GROUP ASSIGNMENT AND ANNUAL AVERAGE DAILY TRAFFIC ESTIMATION OF SHORT-TERM TRAFFIC COUNTS USING GAUSSIAN MIXTURE MODELING AND NEURAL NETWORK MODELS

by

SUNIL KUMAR MADANU

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

August 2016

In Loving Memory of

Smt. Mary Jecintha

Dedicated to:

My parents Smt. Mary Jecintha and Sri. Arogyam

Acknowledgements

I am taking a moment to thank many people helped me in completing this work. First, I am sincerely thankful to my supervising professor Dr. Stephen Mattingly who mentored me through this journey. In the past few years, his care, friendship, treatment like a family member, support, guidance, critique and encouragement has given me motivation and strength to advance my degree progressively and successfully to reach my goals. I want to thank Dr. Siamak Ardekani, Dr. James Williams, and Dr. Mahmut Yaşar for their guidance, being committee members and providing valuable suggestions.

I am very grateful for the support of my Ph.D. colleagues. I am thankful for the Oregon DOT staff for their suggestions and data support.

My love and gratitude are towards my wife Jyothsna and little one Shawn. Her constant support and encouragement make this journey possible. I enjoyed playing and read a book to my little friend who makes my life moving. They supported me immensely throughout the study.

I want to convey my respect to my parents Smt. Mary Jecintha and Sri. Arogyam for their care and love. I remember invaluable blessings from my mom and dad who strive hard for getting an education for me and my sisters and brother. My heartfelt thanks to my sisters: Joshi, Anitha, Asha and my brother Sudheer. They are praying, wishing and patiently waiting for my success. I specially appreciate blessings from Smt. Mary Josephena, Sri. Francis Showraiah and Sandeep.

September 7, 2016

Abstract

GROUP ASSIGNMENT AND ANNUAL AVERAGE DAILY TRAFFIC ESTIMATION OF SHORT-TERM TRAFFIC COUNTS USING GAUSSIAN MIXTURE MODELING AND NEURAL NETWORK MODELS

Sunil Madanu, PhD

The University of Texas at Arlington, 2016

Supervising Professor: Stephen P. Mattingly

The grouping of similar traffic patterns and cluster assignment process represent the most critical steps in AADT estimation from short-term traffic counts. Incorrect grouping and assignment often become a significant source of AADT estimation errors. For instance, grouping a commuter traffic trend pattern into a recreational traffic trend may produce an erroneous AADT value. The traditional knowledge-based methods, often aided with visual interpretation, introduce subjective bias while grouping traffic patterns. In addition, the grouping requires personnel resources to process large amounts of data and remains inefficient with unapparent traffic patterns. The functional class grouping, a traditional method, also produces larger errors. Under limited resources and constraints, better methods and techniques may group sites with similar characteristics.

The study uses Gaussian Mixture Modeling (GMM) for clustering and an enhanced neural network model (OWO-Newton or ONN) for classification of continuous count data. The researchers compare this modified approach with volume factor grouping and a traditional approach. The study uses Automatic Traffic Recorder (ATR) data from the Oregon Department of Transportation (ODOT) as a comparative case study. Overall, the proposed two-step approach, GMM-ONN, exhibits improved performance. The study observes an error difference of 6% to 27%, which is statistically significant at 5 percent level, between the GMM-ONN and other methods. The GMM-ONN method produces less than five percent error for urban interstates and less than ten percent for urban arterials and freeways. The study method meets the FHWA recommended AADT forecasting error of less than ten percent for commuter patterns. The GMM-ONN also produces less error when compared to studies based on the national average and Minnesota and Florida DOT count data. The lower AADT estimation errors and its distribution show an effective and reliable approach for AADT estimation using short-term traffic counts. Moreover, the lower standard deviation of errors shows the satisfactory accuracy of the AADT estimates. The study recommends the improved two-step process due to its accuracy, economical approach by using daily patterns, and ability to meet the agency's need for a low-cost traffic counting program. The GMM-ONN method not only minimizes judgment errors but also supplements the FHWA guidelines on recommending clustering techniques for grouping the traffic patterns.

Table of Contents

List of Illustrations

List of Tables

List of Acronyms and Abbreviations

Abbreviation	Description
AADT	Annual Average Daily Traffic
AASHTO	American Association of State Highway and Transportation Officials
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
ARI	Adjusted Rand Index
ATR	Automatic Traffic Recorders
AUC	Area Under Curve
BIC	Bayesian Information Criterion
BP	Back Propagation
CG	Conjugate Gradient
CH	Calinski and Harabasz Index or Criterion
CI	Confidence Interval
CV	Coefficient of Variation or Cross-Validation
DAPE	Daily variation of Absolute Percent Error
DHV	Design Hourly Volume
DOT	Departments of Transportation
DT	Daily Traffic volume
EM	Expected-Maximization
FC	Functional Class
FHWA	Federal Highway Administration
FN	False Negatives
FNR	False Negative Rate
FP	False Positives
FPR	False Positive Rate
GMM	Gaussian Mixture Modelling
GWR	Geographically Weighted Regression
HCA	Hierarchical Clustering Analysis
KM	K-means
k-NNC	k-Nearest Neighbor Classifier
LDA	Linear Discriminant Analysis
LL	Log-likelihood

MADT	Monthly Average Daily Traffic
MAP	Maximum a Posteriori
MAPE	Mean Absolute Percent Error
MLP	Multiple Layer Perceptron
MSE	Mean Squared Error
NB	Naïve Bayes
NN	Neural Network
OAA	One-Against-All
OAO	One-Against-One
ODOT	Oregon Department of Transportation
OLF	Optimal Learning Factor
ONN	OWO-Newton Neural Network
OWO	Output Weights Optimization
PPV	Positive Prediction Value
PR	Precision-Recall
QDA	Quadratic Discriminant Analysis
ROC	Receiver Operating Characteristic
SAF	Seasonal Adjustment Factors
STTC	Short Term Traffic Counts
SVM	Support Vector Machines
TMAS	Travel Monitoring Analysis System
TMG	Traffic Monitoring Guide
TN	True Negatives
TNR	True Negative Rate
TP	True Positives
TPR	True Positive Rate
VMT	Vehicle Miles Traveled
WCV	Weighted Coefficient of Variation

Chapter 1 Introduction and Methodology

Transportation agencies need reliable estimates of traffic volumes for planning, designing, operating and maintaining highway infrastructure. The American Association of State Highway and Transportation Officials (AASHTO) *Guidelines for Traffic Data Programs* identifies key areas of traffic data use in safety analysis, air quality, capacity analysis, pavement design, operational analysis and project evaluation and selection (*1*). Agencies use the traffic data on Annual Average Daily Traffic (AADT), Vehicle Miles Traveled (VMT) and Design Hourly Volume (DHV) in most of their projects (*2*).

INTRODUCTION

According to the Federal Highway Administration (FHWA) *Traffic Monitoring Guide* (TMG), monitoring traffic volume trends represents a key task for continuous traffic count program (*4*). The deployment of Automatic Traffic Recorders (ATR) on a state-wide network helps state DOTs (Department of Transportation) to collect and monitor traffic patterns.

The agencies allocate significant resources to collect the traffic data on their networks using Automatic Traffic Recorders (ATRs). The ATRs collect the traffic data at a section of highway continuously for 365 days of the year. The agencies study temporal variation of traffic, like month-of-year, day-of-week or hour-of-day patterns, using the ATR counts and, they use these temporal patterns to convert short-term counts to AADT values.

The cost and maintainability restrict ATR deployment to limited strategic locations on the highway network. State agencies face a tough decision on how many ATRs to deploy, where to deploy, and how frequently to collect the data with limited resources. In addition, decisions on the type of traffic patterns to monitor, either by vehicle type, monthly, the day-of-week, or hourly distribution complicates the monitoring process.

The FHWA and state agencies have developed some guidelines for the Traffic Monitoring Analysis System (TMAS) to evaluate volume trends over a specified time period. Moreover, monitoring the AADT and its trends requires continuous data from only a limited number of locations. In order to cover specific locations of interest, Short Term Traffic Counts (STTC) are taken and seasonally adjusted using factor groups (*3*). Factor groups have reasonably homogeneous patterns, usually but not necessarily, calculated based on monthly traffic patterns. In addition, the TMG also suggests the use of day-of-week, hourly patterns, and patterns by vehicle type (passenger or trucks) or geographic region when clustering. Many state DOTs group traffic patterns from ATR sites and calculate AADT values using seasonal factors (*3*).

In lieu of ATRs, the agencies commit to Short-term Traffic Counts (STTCs). The STTCs can be cost effective and deployed almost everywhere on the network to comprehensively study the traffic data. Typically, the short-term counts are collected on a road segment every few years and the collection periods vary from 1 to 7 days (*2*). Qi *et al.* (*4*) report that the Montana DOT ATR's total setup cost of $8,700 for two lanes and $15,700 for four-lane sections of highways. During 2012, the annual short-term counts operating costs account for $100 per counter and ATRs cost about $ 5,581 to operate annually in Montana DOT (*4*).

Even though, the traditional method is the most widely used method among many Departments of Transportation (DOTs), the assignment and grouping based on only the functional class affects the accuracy of the AADT estimates. Many studies are proposed to address the issue of accuracy from short-term counts and obtain improved accuracy over the traditional method.

Purpose

While previous studies have developed AADT estimation methods based on regression analysis, statistical analysis, neural networks and other machine learning techniques, these techniques often remain inadequate to meet the AADT estimation needs of engineering design and planning. Even though non-regression methods may provide improved estimates, the increase in the application of new paradigms from machine learning in pattern recognition and classification provides an opportunity to apply a new methodology for clustering and assigning traffic patterns. However, very little effort is focused on assessing relative strengths of different methods. In addition, the area of robustness of the proposed methods is often neglected in previous studies. Despite challenges, the researchers in the transportation field, in general, are looking for a way to improve the AADT estimates from short-term counts. Still, the process of collecting short-term counts and adjusting them to obtain the AADT values appears valuable to many DOTs due to a reduction in traffic count costs. The study tries to enhance the AADT accuracy using improvements from both grouping and classification stage. The accuracy of the AADT estimates obtained from the STTCs becomes important for correctly planning a traffic monitoring program.

Objectives

The study considers following objectives to address the needs:

- Provide an improved clustering process using Gaussian Mixture Modelling (GMM)
- Assess the Robustness of the clustering solutions using different scenarios
- Perform comparative analysis of GMM clustering with the K-means (KM) and agglomerative Hierarchical Clustering Analysis (HCA) solutions

- Introduce a modified Back Propagation learning algorithm (OWO-BP) and a new training algorithm OWO-Newton method (ONN) for classification of traffic pattern groups
- Conduct a thorough evaluation of classifiers using different performance measures
- Perform comparative analysis with the Oregon DOT seasonal trend grouping and traditional functional class grouping
- Conduct statistical testing between the methods for difference in estimation errors

STUDY METHODOLOGY

The FHWA methodology has four stages. The first step called *grouping or clustering*, combines traffic patterns based on the similarity in their 24-hour traffic patterns. The second stage, called *factoring*, computes the Seasonal Adjustment Factors (SAFs). Next, the process must *assign* the short-term counts to one of the groups. The final step estimates the AADT using the short-term counts and corresponding seasonal adjustment factors.

This research focuses on developing an innovative strategy for the first and third stages in the FHWA methodology. The author introduces the new approaches in the next few subsections. This study uses data sets from the ATR counters throughout the State of Oregon for the years 2011 and 2012. For the purpose of analysis, a Short Term Traffic Count (STTC) is defined as a complete 24-hour count on a given day. At each ATR station, the investigation samples the STTCs and uses them for clustering and classification analysis. The 2011 data is used only for the clustering and classifier design. The author tests the proposed method and its relative performance against other strategies using the 2012 data.

First Stage: Clustering

The first stage applies the innovative clustering techniques to the Oregon Department of Transportation (ODOT) ATR data. The 2011 data has 30,393 STTCs and each STTC has a continuous 24-hour traffic volume (veh/hour) data. After developing and presenting the novel GMM clustering of traffic count data, the research presents the k-means and hierarchical clustering analysis for a comparative analysis. After finding the clustering solutions, the study investigates the robustness of the solutions using bootstrapping, replacing points by noise, jittering and subsetting resampling schemes for cluster stability analysis. The comparative analysis includes the robustness assessment as well as a comparison of the AADT estimation errors from the GMM KM and HCA clustering methods, with ODOT's seasonal trend grouping method and functional class grouping method. Chapter 2 presents a detailed analysis of the innovative clustering methodology and robustness inquiry.

Second Stage: Seasonal Adjustment Factors

After the clustering step, each cluster group must be sub-grouped by month to compute the average Seasonal Adjustment Factors (SAFs). The seasonal factors are computed by averaging the ratios of annual average daily traffic (AADT) to daily traffic (DT) in a given month. Subgrouping helps to address the monthly and seasonal variation of traffic data. Chapter 2 also presents the SAFs obtained from the GMM clustering.

Third Stage: Classification

The cluster assignment process represents the most critical step in the AADT estimation process. The study proposes an optimal neural network structure and two variants of learning algorithms to improve the estimation error. In the classification step, the study presents an inventive approach for building a better classifier for assigning the traffic patterns. The study adopts Multiple Layer Perceptron (MLP) neural networks for the assignment step and proposes changes to the network structure, learning process and learning algorithms. Each algorithm is trained on a neural network with one input layer, one hidden layer, and one output layer. The network selection process outputs the required number of units in the hidden layer. The investigation considers different training algorithms and obtains their training and testing mean square errors (MSEs).

The study adopts a 10×10 stratified cross-validation approach, which performs ten runs of 10-fold stratified cross-validation on the 2011 data set. The author also develops Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Naïve Bayes (NB) classifier, and ONN classifiers for another comparative analysis. Chapter 3 provides details on classification step and presents the evaluation of different classifiers in estimating the AADT from short-term counts.

Final Stage: Two-Step Process and AADT Estimation

The study uses the 2012 ATR data set with a sample of 32,289 data patterns for testing. The 2012 dataset has hourly traffic data showing time of day variation and the groupings according to the ODOT seasonal trend grouping method and highway functional class. The classifiers assign the group number for the test data according to the GMM, KM, and the HCA solution. The AADT is calculated using the SAFs and the sum of 24-hour traffic volume (daily traffic or DT). The computed AADT value is compared with the actual AADT value to obtain Mean Absolute Percent Error (MPAE). In addition, the study computes and reports the standard deviation of the errors. Chapter 4 provides details on the AADT accuracy among different methods and presents the practical implications of the innovative methodology.

STUDY ORGANIZATION

The remainder of the report is arranged as follows. Chapter 2 presents the clustering analysis and partitions robustness. Next chapter deals with the innovative assignment methodology for test patterns; this analysis includes neural networks, discriminant analysis, and Naïve Bayes classification methods. Chapter 4 combines the clustering and classification methods to improve the AADT estimation accuracy and presents a detailed analysis of errors by month, day-of-week, and highway functional class. The last chapter provides concluding remarks and study limitations with directions for future study.

REFERENCES

1. AASHTO Guidelines for Traffic Data Programs. 2009. American Association of State Highway and Transportation Officials (AASHTO), 444 North Capitol Street, NW Washington, DC 20001 USA.
2. Zhong, M., Bagheri, E., and Christie, J. (2012). Improving Group Assignment and AADT Estimation Accuracy of Short-term Traffic Counts using Historical Seasonal Patterns & Bayesian Statistics. 8th International Conference on Traffic and Transportation Studies, Changsha, China, August 1–3, 2012.
3. Traffic Monitoring Guide. Federal Highway Administration (FHWA). U.S. Department of Transportation, 2013. http://www.fhwa.dot.gov/policyinformation/tmguide. Accessed July 7, 2015.
4. Qi, Y., D. McCarthy, J. Stephens, S. Forsythe, and D. Veneziano. (2013). Montana Weigh-in-Motion (WIM) and Automatic Traffic Recorder (ATR) Strategy. Report for Task II: MDT Traffic Data Collection Program Description/Inventory. https://www.mdt.mt.gov/other/webdata/external/research/docs/research_proj/wim/task2_may14.pdf , Accessed on September 09, 2016.

Chapter 2 Clustering

INTRODUCTION

The grouping and classification of traffic patterns play an important role when estimating Annual Average Daily Traffic (AADT) values from Short Term Traffic Counts (STTCs). Incorrect grouping and classification often become a significant source of AADT estimation errors. According to Gadda *et al*. (*1*): “the errors are estimated to be on the order of 20% or even higher”. For instance, grouping a commuter traffic trend pattern into a recreational traffic trend may produce an erroneous AADT value.

The state Departments of Transportation (DOTs) allocate significant resources to collecting historic traffic data throughout the state-wide networks. The estimation of AADT has many practical applications in planning, operations, maintenance, and decision making. The literature proposes a wide range of AADT estimation techniques like regression analysis, geographically weighted regression, artificial neural networks, time-series analysis, genetic algorithms, and kriging-based methods (see references (*2*) and (*3*) for review).

According to the Federal Highway Administration (FHWA) Traffic Monitoring Guide (TMG), monitoring traffic volume trends represents a key task for a continuous traffic count program (*4*). The deployment of Automatic Traffic Recorders (ATR) on the state-wide network helps state DOTs to identify and monitor traffic patterns. State agencies must decide how many ATRs to deploy, where to deploy, and how frequently to collect the data with limited resources. In addition, decisions on the type of traffic patterns to monitor, either by vehicle type, monthly, the day-of-week, or hourly distribution complicates the monitoring process. The FHWA and state agencies have developed some guidelines for the Traffic Monitoring Analysis System (TMAS) to evaluate volume trends over a specified time period. Moreover, monitoring the AADT and its trends requires continuous data from only at a limited number of locations. In order to cover a specific location of interest, Short Term Traffic Counts (STTC) may be taken and seasonally adjusted using factor groups (*4*).

Background

Factor groups have reasonably homogeneous patterns, usually but not necessarily, calculated based on monthly traffic patterns. In addition, the TMG suggests the use of day-of-week, hourly patterns, and patterns by vehicle type (passenger or trucks) or geographic region when clustering. Many state DOTs group traffic patterns from ATR sites and calculate AADT values using seasonal factors (*4*).

The FHWA and state agencies have developed some guidelines for Traffic Monitoring Analysis System (TMAS) to evaluate volume trends over a specified time period. The FHWA procedure consists of four steps (*4*):

1. Grouping ATR sites with similar traffic volume variations
2. Determining average seasonal adjustment factors for each road group
3. Assigning the road section, monitored with a STTC, to one of the groups defined in step 1
4. Applying the appropriate seasonal adjustment factor to the STTC to produce the AADT estimate for the road section in question

The application of the FHWA procedure may be affected by three sources of error (*3*):

- Error due to day-to-day variations in traffic volumes
- Error in grouping road segments (ATR sites) into significant road groups
- Error in assigning the road segment associated with a STTC to the right road group

The traffic volumes over the given state-wide network fluctuate over time due to a variety of factors. Any kind of estimation in the transportation field must deal with this common problem. The TMG outlines three types of analysis for grouping: Traditional Approach, Cluster Analysis, and Volume Factor Grouping. The traditional approach uses general knowledge of the road system with visual interpretation to identify groups. Cluster analysis is a procedure to group the patterns, often, using monthly factors (ratio of AADT to MADT-Monthly Average Daily Traffic) at continuous count stations. Volume factor grouping maintains separate volume factor groups by highway functional category. Finding groups through knowledge and functional class seem neither practical nor likely to produce better results because of a large amount of continuous count data, and dynamic changes in travel activity patterns (irrespective of highway functional class). In addition, bias due to subjectivity, difficulty in analyzing these large datasets, and significant time resource requirements elevate the problems of the conventional methods (*4*). A few alternative methods, generally labeled as clustering techniques, have evolved for the automatic grouping of traffic patterns.

This chapter aims to provide an improved clustering process using Gaussian Mixture Modelling (GMM). In addition, the GMM solution is compared with the k-means clustering, hierarchical cluster analysis, traditional approach and volume factor grouping to assess its relative performance.

After the background information, this chapter outlines Gaussian mixture modeling and the cluster selection process. The following section describes the classification of test data according to the cluster solutions. Finally, a comparative case study of clustering solutions for estimating AADT values is presented.

CLUSTERING METHODS

Clustering techniques try to determine the structure of data when no information is available except observational data. Cluster analysis partitions the data into meaningful subgroups without knowing its components and structure. Cluster analysis is broadly divided into heuristic methods and statistical models, which follow either hierarchical or relocation strategies (*5*).

Hierarchical Clustering

Hierarchical clustering does not partition the data into a specific number of clusters. Instead, one form of the method starts from as many clusters as the size of the dataset to a single cluster containing all data. Hierarchical clustering may be broadly divided into agglomerative and divisive methods. Agglomerative methods successively fuse individual data points into groups until reaching a single cluster containing all data points. Divisive methods start with all data points in a single cluster and separate them into groups in steps. Both methods use some kind of proximity measure to either divide or separate data points in clusters. The clustering solutions produced by any of these methods are represented using a two-dimensional diagram called as a *dendrogram*. The dendrogram shows either grouping or separation at each stage of the analysis. Figure 2-1 shows a dendrogram and two broad methods of hierarchical clustering.

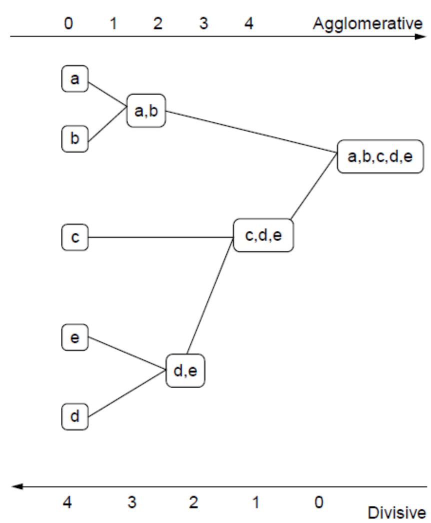


Figure 2-1 Dendrogram and Hierarchical Clustering Methods (Source: Everitt *et al.*(*6*))

Agglomerative methods represent a widely referenced method in hierarchical clustering (*6*). They subdivide the data into groups, where the first group consists of assigning a single cluster to *n* number of data points (equal to the size of the data set) and the last one consists of a single cluster containing all *n* data points. At each stage of fusion, the data points are grouped based on the similarity measure, for instance, single linkage, average linkage, centroid linkage or Ward's method.

Knowledge of the closeness of data points to each other remains essential when identifying the clusters in the given data set; the dissimilarity (similarity) distance or proximity measures typically measure closeness. Two types, namely distance measures and correlation type measure, exist for measuring dissimilarity among multivariate continuous data sets. Euclidean, City block, Minkowski, Canberra distance measures represent examples of distance-based measures. Pearson correlation and angular separation measures belong to correlation type measures.

Single linkage, also known as the nearest neighbor technique, groups the closest pair of data points where only one data point from each group is considered. The *Complete linkage* (or furthest neighbor) method looks for the farthest distanced pair. *Average linkage* considers the average distance between all pairs of individuals from each group. However, the centroid and median linkage methods use a data matrix rather a proximity matrix for fusing the clusters (*6*). The *centroid method* involves merging clusters with similar mean vectors. The *median linkage* method is similar to the centroid method except that centroids of the constituent clusters are considered while merging. The *weighted average linkage* method introduces weighting the average inter-cluster distances according to the inverse of the number of data points in each group or class. *Ward's method* is based on the size of an error sum-of-squares criterion. Each stage seeks to minimize the increase in the total within-cluster error sum of squares (*6*).

Agglomerative methods produce different cluster solutions for any given data under different distance (or proximity) measures and linkage methods. In general, agglomerative methods have the following limitations (*6*):

- Tends to produce unbalanced especially in large data sets
- Does not take account of cluster structure
- Tends to find compact clusters with equal diameters
- Tends to join clusters with small variances
- Subject to reversals
- Sensitive to outliers

In addition, choosing a correct number of clusters and plotting cluster solution (or dendrogram) for large data sets becomes critical. In the case of ties (due to a similar linkage function value), the decision on which cluster group to assign the data pattern may affect the clustering solution output (*6*). Due to its wide applicability to a variety of situations, the study uses agglomerative Hierarchical Clustering Analysis (HCA) with Ward's similarity measure for clustering.

K-means clustering

Relocation based methods assign the observations iteratively among the groups. The number of clusters or groups has to be specified in advance, and they do not change during the course of iterations. However, at each iteration, observations move from one group to other groups, usually, using some form of distance criteria (the study uses Euclidean distance). K-means (KM) clustering (and finite mixture modeling) represent an example of relocation based methods. K-means clustering selects the number of clusters that minimize within-group variance (*5*).

Finite Mixture Models

Mixture modeling involves probability based cluster analysis, where observational data is assumed to come from a mixture of probability distributions. The Bayesian Information Criterion (BIC) is used to determine the number of components and model form for clustering. Often, mixture components are modeled to follow a Gaussian distribution where its maximum likelihood parameters are found using an Expected-Maximization (EM) algorithm (*5*).

GMM Clustering

Model-based clustering, an alternative approach to clustering, assumes that any given dataset consists of a number of sub-populations or clusters. The variables in each cluster will have different multivariate probability density functions that constitute a finite mixture density for a population. The clustering problem becomes estimating the parameters of these mixtures and calculating the posterior probabilities for assessing the cluster membership. Model-based clustering provides a meaningful statistical model for the clustering process (*6*).

Given observations (or patterns) of hourly traffic volumes, $\mathbf{x} = \{\mathbf{x}_1,...,\mathbf{x}_n\}$, let $f_k(\mathbf{x}_i \mid \boldsymbol{\theta}_k)$ be the density of an hourly pattern $\mathbf{x}_i$ (with d=24 dimensions) from the kth component, where $\boldsymbol{\theta}_k$ are the corresponding parameters, and let G be the number of components in the mixture. The probability density function can be expressed as:

$$f(\boldsymbol{x};\ \boldsymbol{\theta}, \boldsymbol{\tau}) = \prod_{i=1}^{n} \sum_{k=1}^{G} \tau_k\ g_k(\boldsymbol{x}_i | \boldsymbol{\theta}_k) \tag{1}$$

Where f is a maximum likelihood parameter and τ_k is the probability (mixing proportion) that an hourly pattern belongs to the kth cluster or component. The mixing proportions are non-negative and are such that $\tau_k > 0$; $\sum_{k=1}^{G} \tau_k = 1$, and parameters $\boldsymbol{\theta}' = \left(\boldsymbol{\theta}'_1, \cdots, \boldsymbol{\theta}^{\mathbf{1}}_G\right)$ and mixing proportions $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_G)$.

Finite mixtures in model-based clustering, generally, assume that each group in a given data set comes from a different probability distribution. However, they may also come from the same family of distribution functions with different parameter values. The present study assumes that each cluster or sub-population comes from a Gaussian or Normal distribution with different parametric values. The authors choose Gaussian mixture modeling (GMM) due to its considerable success in a number of other applications (*5*, *6*).

In the GMM, each cluster follows a normal distribution with parameters $\boldsymbol{\theta}_k$ that consists of a mean vector $\boldsymbol{\mu}_k$ and a covariance matrix Σ_k ,and a density function of the form:

$$g_k(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \Sigma_k) = \frac{e^{\left\{-\frac{1}{2}(\boldsymbol{x}_i-\boldsymbol{\mu}_k)^T \Sigma_k^{-1} (\boldsymbol{x}_i-\boldsymbol{\mu}_k)\right\}}}{(2\pi)^{\frac{d}{2}} |\Sigma_k|^{\frac{1}{2}}} \quad (2)$$

In the GMM, a given cluster k is centered at mean μ_k, and covariance matrix Σ_k determines the cluster geometric characteristics (shape, orientation and volume). Banfield and Raftery (*7*) developed a model-based framework for clustering by parameterizing the covariance matrix. The covariance matrix of a cluster in terms of its eigenvalue decomposition is given by:

$$\Sigma_k = \lambda_k D_k \Lambda_k {D_k}^T \quad (3)$$

Where Dk is the orthogonal matrix of eigenvectors, Λ_k is a diagonal matrix whose elements are proportional to the eigenvalues of Σ_k and λk is a scalar. The Dk matrix determines the orientation of the principal components of Σ_k, while Λ_k determines the shape of the density contours, and λ_k specifies the volume of the corresponding ellipsoid. GMM yields different clustering models by restricting the cluster orientation, shape and volume (*8*). These models are broadly categorized into three main model families: spherical, diagonal, and general. A three-letter code describes the volume, shape, and orientation of the cluster groups. Each letter in the code may take a value of E for equal, V for variable, or I for identity (8). For instance, model VEV stands for a cluster model with variable volume, equal shape, and variable orientation.

For a given *n* observations, the log-likelihood function (ℓ) is defined as:

$$\ell(\boldsymbol{\tau}, \boldsymbol{\theta}) = \sum_{i=1}^{n} \ln f(\boldsymbol{x};\ \boldsymbol{\theta}, \boldsymbol{\tau}) \quad (4)$$

Parameters are estimated using solutions obtained from likelihood equations of

$$\frac{\partial \ell(\phi)}{\partial \phi} = 0 \quad (5)$$

Where $\phi' = (\theta', \tau')$. The nature of a likelihood function complicates the use of regular methods to obtain the solution. Hence, the Expectation-Maximization (EM) iterative approach is widely used to obtain parameters in the GMM.

The group or class, $\boldsymbol{z} = (\boldsymbol{z}_1, \cdots, \boldsymbol{z}_n)$, to which a data point belongs is unknown or to be determined. The group of a data point $\mathbf{x}_i$ is expressed as $\boldsymbol{z_i} = (z_{i1}, \cdots, z_{ik})$, where z_{ik} takes a value of one if $\mathbf{x}_i$ belong to group *k*. The density of an observation $\mathbf{x}_i$ is given by $\prod_{k=1}^{G}[\, g_k(\boldsymbol{x}_i|\boldsymbol{\theta}_k)]^{z_{ik}}$ and the resulting log-likelihood function is

$$\ell(\theta_k, \tau_k, z_{ik}|\boldsymbol{x}) = \sum_{i=1}^{n}\sum_{k=1}^{G} z_{ik}\left[log\left(\tau_k \cdot\ g_k(\boldsymbol{x}_i|\boldsymbol{\theta}_k)\right)\right] \qquad (6)$$

The conditional expectation of z_{ik} given the observation $\mathbf{x}_i$ and parameter values are given by $\widehat{z_{ik}} = E[z|\boldsymbol{x}_i, \boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_G]$. Once the parameters are estimated, the observations can be assigned to a particular cluster *j* based on the maximum value of the estimated posterior probabilities given by $\{\, j\, |z_{ij}{}^* = max_k\, \widehat{z_{ik}}\}$.

The following steps are involved in finding the parameter of the GMM using the EM algorithm (*9*)

1. Initialize $\widehat{z_{ik}}$ (using a classification scheme)
2. Repeat M-step and E-step
   a. M-Step: $n_k \leftarrow \sum_{i=1}^{n} \widehat{z_{ik}}$; $\widehat{\tau_k} \leftarrow \frac{n_k}{n}$; $\widehat{\mu_k} \leftarrow \frac{\sum_{i=1}^{n} x_i \cdot \widehat{z_{ik}}}{n_k}$; $\widehat{\Sigma_k}$ (use Eq. 3)
   b. E-step: using parameters from M-step, compute $\widehat{z_{ik}} = \frac{\widehat{\tau_k} \cdot g_k(x_i|\widehat{\mu_k}, \widehat{\Sigma_k})}{\sum_{j=1}^{G} \widehat{\tau_j} \cdot g_j(x_i|\widehat{\mu_j}, \widehat{\Sigma_j})}$
3. Verify convergence criteria, if not satisfied, repeat step 2

Model selection

The selection of mixture models usually uses Bayesian Information Criteria (BIC). In model-based clustering, a decisive first local maximum of BIC indicates strong evidence for a good model (*9*). The larger the value of the BIC, stronger the evidence for the model. Usually, the number of mixture components is taken as the number of clusters (*9*). The BIC to choose the best model given G groups is 2 log (*L*) + *m* log (*n*), where *L* is the likelihood function and *m* is the number of free parameters to be estimated (*6*).

Fraley and Raftery outlined the following steps in model-based strategy for clustering (*9*):

- Determine a maximum number of clusters to consider (G)

- Perform initialization, usually, using agglomerative hierarchical clustering and get initial parameters of mixture models
- Perform EM algorithm and find each parameterization for number of clusters starting from 2 to G under different models
- Compute the BIC for all clusters and models trails and develop a matrix of BIC values
- Plot the BIC values for each model
- A decisive first local maximum indicates strong evidence for a model (parameterization and number of clusters)

Selecting of Number of Clusters

Three clustering solutions can potentially (and generally do) produce a different number of clusters. Often, the parameters associated with these clustering patterns are different and affect the solution according to their fine tuning. Moreover, the objectives of clustering change by the clustering method. Despite the difference, the quality of a given clustering method needs to be tested. In a general sense, cluster validation refers to evaluating the quality of a clustering. Cluster validation without knowledge of *true clusters* is a bit more challenging. Henning *et al.* (*10*) present a variety of indices for quantifying the clustering's quality. These indices were used, mostly, to find the optimal number of clusters for which an index value is either maximum or minimum. Measuring the quality of clustering represents an ongoing field of research with new measures being proposed regularly. The problem with cluster validation using indices is that different indices yield a different optimal number of clusters. This makes the cluster selection process more challenging for a given problem (*10*).

Milligan and Cooper (*11*) conducted a comparative study of clustering indices' performance. In this study, both the Calinski and Harabasz (*12*) and Duda and Hart (*13*) criterion are the top two performers under multiple scenarios. Hence, the study adopts the Calinski and Harabasz criterion for finding an optimum number of clusters for the HCA and KM clustering solutions. The GMM clustering uses the BIC criterion for selecting a model and the number of clusters. The best number of clusters for a given clustering solution is based on the clusters corresponding to the maximum criterion values.

Calinski and Harabasz Index or Criterion

Assuming a traffic data matrix **X** of $n \times p$ size with n number of patterns and p dimensional (p=24) hourly traffic data. Let the $\boldsymbol{d}$ matrix represent the dissimilarity matrix between hourly traffic data with $p \times p$ dimension. For a given clustering solution having K number of clusters, $\mathcal{C}_K = \{C_1, \cdots, C_K\}$ represents an exhaustive partition of the clusters that belong to the input traffic patterns. Let $c(i)=j$ denote that a traffic pattern $\boldsymbol{x}_i$ belongs to the cluster j . Let $\overline{\boldsymbol{x}}_j$ be the mean vector of cluster C_j where $j = 1,\ldots, K$, and n_j be the number of patterns that belong to the cluster j or called as the cluster size. Let $\overline{\boldsymbol{x}}$ be the overall mean the traffic patterns. The CH criterion is based on the cluster variance, and hence it is called a variance-ratio test (*10*). The with-in cluster variation of the cluster solution ($\boldsymbol{W}_{\mathcal{C}_K}$) can be written as

$$\boldsymbol{W}_{\mathcal{C}_K} = \sum_{j=1}^{K} \sum_{c(i)=j} \left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_j\right) \left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_j\right)^T \tag{7}$$

and the between-cluster variation is written as

$$\boldsymbol{B}_{\mathcal{C}_K} = \sum_{j=1}^{K} n_j \left(\overline{\boldsymbol{x}}_j - \overline{\boldsymbol{x}}\right) \left(\overline{\boldsymbol{x}}_j - \overline{\boldsymbol{x}}\right)^T \tag{8}$$

The Calinski and Harabasz criterion (or variance ratio criterion) is defined as

$$CH(\mathcal{C}_K) = \frac{trace(\boldsymbol{B}_{\mathcal{C}_K})}{trace(\boldsymbol{W}_{\mathcal{C}_K})} \times \frac{n-K}{K-1} \tag{9}$$

PREVIOUS WORK

Past research has categorized ATR data using a variety of techniques: agglomerative hierarchical grouping (*14,15*), k-means clustering (*16,17*), model-based clustering (*18*), fuzzy C-means method (*19-21*), regression models (*22*), Bayesian statistics (*23*), mixture of regression models (*24*), neural networks (*25*), genetic algorithms (*26*), quantum-frequency algorithm for automated identification of traffic patterns (*27*), fuzzy logic (*28*), Support Vector Machines (SVM) (*29*), and wavelets (*30*).

PURPOSE AND OBJECTIVES

Most of the previous studies focus more on the assignment (classification) of traffic patterns to a given traffic group. In addition, clustering methods present a variety of drawbacks, which include inconsistent clustering results for different locations and time periods, sensitivity to missing data, inability to provide semantic meaning (e.g. summer, commuter, recreational, etc...), theoretical nature of methods, and ill-suitability for wider implementation (*27*). The groups in the cluster analysis, unlike groupings based on expert judgment, avoid biases because they are chosen by their data-driven similarity measure (*4*). However, cluster analysis lacks guidelines on establishing the optimal number of clusters for a given data set. A *priori* information on the number of clusters, cluster initializations, and cluster evaluation criteria play an important role in cluster analysis outputs. Unfortunately, no standard procedure exists for selecting *a priori* information. Cluster analysis output groups, often, cannot be adequately identifiable on a given state-wide network due to their pure mathematical nature (*4*). In addition, day to day traffic variation, missing or bad data from ATR malfunction and incorrect assignment of STTCs to seasonal factor groups make significant contributions to AADT estimation errors (*3*). Cluster analysis may compute a different number of clusters across years and ATR sites may change cluster year by year (*31*). The cluster solutions reported in the literature often lack statistical inference. These drawbacks highlight the challenges associated with cluster analysis and the difficulty of identifying clusters for practical situations. In addition, the partial success of cluster analysis depends on how well the clusters behave for robustness analysis. For instance, missing data and noise conditions affect the clustering solution. This paper seeks to develop a comprehensive methodology that addresses all of these challenges and shortcomings of current clustering strategies.

Need

Regression models often seem inadequate to meet the AADT estimation needs of engineering design and planning. Reducing traffic count costs using limited traffic counts for site-specific studies (at particular periods of the year) appears valuable for state agencies. In addition, the accuracy of AADT estimates obtained from STTCs becomes important for correctly planning a traffic monitoring program. Despite the previous criticism, clustering analysis with the evolution of machine learning algorithms provides an opportunity to improve AADT estimation outputs. This means that the FHWA procedure and its improvements provide important guidance to transportation agencies for estimating the AADT of road networks. The lack of robustness analysis in the cluster analysis from previous efforts provides an opportunity to test the clustering solution under different analysis scenarios. The study considers following objectives to address these needs:

Objectives

This paper seeks to develop an improved clustering process using Gaussian Mixture Modelling. This procedure will include:

- Assessing the stability of the clustering solution using difference resampling schemes
- Performing comparative analysis of GMM clustering with the KM and HCA solutions

CONTRIBUTION

The study addresses key questions on performance, stability and variability of clustering when grouping the traffic patterns. The study proposes the GMM framework that provides a statistical inference for the obtained clusters. The framework explores the reproducibility and variability of cluster parameters like cluster proportions, means, and variance. The study makes an effort to formally and qualitatively label the clustering groups using traffic pattern characteristics. The study introduces cluster-wise stability assessment for the clustering solutions using four different resampling methods. In addition, the study performed missing value analysis and variable size analysis for clustering solutions. A thorough evaluation of clustering performance across months, days of week and highway functional class is performed. The study also conducts a comparative analysis of AADT estimation errors by error size and cluster.

STUDY METHODOLOGY

The study methodology has five stages. The first stage applies the clustering techniques to Oregon Department of Transportation (ODOT) ATR data. After finding the clustering solution, the study conducts a robustness analysis of the partitions to assess their relative stability. The study develops seasonal adjustment factors by cluster and month for all solutions in the third stage. The next stage designs a data classifier using the group labels assigned during the first stage. In the final stage, the classifier helps in assigning a group number to the test data. Furthermore, the study assesses the relative merits of the clustering solutions by comparing estimated and actual AADT values. In particular, the paper compares the GMM solution with the KM method, HCA method, ODOT's seasonal trend grouping method, and functional class grouping method. The following sections present these steps after the case study data description.

CASE STUDY DATA

ATR counters throughout the State of Oregon for the years 2011 and 2012 (see Figure 2-2) provide the data sets. This study defines a Short Term Traffic Count (STTC) as a complete 24-hour count on a given day. At each ATR station, STTCs are sampled and used for clustering analysis. The Oregon DOT has five regions covering different parts of the state-wide highway network. Region 1 is home to the largest and heavily traveled Portland metropolitan area. Region 2 covers the capital city - Salem, a few popular coastal destinations, and university based cities like Corvallis and Eugene. Region 3 covers the Southwest portion of the state and includes Medford. Both Region 4 and 5 have sparsely populated counties and spread to the East of the state (and East of the Cascade mountain ranges). Figure 2-3 shows the regions of the Oregon DOT.

Table 2-1 provides a summary of the ATR data characteristics. Regions 1 and 2 share more ATRs and have higher travel activity. Figure 2-2 shows the distribution of the ATRs within the state of Oregon. The researchers try to generalize the grouping methodology by considering all possible variations in the traffic trends. Thus, the analysis does not subset the data by vehicle type, weekdays, weekends, and seasonal patterns when establishing groupings. The 2011 data is used only for clustering and designing a classifier. The clustering solution and its relative performance are tested with the 2012 data. The next section presents the exploratory analysis of the datasets.

Table 2-1 Study Data Characteristics

Year	ATRs per Region					Total ATRs	AADT	Number of STTCs per year
	1	2	3	4	5			
2011	20	27	12	9	22	90	22,277	30,393
2012	20	28	12	9	22	91	22,814	32,289

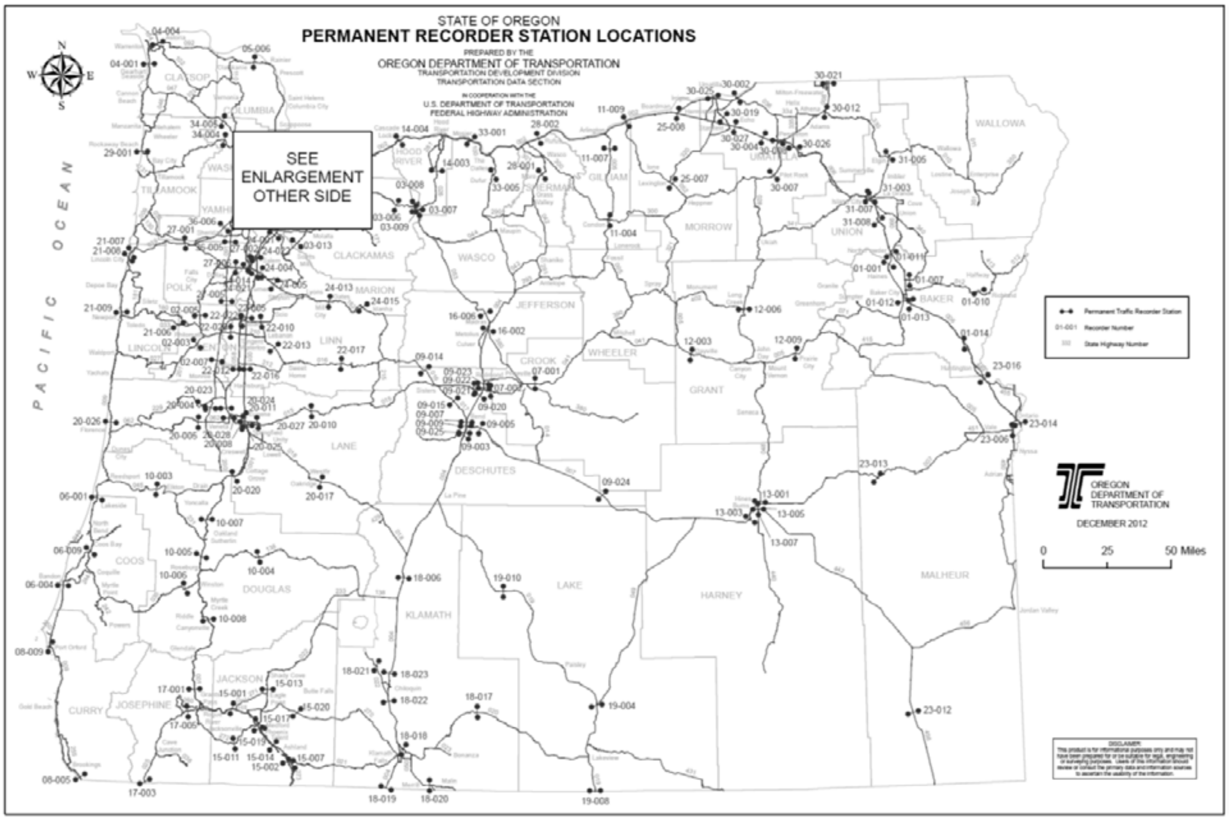


Figure 2-2 Permanent Recorder Station Locations for the State of Oregon (year 2012, source: ODOT)

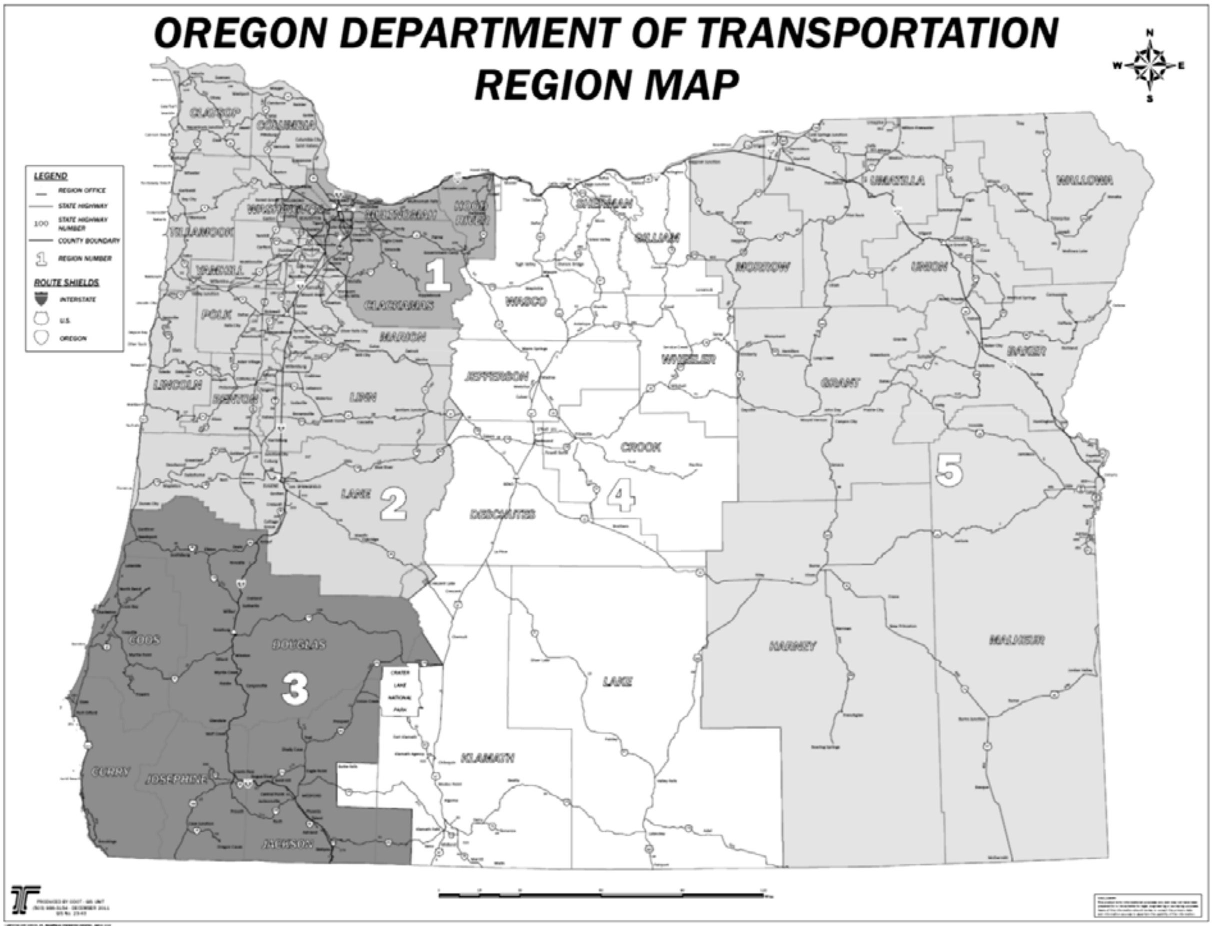


Figure 2-3 ODOT Region Map (source: Oregon DOT)

AADT by Functional Class and Region

The datasets have traffic patterns that belong to seven highway functional classes: rural/urban interstates, other freeway and expressways, other principal arterials, other principal arterials-urban, minor arterial and major collectors. The ODOT follows the federal guidelines of *Highway Functional Classification: Concepts, Criteria and Procedures* (*32*) for classifying the highways. Figure 2-4 shows the mean AADT values among highway functional classes for the years 2011 and 2012. Year 2012 mean traffic values appear slightly lower than the 2011 traffic except for urban interstates and freeways. More traffic growth occurs for freeways/expressways. Region two shows traffic growth and rest of them show decreased traffic from 2011 to 2012 (see Figure 2-5). The quartile distribution of traffic for each functional class is shown in Table 2-2 and Table 2-3.

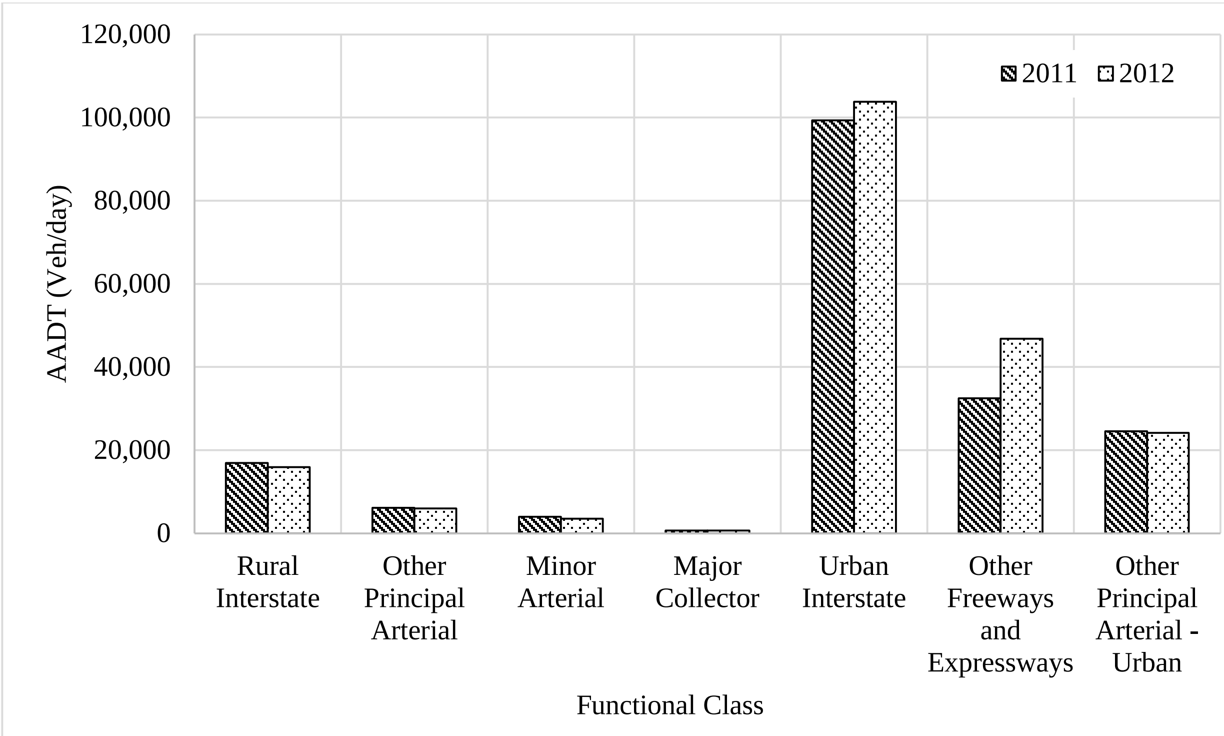


Figure 2-4 AADT by Highway Functional Class

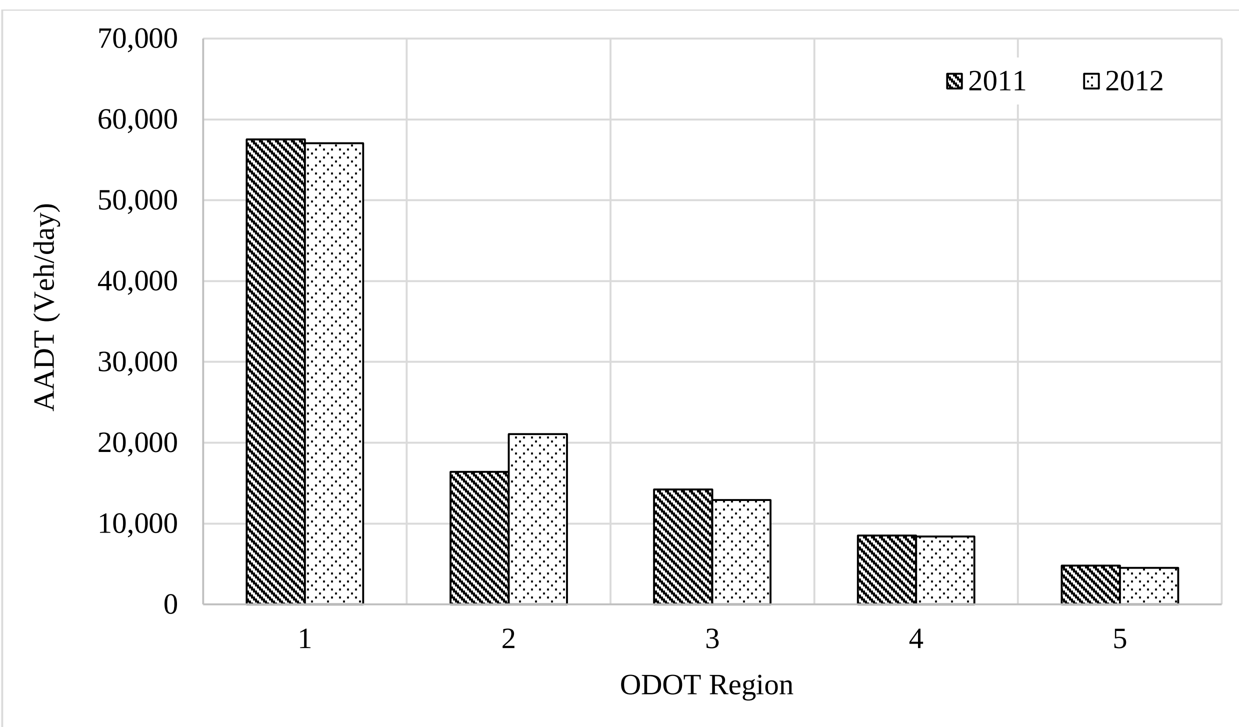


Figure 2-5 AADT by ODOT Region

Table 2-2 Quartile Distribution of AADT by Highway Functional Class

Functional Class	Year	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
Rural Interstate	2011	8,322	9,339	14,240	16,890	20,870	32,840
	2012	7,596	9,337	14,250	15,910	20,750	32,000
Other Principal Arterial	2011	343	1,152	4,682	6,104	9,998	27,220
	2012	346	1,146	4,670	6,015	9,852	26,630
Minor Arterial	2011	976	1,077	2,902	3,981	5,553	12,980
	2012	887	1,312	2,274	3,519	5,079	12,710
Major Collector	2011	193	204	557	697	841	1,742
	2012	191	209	556	656	784	1,711
Urban Interstate	2011	16,400	57,440	123,900	99,310	145,400	155,500
	2012	16,520	57,100	124,300	103,800	150,300	154,400
Other Freeways and Expressways	2011	20,600	25,700	30,160	32,470	35,150	50,640
	2012	20,430	25,590	35,350	46,820	84,200	84,200
Other Principal Arterial - Urban	2011	8,673	18,190	21,720	24,560	32,240	42,480
	2012	8,423	18,190	21,170	24,180	32,800	41,600

Table 2-3 Quartile Distribution of AADT by ODOT Region

Region	Year	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
1	2011	768	5,269	32,240	57,520	123,900	155,500
	2012	703	4,837	29,940	57,050	124,300	154,400
2	2011	1,028	3,280	7,736	16,390	25,700	69,540
	2012	1,030	3,981	9,369	21,080	26,630	84,200
3	2011	988	2,902	8,673	14,230	19,050	42,480
	2012	887	2,873	8,423	12,910	18,740	41,600
4	2011	193	953	4,021	8,506	18,190	21,720
	2012	191	790	3,888	8,389	18,190	21,170
5	2011	204	805	1,152	4,796	9,339	16,400
	2012	209	789	1,312	4,522	8,335	16,520

Figure 2-6 shows the mean hourly traffic for both years covering all traffic patterns with 95% confidence intervals. The hourly distribution shows two peaks with a pronounced evening peak. The traffic increases steadily from morning peak to the evening peak. The hourly distribution for the year 2012 is slightly higher than the year 2011. The 95 percent confidence intervals for both years show a stable trend in the hourly traffic.

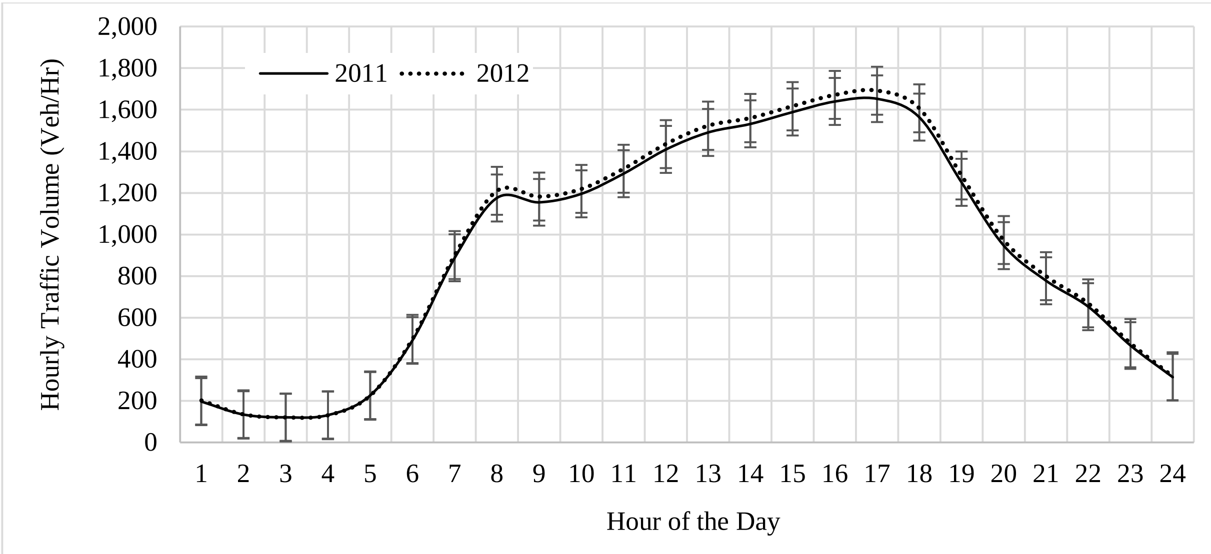


Figure 2-6 Mean Hourly Distribution of Traffic Patterns

FIRST STAGE: CLUSTERING

The clustering analysis uses the 2011 data with a sample size of 30,393. Each sample has continuous 24-hour traffic volume (veh/hour) data, daily traffic volume (DT in veh/day), and the ratio of AADT to DT. At the beginning of the first stage, GMM clustering is fit to the data and refined to obtain a final cluster solution. Then, the researchers present the k-means and hierarchical clustering analysis, and other clustering solutions for comparative analysis. The study uses the *mclust* package available in the R programming language to analyze clustering. The following sections present these steps.

GMM

Sample hourly traffic data (with dimensionality d = 24) represents the input data for the GMM clustering. The study evaluates different models with clusters between 2 and 30 to select a best one. The model-based clustering chooses 15 mixture components using BIC criterion. An unconstrained model "VVV" (variable shape, volume, and orientation of covariance matrix) is chosen (see Figure 2-7). The mix proportions of the fifteen components are not uniform; five components have a proportion between 2 and 4 percent, three components between 4 and 6 percent, four components between 8 and 10 percent, and the proportion varies for the other three components. The clustering solution is not perfectly balanced but shows fewer variations in the produced clusters. The Coefficient of Variation (CV) (ratio of standard deviation to the mean) of all clusters is less than one, which indicates a low variance among the clusters. Table 2-4 summarizes the class proportions and related AADT statistics.

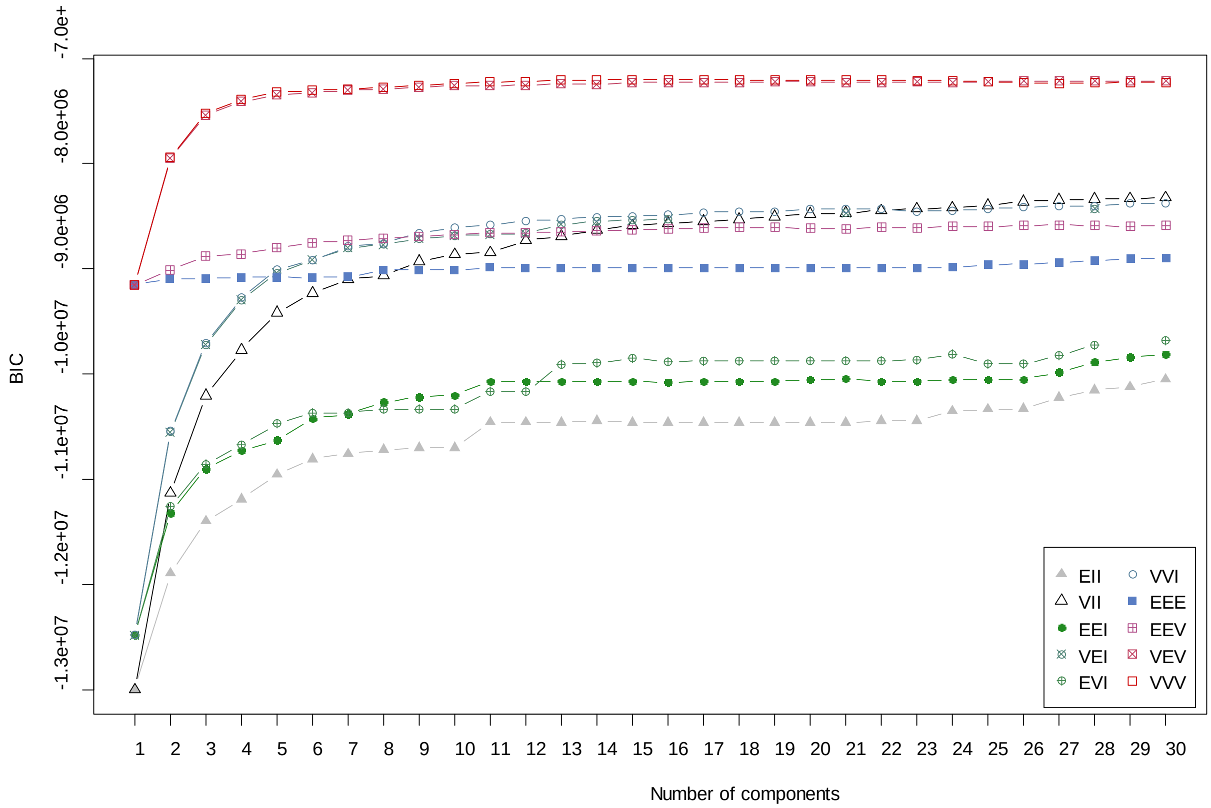


Figure 2-7 BIC Criterion by GMM model Type

Table 2-4 GMM Cluster Proportions and Related AADT statistics

GMM Cluster	Number of Patterns	Proportion of Patterns (%)	AADT					
			Min.	Max.	Mean	Standard Deviation	Standard Error	Coefficient of Variation
1	3616	11.9	204	4,666	1,352	607	10	0.45
2	1564	5.1	193	8,322	2,422	1,295	33	0.53
3	1113	3.7	443	8,748	3,330	1,334	40	0.40
4	1358	4.5	1,172	8,748	4,454	970	26	0.22
5	2646	8.7	443	13,517	5,743	1,665	32	0.29
6	3970	13.1	193	4,519	656	323	5	0.49
7	2071	6.8	1,172	69,540	28,380	12,438	273	0.44
8	1611	5.3	1,172	13,517	9,747	1,821	45	0.19
9	2841	9.3	768	27,224	11,415	4,362	82	0.38
10	2739	9.0	5,553	21,721	14,556	3,138	60	0.22
11	2879	9.5	7,736	35,153	27,673	6,342	118	0.23
12	1165	3.8	101,560	155,531	134,898	19,757	579	0.15
13	738	2.4	27,224	155,531	132,591	26,391	971	0.20
14	936	3.1	4,666	155,531	104,542	47,651	1,558	0.46
15	1146	3.8	42,485	69,540	54,189	8,902	263	0.16

Figure 2-8 shows the monthly (or seasonal variation) variation of traffic among clusters. A general trend of summer peaks occurs among the clusters. However, some patterns produce almost flat trends (only mild summer peaks). Cluster 2 and 6 show a steep increase and decrease in traffic levels between months. However, cluster 6 has a flat peak during months from July to September compared to a single peak month (July) for cluster 2.

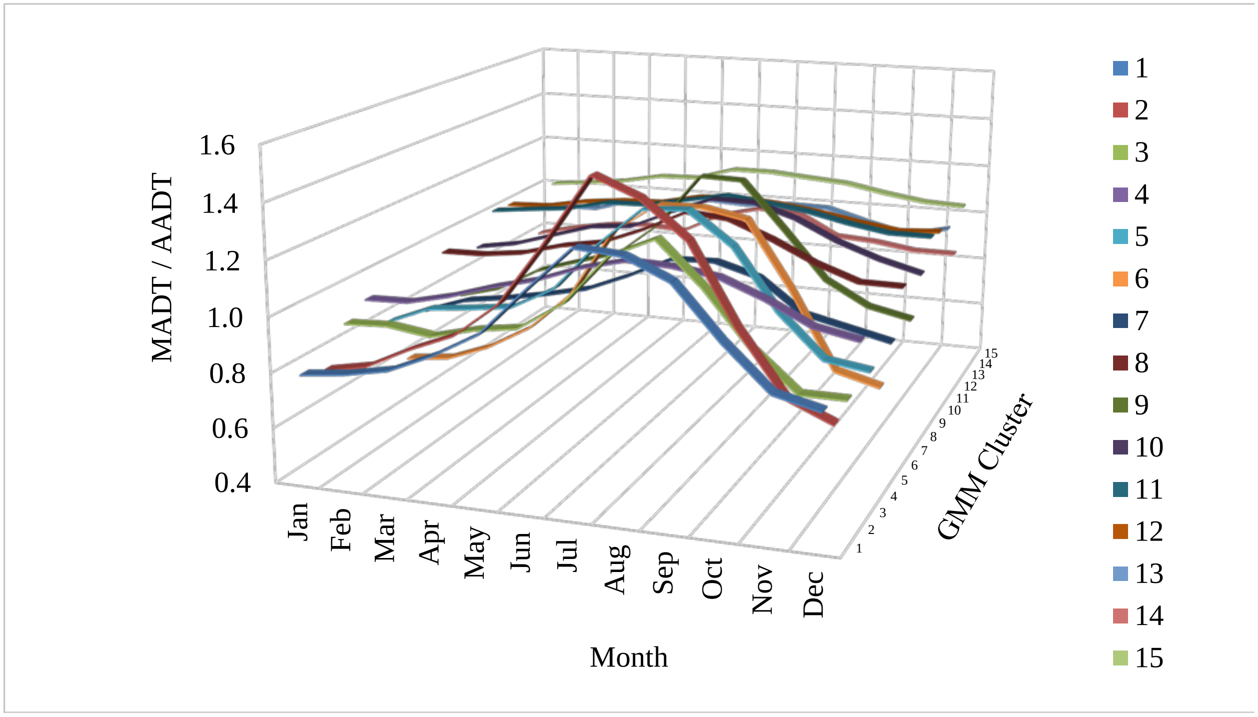


Figure 2-8 Monthly Variation of Traffic Patterns for Each GMM Cluster

The day-of-week patterns show that most of the clusters have stable weekday (Monday to Thursday) patterns and reduced traffic for the rest of the week (see Figure 2-9). Clusters 8 and 11 show steady weekend traffic growth. Some clusters exhibit uneven variations among the days in a week. Clusters 12 and 15 do not have any weekend patterns. In general, the study observes more traffic on Friday compared to other days among the clusters.

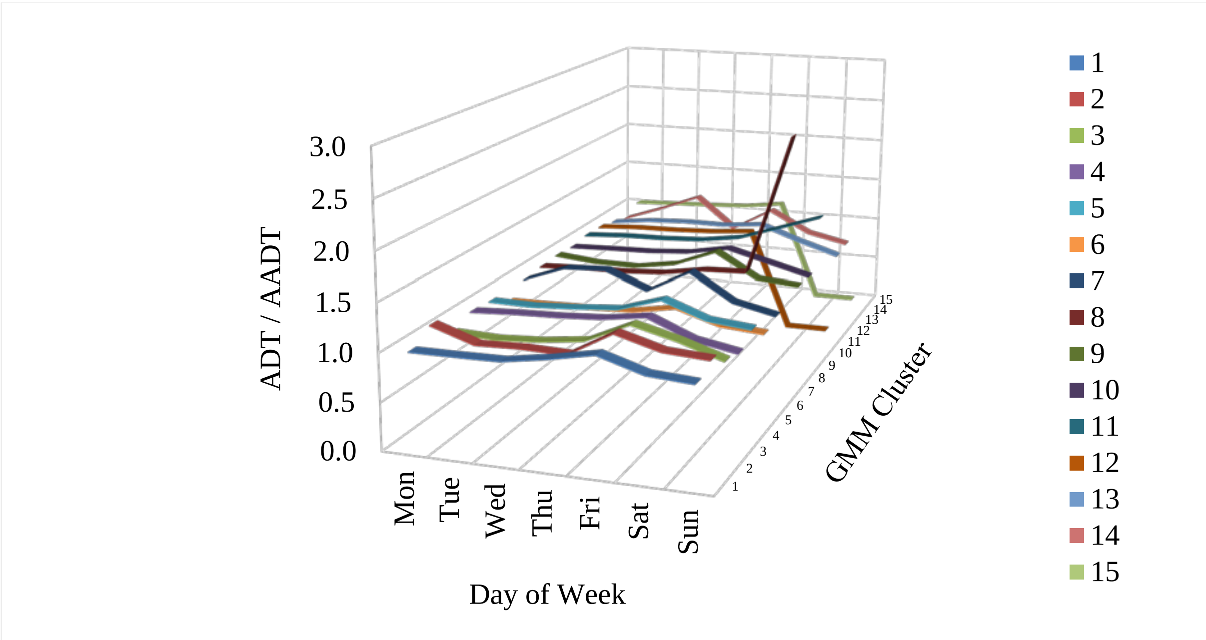


Figure 2-9 Day-of-Week Variation of Traffic Patterns for Each GMM Cluster

Clustering methods output clusters and assign (or label) a class number to each pattern. The clusters would not provide a qualitative description of the pattern types. Moreover, the objective of clustering methods is to explore meaningful patterns that exist in the dataset. Hence, traditionally clustering methods are not given any qualitative description for the produced clusters. However, this study tries to describe clustering using traffic pattern characteristics like functional class, area type, and either weekday or weekend type (see Table 2-5). The study uses the hourly distribution of traffic in each cluster (Figure 2-10), proportion of pattern by highway functional class and area type (Table 2-5), proportion of pattern by the ODOT region (Figure 2-11), monthly patterns (Figure 2-8), and day of week patterns (Figure 2-9) when describing the cluster patterns.

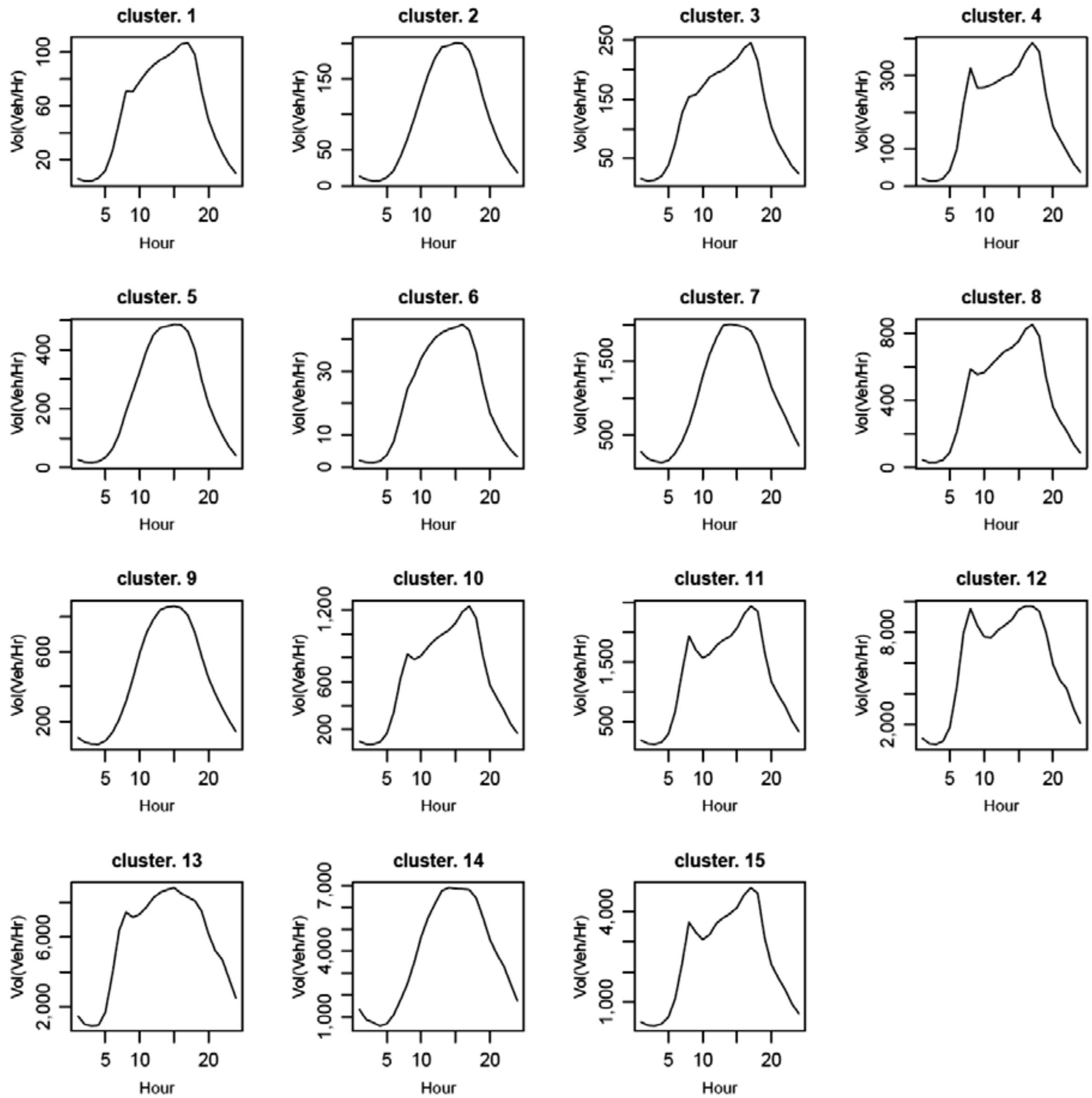


Figure 2-10 Mean Hourly Distribution of Traffic for the GMM Clustering Solution

Cluster 12 exhibits two pronounced peaks during morning and evening peak times. However, the two peaks carry almost the same traffic. All of the patterns in this cluster belong to only urban interstates, with relatively stable AADT values from a minimum of 101,560 vehicles per day to a maximum of 155,531 per day. Hourly traffic patterns in cluster 12 constitute roughly four percent of total traffic patterns and occur in ODOT region 1 (Portland metropolitan area). Hence, cluster 12 can be labeled as an *urban interstate commuter* type.

Table 2-5 GMM Clustering Trend Names and Patterns by Functional Class and Area Type

GMM Cluster	Number of Patterns (Proportion of Patterns (%))	Proposed Trend Name	Proportion of Patterns (%) by Functional Classification: Rural Interstate	Other Principal Arterial	Minor Arterial	Major Collector	Urban Interstate	Other Freeways and Expressways	Other Principal Arterial - Urban	Area Type: Rural	Urban
1	3616 (11.9)	summer	0	40	46	14	0	0	0	100	0
2	1564 (5.1)	summer	0	57	30	13	0	0	0	100	0
3	1113 (3.7)	recreational summer	0	37	62	1	0	0	0	100	0
4	1358 (4.5)	rural commuter	0	47	53	0	0	0	0	100	0
5	2646 (8.7)	coastal destination	0	84	15	0	0	0	1	98	2
6	3970 (13.1)	recreational summer	0	44	13	43	0	0	0	100	0
7	2071 (6.8)	commuter	21	13	3	0	13	23	28	32	68
8	1611 (5.3)	summer	0	69	16	0	0	0	15	60	40
9	2841 (9.3)	non-urbanized interstate	40	33	12	0	3	1	11	70	30
10	2739 (9)	commuter	22	27	18	0	9	0	23	42	58
11	2879 (9.5)	commuter	10	12	0	0	7	34	36	21	79
12	1165 (3.8)	urban interstate commuter	0	0	0	0	100	0	0	0	100
13	738 (2.4)	urban interstate	0	0	0	0	99	1	0	0	100
14	936 (3.1)	urban interstate	2	3	1	0	83	6	4	6	94
15	1146 (3.8)	urban commuter	0	0	0	0	59	21	20	0	100

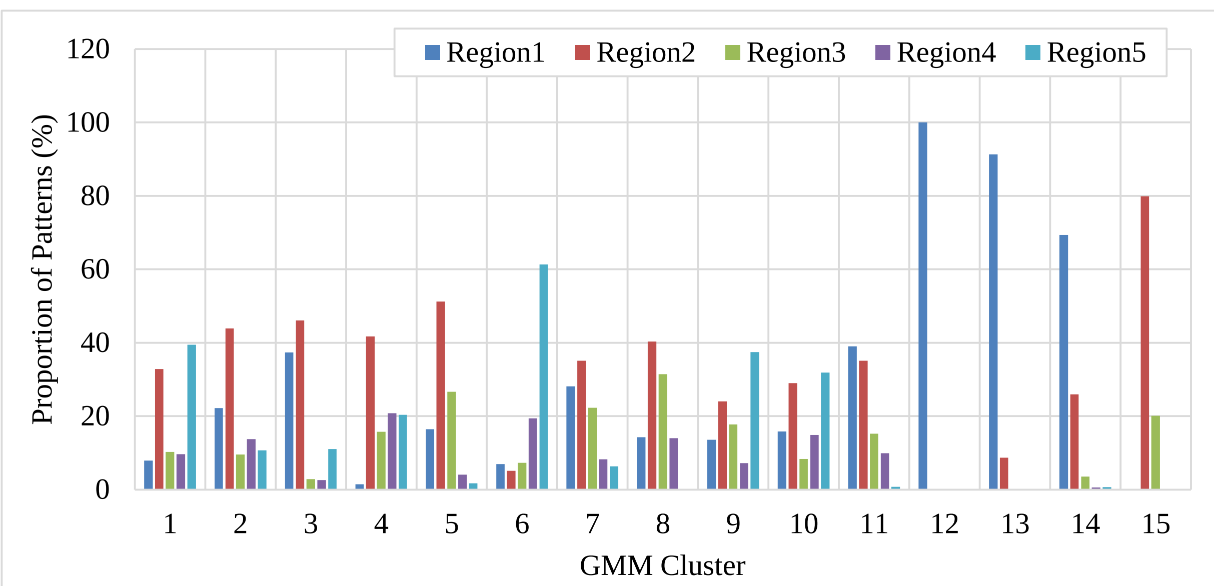


Figure 2-11 GMM Clusters by ODOT Region

Clusters 11, 15 and 4 also exhibit two peaks with a more prominent evening than morning peak. More than 50 percent of the cluster 15 patterns belong to urban interstates and the rest of the pattern remain almost equally distributed between freeways/expressways and urban principal arterials. Clearly, the patterns show commuting patterns with more weekday traffic originating in the urban areas of ODOT regions 2 and 3. Clusters 11 and 4 also exhibit commuting patterns with mean AADT values of 27,673 vehicles per day and 4,454 vehicles per day. Cluster 4 patterns come from the functional class of minor arterials, whereas most of the cluster 11 patterns belong to urban principal arterials/expressways. In addition, cluster 4 patterns carry traffic in rural areas. Cluster 15 can be labeled as *urban commuter* type, cluster 4 as *rural commuter* type, and cluster 11 as *commuter* type.

Clusters 3, 8, and 10 show predominant evening peaks along with smaller morning peaks. However, the traffic grows steadily from morning to evening peak in these clusters. Comparing the average AADT values among them, cluster 3 carries less traffic followed by cluster 8 and cluster 10. Cluster 3 patterns emerge mostly from arterial highways in a rural area. Cluster 3 patterns also exhibit higher weekend traffic with the peak traffic in warmer months (April to October) and high CV value. Hence, cluster 3 can be labeled as *recreational summer* pattern type. Cluster 8 carries patterns that belong to principal arterial highways and they cover both rural and urban area types. Cluster 8 also carries high weekend traffic, especially during warmer months. Cluster 10 exhibits similar monthly patterns as cluster 8; however, the weekend traffic in cluster 10 remains less than cluster 8. The patterns cover mostly arterials and interstates in both urban and rural areas. The patterns also come from all five regions of the ODOT. Cluster 10 patterns can fall under *commuter* type and cluster 8 can be labeled as *summer* pattern type.

Cluster 1 has a predominant evening peak and covers mostly rural area patterns. The average traffic volume is higher for Thursdays and Fridays. The patterns show typical warmer month's peak and patterns belong to traffic on principal, minor arterials, and major collectors. Most patterns come from Region 5 and they can be labeled as *summer* patterns type.

Clusters 2, 5, 7 and 9 have only evening peak traffic. However, these clusters have different average AADT values with relatively higher variability. The patterns constitute a third of total patterns. Clusters 2 and 5 cover a pattern from arterial streets mostly from rural or rural populated areas of regions 1, 2 and 3. Both clusters carry higher traffic during warmer months. However, cluster 2 has peak traffic in the month of July. Cluster 5 has less peak traffic compared to cluster 2. Cluster 2 and 5 have more weekend patterns than weekday patterns. As most of the patterns belong to region 2, which has more coastal destinations, cluster 2 can be labeled as *recreational summer* type patterns and cluster 5 as *coastal destination* type patterns.

Cluster 7 carries more traffic compared to cluster 9. The variability is also high due to a large number of weekend patterns among these clusters. In addition, cluster 7 patterns mostly belong to urbanized areas compared to predominately rural area type of cluster 9 patterns. The monthly variation of traffic for cluster 7 does not exhibit any significant peaks, and weekday traffic is more compared to weekends. Hence, cluster 7 patterns can be labeled as *commuter* traffic pattern type. Cluster 9 has patterns mostly from rural interstates and they can be labeled as *non-urbanized interstate* pattern type.

Cluster 6 carries traffic volumes less than 4,550 vehicles per day. The patterns cover the traffic from arterials and collectors in rural areas. The monthly variations show steady and higher traffic in the summer months. The patterns reflect *recreational summer* trend in cluster 6. Cluster 13 carries a lot of urban interstates traffic with a mean AADT of 133,000 vehicles per day. This cluster has low variability in AADT and patterns exhibit more workweek traffic with stable monthly patterns. Cluster 13 can be categorized as *urban interstate* traffic trend. Like cluster 13, cluster 14 has more urban interstate traffic patterns with steady monthly trends and flat evening peaks. Cluster 14 can also be classified as *urban interstate* traffic trend. Table 2-5 lists the patterns' trend names.

The CV (ratio of standard deviation to the mean) of hourly traffic volumes shows less variability among most of the clusters except for cluster 6. The off-peak period shows more variability than peak period in which the traffic facilities are operated near capacity conditions (see Figure 2-12). Cluster 12 shows low relatively variability consistently across the 24 hour period.

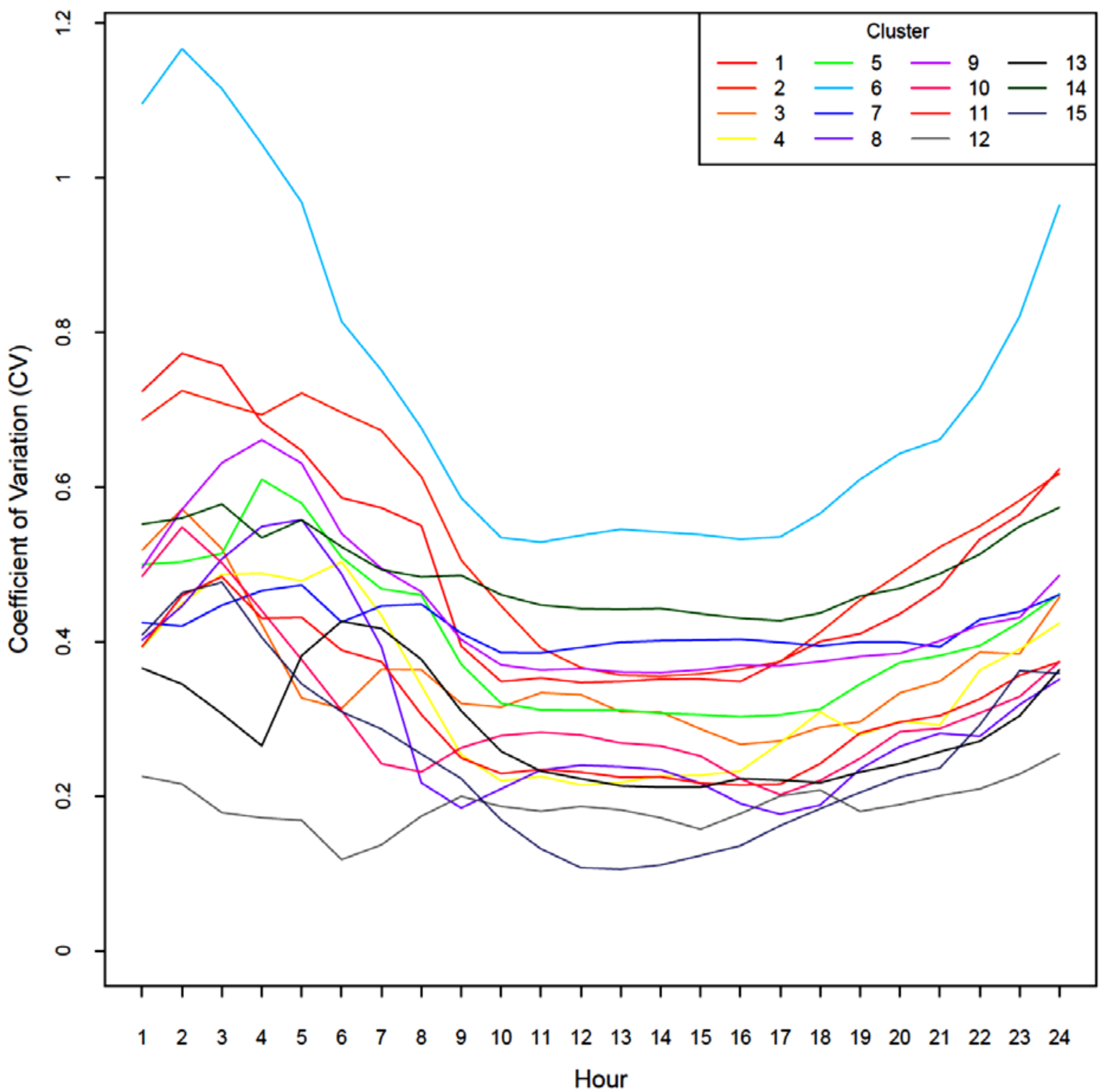


Figure 2-12 Coefficient of Variation of GMM Cluster

k-means

The KM clustering solution produces six clusters using the Calinski and Harabasz criterion (see Figure 2-13). The mean hourly traffic volumes are plotted in Figure 2-14. Clusters 1 and 2 exhibit two peaks with more pronounced evening peaks. However, the peak traffic is different in both clusters. The traffic is growing steadily from morning to evening peaks. Cluster 3 only produces an evening peak with a flat peak region. Cluster 4 and 5 shows similar hourly traffic variations, but cluster 5 carrying more hourly traffic volumes. Cluster 6 carries higher hourly traffic volumes compared to other groups. Cluster 6 has both morning and evening peaks carrying almost equal peak hour volumes. Cluster 4 has the highest proportion and almost half of the patterns (48.7%), followed by cluster 5 with 26 percent of the patterns and cluster 1 with 13 percent of the patterns. The rest of the clusters have less than 5 percent of the patterns.

Figure 2-13 Calinski-Harabasz Criterion for the KM and HCA Clustering

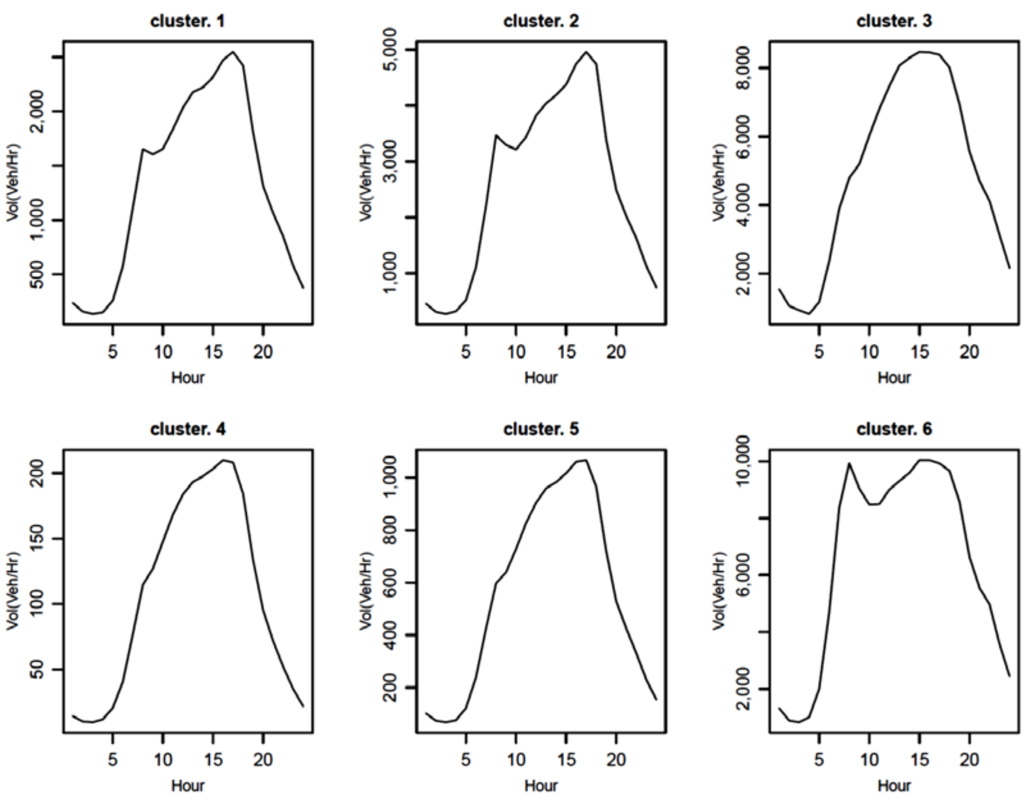


Figure 2-14 Mean Hourly Distribution of Traffic Patterns by K-means

HCA

The HCA clustering selects only three clusters in the solution (see Figure 2-15). The clusters exhibit three different hourly patterns with varying hourly traffic volumes. Cluster 1 has an evening peak and carries low traffic. Cluster 2 has two peaks with a distinct evening peak and carries moderate traffic (hourly volume less than 3,000 vehicles). In cluster 2, the traffic steadily grows from the morning to evening peak. Cluster 3 has a predominant evening peak and carries higher traffic volumes compared to the other clusters. Cluster 1 covers more than 70 percent of the patterns and cluster 2 has 22 percent of the patterns.

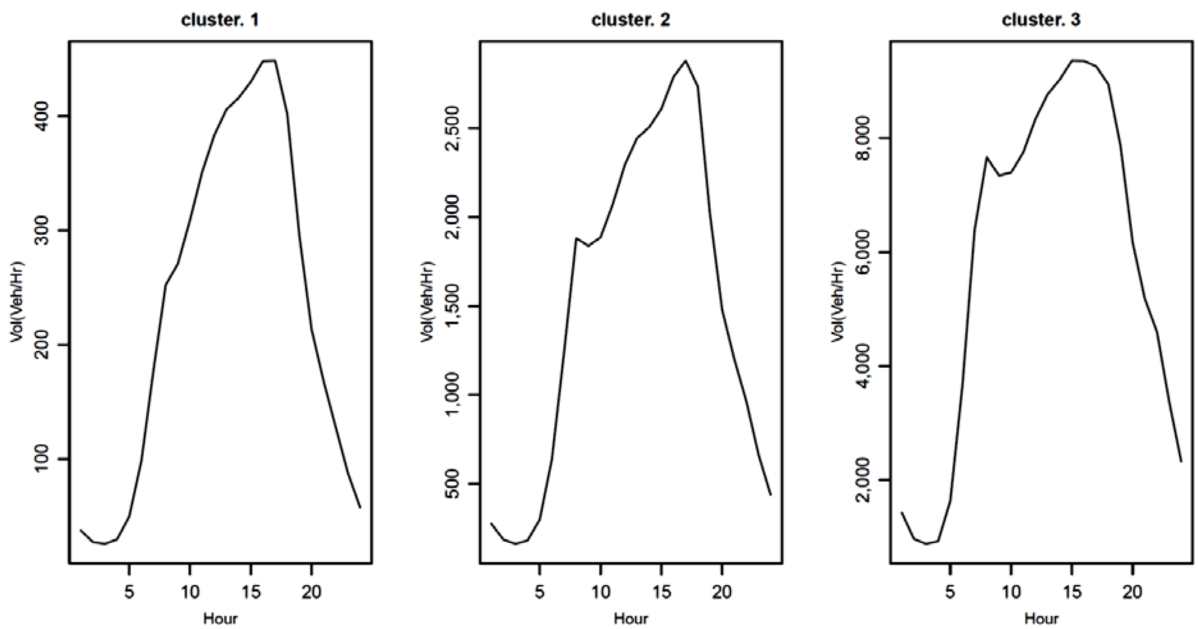


Figure 2-15 Mean Hourly Distribution of Traffic Patterns by HCA

OTHER TRADITIONAL CLUSTERING

In addition, the clustering solutions are compared with the ODOT seasonal trend grouping method (TMG labeled traditional approach) and Functional Class (FC) grouping method (TMG labeled volume factor grouping method). The ODOT seasonal trend grouping method uses knowledge of the road system and visual interpretation. The ODOT, historically, identifies ten seasonal trends: Summer < 2500 AADT, Recreational Summer, Interstate Non-Urbanized, Interstate Urbanized, Agricultural, Commuter, Recreational Summer /Winter, Coastal Destination, Summer, and Coastal Destination Route. The functional class grouping bundles the data patterns based on ODOT highway functional class: Rural Interstate, Urban Interstate, Freeways and Expressways, Principal Arterial, Principal Arterial – Urban, Minor Arterial, and Major Collector. Next, the study performs a robustness analysis on the clustering solutions.

SECOND STAGE: ROBUSTNESS ANALYSIS

The clustering solutions, including traditional methods, have the potential to change or switch cluster membership. For a given clustering method, various clustering validation criteria produce different clustering solutions with no clear winner that suits all situations. The authors are not aware a single criterion that will consistently produce the same clustering solutions for a given problem. The robustness analysis tries to gauge the reproducibility of a given clustering solution under simulated conditions like noise and missing data.

Stability Analysis

Apart from knowing the means to calculate a given cluster validation criterion (like BIC and CH criterion), the knowledge of the distribution properties of these criteria or indices is not possible in most cases (*33*). The distribution of these indices plays a vital role in assessing clustering solution quality and stability. In addition, the question of reproducing the same clustering solution for a new sample drawn from the same input data set remains. The stability analysis may not necessarily reflect the existence of a well-separated cluster (*33, 34*). Assuming that a given criterion assesses the clustering quality, the stability analysis tries to study the distribution of that criterion for different resamples drawn from the sample population. Resampling schemes facilitate a good framework to study the distribution of the interesting criteria/indices, and thereby describe the clustering solution quality.

Resampling methods

When clustering, the sample of traffic patterns (which is a random subset of the population) and clustering methods represent the two main sources of randomness that will potentially affect clustering solutions. The first type of randomness, called sample randomness, comes from the input data. Sample randomness reflects obtaining either similar or very different clustering partitions if another data sample is used for clustering using the same clustering method. The algorithm randomness, the second type of randomness, comes from the application of a clustering algorithm for the same traffic patterns for more than one time. If a given algorithm is repeated several times, due to the stochastic nature of the clustering algorithms, a different partition may be obtained.

Despite randomness, for a clustering solution to be used in practice, a stable partition appears critical for transportation agencies. One way of obtaining multiple samples is to collect multiple yearly traffic patterns for the entire state of Oregon. The ATR data collection program serves that purpose. However, the research team, at the time of analysis, has only two years of complete data. The team is using the entire 2011 dataset for clustering and the 2012 data for AADT validation. In lieu of multiple years of traffic data, resampling methods can help to generate additional samples from the given traffic patterns. The study tests the obtained clustering solution using four different resampling schemes. Each resampling technique, presented in the following sections, introduces a different stability analysis scenario.

Bootstrapping

One of the most widely used resampling methods is bootstrapping. Assume that the study has only one set of yearly traffic patterns $\mathcal{D}$ of fixed size *n* with an unknown distribution. The simplest approximation of the given traffic patterns is to use an empirical distribution $\widehat{F_n}$ of $\mathcal{D}$. Bootstrapping draws samples from $\widehat{F_n}$, which is the same as drawing samples from $\mathcal{D}$ with replacement. Generally, B bootstrap samples,$\{\mathcal{D}^i \cdots \mathcal{D}^B\}$, are drawn from $\mathcal{D}$ with replacement. Clustering algorithms gives partitions $\mathcal{C}^i \cdots \mathcal{C}^B$. The stability is assessed by comparing each of the $\mathcal{C}^i$ partition with the original partition $\mathcal{C}$ using a cluster evaluation criterion.

The general framework for bootstrapping is listed in steps below (*10*):

1. Set bootstrap run i=1
2. Draw a sample $\mathcal{D}^i$ of size *n* from dataset $\mathcal{D}$
3. Apply cluster analysis and obtain a partition $\mathcal{C}^i$ with *K* clusters
4. Compare partition $\mathcal{C}^i$ and original partition $\mathcal{C}$ having *K* clusters, and compute criterion/index/statistic s^i
5. When $i<B$, increase *i* by one and repeat from step 2
6. Summarize criterion/index/statistic for all samples $\{s^i \cdots s^B\}$ for $1 \leq i \leq B$

This section only considers non-parametric bootstrapping. Parametric bootstrapping seems suitable if the clustering is performed by an underlying model, like the GMM solution. When comparing the stability of clusters that include both model (GMM) and non-model (KM and HCA solutions) based solutions, non-parametric bootstrapping can be used. Like bootstrapping, three other resampling schemes, *replacing points by noise, jittering, and subsetting*, are commonly used for cluster stability analysis (*33*).

Replacing Data Points by Noise

Replacing data points by noise explores the strength of the cluster patterns. If the clustering produces a similar partition as an original solution in spite of random perturbations, then the clustering appears stable. A certain number of traffic patterns *m* ($m << n$) are drawn from $\mathcal{D}$ without replacement. These *m* patterns are replaced with points drawn from a noise distribution. The general framework for stability is applied for the remaining (*n*-*m*) non-noise patterns. However, the choice of *m* (number of data patterns to be replaced with noise) and noise distribution affects the clustering solution. The study adopts the guidelines provided in Hennig (*33*) for noise distribution. The study considers ten percent of data patterns to be replaced with noise data (or $m = 0.1n$).

Jittering

Jittering adds a small amount of noise to every single traffic data pattern. This small noise may represent the ATR measurement error or logging error. With jittering, traffic patterns data $\mathcal{D}^i$ for a resample i is represented as $\{\mathbf{y}_1,...,\mathbf{y}_n\}$ where $\mathbf{y}_j = \mathbf{x}_j + \mathbf{e}_j$ (for $j = 1,...,n$) with $\mathbf{e}_j$ being the simulated measurement error. The study adopts the guidelines provided in Hennig (*33*) for introducing the measurement error. The normal distribution is traditionally used as the measurement error distribution (*33*).

Subsetting

Instead of drawing n sample from the dataset (same size as original data set) in bootstrapping, subsetting allows drawing a subsample of $\mathcal{D}$ *without* replacement. The size of subset m is less than n. The value of m cannot be too large (large subsample will not be informative due to lack of variation) or too small (can produce a worse solution than the original). Hennig (*33*) suggests using half of the original data (but randomly selected without replacement) for performing subsetting analysis.

Criteria for Stability Analysis

Stability analysis generally uses two types of criteria, either distances or indices, for cluster comparisons. For understandability and interpretability, the study considers index based measures for stability analysis. However, most of the distance measures are calculated by subtracting the index value from one (*10*). The indices take values between 0 and 1. A larger index value indicates more similarity between two clusters with a value of one being a perfect match. Most of the studies use either a *Jaccard Index* or *Adjusted Rand Index* for stability analysis. For a complete review of other indices and distance based measures refer to Hennig *et al*. (*10*).

For a given dataset $\mathcal{D}$ with n patterns and two clustering solutions $\mathcal{C}$ and $\mathcal{C}'$ of $\mathcal{D}$. Let N_{11} be the number of data patterns that are in the same cluster under both $\mathcal{C}$ and $\mathcal{C}'$. N_{00} be the number of data patterns that are in different clusters under both $\mathcal{C}$ and $\mathcal{C}'$. N_{10} be the number of data patterns that are in the same clusters under $\mathcal{C}$ but not under $\mathcal{C}'$. N_{01} be the number of data patterns that are in the same clusters under $\mathcal{C}'$ but not under $\mathcal{C}$.

The Jaccard index or coefficient is defined as:

$$\mathcal{J}(\mathcal{C},\mathcal{C}') = \frac{N_{11}}{N_{11}+N_{10}+N_{01}} \text{ or } \mathcal{J}(\mathcal{C},\mathcal{C}') = \frac{|\mathcal{C}\cap\mathcal{C}'|}{|\mathcal{C}\cup\mathcal{C}'|} \quad (10)$$

The Adjusted Rand Index (ARI) is defined as:

$$\mathcal{AR}(\mathcal{C},\mathcal{C}') = \frac{\mathcal{R}(\mathcal{C},\mathcal{C}') - E[\mathcal{R}]}{1 - E[\mathcal{R}]} \tag{11}$$

$$\mathcal{R}(\mathcal{C},\mathcal{C}') = \frac{N_{11} + N_{00}}{n(n-1)/2} \tag{12}$$

Where $\mathcal{R}(\mathcal{C},\mathcal{C}')$ is a rand index between clusters $\mathcal{C}$ and $\mathcal{C}'$. E[$\mathcal{R}$] is the expected values of the rand index.

Cluster-wise Stability Assessment

Cluster-wise stability assessment measures the stability of each cluster in a given partition. Let K be the number of clusters given by the original partition $\mathcal{C}$ with $\{C_1,\ldots,C_K\}$ representing the clustering groups 1 to *K* of the original dataset $\mathcal{D}$. When a resampling scheme draws a sample $\mathcal{D}^i$ of size *n* from the original data and performs clustering on $\mathcal{D}^i$ with *K* clusters, it obtains a partition $\mathcal{C}^i$ with $\{{C_1}^i, \cdots, {C_K}^i\}$ clustering groups. The intersection of resampled data and original sample ($\mathcal{D} \cap \mathcal{D}^i$) is used in evaluation (this way observations contained more than once in the sample do not count multiple times). The Jaccard coefficient measures the stability or agreement between the original and resamples data for each cluster as shown below (*10*):

$${s_k}^i = \max_{1 \le k' or\, k \le K} \frac{\left|C_k \cap C^i_{k'}\right|}{\left|C_k \cup C^i_{k'}\right|}, i = 1, \cdots B \tag{13}$$

The mean value given below is used as a cluster-wise (C_k) stability indicator (*10*).

$$s_k = \frac{1}{n} \sum_{i=1}^{B} {s_k}^i \in [0,1] \tag{14}$$

Number of Resampling Runs

The resampling schemes give an idea of variations in the clustering. However, the study does not have an idea of underlying distribution and more importantly *the true clusters* of the traffic data pattern. In the absence of them, resampling uses the empirical distribution of the observed data to draw samples and help to compute the stability indicators. The bootstrapping/resampling samples (*B*) do not have to be very large. Hennig (*33*) suggests that data mining applications with a large sample size may only need five replications to be informative on criteria distribution. The GMM clustering for a dataset of 30,393 patterns with 24-hour traffic data (30,393×24) demands more computational resources and execution time. Hence, the study limits replications to only twenty while performing the stability analysis. However, more resampling datasets (usually *B* value between 50 and 100) may truly assess the variations in the clustering.

Analysis

The study checks how the fifteen clusters in the GMM solution behave when applying different resampling schemes. The investigation replicates the four resampling schemes 20 times and calculates the Jaccard coefficient of each cluster for each replication. Figure 2-16 shows the variation of the Jaccard coefficient for the GMM clusters under different resampling schemes. Henning (*34*) suggests that a stable cluster should yield a mean Jaccard similarity value of 0.75 or more. In addition, the coefficient values between 0.6 and 0.75 show that clusters indicate patterns in the data.

Table 2-6 shows the cluster-wise stability assessment for all three clustering solutions. The analysis shows that clusters 1, 6 and 10 remain stable in the GMM solution for all resampling schemes. Cluster 9 is stable under noise and jittering conditions, but unstable for bootstrapping and subsetting. Bootstrapping produces very unstable clusters for cluster 8 and 15. However, the introduction of noise to the original data shows improved cluster performance in terms of better coefficient values compared to bootstrapping. In addition, the noise and jittering schemes produce mostly similar coefficient values except for clusters 7, 8, 9 and 10. Subsetting, like bootstrapping, produces mostly unstable clusters.

Both k-means and HCA show stable clusters under all four conditions. However, the introduction of noise produces instability in cluster 6 of the k-means solution. In addition, noise produces lower Jaccard coefficient values relative to other schemes for all clusters. Jittering does not have any effect on coefficient values. Bootstrapping and sub-setting almost yield similar Jaccard coefficient values. The HCA clustering solution with only three clusters is expected to show stable clusters. The coefficient values are greater than 0.6 for all schemes. However, cluster 2 shows a lower Jaccard coefficient values for all four schemes.

Seven clusters, almost half of the clusters, of the GMM solution perform better under noise conditions. However, the k-means solution and HCA solution do not perform well under noise conditions compared to the other resampling schemes.

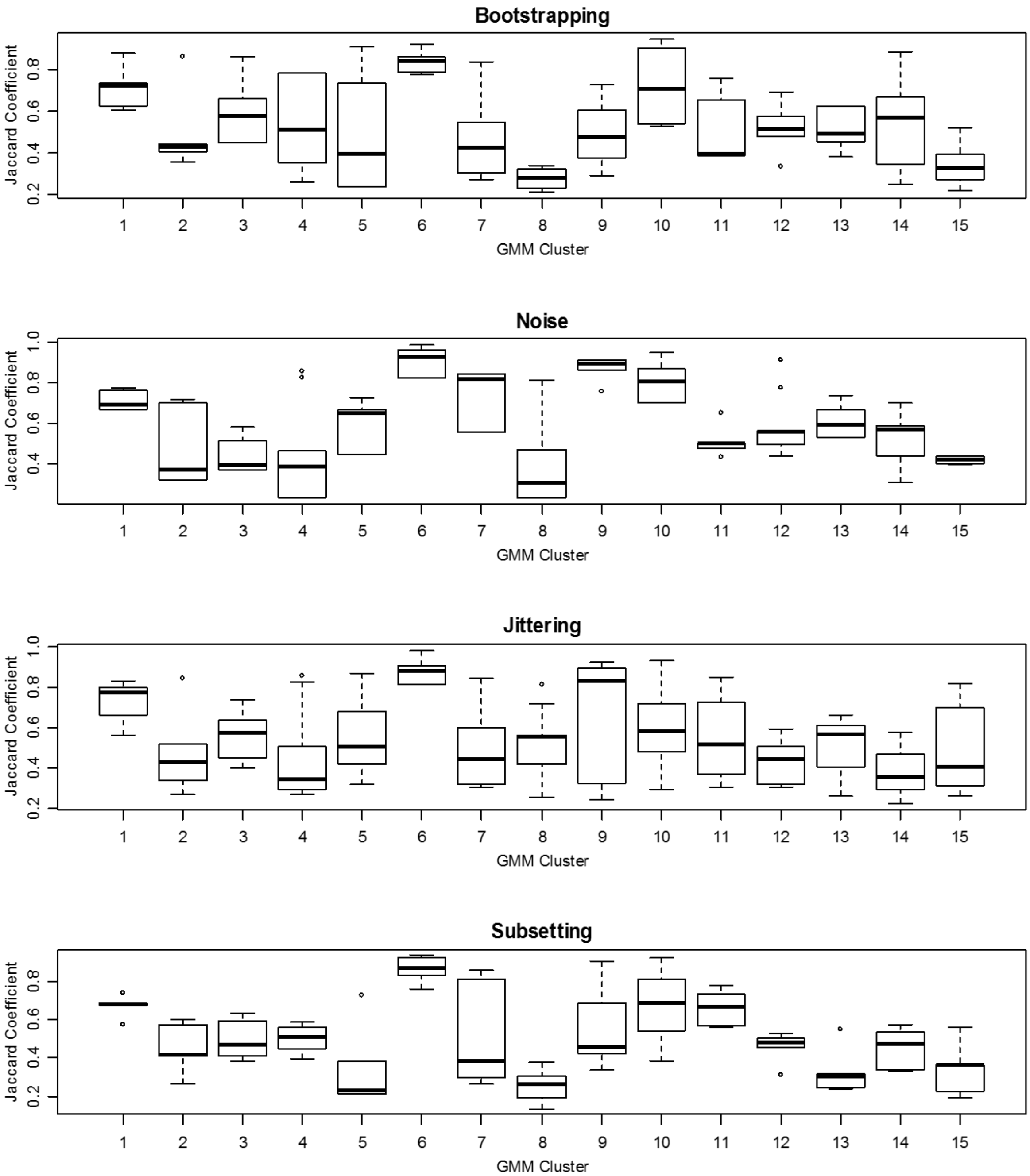


Figure 2-16 Jaccard Coefficient for the GMM Clusters under Stability Analysis

Table 2-6 Cluster-wise Stability Assessment

Clustering Method	Cluster No.	Jaccard Coefficient			
		Bootstrapping	Replacing Data Points by Noise	Jittering	Subsetting
GMM	1	0.69	0.71	0.73	0.68
	2	0.46	0.46	0.46	0.47
	3	0.59	0.45	0.56	0.50
	4	0.52	0.43	0.46	0.50
	5	0.48	0.60	0.54	0.34
	6	0.83	0.90	0.87	0.87
	7	0.44	0.74	0.51	0.48
	8	0.27	0.35	0.52	0.25
	9	0.48	0.88	0.66	0.55
	10	0.71	0.81	0.60	0.66
	11	0.51	0.51	0.55	0.67
	12	0.50	0.57	0.44	0.46
	13	0.52	0.61	0.52	0.31
	14	0.52	0.53	0.38	0.44
	15	0.35	0.42	0.50	0.32
KM	1	0.99	0.87	1.00	0.98
	2	0.98	0.85	1.00	0.96
	3	0.99	0.66	1.00	0.99
	4	1.00	0.71	1.00	1.00
	5	0.98	0.89	1.00	0.97
	6	0.99	0.27	1.00	0.99
HCA	1	0.90	0.92	0.96	0.91
	2	0.72	0.73	0.87	0.74
	3	0.97	0.96	1.00	0.98

The stability analyses do not always reflect the validity of the clusters, but they provide more information on the clusters under a given cluster modeling framework. Large stability values do not guarantee clusters that are more valid. Small stability values, often clusters with a Jaccard coefficient less than 0.6, may correspond to either meaningless clusters for a given cluster model or actual instabilities in the clusters themselves or clustering method. However, Table 2-6 shows Jaccard coefficient values across alternative resampling schemes, the ranking among schemes depends on the data and clustering method. The study cannot recommend the best scheme to use in the stability analysis of clusters. However, different schemes bring valuable insights to the clustering solution under varying conditions. In general, the noise methods reflect more valuable information than the other methods. The stability analysis suggests the use of partitions from either HCA or KM methods to perform clustering analysis. However, stable clusters do not necessarily produce less AADT estimation errors. The true ability of a given clustering method is only realized if a clustering model framework yields lower AADT estimation errors.

Missing values

The traffic data may contain missing information due to malfunctioning detectors. In addition, data logging and data storage may potentially cause missing data. The improper maintenance of detectors or extreme weather conditions may also contribute to detector outages. In any case, the missing data influences the stability of the cluster solutions, as well as the development of seasonal factors and thereby the AADT estimations. The clustering solution with missing traffic data may produce biased AADT estimates and affect the forecasting error. Methods to impute the missing data generally exist in the literature; however, the study does not seek to test the missing values' impact on AADT estimation error. Instead, the authors test the stability (or label switching) of the current GMM clustering solution due to missing patterns.

The study (deliberately and) randomly chooses ten percent of the traffic data and marks it as missing. The authors pick ten percent of the data and exclude it without replacement. The trial uses the rest of the data (90 percent of data) for performing a clustering solution stability analysis. The researchers replicate this analysis ten times so that all traffic patterns can be missed exactly once. Each time a completely different random dataset is chosen as the missing patterns. When performing clustering analysis on the reduced data sets, the study restricts the GMM model type (model "VVV") and number of clusters to fifteen (as that of original number of clusters). These restrictions can help to see the deviation from the original clustering solution. The study adopts the ARI to study the similarity between the clustering solutions. An ARI value of one indicates identical clusters and zero indicates that partitions are independent (*10*). The missing data analysis gives ARI values between 0.44 and 0.55. ARI values less than one show that the missing value does effect the GMM clustering solution.

The study also tests the missing data impact on the KM and HCA solutions. Figure 2-17 The Adjusted Rand index (ARI) values from Missing Data Analysis compares the ARI values of ten replications for all three clustering methods. Even though the HCA solution shows stability previously, the missing data produces ARI values of almost zero (less than 0.12). The ARI value for the HCA solution shows that the reduced data partitions significantly differ from the original classification. The KM solution shows an improvement in ARI values compared to the HCA solution. Hence, missing values affect the KM and HCA clustering solutions more significantly than the GMM solution.

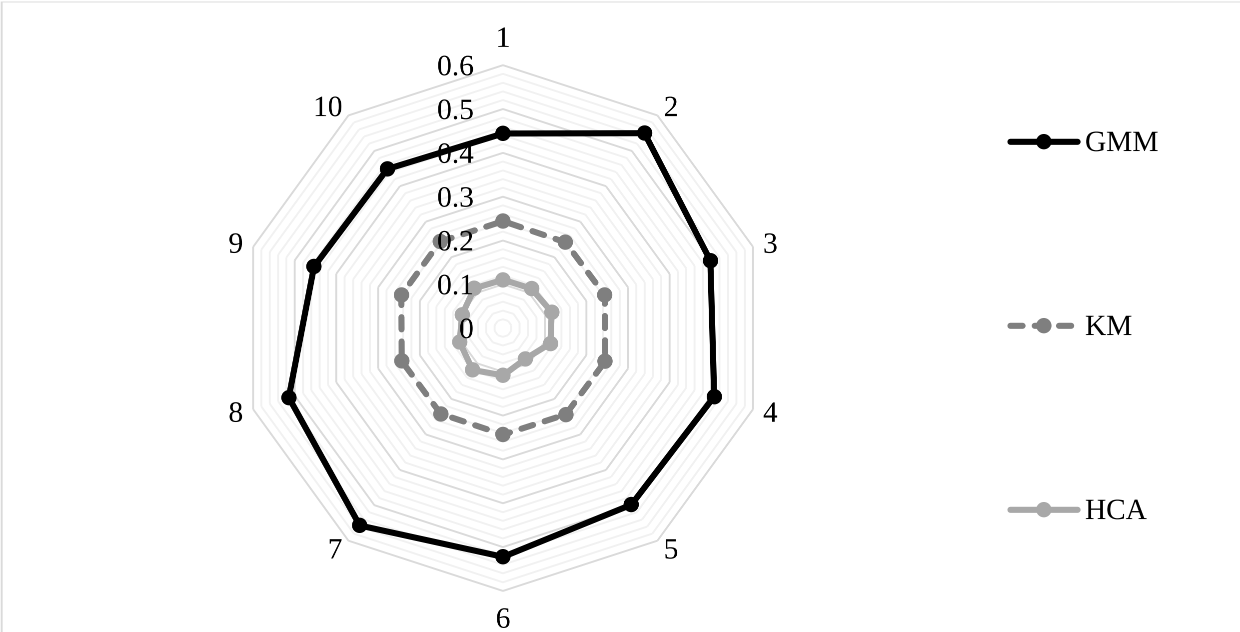


Figure 2-17 The Adjusted Rand index (ARI) values from Missing Data Analysis

Variable Size

The agencies, usually, collect STTCs for multiple days (or multiple 24 hours). In certain circumstances, depending on project or study need, shorter than 24- hour counts may be warranted to effectively use limited resources. Either 16-hour counts (usually from 600 to 2200 Hrs) or 12-hour counts (600 to 1800 Hrs) may be substituted for 24-hour counts. This study tries to find, whether the clustering solution for 16-hour or 12-hour pattern data differs from of the solution based on 24-hour patterns. The variable impact study checks the extent of the domination of certain variables on the cluster solution. To realize this, the investigation employees the clustering methods on a reduced dataset with 16-hours and 12-hours traffic patterns. An ARI value close to one infers that the reduced data set does not change the clustering much, and the omitted hourly patterns have a low impact on the clustering.

The GMM clustering produces different clusters for the reduced data. The ARI shows that the GMM solution for the reduced data does not agree with the solution on full data sets (see Table 2-7). Even though the number of clusters differs for datasets, the GMM solution agrees on the suitability of an unconstrained model type (model "VVV") for both the full and reduced hourly traffic patterns. K-means clustering seems insensitive to the reduced datasets and produces a six-cluster solution with an ARI value close to one (see Table 2-7). The K-means clustering suggests that the clustering solution for full data sets and either 16-hour or 12-hour datasets do not differ much. In addition, the reduced hourly traffic patterns have a low impact on the clustering output. The HCA method produces different clustering solutions for 24-hour, 16-hour and 12-hour data sets. Both reduced data sets have eight cluster solutions.

Though different clustering methods show the importance of variables and also produce different clustering solutions, the study takes the original clustering solutions from all clustering methods obtained on the full data set (or 24-hour patterns). The variable size analysis only provides insight into the impact of the variables on the clustering solution. Any decision and related analysis on which dataset (24-hour, 16-hour or 12-hour) produces less AADT estimation error should be investigated in a future study.

Table 2-7 The ARI values and Number of Clusters for Reduced Hourly Datasets

Clustering Method	ARI		Number of Clusters		
	Hr16	Hr12	Hr24	Hr16	Hr12
GMM	0.53	0.49	15	23	20
KM	0.99	0.98	6	6	6
HCA	0.49	0.42	3	8	8

Checking Need for Clustering

Without clustering, the 2011 dataset can be grouped by functional classification (volume factor approach suggested by the FHWA TMG). If any of the clustering solution appears close to the grouping by functional class, the clustering solution becomes trivial compared to the existing default or natural grouping. The ARI is computed between the groups in the functional class and the partition obtained using clustering methods. The ARI between GMM, KM, and HCA versus functional class is 0.13, 0.24 and 0.25. The ARI values show that the clustering solution differs from the default or natural grouping. However, these methods need to be tested against the AADT estimation error to select the best clustering method.

THIRD STAGE: SEASONAL ADJUSTMENT FACTORS

After the clustering step, each cluster group is again sub-grouped by month to compute the average seasonal factors (ratio of AADT to DT). For instance, the GMM's fifteen cluster solution produces a table of 180 average ratios of AADT to DT (15 groups × 12 months) after sub-grouping. The subgrouping helps to address monthly and seasonal variation of the traffic data. The subgrouping of each cluster by month continues for the KM, HCA, seasonal trend grouping (ODOT), and highway functional class (FC) grouping methods (see Table 2-8 to Table 2-12).

Table 2-8 Seasonal Adjustment Factors for the GMM Clustering

GMM Cluster	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	0.79	0.80	0.84	0.91	1.00	1.16	1.32	1.31	1.23	1.05	0.90	0.86
2	0.75	0.79	0.86	0.93	1.06	1.28	1.53	1.46	1.33	1.05	0.83	0.76
3	0.88	0.89	0.87	0.91	0.93	1.03	1.22	1.29	1.13	0.94	0.80	0.80
4	0.93	0.93	0.97	1.02	1.06	1.12	1.16	1.15	1.13	1.06	0.98	0.95
5	0.80	0.87	0.88	0.90	0.99	1.17	1.31	1.32	1.20	0.98	0.82	0.80
6	0.61	0.63	0.69	0.78	0.93	1.18	1.29	1.29	1.26	1.02	0.73	0.69
7	0.76	0.81	0.84	0.87	0.90	0.97	1.05	1.05	1.01	0.88	0.85	0.81
8	0.96	0.96	0.98	1.02	1.05	1.11	1.19	1.18	1.11	1.04	0.98	0.98
9	0.74	0.78	0.87	0.92	0.97	1.10	1.30	1.29	1.12	0.93	0.84	0.81
10	0.90	0.93	0.98	1.03	1.04	1.11	1.18	1.17	1.12	1.05	0.99	0.94
11	1.02	1.04	1.05	1.09	1.09	1.13	1.16	1.14	1.12	1.08	1.05	1.05
12	1.01	1.02	1.05	1.07	1.07	1.11	1.12	1.10	1.08	1.05	1.03	1.04
13	0.96	0.98	0.99	1.03	1.00	1.06	1.05	1.06	1.06	1.02	0.98	1.01
14	0.82	0.85	0.88	0.89	0.88	0.96	1.00	0.99	0.91	0.90	0.87	0.87
15	1.01	1.02	1.04	1.07	1.08	1.13	1.13	1.11	1.10	1.07	1.05	1.04

Table 2-9 Seasonal Adjustment Factors for the KM Clustering

KM Cluster	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	0.95	0.97	1.00	1.01	1.02	1.08	1.10	1.11	1.06	1.01	0.99	0.97
2	0.96	0.99	1.02	1.05	1.05	1.11	1.12	1.11	1.09	1.05	1.03	1.00
3	0.87	0.91	0.94	0.95	0.93	0.98	0.95	0.99	0.94	0.93	0.90	0.91
4	0.74	0.77	0.81	0.88	0.98	1.16	1.32	1.30	1.22	1.02	0.83	0.78
5	0.86	0.91	0.96	0.99	1.01	1.10	1.23	1.23	1.12	0.99	0.93	0.91
6	1.03	1.05	1.06	1.07	1.07	1.10	1.10	1.09	1.08	1.06	1.04	1.05

Table 2-10 Seasonal Adjustment Factors for the HCA Clustering

HCA Cluster	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	0.78	0.81	0.85	0.91	0.99	1.14	1.29	1.28	1.19	1.01	0.86	0.82
2	0.93	0.96	0.99	1.01	1.02	1.08	1.11	1.12	1.06	1.01	0.99	0.97
3	0.95	0.98	1.01	1.02	1.01	1.05	1.03	1.05	1.02	1.00	0.97	0.98

Table 2-11 Seasonal Adjustment Factors for the ODOT Grouping

ODOT Seasonal Trend	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Agricultural	0.81	0.84	0.86	0.93	1.03	1.14	1.22	1.20	1.28	1.08	0.89	0.78
Coastal Destination	0.82	0.88	0.91	0.94	0.96	1.10	1.23	1.26	1.16	0.97	0.88	0.85
Coastal Destination Route	0.71	0.76	0.84	0.88	0.97	1.12	1.43	1.44	1.25	0.96	0.80	0.75
Commuter	0.92	0.96	0.98	1.02	1.01	1.05	1.03	1.06	1.05	1.01	0.97	0.94
Interstate Non-Urbanized	0.77	0.83	0.95	0.98	1.00	1.13	1.21	1.20	1.10	1.00	0.94	0.85
Interstate Urbanized	0.93	0.97	1.00	1.01	1.00	1.06	1.04	1.06	1.03	1.00	0.97	0.96
Recreational Summer	0.62	0.64	0.69	0.78	0.99	1.26	1.56	1.50	1.36	1.04	0.76	0.71
Recreational Summer / Winter	1.04	1.11	1.02	0.87	0.79	1.03	1.43	1.34	1.06	0.83	0.70	1.03
Summer	0.85	0.88	0.92	0.97	1.01	1.10	1.18	1.19	1.09	0.99	0.91	0.88
Summer < 2500	0.80	0.82	0.86	0.97	1.03	1.14	1.18	1.18	1.14	1.03	0.93	0.87

Table 2-12 Seasonal Adjustment Factors for the Functional Class Grouping

Functional Class	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Rural Interstate	0.76	0.81	0.95	0.98	1.01	1.13	1.23	1.22	1.11	1.00	0.94	0.85
Other Principal Arterial	0.79	0.82	0.85	0.88	0.98	1.14	1.32	1.31	1.22	1.01	0.84	0.82
Minor Arterial	0.82	0.85	0.89	0.94	0.99	1.11	1.22	1.22	1.16	1.04	0.93	0.86
Major Collector	0.67	0.69	0.73	0.89	1.04	1.28	1.44	1.40	1.27	1.00	0.75	0.68
Urban Interstate	0.91	0.95	0.99	1.00	1.00	1.06	1.06	1.07	1.03	1.00	0.97	0.95
Other Freeways and Expressways	0.93	0.97	0.99	1.01	1.01	1.05	1.02	1.05	1.03	1.00	0.97	0.97
Other Principal Arterial - Urban	0.91	0.95	0.96	1.01	1.01	1.07	1.07	1.07	1.03	1.00	0.95	0.95

FOURTH STAGE: CLASSIFICATION

Next, the research team tests the relative merit of the cluster solutions for estimating the AADT. The study adopts respective cluster solutions for estimation. First, the cluster solution classifies each data pattern into one of the groups (for instance, one of fifteen clusters for the GMM solution). However, the study needs a classifier that assigns the data patterns into one of the groups. In the past, researchers have developed numerous methods for assigning traffic data to factor groups (see reference (*3*, *35*)). Both reasonably grouping the traffic data (the current chapter's focus) and assigning traffic data to the correct groups (classification) play an important role when computing AADT values.

The study uses the commonly adopted Quadratic Discriminant Analysis (QDA) technique for classifying test data patterns. The QDA statistically assigns a given data pattern to the groups of known characteristics. The misclassification rate (percent of data patterns whose groups are misclassified) shows the predictability of the QDA. The study adopts a 10×10 stratified cross validation approach for the performance evaluation of the classifiers. The 10×10 cross-validation approach makes 10 runs of 10-fold stratified cross-validation of a given data set (*36*). Next, the QDA is performed to classify the training data patterns. The trained QDA classifier has an error rate of 2.8 percent for the GMM clustering, 6.3 percent for the HCA and 7.8 percent for the KM clustering solution.

FINAL STAGE: AADT ESTIMATION

The 2012 ATR data set with a sample of 32,289 data patterns is used for testing. The 2012 dataset has hourly traffic data showing time of day variation and the groupings according to the ODOT seasonal trend grouping method and highway functional class. The QDA classifies the test data and assigns the group number according to the GMM, KM, and the HCA solution.

The AADT is calculated by matching the *group number and month* of the test patterns (using the 2012 data). The product of the matched average ratio of AADT to DT (corresponding to a matched *group number and month*) and the sum of 24-hour traffic volume (daily traffic or DT) estimates the AADT. The computed AADT value is compared with the actual AADT value to obtain an error. The Mean Absolute Percent Error (MPAE) given in Equation (15) is used to compare the estimates from the clustering methods (*37, 38*):

$$MAPE = \sum_{n=1}^{N}\left(\left|\frac{AADT_{Estimated} - AADT_{Actual}}{AADT_{Actual}}\right|_{n} \times 100\right) \Big/ n \qquad (15)$$

In addition, the standard deviation of the errors is also computed and reported.

Monthly Variation of Errors

Figure 2-18 shows a variation of the MAPE among the all clustering solutions. In general, winter months (November to March) show more errors than warmer months (April to October). All clustering solutions yield lower error values for the months of June and October. The GMM solution shows lower AADT estimation errors than the KM and HCA solutions (see Figure 2-19). Though the HCA solution shows stable clustering solution, when tested for new data the solution produces more errors. Among the clustering solutions, the errors produced by the HCA solution appear higher for each month. However, all clustering solutions perform better than the default clustering solution using the functional class. The error difference between the GMM and the other clustering solutions appears higher during the winter months (see Figure 2-19). The Functional class grouping consistently produces high error values.

The GMM solution also performs better than the ODOT seasonal trend grouping. When comparing with the ODOT seasonal trending, the AADT estimation errors for all clustering solutions remain smaller for the months October to June. Both, the HCA and KM solutions produce more errors than the ODOT method during July and August. Furthermore, the GMM solution produces a slightly higher error (+0.1%) than the ODOT method during the month of July. In general, the summer months (especially July and August) attract more recreational trips during weekends, which will produce large error variations during these months for all clustering solutions. The error variation by day of the week may provide some insight into the trends that reflect summer/recreational trips.

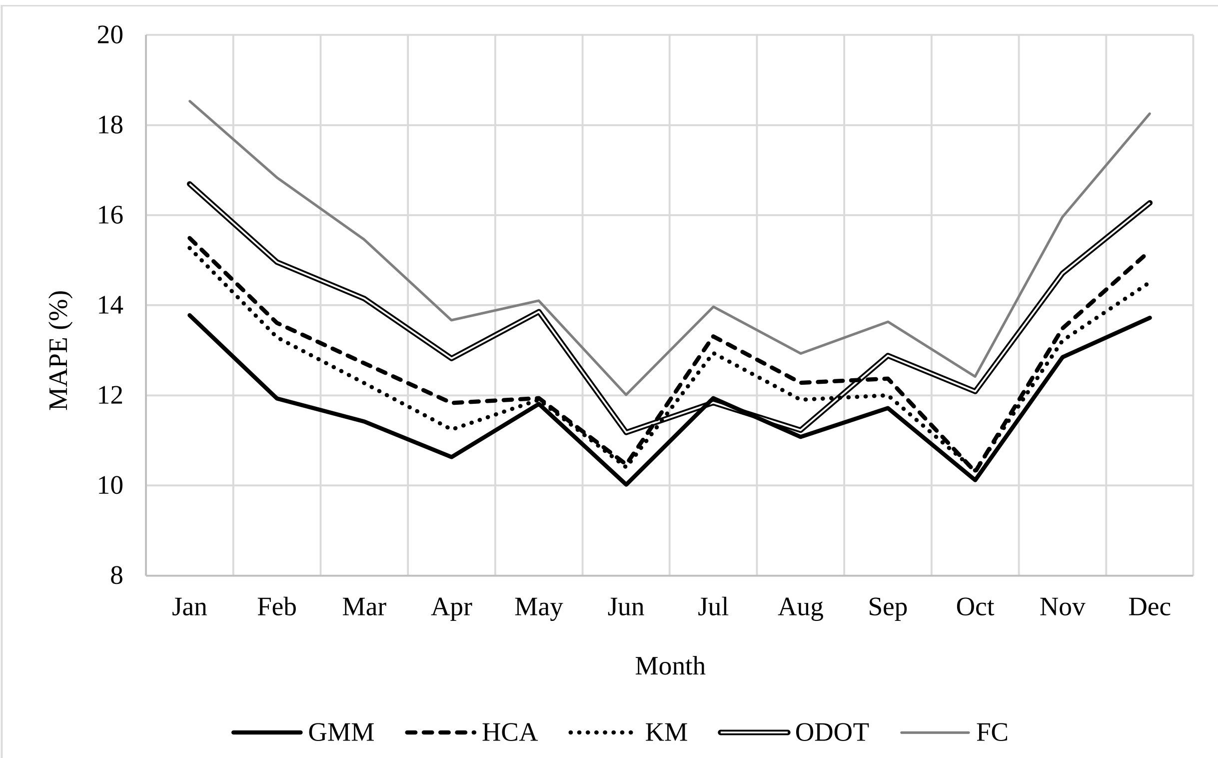
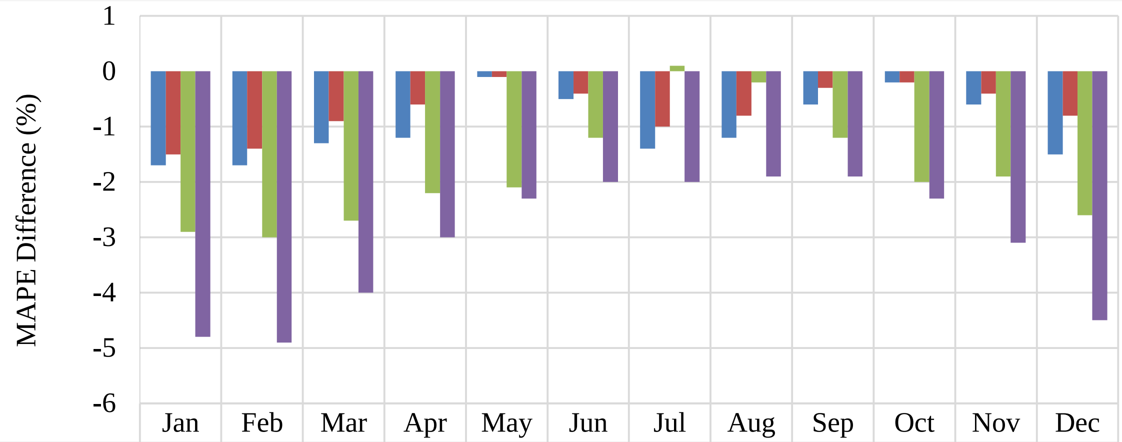


Figure 2-18 Monthly Variation of Errors for Different Clustering Solutions



	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
GMM vs HCA	-1.7	-1.7	-1.3	-1.2	-0.1	-0.5	-1.4	-1.2	-0.6	-0.2	-0.6	-1.5
GMM vs KM	-1.5	-1.4	-0.9	-0.6	-0.1	-0.4	-1	-0.8	-0.3	-0.2	-0.4	-0.8
GMM vs ODOT	-2.9	-3	-2.7	-2.2	-2.1	-1.2	0.1	-0.2	-1.2	-2	-1.9	-2.6
GMM vs FC	-4.8	-4.9	-4	-3	-2.3	-2	-2	-1.9	-1.9	-2.3	-3.1	-4.5

Figure 2-19 MAPE Difference between the GMM versus Other Methods

The standard deviation of errors varies between 10 and 18 percent (Figure 2-20). The standard deviation of the MAPEs remains lower for the GMM solution compared to the HCA and KM solutions except for the month of November. The GMM solution also shows lower error deviations than the default functional class clustering for eight months. However, the deviation of errors for the ODOT method appears smaller than all other methods from June to September. The possible reason could be a large contribution of errors, from relatively a small fraction of patterns, with greater than 25 percent error size (or MAPE > 25%). The error distribution analysis presented in the upcoming sections provides more insight on this issue.

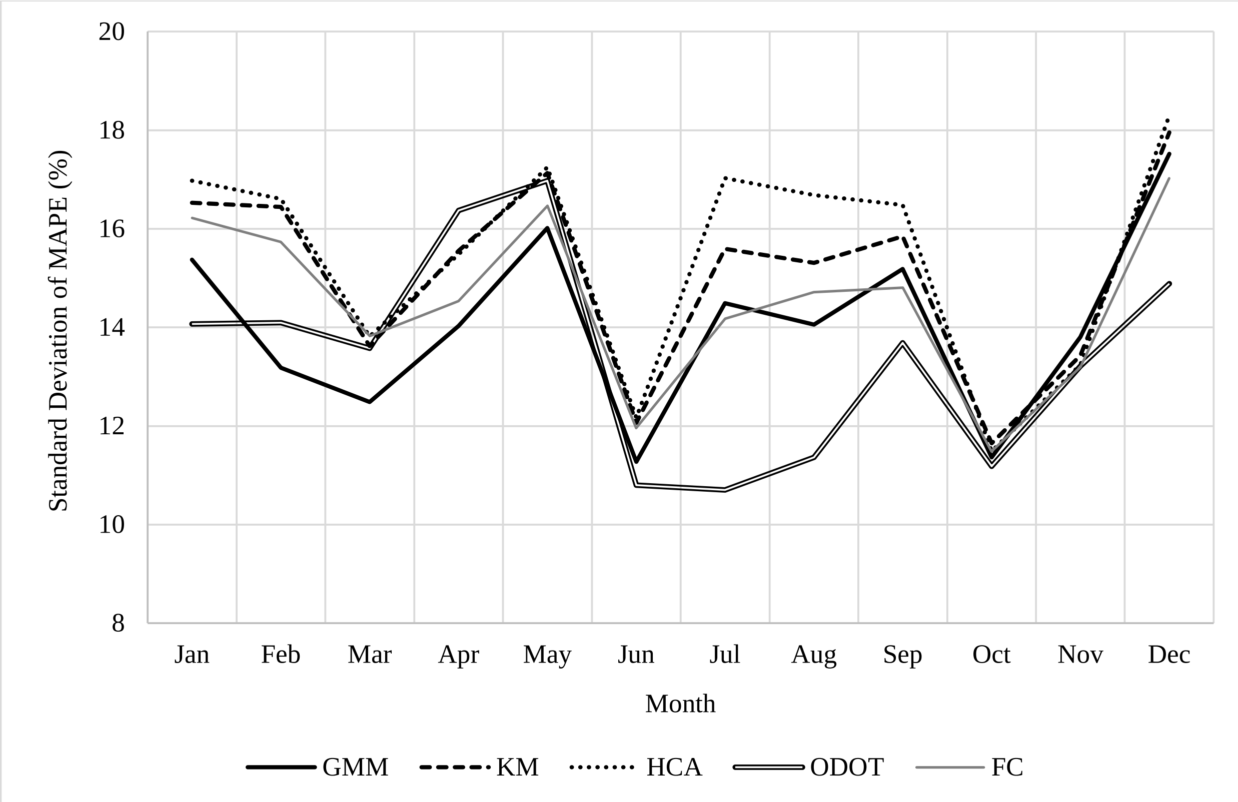


Figure 2-20 Standard Deviation of the Errors among Different Clustering Solutions

Daily Variation of Errors

Figure 2-21 shows the Daily variation of the Absolute Percent Error (DAPE). The error is obtained by averaging the absolute percent errors of a particular day of a week in a given year. The GMM solution produces lower DAPE values compared to the KM and HCA clustering solutions for all days of a week. The functional class grouping consistently gives more error than all other methods. Except for Monday, the ODOT method produces more error than the GMM clustering; however, the ODOT error rate also appears higher than the KM and HCA methods during Fridays and Sundays. The clustering solutions GMM, HCA, and KM produce error rates of less than 15 percent and exhibit higher weekend error rates compared to weekdays.

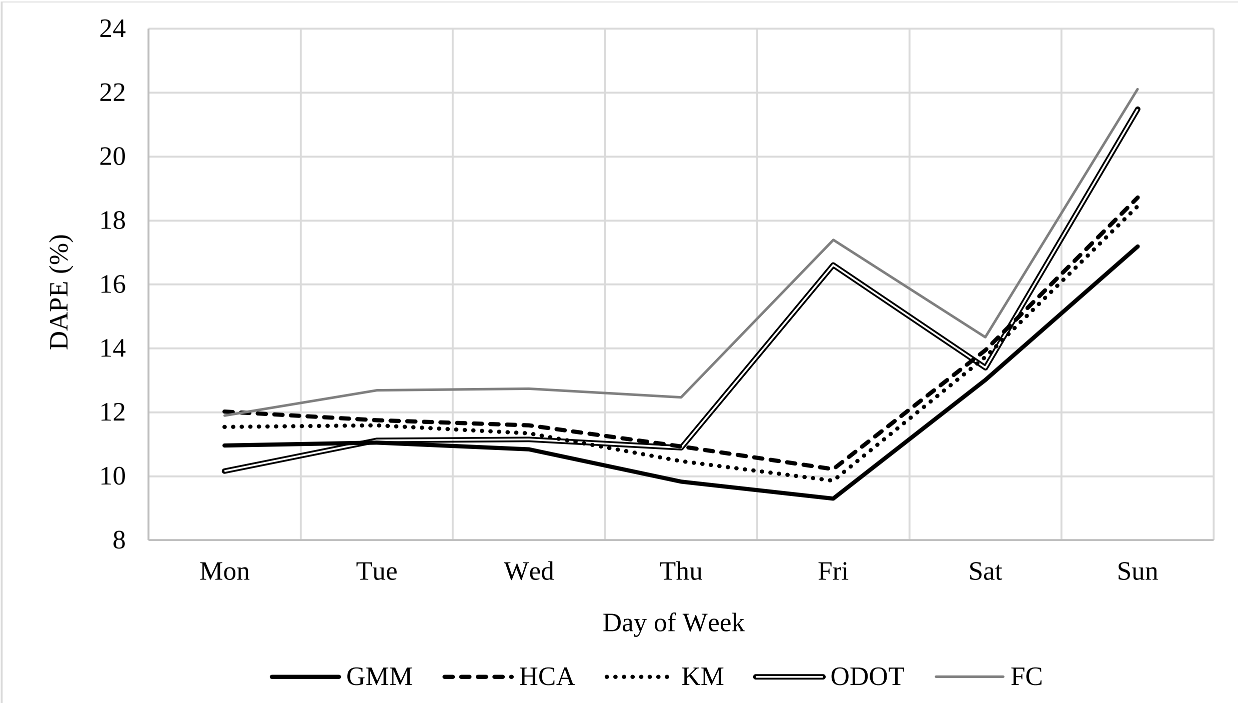


Figure 2-21 Monthly Variation of Errors for Different Clustering Solutions

Error Summary Statistics

Instead of only observing the mean error, looking at the quartiles and median of the error provides a better a picture of the distribution of MAPEs. All clustering solutions produce errors where the mean is always greater than the median, and the error distribution is skewed and not symmetrical around the mean. However, first quartile, median and third quartile errors by the GMM clustering solution error remain lower than all other methods. Table 2-13 lists the summary statistics of errors with the 95 percent confidence intervals for mean.

Table 2-13 Summary Statistics of MAPE Errors

Clustering Method	1st Quartile	Median	3rd Quartile	Standard Deviation	Mean	95 % CI of Mean	
						Lower	Upper
GMM	3.0	7.3	15.4	14.2	11.8	11.6	11.9
KM	3.3	7.9	16.2	15.3	12.5	12.3	12.6
HCA	3.4	8.2	16.7	15.7	12.8	12.6	12.9
ODOT	4.8	10.0	18.2	13.7	13.6	13.4	13.7
FC	5.3	11.0	19.8	14.8	14.8	14.7	15.0

Error Distribution

In addition to the quartiles of the errors, the study also explores the error distribution of MAPEs. Other studies on AADT estimation report obtaining an *average* MAPE value of less than 15 percent from 24-hour traffic data (*1*). Hence, the study creates an initial set of closed groups with an error increment of 5 percent until 25 percent (i.e. MAPE between 0-5, 5-10, 10-15, and 20-25), and sparse error groups beyond the 25 percent error mark (25-50, 50-100 and >100). However, an additional error group where MAPE is just greater than 15 percent is also created to perform a comparative analysis with other studies. Table 2-14 outlines the error distribution among the clustering methods. The GMM solution reports the highest number of patterns with an error of less than 5 percent than other clustering solutions. All three clustering solutions have more patterns that exhibit less than five percent error than the ODOT and functional grouping methods. As the percent error increases, the clustering solutions produce fewer patterns in that particular group than the default clustering. If a MAPE value of fifteen percent represents the benchmark, the GMM solution again produces fewer patterns having a MAPE value greater than 15 percent than all other clustering solutions. All clustering solutions produce fewer patterns than the conventional clustering (ODOT or FC) in this error category. The error distribution shows that the GMM solution error rate mostly stays below 15 percent. Roughly, a quarter of the patterns under the GMM solution produce a MAPE value of greater than 15 percent.

Table 2-14 Distribution of MAPE by Error Size and Clustering Method

Clustering Method	Percentage Patterns with a MAPE size (in %) of								
	0-5	5-10	10-15	15-20	20-25	25-50	50-100	>100	>15
GMM	37.9	22.7	13.7	8.3	5.5	9.9	1.8	0.2	25.8
KM	35.3	23.1	14.0	9.1	5.5	10.5	2.2	0.4	27.6
HCA	34.7	22.3	14.5	9.2	6.4	10.3	2.3	0.4	28.6
ODOT	25.9	23.8	17.5	11.7	7.4	11.6	1.9	0.2	32.8
Functional Class	23.7	22.3	17.0	12.3	8.0	14.2	2.2	0.3	37.0

The study also compares the error distribution for the months of July and August between the GMM and the ODOT solutions. The GMM solution produces more patterns with a MAPE less than five percent (Table 2-15). As the error size increases, the percent of patterns decreases, as expected, for the GMM solution compared to the ODOT solution. However, when the error size is beyond 25 percent, the GMM produces more patterns. Even though the difference between the patterns produced is less than 2.5 percent, the size of the error - when averaging – shows a slightly higher error rate for the GMM in the month of July.

Table 2-15 Selective Monthly Distribution of MAPE by Error Size between the GMM and ODOT

Method	Month	Percentage Patterns with a MAPE size of (%)							
		0-5	5-10	10-15	15-20	20-25	25-50	50-100	>100
GMM	Jul	38.6	21.9	13.6	8.0	5.2	10.6	2.0	0.2
ODOT		27.1	27.2	17.6	11.5	7.2	8.2	1.1	0.0
GMM	Aug	41.0	22.8	13.5	7.7	4.3	8.5	2.0	0.2
ODOT		32.6	26.3	15.9	9.8	6.3	7.6	1.3	0.1

Cluster-wise Error Assessment

Figure 2-22 shows cluster-wise assessment of monthly MAPE values for the GMM clustering solution. Clusters 4, 8, 10, 11, 12, 13, 14 and 15 show less than fifteen percent average error for all months. Except for a few months, Clusters 1, 7 and 9 also show less than fifteen percent error rate. The rest of the clusters mostly have MAPE values greater than fifteen percent. However, the stability analysis presented earlier shows that clusters 1, 6 and 10 appear stable in the GMM solution. The stable clusters, in general, but not necessarily, show lower error rates. Clusters 1 and 10 follow the trend of reduced errors; however, cluster 6 shows a higher error for all months except for August.

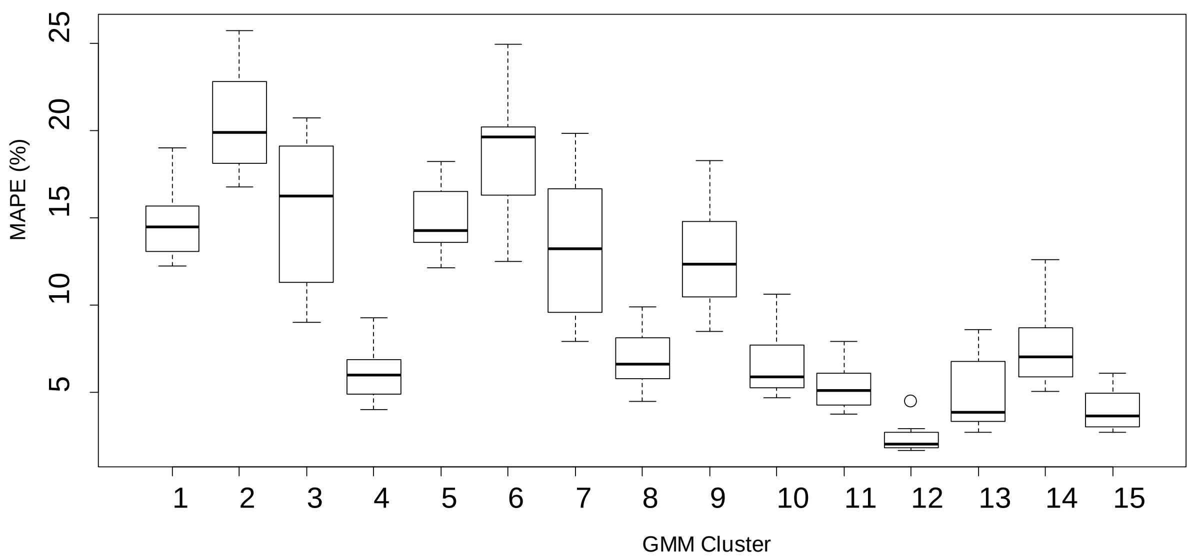


Figure 2-22 The GMM Cluster-wise Monthly Distribution of MAPE values

Stability analysis shows that all clusters produced by either HCA or KM solutions appear stable. However, the monthly error variation shows that cluster 4 of the KM solution and cluster 1 of the HCA solution produces errors mostly greater than 15 percent (see Figure 2-23 and Figure 2-24).

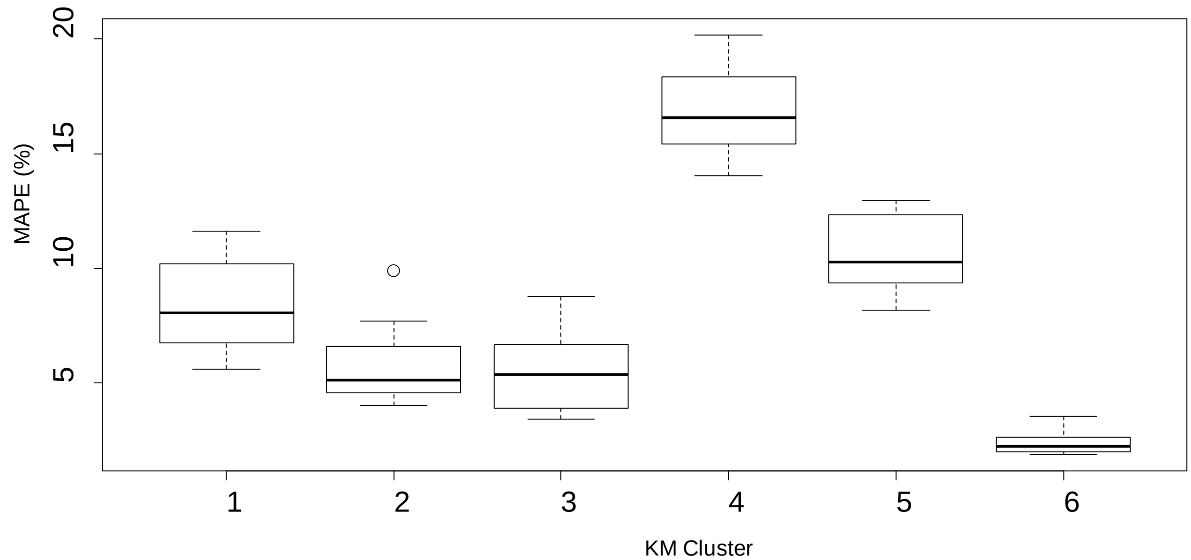


Figure 2-23 Cluster-wise Monthly Distribution of MAPE values for the KM

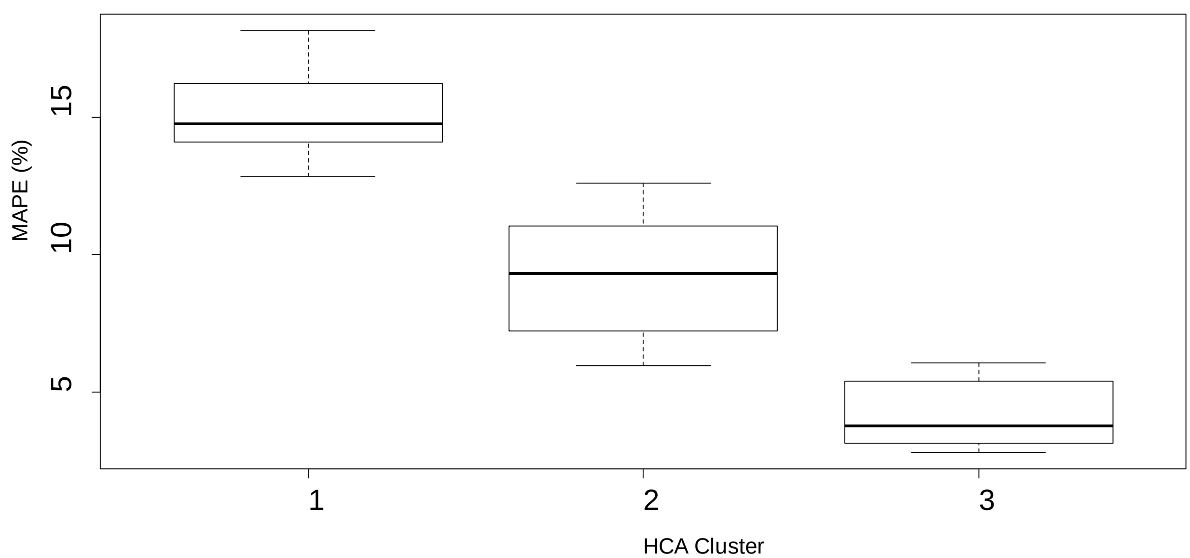


Figure 2-24 The HCA Cluster-wise Monthly Distribution of MAPE values

Statistical Test for Error Differences

The study conducts a one-tailed t-test to statistically test the error differences produced between the GMM solution and other clustering methods for its significance. The test uses a null hypothesis of zero error difference between the GMM and other clustering solutions. Table 2-16 lists the results of t-test between the clustering methods. The GMM produces statistically significant and less error compare to another clustering. An error reduction of 6% to 26% is observed.

Table 2-16 *t*-test Results between the GMM and other Clustering Methods

Comparison	*t* statistic	*p*-value	95 % CI of Mean Error Difference		Mean Error Difference	Percentage Error Reduction (%)
			lower	upper		
GMM-QDA vs HCA-QDA	-21.76	2.20E-16	-1.07	-0.90	-0.98	-8.5%
GMM-QDA vs KM-QDA	-16.17	2.20E-16	-0.75	-0.59	-0.67	-5.9%
GMM-QDA vs ODOT	-29.35	2.20E-16	-1.92	-1.68	-1.80	-15.3%
GMM-QDA vs FC	-53.40	2.20E-16	-3.17	-2.95	-3.06	-25.4%

SUMMARY

The study applies a GMM technique for obtaining clusters in a state-wide network traffic data. The study uses only 24 hourly traffic patterns of all vehicles in clustering process a general trend of summer peaks is observed among the clusters. The day-of-week patterns show the most of the clusters has stable weekday (Monday to Thursday) patterns and reduced traffic for the rest of the week. The study addresses, through robustness analysis, a question of producing same clustering solution for a new sample drawn from the same population from which the input data is generated. The study compares the performance of the GMM, KM, HCA, ODOT, and FC grouping using mean absolute percent error values. The GMM clustering solution produces less error values than the HCA and KM methods.

CONCLUSIONS

The GMM clustering solution shows low variance among clusters with a stable weekday and general summer peaks in the clusters. The clusters show trends of both commuter and recreational patterns and they typically reflect different geographical regions and highway functional classes of the ODOT. Often, clusters that exhibit non-commuter trends show more variability of cluster proportions, which makes them susceptible to label switching if the GMM algorithm is applied on resampled datasets.

The stability analysis addresses the likelihood of obtaining a similar clustering when drawing a new sample from the same population. The stability analysis using bootstrapping, replacing points by noise, jittering, and subsetting resampling schemes provides insights on cluster-wise stability assessment for the GMM, KM, and HCA methods. Both the k-means and HCA show stable clusters under all resampling methods. The stability analysis does not always reflect the validity of clusters. Nevertheless, they provide more information on clusters under a given modeling framework.

The missing data analysis shows that the missing data affects the GMM clustering solution. Even though the KM and HCA methods show stable clusters, the missing data affects their clustering solutions more significantly than the GMM solution.

The winter months (November to March) show more AADT estimation errors than the warmer months (April to October). The GMM solution shows less AADT estimation errors than the KM and HCA solutions. Although the HCA and KM solutions show stable clustering solutions, when tested for new data, they produce more errors. The GMM solution also performs better than the ODOT seasonal trend grouping. The Functional class grouping consistently produces higher error values. An error reduction of 6% to 26%, which is statistically significant at the 5 percent level, is observed for the GMM solution. The standard deviation of errors varies between 10 and 18 percent. The standard deviation of MAPEs is lower for the GMM solution compares to the HCA and KM solutions.

The GMM solution also produces lower DAPE values compared to the clustering solutions KM and HCA for all days of a week. The GMM solution reports the highest number of patterns with an error of less than five percent. The GMM solution again produces fewer patterns having a MAPE value of greater than 15 percent than all other clustering solutions.

REFERENCES

1. Gadda, S. C., A. Magoon, and K.M.Kockelman. Quantifying theuncertainty in annual average daily traffic (AADT) count estimates." *Proc., Eleventh World Conf. on Transportation Research*, 2007. World Conference on Transport Research Society, Lyon, France.
2. Duddu, V. R., and S. S. Pulugurtha. Principle of Demographic Gravitation to Estimate Annual Average Daily Traffic: Comparison of Statistical and Neural Network Models. Journal of Transportation Engineering, Vol. 139, No. 6, 2013, pp. 585–595.
3. Gastaldi, M., G. Gecchele, and R. Rossi. Estimation of Annual Average Daily Traffic from One-week Traffic Counts - A Combined ANN-Fuzzy Approach. Transportation Research: Part C, vol. 47, 2014, PP. 86-99.
4. Traffic Monitoring Guide. Federal Highway Administration (FHWA). U.S. Department of Transportation, 2013. http://www.fhwa.dot.gov/policyinformation/tmguide. Accessed July 7, 2015.
5. Fraley, C., and A. E. Raftery. How Many Clusters? Answers via Model-Based Cluster Analysis. The Computer Journal, vol. 41, 1998, pp. 578–588.
6. Everitt, B. S., Landau, Leese, M. and Stahl, D. (2011) Cluster Analysis. John Wiley and Sons, Ltd. United Kingdom.
7. Banfield, J.D. and Raftery, A.E. (1993). Model-based Gaussian and non-Gaussian clustering. Biometrics, vol. 49, pp: 803-821.

8. Celeux, G. and Govaert, G. (1995) Gaussian parsimonious clustering models. Pattern Recognition, vol. 28, pp: 781-793.
9. Fraley, C., and A. E. Raftery. How Many Clusters? Answers via Model-Based Cluster Analysis. Technical Report No. 329, Department of Statistics, University of Washington, Box 354322, Seattle, WA 98195-4322 USA.
10. Hennig, C., Meila, M., Murtagh, F. and Rocci, R. (2015) Handbook of Cluster Analysis. Chapman and Hall/CRC, Boca Raton, FL.
11. Milligan, G. W. and Cooper, M. C. (1985) An examination of procedures for determining the number of clusters in a data set. Psychometrika, vol. 50, pp: 159–179.
12. Calinski, R. B. and Harabasz, J. (1974) A dendrite method for cluster analysis. Communications in Statistics, vol. 3, pp: 1–27.
13. Duda, R. O. and Hart, P. E. (1973) Pattern Classification and Scene Analysis. John Wiley & Sons, Inc., New York.
14. Sharma, S. C., and A. Werner. Improved Method of Grouping Province Wide Permanent Traffic Counters. In Transport Research Record: Journal of the Transportation Research Board, No. 815, Transportation Research Board of the National Academies, Washington, D.C., 1981, pp.12–18.
15. Faghri, A., and J. Hua. Roadway Seasonal Classification using Neural Networks. Journal of Computing in Civil Engineering, Vol. 9, No. 3, 1995, pp. 209–215.
16. Flaherty, J. Cluster Analysis of Arizona Automatic Traffic Record Data. In Transport Research Record: Journal of the Transportation Research Board, No. 1410, Transportation Research Board of the National Academies, Washington, D.C., 1993, pp. 93–99.
17. Tsapakis, I., W. H. Schneider IV, and A. P. Nichols. Improving the Estimation of Total and Direction-based Heavy-Duty Vehicle Annual Average Daily Traffic, Transportation Planning and Technology, Vol. 34, Issue. 2, 2011, pp. 155-166.
18. Rossi, R., M. Gastaldi, and G. Gecchele. Comparison of Clustering Methods for Road Group Identification in FHWA Traffic Monitoring Approach: Effects on AADT Estimates. Journal of Transportation Engineering, Vol. 140, No. 7, 2014, 04014025.
19. Gecchele, G., S. Kikuchi, R. Rossi, and M. Gastaldi. Advances in Uncertainty Treatment in FHWA Procedure for Estimating Annual Average Daily Traffic Volume. Presented at 91st Annual Meeting of the Transportation Research Board, Washington, D.C., 2012.
20. Rossi, R., M. Gastaldi, G. Gecchele, and S. Kikuchi. Estimation of Annual Average Daily Truck Traffic Volume. Uncertainty Treatment and Data Collection Requirements. Procedia Social and Behavioral Sciences, vol. 54, 2012, pp. 845–856.
21. Gastaldi, M., R. Rossi, G. Gecchele, and L. Della Lucia. Annual Average Daily Traffic Estimation from Seasonal Traffic Counts. Procedia Social and Behavioral Sciences, vol. 87, 2013, pp. 279–291.

22. Faghri, A., and P. Chakroborty. Development and Evaluation of a Statistically Reliable Traffic Counting Program, Transportation Planning and Technology, Vol. 18, No. 3, 1994, pp. 223-237.

23. Davis, G. A. Accuracy of Estimates of Mean Daily Traffic: A Review." In Transport Research Record: Journal of the Transportation Research Board, No. 1593, Transportation Research Board of the National Academies, Washington, D.C., 1997, pp. 12–16.

24. Chen, P., K. Yin, and J. Sun. Application of Finite Mixture of Regression Model with Varying Mixing Probabilities to Estimation of Urban Arterial Travel Times. In Transportation Research Record: Journal of the Transportation Research Board, No. 2442, Transportation Research Board of the National Academies, Washington, D.C., 2014, pp. 96–105.

25. Lingras, P. Traffic Parameter Estimation and Highway Classification: Rough Patterns Using A Neural Networks Approach, Transportation Planning and Technology, Vol. 21, No. 3, 1998, pp. 155-179.

26. Lingras, P. Statistical and Genetic Algorithms Classification of Highways, Journal of Transportation Engineering, Vol. 127, No. 3, 2001, pp. 287-243.

27. Venkatanarayana, R., B. L. Smith, and M. J. Demetsky. Quantum-Frequency Algorithm for Automated Identification of Traffic Patterns. In Transportation Research Record: Journal of the Transportation Research Board, No. 2024, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 8–17.

28. Kirshfink, H., and C. Chadenas. Traffic Situation Prediction Applying Pattern Matching and Fuzzy Classification. European Symposium on Intelligent Techniques, Orthodox Academy of Crete, Greece, June 3–4, 1999.

29. Ren, J. T., X. L. Ou, Y. Zhang, and D. C. Hu. Research on Network-Level Traffic Pattern Recognition. IEEE 5th International Conference on Intelligent Transportation Systems, Singapore, Sept. 3–6, 2002.

30. Venkatanarayana, R., B. L. Smith, and M. J. Demetsky. Traffic Pattern Identification Using Wavelet Transforms. Presented at 84th Annual Meeting of the Transportation Research Board, Washington, D.C., 2005.

31. Ritchie, S. A Statistical Approach to Statewide Traffic Counting. In Transportation Research Record: Journal of the Transportation Research Board, No. 1090, Transportation Research Board of the National Academies, Washington, D.C., 1986, pp. 14–21.

32. Federal Highway Admistration. Highway Functional Classification: Concepts, Criteria and Procedures. 2013 Edition. report no: FHWA-PL-13-026 https://www.oregon.gov/ODOT/TD/TDATA/rics/docs/HighwayFCConceptsCriteriaProcedures.pdf retrieved on August 12, 2016.

33. Hennig, c. (2007) Cluster-wise assessment of cluster stability. Computational Statistics & Data Analysis. Vol. 52, pp: 258 – 271.

34. Hennig, C. (2008) Dissolution point and isolation robustness: robustness criteria for general cluster analysis methods. Journal of Multivariate Analysis, Vol. 99, pp: 1154-1176.

35. Tsapakis, I., W.H.Schneider, A.P.Nichols, and J. Haworth. Alternatives in assigning short -term counts to seasonal adjustment factor groupings. Journal of Advanced Transportation, Vol. 48, 2014, pp. 417-430.

36. Japkowicz, N. and Shah, M. (2011). Evaluating Learning Algorithms: A Classification Perspective. Cambridge University Press, New York, USA.

37. Tsapakis, I., W. H. Schneider IV, and A. P. Nichols. Improving the Estimation of Total and Direction-based Heavy-Duty Vehicle Annual Average Daily Traffic, Transportation Planning and Technology, Vol. 34, Issue. 2, 2011, pp. 155-166.

38. Rossi, R., M. Gastaldi, and G. Gecchele. Comparison of Clustering Methods for Road Group Identification in FHWA Traffic Monitoring Approach: Effects on AADT Estimates. Journal of Transportation Engineering, Vol. 140, No. 7, 2014, 04014025.

39. Chen, P., K. Yin, and J. Sun. Application of Finite Mixture of Regression Model with Varying Mixing Probabilities to Estimation of Urban Arterial Travel Times. Transportation Research Record: Journal of the Transportation Research Board, No. 2442, Transportation Research Board of the National Academies, Washington, D.C., 2014, pp. 96–105.

Chapter 3 Classification

INTRODUCTION

Transportation agencies need a reliable estimate of traffic volumes for planning, designing, operating and maintaining highway infrastructure. The American Association of State Highway and Transportation Officials (AASHTO) *Guidelines for Traffic Data Programs* identifies key areas of traffic data use in safety analysis, air quality, capacity analysis, pavement design, operational analysis and project evaluation and selection (*1*). Traffic data such as Annual Average Daily Traffic (AADT), Vehicle Miles Traveled (VMT) and Design Hourly Volume (DHV) represent some of the key inputs that agencies use most in their projects (*2*).

The agencies allocate significant resources to collect the traffic data on their networks using Automatic Traffic Recorders (ATRs). However, the cost and maintenance restrict the ATR deployment to few strategic highway network locations. The ATRs collect the traffic data for a section of highway continuously for all 365 days of the year. The ATR collects and stores the data in 15-min intervals with an option of recording vehicle classes and weights. The agencies study the temporal variation of traffic, like month-of-year, day-of-week or hour-of-day patterns, using the ATR counts. Moreover, the Federal Highway Administration (FHWA) *Traffic Monitoring Guide* (TMG) insists that monitoring traffic volume trends represents a key task for the continuous traffic count program (*3*). Both continuous count data and short-term counts form two key components of a successful Traffic Monitoring Program (TMP) (*3*).

In lieu of ATRs, the agencies may use Short-term Traffic Counts (STTCs). The STTCs can be cost effective and deployed almost everywhere on the network to comprehensively study the traffic data. Typically, an agency may collect STTCs on a road segment every few years and the collection periods vary from one to seven days (*2*). However, the recommended minimum counting period is 48 hours for rural roads and 24 hours for urban roads (*2*).

Seasonal Adjustment Factors (SAFs or expansion factors) convert the STTCs to AADT estimates. This necessitates a method to identify an ATR or groups of ATRs that exhibit similar seasonal characteristics so that the appropriate factors may be applied to transform the short-term counts to an AADT value. Most transportation agencies use a traditional FHWA factoring method. The traditional method, introduced by Drusch (*4*), has four steps. The first step, called factoring, computes a seasonal adjustment for each ATR Then, the grouping step combines the ATRs based on highway functional class or the geographic region. The short-term counts may be assigned to one of the groups based on functional class, spatial characteristics or both. The final step estimates the AADT using counts and corresponding seasonal adjustment factors.

Although the traditional method remains a widely used method among many Department of Transportation (DOTs), the assignment based on functional class can affect the accuracy of the estimates. Because any two patterns belong to same functional class cannot guarantee that they exhibit similar trends. This chapter presents an improved classification methods based on neural networks. The authors compare the proposed method with the performance of discriminant analysis and traditional methods.

Previous Work

The assignment step in the traditional method plays a critical role in the AADT estimation process and earlier studies observed large errors for incorrect assignment of traffic patterns (*5, 6*). The AADT estimation errors can be minimized if the assignment step minimizes the risk of misclassification (*7, 8*). Davis and Guan (*7*) use Bayesian statistics to assign short-term counts to factor groups and the study reports an average error of within ±5% to ±20% for a state-wide network. The FHWA TMG suggests collecting multiple short-term counts at different periods to minimize the estimation errors (*3*). Zhong *et al*. (*9*) use historical seasonal patterns and Bayesian statistics to improve the group assignment and AADT estimation errors. The study obtains less than 12% error (95th percentile) using the new approach.

Sharma and Werner (*10*) use monthly traffic patterns to classify traffic count sites based on the hierarchical grouping. Sharma and Allipuram (*11*) refine the hierarchical grouping approach for highway classification further. Lingras (*12*) uses neural networks to obtain classification similar to hierarchical grouping. Lingras (*13*) compares a conventional statistical method and a genetic algorithm approach for classifying temporal traffic patterns.

Tsapakis *et al*. (*14*) use Linear Discriminant Analysis (LDA) to assign 24-hour short-term counts; they observe an average reduction of 58% in mean absolute percentage error for LDA over the traditional functional group classification. A later study extends this methodology using a new statistical approach based on a weighted coefficient of variation (WCV) (*15*). The comparison between the WCV, LDA and functional group classification shows that the WCV method reduces the mean absolute percentage error by 58% and standard deviation of error by 70% for direction traffic. Lingras (*16*) uses neural networks based on rough patterns for highway classification and obtains 10% (95^{th} percentile error) AADT estimation errors using two days of data collection in both July and December.

Caceres *et al.* (*17*) use attractiveness of a given road section when associating patterns of road groups. Gecchele *et al.* (*18*) measure uncertainty when assigning the groups using fuzzy and neural network framework. Li *et al.* (*19*) use a fuzzy based decision tree to assign short-term counts. Jin *et al*. (*20*) and

Lam *et al.* (*21*) use a k-nearest neighbor algorithm (k-NNC) to classify roadways. Tsapakis and Schneider (*22*) adopt Support Vector Machines (SVM) to assign short-term counts to seasonal factor groups. The SVM models improve the AADT accuracy by 65% and decrease the error standard deviation by 73.7% when compared to the traditional method.

Even though different approaches improve the AADT estimates, with the rapid adoption of machine learning algorithms in other fields for pattern recognition and classification problems, this study investigates the performance of these innovative methods and assess their improvement in the assignment of traffic patterns to correct groups. The study starts by developing a modified neural network structure and training algorithms for the classification problem. Neural networks, due to their adaptability and ability to map non-linear relationships, applied in many studies analyzing different kinds of problems. Neural networks with backpropagation algorithms are extensively used in many areas of transportation; however, this study expects that a few changes to the neural network structure and adaptation of enhanced training learning can improve the accuracy of the AADT estimates. In addition, the study also introduces the Quadratic Discriminant Analysis and Naïve Bayes classification and compares their performance with traditional methods.

Objectives

The study defines following objectives for analysis:

- Building a better classifier for assigning traffic patterns based on a clustering solution
- Introducing a modified neural net structure and enhanced network learning algorithms
- Conducting a thorough evaluation of classifiers using different performance measures

Chapter Organization

The remainder of this chapter starts with the presentation of the modified neural network structure, network selection, and learning process. The study evaluates two variants of network algorithms for training and testing errors. Next, the paper presents the typical classifier performance metrics and uses them in evaluating different classifiers. Finally, the study presents the accuracy and estimation errors between the classifiers and statistically tests the performance difference between them.

CONTRIBUTION

The study proposes changes to the neural network structure and learning algorithms. The study introduces two improved algorithms, the OWO-BP and OWO-Newton (ONN) methods, and obtains better performance than the regular BP algorithm. The author also introduces a fully connected network and a non-heuristic optimal learning factor to improve the network performance by optimally adjusting the learning factors between iterations. Moreover, the network selection process helps to find the best value for the number of hidden units in a hidden layer as opposed to the typical heuristic approach of selecting units. The study tests and evaluates the developed ONN framework using multiple performance measures. In particular, the study uses the AUC measure of a multi-class problem obtained using pairwise computations. In addition, the study introduces a 10×10 stratified cross validation approach for performance evaluation of classifiers. The ROC analysis presented in this study can help to study the behavior of a classifier, aid comparative analysis, and model selection process. In addition, the study presents the classification performance of each cluster for the GMM solution. The error analysis across months, highway functional class, and by error size provides more insight than presenting just average error values.

METHODOLOGY

The assignment methodology involves five stages. In the first stage, the 24-hour traffic patterns are grouped based on similarity using Gaussian Mixture Modeling (GMM). The GMM clusters' stability, sensitiveness to missing data, and relative performance is evaluated in chapter 2. Chapter 2 concludes that the GMM performs better compare to other clustering solutions. Hence, this chapter uses the GMM clustering solution to compute the Seasonal Adjustment Factors (SAFs) in the second stage. The SAFs are later used to estimate the AADT values for test patterns. The third stage presents the modified neural network models and studies the network performance for classifying the GMM clusters. Next stage assesses the discriminant analysis and Naïve Bayes classifiers and comparing them with the neural network using a variety of performance metrics. Final stage deploys trained classifiers on test datasets to assign a cluster label and subsequently estimate the AADT values. The error analysis explores the variation of the AADT estimates by month, day-of-week and highway functional class.

NEURAL NETWORK MODELS

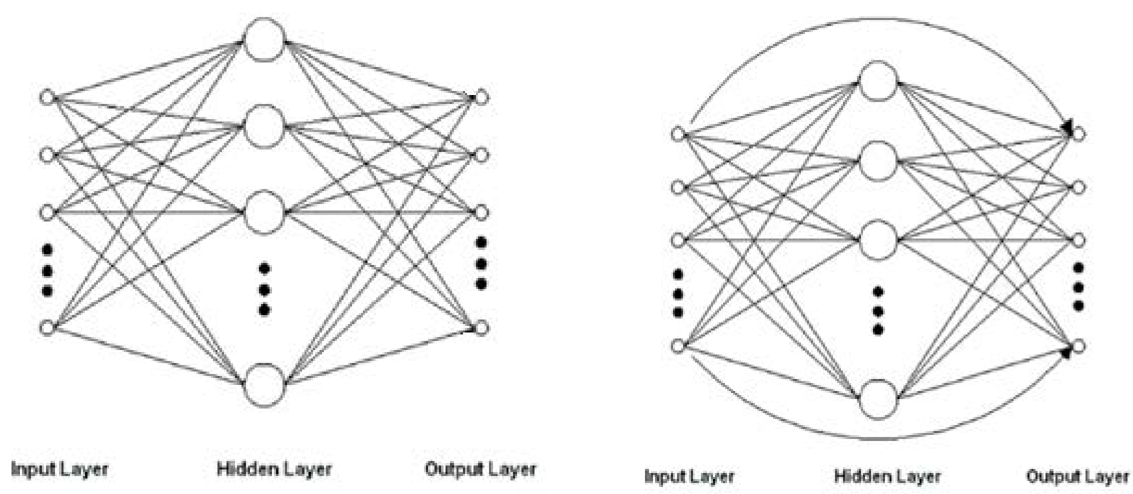
A neural network made up of artificial neurons creates a framework designed to mimic the brain performance of a particular task (*23*). The Artificial Neural Networks (ANN) performs necessary computations through a process of learning. Neural networks have interconnected units often called neurons or processing units. Through the learning algorithm, the ANN deploys a learning mechanism to map the input and target outputs with an objective of minimizing the error. In achieving the desired objectives, the network modifies or "trains" the connection weights in a systematic manner. Numerous fields apply neural networks due to their ability to perform non-linear signal processing, input-output mapping, adaptively and closely mimicking an optimal process (*23*).

Multiple Layer Perceptron (MLP)

A MLP is a neural network with one or more hidden layers. A single-layer neural network, as also called a perceptron, remains limited to classifying linearly separable patterns, but the MLP extends the neural network capabilities to map even non-linear relations. Haykin (*23*) highlights the main features of a MLP as a differentiable nonlinear activation function, one or more hidden layers with input and output layers, and a high degree of connectivity. The structure of an MLP, usually, has one input layer, one hidden layer, and an output layer. However, this study adopts some changes to the network structure, learning process and learning algorithms.

Connectivity

The signals that come from the input layer propagate forward through the network to the output units. The connections between various units of different layers carry certain weights (also called a connection strength or synaptic weight). Most studies in the transportation literature (*24, 25, 18*) consider cascade connectivity where every unit connects only to units in the previous layer (Figure 3-1 (a)). However, this study adopts a fully connected network (Figure 3-1(b)). When mapping an input towards the target output, direct connections between both the input and output, which are not present in cascade connectivity, appear intuitive. These additional input to output connections increase the computational burden due to the additional weights but they also improve the performance and inference from the neural network models.



(a) Cascade Connectivity (b) Full Connectivity

Figure 3-1 MLP Network Connectivity (source: Malalur and Manry (*26*))

Learning Factors

When adjusting weights iteratively using steepest descent, the learning rate or learning factor controls the changes to the weights and convergence of the learning algorithm. If the learning rate remains too small, the weights' change is small. The learning rate can increase by choosing a large value for the learning rate, but this may result in large changes that often make the network unstable or oscillatory. Typically, most neural networks use either a fixed constant or heuristically scaling approach for learning factors. Fixed learning factors potentially contribute to slow convergence while a heuristically scaling approach between iterations increases the convergence rate. However, this study adopts a non-heuristic optimal learning factor (OLF). The OLFs can optimally adjust the learning factors between iteration so that error can be minimized.

Learning Algorithms

Back-propagation (BP) is a popular and efficient method of training MLP networks. The algorithm uses a partial derivative of the error function with respect to the connection weights determined using a back propagation of the error through the network. The BP uses a first order error gradient vector while training. The OWO-BP (output weights optimization-BP) method develops an improved first order algorithm and has shown better performance than the regular BP algorithm for five different datasets (*26*). The first order algorithms need fewer multiplications, data passes and less execution time per iteration (*26*). However, the BP and OWO-BP remain sensitive to the input patterns' means and gains (*27*). With the advent of machine learning algorithms, second order training algorithms are proposed and applied to multiple studies in various fields. The second order algorithms provide better performance than BP in multiple comparative studies (*28, 29*). With the lack of applications of second-order training algorithms in the transportation engineering field, the study examines their performance for the group assignment of traffic patterns. In particular, the study adopts two improved algorithms, OWO-BP (first order) and OWO-Newton (second order) methods (*28, 29*), for classification and compare them with BP.

TRAINING ALGORITHMS

A fully connected neural network with a single hidden layer is shown in Figure 3-2. The training data of N_v training patterns consists of N-dimensional input vectors $\boldsymbol{x}_p$ and M-dimensional desired output vectors $\boldsymbol{t}_p$, where p represents one of the traffic patterns. An extra input 1 is added to handle the bias or threshold in the hidden and output layer.

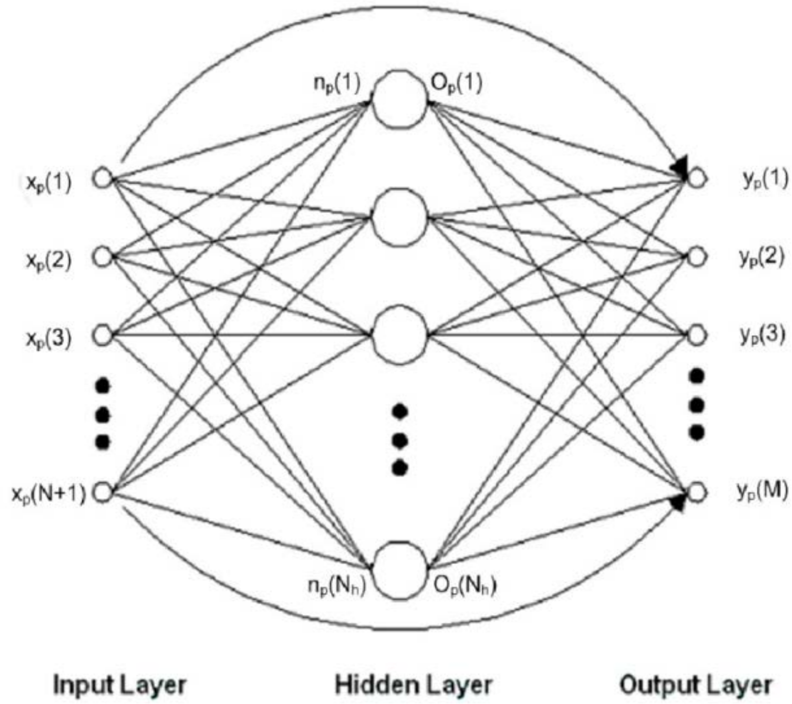


Figure 3-2 Fully Connected MLP Neural Network (source: Malalur and Manry (*26*))

MLP notation

- $\mathbf{x}$ – Input vector
- $\mathbf{x}_p$ – Input vector for p^{th} pattern
- p – Traffic pattern number
- $x_p(n)$ – n^{th} element of $\mathbf{x}_p$
- $\mathbf{t}_p$ – desired output vector for p^{th} pattern
- $t_p(i)$ – i^{th} element of $\mathbf{t}_p$
- $\mathbf{y}_p$ – Actual output vector when $\mathbf{x} = \mathbf{x}_p$
- $y_p(i)$ – i^{th} element of $\mathbf{y}_p$
- N – Number of network inputs, dim($\mathbf{x}$)
- M – Number of network outputs=dim($\mathbf{y}$) = dim($\mathbf{t}$)
- N_V – Number of patterns

$w(k,n)$ are the input weights that connect the n^{th} input to the k^{th} hidden unit. Output weights $w_{oh}(i,k)$ connect the k^{th} hidden unit's activation ($O_p(k)$) to the i^{th} output $y_p(i)$. The study assumes that the hidden unit has a sigmoidal activation function and the output unit has a linear activation. The bypass weight $w_{oi}(i,n)$ connects the n^{th} input *directly* to the i^{th} output. N_h denotes the number of hidden units in the network. The input weights **W** is N_h by N+1, and output weights $\mathbf{W}_{oh}$ and $\mathbf{W}_{oi}$ are M by N_h and M by N+1 respectively.

The training data {x_p, t_p} consists of both inputs and target outputs. The extra input given to the input data is denoted by $x_p(N+1)$ takes a value of 1, so the complete input data for the pth pattern is given by $\boldsymbol{X}_p = \left[x_p(1), x_p(2), \cdots, x_p(N+1)\right]^T$. For the pth pattern, the kth hidden unit's net function is

$$n_p(k) = \sum_{n=1}^{N+1} w(k,n) \cdot x_p(n) \tag{16}$$

And can be summarized as $\boldsymbol{n}_p = \boldsymbol{W}\,\boldsymbol{X}_p$. The kth hidden unit's activation output is denoted as $O_p(k)$ and can be expressed as

$$O_p(k) = f(n_p(k)) = \frac{1}{1 + e^{-n_p(k)}} \tag{17}$$

For the pth pattern, the ith element of the M-dimensional output vector $\mathbf{y}_p$ is

$$y_p(i) = \sum_{n=1}^{N+1} w_{oi}(i,n) \cdot x_p(n) + \sum_{k=1}^{N_h} w_{oh}(i,k) \cdot O_p(k) \tag{18}$$

or can be summarized as

$$\boldsymbol{y}_p = \boldsymbol{W}_{oi}\,\boldsymbol{X}_p + \boldsymbol{W}_{oh}\,\boldsymbol{O}_p \tag{19}$$

Back Propagation (*30, 31*)

The typical error function used in training the MLP is the mean-squared error (MSE) that is described as

$$E = \frac{1}{N_v} \sum_{p=1}^{N_v} \sum_{i=1}^{M} \left[t_p(i) - y_p(i)\right]^2 \tag{20}$$

For the pth pattern, output and hidden layer delta functions (*31*) are respectively found as

$$\delta_{po}(i) = 2\big(t_p(i) - y_p(i)\big) \tag{21}$$

$$\delta_p(k) = f'\big(n_p(k)\big) \sum_{i=1}^{M} \delta_{po}(i).\, w_{oh}(i,k) \tag{22}$$

Now, the negative input weight gradient of E (Jacobian matrix) is

$$g(k,n) = \frac{-\partial E}{\partial w(k,n)} = \frac{1}{N_v}\sum_{p=1}^{N_v} \delta_p(k).\, x_p(n) \quad (23)$$

The matrix of negative partial derivatives can be written as

$$\boldsymbol{G} = \frac{1}{N_v}\sum_{p=1}^{N_v} \boldsymbol{\delta_p}\left(\boldsymbol{X_p}\right)^{\boldsymbol{T}} \quad (24)$$

Where $\delta = \left[\delta_p(1), \cdots, \delta(N_h)\right]^T$. If the steepest descent method is used to modify the input weights, then **W** in a given iteration is updated using

$$\boldsymbol{W} \leftarrow (\boldsymbol{W} + z \cdot \boldsymbol{G}) \; or \; \Delta\boldsymbol{W} = z \cdot \boldsymbol{G} \quad (25)$$

Where *z* is the learning factor. The details on calculating learning factors are given in (*32*). Like the input weights, the negative output weight gradients G_{oh} and G_{oi} are also found using BP algorithm and accordingly the output weights $\mathbf{W}_{oh}$ and $\mathbf{W}_{oi}$ are updated. Both the input and output weights are updated using the OLF and three directional vectors g(*k*,*n*), $g_{oi}(i,n)$ and $g_{oh}(i,k)$. Backpropagation training algorithm is a first order training algorithm. Second order algorithms, presented in the following sections, perform well compared to the BP algorithm (*26, 28, 29*).

OWO-BP

The Output Weight Optimization – Back Propagation (OWO-BP) finds the output weights $\mathbf{W}_{oh}$ and $\mathbf{W}_{oi}$ using optimization and trains the input weights **W** using a BP algorithm. The OWO technique finds the weights connected to the output units by solving a system of linear equations. Equation (19) can be re-written as $\boldsymbol{y}_p = \boldsymbol{W}_o\, \overline{\boldsymbol{X}}_p$. Where $\overline{\boldsymbol{X}}_p = \left[\boldsymbol{X}_p^T, \boldsymbol{O}_p^T\right]$ is the augmented input vectors that connect to the output units and $\boldsymbol{W}_o = [\boldsymbol{W}_{oi}\colon \boldsymbol{W}_{oh}]$ denotes all the weights connected to the outputs. $\overline{\boldsymbol{X}}_p$ is a column vector of size N_u=N+N_h+1 and $\boldsymbol{W}_o$ is M by N_u. By setting $\frac{\partial E}{\partial \boldsymbol{W}_o} = 0$, the output weights can be solved using orthogonal least squares (OLS) for a set of linear equations given by

$$\boldsymbol{C} = \boldsymbol{W}_o \cdot \boldsymbol{R}^T \quad (26)$$

Where $\boldsymbol{C} = \frac{1}{N_v}\sum_{p=1}^{N_v} \boldsymbol{y}_p\, \overline{\boldsymbol{X}}_p^{\;T}$ and $\boldsymbol{R} = \frac{1}{N_v}\sum_{p=1}^{N_v} \overline{\boldsymbol{X}}_p\, \overline{\boldsymbol{X}}_p^{\;T}$

As before, equations (24) and (25) are used to compute the input weights.

Optimal Leaning Factors (*33*)

The learning factor *z* has a direct effect on the convergence of OWO-BP algorithm (*34*). Usually, either a fixed constant or heuristically scaling approach is used for learning factors. However, a non-heuristic optimal learning factor (OLF) for OWO-BP can be derived using Taylor's series for error. The OLFs can optimally adjust the learning factors between iteration so that error can be minimized.

The Taylor series expansion of error is written as:

$$E(z) = E(0) + z \cdot \frac{-\partial E}{\partial z} + \frac{1}{2}\frac{\partial^2 E}{\partial z^2}\, z^2 \tag{27}$$

To get an optimal learning factor z,

$$\frac{\partial E(z)}{\partial z} = 0 = \frac{\partial E}{\partial z} + \frac{\partial^2 E}{\partial z^2}\, z \;\text{ or }\; z^* = \frac{-\partial E/\partial z}{\partial^2 E/\partial z^2} \tag{28}$$

Rewriting the output at the i^{th} output unit for the p^{th} pattern in terms of learning factor z and direction vector d(k,n)

$$y_p(i) = \sum_{n=1}^{N+1} w_{oi}(i,n)\cdot x_p(n) + \sum_{k=1}^{N_h} w_{oh}(i,k)\cdot\left\{f\Big(\sum_{n=1}^{N+1}(w(k,n) + z\cdot d(k,n))\cdot x_p(n)\Big)\right\} \tag{29}$$

And

$$\frac{\partial E}{\partial z} = \frac{-2}{N_v}\sum_{p=1}^{N_v}\sum_{i=1}^{M}\left[t_p(i) - y_p(i)\right]\cdot\frac{\partial y_p(i)}{\partial z} \tag{30}$$

Where

$$\frac{\partial y_p(i)}{\partial z} = \sum_{k=1}^{N_h} w_{oh}(i,k)\cdot f'(n_p(k))\cdot\sum_{n=1}^{N+1} d(k,n)\cdot x_p(n) \tag{31}$$

The gauss-newton approximation of the second derivative of error with respect to z is:

$$\frac{\partial^2 E}{\partial z^2} = \frac{2}{N_v}\sum_{p=1}^{N_v}\sum_{i=1}^{M}\left[\frac{\partial y_p(i)}{\partial z}\right]^2 \tag{32}$$

The optimal z obtain from equation (28) is used when training the inputs.

OWO-Newton

The output weight optimization – Newton's method finds the output weights $\mathbf{W}_{oh}$ and $\mathbf{W}_{oi}$ using optimization and trains the input weights **W** using Newton's method (*35*). The OWO technique solves for the outputs weights as described in the previous section.

Assume that the vector **W** of dimension N_w stores the input weights during network training. Let **g** be the negative gradient vector (negative Jacobian) for the training error E. The Hessian matrix **H** represents the second derivative of the training error. An element of a negative Jacobian matrix g(k,n) gives the first partial of E with respect to input weights w(k,n)

$$\frac{\partial E}{\partial w(k,n)} = \frac{-2}{N_v} \sum_{p=1}^{N_v} \sum_{i=1}^{M} \left[t_p(i) - y_p(i)\right] \cdot \frac{\partial y_p(i)}{\partial w(k,n)} \tag{33}$$

Where $\frac{\partial y_p(i)}{\partial w(k,n)} = w_{oh}(i,k) \cdot f'(n_p(k)) \cdot x_p(n)$, and an element of the Hessian matrix is given by

$$\frac{\partial^2 E}{\partial w(k,n)\, \partial w(u,v)} = \frac{2}{N_v} \sum_{p=1}^{N_v} \sum_{i=1}^{M} \frac{\partial y_p(i)}{\partial w(k,n)} \cdot \frac{\partial y_p(i)}{\partial w(u,v)} \tag{34}$$

Quadratic Approximation of Error

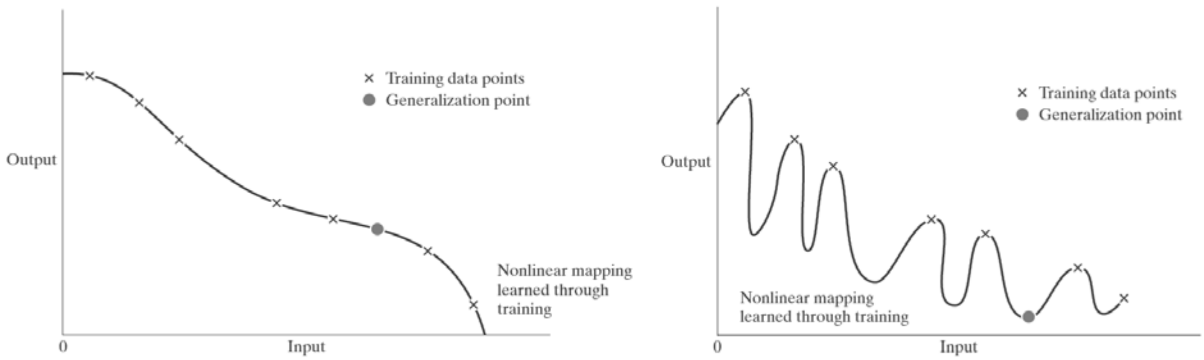
Consider the weight change vector (**e)**, which is computed as **W – W'**, where **W'** is the new version of the weight vector **W**. Using a multivariate Taylor's theorem, the error can be approximated as

$$E' \approx E - \boldsymbol{e}^T.\boldsymbol{g} + 1/2\ \ \boldsymbol{e}^T \cdot \boldsymbol{H}.e \tag{35}$$

By setting ${\partial E'}/{\partial e} = 0 = -\boldsymbol{g} + \boldsymbol{H} \cdot \boldsymbol{e}$, the weight change vector is obtained using $\boldsymbol{H}.\boldsymbol{e} = \boldsymbol{g}\ \ \boldsymbol{or}\ \boldsymbol{e} = \boldsymbol{H}^{-1} \cdot \boldsymbol{g}$, which can be solved using the OLS. The weights are updated using $\boldsymbol{W}' = \boldsymbol{W} + z \cdot \boldsymbol{e}$ where z is the learning factor.

Generalization

A neural network achieves generalization when a given neural network maps (almost) correctly the input and output for a test data set (drawn from the same population as the training data set) that has never used in training and validation of the network. Training (or learning) of the neural network can be treated as a curve fitting problem between the input and output. The input-output mapping represented by a smooth curve in Figure 3-3(a) shows a well-generalized network. If a test input (shown as a dot) other than one used in training (showed as an ×) is given, a well-generalized network could interpolate the test pattern. If the network is over trained, it memorizes the input and output pattern, which leads to overfitting or overtraining. An over fitted network loses the ability to generalize and produces more error when given new testing patterns (see Figure 3-3 (b)).



(a) Well Generalized Network Mapping (b) Poorly Generalized Network Mapping

Figure 3-3 Examples of Generalization in Neural Networks (Source: Haykin (*23*))

As Haykin(*23*) points out , three factors influence the generalization of a neural network: the size of the training sample, the architecture of the neural network and complexity of the problem. The study assumes that size of the sample remains fixed due to data availability limitations. As mentioned earlier, the study uses only 2011 data for training and testing the neural network performance. However, nothing precludes obtaining additional data as a future investigation. The team has no control over the complexity of the classification problem at hand; therefore, the determination of the *best architecture* for the neural network becomes the issue of interest.

The main objective of learning in a neural network is to transform the input-output relationship into synaptic weights and thresholds of an MLP so that network becomes well trained to generalize the patterns. Hence, the training should emphasize the network parameterization for a given set of data. Network selection, a part of learning process, deals with choosing the best parameters according to a certain criterion within candidate model structures for a given data set. One of the important parameters of the network is the choice of a number of units in the hidden layer (N_h) or a number of network coefficients (or weights) (N_w). Given a fixed input size (24-hour data) and target outputs (15 cluster GMM solution), the problem of choosing the best N_h and N_w essentially becomes the same. Hence, the network selection process deals with finding the best value for N_h.

Cross-validation provides a framework to aid the network selection process. In cross-validation, the data set is divided into training and testing samples usually using a split of 80 and 20 percent (*36*). However, the study conservatively adopts a 70/30 (training/testing) rule to keep the factor of misclassification error permitted on the test data around ten percent. The study uses a *stratified* cross-validation scheme so that both training and testing samples carry the same proportion of the cluster labels as the original data set. The study varies the number of hidden units in the hidden layer in increments of ten, starting from ten units increasing to 100 units. For each N_h value, both training error, $E_t(N_h)$ and testing error $E_v(N_h)$ are calculated respectively. When the training data $E_t(N_h)$ is plotted against the N_h values, a monotonically decreasing curve for an increase in N_h value is usually, but not necessarily, observed. However, the testing error curve decreases monotonically to a minimum value and starts to increase as the training continues (see Figure 3-4).

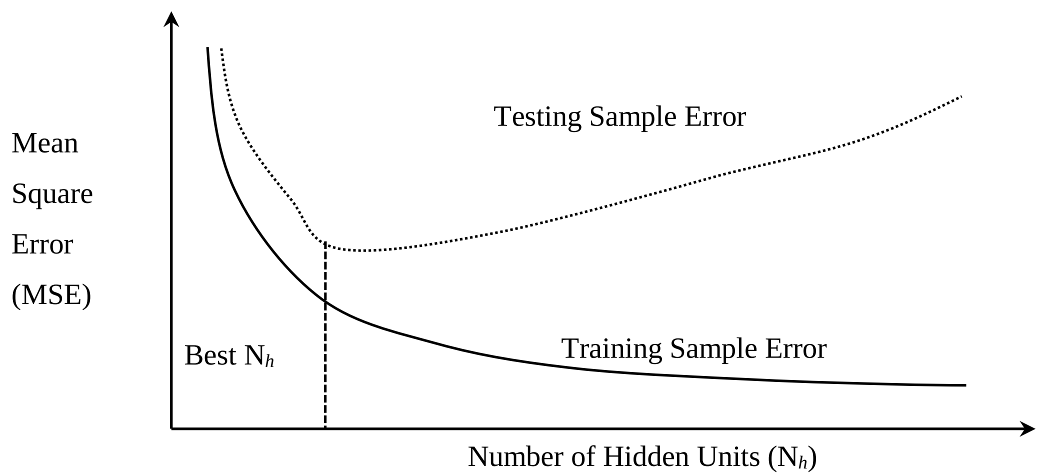


Figure 3-4 Network Selection Process to Find Best N_h Value

As mentioned in Haykin (*23*), the testing sample error curve may not evolve as smoothly as shown in Figure 3-4. The test error curve may show a few local minima before exhibiting an increase in error after a certain number of hidden units. In the presence of two or more local minima, the selection of more hidden units is preferred for a small improvement in the generalization performance (*37*).

Training Neural Networks

The MLP network structure is used to train the neural network classifier for classification of the fifteen cluster solution provided by the GMM. The study compares the performance of the OWO-Newton method to the OWO-BP, BP and CG (Conjugate Gradient) training algorithms. Each algorithm is trained on a network with one input layer, one hidden layer, and one output layer. The input layer has 24 units (representing the 24-hour traffic pattern) and an output layer with 15 units (for 15 clusters given by the GMM clustering). The study adopts the one-against-all (OAA) classification scheme when training the network. The study, using network selection guidelines, decides the best number of units in the hidden layer.

During training, the inputs values are preprocessed so that mean of the input vector is close to either zero or small when compared to the standard deviation (*38*). Inputs with larger standard deviation than others can dominate the training even if they are relatively not useful. In addition, initial synaptic weights and threshold will be a key to successful network design. It is desirable to have initial synaptic weights drawn from a uniform distribution with zero mean and variance equal to the reciprocal of the number of synaptic connections of a neuron (*23*). The training of the input weights strongly depends on the slopes of the activation function (sigmoid function). The Net control process adjusts the mean and standard deviations of all hidden units so that hidden units have desired mean and standard deviation (usually mean of 0.5 and standard deviation of one is adopted). Net control helps to prevent the loss of training because of zero activation function derivative due to input patterns (*39*). The study adopts a non-heuristic optimum learning factor for learning, a sigmoid activation function at hidden units and a linear activation function at output units. The study trains the network using 100 epochs.

Analysis

The number of units in the hidden layer varies by the training algorithm. The study iteratively trains the input-output relationships using training data and tests the network on the rest of the data. At each step of the N_h increments, both training and testing errors are calculated. Figure 3-5 shows the training and testing mean square error (MSE) by a number of hidden units for the OWO-BP algorithm. The testing error decreases with N_h (with almost a flat region between 40 and 50 units) and increases after a value of 80 units. The N_h value of 80 can be considered as the best value for MLP training using the OWO-BP algorithm. Likewise, the study obtains the best N_h value of 40 for BP, 50 for OWO-Newton and 60 for CG algorithm.

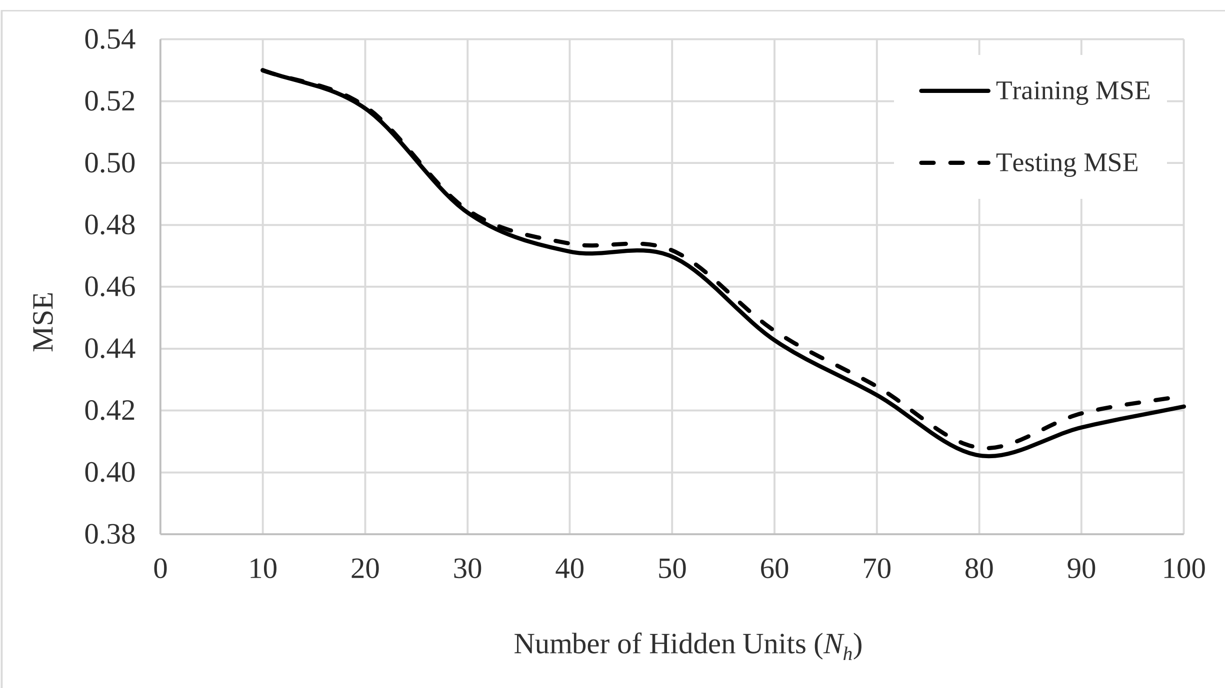


Figure 3-5 Selecting Best N_h value for the OWO-BP Training Algorithm

Once the network selection is completed, the training algorithm is applied to each network to obtain training and testing MSEs (see Table 3-1). The OWO-Newton produces less error compared to the others. The BP even performs more poorly than the CG method. The OWO-BP shows improved performance over the BP method. Figure 3-6 shows plots between the training MSE versus a number of iterations. The training error for CG and BP appears relatively flat. These methods experience a very slow rate of error decay because of heuristic learning rates. For instance, the BP algorithm produces an initial error of 0.714985, and by the end of 100 iterations error, it only reduces to 0.714978 (only a difference of 0.7E-06). Similarly, an error difference of 3.6E-05 is reported for the CG training algorithm.

Table 3-1 Performance of Neural Network Training Algorithms

Training Algorithm	Input Units	Hidden Units	Output Units	Training MSE	Testing MSE	Average Cost
CG	24	60	15	0.618	0.620	27.82
BP	24	40	15	0.715	0.722	1.80
OWO-BP	24	80	15	0.406	0.408	0.95
OWO-Newton	24	50	15	0.252	0.260	0.89

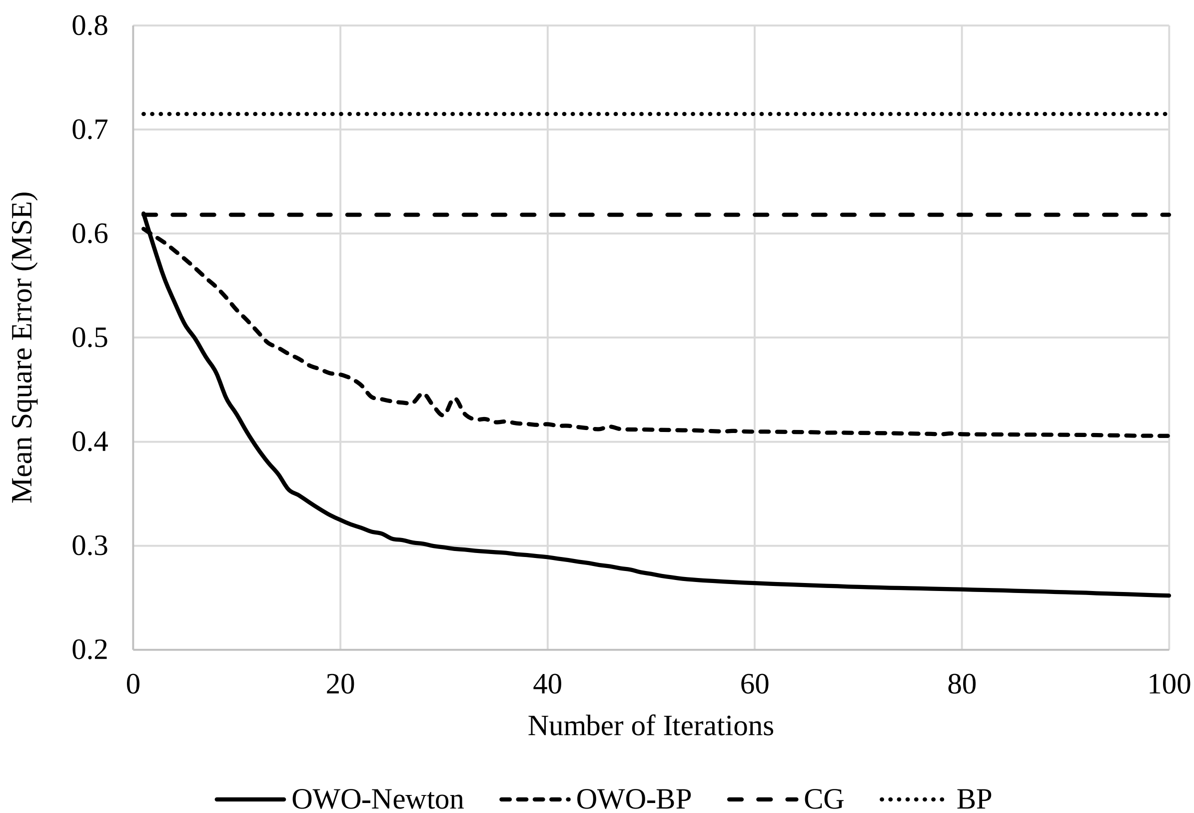


Figure 3-6 MSE for Different Training Algorithms

Cost Function

In addition to the MSE, the cost functions for the classifiers provide an alternate means to compare classifier performance. The cost assignment to a misclassified pattern requires an expert opinion. In lieu of actual cost, alternative cost functions, for instance, use of likelihood functions is suggested (*40*). A classifier with K classes, maps the input vector $\mathbf{X}_p$ to output vector $\mathbf{y}_p$ for all traffic patterns from 1 to N_v. The probability that a traffic pattern $\mathbf{X}_i$ belongs to one of K classes is denoted by

$$P_k(\boldsymbol{X}_i) = \frac{e^{y_i(k)}}{\sum_{l=1}^{K} e^{y_i(l)}} \tag{36}$$

The maximum value of $P_k(\boldsymbol{X}_i)$ for k between 1 and K yields the predicted class label for the pattern i. The maximum value of $P_k(\boldsymbol{X}_i)$ represents the likelihood that a pattern i belongs to the predicted class label, often represented by $P_{\boldsymbol{y}_i}(\boldsymbol{X}_i)$. The training algorithms predict the class label to maximize the likelihood for a given training data set. The log-likelihood function (*LL*) for the training data set can be written as:

$$LL = -\sum_{i=1}^{N_v} \log P_{\boldsymbol{y}_i}(\boldsymbol{X}_i) \tag{37}$$

The value of $P_{\boldsymbol{y}_i}(\boldsymbol{X}_i)$ lies between 0 and 1, and $\log P_{\boldsymbol{y}_i}(\boldsymbol{X}_i)$ produces negative values. Hence, the Log-likelihood function (*LL*) has a negative value. The average value of the *LL* function can serve as cost function (*J*).

$$J = -\frac{1}{n}\sum_{i=1}^{N_v} \log P_{\boldsymbol{y}_i}(\boldsymbol{X}_i) \tag{38}$$

Ideally, a *J* value of zero indicates the best classifier, and the worst possible value for *J* is infinity. The OWO-Newton method has the least average cost compared to other methods. The cost of the OWO-BP algorithm is half of the BP algorithm. Even though the BP algorithm produces more training and testing MSE, the average cost is much less than the CG algorithm. The study selects the OWO-Newton (ONN) method with a modified neural structure for classification due to its low error and cost of misclassification.

OTHER CLASSIFIERS

Previous studies (*14, 15*) obtain better classification performance over traditional methods using linear discriminate analysis and statistical methods. Hence, the study trains other classifiers using Linear Discriminant Analysis (LDA), its extension Quadratic Discriminant Analysis (QDA), and the extensively studied simple Naïve Bayes (NB) classification. In addition, this chapter evaluates and compares the classifier performance based on the GMM clustering solution. The GMM solution's better performance compared to the k-means and hierarchical clustering analysis justifies its selection for the training and validation of classifiers.

Linear Discriminant Analysis (LDA)

The LDA assigns short-term counts to traffic pattern groups with known attributes. The LDA models the linear combination of variables (hourly traffic data) that identify the traffic groups. The linear relationship separates the groups in the datasets. A separate discriminate function establishes the linear relationship between the traffic data and groups with known characteristics. If a clustering solution (based on 24-hour traffic data) has *K* number of clusters, the discriminant function for class *i* is expressed as (*14*):

$$D_i = d_{i0} + d_{i1}\, V_1 + d_{i2}\, V_2 + \cdots + d_{i24}\, V_{24} \tag{39}$$

Where D_i represents the discriminate score for the i^{th} discriminant function, V_h is the hourly volume for hour h (veh/hr), and d_{i0} and d_{ih} are the constant and function coefficient for hour h.

The LDA takes hourly traffic data for input and builds *K* discriminant functions (equal to the number of group in the input datasets). The group that results in the highest discriminate score is assigned to test patterns. The details of the model development and analysis are presented elsewhere (*14*, *40*).

Quadratic Discriminant Analysis (QDA)

The QDA assumes that the traffic patterns in each group follow a normal distribution. Even though the QDA is related to the LDA, there is no assumption of identical covariance for each cluster group. The QDA tries to incorporate the non-linear combination of variables in discriminant functions. The group that receives the highest discriminant score will be assigned to a given test pattern. In addition, the QDA takes the group (or class) specific covariance structure and forms non-linear (quadratic surfaces) class boundaries. The additional complexity in the QDA discriminant functions may increase the classifier performance. The details of model development and analysis are found in Kuhn and Johnson (*40*).

Naïve Bayes Classification

The NB classifier is a non-linear classification method based on the Bayes Rule. The model outputs the probability that a given classifier belongs to a particular class *i*, and is given by:

$$p(y = C_i|\boldsymbol{X}) = \frac{p(y)\, p(\boldsymbol{X}|y = C_i)}{p(\boldsymbol{X})} \tag{40}$$

$p(y = C_i|\boldsymbol{X})$ is the *posterior probability*, $p(y)$ is the *prior* probability of the classifier outcome, $p(\boldsymbol{X})$ is the probability of the predictor, and $p(\boldsymbol{X}|y = C_i)$ is the conditional probability that represents the probability that the predictor data comes from the data associated with class C_i.

The NB classifier assumes the independence of the input data, which is difficult to claim for the ATR traffic data. However, the independence assumption makes performing complex calculations to obtain the posterior probabilities simple. The MAP (Maximum a Posterior Probability) rule assigns the test patterns to a class. The details of model development are presented in Kuhn and Johnson (*40*).

Training and Validation

The study adopts a 10×10 validation scheme for training and validation of the LDA, QDA, and NB classifiers. The 10×10 cross-validation runs 10 folds of training and validation of classifiers and repeats the process ten times. The model that produces the highest accuracy from the 100 runs is chosen as the trained classifier model. Subsequently, the trained classifiers assign the class labels for the test patterns in the 2012 dataset. However, the cross-validation, irrespective of classification method, requires performance metrics. The following section presents a few measures to evaluate the classifier performance.

PERFORMANCE MEASURES

The learning algorithm type may affect the issue of choosing an appropriate performance evaluation criterion. Japkowicz and Shah (*41*) provide detailed performance measures, error estimation and statistically significant tests for evaluating learning algorithms. The following sections provide a brief overview of some relevant measures obtained from (*41*).

The deterministic classifiers measure outputs as a binary response in terms of zero-one loss. Confusion matrix based measures may typically be applied for deterministic classifiers (*41*). Probabilistic classifiers output the class membership in the form of a probability estimate. Typically, a *maximum a posteriori* (MAP) or a Bayesian estimate obtains the deterministic class assignment (class labeling) (*41*). After obtaining the class labels for the test patterns, the results are arranged in a confusion matrix. Whereas, scoring based algorithms, like neural networks, need thresholds to classify a test pattern to either a positive or a negative case. The scores operate in continuous space, which makes many possible threshold values possible. However, an optimal threshold is usually obtained to make a distinction between positive and negative class. In addition to confusion matrix, graphical performance measures, like Receiver Operating Characteristic (ROC) curve analysis, precision-recall (PR) curves, and cost curves may be used for scoring classifiers (*41*).

Multiple-class Performance Measures

For a given classifier *f*, denote the confusion matrix by **C** where the C matrix has elements $\{C_{ij}\}, i, j \in \{1,2,\cdots,l\}$ with *i* as the row index and *j* as the column index, and *l* as the total number of classes. Usually, a given training algorithm (classifier *f*) is trained on the training data and tested on the test set to develop a confusion matrix **C**(*f*). The confusion matrix **C**(*f*) is a square matrix *l*×*l* for a dataset with *l* classes. Each element $\mathbf{C}_{ij}(f)$ denotes the number of patterns whose actual class label is *i* , but classified as class *j*. Hence, for a test patterns set **T**, the confusion matrix **C**(*f*) for a classifier *f* can be defined as (*41*):

$$\boldsymbol{C}(f) = \left\{ c_{ij}(f) = \sum_{\boldsymbol{X} \in \boldsymbol{T}} [(y = i) \wedge (f(\boldsymbol{X}) = j)] \right\} \tag{41}$$

Where **x** represents the traffic patterns that belong to the test set, *y* is the corresponding label such that $y \in \{1, 2 \dots l\}$.

The performance metrics for learning algorithms dealing with multiple classes focus more on overall performance. Measures like error rate or accuracy of the classifier are suggested for evaluation (*41*). The error rate $R_{\boldsymbol{T}}(f)$ (or misclassification rate) gives the percent of test patterns misclassified, irrespective of class, by a classifier *f* (*41*):

$$R_{\boldsymbol{T}}(f) = \frac{\sum_{i,j:i \neq j} c_{ij}(f)}{\sum_{i=1,j=1}^{l} c_{ij}(f)} \text{ or } R_{\boldsymbol{T}}(f) = \frac{1}{|\boldsymbol{T}|} \sum_{i=1}^{|\boldsymbol{T}|} I(y_i \neq f(\boldsymbol{X}_i)) \tag{42}$$

Where I(*a*) is an indicator function that gives 1 if *a* is true and otherwise it outputs zero. Let $\boldsymbol{X}_i$ be the test pattern *i* and $|\boldsymbol{T}|$ is the size of the test set **T**.

The accuracy measure, a complement to the error rate, gives the percentage of correctly classified.

$$Acc_{\boldsymbol{T}}(f) = \frac{\sum_{i=1}^{l} c_{ii}(f)}{\sum_{i=1,j=1}^{l} c_{ij}(f)} \text{ or } Acc_{\boldsymbol{T}}(f) = \frac{1}{|\boldsymbol{T}|} \sum_{i=1}^{|\boldsymbol{T}|} I(y_i = f(\boldsymbol{X}_i)) \tag{43}$$

Accuracy or error rate characterizes the overall performance of a classifier and focuses on general behavior (*41*). The study of general behavior using these measures appears more effective for a balanced class distribution or all classes have equal importance. The GMM solution's class distribution is not equal. However, the study assumes that every class in the GMM solution remains important in exploring the patterns in the traffic data set. In essence, the accuracy and error rate measure continue to be relevant for assessing the performance of a classifier based on the GMM solution.

The Area Under curve (AUC) measure of the Receiver Operating Characteristic (ROC) curve exhibits the discriminating power of a classifier (*41*). Usually, the ROC analysis is performed for binary class classification problems. However, the ROC analysis needs an extension to the much more complex multi-class problem. AUC for the multiple class problem can be found using (*42*):

$$AUC_{multi-class}(f) = \frac{2}{l(l-1)} \sum_{l_i, l_j \in \mathcal{L}} AUC_{l_i, l_j}(f) \tag{44}$$

Where $AUC_{multi-class}(f)$ is the AUC for the multiclass ROC of a classifier *f*, $\mathcal{L}$ is the set of classes with size of $|\mathcal{L}|$ is *l*, and $AUC_{l_i,l_j}(f)$ is the area under ROC for classes l_i and l_j . The AUC is computed using a one-against-one (OAO) scheme, where a particular class is tested against all other classes on a pair-by-pair basis. In the OAO process each of the *K* pattern classes are tested against every one of other classes (*43*). The OAO process forms a system of *K (K-1)/2* binary classifiers and computes the AUC between each pair. The overall AUC for the multi-class classifier is computed using equation (44).

Single-class Performance Measures

However, when only investigating a single class (say, class *i*), the performance measure may be computed for a single class. Measures on individual classes can also serve the purpose of measuring the overall performance of a classifier.

Consider a single class (class *i*) problem as a binary classification case where a classifier takes a value of one if the output belongs to a class *i* or takes zero for all other remaining classes. The confusion matrix, in this case, has four characteristic values: true positives (TPs), false positives (FPs), false negatives (FNs), and true negatives (TNs). False positives and false negatives represent the negative and positive labels that are erroneously labeled as positives and negatives respectively. True positives and true negatives have labels correctly classified as their original labels positives and negatives. A Confusion matrix **C**(*f*) of a generic binary classifier *f* is shown in Table 3-2.

Table 3-2 Confusion Matrix of a Binary Classifier (Source: Japkowicz and Shah (*41*))

Class *i*	Predicted_Negative	Predicted_Positive	Quantity
Actual_Negative	True Negative (TN)	False Positive (FP)	N = TN + FP
Actual_Positive	False Negative (FN)	True Positive (TP)	P = FN + TP

The True Positive Rate (TPR) remains the most measured metric for a single class problem. This measure computes the proportion of patterns with class label *i* actually predicted as class *i* by the classifier *f*.

$$TPR_i(f) = \frac{c_{ii}(f)}{\sum_{j=1}^{l} c_{ij}(f)} = \frac{TP}{TP + FN} = sensitivity \quad (45)$$

The False Positive Rate (FPR) gives the instances that a classifier assigns patterns to class *i* that do not actually belong to this class.

$$FPR_i(f) = \frac{\sum_{j:j\neq i} c_{ji}(f)}{\sum_{j,k:j\neq i}^{l} c_{jk}(f)} = \frac{FP}{FP + TN} \quad (46)$$

Similar measures, True Negative Rate (TNR) and False Negative Rate (FNR) can also be obtained for the negative class.

$$TNR(f) = \frac{TN}{FP + TN} = specificity \quad (47)$$

$$FNR(f) = \frac{FN}{FN + TP} \quad (48)$$

The true positive rate is also called *sensitivity* and the true negative rate is called *specificity*. In the multi-class scenario, sensitivity reflects the accuracy of a classifier. The precision of a classifier measures how precise a classifier *f* identifies the patterns of a given class. The positive prediction value (PPV) describes precision, and measures the proportion of correctly assigned positive class (or class *i*).

$$Prec_i(f) = PPV_i(f) = \frac{c_{ii}(f)}{\sum_{j=1}^{l} c_{ji}(f)} = \frac{TP}{TP + FP} \quad (49)$$

The Recall of a classifier is equivalent to the definition of TPR or sensitivity.

$$Rec(f) = TPR(f) = \frac{TP}{TP + FN} \quad (50)$$

ROC Analysis

In addition to measures based on the confusion matrix, the ROC analysis continues to be the most widely used performance evaluation method. The ROC graphs provide a better tool to visualize the performance of a classifier over a variety of decision criteria (*41*). The ROC analysis can help study the behavior of the classifier, aid comparative analysis, and model the selection (threshold selection through trade-off analysis) process (*41*).

The ROC plot has FPR (or *1-specificity*) on the x-axis and TPR (or sensitivity) on the y-axis. The ROC analysis, in a way, studies the relationship between sensitivity and specificity of the classifier (*41*). Both TPR and FPR have values between 0 and 1, hence the ROC space is a unit square. The point (0, 0) in the ROC denotes a trivial classifier that misclassifies (or gives the negative class) all test pattern instances. The point (1, 1) also denotes a trivial classifier because it labels all instances positive. The diagonal that connects (0, 0) and (1, 1) has TPR=FPR. Any classifier that produces values along the diagonal is a random classifier like a coin toss. The points (1, 0) and (0, 1) are other extremes of the ROC. The point of (1, 0) denotes the worst classifier due FPR=1 and TPR=0. However, an ideal classifier should be at the corner of (0, 1). Figure 3-7 denotes the ROC space.

The classifiers use operating points to denote whether a traffic pattern belongs to class *i* (positive class) or the other class (negative class). An operating point refers to a particular decision threshold in the ROC space to assign discrete labels to the test patterns.

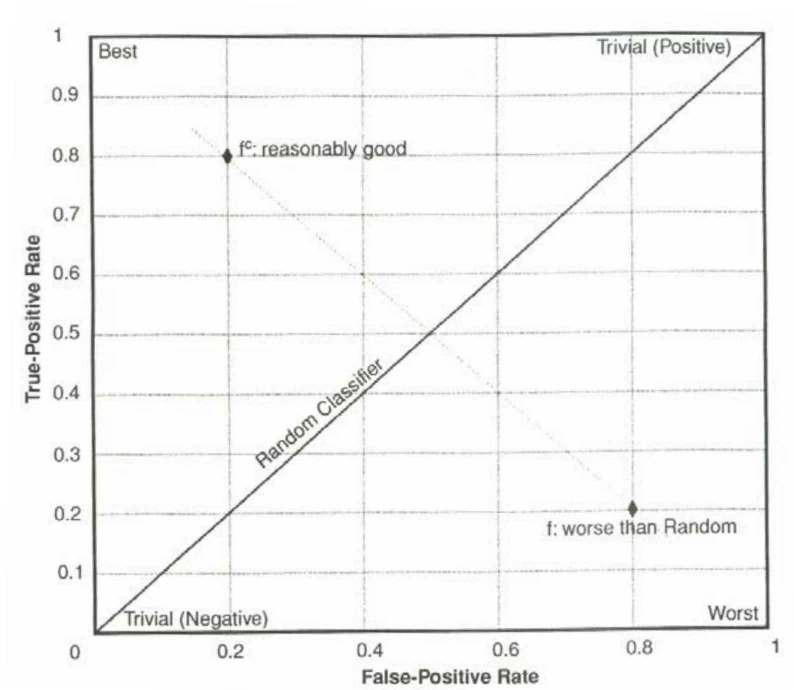


Figure 3-7 ROC Curve Space (Source: Japkowicz and Shah (*41*))

The patterns having classifier scores above the threshold are labeled as positives. The optimal operating point (or threshold) that distinguishes positive and negative class has implications on the classifier's performance. One way of selecting an optimal threshold value is to draw a graph between the sensitivity and specificity versus each possible threshold value. The threshold that maximizes both sensitivity and specificity (the threshold at which both sensitivity and specificity are equal) is selected as the optimal value (*44*).

In the ROC analysis, generating a ROC curve and finding its AUC describes the discriminating power of a classifier. If distinct classifier scores, over its entire range, are tuned as decision thresholds, the analysis obtains different sets of TPR and FPR for each threshold. These sets form a continuous curve in the ROC space. Using area under the ROC curve and the following guidelines, the strengths of a classifier may be assessed (*44*):

- AUC = 0.5, suggests no discrimination like a coin flip
- 0.5 < AUC < 0.7, poor discrimination, not much better than a coin toss
- 0.7 ≤ AUC < 0.8, acceptable discrimination
- 0.8 ≤ AUC < 0.9, excellent discrimination
- AUC ≥ 0.9, outstanding discrimination

A separate classifier and ROC for each class can assess the discriminating performance by class.

Precision-Recall (PR) curves explore the trade-off between the well-classified positive patterns and misclassified negative patterns (*41*). PR curves measure the amount of precision at various degrees of recall. Precision decreases as recall increases. PR curves seem particularly more important than ROC curves when the classes appear highly imbalanced in the data (*45*).

Cohen's κ (kappa) statistic

The classifiers are trained using given (called as true labels) labels generated by a process (for instance using the GMM methodology). The process is assumed to generate labels in an unbiased and correct manner and hence the labels do not occur by chance. However, the label generating process (or clustering process) may not generate true labels. Moreover, the true labels for traffic data seem neither defined nor unquestionably established. The authors believe that the clustering process approximately generates labels that may potentially reflect the ground truth. The study needs to test the labels that are used in training and validation of classifiers for the occurrence by chance. Despite the strengths of the label generating process, the performance measures should correct for chance (*41*). The performance measures listed previously do not account for the coincidence of concordance between the classifier output and label generating process. Cohen's κ (kappa) statistic measures the chance corrected agreement between any two label generating processes. The κ Statistic uses two estimates P_0 and P_e^C as shown below (*41*):

$$\kappa = \frac{P_0 - P_e^C}{1 - P_e^C} \tag{51}$$

P_0 denotes the probability of overall agreement between the classifier and true class labels. P_e^C denotes chance agreement over the labels.

$$P_0 = \frac{TN+TP}{m}, \text{ where } m\text{=TN+FN+TP+FP and} \tag{52}$$

$$P_e^C = \frac{(FN+TP)}{m} \cdot \frac{(FP+TP)}{m} + \frac{(TN+FP)}{m} \cdot \frac{(TN+FN)}{m} \tag{53}$$

The chance agreement is a product of the proportion of the actual instance of a class and its predicted instances summed over all classes.

EVALUATION OF CLASSIFIERS

The study evaluates trained classifier performance on the test data sets using the metrics presented in the previous section. The author uses both validation and testing terms synonymously. Multiple methods exist for testing a classifier. The *holdout* method trains a given classifier on a certain amount of data (80 percent) and tests it on the remaining data set (20 percent) (*23*). *Leave-one-out* (or *Jackknife*) approach, a computationally expensive approach, tests every data pattern and train the classifier with one pattern less than that of the given data (*41*).

Cross Validation (CV) remains the most popular method for testing learning algorithms. *K-fold cross-validation* divides the data set into *k* disjointed subsets (or folds) of equal size. The learning algorithm is trained on *k*-1 subsets and tested on the k^{th} subset. Each of the *k* folds becomes a test set once and the process outputs *k* different values of chosen performance measures (*41*). Ten appears to be the most commonly used value for the number of folds (*41*).

Even though the *k*-fold cross validation considers all data patterns, a given fold may not represent all class labels. Either over or under representation of classes in a given fold gives erroneous performance measure values. Hence, the cross-validation should account for class distribution when generating the training and testing folds. A *Stratified k-fold cross validation* splits the data into *k* disjointed subsets such that each subset has the same class proportion as the original dataset.

In a 10-fold cross validation, the learning classifier yields only ten different performance values. Comparing the performance of multiple classifiers on a limited set of performance measures may not yield significant results. In lieu of small samples, multiple runs of validation (like multiple *k*-fold cross validation) is recommended (*41*). For instance, by running (or repeating) stratified *k*-fold cross validation multiple times, the classifier generates different sets of *k*-folds from the datasets and gets more sets of performance metrics.

Multiple resampling methods generate performance measures based on multiple sampling from the data sets. Random subsampling, bootstrap, and randomization approaches represent a few prominent approaches used for multiple resampling (*41*). The bootstrapping method assumes that a given dataset is representative of the original distribution of the population. Bootstrapping creates new samples from a given dataset by randomly drawing samples of equal size as an original dataset with replacement. For a *B* number of bootstrap samples on a dataset $\mathcal{D}$, the test sets T_{boot}^{i} (*i=1… B*) are formed and a classifier *f* is trained and tested on every bootstrap sample while computing performance measures. The average performance values (for instance, accuracy of a classifier) for bootstrap samples appear below (*41*):

$$Acc(f) = \frac{1}{B}\sum_{i=1}^{B}\left[\frac{1}{\left|\boldsymbol{T_{boot}^{i}}\right|}\sum_{j=1}^{\left|\boldsymbol{T_{boot}^{i}}\right|} I\left(y_j = f_{boot}^{i}(\boldsymbol{X}_j)\right)\right] \tag{54}$$

Where f_{boot}^{i} is the classifier trained on the training sample from i^{th} bootstrapping step and testing the classifier performance on T_{boot}^{i} test dataset.

10×10 Stratified Cross Validation

Validation using multiple resampling through bootstrapping and k-fold stratified cross validation on each bootstrap run provides a framework to perform associated significance tests related to classifier performance (*41*). This type of validation represents a $r \times k$ cross-validation scheme. In this approach, r runs of k-fold stratified cross validation are performed on the metrics between any two given classifiers. The difference between the performance measures of the two classifiers is tested for a statistical significance. Bouckaert (*46, 47*) studies different $r \times k$ schemes and recommends a 10×10 CV scheme due to the increased power of the test.

The Z-score is computed to compare the performance of two classifiers for statistical significance. Let d_{ij} be the difference in performance measure (for instance, accuracy) between the two classifiers f_1 and f_2 reported when testing the data of i^{th} fold in the j^{th} run. For r runs and k folds, the average difference is computed as (*41*):

$$\bar{d} = \frac{1}{k}\sum_{i=1}^{k}\frac{1}{r}\sum_{j=1}^{r} d_{ij} \tag{55}$$

The test statistic calculation requires a variance estimate. However, the variance estimate in the $r \times k$ scheme depends on the manner used for obtaining the average over the folds and runs. Even though four ways of variance estimation exist, the *use all the data* scheme (equation 55) appears appropriate when a test with high power remains desirable (*47*). The variance estimate from the *use all the data* scheme is given be (*41*):

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{k}\sum_{j=1}^{r}\left(d_{ij} - \bar{d}\right)^2}{kr - 1} \tag{56}$$

The Z-score (test static) is computed using the variance and mean of the validation scheme as (*41*):

$$Z = \frac{\bar{d}\,\sqrt{f_d + 1}}{\hat{\sigma}} \tag{57}$$

Where f_d is the degree of freedom, which is equal to $k \times r$-1 for *use all the data* scheme (*41*).

PERFORMANCE COMPARISONS

The study adopts a 10×10 stratified cross validation approach for the performance evaluation of the classifiers. The 10×10 CV approach makes ten runs of 10-fold stratified cross-validation for a given data set. The study performs a comparative analysis on the classifiers: Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Naïve Bayes (NB) classifier, and the ONN. For statistical testing of the performance difference, the study uses a classifier's overall accuracy in predicting the labels for test data sets.

This section seeks to compare the performance of ONN classifier with the rest of the commonly adopted classifiers. The study also observes the variations of the performance measures across multiple runs. Figure 3-8 (*a-f*) shows the box plots for performance metrics from the 10×10 cross-validation scheme. Smaller intervals from the mean reflect less variation of the performance measures across multiple runs. However, the plots show differences of performance between classifiers for each measure. Table 3-3 lists the average classifier performances.

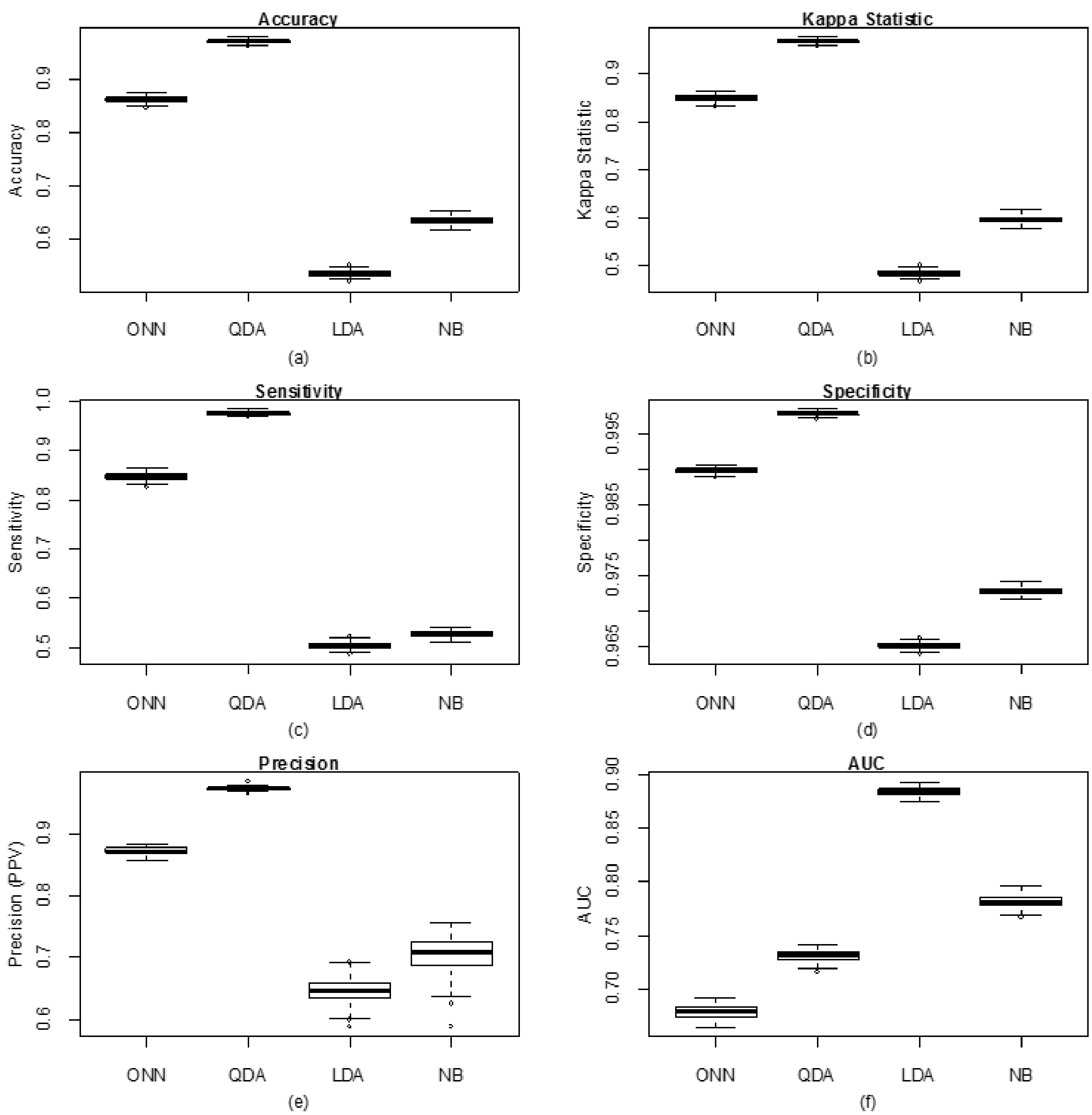


Figure 3-8 Box Plots of Performance Measures of Different Classifiers

Table 3-3 Performance Metrics of Different Classifiers

Performance Measure	ONN	QDA	LDA	NB
Accuracy	86%	97%	54%	63%
Sensitivity (TPR or Recall)	0.85	0.98	0.50	0.53
Specificity (TNR)	0.99	1.00	0.97	0.97
Precision (Positive Prediction Value)	0.87	0.97	0.65	0.70
AUC	0.68	0.73	0.88	0.78
Kappa Statistic	0.85	0.97	0.48	0.60

Based on the overall performance metrics, the accuracy for QDA classifier appears better than the other classifiers. The ONN makes a correct prediction for 86 percent of the patterns. The LDA performs poorly compared to QDA for the accuracy measure, and the NB classifier even performs better than LDA. In addition, the accuracy for ONN among 100 runs varies between 85% and 87%.

A similar trend continues for the isolated metrics: sensitivity, specificity, and precision. Sensitivity values of 0.85 suggest that the ONN rightly predicts 85 percent of the actual labels of the test patterns. In other words, the ONN misses 15 percent of the actual class labels or misclassifies them to a different class. For the labels of non-occurrence, the ONN classifier rightly predicts 99 percent of the patterns. Sensitivity and specificity remain higher for QDA than all other methods. The QDA method receives a sensitivity value of 0.98 and a specificity of unity. The QDA method almost predicts rightly both actual (positive) and negative classes. The NB classifier shows a little bit more sensitivity than the LDA and the specificity is equal among both methods. However, both the LDA and NB can only be trusted 50 percent in classifying the positive classes; however, both these measures appear capable when classifying the negative classes. The precision metric (positive prediction value) remains higher for the QDA method followed by the ONN classifier. A precision value of 0.87 indicates the positive prediction seems more reliable for the ONN. The LDA and NB methods' positive prediction values stay lower than the ONN and QDA.

The AUC for the multi-class problem is computed using a pair-wise AUC for all class pairs in the clustering solution. The AUC is higher for the LDA and NB classifiers, and all classifiers yield an AUC value of greater than 0.5, which indicates the performances are better than random guessing. The authors did not encounter and remain unaware of any guidelines, unlike binary classifiers, for assessing the strength of a classifier when performing multi-class ROC analysis. However, more AUC (at least 0.5) shows the good discriminatory power of a classifier. The LDA has higher AUC values than other classifiers. Moreover, the QDA obtains a larger AUC value than the ONN classifier.

The accuracy estimate of 0.86 (or 86%) for the ONN does not appear overly optimistic, as the accuracy and κ statistic produces almost similar values. Other classifiers' kappa statistic is also close to the accuracies reported; hence, the accuracy values do not appear to be due to randomness. The label generating process, the GMM method, seems to produce true clusters (only believed to be true in lieu of ground truth) and clustering labels do not exactly happen by a chance.

The author computes the ROC curves for the multi-class problem using a one-against-all (OAA) scheme. The OAA scheme tests each class against all other classes. When implemented. The OAA produces K (or a number of classes in the clustering solution) binary classifiers. Each classifier tests a particular class i against all other classes j (Where $j \in \{1,2,\cdots,K\}$ and $j \neq i$). The ROC curves generated for the GMM clustering solution is shown in Figure 3-9.

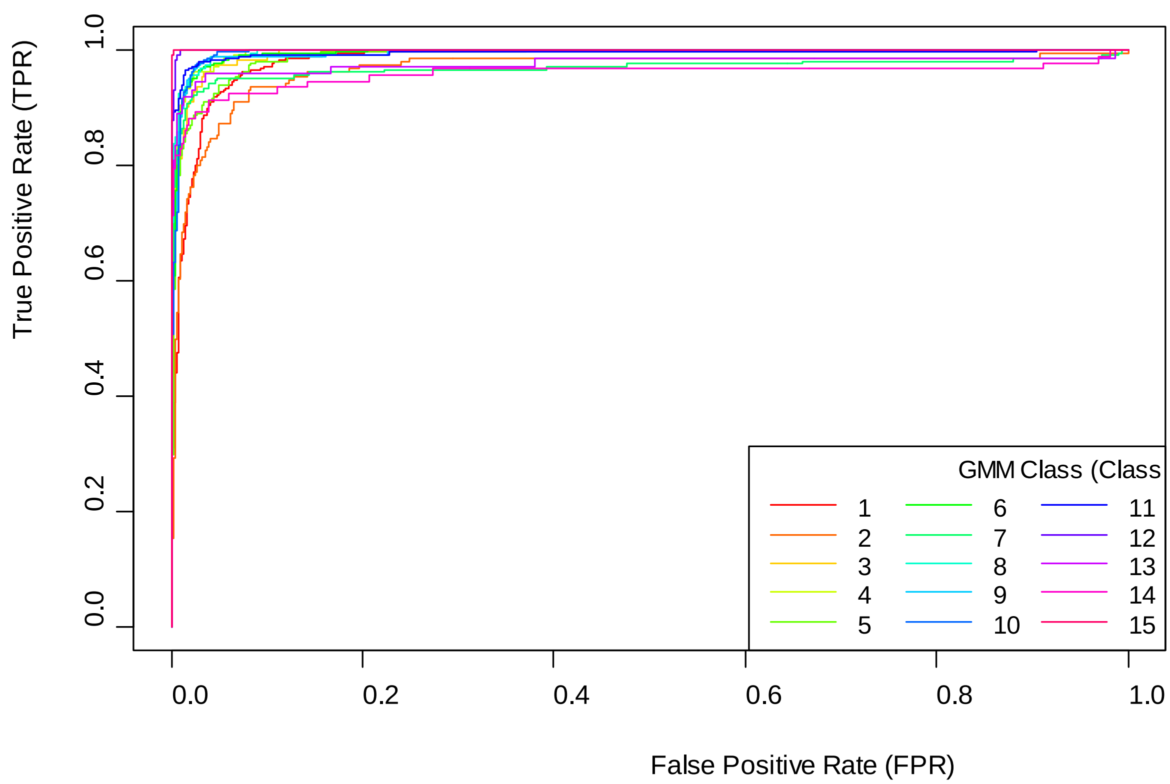


Figure 3-9 ROC Curves for GMM Classes

The scoring classifiers presented in this study output continuous values typically in the range of [0, 1]. Ideally, an output value closer to one gives a positive class and to zero yields a negative class. Practically, any value (not only 0.5) between zero and one can act as a decision threshold. Different values for TPR and FPR can be computed for various threshold values. Each curve depicts the classification potential of the ONN by class. The AUC for curves is greater than 0.9; hence, the ONN outstandingly distinguishes each class against all other classes. Figure 3-10 shows the AUC for each binary classifier produced by the ONN and QDA classifiers. The ONN produces slightly higher values of AUCs than the QDA method.

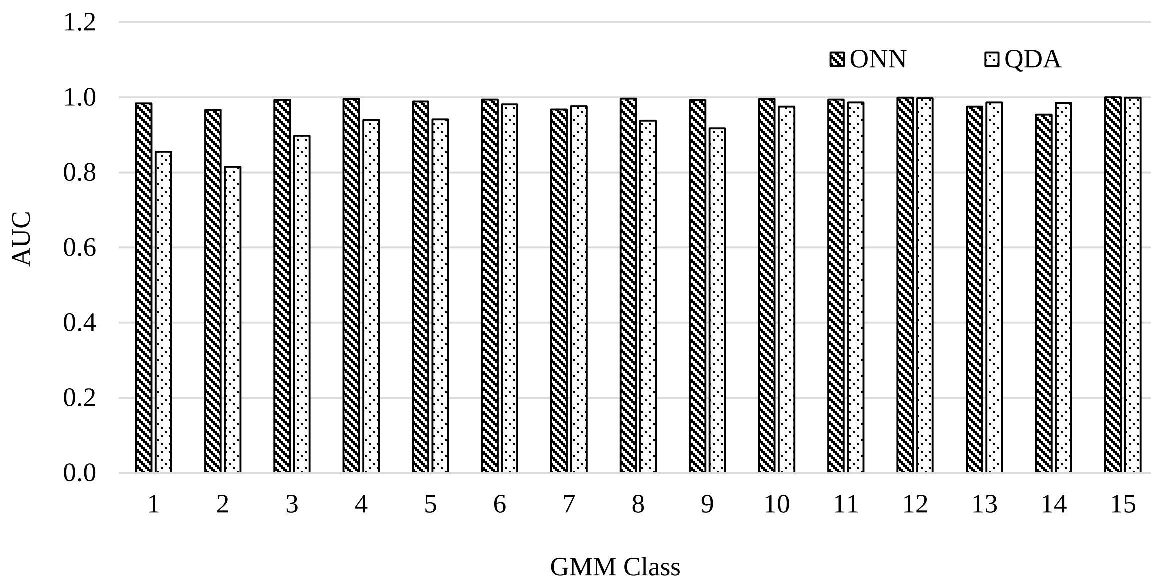


Figure 3-10 AUC by the GMM Classes between the ONN and QDA methods

The study assumes that not only detecting many possible true classes but also identifying the negative class is important. The Precision-Recall (PR) curves, similar to the ROC curves, measure the amount of precision at various degrees of recall. The curves plot the relationship between the precision of a classifier and its recall. The curves explore the trade-off between well classified positive examples and misclassified negative examples (*41*). Unlike the ROC curves, the PR curves have negative slopes because precision decreases as recall increases (see Figure 3-11).

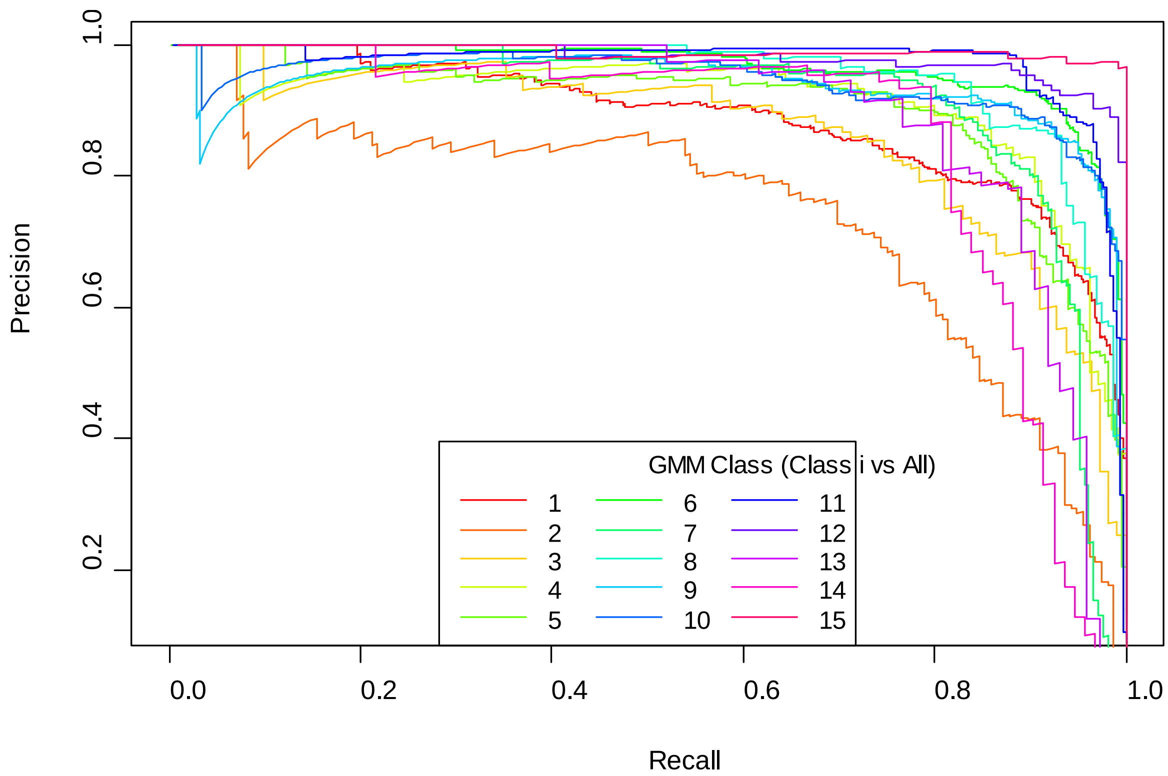


Figure 3-11 Precision and Recall Curves between the GMM Classes

Statistical Significance Tests

The study performs statistical tests on the significance of the performance differences between the classifiers using a one-tailed *z*-test. The accuracy measure, the most widely used metric, represents the overall performance of the classifiers. Hence, the one-tailed z-test uses accuracy differences between the classifiers to perform the statistical testing. However, other measures listed in the previous sections can also be used for statistical testing.

The null hypothesis assumes no difference in the accuracy between the classifiers. The study compares the ONN classifier accuracy with other classifiers' accuracy (Table 3-4). The ONN accuracy is greater than the LDA and NB classifier in the order of 38 and 27 percent, respectively. However, the QDA performance is better than the ONN method. All the differences are statistically significant at a five percent significance level.

Table 3-4 z-tests on Performance Comparison between the Classifiers

Comparison	Accuracy Difference, $\bar{d}$	Percentage Change in Mean Accuracy (%)	p-value	95% Confidence Interval
ONN vs LDA	0.327	38	<0.001	[0.326 , 0.328]
ONN vs QDA	-0.109	-13	<0.001	[-0.110 , -0.108]
ONN vs NB	0.228	27	<0.001	[0.227 , 0.230]

AADT ESTIMATION

The 2012 ATR data set with a sample of 32,289 data patterns is used for testing. The 2012 dataset has hourly traffic data showing time of day variation and the groupings according to the ODOT seasonal trend grouping method and highway functional class. The trained classifiers assign the group number according to the GMM clustering solution.

The AADT is calculated by matching the *group number and month* of the test patterns (using 2012 data). The product of the matched average ratio of AADT to DT (corresponding to a matched *group number and month* from the seasonal adjustment factors table) and the sum of 24-hour traffic volume (daily traffic or DT) estimates the AADT. The computed AADT value is compared with the actual AADT value to obtain an error. The Mean Absolute Percent Error (MPAE) given in Equation (58) compares the estimates from the clustering methods (*48, 49*):

$$MAPE = \sum_{n=1}^{N} \left(\left| \frac{AADT_{Estimated} - AADT_{Actual}}{AADT_{Actual}} \right|_{n} \times 100 \right) / n \tag{58}$$

In addition, the study computes and reports the standard deviation of the errors.

Monthly Variation of the Error

Figure 3-12 shows the monthly variation of error for the different classifiers. The ONN method's average monthly percentage error variation almost matches the QDA classification. Both these methods produce lower MAPE values compared to the LDA and NB methods. The QDA performs better than the LDA method across all months. The monthly variation of errors for the NB classifier exhibits more error than the LDA method. Except for the month of July, the ONN produces fewer errors than the ODOT seasonal trend grouping. The FC grouping also produces more monthly average error than the ONN classifier.

The NB classifier produces higher values of standard deviation of the Absolute Percent Error (APE) for most of the months. The deviation remains lower for the both ONN and QDA methods (see Figure 3-13). However, the ODOT method shows less deviation than the ONN method during the months of July, August and December. The LDA method consistently shows more variation than the QDA method. The FC grouping produces more variation than the ONN and ODOT methods.

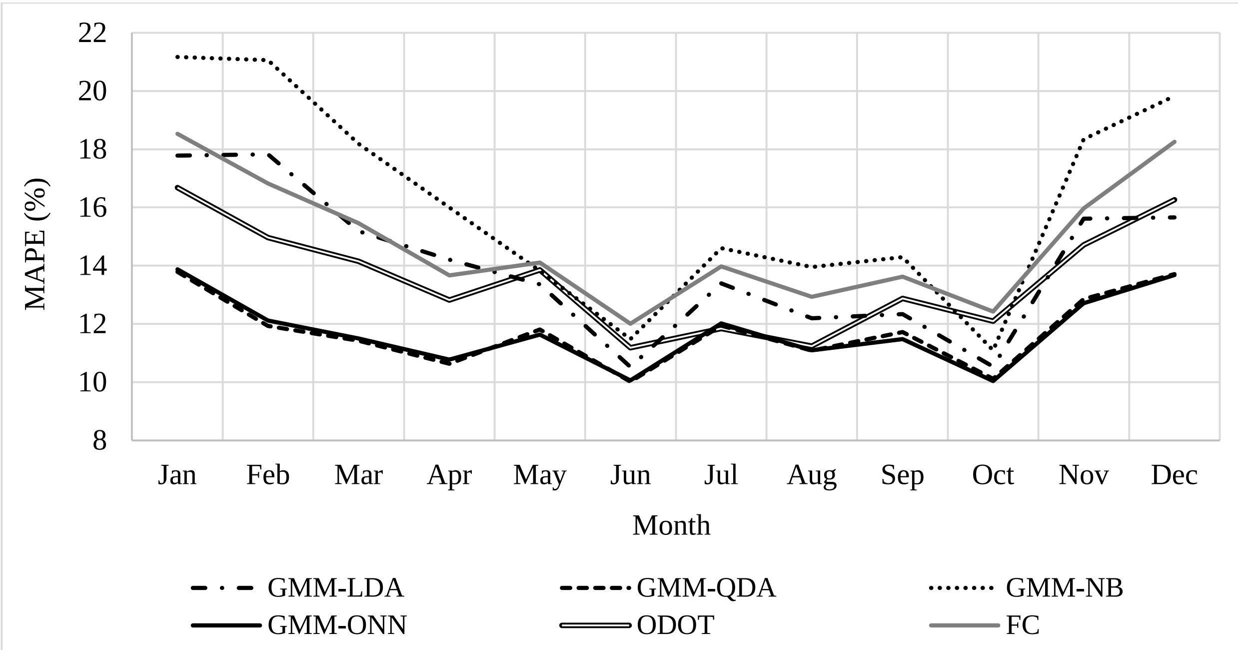


Figure 3-12 Monthly Variation of MAPE for Different Classifiers

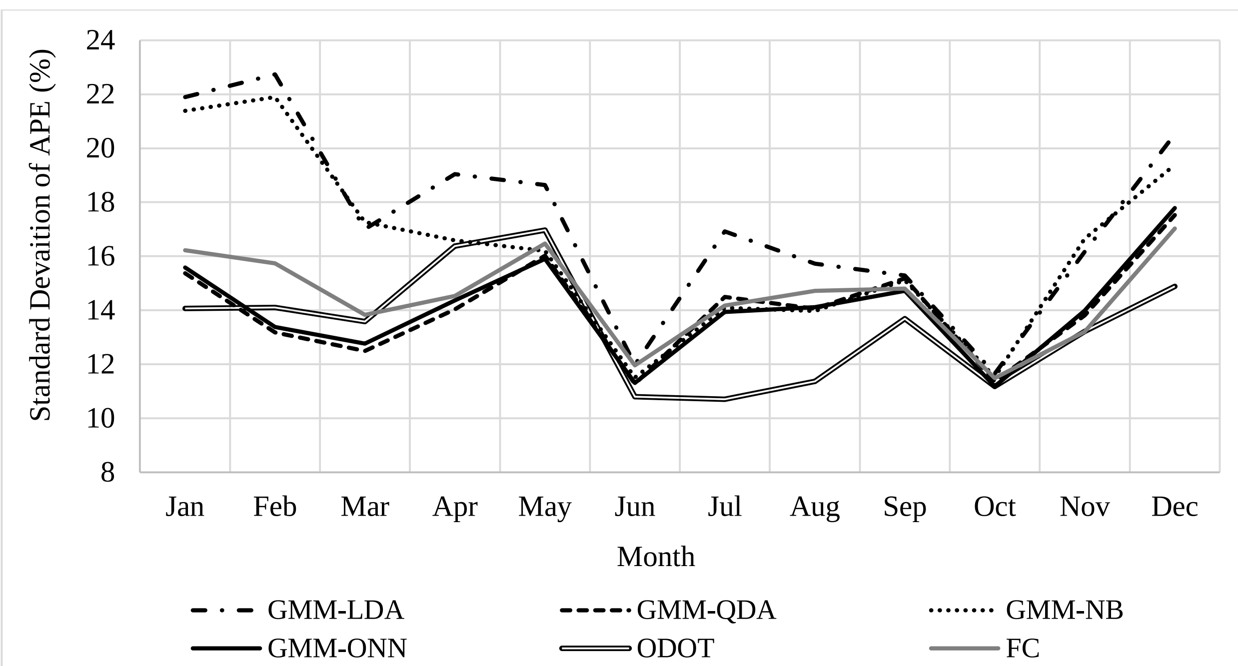


Figure 3-13 Standard Deviation of Errors for Different Classifiers

Error by Cluster

Table 3-5 presents the average errors by GMM class for each classification method. The ONN and QDA methods produce almost equal values of error for the GMM classes. The clusters labeled originally as *commuter* pattern types by this study (clusters 10 to 15) show less than ten percent error. This observation meets the FHWA recommendation of ten percent AADT forecasting error (*3*). The LDA produces more error compared to the QDA classification. The clusters that exhibit summer/recreation patterns exhibit large error differences. The NB classifier produces a lot more error than the other three methods for the clusters 10 to 15.

Table 3-5 MAPE by the GMM Cluster and Classification Method

Classifier	GMM Clusters														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
LDA	19.3	28.1	18.9	20.3	17.6	19.4	13.9	9.7	13.9	6.9	6.2	2.6	4.1	8.4	4.0
QDA	14.4	20.5	15.4	6.1	14.9	19.4	13.9	7.0	12.9	6.6	5.3	2.4	4.9	7.9	3.9
NB	14.8	21.8	15.9	8.2	16.1	19.3	16.0	6.6	14.1	6.5	10.3	30.8	28.1	24.8	28.7
ONN	14.5	21.4	15.1	6.5	15.1	19.2	13.4	7.0	12.9	6.6	5.3	2.4	4.1	7.7	3.9

Error by Functional Class

Figure 3-14 shows the average error rate among different highway functional classes. The ONN method produces less than five percent error for urban interstates and less than ten percent error for urban freeways/expressways and arterials. Typically, the traffic patterns on these facilities represent commuter pattern trends. In addition, the error achieves the FHWA recommended value of less than ten percent. The error varies between ten and fifteen percent for other arterials and rural interstates. However, the major collectors, irrespective of classification method, produce higher than twenty percent error. The ONN and QDA methods' errors appear almost similar across each different functional class. The ODOT method shows more error than the ONN method for all functional classes except for major collector streets. The FC grouping consistently produces more error than the ONN and ODOT methods. The LDA method produces more error than the QDA method.

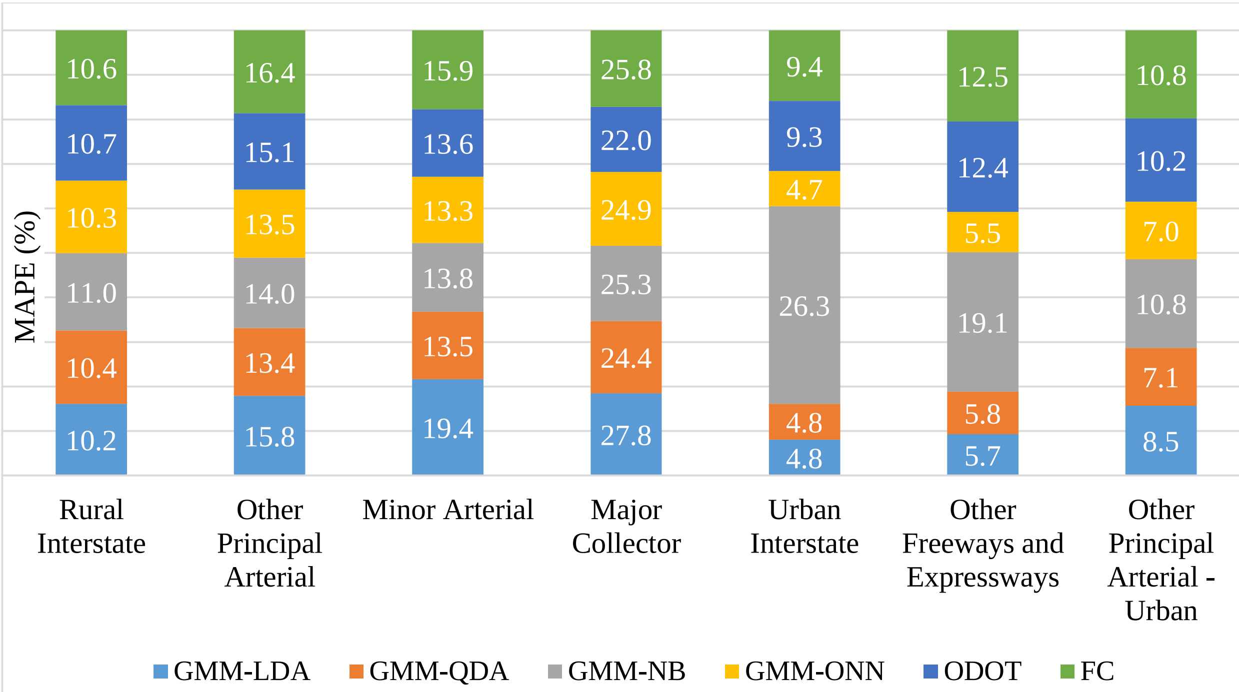


Figure 3-14 MAPE by Highway Functional Class

Error Distribution by Error Size

Figure 3-15 shows the error distribution of the classification schemes. The ONN method has the highest proportion of patterns with less than five percent MAPE value. The error distributions of both the ONN and QDA methods exhibit similar trends. However, the ONN method produces a slightly lower number of patterns for increased error sizes. The LDA and NB classifiers perform poorly compared to the QDA methods. The ODOT method has a lower number of patterns for the less than five percent error range compared to the ONN classifier. Finally, the study checks the percent of patterns produced for benchmark error size of fifteen percent (*50*). Both the ONN and QDA have the lowest number of the patterns that exceed fifteen percent error, and they differ by seven percent compared to the ODOT percent of patterns.

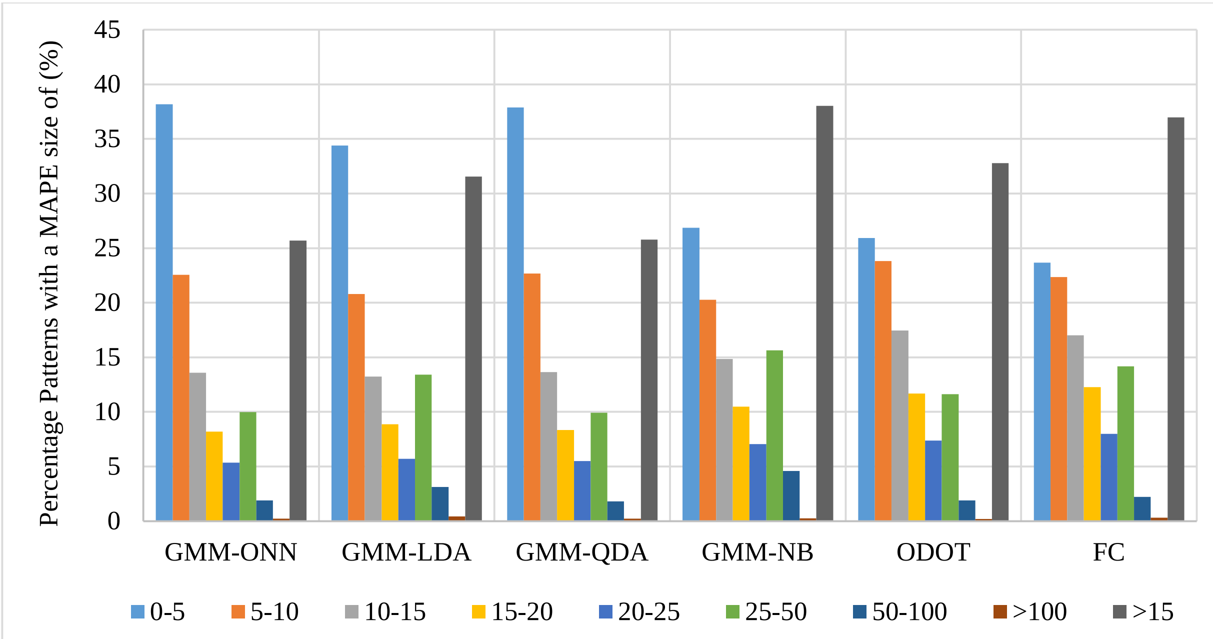


Figure 3-15 Error Distribution of Classification Methods

Table 3-6 shows the quartile distribution of the errors. The ONN and QDA methods show a similar distribution of errors with an equal mean and standard deviation. However, the standard deviation of the ODOT method appears lower than the ONN and QDA methods. The 95 percent confidence intervals show a tighter mean error interval. The LDA and NB classifiers carry a larger deviation of errors. The mean and deviation of error for the LDA method seem higher than the QDA method.

Table 3-6 Summary Statistics of Errors by Classification Methods

Method	1st Quartile	Median	3rd Quartile	Mean	SD	95% CI of Mean	
						lower	upper
ONN	3.0	7.3	15.3	11.8	14.3	11.6	11.9
LDA	3.3	8.5	18.5	14.1	17.8	13.9	14.3
QDA	3.0	7.3	15.4	11.8	14.2	11.6	11.9
NB	4.6	10.9	21.6	16.2	17.0	16.0	16.4
ODOT	4.8	10.0	18.2	13.6	13.7	13.4	13.7
FC	5.3	11.0	19.8	14.8	14.8	14.7	15.0

The study conducted a one-tailed paired t-test between the ONN method and errors produced by other classifiers. The null hypothesis assumes no difference between the errors produced by the ONN and another classifier. Table 3-7 shows the results of the t-tests conducted on the 2012 data sets.

Table 3-7 *t*-tests for Mean Error Difference between Classifiers

Comparison (GMM with)	*t* statistic	*p*-value	95 % CI of Mean Error Difference		Mean Error Difference	Percentage Error Reduction (%)
			lower	upper		
ONN vs LDA	-41.08	2.20E-16	-2.37	-2.15	-2.26	-19.6
ONN vs QDA	0.18	> 0.05	-0.04	0.05	0.005	0.0
ONN vs NB	-62.04	2.20E-16	-4.54	-4.27	-4.41	-37.4
ONN vs ODOT	-29.29	2.20E-16	-1.90	-1.66	-1.78	-15.4
ONN vs FC	-54.03	2.20E-16	-3.13	-2.91	-3.02	-26.1

The mean error rate between the ONN and QDA remains almost equal. The ONN produces lower errors, a reduction of 15% to 38%, compared to other classifiers. The error difference between the ONN and other methods (except the QDA) show a statistically significant error reduction. However, the study fails to reject the hypothesis of equal means between the ONN and the QDA methods at a five percent significance level.

ONN VS QDA

During training and validation, the QDA method shows the highest accuracy and other performance measures than the ONN and other classifiers. The cluster-wise ROC analysis shows better AUC values for the ONN classifier. Although the QDA method receives a validation accuracy of 97% compared to the 85% accuracy for the ONN method, the 12% accuracy advantage for the QDA method does not translate into the AADT estimates on the new datasets. When tested for the 2012 data, both methods show statistically significant *indifference* on the AADT estimates. However, the ONN method has slightly improved error estimates. The question, then, becomes which method to use when classifying the traffic patterns.

As QDA and ONN are score based classifiers, soft-max transformations coerce the outputs into probability-like values (*40*). In both cases, the test pattern's class corresponds to the maximum probability class.

The study adopts the ONN method in the subsequent analysis for the following reasons:

- The discriminate functions in the QDA are separated by quadratic surfaces. The model development, selection, and testing need considerable amount of knowledge on probability and statistics. In addition, model validation requires additional tests to verify the model assumptions. The model structure and its interpretation are not straightforward and are often based on variance measures. The neural network models appear simple to understand, train, and use. For instance, the ONN models are based on ordinary least square and second order error gradients. The neural networks do not require strong assumptions or any tests for checking the model form and structure.

- The ONN method seems relatively adaptable to changes in the datasets. The agencies typically make enhancements to the developed models to accommodate changes in the traffic data collected in subsequent years. The developed classifier should function and respond to the changes- like adopting new datasets from different geographic regions (data from other DOTs)-smoothly. For instance, if new datasets of 24-hour traffic patterns yield a different clustering solution, say other than the fifteen cluster solution, then the output layer in the neural network model changes the number of units and the network selection process again chooses the best number of units in the hidden layer. However, the QDA analysis needs quite a bit of remodeling efforts. The steps: model development, selection and validation of assumptions must be repeated.
- Both the ONN and QDA models belong to the non-linear discriminant analysis family.
- The ONN does not directly give input-output mapping in terms of relationship (or equations) like the discriminant functions in the QDA. If necessary, the relation between the input and output can implicitly be formulated using connection weights and thresholds. The lack of equations may not hinder the network performance for classification of the test patterns.

The attributes like adaptability to new datasets, easy to train and interpretable, and producing comparable and slightly fewer error values encourage the study select the ONN method for classification of 24-hour traffic patterns.

SUMMARY

The cluster assignment process is the most critical step in the AADT estimation process. The present study addresses the issues of cumulative errors that are inherent with the traditional approach during assignment step. In particular, the study proposes an optimal neural network structure and two variants of learning algorithms to improve the estimation error. The study provides a framework to perform validation using multiple resampling through bootstrapping and k-fold stratified cross validation on each bootstrap sample to assess the classifier performance. The study compares Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Naïve Bayes (NB) classifier, and the ONN. For statistical testing, the study uses the classifier's overall accuracy in predicting the labels on the test data sets. Each classifier is tested again for its AADT estimation. The OWO-Newton (ONN) method produces lower errors, a reduction of 15% to 38%, compared to the other classifications. The error difference between the ONN and other methods (except the QDA) show statistically significant error reductions.

CONCLUSIONS

The changes to the neural network structure, learning process and learning algorithms show improved performance compared to the back-propagation algorithm. The network selection process outputs help to select the best number of units in the hidden layer. The OWO-Newton (Output Weights Optimization-Newton method) reduces assignment errors and misclassification costs. The overall performance measures, ROC curve analysis and performance measures based on single class aid in the performance analysis. In particular, the study eliminates biases in the estimated measures by repeating stratified cross validation for multiple runs. The 10×10 cross-validation evaluates the performance from 100 different samples by maintaining the same cluster proportion as the original clustering solution in both training and testing datasets. The ONN method shows an overall accuracy of 86 percent with an AUC value of more than 0.5. The ONN method's average monthly percentage error produces slightly reduced errors than the QDA classification, but significantly, fewer errors compared to other classifiers. The error reduction of 15% to 38% is observed for the ONN classification. The ONN method produces less than 5 percent error for urban interstates and less than ten percent for urban freeways/expressways and arterials that are within in the limits recommend by the FHWA for commuter patterns. The ONN method has the highest proportion of the patterns with less than five percent MAPE value.

The proposed ONN classification provides an improved way of assigning traffic patterns; its higher accuracy and better classification capabilities reflect an effective approach for labeling the unknown traffic patterns. The statistically significant error reduction for the ONN classification, based on the GMM clustering solution, shows a reliable approach for AADT estimation. The mean error rate of less than ten percent for commuter routes and lower standard deviation of errors show the satisfactory accuracy of the AADT estimates. The ONN classification helps to minimize the judgment errors emanated from the traditional and factor group approaches.

REFERENCES

1. *AASHTO Guidelines for Traffic Data Programs*. 2009. American Association of State Highway and Transportation Officials (AASHTO), 444 North Capitol Street, NW Washington, DC 20001 USA.
2. Zhong, M., Bagheri, E., and Christie, J. (2012). Improving Group Assignment and AADT Estimation Accuracy of Short-term Traffic Counts using Historical Seasonal Patterns & Bayesian Statistics. *8th International Conference on Traffic and Transportation Studies*, Changsha, China, August 1–3, 2012.
3. *Traffic Monitoring Guide*. Federal Highway Administration (FHWA). U.S. Department of Transportation, 2013. http://www.fhwa.dot.gov/policyinformation/tmguide. Accessed July 7, 2015.

4. Drusch, R. L. Estimating Annual Average Daily Traffic from Short- Term Traffic Counts. In *Highway Research Record 118*, HRB, National Research Council, Washington, D.C., 1966, pp. 85–95.

5. Gulati, B. M. (1995). “Precision of AADT estimates from short-period traffic counts.” MASc thesis, University of Regina, Regina, Sask., Canada.

6. Davis, G. A. (1996). ‘‘Estimation theory approaches to monitoring and updating average daily traffic.’’ Report No. 97-05. Center of Transportation Studies, University of Minnesota, Minneapolis.

7. Davis, G. A., Guan, Y. (1996). Bayesian assignment of coverage counts location to factor groups and estimation of mean daily traffic. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1542, Transportation Research Board, Washington, D.C., pp. 30–37.

8. Davis, G. A. (1997). ‘‘Accuracy of estimates of mean daily traffic: A review.’’ In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1593, Transportation Research Board, Washington, D.C., 1996, pp. 12–16.

9. Zhong, M., Bagheri, E., and Christie, J. (2012). Improving Group Assignment and AADT Estimation Accuracy of Short-term Traffic Counts using Historical Seasonal Patterns & Bayesian Statistics. *8th International Conference on Traffic and Transportation Studies*, Changsha, China, August 1–3, 2012.

10. Sharma, S.C., Werner, A., 1981. Improved method of grouping province-wide permanent traffic counters. Transportation Research Record 815. Transportation Research Board, National Research Council, Washington, DC, pp. 12–18.

11. Sharma, S. C., and Allipuram, R. R. (1993). ‘‘Duration and frequency of seasonal traffic counts.’’ *Journal of Transportation Engineering*, 119(3), pp: 344–359.

12. Lingras, P. J. (1995). ‘‘Classifying highways: Hierarchical grouping versus Kohonen neural networks.’’ *Journal of Transportation Engineering,* 121(4), pp: 364–368.

13. Lingras, P. (2001). Statistical and Genetic Algorithms Classification of Highways. Journal of Transportation Engineering, Vol. 127, No. 3, pp: 237-243.

14. Tsapakis, I., Schneider, W., Bolbol, A., Skarlatidou, A., 2011. Discriminant analysis for assigning short-term counts to seasonal adjustment factor groupings.Transportation Research Record, No. 2256, Transportation Research Board, National Research Council, Washington, DC, pp. 112–119.

15. Tsapakis, I., W. H. Schneider IV, A. P. Nichols, and J. Haworth. Alternatives in Assigning Short-term Counts to Seasonal Adjustment Factor Grouping, *Journal of Advanced Transportation*, Vol. 48, 2014, pp. 417-430.

16. Lingras, P. (1998) Traffic parameter estimation and highway classification: rough patterns using a neural networks approach, Transportation Planning and Technology, 21(3), pp: 155-179.

17. Caceres, N., Romero, L.M., Benitez, F.G., 2012. Estimating traffic flow profiles according to a relative attractiveness factor. Procedia of Social Behavioral Science, vol. 54, pp: 1115–1124.

18. Gecchele, G., Rossi, R., Gastaldi, M., Kikuchi, S., 2012. Advances in uncertainty treatment in the FHWA procedure for estimating annual average daily traffic volume. Transportation Research Record 2308. Transportation Research Board, National Research Council, Washington, DC, pp. 148–156.

19. Li, M. T., F. Zhao, and L. F. Chow. Assignment of Seasonal Factor Categories to Urban Coverage Count Stations Using a Fuzzy Decision Tree. *Journal of Transportation Engineering,* 132 (8), 2006, pp. 654-662.

20. Jin, L., C. Xu, and J. D. Fricker. Comparison of Annual Average Daily Traffic Estimates: Traditional Factor, Statistical, Artificial Neural Network, and Fuzzy Basis Neural Network Approach. *Transportation Research Board 87th Annual Meeting,* TRB, Washington D.C., 2008.

21. Lam, W., Tang, Y. F., Chan, K. S. and M. L. Tam. Short-term Hourly Traffic Forecasts using Hong Kong Annual Traffic Census. *Transportation*, 33 (3), 2006, pp. 291-310.

22. Tsapakis, I., and W.H.Schneider. Use of Support Vector Machines to Assign Short-Term Counts to Seasonal Adjustment Factor Groups. In *Transport Research Record: Journal of the Transportation Research Board, No. 2527*, Transportation Research Board, Washington, D.C., 2015, pp. 8-17.

23. Haykin, S. (2009). *Neural Networks and Learning Machines*. 3/e. Prentice Hall, Pearson Education, Inc., Upper Saddle River, New Jersey 07458.

24. Sharma, S., Lingras, P., Xu, F., & Kilburn, P. (2001). Application of neural networks to estimate AADT on low-volume roads. *Journal of Transportation Engineering*, 127(5), 426-432.

25. Gastaldi, M., G. Gecchele, and R. Rossi. Estimation of Annual Average Daily Traffic from One-week Traffic Counts - A Combined ANN-Fuzzy Approach. *Transportation Research: Part C*, vol. 47, 2014, PP. 86-99.

26. Malalur, S. S. and M. Manry. (2009). Feed-forward Network Training using Optimal Input Gains. *Proceedings of International Joint Conference on Neural Networks*, Atlanta, Georgia, USA, June 14-19, pp: 1953-1960.

27. Yu, C., M.T. Manry., and J. Li. (2005). "Effects of nonsingular pre-processing on feed-forward network training ". *International Journal of Pattern Recognition and Artificial Intelligence,* Vol. 19, No. 2, pp. 217-247.

28. Jesudhas, P., M.T. Manry, R.Rawat, and S.Malalur. (2011). Analysis and Improvement of Multiple Optimal Learning Factors for Feed-Forward Networks. *Proceedings of International Joint Conference on Neural Networks*, San Jose, California, USA, July 31 – August 5, 2011.

29. Malalur, S.S. and M. T. Manry. (2010). Multiple optimal learning factors for feed-forward networks, *Proc. of SPIE: Independent Component Analyses, Wavelets, Neural Networks, Biosystems, and Nanoengineering VIII*, Orlando, Florida, vol. 7703, pp. 77030F-1 - 77030F-12, April 7-9.

30. P. Werbos, "Beyond regression: New tools for prediction and analysis in the behavioral sciences," Ph.D. dissertation, Committee on Appl. Math., Harvard Univ., Cambridge, MA, Nov. 1974.

31. D.E. Rumelhart, G.E. Hinton, and R.J. Williams, "Learning internal representations by error propagation," in D.E. Rumelhart and J.L. McClelland (Eds.), *Parallel Distributed Processing*, Vol. I, Cambridge, Massachusetts: The MIT Press, 1986.

32. H-H Chen, M.T. Manry, and H. Chandrasekaran, "A Neural Network Training Algorithm utilizing Multiple Sets of Linear Equations," Neurocomputing, April 1999, vol. 25, pp. 55-72.

33. S. S. Malalur and M. T. Manry, "Multiple optimal learning factors for feed-forward networks," Proc. of IJCNN'09, Atlanta, Georgia, June 14-19, 2009, pp. 1953-1960.

34. Marquardt, D., "An algorithm for least-squares estimation of nonlinear parameters," *SIAM J. Appl. Math.*, vol. 11,431-441, (1963).

35. " Newton's Method in Optimization", *Wikipedia*, http://en.wikipedia.org/wiki/Newton%27s_method_in_optimization

36. Kearns, M., 1996. "A bound on the error of cross validation using the approximation and estimation rates, with consequences for the training-test split," *Advances in Neural Information Processing Systems*, vol. 8, pp. 183‑189, Cambridge, MA: MIT Press.

37. Prechelt, L., 1998. *Early Stopping—But When? in Neural Networks: Tricks of the Trade*, ed. G. Orr and K. Muller, Lecture Notes in Computer Science, no. 1524. Berlin: Springer, pp. 55‑69.

38. LeCun,Y., 1993. *Efficient Learning and Second-order Methods*, A Tutorial at NIPS 93, Denver.

39. J. Olvera, X. Guan and M.T. Manry, "Theory of Monomial Networks," *Proceedings of SINS'92*, Automation and Robotics Research Institute, Fort Worth, Texas, pp. 96-101, December 1992.

40. Kuhn, M. and K. Johnson. (2013). *Applied Predictive Modelling*. Springer, New York.

41. Japkowicz, N. and M. Shah. (2011). *Evaluating Learning Algorithms: A Classification Perspective*. Cambridge University Press, 32 Avenue of the Americas, New York, NY 10013-2473, USA.

42. Hand., D. J. and R. J. Till. (2001). A Simple Generalisation of the Area under the ROC Curve for Multiple Class Classification Problem. *Machine Learning*, 45: 171-186

43. Ou, G. and Y.L. Murphey. (2007). Multi-class Patterns Classification using Neural Networks. Pattern Recognition 40, pp. 4-18.

44. Hosmer., D.W., S. Lemeshow and R.X. Sturdivant. 2013. *Applied Logistic Regression*, 3/e, John Wiley & Sons, Inc., Hoboken, New Jersey.

45. Davis, J. and M. Goadrich. (2006). The Relationship between Precision-recall and ROC curves. In *Proceedings of the International Conference on Machine Learning*, pp. 233-240. Association for Computing Machinery, New York.

46. Bouckaert, R.R. (2003). Choosing between Two Learning Algorithms based on Calibrated Tests. In T.Fawcett and N. Mishra, editors, *Proceedings of the 20th International Conference on Machine Learning*. American Association for Artificial Intelligence, Menlo Park, CA.

47. Bouckaert, R.R. (2004). Estimating Replicability of Classifier Learning Experiments. In C. Brodley, editor, *Proceedings of the 21st International Conference on Machine Learning*. American Association for Artificial Intelligence, Menlo Park, CA.

48. Tsapakis, I., W. H. Schneider IV, and A. P. Nichols. Improving the Estimation of Total and Direction-based Heavy-Duty Vehicle Annual Average Daily Traffic, *Transportation Planning and Technology*, Vol. 34, Issue. 2, 2011, pp. 155-166.

49. Rossi, R., M. Gastaldi, and G. Gecchele. Comparison of Clustering Methods for Road Group Identification in FHWA Traffic Monitoring Approach: Effects on AADT Estimates. *Journal of Transportation Engineering*, Vol. 140, No. 7, 2014, 04014025.

50. Gadda, S. C., Magoon, A., and Kockelman, K. M. (2007). "Quantifying the uncertainty in annual average daily traffic (AADT) count estimates." Proc., Eleventh World Conf. on Transportation Research, World Conference on Transport Research Society, Lyon, France.

Chapter 4
Two-Step Process: Clustering and Classification

INTRODUCTION

Transportation system performance, not only depends on the current demands for travel, but also on the future demand that emanates from changes within the region. Transportation agencies deploy significant resources to monitor traffic patterns and evaluate the network capabilities and performance to facilitate efficient operations. The main data source for the traffic monitoring program comes from the installation of Automatic Traffic Recorders (ATRs) over the network. Due to cost and maintenance, agencies can only deploy the ATRs at a few strategic locations of the network. Agencies often use short-term counts, in lieu of ATRs, to estimate AADT values while saving resources. The Federal Highway Administration (FHWA) *Traffic Monitoring Guide* (TMG) outlines a traditional method to estimate the AADT from short-term counts (*1*). The traditional method has four steps: development of seasonal adjustment factors, grouping, assignment and AADT estimation. Previous studies focus more on either grouping or assignment steps. In addition, previous research also attempts to estimate the AADT directly (or bypassing the grouping and assignment steps) from the traffic data.

Background

Sharma *et al.* (*2, 3*) developed neural network models with a multilayered, feed-forward, and back-propagation design. The input for the neural network uses a ratio of hourly volumes to the sample average daily traffic and the network outputs a single factor (which is the ratio of daily traffic to the AADT). The study reported 95th percentile error values of about 25% under different short term counts and frequency. Lingras (*4*), Lingras and Sharma (*5*), and Xu (*6*) applied neural networks to classify the ATR sites into factor groups and estimated AADT from sample traffic counts. Sharma *et al.* (*7*) carried out a comparative analysis of the traditional factor approach and the neural network approach to estimate the AADT from 48-hour counts. The study obtained 95th percentile errors between 14.14% and 16.68% for different months and days. McCord *et al.* (*8*) used aerial photos and satellite images for estimating the AADT for a few highway segments in Ohio. Jiang (*9*) and Jiang *et al.* (*10*) also used image based traffic information together with traffic data for AADT estimation. Zhong *et al.* (*11*) developed models using neural network and regression analysis to evaluate missing traffic counts from permanent traffic stations. Fricker *et al.* (*12*) evaluated three different methods: neural networks, fuzzy logic, and analysis of variance (ANOVA) approach and compared the AADT estimates with the estimates from the traditional factor method. The results showed that all three approached performed better than the traditional factor method.

In addition to the FHWA traditional approach, some studies estimate AADT using alternative means from short-term counts. These methods use land use and socio-economic characteristics, demographics and historical counts to estimate the AADT values. For instance, multiple regression analysis (*13*-*16*), geographically weighted regression (GWR) technique by Zhao and Park (*17*), and Kriging-based methods (*18, 19*) represent some of the methods used to directly estimate the AADT. Duddu and Pulugurtha (*20*) use the principle of demographic gravitation and neural networks to estimate AADT values and obtain improved estimates compared to statistical methods and tradition four-step models. In addition, other reserchers propose Gaussian maximum likelihood methods (*21*), fuzzy decision trees (*22*), and the smoothly clipped absolute deviation penalty procedure (*23*).

Gastaldi *et al.* (*24*) use fuzzy c-mean clustering and neural networks to estimate the AADT values from a one-week seasonal traffic count of a road section and obtains less than 10 percent of error for commuter groups. Gecchele *et al.* (*25*) also uses fuzzy c-mean clustering and neural networks to estimate the AADT values but using 48-hour weekday counts. Their study obtains less mean absolute errors than neural networks (*7*) and linear discriminant analysis (*26*). Gadda *et al.* (*27*) evaluate and quantify the uncertainties in annual average daily traffic (AADT) from count data. Their study estimates the errors from extrapolating short-term counts to be on the order of 20% to 100% or even higher. Pulugurtha and Kusam (*28*) use multiple bandwidths from a highway segment and off-network characteristics, like demographic, socio-economic and land-use characteristics, to estimate AADT.

Purpose

Past studies suggest that direct estimation of AADT from traffic counts or using land-use, socio-economic and demographic factors appears attractive because of its resource effectiveness. Moreover, previous studies focus more on either grouping (clustering) traffic patterns or assignment (classification) steps. The FHWA two-stage AADT estimation approach, which performs both grouping and assignment, seems less frequently attempted. One reason points to the requirement of two different modeling approaches (clustering/grouping and classification/assignment) to estimate AADT. The objective of clustering and classification differs from direct AADT estimation. For instance, regression analysis minimizes the error between the observed and modeled AADT, but cluster analysis minimizes the within cluster distance and tries to maximize distance across different clusters. Classification only aims to minimize the percent of misclassification. However, the FHWA four step methodology (two-stage approach) remains the widely *recommended* method to estimate AADT from continuous count data (*1*). The two-step improvement process has the following objectives:

Objectives

- Present an improved two-step process for clustering using Gaussian Mixture Modelling (GMM) and classification using the OWO-Newton method (ONN)
- Evaluate the performance of the combined GMM clustering with ONN classification and the KM and HCA clustering solutions with Quadratic Discriminant Analysis (QDA)
- Perform comparative analysis with the Oregon DOT seasonal trend grouping and traditional functional class grouping
- Assess the accuracy in AADT estimation errors

Chapter Organization

The remainder of this chapter starts with a layout of the steps involved in the study methodology. The next section presents the results from the clustering step followed by a brief discussion of the classification methods. The last section presents the AADT estimation error analysis between the proposed and traditional methods. In addition, the paper presents statistical tests on the performance difference between the methods.

CONTRIBUTION

The study proposed a two-step improvement framework for both grouping and assignment of short-term traffic counts. The Gaussian Mixture Modelling (GMM) for clustering and OWO-Newton (ONN) neural network method of assignment produce small errors than other clustering methods. The GMM-ONN method consistently produces lower errors and an error difference of 6% to 27% percent. The study successfully shows that the possible reduction of errors at both clustering and classification steps will enhance the AADT estimates.

METHODOLOGY

The assignment methodology involves four steps. The first stage groups the 24-hour traffic patterns based on similarity using Gaussian Mixture Modeling (GMM), k-means (KM) and the agglomerative Hierarchical Clustering Analysis (HCA). The next stage computes the Seasonal Adjustment Factors (SAFs) because they must be used to estimate the AADT values. The third step presents classification methodologies. The two-stage framework relies heavily on clustering and classification steps presented in chapter 2 and 3 of this report. The final stage deploys the combined methods (clustering and classification) on test datasets to estimate the AADT values from short-term counts. The error analysis explores the variation of the AADT estimates by month, day-of-week and highway functional class.

Case Study Data

The data sets are developed from ATR counters throughout the State of Oregon for the years 2011 and 2012. The 2011 data is used only for clustering and design of a classifier. The relative performance between combined methods is tested with the 2012 data. Chapter 2 provides more explanatory analysis of the case study data.

STEP 1: CLUSTERING

The clustering analysis uses the 2011 data with a sample size of 30,393. Sample hourly traffic data (with dimensionality d = 24) represents the input data for the clustering. The study evaluates different models with clusters between 2 and 30 to select a best one. The model-based clustering chooses fifteen mixture components using BIC criterion. The K-means clustering solution produces six clusters and agglomerative Hierarchical Clustering Analysis (HCA) selects a three cluster solution. Chapter 2 provides a detailed analysis of clustering process.

STEP 2: SEASONAL ADJUSTMENT FACTORS

After the clustering step, each cluster group is again sub-grouped by month to compute the average seasonal adjustment factors (SAFs). The seasonal factors are computed by averaging the ratios of annual average daily traffic (AADT) to daily traffic (DT) in a given month. Subgrouping helps to address monthly and seasonal variation of traffic data (see Table 2-8 to Table 2-12 in Chapter 2).

STEP 3: CLASSIFICATION

The cluster assignment process represents the most critical step in the AADT estimation process. The study adopts quadratic discriminant analysis (QDA) and neural network (ONN) classification methods and creates an improved strategy for assigning traffic patterns. The higher accuracy and better classification capabilities reflect an effective approach for labeling the unknown traffic patterns. The statistically significant error reduction for the ONN and QDA classification, based on the GMM clustering solution, shows a reliable approach for AADT estimation. Hence, the study uses the ONN and QDA classification in the combined method. Chapter 3 provides more details on classifiers' evaluation.

STEP 4: AADT ESTIMATION

This section tests the performance of the AADT estimation using a combined GMM and ONN approach versus other clustering methods with QDA classification. In addition, the study evaluates two default methods of estimation, the ODOT and FC grouping methods.

Monthly Variation of Errors

The GMM-ONN method produces less monthly average errors consistently when compared to the HCA-QDA and KM-QDA solution. The errors remain less than the ODOT method except for July where a 0.2% overestimation error occurs. The FC groups always produce more error than all other AADT estimation methods. Except for July and August, the KM-QDA and HCA-QDA methods produce less error than the ODOT seasonal trend grouping. Figure 4-1 shows the monthly variation of the errors for the different methods.

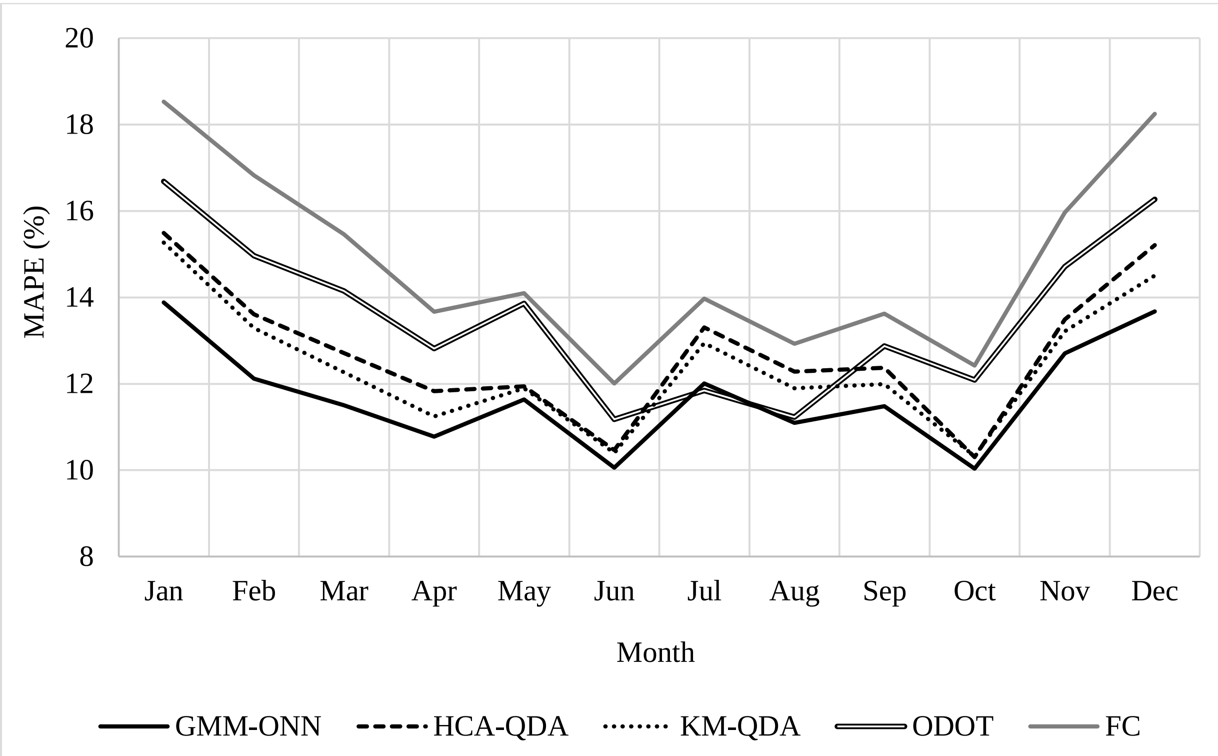


Figure 4-1 Monthly Variation of MAPE by Estimation Method

All methods exhibit more variation during winter months (December and January). The HCA-QDA method produces a larger standard deviation of errors compared to the other clustering methods. The GMM-ONN solution has less deviation than the ODOT method except for the months between July and September. In addition, the GMM-ONN method produces less variation compared to the HCA and KM methods with QDA classification. Figure 4-2 shows the standard deviation of errors among the methods.

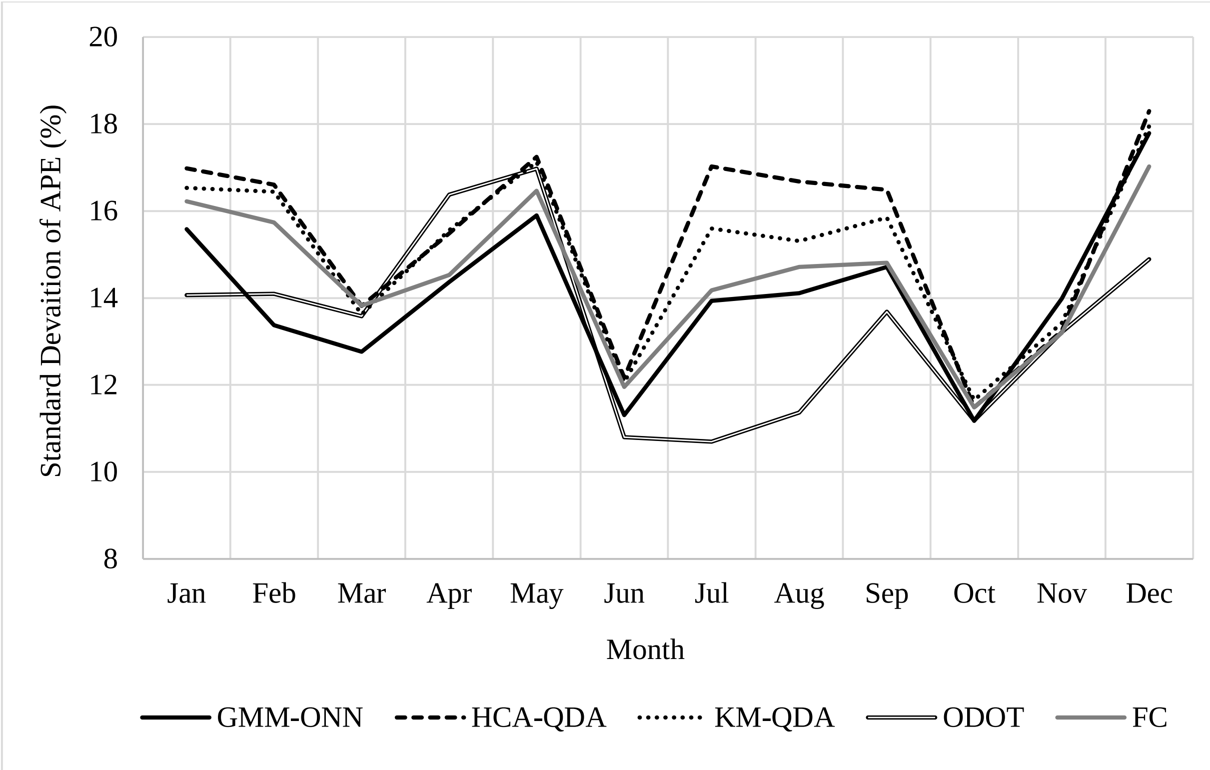


Figure 4-2 Standard Deviation of Errors by Estimation Method

Error Distribution by Size

The GMM-ONN method has the highest proportion of patterns with a MAPE value of less than five percent (Table 4-1). As the error size increases, the proportion of patterns decreases as expected. The distribution of patterns across multiple error sizes shows that the GMM-ONN performs better than the HCA-QDA and KM-QDA methods. In addition, the GMM-ONN approach exhibits less error compared to the ODOT and FC grouping. More than a third of patterns produce error values greater than 15 percent for the FC grouping. For the error size beyond 50 percent, the GMM-ONN and the ODOT show fewer patterns than other methods. A few patterns with a large error size may affect the average error value.

Table 4-1 Error Distribution by Size

Estimation Method	Percentage Patterns with a MAPE size of (%)								
	0-5	5-10	10-15	15-20	20-25	25-50	50-100	>100	>15
GMM-ONN	38.2	22.6	13.6	8.2	5.4	10.0	1.9	0.2	25.7
HCA-QDA	35.3	23.1	14.0	9.1	5.5	10.5	2.2	0.4	27.6
KM-QDA	34.7	22.3	14.5	9.2	6.4	10.3	2.3	0.4	28.6
ODOT	25.9	23.8	17.5	11.7	7.4	11.6	1.9	0.2	32.8
FC	23.7	22.3	17.0	12.3	8.0	14.2	2.2	0.3	37.0

Analysis of MAPE >15%

Figure 4-3 shows the distribution of MAPE values with error size greater than 15%. Each radial line represents a month and concentric circles show the proportion of patterns having a MAPE value greater than 15 percent. The monthly distribution of patterns shows relative similar performance among the methods. Both winter and summer months show a larger concentration of MAPE values greater than fifteen percent.

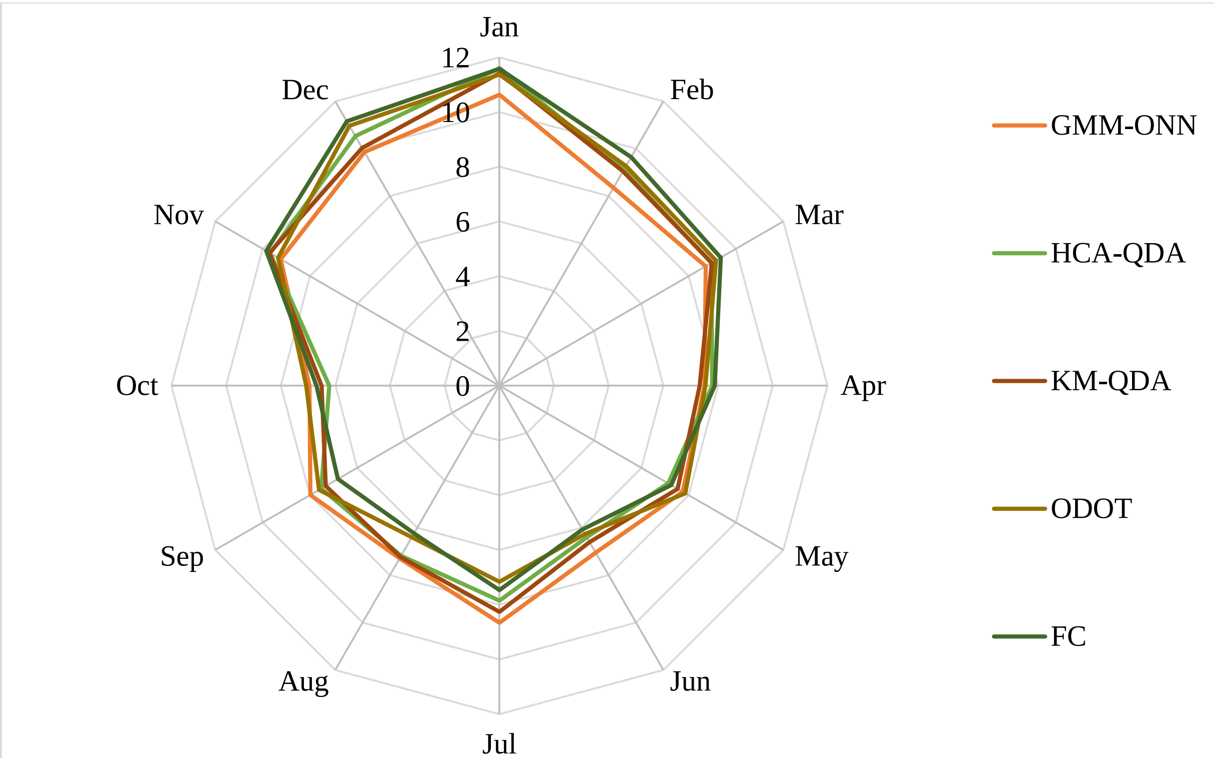


Figure 4-3 Monthly Error Distribution for Patterns with MAPE > 15%

MAPE during Summer Months

The clustering methods exhibit larger deviation of errors from the mean during summer months (July to September). Hence, the study analyzes the distribution of errors by size for these months (see Table 4-2). The GMM-ONN method, for all three months, produces more patterns with error less than five percent. The KM-QDA and HCA-QDA methods comparatively produce more patterns with larger errors than the GMM-ONN method. The ODOT method produces more patterns with larger errors than the three clustering techniques. The low number of patterns for errors greater than 25 percent may affect (or reduce) the monthly average of the MAPE in July. Moreover, the ODOT method and the GMM-ONN method only differ by two percent of the patterns with more than the 15 percent error benchmark.

Figure 4-4 MAPE between the GMM-ONN and ODOT methods during July-September compares the error distribution between the GMM-ONN method and the ODOT method for all three months.

Table 4-2 MAPE during Summer Months (July-September)

Month	Method	Percentage Patterns with a MAPE size of (%)								
		0-5	5-10	10-15	15-20	20-25	25-50	50-100	>100	>15
July	GMM-ONN	37.9	22.1	13.3	7.9	5.3	11.5	1.9	0.1	26.7
	KM-QDA	34.3	25.5	12.9	8.2	4.4	11.8	2.4	0.5	27.3
	HCA-QDA	33.6	24.0	15.5	7.6	5.2	11.1	2.4	0.6	26.9
	ODOT	27.1	27.2	17.6	11.5	7.2	8.2	1.1	0.0	28.1
	FC	25.6	25.3	15.9	11.7	6.9	11.8	2.5	0.3	33.1
August	GMM-ONN	41.5	22.2	14.0	7.3	4.1	8.9	1.9	0.2	22.5
	KM-QDA	37.6	24.3	14.1	8.1	4.2	8.7	2.5	0.5	23.9
	HCA-QDA	38.1	23.1	14.2	8.1	4.5	9.1	2.3	0.7	24.6
	ODOT	32.6	26.3	15.9	9.8	6.3	7.6	1.3	0.1	25.1
	FC	30.4	25.0	16.9	10.0	4.9	9.5	2.9	0.3	27.6
September	GMM-ONN	39.6	20.6	12.1	8.7	5.1	8.8	1.7	0.4	24.6
	KM-QDA	38.9	20.9	12.6	7.3	4.7	9.5	2.2	0.6	24.3
	HCA-QDA	37.0	21.8	12.1	8.1	5.4	9.5	2.0	0.7	25.7
	ODOT	27.3	22.6	16.9	11.7	6.8	9.6	1.5	0.3	29.9
	FC	26.0	23.7	16.7	10.6	5.9	11.6	1.6	0.4	30.2

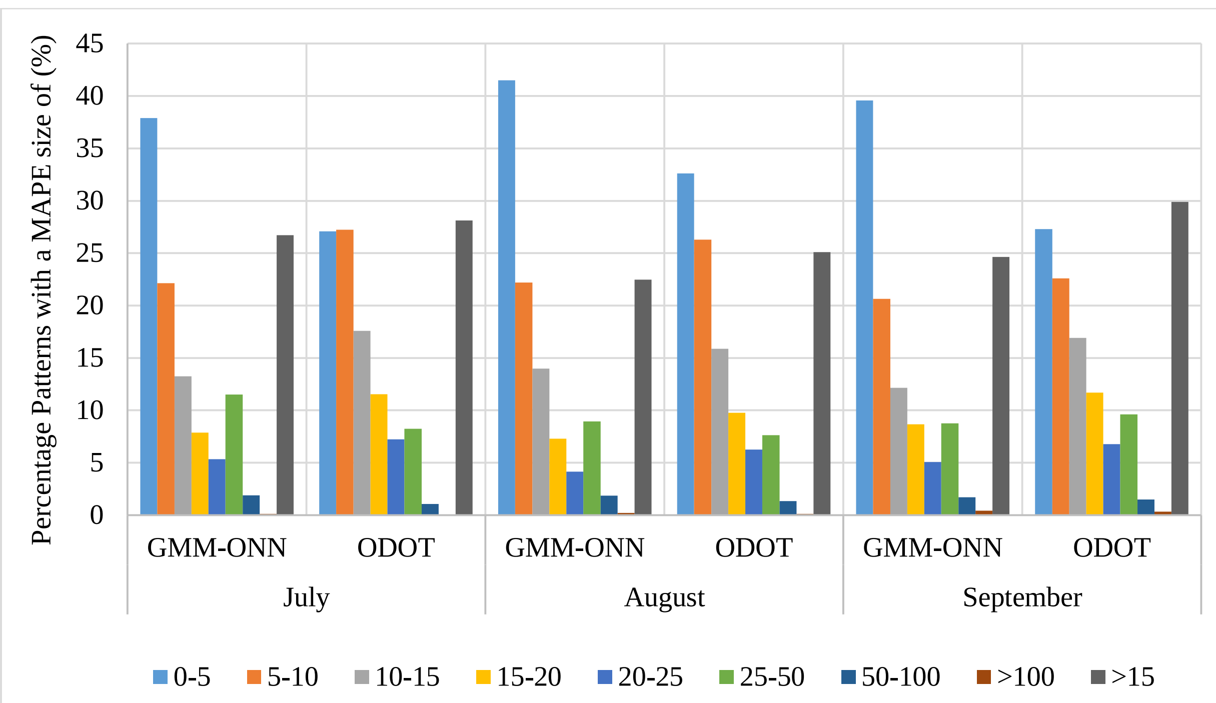


Figure 4-4 MAPE between the GMM-ONN and ODOT methods during July-September

Error Summary Statistics

The quartile distribution of absolute percent errors is shown in Table 4-3. The GMM-ONN method produces lower mean error values compared to the HCA-QDA and KM-QDA methods. Moreover, the error rate remains lower than the traditional ODOT and FC techniques; however, the ODOT approach has a lower standard deviation of errors. The standard deviation of the HCA-QDA and KM-QDA methods remain larger than the GMM-ONN method. The 95 percentile of the confidence interval is also listed in Table 4-3.

Table 4-3 Error Summary Statistics by Estimation Methods

Method	1st Quartile	Median	3rd Quartile	Mean	SD	95% CI of Mean	
						lower	upper
GMM-ONN	3.0	7.3	15.3	11.8	14.3	11.6	11.9
HCA-QDA	3.4	8.2	16.7	12.8	15.7	12.6	12.9
KM-QDA	3.3	7.9	16.2	12.5	15.3	12.3	12.6
ODOT	4.8	10.0	18.2	13.6	13.7	13.4	13.7
FC	5.3	11.0	19.8	14.8	14.8	14.7	15.0

Error by Functional Class

Figure 4-5 shows the error variation for different methods across highway functional classes. In general, the GMM-ONN method produces fewer errors than the KM-QDA and HCA-QDA methods. In addition, the GMM-ONN method usually outperforms both the ODOT and FC methods. Notably, the major collectors carry higher error rates for all methods, and the GMM-ONN method experiences larger estimation errors than the strategies other than the FC method. The GMM-ONN method produces less than five percent error for urban interstates and less than ten percent error for urban arterials and freeways. The KM-QDA and HCA-QDA methods also follow a similar trend for these facilities. The study methods also meet the FHWA recommended AADT forecasting error of less than 10 percent for commuter patterns. In contrast, the traditional methods, the ODOT and FC grouping, show larger than ten percent errors for commuter type facilities. Moreover, all methods produce an error of less than 15 percent for interstates and arterials. The rural interstates experience almost double the error rate compared to the urban interstate when using the GMM-ONN, HCA-QDA and KM-QDA methods. However, the ODOT and FC methods produce almost equal errors for both rural and urban interstates.

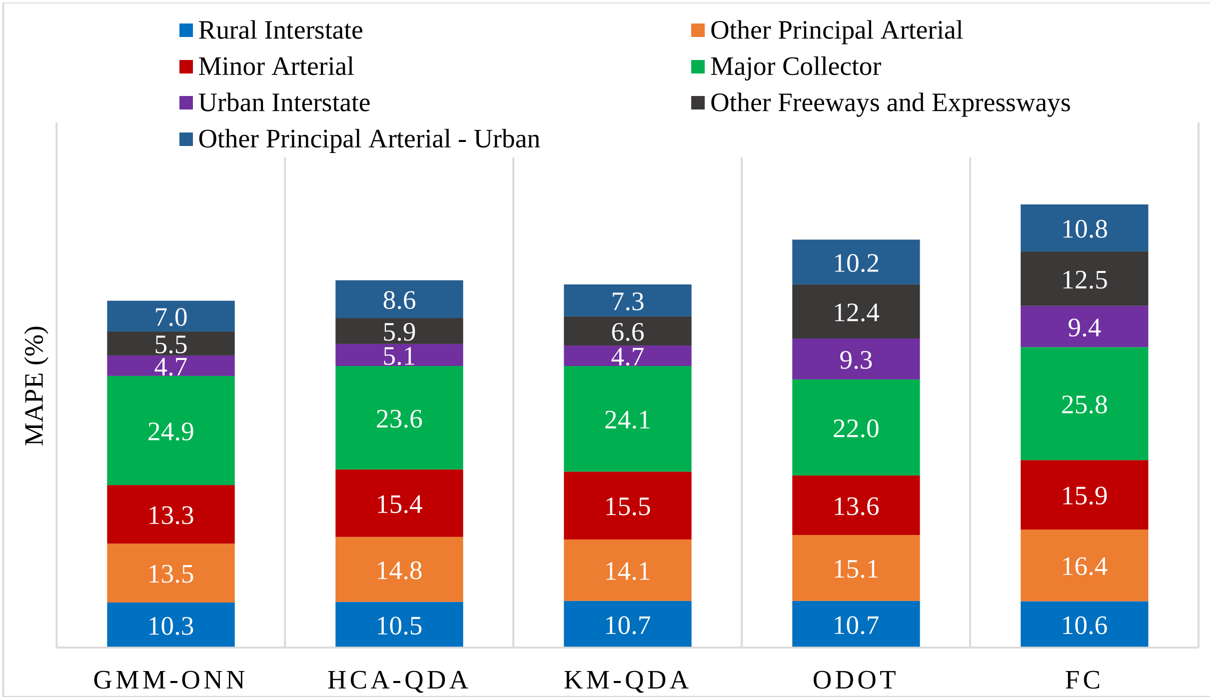


Figure 4-5 Error by Highway Functional Class

Statistical Significance Test

The study conducts a one-tailed paired t-test between the GMM-ONN method and the errors produced by other methods. The null hypothesis assumes no difference between the errors produced by the GMM-ONN and other methods. Table 4-4 shows the results of the t-tests conducted on the 2012 data sets. The GMM-ONN method consistently produces lower errors and an error difference of -6% to -27% percent. All error differences remain statistically significant at a five percent significance level.

Table 4-4 Statistical Significance Tests for Error Difference

Comparison	t statistic	*p*-value	95 % CI of Mean Error Difference		Mean Error Difference	Percentage Error Reduction (%)
			lower	upper		
GMM-ONN vs HCA-QDA	-21.42	2.20E-16	-1.04	-0.87	-0.96	-8.5
GMM-ONN vs KM-QDA	-15.91	2.20E-16	-0.72	-0.56	-0.64	-5.9
GMM-ONN vs ODOT	-29.29	2.20E-16	-1.90	-1.66	-1.78	-15.4
GMM-ONN vs FC	-54.03	2.20E-16	-3.13	-2.91	-3.02	-26.1

COMPARISON WITH OTHER METHODS

The study compares the mean error estimates with other methods that use 24-hour counts. Gadda *et al.* (*27*) obtain average errors between 12.5% and 13% from 24-hour counts from the Minnesota DOT and

Florida DOT. In addition, their weekday arterial and freeway AADT estimates show error values of 10 to 15%. Battelle (*29*) uses Federal Highway Administration (FHWA) Travel Monitoring Analysis System (TMAS) data from thirteen years (2000-2012) for analyzing a national summary of AADT estimation errors from multiple short duration counts. The study uses 24-hour traffic data from 43,000 ATR sites covering many DOTs (see Figure 4-6). The study uses highway functional class grouping, volume factor grouping, and k-means clustering method and assesses the errors by day-of-week, month-of-year and yearly basis. The study obtains 95% confidence intervals for the national average weekday errors (*29*): -17.4 to 19.0% (clustering), -25.6 to 31.5% (volume factor grouping) and -24.9 to 30.5% (highway functional class grouping). The proposed GMM-ONN method obtains 95% confidence intervals for error between 11.6% and 11.9%.

FC	State	Year	Number of Sites per Group		
			Complete	AADT	Total
1R	Indiana	2007	1	14	15
	Pennsylvania	2004	3	5	8
1U	Hawaii	2010	3	9	12
	Massachusetts	2000	9	1	10
	Michigan	2006	3	11	14
	Minnesota	2008	10	4	14
	Wisconsin	2009	1	13	14
2U	California	2003	3	10	13
	New York	2006	3	12	15
	Oklahoma	2008	1	7	8
3R	Idaho	2004	7	8	15
	Kansas	2004	2	13	15
	Kentucky	2003	4	11	15
	Maine	2008	15	0	15
	Mississippi	2008	1	10	11
	Nebraska	2008	7	8	15
	New Mexico	2005	2	7	9
	Washington	2005	8	6	14
	Wyoming	2000	3	12	15
3U	Alabama	2000	2	7	9
	California	2003	4	11	15
	Florida	2005	1	14	15
	Illinois	2011	2	9	11
	Nevada	2008	1	9	10
	Utah	2005	9	2	11
	Wisconsin	2010	2	13	15
	Wyoming	2000	4	11	15

FC	State	Year	Number of Sites per Group		
			Complete	AADT	Total
4R	Colorado	2011	2	7	9
	Florida	2006	1	14	15
	Michigan	2003	2	8	10
	Montana	2006	1	12	13
	Oregon	2005	9	6	15
	Pennsylvania	2011	6	5	11
	Texas	2002	15	0	15
	Vermont	2011	3	9	12
4U	Alaska	2010	5	10	15
	Iowa	2004	7	0	7
	Nevada	2008	3	6	9
5R	Kansas	2009	2	8	10
	Maine	2011	15	0	15
	Ohio	2005	2	7	9
	Texas	2001	11	4	15
	Vermont	2012	9	6	15
5U	Kentucky	2006	2	7	9
Grand Total			206	346	552

Figure 4-6 Summary of Sites Selected for Nationwide Study on Assessing AADT Accuracy from Short-Term Count Durations (source: Battelle (*29*))

Sharma *et al.* (*2, 3*) report 95th percentile error values of about 25% under different short term counts and frequency. Sharma *et al.* (*7*) obtain 95th percentile errors between 14.14% and 16.68% for different months and days from 48-hour counts. The GMM-ONN obtains 54 percent error reduction compared to the national average. Moreover, the proposed method shows 6 to 10 percent reduction than methods using Minnesota and Florida DOT datasets. However, the GMM-ONN reduces error by 31 percent compared to direct estimation method using neural networks. Hence, the proposed GMM-ONN method reflects a better alternative to enhance the accuracy of AADT estimation from short-term counts.

SUMMARY

This chapter presents the performance evaluation of a combined clustering and classification framework (the GMM clustering with ONN classification) and other clustering methods with QDA classification. Moreover, the evaluation includes the ODOT and FC grouping methods to assess the merits of the proposed two-step approach. The study uses the GMM, KM, and HCA methods for clustering. Based on these clustering solutions the study designs different classifiers. First, the clustering solution using 24-hour traffic patterns is developed from the 2011 datasets. The obtained clustering solution, split into training and testing sets, becomes an input for training the classifiers. The authors use the trained classifiers to classify the 2012 test datasets. Next, the study obtains the MAPE between the estimated AADT and observed AADT values. The GMM-ONN method produces lower monthly average errors compared to the HCA-QDA, KM-QDA, and the ODOT solutions. The FC groups always produce more error than all other AADT estimation methods. All methods exhibit more error variation during winter months. The GMM-ONN produces fewer error variations compared to the HCA and KM methods with QDA classification. The distribution of patterns across multiple error sizes shows the GMM-ONN outperforms the HCA-QDA and KM-QDA methods. The GMM-ONN method meets the FHWA recommended AADT forecasting error of less than 10 percent for commuter patterns.

CONCLUSION

The proposed two-step approach exhibits improved performance. The frameworks provide a better way for both grouping and assignment of 24-hour traffic patterns. The errors produced in the clustering step carry forward during the assignment and impact the forecasting error; however, the reduction of errors at both the clustering and classification steps will enhance the AADT estimates. The proposed GMM-ONN methodology produces less error than other clustering methods and traditional approaches with an error difference of 6% to 27%, which is statistically significant at a five percent level. The GMM-ONN method produces less than five percent error for urban interstates and less than ten percent for urban arterials and freeways. The study method meets the FHWA recommended AADT forecasting error of less than ten percent for commuter patterns. The GMM-ONN solution has less variation than other methods for most of the months. The GMM-ONN approach provides an effective and reliable alternative for estimating the

AADT using short-term traffic counts. The GMM-ONN produces less error compared to the national average and studies based on Minnesota and Florida DOT count data.

REFERENCES

1. *Traffic Monitoring Guide*. Federal Highway Administration (FHWA). U.S. Department of Transportation, 2013. http://www.fhwa.dot.gov/policyinformation/tmguide. Accessed July 7, 2015.
2. Sharma, S. C., Lingras, P., Liu, G. X., and Xu, F. (2000). "Estimation of annual average daily traffic on low-volume roads factor approach versus neural networks." Transportation Research Record 1719, Transportation Research Board, Washington, DC, 103–111.
3. Sharma, S. C., P. Lingras, F. Xu, and P. Kilburn. (2001) Application Of Neural Networks To Estimate AADT On Low-Volume Roads. Journal of Transportation Engineering, Vol. 127, No. 5, pp. 426–432.
4. Lingras, P. J. (1995). ''Classifying highways: Hierarchical grouping versus Kohonen neural networks.'' *Journal of Transportation Engineering,* 121(4), pp: 364–368.
5. Lingras, P., and Sharma, S. C. (1997). ''Determination of traffic volume parameters using neural networks.'' *Proc., 5th Int. Conf. on Comp. in Urban Plng. and Mgmt.*, 16–19.
6. Xu, F. (1998). ''Estimation of annual average daily traffic using the traditional and neural network methods.'' MASc thesis, University of Regina, Regina, Saskatchewan, Canada.
7. Sharma, S., Lingras, P., Xu, F., & Liu, G., 1999. Neural networks as alternative to traditional factor approach to annual average daily traffic estimation from traffic counts. Transportation Research Record 1660. Transportation Research Board, National Research Council, Washington, DC, pp. 24–31.
8. McCord, M., Yang, Y., Jiang, Z., Coifman, B., and Goel, P. (2003). "Estimating AADT from satellite imagery and air photos: Empirical results." Transportation Research Record 1855, Transportation Research Board, Washington, DC, 136–142.
9. Jiang, Z. (2005). "Incorporating image-based data in AADT estimation Methodology and numerical investigation of increased accuracy." Ph.D. dissertation, The Ohio State Univ., Columbus, OH.
10. Jiang, Z., McCord, M. R., and Goel, P. K. (2006). "Improved AADT estimation by combining information in image and ground-based traffic data." J. Transp. Eng., 132(7), 523–600.
11. Zhong, M., Lingras, P., and Sharma, S. (2004). "Estimation of missing traffic counts using factor, genetic, neural, and regression techniques." Transport. Res. C Emerg. Tech., 12(2), 139–166.
12. Fricker, J. D., Xu, C., and Jin, L. (2008). "Comparison of annual average daily traffic estimates: Traditional factor, statistical, artificial neural network, and fuzzy basis neural network

approach." Proc., Transportation Research Board 87th Annual Meeting, Transportation Research Board, Washington, DC.

13. Mohamad, D., K. C. Sinha, T. Kuczec, and C. F. Scholer. Annual average daily traffic prediction model for county roads. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1617, Transportation Research Board of the National Academies, Washington, D.C., 1998, pp. 69-77.
14. Xia, Q., Zhao, F., Chen, Z., Shen, L. D., and Ospina, D. (1999). "Development of a regression model for estimating AADT in a Florida County." Transportation Research Record 1660, Transportation Research Board, 32–40.
15. Zhao, F., and Chung, S. (2001). "Contributing factors of annual average daily traffic in a Florida county." Transportation Research Record 1769, Transportation Research Board, Washington, DC, 113–122.
16. Seaver, W. L., A. Chatterjee, and M. L. Seaver. Estimation of traffic volume on rural local roads. *Transportation Research Record*, Journal of the Transportation Research Board, Transportation Research Board of the National Academies, Washington, D.C., No. 719, 2000, pp. 121-128.
17. Zhao, F., and Park, N. (2004). "Using geographically weighted regression models to estimate Annual Average Daily Traffic." Transportation Research Record 1879, Transportation Research Board, Washington, DC, 99–107.
18. Wang, X., and Kockelman, K. M. (2009). "Forecasting network data: Spatial interpolation of traffic counts using Texas data." Transportation Research Record 2105, Transportation Research Board, Washington, DC, 100–108.
19. Selby, B., and Kockelman, K. M. (2011). "Spatial prediction of AADT in unmeasured locations by universal Krigging." Proc., Transportation Research Board 90th Annual Meeting, Transportation Research Board, Washington, DC.
20. Duddu, V. R., and S. S. Pulugurtha. Principle of Demographic Gravitation to Estimate Annual Average Daily Traffic: Comparison of Statistical and Neural Network Models. *Journal of Transportation Engineering*, Vol. 139, No. 6, 2013, pp. 585–595.
21. Tang, Y. F., Lam, W. H. K., and Pan, L. P. (2003). "Comparison of four modeling techniques for short-term AADT forecasting in Hong Kong." J. Transp. Eng., 129(3), 223–329.
22. Li, M. T., F. Zhao, and L. F. Chow. Assignment of Seasonal Factor Categories to Urban Coverage Count Stations Using a Fuzzy Decision Tree. *Journal of Transportation Engineering* 132 (8), 2006, pp. 654-662.
23. Yang, B., Wang, S., and Bao, Y. (2011). "Efficient local AADT estimation via SCAD variable selection based on regression models." Proc., Control and Decision Conference (CCDC), IEEE, New York, 1898–1902.

24. Gastaldi, M., R. Rossi, G. Gecchele, and L. Della Lucia. Annual Average Daily Traffic Estimation from Seasonal Traffic Counts. *Procedia Social and Behavioral Sciences*, vol. 87, 2013, pp. 279–291.

25. Gecchele, G., Rossi, R., Gastaldi, M., Kikuchi, S., 2012. Advances in uncertainty treatment in the FHWA procedure for estimating annual average daily traffic volume. Transportation Research Record 2308. Transportation Research Board, National Research Council, Washington, DC, pp. 148–156.

26. Tsapakis, I., Schneider, W., Bolbol, A., Skarlatidou, A., 2011. Discriminant analysis for assigning short-term counts to seasonal adjustment factor groupings.Transportation Research Record 2256. Transportation Research Board, National Research Council, Washington, DC, pp. 112–119.

27. Gadda, S. C., Magoon, A., and Kockelman, K. M. (2007). "Quantifying the uncertainty in annual average daily traffic (AADT) count estimates." Proc., Eleventh World Conf. on Transportation Research, World Conference on Transport Research Society, Lyon, France.

28. Pulugurtha, S. S., and Kusam, P. R. (2012). "Modeling AADT using integrated spatial data from multiple network buffer bandwidths." Proc., Transportation Research Board 91st Annual Meeting, Transportation Research Board, Washington, DC.

29. Battelle (Robert Krile, Fred Todt and Jeremy Schroeder). 2015. *Assessing Roadway Traffic Count Duration and Frequency Impacts on Annual Average Daily Traffic Estimation Assessing AADT Accuracy Issues Related to Short-Term Count Durations.* http://www.fhwa.dot.gov/policyinformation/presentations/his2015/AADT_Task_3_Report_Jessberger.pdf Accessed on August 29, 2016.

30. Davis, G. A., Guan, Y. (1996). Bayesian assignment of coverage counts location to factor groups and estimation of mean daily traffic. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1542, Transportation Research Board, Washington, D.C., pp. 30–37.

31. Tsapakis, I., and W.H.Schneider. Use of Support Vector Machines to Assign Short-Term Counts to Seasonal Adjustment Factor Groups. In *Transport Research Record: Journal of the Transportation Research Board, No. 2527*, Transportation Research Board, Washington, D.C., 2015, pp. 8-17.

Chapter 5 Conclusions and Future Directions

The deployment of Automatic Traffic Recorders (ATR) on a state-wide network helps state DOTs (Department of Transportation) to collect and monitor traffic patterns. The cost and maintainability restrict ATR deployment to limited strategic locations on the highway network. In lieu of ATRs, the agencies commit to Short-term Traffic Counts (STTCs). In order to cover specific locations of interest, short-term traffic counts are taken and seasonally adjusted using factor groups. The present study addresses the issue of accuracy from short-term counts and obtains improved accuracy over the traditional method. The study tries to enhance the AADT accuracy using improvements from both grouping and classification stage.

The study addresses key questions on performance, stability, and variability of clustering when grouping the traffic patterns in chapter two. The study proposes the GMM framework that provides a statistical inference for the obtained clusters. The study makes an effort to formally and qualitatively label the clustering groups using traffic pattern characteristics. The study introduces cluster-wise stability assessment for the clustering solutions using different resampling methods. In addition, the study performed missing value analysis and variable size analysis for clustering solutions. The study also conducts a comparative analysis of GMM clustering with other clustering methods.

The study introduces, in chapter three, two improved classification algorithms, the OWO-BP and OWO-Newton (ONN) methods, and obtains better performance than the regular BP algorithm. The author also introduces a fully connected network and a non-heuristic optimal learning factor to improve the network performance by optimally adjusting the learning factors between iterations. Moreover, the network selection process helps to find the best value for the number of hidden units in a hidden layer as opposed to the typical heuristic approach of selecting units. The study tests and evaluates the developed ONN framework using multiple performance measures. In addition, the study introduces a 10×10 stratified cross validation approach for performance evaluation of classifiers to minimize the bias. Later, the study presents the classification performance of each cluster for the GMM solution.

The study utilized two-step improvement framework for estimating AADT from short-term counts. The Gaussian Mixture Modelling (GMM) for clustering and OWO-Newton (ONN) neural network method of assignment produce small errors than other clustering methods. The study successfully shows that the possible reduction of errors at both clustering and classification steps will enhance the AADT estimates.

CONCLUSIONS

The study uses 24-hour traffic patterns of all vehicles in a clustering process to develop characteristic patterns of a traffic counts program. The GMM clustering selects fifteen clusters, K-means (KM) obtains six, and Hierarchical Clustering Analysis (HCA) identifies three clusters. The GMM solution provides an error reduction of 6% to 26%, which is statistically significant at a five percent level, over the other clustering approaches and traditional methods.

The study introduces several enhancements to an ANN to create a modified neural network model, OWO-Newton (ONN). The author uses this model to classify 24-hour traffic counts into particular clusters. The research assesses the ONN model's performance as a classifier versus other potential classifiers (discriminant analysis and Naïve Bayes (NB)). The ONN makes a correct prediction for 86 percent of the patterns, and it produces less error, a reduction of 15% to 38%, compared to most of the other classifiers. ONN and QDA provide similar performance, but QDA has numerous underlying assumptions that must be verified before its use as a classifier. While this level of attention to detail may be reasonable for research, DOTs appear less likely to invest the resources to ensure the validity of the model when the easier to apply ONN offers similar overall performance. The innovative methodology proposed in this study shows a significant improvement to the AADT accuracy when compared to alternative techniques.

The errors produced in the clustering step will carry forward during the assignment and affect the forecasting error. However, the possible reduction of errors at both the clustering and classification steps provides an opportunity to enhance the AADT estimates. Hence, the study uses combined clustering and classification, a two-step approach, for AADT estimation. The two-step framework provides a better approach for both grouping traffic patterns and assigning 24-hour traffic patterns to a cluster of similar patterns. The proposed methodology produces less error than other methods and traditional approaches. The comparison indicates an error difference of 6% to 27%, which is statistically significant at 5 percent level. The GMM-ONN method produces less than five percent error for urban interstates and less than ten percent for urban arterials and freeways. The study method meets the FHWA recommended forecasting error of less than ten percent for commuter patterns. The solution has less deviation of error compared to the other methods for most of the months. The GMM-ONN approach provides an effective and reliable technique for estimating the AADT using short-term traffic counts. Moreover, the GMM-ONN produces less error compared to the national average and studies based on Minnesota and Florida DOT count data. The study recommends the improved two-step process due to its accuracy, economical approach, and suitability to meet the agencies' need for a low-cost traffic counting program.

PRACTICAL IMPLICATIONS

The proposed GMM-ONN method provides a better way of grouping traffic patterns and assigning STTCs to traffic patterns. The lower AADT estimation errors and their distribution show an effective and reliable approach for AADT estimation using short-term traffic counts. The mean error rate of less than ten percent for commuter routes and lower standard deviation of errors show the satisfactory accuracy of the AADT estimates. The uses of daily patterns reflect economical approach and suit the agency's need for the low-cost traffic counting program. The knowledge on traffic patterns, its temporal variation, and the accuracy of AADT estimation is of a particular interest to transportation agencies for planning traffic count programs and allocating its resources in a cost effective way. The proposed method not only minimizes judgment errors but also supplements the FHWA guidelines on recommending clustering techniques when estimating AADT from short-term counts. The GMM-ONN method can distinguish less apparent traffic patterns and minimizes subjectivity and bias when grouping patterns. In addition, the proposed method requires fewer personnel and time resources compares to the ODOT seasonal trend grouping.

LIMITATIONS AND FUTURE DIRECTIONS

The study uses 24-hour daily traffic patterns for clustering using the GMM, KM, and HCA. However, the variable size analysis shows that K-means clustering for full data sets and either 16-hour or 12-hour datasets does not differ much. The HCA method produces different clustering solutions for 24-hour and 16-hour/ 12-hour data sets. The reduced data sets have eight cluster solutions for the HCA method. The clustering solution and error analysis using full and reduced datasets, and evaluation of both sets may provide more insight into the usability of reduced hourly patterns for estimating AADT using STTCs. The recommendation of reduced hours, if proven to produce a low error rate, for collecting STTCs has more implications for agencies in planning, scheduling and resource allocation.

Instead of daily patterns, a separate analysis for weekdays and weekends may be strategically helpful to stage the traffic counts. The MAPE monthly variation, irrespective of clustering method, shows large errors during typical winter and summer months. Due to the different geographic conditions in the state of Oregon, a separate weekday and weekend analysis may reduce the AADT errors especially among the recreational pattern types. In fact, the ODOT seasonal trend grouping method uses both weekday and weekend patterns for grouping the traffic patterns. This may be one of the reasons that the ODOT method shows less standard deviation of errors during summer months. In addition, if a particular route carries predominantly weekend traffic, a short-term traffic count should consider staging the surveys during weekends.

While clustering, the study uses patterns from a single year. Additional years, at least five years of data reported in other studies, may enhance the stability of the traffic trends and thereby assist the agency's decision-making process. The results reported here belong to a particular state. The consistency of these results using data from other state DOTs must be confirmed before recommending a single clustering technique.

The study considers a purely temporal variation of traffic for the entire state of Oregon when clustering. Several conditions like weather events, detector loops malfunctioning, special and local events, incidents, abrupt land use changes and traffic diversions may potentially disrupt the traffic patterns. These events may potentially isolate the traffic patterns when grouping and thus produce large AADT estimation errors. In the absence of knowledge on underlying events, the study reports all errors and does not put an emphasis on outlier analysis. However, pre-processing input traffic patterns coupled with outlier analysis improve the AADT estimation process. AADT estimation errors, especially larger than 15 percent and observed in some periods like December and July, should further be investigated to assess the impact.

The study considers all vehicle patterns when clustering. However, the spatial and temporal variation of passenger and truck traffic differs and therefore a separate analysis should improve the AADT estimation. Similarly, clustering using directional traffic provides more insights and yields better estimates than the total traffic patterns (combining all directions) used in this study. The study does not consider land-use characteristics when grouping and assigning short-term traffic counts. These characteristics appear particularly important when analyzing the recreational patterns (*1*).

The GMM clustering uses only the Gaussian family for all mixtures. However, replacing some components with other types of distribution that better fit the dataset may improve the performance compared to the standard GMM. The use of multiple distributions in finite mixture modeling brings additional complexity to the clustering problem. Hence, a trade-off analysis between the model complexity and its accuracy is warranted to select the best model-based clustering framework (2). The GMM clustering uses hierarchical clustering solution as an initial partition during the EM algorithm. The study needs to check the effect of different GMM clustering initializations on clusters partitions' convergence to the same solution.

The study can expand the evaluation to other classifiers. For instance, support vector machines (*3*), fuzzy logic decision tree (*4*) and k-nearest neighborhood classification (*5, 6*) represent some of the methods proposed in the literature. The classifiers are trained and tested only based on the GMM solution. However, the evaluation using other clustering solutions like the KM and HCA methods provides a robust way to assess the proposed classifier.

The neural networks are trained using a one-against-all approach. However, Ou and Murphy (7) conclude that the one-against-one approach seems to be a better approach for large training datasets with a large number of classes. In the one-against-one approach, each of the *K* (number of clusters) pattern classes can be trained against every one of the other pattern classes. Thus, this approach forms *K (K-1)*/2 binary neural networks. The decision function to assign a class i is either based on a majority vote or max-win scheme (*7*). However, the one-against-one approach, due to its large number of networks, requires more computational resources and time to train the networks.

This study only tests the framework using 24-hour traffic patterns. The extension to study the monthly patterns, the day-of-week patterns, and a combined hourly and daily traffic patterns may truly assess the strengths and limitations of the new method. However, only inputs to the GMM when applying the clustering and neural network structure (number of units in all layers) need to be changed.

The traffic patterns, especially in Oregon, vary by region. The clustering analysis by region brings the regional difference in AADT accuracy values. The study applies the GMM-ONN method on patterns in each region and analyzes the error values by region.

The study considers only 24-hour patterns. However, the optimal duration and frequency of short-term counts provide valuable guidelines when planning traffic count programs. The knowledge on how many days and which months that the agency collects traffic counts have larger implications on resource allocations. The study needs to apply the GMM-ONN method for all combinations of traffic count durations (for instance, 24-hr, 48-hr, 72-hr, 96-hr, etc…) and months (either every month, every two months, or by seasons). The combination that gives the lowest MAPE value is considered as the better option for the count program.

When performing the evaluation, the study did not perform an assessment of a suitable classifier for the KM and HCA methods. Instead, the GMM-ONN method is compared with the KM-QDA and HCA-QDA methods. However, a future study should evaluate and consider a better classifier for the KM and HCA solutions before comparing to the GMM-ONN method.

References

1. Gastaldi, M., G. Gecchele, and R. Rossi. Estimation of Annual Average Daily Traffic from One-week Traffic Counts - A Combined ANN-Fuzzy Approach. Transportation Research: Part C, vol. 47, 2014, PP. 86-99.
2. Chen, P., K. Yin, and J. Sun. Application of Finite Mixture of Regression Model with Varying Mixing Probabilities to Estimation of Urban Arterial Travel Times. Transportation Research

Record: Journal of the Transportation Research Board, No. 2442, Transportation Research Board of the National Academies, Washington, D.C., 2014, pp. 96–105.

3. Tsapakis, I., and W.H.Schneider. Use of Support Vector Machines to Assign Short-Term Counts to Seasonal Adjustment Factor Groups. In Transport Research Record: Journal of the Transportation Research Board, No. 2527, Transportation Research Board, Washington, D.C., 2015, pp. 8-17.
4. Li, M. T., F. Zhao, and L. F. Chow. Assignment of Seasonal Factor Categories to Urban Coverage Count Stations Using a Fuzzy Decision Tree. Journal of Transportation Engineering, 132 (8), 2006, pp. 654-662.
5. Jin, L., C. Xu, and J. D. Fricker. Comparison of Annual Average Daily Traffic Estimates: Traditional Factor, Statistical, Artificial Neural Network, and Fuzzy Basis Neural Network Approach. Transportation Research Board 87th Annual Meeting, TRB, Washington D.C., 2008.
6. Lam, W., Tang, Y. F., Chan, K. S. and M. L. Tam. Short-term Hourly Traffic Forecasts using Hong Kong Annual Traffic Census. Transportation, 33 (3), 2006, pp. 291-310.
7. Ou, G. and Y.L. Murphey. (2007). Multi-class Patterns Classification using Neural Networks. Pattern Recognition 40, pp. 4-18.