parameters. Epoch is set as 500 in this study.

6.5.3 Discussion of the results

CAFM has an RMSE value of 51,621.53, MRE of 172.44, and R^2 of 0.7382. All statistical evaluations are better than the conventional gravity model even through the regression based model. In the picture below, the signs of all parameters agree with the expected signs. Zone production and zone attraction are directly related with trip while distance is inversely related with trip. These results confirm that the model structure accords with the behavior of freight distribution.

Table 6.6 Input variable signs: initial expectations of CAFM

Variable	**Description**	**Expected sign**	**CAFM**
Pro	Zone production	+	+
Att	Zone attraction	+	+
Dist	Distance	-	-

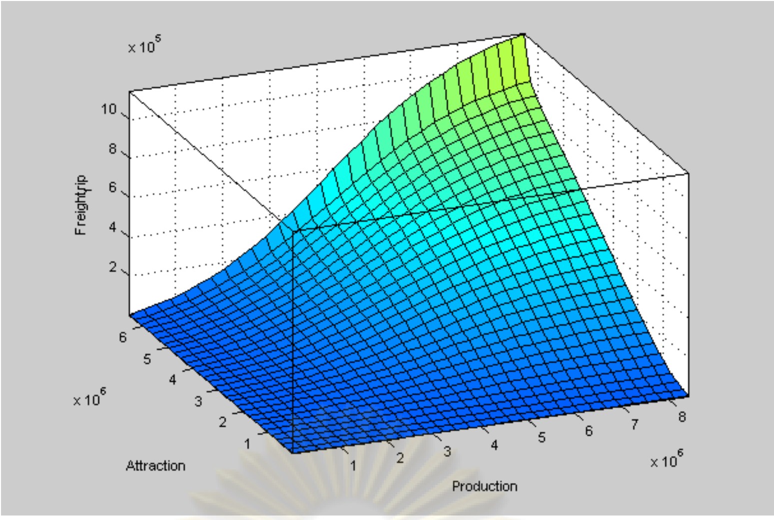

Fig 6.7 Influence of production and attraction on trip

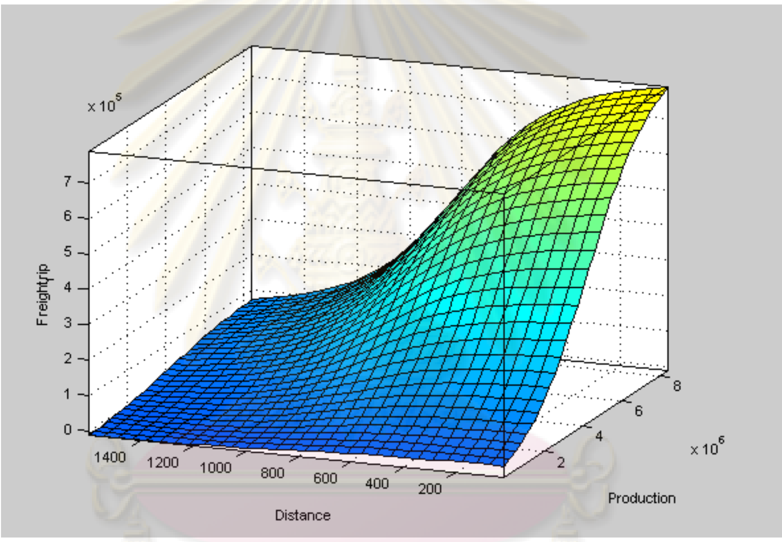

Fig 6.8 Influence of production and distance on trip

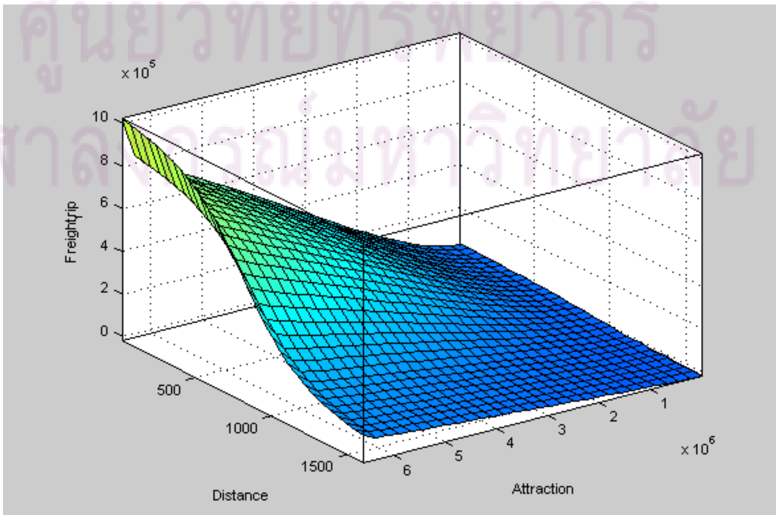

Fig 6.9 Influence of attraction and distance on trip

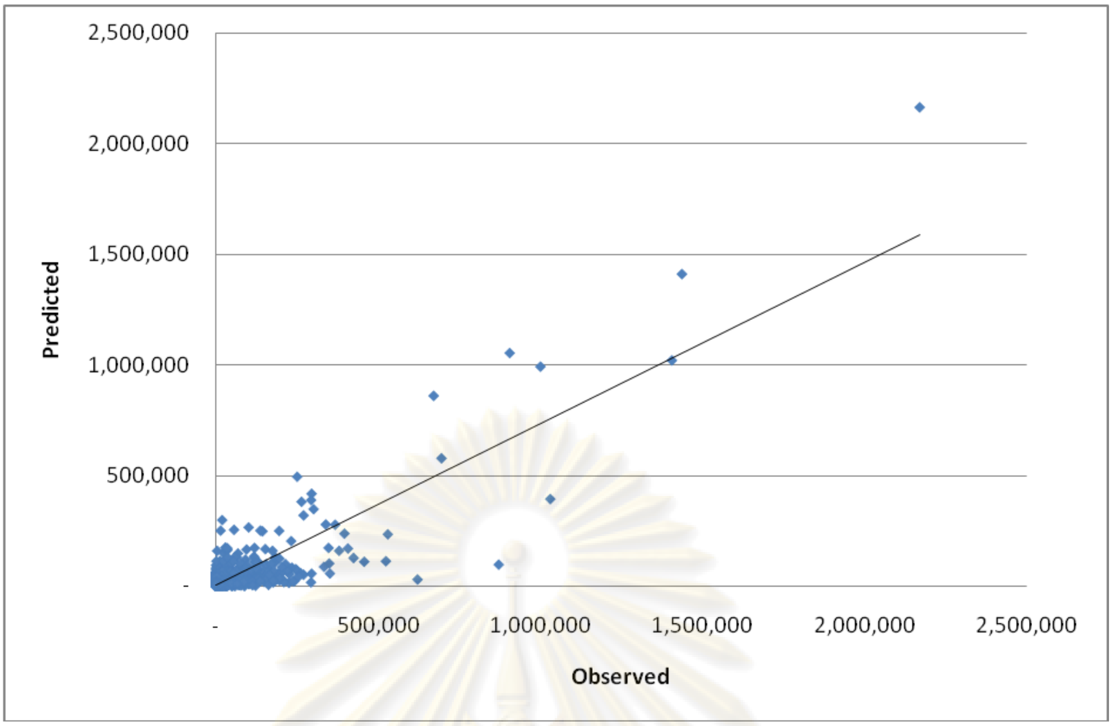

Fig 6.10 Comparing predicted and observed data of CAFM

SAFM has an RMSE of 80,791.520, MRE of 211.27, and R^2 of 0.3471. All statistical evaluations are better than the conventional gravity model but lower than the regression based model and ANFIS using conventional variables. Compared with SBCM, which use socio economic data as does SAFM, the performance of SAFM is better than SBCM. In others words, ANFIS improved the performance of the model. The picture below shows that the signs of all parameters agree with expected signs. Distance, which is impedance or friction of transportation, is inversely related to trip. On another hand, wholesale & retail sectoral employment at origin (Oemp), which represents the production of the zone, is directly related with trip. Moreover, total population at destination (Dpop) and personal income per capita at destination (Dinc), which represent consumption or attraction of the zone, are also directly related with trip. The result confirms that the model structure accords with the behavior of freight distribution.

Table 6.7 Input variable signs: initial expectations of SAFM

Variable	Description	Expected sign	SAFM
Dist	Distance	-	-
Oemp	Wholesale & retail sectoral employment at origin	+	+
Dpop	Total population at destination	+	+
Dinc	Personal income per capita at destination	+	+

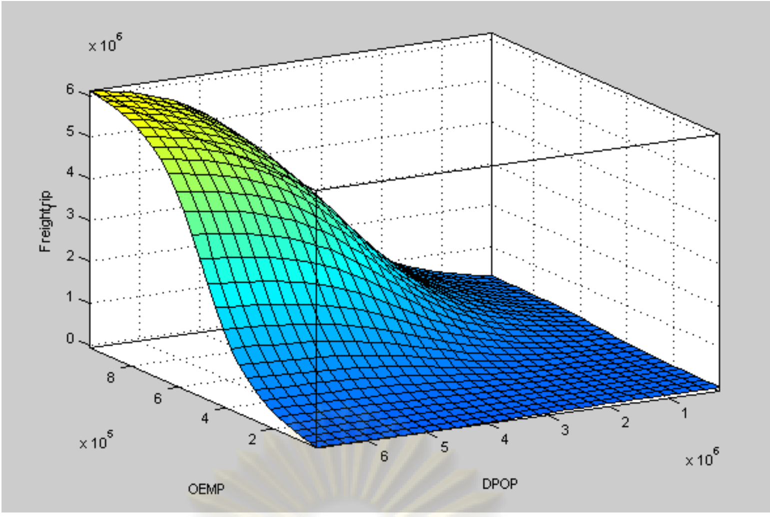

Fig 6.11 Influence of origin employment and destination population on trip

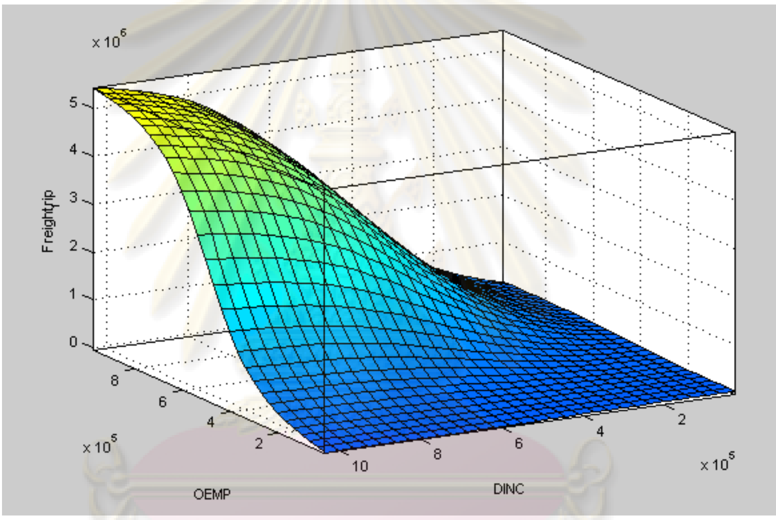

Fig 6.12 Influence of origin employment and destination income on trip

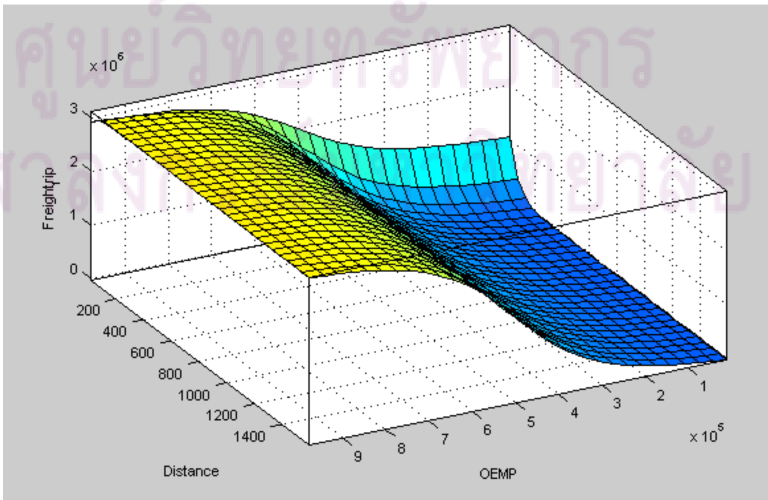

Fig 6.13 Influence of origin employment and distance on trip

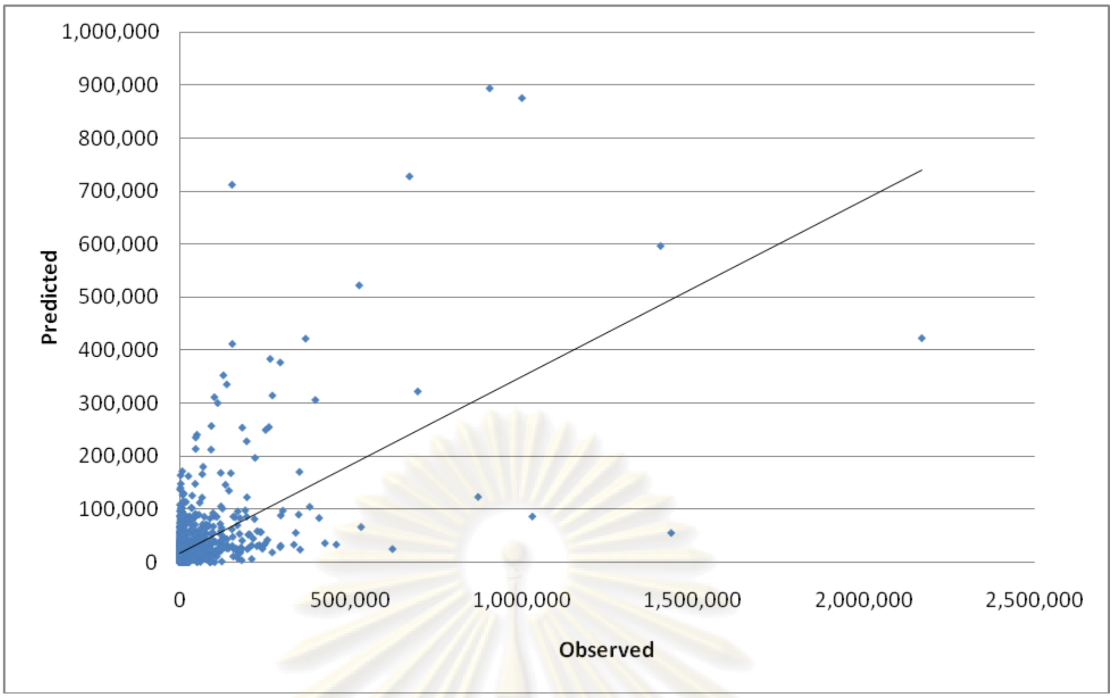

Fig 6.14 Comparing predicted and observed data of SAFM

6.6 Model evaluation

The results of Conventional gravity model (CGM), two Box-Cox Regression models, Conventional Variables (CBCM), Significant Variables (SBCM), and two ANFIS models [Conventional Variables (CAFM) and Significant Variables (SAFM) model] are compared and evaluated using root mean squared error (RMSE), mean relative error (MRE), and R^2 statistics.

The purpose of the performance evaluation in this study is to test whether an ANFIS and Box-Cox model outperform the conventional gravity model.

- Test whether the Box-Cox model using the three conventional gravity model variables outperforms the conventional gravity model
- Test whether the ANFIS model using the three conventional gravity model variables outperforms the conventional gravity model and the Box-Cox model
- Test whether this performance can be improved upon by inputting the statistically significant variables into these models

The comparison of the statistics evaluation of the developed models as is follows.

Table 6.8 Performance comparison of the models

Model	RMSE	MRE	R^2
CGM	101,632.03	215.61	0.5890
CBCM	89,413.20	36.92	0.4581
SBCM	94,216.32	30.56	0.2910
CAFM	51,621.53	172.44	0.7382
SAFM	80,791.52	211.27	0.3471

The evaluation indicates that, using conventional variables including zone production, zone attraction, and distance, CAFM has the best performance outperforms CGM. Although SAFM has lower statistical performance evaluation than CAFM, its performance is better than that of SBCM. These results show the performance of ANFIS for modeling complex systems. Many researchers have reported the superior performance of ANFIS, such as Aqil (2007), Chang and Chang (2006), and Firat and Gungor (2007). Moreover, the results indicate that Box-Cox transformation outperforms the conventional gravity model.

However, the performance of the model using socio economic data does not outperform the conventional variable model. The zonal total variable consolidates the robustness and error of the data, represented by a number of trips which are generated from or attract to the zone. Its use of socio economic data increases data robustness. Moreover, trip generation and attraction may be explained by a number of socio economic variables.

6.7 Summary and concluding remarks

To develop a gap-filling method, a model using Box-Cox transformation and ANFIS was developed and verified against the conventional gravity model. Two types of model, using conventional gravity variables and using socio economic variables, were developed.

The socio economic variable model used only the statistically significant variables, at the 95% confidence level, among the set of variables identified by Box-Cox regression based model:

- Origin employment in wholesale and retail sector (Oemp)
- Destination population (Dpop)
- Destination income (Dinc)
- Distance between province (Dist)

The purpose of performance evaluation in this study is to test whether an ANFIS and Box-Cox model outperforms the conventional gravity model.

- Test whether the Box-Cox model using the three conventional gravity model variables outperforms the conventional gravity model
- Test whether the ANFIS model using the three conventional gravity model variables outperforms the conventional gravity model and the Box-Cox model

- Test whether this performance can be improved upon by inputting the statistically significant variables into these models

The results of this evaluation show that the ANFIS model outperforms both the conventional gravity model and the Box-Cox model. These results confirm the superior performance of the Adaptive Neuro Fuzzy Inference System (ANFIS) for modeling complex systems. These results agree with many researchers who have reported on the superior performance of ANFIS, such as Aqil (2007), Chang and Chang (2006), and Firat and Gungor (2007). Moreover, these results indicate that the Box-Cox transformation model outperforms the conventional gravity model.

Chapter VII
Freight origin destination matrix development

This chapter presents a comprehensive framework for developing the freight origin destination matrix and presents details of the developed matrix. The last part is dedicated to concluding remarks.

7.1 Final matrix development

The model evaluation results in Chapter 6 indicate that the ANFIS model using conventional variables outperform other models. Therefore the ANFIS model is employed to develop the origin and destination matrix of combined food and consumer goods in Thailand. The trained ANFIS model is applied to empty cells data. The results reveal the following:

A total of 28,054,800 tons of freight are filled in the 91,912,662 tons of the whole matrix, or 29.69%. A total of 2,830 empty cells are filled while the remaining 968 cells had volume less than one ton per year. These cells are classified as true zero cells. There are two methods for augmenting the empty cells in the matrix. The first is to add a given value to all of the matrix's cells, which will place values in all empty cells but which will also change marginal total and trip length distribution. This method assumes that an adjusted matrix is still under-reported, and therefore the marginal total must also be adjusted. The second method is to fill the empty cells while keeping the marginal total and the trip length distribution. This method assumes that the marginal total is accurate. The differences of trip length distribution of these two methods are shown below.

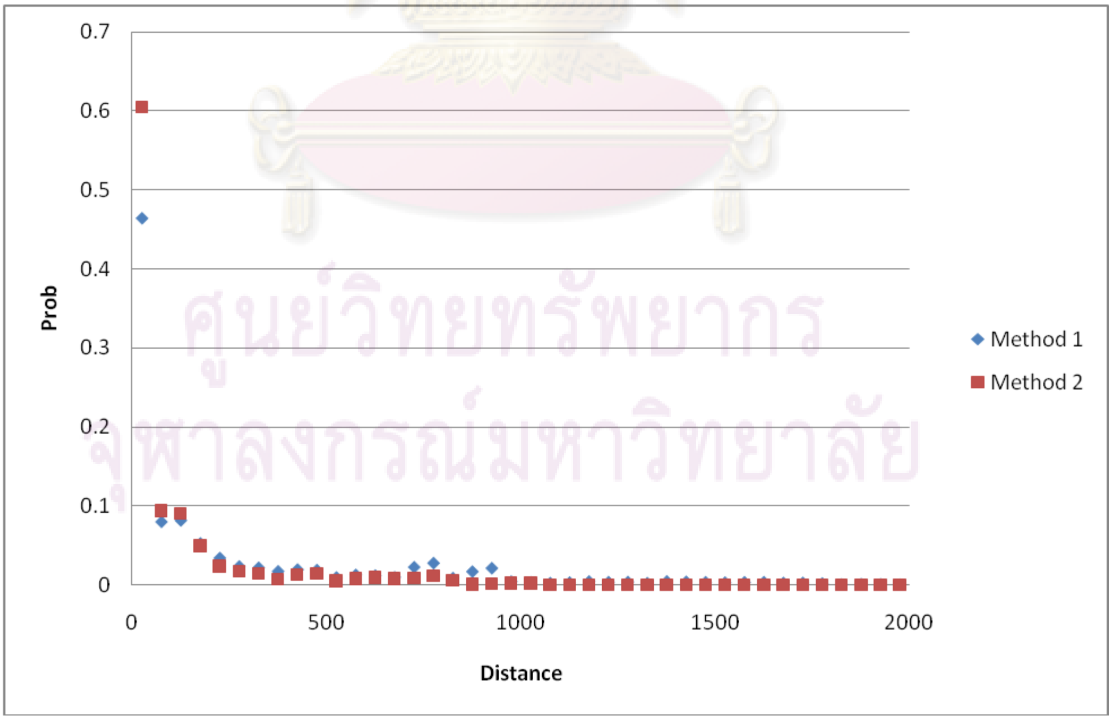

Figure 7.1 Comparing trip length distributions of zero-cell augmentation methods

To evaluate these methods, the marginal total of each zone is employed to develop the trip generation models. The trip production model uses wholesale & retail sectoral employment (EMP) as input variables while the trip attraction model uses gross province product (GPP) as input variables.

Our results show that the second method has a higher R^2 and stronger statistically significant parameters. Moreover, the second method has less bias. It can therefore be concluded that Method 2 outperforms Method 1 and is preferable to Method 1.

Table 7.1 Trip production models

Item	Method 1	Method 2
Model	395,348+7.564(EMP)	98,058+6.639(EMP)
t-stat of variables	2.256 , 5.854	0.841 , 7.721
R^2	0.316	0.446

Agreeing with the trip production models, the calibrated trip attraction models also indicate that Method 2 has a higher R^2 and statistically significant parameters than Method 1, as well as less bias, making Method 2 preferable to Method 1.

Table 7.2 Trip attraction models

Item	Method 1	Method 2
Model	624,430+ 3.085(GPP)	301,972+ 2.682(GPP)
t-stat of variables	4.571 , 6.570	3.853 , 9.957
R^2	0.368	0.573

The results of this evaluation show that augmenting zero cells while keeping the marginal total and trip length distribution is preferable to adding to zero cells while superimposing values over reliable marginal totals and abandoning the existing trip length distribution. Thus, to create the final adjusted matrix in the next step, zero cells are augmented while keeping marginal total and trip length distribution.

7.2 Details of combined food and consumer goods transportation

Using the method developed in Chapters 5 and 6, the final matrix of combined food and consumer goods origin destination matrix is developed. This matrix reveals the importance characteristics of combined food and consumer goods transportation within the whole kingdom of Thailand.

The average trip length reported by this final adjusted matrix is 125.33 kilometers, which is less than reported from original CFS matrix or the original RS matrix, which reported 143.64 and 168.65 kilometers respectively. There is a slight difference in average trip length from the final adjusted matrix and the original CFS matrix. The details of combined food and consumer goods shipped in Thailand are shown below.

Bangkok generates the most transported freight. A total of combined food and consumer goods of 14,197,631 tons per year were produced in Bangkok. A total of

7,268,412 tons were transported to other provinces while 6,929,219 tons were consumed in Bangkok. These results are not surprising because many wholesale and retail establishments are located in Bangkok.

Table 7.3 Top ten origin provinces of combined food and consumer goods

Province	Volume (tons/year)		
	All	Internal	External
Bangkok	14,197,631	7,268,412	6,929,219
Saraburi	9,507,835	8,454,166	1,053,669
Pathum Thani	6,571,348	493,653	6,077,695
Surat Thani	3,688,476	1,869,241	1,819,235
Samut Sakhon	2,506,283	452,431	2,053,852
Chonburi	2,310,804	1,254,708	1,056,096
Khon kaen	2,198,762	852,031	1,346,731
Nakhon Pathom	2,111,078	778,677	1,332,401
Uttaradit	1,963,314	1,459,065	504,249
Nakhon Ratchasima	1,806,789	967,420	839,368

Bangkok attracts the most freight transport. A total of combined food and consumer goods of 12,837,521 tons per year were shipped to Bangkok. A total of 7,268,412 tons originated within Bangkok while 5,569,109 tons were shipped from other provinces.

Table 7.4 Top ten destination provinces of combined food and consumer goods

Province	Volume (tons/year)		
	All	Internal	External
Bangkok	12,837,521	7,268,412	5,569,109
Saraburi	11,870,750	8,454,166	3,416,583
Ayudthaya	3,750,446	77,221	3,673,225
Pathum Thani	3,749,746	493,653	3,256,093
Samut Prakan	3,445,261	400,815	3,044,446
Chonburi	3,383,725	1,254,708	2,129,017
Surat Thani	2,633,618	1,869,241	764,377
Ratchaburi	1,983,900	906,832	1,077,068
Uttaradit	1,736,204	1,459,065	277,140
Nakhon Ratchasima	1,735,312	967,420	767,891

Bangkok also has the highest volume shipped from origin to other provinces, with 6,929,219 tons shipped from Bangkok to other provinces. This is followed by Pathum Thani with an export volume to other provinces of 6,077,695 tons. Third is Samut Sakhon, which ships 2,053,852 tons to other provinces. The export volumes of Surat Thani and Khon kaen, which are central provinces of the south and northeast, are 1,819,235 tons and 1,346,731 respectively.

Table 7.5 Top ten transports out of provinces of combined food and consumer goods

Province	Volume (tons/year)
Bangkok	6,929,219
Pathum Thani	6,077,695
Samut Sakhon	2,053,852
Surat Thani	1,819,235
Khon kaen	1,346,731
Nakhon Pathom	1,332,401
Krabi	1,323,871
Samut Prakan	1,267,699
Ayudthaya	1,120,469
Chonburi	1,056,096

Bangkok has highest volume of freight shipped into the province, 5,569,109 tons. This is followed by Ayudthaya, which has 3,673,225 tons of shipped volume. Third is Saraburi with 3,416,583 tons. The data reveals that the six most attractive freight destinations are Bangkok and nearby provinces: Pathum Thani, Bangkok, Phra Nakhon Si Ayudthaya, Samut Prakan, Saraburi, and Chonburi.

Table 7.6 Top ten transports into provinces of combined food and consumer goods

Province	Volume (tons/year)
Bangkok	5,569,109
Ayudthaya	3,673,225
Saraburi	3,416,583
Pathum Thani	3,256,093
Samut Prakan	3,044,446
Chonburi	2,129,017
Ratchaburi	1,077,068
Nonthaburi	1,022,133
Nakhon Sawan	837,625
Nakhon Pathom	822,100

The destination of freight shipped from Bangkok is presented in the table below. Most freight from Bangkok was transported to provinces in its vicinity. The most volume from Bangkok was transported to Pathum Thani (1,402,319 tons). The second destination is Phra Nakhon Si Ayudthaya (902,328 tons). Moreover, significant volumes of freight were transported from Bangkok to two eastern provinces, Chonburi and Rayong, in the amounts of 816,776 tons and 360,532 tons respectively.

Table 7.7 Top ten destinations of freight shipped from Bangkok

Province	Volume (tons/year)
Pathum Thani	1,402,319
Phra Nakhon Si Ayudthaya	902,328
Chonburi	816,776
Samut Prakan	669,424
Nonthaburi	394,966
Rayong	360,532
Saraburi	304,350
Kanchanaburi	289,203
Ratchaburi	207,289
Samut Sakhon	194,214

The origin of freight shipped to Bangkok is presented in the table below. Most freight shipped to Bangkok was transported from provinces in its vicinity. The largest volume shipped to Bangkok was from Pathum Thani (2,164,553 tons). The second most common origin is Samut Sakhon (693,519 tons). Moreover, significant volumes of freight were transported to Bangkok from the eastern and northeastern provinces of Chonburi and Khonkaen, in the amounts of 243,505 tons and 162,115 tons respectively.

Table 7.8 Top ten origins of freight shipped to Bangkok

Province	Volume (tons/year)
Pathum Thani	2,164,553
Samut Sakhon	693,519
Nakhon Pathom	264,627
Samut Prakan	262,940
Chonburi	243,505
Samut Songkhram	183,190
Prachin Buri	177,369
Phra Nakhon Si Ayudthaya	164,312
Khon kaen	162,115
Nonthaburi	152,542

7.3 Discussion and concluding remarks

In order to develop the combined food and consumer goods origin destination matrix, the best performance model from Chapter 6 was used. A model using Adaptive Neuro Fuzzy Inference System (ANFIS) with conversion variables consisting of zone production, zone attraction, and distance was applied to empty cell data. A total of 28,054,800.00 tons of freight are filled in the 91,912,662 tons of the whole matrix, or 29.69%. A total of 2,830 empty cells are filled while the remaining 968 cells are true zero cells.

There are two alternative methods to augment empty cells. The first is to add a value to both the empty cells and the non-empty cells throughout the matrix, thus changing the marginal total and the trip length distribution. The latter is to fill the empty cells while keeping marginal total and trip length distribution. To compare these methods, marginal totals were used to develop two trip generation models. The results of this comparison show that augmenting the empty cells while keeping marginal total and trip length distribution is preferable to augmenting all cells while sacrificing existing marginal total and trip length distribution.

The developed matrix reveals that Bangkok generates the most transported freight. A total of combined food and consumer goods of 14,197,631 tons per year were produced in Bangkok, 7,268,412 tons of which were transported to other provinces and 6,929,219 tons which were consumed in Bangkok. Moreover, Bangkok attracts the most freight transport. A total of combined food and consumer goods of 12,837,521 tons per year were shipped to Bangkok. A total of 7,268,412 tons originated within Bangkok while 5,569,109 tons were shipped from other provinces.

Chapter VIII
Conclusion and recommendations

8.1 Conclusion

Freight flow data, as well as passenger flow data, plays an important role in transportation planning. At present, limited freight data has led to more extensive passenger flow studies than freight flow studies. Since the flow of freight is relatively about economic activities of origin and destination, the data for freight models involve many shippers, manufacturers, and receivers, making it costlier and more difficult to collect than passenger flow data.

The Thai government has attempted to solve this lack of freight data by setting up a national logistics data strategy. This strategy propels the two governmental agencies, National Statistics Office (NSO) and Department of Land Transport (DLT), to collect the important transport and logistics data including Commodity Flow Survey (CFS) and truck O-D survey by the roadside interview method.

Since CFS is a pioneering project, the published data provided by NSO astonished the well-informed, especially the total volume shipped in Thailand. Moreover, due to budget limitations of the truck O-D survey, data were collected at 10 target provinces, and they were collected in the harvest season and out of harvest season rather than quarterly throughout the year. For these reasons, these two sets have a number of weak points. Development of freight flow data must use the strengths of these two data. CFS collects data from sample establishments across the country, making its distribution pattern acceptable, while the marginal total of truck O-D survey has been accepted and verified with available data source.

Moreover, empty cells are a major problem in the freight origin destination matrix. These empty cells are a gap in O-D matrices which must be explored to determine which cells should contain positive flows and what size of flow should be filled. The most commonly used method for filling gaps is the gravity model. The weaknesses of this method are its inability to explain the relationship of explanatory factors and socio economic variable of study area, and that it requires a large amount of data for calibration. Another method is the regression-based method, which can overcome a major inefficient point of the gravity technique but which still suffers from performance of the model. Recently, Adaptive Nero Fuzzy Inference System (ANFIS), a soft computing technique, has been accepted as an efficient alternative tool for modeling complex non-linear systems and widely used for prediction. The new method has proven to be an efficient tool in many disciplines, including transportation engineering. Since studies using ANFIS as a tool for this approach is limited, one merit of this research is its use of ANFIS as an efficient gap-filling method.

This research has two main purposes: developing a new method for combining CFS with roadside survey data, and evaluating an Adaptive Neuro Fuzzy Inference System (ANFIS) gap-filling method against a conventional gravity model and a regression-based model using Box-Cox transformation.

In the first context, this research develops two new methods for combining CFS and roadside survey data. The first method is Trip Length Distribution Adjustment (TLDA) which aims to adjust the bare CFS original destination matrix which under-reports marginal totals to meet the more reliable marginal totals of roadside interview data. Two data sets, RS1 (Department of Land Transport roadside survey data) and RS2 (Roadside Survey data from Chonburi road network strategic planning for supporting logistics development project), were used to calibrate the method and test its performance in adjusting data from difference sources. The results indicate that the adjusted volumes of RS1 and RS2, which collected data at different times using different authors, are in agreement. The difference in total adjusted volumes of RS1 and RS2 is quite small. The second method is using Gravity Model Approach (GMA) to adjust a 10x76 matrix of RS1 using the CFS distribution pattern because the CFS distribution pattern is preferable to available data. The 76x76 matrix of CFS is used to calibrate impedance function. The results indicate that the freight transportation adjustment agreed with the TLDA method.

For these reasons, it can be concluded that the developed method can be reliably used to adjust the data. Using the developed method, two matrices were developed. The first is CFS based using a TLDA adjustment factor to produce a 76x76 combined food and consumer goods matrix. The other is an RS1 based matrix using GMA to produce a 10x76 matrix which uses an auxiliary matrix for empty cell verification and augmentation in the second context.

However, a major asset to this combination method is that it reliably reflects the trip length distribution of the CFS results. Thus, the quality of the CFS data is obviously a key part of the analysis. On the other hand, the high quality of the CFS data may reduce the magnitude of the adjustment factor and thus the reliability of the final adjusted matrix. For this reason, improving the CFS survey method in subsequent surveys to further improve data quality is an important task for future research.

In the second context, the model using Box-Cox transformation and ANFIS are developed and verified against a conventional gravity model. Two types of model, using conventional gravity variables and using socio economic variables, were developed. The results show that the ANFIS model outperformed the conventional gravity and Box-Cox models. The results prove the performance of ANFIS for modeling complex systems, which agrees with many researchers who report on the superior performance of ANFIS, such as Aqil (2007), Chang and Chang (2006), and Firat and Gungor (2007). Moreover, the results indicate that Box-Cox transportation outperforms the conventional gravity model. Although ANFIS has the same limitations for modeling trip distribution as other soft computing techniques, such as neural network or fuzzy logic, ANFIS is preferable to conventional methods of studying distribution behavior. The important task for any gap-filling method is to understand the behavior of the freight distribution pattern; therefore ANFIS is more successful at filling gaps in the trip distribution matrix.

8.2 Recommendations

This research proposes a new method for combining CFS and roadside interview data. Combined food and consumer goods freight was selected among 37 commodity categories for calibrating and verifying the method because it was generated and distributed in all provinces. However, to develop freight flow data for modeling in Thailand, all commodities must be adjusted. Moreover, the quality of CFS data is important part of an analysis, so improving the CFS survey method in the next survey to further improve the quality of the data is an important task for future research.

References

Aqil, M. et al. 2007. Analysis and prediction of flow from local source in river basin using a neuro-fuzzy modeling tool. Journal of Environment Management 85 (October): 215-223.

Bergkvist, E., Westin L. 1997. Estimation of gravity models by OLS estimation, NLS estimation, Poisson and Neural Network specifications[Online]. UMEÅ University CERUM, Sweden. Available from: http://www.umu.se/cerum[2008, October 20]

Black, W.R. 1995. Spatial interaction modeling using artificial neural network. Journal of Transport Geography 3 (September): 156-166.

Box, G. E. P., Cox D. R. 1964. An analysis of transformation. Journal of the Royal Statistical Society. Series B (Methodological) 26: 211-252. Available from: http://www.jstor.org/stable/2984418?cookieSet=1[2009, May 31]

Celik, H.M. 2004. Modeling freight distribution using artificial neural networks. Journal of Transport Geography 12 (June): 141-148.

Celik, H.M. 2010. Sample size needed for calibrating trip distribution and behavior of the gravity model. Journal of Transport Geography 18 (January): 183-190.

Celik, H.M. Guldmann J.M., 2007. Spatial interaction modeling of interregional commodity flows. Journal of Socio-Economic Planning Sciences 41 (June): 147-162.

Cesario, F.J. 1976. The calibration of trip distribution model: Comments on the new philosophy of Rose[Online]. Urban study 14 (June): 229-231. Available from: http://www.sagepublications.com[2009, May 31]

Chang, F.F., Chang Y.T. 2006. Adaptive neuron-fuzzy inference system for prediction of water level in reservoir. Advance in Water Resources 29 (January): 1-10.

Dougherty, M. 1995. A review of neural network applied to transport. Transportation Research Part C: Emerging 3 (August): 247-260.

Erlander S, Jornsten, K.O., and Lundgren, J.T. 1985. On the estimation of trip matrices in the case of missing and uncertain data. Transportation research part B: Methodological 19 (April): 123-141.

Firat, M., Gungor M. 2007. River flow estimation using adaptive neuron fuzzy inference system. Mathematics and Computers in Simulation 75 (July): 87-96.

Fotheringham, A.S., O'Kelly, M.E. 1989. Spatial Interaction Models: Formulation and Applications. Kluwer Academic Publishers, Dordrecht/Boston/London.

Guldmann, J.M. 1998.Competing destinations and intervening opportunities interaction models of inter-city telecommunication flows[Online]. Paper prepared for presentation at the 38th Congress of the European Regional Science Association. Available from: http://www.ersa.org[2009, July 25]

Gunyoung, K. 2003.Analysis of trip length distribution between commodity-based model and truck trip based model in Seoul metropolitan area. Proceeding of the Eastern Asia Society for Transportation Studies 4, pp.1242-1255.

Esen, H. et al. 2008. Artificial neural networks and adaptive neuro-fuzzy assessments for ground-coupled heat pump system. Energy and Building 40: 1074-1083.

Hall, D.L., Llinnas, J. 1997. An introduction to multisensor data fusion. Proceedings of IEEE 85 (1): 6-23.

Jang, J.S.R., Sun C.T., and Mizutani E. 1997. Neuro-Fuzzy and soft computing: a computation approach to learning and machine intelligence. Prentice-Hall, Eaglewood Cliffs, NJ.

Jou, Y.J. et al. 2006. Incomplete Information Analysis for the Origin-Destination Survey Table. Journal of Urban Planning and Development 132 (December): 193-200.

Kalic, M., Teodorvic D. 2003. Trip distribution modeling using fuzzy logic and genetic algorithm. Journal of Transportation Planning and Technology 26 (June): 213-238.

Kam, M., Zhu, X., Kalata, P. 1997. Sensor fusion for mobile robot navigation. Proceedings of IEEE 85 (1): 108-119.

Kirby, H. R. 1979. Partial matrix techniques. Traffic engineering & control.

Kuwahara, M., Sullivan E. 1987. Estimating origin-destination matrices from roadside survey data. Transportation research part B: Methodological 21 (June): 233-248.

Ogden, K.W. 1978. The distribution of truck trips and commodity flow in urban areas: a gravity model analysis. Transportation research 12 (April): 131-137.

Southworth, F. 2005. Filling Gaps in the U.S. Commodity Flow Picture: Using the CFS with Other Data Sources. Presented at The 2005 Commodity Flow Survey Users' Conference Boston, MA. pp.29-46.

Tellema, F et al. 2006. Comparison of neural networks and gravity models in trip distribution. Computer-Aided Civil and Infrastructure Engineering 21(2): 104-119.

Teodorovic, D. 1999. Fuzzy logic systems for transportation engineering: the state of the art. Transportation research Part A: Policy and Practice 33 (June): 337-364.

The Office of National Statistics of Thailand. 2008. Development of commodity flow database: full scale survey action plan report (in Thai). Bangkok : The Office of National Statistics of Thailand.

Wang, Y.M, Elhag T. 2008. An adaptive neuron-fuzzy inference system for bridge risk assessment. Expert System with Applications 34 (May): 3099-3016.

Appendix

Table A1 Population and employment data Unit : person

Jurisdiction Code	Province	Population	Employment				
			Total	Production	Retail and	Manufacture	Agriculture
10	Bangkok	6,841,756	3,894,464	796807	978006	567710	42556
11	Samut Prakan	1,270,142	746,155	328107	111600	487225	4154
12	Nonthaburi	952,298	529,456	93385	105820	78141	21423
13	Pathum Thani	805,803	454,304	150836	87497	221534	40303
14	Phra Nakhon Si Ayudthaya	761,077	413,978	170130	50740	175635	47120
15	Ang Thong	273,341	151,521	41665	21747	7601	44186
16	Lop Buri	771,263	444,229	93345	69596	35114	149950
17	Sing Buri	235,321	129,568	20442	24268	12387	40841
18	Chai Nat	364,684	207,715	22321	28029	8798	101735
19	Saraburi	602,575	344,679	134021	50644	73927	50927
20	Chonburi	1,169,285	676,044	216789	116467	226674	35024
21	Rayong	584,138	341,453	101135	48362	131847	78040
22	Chanthaburi	526,656	321,773	30211	43473	11946	162163
23	Trat	239,650	145,309	7070	19130	4570	69720
24	Chachoengsao	697,587	384,941	121711	50581	124190	101962
25	Prachin Buri	443,189	246,613	81204	24813	73063	65384
26	Nakhon Nayok	257,879	143,222	32099	22582	7541	40378
27	Sa Keaw	531,884	299,446	35268	45552	8747	165353
30	Nakhon Ratchasima	2,773,326	1,486,587	266336	251986	116145	597247
31	Buri Ram	1,622,157	898,856	88755	93834	14229	542125
32	Surin	1,422,754	773,304	46004	85860	10872	496805
33	Si Sa Ket	1,514,723	950,891	15989	71139	6640	759455
34	Ubon Ratchathani	1,837,171	1,014,638	29647	114983	15958	655923
35	Yasothon	608,910	379,667	33259	29435	4431	238837

99

Jurisdiction Code	Province	Population	Employment				
			Total	Production	Retail and	Manufacture	Agriculture
36	Chaiyaphum	1,182,709	666,147	58291	60995	39351	446248
37	Amnat Chareon	392,436	238,625	15583	26842	2719	157570
39	Nong Bue Lumphu	526,047	296,032	15739	29992	5046	196170
40	Khon kaen	1,866,131	995,880	143748	147198	53215	470478
41	Udon Thani	1,599,775	824,410	20976	105475	22854	512593
42	Loei	652,835	351,646	22022	47495	9656	202928
43	Nong Khai	957,696	516,433	27632	62585	6135	260125
44	Maha Sarakham	1,016,667	582,488	73839	82933	14185	305542
45	Roi Et	1,344,393	821,429	86573	91066	11659	506295
46	Kalasin	995,003	545,526	22542	55385	18420	369749
47	Sakon Nakhon	1,133,470	636,005	17380	71277	6630	436121
48	Nakhon Phanom	738,184	421,087	8471	59768	4422	261908
49	Mukdahan	337,714	208,572	8345	25223	3185	143810
50	Chiang Mai	1,583,657	964,674	151009	162994	43218	274912
51	Lamphun	432,646	266,518	77347	40714	37326	89956
52	Lampang	814,889	467,544	54189	59513	29134	228931
53	Uttaradit	486,214	266,692	19391	37261	5580	140908
54	Phrae	514,200	292,720	44447	38488	12678	133178
55	Nan	484,696	291,888	15978	26267	2914	177692
56	Phayao	529,554	313,471	28767	41545	4936	168874
57	Chiang Rai	1,194,409	697,872	41982	84002	10926	372998
58	Mae Hong Son	227,854	143,282	1309	9824	880	104990
60	Nakhon Sawan	1,142,397	662,812	40530	127919	21327	340458
61	Uthai Thani	317,289	185,547	11501	21019	3673	109755
62	Kamphaeng Phet	710,093	419,264	20061	46619	8483	275890
63	Tak	519,966	293,649	24183	30257	46085	168349

100

Jurisdiction Code	Province	Population	Employment				
			Total	Production	Retail and	Manufacture	Agriculture
64	Sukhothai	623,293	347,788	29064	40614	7498	204902
65	Phitsanulok	838,119	483,788	36358	71799	29318	216426
66	Phichit	594,611	325,879	27980	41656	8142	170319
67	Phetchabun	1,023,194	587,083	35425	88596	19486	317535
70	Ratchaburi	825,468	489,723	94078	126465	55556	135783
71	Kanchanaburi	774,016	469,811	65095	64634	27901	235343
72	Suphan Buri	884,385	526,637	52229	74706	19165	260213
73	Nakhon Pathom	948,188	569,280	172537	105173	150918	114242
74	Samut Sakhon	558,866	353,262	190569	57622	298681	18837
75	Samut Songkhram	209,985	125,882	41236	25032	8310	18476
76	Phetchaburi	454,541	269,154	36859	47761	19950	86666
77	Prachuap Khiri Khan	474,376	292,679	31981	50402	24518	101243
80	Nakhon Si Thammarat	1,671,356	913,996	77740	151950	21517	399766
81	Krabi	380,585	208,917	7097	31377	8390	111649
82	Phangnga	259,402	142,197	11575	24081	4979	59760
83	Phuket	289,139	152,493	15025	30132	6987	5379
84	Surat Thani	974,115	552,285	36838	91268	27647	258063
85	Ranong	182,889	100,699	6612	20251	5172	44572
86	Chumphon	494,398	288,714	11749	44686	12583	137269
90	Songkhla	1,409,549	779,546	80218	164263	74461	223483
91	Satun	280,117	156,423	5740	23161	4025	79277
92	Trang	663,193	394,630	24692	62330	21647	192778
93	Phatthalung	548,135	318,352	17951	40724	4204	160239
94	Pattani	663,118	337,894	41445	67531	9274	116538
95	Yala	465,734	234,168	7482	25093	6081	153131
96	Narathiwat	742,002	379,682	46653	55825	3905	132424

Table A2 Gross province product data

Jurisdiction Code	Province	Gross province product (Million bath)			
		Total	Agriculture	Production	Retail and wholesale
10	Bangkok	2,216,996.90	3524.2	477895.2	535796.3
11	Samut Prakan	614,124.50	783.7	411866.6	24778.9
12	Nonthaburi	111,115.50	3238.6	38997.1	15581.7
13	Pathum Thani	192,947.80	4491.1	129614.1	9927.5
14	Phra Nakhon Si Ayudthaya	337,826.30	6804.7	280161.4	10502.7
15	Ang Thong	20,836.60	3124.1	3329.5	4645
16	Lop Buri	70,235.30	9872.4	23552.3	9267.8
17	Sing Buri	22,140.40	3797.2	6199.7	3970.4
18	Chai Nat	26,345.70	7640.8	2611.9	5605.3
19	Saraburi	134,028.70	6805.1	76859.7	10352.7
20	Chonburi	453,885.80	15151.8	263494.8	37127.5
21	Rayong	604,896.10	13499.7	301934.2	12755.7
22	Chanthaburi	38,214.50	11345.2	2328.1	7182.2
23	Trat	20,309.10	4501.8	756.8	1956
24	Chachoengsao	210,530.40	9592.2	160559.4	13111.9
25	Prachin Buri	70,292.10	5643.5	32091.5	18479.4
26	Nakhon Nayok	16,946.20	2723.2	1362.1	3612.2
27	Sa Keaw	29,523.90	7774.5	4169.3	7856.4
30	Nakhon Ratchasima	150,763.10	29342.2	32326.7	22555.6
31	Buri Ram	51,006.60	11760.6	6470.9	11318.7
32	Surin	45,185.40	9127.3	5127.7	10704.4
33	Si Sa Ket	44,191.10	10320.9	2811.9	11477.8
34	Ubon Ratchathani	67,389.00	10103.7	8175.7	17456.3
35	Yasothon	19,508.40	4283.8	1899.8	4441.3

102

Jurisdiction Code	Province	Gross province product (Million bath)			
		Total	Agriculture	Production	Retail and wholesale
36	Chaiyaphum	42,078.80	11136.2	5752.7	9007.2
37	Amnat Chareon	12,153.70	3282.6	492.9	2103.1
39	Nong Bue Lumphu	15,373.30	4393.9	2168	1955
40	Khon kaen	127,088.70	14406.6	46262	20465.3
41	Udon Thani	71,152.00	10216.4	9188.3	15800.6
42	Loei	31,806.80	11510.6	1030.7	5404.4
43	Nong Khai	32,504.90	8519.5	3150.7	6737.2
44	Maha Sarakham	33,983.40	6569.3	3852.3	7401.3
45	Roi Et	47,933.00	8543.8	6477.5	11608.3
46	Kalasin	38,367.50	9462.1	5458.3	7980.3
47	Sakon Nakhon	38,293.00	8170.6	3027.9	9253.7
48	Nakhon Phanom	22,370.50	6349	640.8	4139.5
49	Mukdahan	12,969.70	2835.4	1506.7	2046
50	Chiang Mai	118,020.40	16784.3	11823.1	18596.2
51	Lamphun	65,181.90	5015.1	45001.9	4059.3
52	Lampang	45,614.60	5524.1	5392.4	7935.9
53	Uttaradit	26,900.30	7791.2	3576.5	4114.6
54	Phrae	21,883.50	3685.9	2133	4146.4
55	Nan	20,746.40	6228.7	1176.8	2736.1
56	Phayao	23,298.20	6795.5	1315.8	3872.6
57	Chiang Rai	54,306.10	14869.8	4136.9	9503.2
58	Mae Hong Son	9,430.90	2815.7	326.4	1215.5
60	Nakhon Sawan	73,533.30	19566.9	15320.6	11373.6
61	Uthai Thani	19,237.30	6620.9	2443.2	3105.6
62	Kamphaeng Phet	72,644.20	12777.1	23035.3	7100
63	Tak	35,075.40	9351.4	5546.3	4056.4
64	Sukhothai	29,695.70	9322.1	2129.8	5269.4

103

Jurisdiction Code	Province	Gross province product (Million bath)			
		Total	Agriculture	Production	Retail and wholesale
65	Phitsanulok	54,768.60	13811.2	4862.1	9234.2
66	Phichit	30,620.10	9739.9	2740.1	5992.3
67	Phetchabun	58,443.00	25625.7	4614	8654.9
70	Ratchaburi	102,900.70	15050.8	28811	10568.5
71	Kanchanaburi	69,263.80	14053	15791.2	13296.4
72	Suphan Buri	57,996.70	16899.2	7792.9	11951.1
73	Nakhon Pathom	126,139.70	10637.8	66894.3	11492.6
74	Samut Sakhon	315,473.10	1282.9	267514.2	9956
75	Samut Songkhram	15,398.20	988	3576.5	3326.8
76	Phetchaburi	51,028.00	10458.9	10081	6207.7
77	Prachuap Khiri Khan	53,784.90	11772.2	8386.4	7248.4
80	Nakhon Si Thammarat	122,763.50	25253.5	14690.4	16048.4
81	Krabi	43,957.60	20756.6	4381	3612
82	Phangnga	29,558.50	13026	1783.9	3268.4
83	Phuket	61,904.50	2418.2	2747.5	6406.2
84	Surat Thani	120,749.20	40433.4	19972	11311.3
85	Ranong	16,594.30	3141.1	1016.9	1990.2
86	Chumphon	45,390.30	18194	3949.5	5634.3
90	Songkhla	162,071.50	31225.7	45319.4	16701.5
91	Satun	26,851.10	8066.2	3054.1	2513.8
92	Trang	61,924.20	24657.3	8475	7665.5
93	Phatthalung	32,936.90	12436	2552.2	5970
94	Pattani	37,749.30	6224.3	2733.2	4204
95	Yala	38,537.00	18251.7	2578.3	3835.3
96	Narathiwat	45,623.40	25697.4	2173.7	4146.8

104

Table A3 Others data

Jurisdiction Code	Province	Area	Crop Area	Rice crop	Fruit crop	Number of manufacture
10	Bangkok	978263	120095	97063	17404	18889
11	Samut Prakan	627558	191501	43720	17881	7088
12	Nonthaburi	388939	164836	100081	59468	2035
13	Pathum Thani	953660	461885	328227	109684	2695
14	Phra Nakhon Si Ayudthaya	1597900	1069259	961702	52947	1700
15	Ang Thong	605232	468459	357716	66506	465
16	Lop Buri	3874846	2213113	873709	110366	652
17	Sing Buri	514049	420611	368967	13370	314
18	Chai Nat	1543591	1184328	901497	59770	371
19	Saraburi	2235304	936234	395172	118005	1391
20	Chonburi	2726875	1339579	133530	626966	3378
21	Rayong	2220000	1289252	31832	928043	2129
22	Chanthaburi	3961250	1634903	76228	1096517	670
23	Trat	1761875	532640	64901	402462	385
24	Chachoengsao	3344375	1806218	863527	229283	1595
25	Prachin Buri	2976476	1168327	757600	144476	876
26	Nakhon Nayok	1326250	627598	502270	94674	300
27	Sa Keaw	4496962	2099436	861207	186861	429
30	Nakhon Ratchasima	12808728	7718934	3817384	356679	7288
31	Buri Ram	6451178	3933547	3136344	217016	1536
32	Surin	5077535	3547551	2953732	200444	1112
33	Si Sa Ket	5524985	3441625	2586593	374778	1897
34	Ubon Ratchathani	9840526	4759667	3433109	292522	4014
35	Yasothon	2601040	1605388	1254469	116291	922

105

Jurisdiction Code	Province	Area	Crop Area	Rice crop	Fruit crop	Number of manufacture
36	Chaiyaphum	7986429	3408150	1514182	204704	1645
37	Amnat Chareon	1975785	1342076	1024109	94149	302
39	Nong Bue Lumphu	2411929	1485567	980345	106597	1205
40	Khon kaen	6803744	4147800	2720562	207414	4784
41	Udon Thani	7331439	3689539	2173151	245349	4002
42	Loei	7140382	2295431	410045	651782	1019
43	Nong Khai	4582675	2637309	1292051	772273	1175
44	Maha Sarakham	3307302	2716051	2013796	127574	2709
45	Roi Et	5187156	3194532	2749920	63431	3250
46	Kalasin	4341716	2614708	1554887	150946	1995
47	Sakon Nakhon	6003602	2796829	2013509	192653	2289
48	Nakhon Phanom	3445418	1481894	1079813	160336	348
49	Mukdahan	2712394	919090	408641	157117	428
50	Chiang Mai	12566911	1346371	586826	541631	2303
51	Lamphun	2816176	556512	149151	321483	915
52	Lampang	7833726	1015232	417192	298885	1485
53	Uttaradit	4899120	1255225	610057	285897	355
54	Phrae	4086624	609374	271488	146771	1157
55	Nan	7170045	701150	207420	142222	422
56	Phayao	3959412	1038055	576134	192978	435
57	Chiang Rai	7298981	2004971	1189407	327987	1789
58	Mae Hong Son	7925787	251959	114567	70719	140
60	Nakhon Sawan	5998548	3904651	2248789	173679	1535
61	Uthai Thani	4206404	1338983	538269	103777	346
62	Kamphaeng Phet	5379681	2467932	1115970	105477	556
63	Tak	10254156	991695	286778	114436	553
64	Sukhothai	4122557	1877120	821027	268091	982

Jurisdiction Code	Province	Area	Crop Area	Rice crop	Fruit crop	Number of manufacture
65	Phitsanulok	6759909	2404936	1452434	184651	1282
66	Phichit	2831883	1980228	1531574	125978	740
67	Phetchabun	7917760	3711569	1361562	369289	913
70	Ratchaburi	3247789	1107982	316894	214734	1413
71	Kanchanaburi	12176968	2051830	421193	309284	1412
72	Suphan Buri	3348755	2068951	1157329	139379	1107
73	Nakhon Pathom	1355204	716681	349596	146207	2969
74	Samut Sakhon	545217	155470	28751	83208	4709
75	Samut Songkhram	260442	114792	3251	72002	289
76	Phetchaburi	3890711	657243	338984	167651	663
77	Prachuap Khiri Khan	3979762	1229662	61560	718806	610
80	Nakhon Si Thammarat	6214064	3025699	851534	1740461	1665
81	Krabi	2942820	1279632	58596	1184271	466
82	Phangnga	2606809	836366	11642	803757	364
83	Phuket	339396	134902	1367	127021	418
84	Surat Thani	8057168	2721645	160133	2330072	978
85	Ranong	2061278	467676	24258	414031	307
86	Chumphon	3755630	1970915	130029	1707595	723
90	Songkhla	4621181	2131072	407956	1541413	2054
91	Satun	1549361	594280	86721	453055	267
92	Trang	3073449	1505276	68118	1375689	692
93	Phatthalung	2140296	1225040	397199	704784	834
94	Pattani	1212722	761770	185217	514356	885
95	Yala	2825674	1198218	66932	1087154	330
96	Narathiwat	2797144	1415690	109628	1227965	459

107

Table A4 CFS goodness of fit of distribution

Order	Distribution	K-S
1	Exponential	0.34730
2	Gamma	0.34415
3	Log-Logistic	0.51402
4	Lognormal	0.29246
5	Power Function	0.58384
6	Weibull	0.25957

Table A5 RS1 goodness of fit of distribution

Order	Distribution	K-S
1	Exponential	0.22524
2	Gamma	0.18803
3	Log-Logistic	0.24904
4	Lognormal	0.16346
5	Power Function	0.52347
6	Weibull	0.20349

Table A6 RS2 goodness of fit of distribution

Order	Distribution	K-S
1	Exponential	0.22982
2	Gamma	0.27919
3	Log-Logistic	0.68168
4	Lognormal	0.21921
5	Power Function	0.36788
6	Weibull	0.27927

Biography

The author grew up in Sakonnakon province, a small province in the upper northeastern part of Thailand. He graduated from the first-ranked high school in Sakonnakon province, then went to Khon Kean province to study at Khon Kean University. After four years of studying, the author graduated with a Bachelor's degree in Civil Engineering from Khon Kean University. He worked as civil engineer in a construction company and consultant firm for a few years. Then he studied for his Master's degree in Civil Engineering at Chulalongkorn University. After that, he spent several years working as a transportation engineer at a consulting firm, but the hard life of Bangkok made him miss his home town. Fortunately, Kasetsart University set up a campus in Sakonnakon province. The author returned to his home town and began working as lecturer at Kasetsart University Sakonnakon Chalermpakiat Campus. He earned an AUN-SEEDNET scholarship for a doctoral degree and he studied at Chulalongkorn University. Upon earning his doctoral degree in Civil Engineering (major transportation engineering), he again returned to his home town, where he currently works as a lecturer at Faculty of Science and Engineering Kasetsart University Sakonnakon Chalermpakiat Campus.