
Volume 19C, issue 3, June 2011
ISSN 0968-090X
ELSEVIER
TRANSPORTATION RESEARCH
AN INTERNATIONAL JOURNAL
Part C: Emerging Technologies
Editor-in-Chief: M. Papageorgiou

Transportation Research Part C 19 (2011) 387–399

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Overview Paper

Statistical methods versus neural networks in transportation research: Differences, similarities and some insights

M.G. Karlaftis *, E.I. Vlahogianni

Department of Transportation Planning and Engineering, School of Civil Engineering, National Technical University of Athens, Athens, Greece

ARTICLE INFO

Article history:
Received 5 October 2009
Received in revised form 1 September 2010
Accepted 25 October 2010

Keywords:
Statistical models
Neural networks
Transportation research

ABSTRACT

In the field of transportation, data analysis is probably the most important and widely used research tool available. In the data analysis universe, there are two 'schools of thought'; the first uses statistics as the tool of choice, while the second – one of the many methods from – Computational Intelligence. Although the goal of both approaches is the same, the two have kept each other at arm's length. Researchers frequently fail to communicate and even understand each other's work. In this paper, we discuss differences and similarities between these two approaches, we review relevant literature and attempt to provide a set of insights for selecting the appropriate approach.

1. Introduction

Transportation data are most commonly modeled using two different approaches: statistics and/or Computational Intelligence (CI). The first, statistics, is the mathematics of collecting, organizing and interpreting numerical data, particularly when these data concern the analysis of population characteristics by inference from sampling (Glymour et al., 1997). Statistics have solid and widely accepted mathematical foundations and can provide insights on the mechanisms creating the data. However, they frequently fail when dealing with complex and highly nonlinear data (curse of dimensionality). The second, CI, combines elements of learning, adaptation, evolution and fuzzy logic to create models that are "intelligent" in the sense that structure emerges from an unstructured beginning (the data) (Engelbrecht, 2007; Sadek et al., 2003).

Neural Networks (NN), an extremely popular class of CI models, has been widely applied to various transportation problems, partly because they are very generic, accurate and convenient mathematical models able to easily simulate numerical model components. They have the inherent propensity for storing empirical knowledge and can be used in any of three basic manners (Haykin, 1999): i. As models of biological nervous systems. ii. As real-time adaptive signal processors/controllers. iii. As data analytic methods. In transportation research, NN have been mainly used as data analytic methods because of their ability to work with massive amounts of multi-dimensional data, their modeling flexibility, their learning and generalization ability, their adaptability and their – generally – good predictive ability.[1]

Very frequently, researchers proficient in one of the two approaches argue fervently in support of 'their' chosen method, making the selection of analysis approach one of the hotly debated topics in research gatherings and publications; and, although these arguments provide for interesting scientific debates, they manage to thoroughly confuse both younger researchers and – particularly – practitioners who are more interested in what model they should use rather than concentrating on the philosophical or mathematical underpinnings of the approaches.

* Corresponding author. Tel.: +30 2107721280.
E-mail addresses: mgk@central.ntua.gr (M.G. Karlaftis), elenivl@central.ntua.gr (E.I. Vlahogianni).

[1] Some studies in transportation have also used NN for traffic control; such research includes the work of Bingham (2001), Abdulhai et al. (2003), Srinivasan et al. (2006).

0968-090X/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.trc.2010.10.004

Table 1
Basic NN terminology and statistical equivalent (adapted from Sarle (1994)).

Statistics	Neural networks
Independent/predicted variable	Input/output
Dependent variables	Targets or training values
Residuals	Errors
Estimation	Training, learning, adaptation, or self-organization.
Estimation criterion	Error function, cost function, Lyapunov function
Observations	Patterns or training pairs
Parameter estimates	(Synaptic) weights
Interactions	Higher order neurons
Transformations	Functional links
Regression and discriminant analysis	Supervised learning or hetero-association
Data reduction	Unsupervised learning, encoding or auto-association
Cluster analysis	Competitive learning or adaptive vector quantization

In this paper, we tackle three issues: i. The differences and similarities between the two approaches to answer the question of what their distinctions are and what important components they share. ii. Review the literature to highlight the vast array of NN applications in transportation while focusing on works that compare findings from statistics and NN. iii. Offer some insights and directions on the preferred analysis approach.[2]

2. Differences, similarities and complementarities

2.1. Differences

Statistics and NN have much in common; they are both concerned with planning, combining evidence and assisting in decision making while neither is an empirical science in itself; they both aspire, however, to being general sciences of practical reasoning. Despite potential communalities, the two disciplines rarely communicate and have kept each other at arm's length; we attribute this surprising behavior to five main factors: i. Terminology. ii. Philosophy. iii. Goals. iv. Model development. v. Knowledge acquisition.

The first difference between statistics and NN lies in the *terminology* used by each discipline. Most terms used in NN modeling are entirely different from those in statistics; Sarle (1994) made an effort to codify neural network terminology in statistical terms (or statistical terminology in NN terms!) (see Table 1). Still, however, there is significant ambiguity when it comes to the terminology used that may be attributed to NN applications whose phrasing has lacked continuity on several occasions; this has led to a confusion for a number of researchers, who fail to understand the – clear in our opinion – correspondence between the two approaches and has been a source of criticism and uncertainty between transportation researchers.

The second difference lies in the core of the two approaches, i.e. their very underlying *philosophy* (Table 2). NN emphasize implementation while statistics largely emphasize inference and estimation (Hand, 2000). As Breiman (2001) discusses, statistics treat data as having been generated from a stochastic process, whereas NN assume data to have been generated by a mechanism of unknown dynamics. In general, the process is the same regardless of the approach used; that is, to recognize and define the problem, select a method to solve it, and, then, interpret the results (Wild and Pfannkuch, 1999).

The third difference lies in the *goals* of each approach. Statistics aim at providing a model – a predictor or a classifier for example – and offering insights on the data and its structure; the elements of such a model should be self-explanatory (Wild and Pfannkuch, 1999). Further, statistics explain the phenomena investigated by interpreting marginal effects and signs, studying elasticities and estimator properties. On the other hand, most NN applications do not target interpretation, but rather aim at providing an efficient – in terms of accuracy and development time – representation of the underlying properties of the data and offer good predictions for the phenomenon under study.

The fourth difference lies in the core *model development* process for each approach, viewed through four steps: learning, definition and interpretation, assumptions, and collinearity. A fundamental difference between statistics and NN is the *learning* process in NN which, regardless of the method used (supervised or unsupervised, maximum likelihood or Bayesian, and so on), results in more than one model; this is in stark contrast with (classical) statistics which result in one final model. In NN, the learning curve may have various local minima and a NN model may converge to various sequential architectures that are not necessarily nested (Ripley, 1996); this inherent characteristic makes NN modeling more flexible than statistics, since the functional form is approximated via learning and not a priori assumed as is done in statistics (Warner and Misra, 1996). However, this leads to significant implications regarding the quality of modeling; for example, NN's inference mechanism is hidden and, very frequently, researchers disregard the implicit assumptions made regarding the data (this is possibly the main reason some researchers call NN 'black boxes' that lack transparency and reproducibility). Finally, the learning process leads to time consuming model development, particularly when compared to simpler statistical model structures (Cheng

[2] We note that the conceptual and methodological comparisons we present in this paper are based on applications from the transportation literature and may not be applicable to general statistics and NN theory.

Table 2
Comparison of NN and statistical philosophies.

Steps	Neural networks	Statistical methods
1	Problem definition and formulation	Problem definition and formulation
2	Data collection, preprocessing and partition	Data collection and preprocessing
3	Architecture and learning type selection	Selection of a close form equation
4	Learning: structural and learning optimization via error minimization	Parameter estimation
5	Testing using new data, comparison with other methods (neural networks or statistics)	Testing using goodness of fit tests and residual examination

and Titterington, 1994). The time consuming NN development and training process and their implications on real-time applications have not been – surprisingly – investigated in the transportation literature.

Another model development difference is in parameter *definition and interpretation*. Statistical methods are generally defined in terms of the mathematical model they use and of the statistical properties of their results, whereas NN are often defined in terms of their architecture and their learning algorithms. Researchers implementing NN models, in contrast to statisticians who spend more time analyzing the problem and controlling the validity of their hypotheses, assume that the model will learn the desired relations in the data without intervention or inclusion of a priori knowledge (Flexer, 1996).

An additional model development difference lies in the *assumptions/limitations* of the two approaches; statistics frequently make a number of hypotheses and place restrictions on developing models that are not made in NN. For example, regression cannot deal effectively with nonlinearity while NN are inherently nonlinear nonparametric models that can straightforwardly deal with indefinable nonlinearity (DeTienne et al., 2003). Further, models in statistics are specified a priori and strict hypotheses are made regarding the error term; on the other hand, NN parameters are extremely adaptable, few – if any – assumptions are made regarding the error term and the construction of complex (regression-like) models is feasible without prior model or error distribution specifications (Hanson, 1995). Moreover, a problematic issue in statistical model development is *multicollinearity*, i.e. the high degree of correlation between two or more independent variables, which is much better dealt with in NN as the assumption of independent variables being uncorrelated is not made (DeTienne et al., 2003). Finally, statistical models are fairly inept at dealing with outliers, missing or noisy data (Gupta and Lam, 1996; Principe et al., 2000).

The final difference between statistics and NN relates to the *knowledge acquisition* process for each approach. Statistics are, in the minds of students (frequently justified), a tough discipline, while most times students do not see the need to be involved in the modeling of large datasets (Nicholls, 1999). On the other hand, new NN software packages have simplified the development of extremely sophisticated NN models, while the corresponding statistical models would be almost impossible to develop (Kuan and White, 1994).

2.2. Similarities

Despite their differences, statistics and NN share a surprising similarity: many times, using either statistics or NN, results in the same model! (see Table 3 for some of the modeling analogies). We note that we discuss here prevailing analogies between statistics and Multilayer feed-forward Perceptrons (MLPs), because the structural flexibility of such NN allows for the construction of powerful statistical models.

Most of the literature comparing statistics and NN bases its (statistical) analyses on linear/nonlinear regression, a well established method in transportation applications (a detailed study of the analogies between regression and MLPs is Sarle (1994)). In essence, a single Perceptron with one input variable (independent variable), one output (dependent variable) and a linear activation function (linear combination of inputs) resembles a simple linear regression. Choosing a logistic rather than a linear function results in a nonlinear perception analogous to a logistic regression, whereas a threshold function (0 if $x < 0$, 1 otherwise) results in a linear discriminant function. We note, however, that the difference between these approaches is that regression has a closed form solution for the coefficients while NN use an iterative process for identification (Warner and Misra, 1996).

MLP for function approximation have been previously compared to Splines and polynomials. Eubank (1988) showed that a simple MLP with one hidden layer, a logistic function and a linear output resembles a polynomial regression or a least squares smoothing Spline. Sarle (1994) suggests that the flexibility of neural networks in straightforwardly extending the models to include multiple inputs and outputs without an exponential increase in the number of parameters to be fit is one of their very attractive properties. Cheng and Titterington (1994) discuss the analogies between single unit Perceptrons and Fisher's linear discriminant and linear logistic regression and argued that, although hidden units may result in complex discriminant rules, classical statistics can also produce sophisticated models, such as kernel-based regression, regression trees, linear vector quantization, *k*-nearest neighbor projection pursuit regression that could be faster to converge and more efficient than MLP.

Furthermore, MLP can be used for forecasting by adding memory to its neurons (Principe et al., 2000); for example, memory can be introduced in any layer of a typical MLP and be compared to a moving average (MA) or an autoregressive (AR) time-series process. An MLP with a time delayed input structure – known as time-delayed NN – is a feed-forward process and effectively a nonlinear moving average (NMA) structure. A MLP with multiple inputs and memory in the output layer

Table 3
Basic neural network topologies and their statistical equivalent (adapted from Sarle (1994)).

Neural networks	Statistical models
Feed-forward NN with no hidden layer	Generalized linear models
• Simple linear Perceptron	• (Multivariate) linear regression
• Simple nonlinear Perceptron	• Logistic regression
• adaline	• Linea discriminant function
Feed-forward NN with one hidden layer	Projection pursuit regression
Generalized regression NN	Kernel regression
Probabilistic NN	Kernel discriminant analysis
Competitive learning networks	*k*-means clustering
Kohonen self-organizing maps	Discrete approximations to principal curves and surfaces
Hybrid networks (supervised and unsupervised learning)	Principal components regression
Learning vector quantization	Variation of nearest-neighbor discriminant analysis
Hebbian learning	Principal component analysis

that feeds back (global feedback) previous outputs to the NN, implements a nonlinear autoregressive with exogenous inputs (NARX) process. Similarly, a MLP with multiple inputs feeding forward previous input data and also feeding backward previous output data resembles a NARMAX structure (Mandic and Chambers, 2001).

2.3. What are possible synergies?

Despite their similarities and differences, the approaches discussed can be combined into a solid and powerful methodological platform. As Cheng and Titterington (1994) argue, NN provide a representational framework for familiar statistical constructs, while statistical methodology can be directly applicable to NN models through estimation criteria, confidence intervals, and diagnostics and so on. In general, there are three basic areas where statistics and NN can act complementarily: i. Core model development. ii. Analysis of large data sets. iii. Causality investigation.

In *model development* and evaluation, statistics can provide the theoretical means for testing for optimality and suitability of the learning algorithms and NN models (Yang et al., 1998); from the standpoint of statistical inference, learning in NN can be viewed as "sequential estimation" (Amari, 1995), while statistical principles may also be widely implemented in NN model selection (White, 1989; Terasvirta et al., 1993; Anders and Korn, 1999; Curry and Morgan, 2006). Besides the commonly used regularization, pruning and early stopping (see textbooks by Bishop (1995) and Principe et al. (2000)), researchers have implemented Wald and LM procedures for testing the hypothesis of parameter significance in NN models (White, 1989; Terasvirta et al., 1993; Anders and Korn, 1999). Moreover, information criteria such as AIC, BIC and NIC and the popular cross-validation have been applied to find the trade-off between unbiased approximation and loss of accuracy by increased parameterization (Stone, 1974; Akaike, 1973; Murata et al., 1994). Interestingly, statistical inference in NN design and training has not been systematically implemented. Few studies have utilized network pruning and growing techniques (Jin et al., 2002; Srinivasan et al., 2004), information criteria (Costa and Markellos, 1997; Jiang and Adeli, 2005), and sensitivity analysis to test for the response of the output variable to variations in inputs (Yasdi, 1999; Yin et al., 2002; Xie et al., 2007; Loizos and Karlaftis, 2006; Ye et al., 2009).

Second of possible synergies is the analysis of *large data sets*. Modern datasets, containing massive amounts of data frequently collected in real-time, are becoming more complex and multi-dimensional requiring "intelligent" algorithms and methodologies for their analysis (Hand, 2000). Computational and inferential implications when working with large datasets are significant and algorithms and models that have traditionally worked well in small datasets may become infeasible with massive data thus requiring 'intelligent', flexible and nonlinear models (such as NN). Further, massive data sets frequently contain erroneous information and NN are more suitable than statistical approaches in these cases (Chen et al., 2001). However, statistics are helpful in handling large datasets and mining critical information; to this end, Glymour et al. (1997) suggest three fundamental statistical concepts that could be helpful: (a) clarity about the modeling goals, (b) assessment of reliability and (c) accounting for sources of uncertainty in the underlying data mechanism and models. In this context, statistical theory contains a vast set of tests and procedures to offer to NN analyses.

The third complementarity point is the extraction of *causalities*. NN have no straightforward manner for investigating causality; although NN can approximate complex dynamics without multicollinearity's restrictions (Chien et al., 2002; Ulengin et al., 2007), many statistical constructs and tests can be very effective in assessing the characteristics of the input–output relationships and investigating causalities that are extremely useful in research studies.

3. Contrasting statistical methods and neural networks in transportation research

Despite the extensive use of NN in transportation research, surprisingly few papers compare findings from statistics to NN. This section focuses on the comparative applications between neural networks and statistical models in various fields of transportation, infrastructure and traffic engineering. The selected literature incorporates peer reviewed journal articles that are related to NN and are summarized in Tables 4–9. Each Table encompasses research efforts from one of six distinct categories of transportation research: traffic operations, infrastructure management, maintenance and rehabilitation, planning, environment and transportation, and safety and human behavior. In each table, research efforts are summarized with respect to the transportation and modeling problem faced, the NN architecture and training algorithm utilized, as well as whether formal approaches were employed to optimize the NN's structure or its training.

Finally, Table 10 summarizes research papers dedicated to comparing statistical approaches with NN models. Comparisons are done for the following problems: classification and clustering, function approximation and time-series analysis.

3.1. Classification and clustering problems

Much of the 'comparative' literature refers to classification problems; MLPs have been shown to perform better than logit models, discriminant analysis, negative binomial regression and stepwise logistic regression in classification problems in incident detection (Ivan and Sethi, 1998; Khan and Ritchie, 1998), gap acceptance modeling at stop controlled intersections (Pant and Balakrishnan, 1994), and safety modeling (Hashemi et al., 1995; Chang, 2005; Sommer et al., 2008). Hashemi et al. (1995) refer to a six steps process for comparing discriminant analysis and neural networks and suggest that modeling with NN places no requirements for a specific functional form and indicate their ability to deal with missing data.

Lingras and Adamo (1996), in modeling average and peak traffic volumes, showed that NN and multiple regression models resulted in similar errors, but NN managed to produce reasonable estimations when the data were insufficient to develop multiple regression models; similar results were reported in McFadden et al. (2001). The effectiveness of NN compared to regression was discussed in Kaseko et al. (1994) and Fwa et al. (1997). Hoogendoorn and Hoogendoorn-Lanser (2001), on modeling travel behavior in transportation networks, explain that the improvement in the performance achieved by NN compared to multinomial logit may be due to the nonlinearity in the choice process. Hunt and Lyons (1994) in their analysis

Table 4
Analysis of literature on traffic operations.

References	Application	Problem[a]	Architecture[b]	Training[c]	GA optimization[d]
Stathopoulos et al. (2008)	Traffic forecasting	P	F	Other	+
Vlahogianni et al. (2008)	Traffic pattern analysis	CLU	Hybrid	U	
Wang et al. (2008)	Incident detection	CLA	SVM	BP	
Vlahogianni et al. (2007)	Traffic forecasting	P	M	LM	+
Oh et al. (2005)	Data aggregation	P	R	Other	
Sazi-Murat (2006)	Vehicle delay modeling	FA	F	BP	
Srinivasan et al. (2006)	Traffic signal control	CLA	F	Other	
Teodorovic et al. (2006)	Signal timing optimization	FA	MLP	BP	
Vlahogianni et al. (2005)	Traffic forecasting	P	MLP	BP	+
Zhong et al. (2005)	Traffic forecasting	P	TD	BP	+
Hooshdar and Adeli (2004)	Variable message signs	CLA	MLP	LM	
Srinivasan et al. (2004)	Incident detection	CLA	P	Other	
Zhong et al. (2004)	Traffic analysis	P	TDNN	BP	+
Teng and Qi (2003a)	Incident detection	CLA	P	BP	
Teng and Qi (2003b)	incident detection	CLA	MLP	BP	
Jin et al. (2002)	Incident detection	CLA	P	Other	
Tong and Hung (2002)	Discharge rate/signalized	FA	MLP	BP	
Jin et al. (2002)	Incident detection	CLA	P	BP	
Yin et al. (2002)	Traffic forecasting	P	F	Other	
Chen et al. (2001)	Traffic forecasting	P	Hybrid	BP	
McFadden et al. (2001)	Traffic forecasting	FA	MLP	BP	
Qiao et al. (2001)	Traffic forecasting	P	MLP	BP	
Lingras et al. (2000)	Traffic forecasting	P	TDNN	BP	
Abdulhai and Ritchie (1999)	Incident detection	CLA	P	BP	
Park et al. (1999)	Traffic forecasting	FA	RBF	BP	
Amin et al. (1998)	Traffic forecasting	P	RBF	Other	
Ivan and Sethi (1998)	Incident detection	CLA	MLP	BP	
Khan and Ritchie (1998)	Incident detection	CLA	M	BP	
Nakatsuji et al. (1998)	Traffic analysis	R	Hybrid	BP	
Smith and Demetsky (1997)	Traffic forecasting	P	MLP	BP	
Van Der Voort et al. (1996)	Traffic forecasting	CLU	Hybrid	BP	
Cheu and Ritchie (1995)	Incident detection	CLA	KSOM	U	
Pant and Balakrishnan (1994)	Gap acceptance/unsignalized	CLA	MLP	BP	

[a] P: prediction, CLA: classification, FA: function approximation, CLU: clustering.
[b] F: fuzzy, M: modular, W: wavelet, H: hopfield, R: recurrent, P: probabilistic, TD: time-delay, KSOM: Kolhonen self-organizing maps, MLP: multi-layer Perceptron, RBF: radial basis function.
[c] BP: back-propagation, LM: Levenberg–Marquardt, U: unsupervised.
[d] +: yes.

Table 5
Analysis of literature on infrastructure management, maintenance and rehabilitation.

References	Application	Problem[a]	Architecture[b]	Training[c]	GA optimization[d]
Loizos and Karlaftis (2006)	Pavement crack modeling	FA	MLP	BP	
Yang et al. (2006)	Pavement crack modeling	FA	MLP	BP	
Mukkamala and Sung (2003)	Intrusion detection	CLA	MLP	LM	
Fwa et al. (1997)	Airfield pavement management	CLA	MLP	BP	
Furuta et al. (1996)	Damage assessment	FA	MLP	BP	
Williams and Gucunski (1995)	Pavement backcalculation	FA	MLP	BP	
Kaseko et al. (1994)	Pavement crack detection	CLA	Other	BP	

[a] P: prediction, CLA: classification, FA: function approximation, CLU: clustering.
[b] F: fuzzy, M: modular, W: wavelet, H: Hopfield, R: recurrent, P: probabilistic, TD: time-delay, KSOM: Kolhonen self-organizing maps, MLP: multi-layer Perceptron, RBF: radial basis function.
[c] BP: back-propagation, LM: Levenberg–Marquardt, U: unsupervised.
[d] +: yes.

Table 6
Analysis of literature on planning.

References	Application	Problem[a]	Architecture[b]	Training[c]	GA optimization[d]
Zhang and Xie (2008)	Travel mode choice modeling	FA	SVM	BP	
Dia and Panwai (2007)	Route choice modeling	CLA	Hybrid	Other	+
Celikoglu (2006)	Travel mode choice modeling	CLA	RBF	LM	
Cheu et al. (2006)	Shared-use vehicle trips	FA	MLP	BP	
Andrade et al. (2006)	Transport mode choice model	FA	F	Other	
Longhi et al. (2005)	Regional employment patterns forecasting	FA	MLP	BP	
Tillema et al. (2006)	Trip distribution modeling	FA	MLP	BP	
Cantarella and de Luca (2005)	Mode choice modeling	CLA	MLP	BP	
Sarvareddy et al. (2005)	Truck trip generation model	P	R	Other	
Mostafa (2004)	Forecasting the Suez Canal traffic	P	MLP	BP	
Tang et al. (2003)	AADT forecasting	P	MLP	BP	
Vythoulkas and Koutsopoulos (2003)	Mode choice modeling	FA	F	BP	
Lingras et al. (2002)	Recreational travel prediction	P	TD	LM	+
Mohammadian and Miller (2002)	Predicting household automobile choices	CLA	MLP	BP	
Tseng et al. (2002)	Tourist arrival forecasting	P	Hybrid	BP	
Lingras and Mountford (2001)	Short term inter-city traffic forecasting	P	TD	BP	
Mozolin et al. (2000)	Trip distribution forecasting	FA	MLP	BP	
Xu et al. (1999)	Level of urban taxi services modeling	FA	MLP	BP	
Lingras and adamo (1996)	Traffic demand modeling	FA	MLP	BP	
Nijkamp et al. (1996)	Mode choice modeling	FA	MLP	BP	
Shmueli et al. (1996)	Analysis of travel behavior	CLA	MLP	BP	
Faghri and Hua (1995)	Roadway seasonal classification	CLU	Other	U	
Lingras (1995)	Highway classification	CLU	KSOM	U	

[a] P: prediction, CLA: classification, FA: function approximation, CLU: clustering.
[b] F: fuzzy, M: modular, W: wavelet, H: Hopfield, R: recurrent, P: probabilistic, TD: time-delay, KSOM: Kolhonen self-organizing maps, MLP: multi-layer Perceptron, RBF: radial basis function.
[c] BP: back-propagation, LM: Levenberg–Marquardt, U: unsupervised.
[d] +: yes.

of NN performance used for classification in lane changing modeling state that "*the extent to which neural network solutions are superior to existing data analysis and predictive techniques, such as multiple regression, remains unclear*" in large part because of the little guidance provided to practitioners regarding the selection of NN architecture (a critical factor in determining the success or failure of a NN).

In mode choice modeling, Shmueli et al. (1996) compared simple MLP to nonlinear classification and regression trees (CART) and provided evidence that both methodologies perform equally well in modeling travel behavior. Sayed and Razavi (2000) show that fuzzy NN have similar classification ability as do logit and probit models, while Mohammadian and Miller (2002) suggest that the MLP has a significant edge over the nested logit model in terms of the percentage of cases correctly classified in modeling household automobile choice. Celikoglu (2006) shows that simple MLP may outperform the utility function calibration in travel choice modeling, while Rao et al. (1998), Xie et al. (2003), Adrande et al. (2006), Zhang and Xie (2008) and Hensher and Ton (2000) reported MLP's predictive capability as superior over multinomial and nested logit models; Vythoulkas and Koutsopoulos (2003) suggested that results from fuzzy NN in mode choice behavior modeling compare favorably to the logit model.

There are, however, some research papers that report the opposite finding; Abdelwahab and Abdel-Aty (2002), for example, provided evidence that the two-level nested logit model outperforms the MLP in analyzing driver injury severity. Further, Teng and Qi (2003a,b) discuss the dominance of wavelets over different NN structures in the area of incident detection.

Table 7
Analysis of literature on environment and transportation.

References	Application	Categorization	Problem[a]	Architecture[b]	Training[c]	GA optimization[d]
Cai et al. (2009)	Air pollutants prediction	Env	P	MLP	BP	
Chelani and Devotta (2007)	Air pollutants prediction	Env	P	MLP	BP	
Sousa et al. (2007)	Air pollutants prediction	Env	P	Hybrid	BP	
Zhang and Xie (2006)	Accident detection	ITS	CLA	MLP	BP	
Ayala Botto et al. (2005)	Intelligent vehicle system	ITS	FA	TD	BP	
Abdulhai and Tabib (2003)	Vehicle tracking	ITS	CLA	TD	BP	
Mantri and Bullock (1995)	Vehicle detection	ITS	CLA	MLP	Other	

[a] P: prediction, CLA: classification, FA: function approximation, CLU: clustering.
[b] F: fuzzy, M: modular, W: wavelet, H: Hopfield, R: recurrent, P: probabilistic, TD: time-delay, KSOM: Kolhonen self-organizing maps, MLP: multi-layer Perceptron, RBF: radial basis function.
[c] BP: back-propagation, LM: Levenberg–Marquardt, U: unsupervised.
[d] +: yes.

Table 8
Analysis of literature on safety and human behavior.

References	Application	Problem[a]	Architecture[b]	Training[c]	GA optimization[d]
Li et al. (2008)	Accidents analysis	FA	SVM	BP	
Sommer et al. (2008)	Fitness to drive	CLA	MLP	Other	
Xie et al. (2007)	Accidents analysis	FA	Other	BP	
Chang (2005)	Accidents analysis	CLA	MLP	BP	
Abdel-Aty and Abdelwahab (2004)	Accidents analysis	FA	F	BP	
Abdelwahab and Abdel-Aty (2002)	Accidents analysis	CLA	MLP	BP	
Al-Alawi et al. (1996)	Accidents analysis	FA	MLP	BP	

[a] P: prediction, CLA: classification, FA: function approximation, CLU: clustering.
[b] F: fuzzy, M: modular, W: wavelet, H: Hopfield, R: recurrent, P: probabilistic, TD: time-delay, KSOM: Kolhonen self-organizing maps, MLP: multi-layer Perceptron, RBF: radial basis function.
[c] BP: back-propagation, LM: Levenberg–Marquardt, U: unsupervised.
[d] +: yes.

Some comparative studies in clustering are those by Lingras (1995) on highway classification and Vlahogianni et al. (2008) on traffic flow regime identification. Both studies compared KSOMs with classical statistical clustering approaches such as hierarchical clustering and *k*-means clustering. Lingras (1995) marked the power of KSOMs in classifying incomplete patterns, while Vlahogianni et al. (2008) tested the KSOM in clustering complex noisy traffic patterns and showed that they performed better than *k*-means clustering in terms of cluster separability.

3.2. Function approximation problems

In function approximation problems, NN have been systematically compared to classical statistical models. In traffic operations, Qiao et al. (2001) showed that MLP perform better than probability based methods in predicting traffic flow dispersion in signalized arterials, particularly under complex traffic flow conditions and are more capable for real-time online traffic control. Moreover, Tong and Hung (2002) studied vehicle discharge rates at signalized intersections and presented evidence that MLP performed better that regression. Moreover, Costa and Markellos (1997) suggested that MLP are more efficient than corrected ordinary least squares in estimating public transport efficiency. Xu et al. (1999) found MLP to perform better than simultaneous equations modeling in determining the level of urban taxi services. Al-Deek (2001) showed that NN work better than linear regression in freight truck traffic modeling. Xie et al. (2007) and Li et al. (2008) tested the performance of NN in safety modeling and reported their applicability and superiority over the negative binomial regression models.

In trip distribution modeling, Mozolin et al. (2000) compared the performance of MLP to maximum likelihood doubly-constrained models for commuter trip distribution and suggested that MLP may fit data better but its predictive accuracy is poor in comparison to maximum likelihood doubly-constrained models due to over-fitting and lack of generalization power. Celik (2004) showed that MLP outperforms gravity models, while Tillema et al. (2006) reported that NN perform better than gravity models with limited data but, when data increase, the performance gap of the two modeling approaches decreases.

3.3. Time-series analysis and forecasting problems

Most comparative studies in time-series analysis deal with traffic parameters such as volume, speed and so on, with some comparative studies in the field of planning (Tseng et al., 2002; Tang et al., 2003; Mostafa, 2004), transit operations (Jeong

Table 9
Analysis of literature on air, transit, rail and freight operations.

References	Application	Categorization	Problem[a]	Architecture[b]	Training[c]	GA optimization[d]
Tsai et al. (2009)	Passenger demand forecasting	Rail	P	M	BP	
Tortum et al. (2009)	Mode choice modeling	Freight	FA	F	BP	
Celikoglu and Cigizoglu (2007)	Public transport flow forecasting	Transit	P	RBF	LM	
Ulengin et al. (2007)	Freight demand prediction	Freight	FA	MLP	BP	
Jeong and Rilett (2005)	Bus arrival time prediction	Transit	FA	MLP	BP	
Celik (2004)	Freight modeling	Freight	P	MLP	BP	
Al-Deek (2001)	Freight modeling	Freight	FA	MLP	BP	
Costa and Markellos (1997)	Public transport efficiency	Transit	FA	MLP	BP	
Hashemi et al. (1995)	Freight safety modeling	Freight	CLA	MLP	BP	

[a] P: prediction, CLA: classification, FA: function approximation, CLU: clustering.
[b] F: fuzzy, M: modular, W: wavelet, H: Hopfield, R: recurrent, P: probabilistic, TD: time-delay, KSOM: Kolhonen Self-organizing maps, MLP: multi-layer perceptron, RBF: radial basis function.
[c] BP: back-propagation, LM: Levenberg-Marquardt, U: unsupervised.
[d] +: Yes.

and Rilett, 2005; Celikoglu and Cigizoglu, 2007), environment and transport (Chelani and Devotta, 2007), and infrastructure management (Yang et al., 2006).

Regarding forecasting (prediction), NN employed range from simple static MLPs to complex time-delayed NN (TDNN) and recurrent NN (RNN). The common approach to comparing NN performance for time-series prediction is to test their accuracy against classic Autoregressive Integrated Moving Average (ARIMA) models; this approach is followed by almost all research studies, but the results are rather 'confusing'. There is a part of the literature that suggests that NN are more accurate than ARIMA models, while other research studies provide evidence for the opposite; Kirby et al. (1997), for example, demonstrate that NN perform similarly well to ARIMA models. Lingras et al. (2000), Tseng et al. (2002), Vlahogianni et al. (2005) and Zhong et al. (2004, 2005) show that NN are better predictors of traffic volume than ARIMA models and locally weighted regression models, and Jeong and Rilett (2005) suggest that NN models outperform both historical data based models and regression in terms of prediction accuracy in bus arrival prediction applications.

There are, however, many exceptions; Smith and Demetsky (1997) provided evidence based on the prediction error distributions that nonparametric regression is more suitable for traffic flow prediction when compared to simple MLPs. Tang et al. (2003), on traffic demand prediction, suggested that classical models such as Gaussian Maximum Likelihood require less data calibration when compared to ARIMA models, but that NN are probably more robust and suitable for extensive practical applications. Further, Yang et al. (2006) provided evidence that recurrent Markov chains tend to produce more consistent forecasts when compared to NN (which tend to over-predict crack deterioration), while Chelani and Devotta (2007) showed that local approximation models outperform NN in predicting air pollutant concentrations.

Interest has also concentrated on hybrid structures of NN in short-term traffic flow prediction problems; these structures always outperform simple autoregressive models, particularly in modeling multi-dimensional datasets and constructing models with various exogenous parameters. Examples are the work of Van der Voort et al. (1996), Chen et al. (2001), Vlahogianni et al. (2007) and Stathopoulos et al. (2008) in traffic flow forecasting. Chen et al. (2001) reported the superiority of the combined Kohonen Self-Organizing Maps (KSOM) and MLP modeling when compared to simple ARIMA and the combined KSOM and ARIMA models. Stathopoulos et al. (2008) demonstrated the predictive power of fuzzy NN over ARIMA models and Kalman filters, while Vlahogianni (2009) suggested that advanced NN structures outperform ARIMA models, particularly when traffic flow approaches capacity. Further, NN have been proven as more reliable for multiple steps ahead predictions (Chen et al., 2001), and Mostafa (2004) showed that the forecasting performance of NN relative to ARIMA models remains strong throughout the 12-month forecast horizon and does not appear to deteriorate as the horizon increases.

4. Discussion and conclusions

In this paper, we compared statistical methods and NN as applied in transportation research, while we also reviewed the literature that 'compares' the performance of models developed with the two approaches. From the work discussed, there are three important findings: i. Transportation researchers rely – almost exclusively – on the estimation/prediction error when deciding on the 'effectiveness' of a modeling approach, while disregarding issues such as underlying hypothesis testing, parameter stability, explanatory power, causality, error distribution and so on. ii. When modeling complex datasets with possible nonlinearities or missing data, NN are often regarded as more flexible compared to statistical models. In such cases, the constraint free form of the NN is frequently preferred over the explanatory power of statistics. iii. In applications where NN 'outperform' classical statistical models, researchers disregard their limited inherent explanatory power, preferring higher prediction accuracy over 'explanation'.

What becomes clear from this review is that despite the countless research papers in transportation that utilize NN, researchers often implement them blindly, ignoring some of their shortcomings such as limited inherent explanatory power or their inherent inability to produce a unique solution to a problem (leading many to refer to NN as "black-boxes"). Moreover, many of the comparisons between statistical and NN models are 'unfair', particularly because complex NN (essentially

Table 10
Analysis of literature for the comparative studies between statistical and neural networks models' performance in transportation research.

References	Categorization[a]	Problem[b]	Architecture[c]	Comparison[d]	References	Categorization[a]	Problem[b]	Architecture[c]	Comparison[d]
Cai et al. (2009)	E	P	MLP	R	Teng and Qi (2003a,b)	T	CLA	MLP	R
Dimitriou et al. (2008)	T	P	FNN	Kf and AR	Vythoulkas and Koutsopoulos (2003)	P	FA	FNN	L
Li et al. (2008)	S	FA	SVM	R	Abdelwahab and Abdel-Aty (2002)	S	CLA	MLP	L
Sommer et al. (2008)	S	CLA	MLP	R	Mohammadian and Miller (2002)	P	CLA	MLP	L
Tortum et al. (2009)	F	FA	FNN	R	Tong and Hung (2002)	T	FA	MLP	R
Vlahogianni et al. (2008)	T	CLU	M	*k*-means	Tseng et al. (2002)	P	P	M	AR
Wang et al. (2008)	T	CLA	SVM	R	Al-Deek (2001)	F	FA	MLP	R
Zhang and Xie (2008)	P	FA	SVM	L	Hoogendoorn and Hoogendoorn-Lanser (2001)	F	CLA	MLP	L
Celikoglu and Cigizoglu (2007)	F	P	RBF	AR	McFadden et al. (2001)	T	FA	MLP	R
Chelani and Devotta (2007)	E	P	MLP	AR	Qiao et al. (2001)	T	P	MLP	PR
Sousa et al. (2007)	E	P	M	R	Lingras et al. (2000)	T	P	TDNN	AR
Xie et al. (2007)	S	FA	BNN	R	Mozolin et al. (2000)	P	FA	MLP	R
Celikoglu (2006)	P	CLA	RBF	R	Xu et al. (1999)	P	FA	MLP	SE
Loizos and Karlaftis (2006)	I	FA	MLP	CART	Amin et al. (1998)	T	P	RBF	AR
Longhi et al. (2006)	P	FA	MLP	R	Ivan and Sethi (1998)	T	CLA	MLP	D
Teodorovic et al. (2006)	T	FA	MLP	DP	Khan and Ritchie (1998)	T	CLA	MNN	PR
Tillema et al. (2006)	P	FA	MLP	GM	Nakatsuji et al. (1998)	T	R	M	R
Ayala Botto et al. (2005)	T	FA	TDNN	AR	Fwa et al. (1997)	I	CLA	MLP	R
Cantarella and de Luca (2005)	P	CLA	MLP	L	Smith and Demetsky (1997)	T	P	MLP	AR and NR
Chang (2005)	S	CLA	MLP	R	Lingras and adamo (1996)	P	FA	MLP	NR
Jeong and Rilett (2005)	F	FA	MLP	R	Nijkamp et al. (1996)	P	FA	MLP	L
Vlahogianni et al. (2005)	T	P	MLP	AR	Shmueli et al. (1996)	P	CLA	MLP	CART
Zhong et al. (2005)	T	P	TDNN	NR	Van Der Voort et al. (1996)	T	CLU	M	AR
Mostafa (2004)	P	P	MLP	AR	Faghri and Hua (1995)	P	CLU	ART	R
Celik (2004)	F	P	MLP	AR	Hashemi et al. (1995)	F	CLA	MLP	*L*
Zhong et al. (2004)	T	P	TDNN	R and AR	Lingras (1995)	P	CLU	KSOM	R
Tang et al. (2003)	P	P	MLP	NR and AR	Kaseko et al. (1994)	I	CLA	LVQ	PR and NR
Teng and Qi (2003a,b)	T	CLA	PNN	W	Pant and Balakrishnan (1994)	T	CLA	MLP	L

[a] T: traffic operations, I: infrastructure management/maintenance/rehabilitation, P: planning, F: air/transit/freight operations, E: environment/transportation, S: safety.
[b] P: prediction, CLA: classification, FA: function approximation, CLU: clustering.
[c] F: fuzzy, M: modular, W: wavelet, H: Hopfield, R: recurrent, P: probabilistic, TD: time-delay, KSOM: Kolhonen self-organizing maps, MLP: multi-layer Perceptron, RBF: radial basis function.
[d] R: regression, PR: probabilistic regression, NR: nonparametric regression, AR: autoregressive models, SE: simultaneous equation, DP: dynamic programming, L: logit family of models, GM: gravity models, Kf: Kalman filters.

highly nonlinear models) are compared to simple linear regression or linear ARIMA models! Additionally, there are few – if any – transportation test problems (benchmark datasets) that could be used for a (fair) comparison between statistical models and NN. Establishing such datasets may facilitate the extraction of useful conclusions regarding the strengths and limitations of both approaches for given – and clearly defined – transportation problems.

Further, most researchers frequently follow the path of least resistance by comparing approaches based solely on model accuracy; however, there is a 'thin line' between modeling accuracy, model simplicity, and model suitability. Kirby et al. (1997) suggested that accuracy is very important but should not be the sole determinant for selecting the proper methodology – statistical or NN – for prediction; other issues should be considered in selecting the appropriate approach such as the time and effort required for model development, skills and expertise required, transferability of the results, adaptability to changing behaviors and so on (Kirby et al., 1997; Smith and Demetsky, 1997; Vlahogianni et al., 2004).

Transportation researchers are often interested in some guidance regarding what the 'best' modeling approach would be for their data; to address this, a researcher should ask some guiding questions such as:

- What are the requirements with respect to accuracy and interpretability of results?
- Is adequate prior knowledge available regarding the problem discussed?
- Is it possible to develop a statistical model that can solve a problem with the desired levels of accuracy?
- How important is interpretability in the problem examined?
- What is the proper design and evaluation of the NN developed?

Although it is very difficult to generalize, the literature largely suggests that statistics are best practiced when: i. There exists a statistical method that solves a given problem better than neural networks. ii. Researchers have knowledge, or a priori information, regarding the functional relationship of the variables in the problem. iii. Researchers need to verify the statistical properties of the underlying mechanism that produced the problem. iv. When interpretability of results and causalities are important. NN are suggested for development when: i. Emphasis is on obtaining 'good' predictions and not so much on how these predictions were obtained. ii. The true data generating process is unknown and hard to identify. iii. The idealized assumptions of statistical models (e.g. normality, linearity, stationarity) are not valid. iv. Traditional (statistical) methods yield results that are extremely tedious and nearly impossible to interpret.

We believe that carefully testing the validity of constraints in statistical modeling is as important as properly selecting, developing and training NN models; if a NN model is not carefully developed and trained, then an overestimated structure can lead to an over-specified model with limited or no generalization power (over-fitting), while an under-estimated structure can lead to limited predictive capabilities. Further, acquiring good command of both statistics and NN before using them as tools for modeling is critical both for selecting the appropriate methodology and for evaluating the efficiency of the model developed (where interactions between the two approaches are possible).

Most transportation research is largely based on the use of modeling tools, be they statistical models or NN. However, both approaches – try to – respond to questions regarding the effects of independent (input) on dependent (output) variables. Interestingly, in recent years, there is an obvious trend in scientific studies with a corresponding trend in mathematical and statistical tools from simpler, linear, low-dimensional to complex, nonlinear, high dimensional systems, largely because of significant increases in computing power, but still not necessarily justified by logic or fundamental research needs. It is our opinion that the goals of the analysis are more important than the tools used, that there are always (implicit) assumptions in all modeling approaches, and that complex, nonlinear tools have both advantages and limitations and frequently simpler models give as good results as complex ones.

Acknowledgments

The authors would like to thank Prof. Halil Ceylan, Ms. Cheryl Richter and Mr. Aristides Karlaftis for initiating the work, offering many useful ideas, and making helpful suggestions on this paper. The constructive comments of the associate editor and three anonymous reviewers are also gratefully acknowledged.

References

Abdel-Aty, M.A., Abdelwahab, H.T., 2004. Predicting injury severity levels in traffic crashes: a modeling comparison. Journal of Transportation Engineering 130 (2), 204–210.

Abdelwahab, H.T., Abdel-Aty, M.A., 2002. Artificial neural networks and logit models for traffic safety analysis of toll plazas. Transportation Research Record 1784, 115–125.

Abdulhai, B., Ritchie, S.G., 1999. Enhancing the universality and transferability of freeway incident detection using a Bayesian-based neural network. Journal of Transportation Research – Part C: Emerging Technologies 7 (5), 261–280.

Abdulhai, B., Tabib, S., 2003. Spatio-temporal inductance-pattern recognition for vehicle re-identification. Journal of Transportation Research-Part C: Emerging Technologies 11 (3–4), 223–239.

Abdulhai, B., Pringle, R., Karakoulas, G.J., 2003. Reinforcement learning for true adaptive traffic signal control. Journal of Transportation Engineering 129 (3), 278–285.

Akaike, H., 1973. Information theory and the extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), Second International Symposium on Information Theory. Academia Kiado, Budapest, pp. 267–281.

Al-Alawi, S., Ali, G., Bakheit, C., 1996. A novel approach for traffic accident analysis and prediction using artificial neural networks. Road Transport Research 5, 118–128.

Al-Deek, H.M., 2001. Which method is better for developing freight planning models at seaports: neural networks or multiple regression? Journal of the Transportation Research Board 1763, 90–97.

Amari, S., 1995. In: Arbib, M.A. (Ed.), The Handbook of Brain Theory and Neural Networks. MIT Press/Bradford Books, pp. 522–526

Amin, S.M., Rodin, E.Y., Garcia-Ortiz, A., 1998. Traffic prediction and management via RBF neural nets and semantic control. Computer-aided Civil and Infrastructure Engineering 13 (5), 315–327.

Anders, U., Korn, O., 1999. Model selection in neural networks. Neural Networks 12, 309–333.

Andrade, K., Uchida, K., Kagaya, S., 2006. Development of transport mode choice model by using adaptive neuro-fuzzy inference system. Transportation Research Record 1977, 295–304.

Ayala Botto, M., Sousa, J.M.C., Sa da Costa, J.M.G., 2005. Intelligent active noise control applied to a laboratory railway coach model. Control Engineering Practice 13, 473–484.

Bingham, E., 2001. Reinforcement learning in neural fuzzy traffic signal control. European Journal of Operational Research 131 (2), 232–241.

Bishop, Ch.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press, ISBN:0-19-853864-2.

Breiman, L., 2001. Statistical modeling: the two cultures. Statistical Science 16, 199–231.

Cai, M., Yin, Y., Xie, M., 2009. Prediction of hourly air pollutant concentrations near urban arterials using artificial neural network approach. Transportation Research Part D: Transport and Environment 14 (1), 32–41.

Cantarella, G.E., de Luca, S., 2005. Multilayer feedforward networks for transportation mode choice analysis: an analysis and a comparison with random utility models. Transportation Research Part C 13, 121–155.

Celik, H.M., 2004. Modeling freight distribution using artificial neural networks. Journal of Transport Geography 12, 141–148.

Celikoglu, H.B., 2006. Application of radial basis function and generalized regression neural networks in non-linear utility function specification for travel mode choice modeling. Mathematical and Computer Modelling 44, 640–658.

Celikoglu, H.B., Cigizoglu, H.K., 2007. Modelling public transport trips by radial basis function neural, networks. Mathematical and Computer Modelling 45, 480–489.

Chang, L.-Y., 2005. Analysis of freeway accident frequencies: negative binomial regression versus artificial neural network. Safety Science 43, 541–557.

Chelani, A.B., Devotta, S., 2007. Prediction of ambient carbon monoxide concentration using nonlinear time series analysis technique. Transportation Research Part D 12, 596–600.

Chen, H., Grant-Muller, S., Mussone, L., Montgomery, F., 2001. A study of hybrid neural network approaches and the effects of missing data on traffic forecasting. Neural Computing & Applications 10, 277–286.

Cheng, B., Titterington, D.M., 1994. Neural networks: a review from a statistical perspective. Statistical Science 9, 2–54.

Cheu, R.L., Ritchie, S.G., 1995. Automated detection of lane blocking freeway incidents using artificial neural networks. Transportation Research Part C: Emerging Technologies 3 (6), 371–388.

Cheu, R.L., Xu, J., Kek, A.G.H., Lim, W.P., Chen, W.L., 2006. Forecasting shared-use vehicle trips with neural networks and support vector machines. Transportation Research Record: Journal of the Transportation Research Board 1968, 40–46.

Chien, S.I-Jy., Ding, Y., Wei, C., 2002. Dynamic bus arrival time prediction with artificial neural networks. ASCE Journal of Transportation Engineering 128 (5), 429–438.

Costa, A., Markellos, R.N., 1997. Evaluating public transport efficiency with neural network models. Transportation Research Part C: Emerging Technologies 5 (5), 301–312.

Curry, B., Morgan, P.H., 2006. Model selection in neural networks: some difficulties. European Journal of Operational Research 170, 567–577.

DeTienne, K.B., Detienne, D.H., Joshi, S.A., 2003. Neural networks as statistical tools for business researchers. Organizational Research Methods 6 (2), 236–265.

Dia, H., Panwai, S., 2007. Modeling drivers' compliance and route choice behavior in response to travel information. Nonlinear Dynamics 49, 493–509.

Dimitriou, L., Tsekeris, T., Stathopoulos, 2008. Adaptive hybrid fuzzy rule-based system approach for modeling and predicting urban traffic flow. Transportation Research Part C 16, 554–573.

Engelbrecht, A.P., 2007. Computational Intelligence. An Introduction, second ed. Wiley, NY.

Eubank, R.L., 1988. Spline Smoothing and Nonparametric Regression of Statistics, Textbooks and Monographs, vol. 90. Marcel Dekker.

Faghri, A., Hua, J., 1995. Roadway seasonal classification using neural networks. Journal of Computing in Civil Engineering 9 (3), 209–215.

Flexer, A., 1996. Statistical evaluation of neural network experiments: minimum requirements and current practice. In: Proceedings of the 13th European Meeting on Cybernetics and Systems Research, Austrian Society for Cybernetic Studies, Vienna, vol. 2, pp. 1005–1008.

Furuta, H., He, J., Watanabe, E., 1996. A fuzzy expert system for damage assessment using genetic algorithms and neural networks. Microcomputers in Civil Engineering 11 (1), 37–45.

Fwa, T.F., Chan, W.T., Lim, C.T., 1997. Decision framework for pavement friction management of airport runways. Journal of Transportation Engineering 123 (6), 429–435.

Glymour, C., Madigan, D., Pregibon, D., Smyth, P., 1997. Statistical themes and lessons for data mining. Data Mining and Knowledge Discovery 1 (1), 11–28.

Gupta, A., Lam, M.S., 1996. Estimating missing values using neural networks. Journal of the Operational Research Society 47 (2), 229–239.

Hand, D.J., 2000. Data mining, new challenges for statisticians. Social Science Computer Review 18 (4), 442–449.

Hanson, S.J., 1995. Back-propagation: some comments and variations. In: Rumelhart, D.E., Yves, C. (Eds.), Back-propagation: Theory, Architecture, and Applications. Lawrence Erlbaum, NJ, pp. 237–271.

Hashemi, R.R., Le bBanc, L.A., Rucks, C.T., Shearry, A., 1995. A neural network for transportation safety modeling. Expert Systems with Applications 9 (3), 247–256.

Haykin, S., 1999. NN: A Comprehensive Foundation. Macmillan, NY.

Hensher, D.A., Ton, T.T., 2000. A comparison of the predictive potential of artificial neural networks and nested logit models for commuter mode choice. Transportation Research Part E 36 (3), 155–172.

Hoogendoorn, S.P., Hoogendoorn-Lanser, S., 2001. Neural networks approach to travel behavior in public transport networks. In: TRB (Ed.), TRB Annual Meeting pre-print CD-rom, Washington DC: Transportation Research Board.

Hunt, J.G., Lyons, G.D., 1994. Modeling dual carriageway lane changing using neural networks. Transportation Research Part C 2 (4), 231–245.

Ivan, J.N., Sethi, V., 1998. Data fusion of fixed detection and probe vehicle data for incident detection. Computer-aided Civil & Infrastructure Engineering 13 (5), 329–337.

Jeong, R., Rilett, L.R., 2005. Bus arrival time prediction model for real-time applications. Transportation Research Record 1927, 195–204.

Jiang, X., Adeli, H., 2005. Dynamic wavelet neural network model for traffic flow forecasting. Journal of Transportation Engineering 131 (10), 771–779.

Jin, X., Cheu, R.L., Srinivasan, D., 2002. Development and adaptation of constructive probabilistic neural network in freeway incident detection. Transportation Research Part C: Emerging Technologies 10 (2), 121–147.

Kaseko, M.S., Lo, Z.-P., Ritchie, S.G., 1994. Comparison of traditional and neural classifiers for pavement–crack detection. Journal of Transportation Engineering 120 (4), 552–569.

Khan, S.I., Ritchie, S.G., 1998. Statistical and neural classifiers to detect traffic operational problems on urban arterials. Transportation Research Part C: Emerging Technologies 6 (5), 291–314.

Kirby, H., Dougherty, M., Watson, S., 1997. Should we use neural networks or statistical models for short term motorway forecasting? International Journal of Forecasting 13, 45–50.

Kuan, Ch.-M., White, H., 1994. Artificial neural networks: an econometric perspective. Econometric Reviews 13 (l), 1–9.

Li, X., Lord, D., Zhang, Y., Xie, Y., 2008. Predicting motor vehicle crashes using support vector machine models. Accident Analysis & Prevention 40 (4), 1611–1618.

Lingras, P., 1995. Classifying highways: hierarchical grouping versus Kohonen neural networks. Journal of Transportation Engineering 121 (4), 364–368.

Lingras, P., Adamo, M., 1996. Average and peak traffic volumes: neural nets, regression, and factor approaches. Journal of Computing in Civil Engineering 10 (4), 300–306.

Lingras, P., Mountford, P., 2001. Time delay neural networks designed using genetic algorithms for short-term inter-city traffic forecasting. Lecture Notes in Artificial Intelligence 2070, 290–299.

Lingras, P.J., Sharma, S.C., Osborne, P., Kalyar, I., 2000. Traffic volume time-series analysis according to the type of road use. Computer Aided Civil and Infrastructure Engineering 15, 365–373.

Lingras, P.J., Sharma, S.C., Zhong, M., 2002. Prediction of recreational travel using genetically designed regression and time delay neural network models. Transportation Research Record 1805, 16–24.

Loizos, A., Karlaftis, M., 2006. Neural networks and non-parametric statistical models: a comparative analysis in pavement condition assessment. Journal Advances and Applications in Statistics 6 (1), 87–110.

Longhi, S., Nijkamp, P., Reggianni, A., Maierhofer, E., 2005. Neural network modeling as a tool for forecasting regional employment. International Regional Science Review 28, 330–346.

Mandic, D.P., Chambers, J.A., 2001. Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures, and Stability. Wiley Series in Adaptive and Learning Systems for Signal Processing, Communications, and Control. John Wiley & Sons, NY.

Mantri, S., Bullock, D., 1995. Analysis of feedforward–backpropagation neural networks used in vehicle detection. Transportation Research Part C: Emerging Technologies 3 (3), 161–174.

McFadden, J., Yang, W.T., Durrans, S., 2001. Application of artificial neural networks to predict speeds on two-lane rural highways. Transportation Research Record 1751, 9–17.

Mohammadian, A., Miller, E.J., 2002. Nested logit models and artificial neural networks for predicting household automobile choices: comparison of performance. Transportation Research Record 1807, 92–100.

Mostafa, M.M., 2004. Forecasting the Suez Canal traffic: a neural network analysis. Maritime Policy & Management 31 (2), 139–156.

Mozolin, M., Thill, J.C., Lynn Usery, E.L., 2000. Trip distribution forecasting with multiplayer perceptron neural networks: a critical evaluation. Transportation Research Part B 34 (1), 53–73.

Mukkamala, S., Sung, A.H., 2003. Feature selection for intrusion detection using neural networks and support vector machines. Transportation Research Record 1823, 33–39.

Murata, N., Yoshizawa, S., Amari, S., 1994. Network information criterion: determining the number of hidden units for an artificial network model. IEEE Transactions on Neural Networks 5, 865–872.

Nakatsuji, T., Tanaka, M., Nasser, P., Hagiwara, T., 1998. Description of macroscopic relationships among traffic flow variables using neural network models. Transportation Research Record 1510, 11–18.

Nicholls, D., 1999. Statistics into the 21st century. Australia & New Zealand, Journal of Statistics 41 (2), 127–139.

Nijkamp, P., Reggiani, A., Tritapepe, T., 1996. Modelling inter-urban transport flows in Italy: a comparison between neural network approach and logit analysis. Transportation Research Part C 4, 323–338.

Oh, C., Ritchie, S.G., Oh, J.-S., 2005. Exploring the relationship between data aggregation and predictability toward providing better predictive traffic information. Transportation Research Record 1935, 28–36.

Pant, P.D., Balakrishnan, P., 1994. Neural network for gap acceptance at stop-controlled intersections. Journal of Transportation Engineering 120 (3), 432–446.

Park, D., Rilett, L., Han, G., 1999. Spectral basis neural networks for real-time travel time forecasting. Journal of Transportation Engineering 125 (6), 515–523.

Principe, J.C., Euliano, N.R., Lefebvre, C.W., 2000. Neural and Adaptive Systems: Fundamentals Through Simulations. John Wiley and Sons Inc.

Qiao, F., Yang, H., Lam, W.H.K., 2001. Intelligent simulation and prediction of traffic flow dispersion. Transportation Research Part B 35, 843–863.

Rao, P.V., Sikdar, P.K., Rao, K.V., Dhingra, S.L., 1998. Another insight into artificial neural networks through behavioural analysis of access mode choice. Computers, Environment and Urban Systems 22 (5), 485–496.

Ripley, B.D., 1996. Pattern Recognition and Neural Networks. Cambridge University Press, Cambridge.

Sadek, A.W., Spring, G., Smith, B.L., 2003. Toward more effective transportation applications of computational intelligence paradigms. Transportation Research Record 1836, 57–63.

Sarle, W.S., 1994. Neural networks and statistical models. In: Proceedings of the Nineteenth Annual SAS Users Group International Conference (April 1–13).

Sarvareddy, P., Al-Deek, H.M., Klodzinski, J., Anagnostopoulos, G., 2005. Evaluation of two modeling methods for generating heavy-truck trips at an intermodal facility by using vessel freight data. Transportation Research Record: Journal of the Transportation Research Board 1906, 113–120.

Sayed, T., Razavi, A., 2000. Comparison of neural and conventional approaches to mode choice analysis. Journal of Computing in Civil Engineering 14 (1), 23–30.

Sazi Murat, Y., 2006. Comparison of fuzzy logic and artificial neural networks approaches in vehicle delay modeling. Transportation Research Part C: Emerging Technologies 14, 316–334.

Shmueli, D., Salomon, I., Shefer, D., 1996. Neural network analysis of travel behavior: evaluating tools for prediction. Transportation Research Part C: Emerging Technologies 4 (3), 151–166.

Smith, B., Demetsky, M., 1997. Traffic flow forecasting: comparison of modeling approaches. Journal of Transportation Engineering 123 (4), 261–266.

Sommer, M., Herle, M., Hausler, J., Risser, R., Schutzhofer, B., Chaloupka, C., 2008. Cognitive and personality determinants of fitness to drive. Transportation Research Part F 11, 362–375.

Sousa, S.I.V., Martins, F.G., Alvim-Ferraz, M.C.M., Pereira, M.C., 2007. Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. Environmental Modelling & Software 22, 97–103.

Srinivasan, D., Jin, X., Cheu, R.L., 2004. Evaluation of adaptive neural network models for freeway incident detection. IEEE Transactions on Intelligent Transportation System 5 (1), 1–11.

Srinivasan, D., Choy, M.C., Cheu, R.L., 2006. Neural networks for real-time traffic signal control. IEEE Transactions on Intelligent Transportation Systems 7 (3), 261–272.

Stathopoulos, A., Dimitriou, L., Tsekeris, T., 2008. Fuzzy modeling approach for combined forecasting of urban traffic flow. Computer-aided Civil and Infrastructure Engineering 23, 521–535.

Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society B 36, 111–147.

Tang, Y.F., Lam, W.H.K., Ng, Pan.L.P., 2003. Comparison of four modeling techniques for short-term AADT forecasting in Hong Kong. Journal of Transportation Engineering 129 (3), 271–277.

Teng, H., Qi, Y., 2003a. Detection-delay-based freeway incident detection algorithms. Transportation Research Part C 11 (3/4), 265–287.

Teng, H., Qi, Y., 2003b. Application of wavelet technique to freeway incident detection. Transportation Research Part C 11 (3/4), 289–308.

Teodorovic, D., Varadarajan, V., Jovan, P., Chinnaswamy, M., Sharath, R., 2006. Dynamic programming—neural network real-time traffic adaptive signal control algorithm. Annals of Operation Research 143 (1), 123–131.

Terasvirta, T., Lin, C., Granger, C.W.J., 1993. Power of the neural network linearity test. Journal of Time Series Analysis 14 (2), 209–220.

Tillema, F., van Zuilekom, K.M., van Maarseveen, M.F.A.M., 2006. Comparison of neural networks and gravity models in trip distribution. Computer-aided Civil and Infrastructure Engineering 21, 104–119.

Tong, H.Y., Hung, W.T., 2002. Neural network modeling of vehicle discharge headway at signalized intersection: model descriptions and the results. Transportation Research Part A 36 (1), 17–40.

Tortum, A., Yayla, N., Gφkdağ, M., 2009. The modeling of mode choices of intercity freight transportation with the artificial neural networks and adaptive neuro-fuzzy inference system. Expert Systems with Applications: An International Journal 36 (3), 6199–6217.

Tsai, T.-H., Lee, C.-K., Wei, C.-H., 2009. Neural network based temporal feature models for short-term railway passenger demand forecasting. Expert Systems with Applications 36, 3728–3736.
Tseng, F.M., Yu, H.C., Tzeng, G.H., 2002. Combining neural network model with seasonal time series ARIMA model. Technological Forecasting & Social Change 69, 71–87.
Ulengin, F., Onsel, S., Topceu, Y., Aktas, E., Kabak, O., 2007. An integrated transportation decision support system for transportation policy decisions: the case of Turkey. Transportation Research Part A 41, 80–97.
Van der Voort, M., Dougherty, M., Watson, S., 1996. Combining Kohonen maps with ARIMA time-series models to forecast traffic flow. Transportation Research Part C 4 (5), 307–318.
Vlahogianni, E.I., 2009. Enhancing predictions in signalized arterials with information on short-term traffic flow dynamics. Journal of Intelligent Transportation Systems 13 (2), 73–84.
Vlahogianni, E.I., Golias, J.C., Karlaftis, M.G., 2004. Short-term traffic forecasting: overview of objectives and methods. Transport Reviews 24 (5), 533–557.
Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C., 2005. Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach. Transportation Research Part C 13, 211–234.
Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C., 2007. Spatio-temporal urban traffic volume forecasting using genetically-optimized modular networks. Computer-aided Civil and Infrastructure Engineering 22 (5), 317–325.
Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C., 2008. Temporal evolution of short-term urban traffic flow: a nonlinear dynamics approach. Computer-aided Civil and Infrastructure Engineering 23 (7), 536–548.
Vythoulkas, P.C., Koutsopoulos, H.N., 2003. Modeling discrete choice behavior using concepts from fuzzy set theory, approximate reasoning and neural networks. Transportation Research Part C: Emerging Technologies 11 (1), 51–73.
Wang, W., Chen, S., Qu, G., 2008. Incident detection algorithm based on partial least squares regression. Transportation Research Part C: Emerging Technologies 16, 54–70.
Warner, B., Misra, M., 1996. Understanding neural networks as statistical tools. The American Statistician 50 (4), 284–293.
Williams, T.P., Gucunski, N., 1995. Neural networks for backcalculation of moduli from SASW test. Journal of Computing in Civil Engineering 9 (1), 1–8.
White, H., 1989. Learning in artificial neural networks: a statistical perspective. Neural Computation 1, 425–464.
Wild, C.J., Pfannkuch, M., 1999. Statistical thinking in empirical enquiry. International Statistical Review 67 (3), 223–265.
Xie, C., Lu, J., Parkany, E., 2003. Work travel mode choice modeling with data mining: decision trees and neural networks. Transportation Research Record 1854, 50–61.
Xie, Y., Lord, D., Zhang, Y., 2007. Predicting motor vehicle collisions using Bayesian neural network models: an empirical analysis. Accident Analysis & Prevention 39 (5), 922–933.
Xu, J., Wong, S.C., Yang, H., Tong, C.-O., 1999. Modeling level of urban taxi services using neural network. Journal of Transportation Engineering 125 (3), 216–223.
Yang, H.H., Murata, N., Amari, S., 1998. Statistical inference. Learning in artificial neural networks. Trends in Cognitive Sciences 2 (1), 4–10.
Yang, J., Lu, J.J., Gunaratne, M., Dietrich, B., 2006. Modeling crack deterioration of flexible pavements: comparison of recurrent Markov chains and artificial neural networks. Transportation Research Record 1974, 18–25.
Yasdi, R., 1999. Prediction of road traffic using a neural network approach. Neural Computation and Application 8, 135–142.
Ye, Z., Shi, X., Strong, C.K., Greenfield, T.H., 2009. Evaluation of the effects of weather information on winter maintenance costs. Transportation Research Record 2107, 104–110.
Yin, H., Wong, S.C., Xu, J., 2002. Urban traffic flow prediction using fuzzy–neural approach. Transportation Research Part C 10 (2), 85–98.
Zhang, Y., Xie, Y., 2006. Application of genetic neural networks to real-time intersection accident detection using acoustic signal. Transportation Research Record 1968, 75–82.
Zhang, Y., Xie, Y., 2008. Travel mode choice modeling with support vector machines. Transportation Research Record 2076, 141–150.
Zhong, M., Sharma, S., Lingras, P., 2004. Genetically designed models for accurate imputation of missing traffic counts. Transportation Research Record 1879, 71–79.
Zhong, M., Sharma, S., Lingras, P., 2005. Short-term traffic prediction on different types of roads with genetically designed regression and time delay neural network models. Journal of Computing in Civil Engineering 19 (1), 94–103.