BUS TRAVEL TIME IN THE MIXED TRAFFIC BASED ON STATISTICA NEURAL NETWORK

Ibrahim Kamaruddin
University Teknology Petronas
Malaysia
Phone: +605-3688191
Fax: +605-3654090
ibrahim@petronas.com.my

Madzlan Napiah
University Teknology Petronas
Malaysia
Phone: +605-3688191
Fax: +605-3654090
madzlan_napiah@petronas.com.my

Suwardo
Gadjah Mada University
Indonesia
suwardo@yahoo.com

Oyas Wahyunggoro
Gadjah Mada University
Indonesia
oyas67@yahoo.com

Abstract

This paper presents the assessment of a number of factors affecting bus travel time and a relationship model between those factors and bus travel time. Statistica Neural Network (SNN) tool is used to solve this complex phenomenon. Data collected include bus travel time, distance, average speed, and number of bus stop. The results show that bus travel time is well predicted using variables of distance, average speed, and number of bus stops. The bus travel time increased due to the increase of distance and number of bus stops, while the higher the average speed from one stop to another, the lower bus travel time.

Keywords: bus travel time prediction, distance, average speed of bus, number of bus stop

Abstrak

Studi ini bertujuan untuk menentukan faktor-faktor yang mempengaruhi waktu tempuh bis dan hubungan antara waktu tempuh bis dengan faktor-faktor tersebut. Perangkat lunak Statistica Neural Network (SNN) digunakan untuk menyelesaikan fenomena yang kompleks ini. Data yang dikumpulkan meliputi waktu tempuh bis, jarak, kecepatan rata-rata, dan jumlah pemberhentian bis. Hasil studi ini menunjukkan bahwa waktu tempuh bis dapat diprediksi dengan bail dengan menggunakan variabel-variabel bebas jarak, kecepatan rata-rata, dan jumlah pemberhentian bis. Waktu tempuh bis meningkat seiring dengan peningkatan jarak dan jumlah pemberhentian bis. Waktu tempuh bis meningkat seiring dengan meningkatnya jarak dan jumlah pemberhentian bis, dan makin tinggi kecepatan rata-rata bis dari satu tempat pemberhentian ke tempat pemberhentian yang lain., makin rendah waktu tempuh bis tersebut.

Kata-kata kunci: prediksi waktu tempuh bis, jarak, kecepatan rata-rata bis, jumlah pemberhentian bis

INTRODUCTION

Travel time is important characteristic of bus service and is highly fluctuated due to the various factors and conditions, such as travel demands level, traffic control devices, incidents, and weather conditions. In advanced bus system operation, the travel time can be better controlled by advance devices than in the regular bus system with traditional operation. In the preliminary study for improvement of the bus service, the current bus system performance was needed to assess. The used of handheld GPS for data collection and performing on-board survey was the simple ways on how the transport planner could collect data while the advanced GPS data was not available.

Bus travel time prediction leading the bus operation timetable design is useful information for both bus operator and users. In fact, bus travel time is fluctuated based upon many factors of bus operation system. The fluctuation of bus travel time is an example of the complexity and multidimensional phenomenon in the mixed traffic. In bus operation, there was a well known simple relationship among travel time, distance and operating speed. Conventionally, the bus travel time is defined as the distance traveled divided by average operating speed.

Many studies have been done for many years in area of travel time prediction using different types of data and methods. The different types of data included detector data, probe vehicle data, and magnetic ticket data. There were many different methods proposed for predicting travel times, ranging from simple statistical approaches to sophisticated artificial intelligence and machine learning based algorithms.

The assessment of a number of factors affecting bus travel time and a study of relationship model between the factors and bus travel time are necessary to be provided for time table design. The determinant variables of travel time for bus travel time prediction are distance, number of bus stop, average speed approaching at bus stop, number of intersection, and delay approaching a bus stop. This study aims to assess some identified variables of bus travel time and to model the bus travel time prediction based on the GPS data.

Two main objectives to be achieved in this study are as follows: (1) To assess a number of factors affecting bus travel time and to study the relationship between those factors and bus travel time, and (2) To model and predict bus travel time based on a number of variables compiled from GPS data collected during on-board survey. The scope of this study covered a number of materials and specific task as briefly explained as follows:

1. The data were collected using handheld GPS from a regular bus system traditionally operated in rural area with minimum congestion, while on-board survey was conducted. The existing bus system was an intercity regular bus operated in the mixed traffic.
2. GPS data included position and time of the operating bus along the route in an intercity corridor. The bus service could stop randomly at any place or at bus stop where the passengers board and alight.
3. At position and time where the bus stopped, a waypoint was entered and the time was recorded automatically into a handheld GPS, and at the same time a number of passengers boarding and alighting were recorded in the prepared worksheet.
4. Data was compiled from records of handheld GPS, including bus travel time, distance, average speed, and number of bus stops.

Neural Networks Overview

The basic concepts of neural networks is briefly described by StatSoft, Inc. (2004). STATISTICA Neural Networks (SNN) is a neural networks application available from a particularly comprehensive neural network tool. Neural networks are being successfully applied in many areas as diverse as finance, medicine, engineering, geology and physics.

Indeed, anywhere that there are problems of prediction, classification or control, neural networks are being introduced.

There are two key factors bringing the neural network successfully applied. First, neural network has a power of use. Neural networks are very sophisticated modeling techniques capable of modeling extremely complex functions, non-linear function, optimization strategies with linear approximation. Neural networks also powerful in attempting to model non-linear functions with large numbers of variables with multiple dimensionality problems. Second, neural network is ease of use. The neural network user gathers representative data, and then invokes training algorithms to automatically learn the structure of the data. The level of user knowledge to successfully apply neural networks is much lower than using traditional non-linear statistical methods. Also, the users can easily select and prepare data, select an appropriate neural network, and interpret the results (Bishop, 1995).

Generally, prediction problems may be divided into two main categories, classification and regression. The objective of classification is to determine to which of a number of discrete classes a given input case belongs. Meanwhile, the objective of regression is to predict the value of a (usually) continuous variable: tomorrow's stock price, the fuel consumption of a car, next year's profits. In this case, the output required is a single numeric variable.

In application, neural networks can actually perform a number of regression and/or classification tasks at once, although commonly each network performs only one. In many cases, therefore, regression needs to define a single network with multiple input variables and single output variables.

Model Profiles in Neural Network

Model profiles in Neural Network indicate the architecture of network and ensembles. A model profile consists of a type code followed by a code giving the number of input and output variables and number of layers and units (network) or members (ensembles). The codes and model types in Neural Network are shown in Table 1. For example, the model profiles given are MLP 1:1-1-1:1, Linear 3:3-1:1, and RBF 3:3-4-1:1. The type of MLP 1:1-1-1:1 indicates a multilayer perceptron network with 1 input variable, 1 output variable, 1 input neuron, 1 hidden neuron, and 1 output neuron. The type of Linear 3:3-1:1 indicates a linear network with 3 input variables, 1 output variable, 3 input neurons, no hidden neuron, and 1 output neuron. The type of RBF 3:3-4-1:1 indicates a radial basis function network with 3 input variables, 1 output variable, 3 input neurons, 4 hidden neurons, and 1 output neuron.

Table 1 Codes and Model Types of Network

Code	Model Type
MLP	Multilayer Perceptron Network
RBF	Radial Basis Function Network
SOFM	Kohonen Self-Organizing Feature Map
Linear	Linear Network
PNN	Probabilistic Neural Network
GRNN	Generalized Regression Neural Network
PCA	Principle Component Network
Cluster	Cluster Network
Output	Output Ensemble
Conf	Confidence Ensemble

MATERIALS AND METHODS

The Ipoh-Lumut corridor is located at Perak State in Peninsular Malaysia which links the Ipoh City (State Center) and Lumut Port (where the Malaysia Navy and Pangkor Island, tourism place located). The land use in this corridor is potential to generate and attract many trips. This corridor is also categorized as a rural area with some important central business district (CBD) including residential and school areas, state buildings, industrial and commercial areas, and some attractive universities. Regular bus services are operated from 7:00 am to 9:00 pm. Figure 1 shows the illustration of bus service operation, such as location, departure time, distance, and travel time. The main bus stop locations are indicated in the figure, but the bus actually can stop any where along the route for boarding and alighting of passengers.

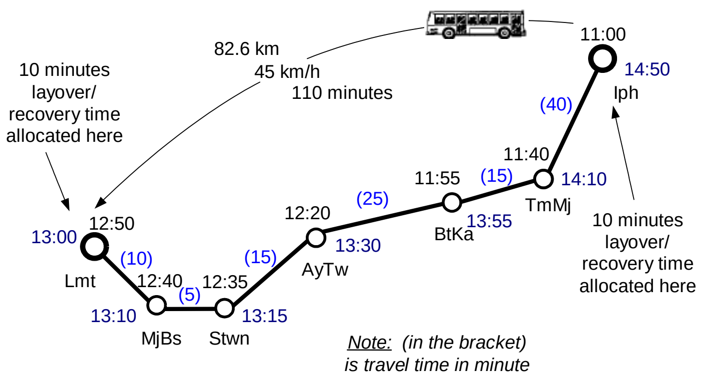


Figure 1 Layout of Bus Operation Schedule

The data were collected during one full day period (07:00-21:00), one week period (11:00-15:00) and one year period (11:00-15:00). The primary data used were collected by on-board survey over Perak Roadways's 14-hour weekday service. The on-board survey was done along with Ipoh-Lumut corridor intercity regular bus service from January to December 2007 during workday and weekend.

Figure 2 showed the time-distance diagram for on-board survey. The number of data points used for analysis were 48 trips per year. The primary data collected include: (1) arrival and departure times of the bus at stop points, (2) distance of location of stop points (bus station, bus stop, and non bus stop), and (3) name of location and environmental situation within the bus route. In addition, secondary data were also used to help surveyor on the primary data collection, such as road network map, timetable and other information on existing bus service.

On-board survey was conducted with more accurate and detailed information gathered about vehicle movement and boarding and alighting of passengers compared to at check point survey method. The survey was conducted by an observer getting on a vehicle traveling over the route for a period of time. The observer recorded the data of location and the time at which the vehicle stopped for boarding and alighting of passengers. The observer also counted the boarding and alighting passengers over the entire route for a specified time period. Since there was no device installed in the bus service system, neither in advanced bus service in other city, there was no advanced GPS data collection in this case.

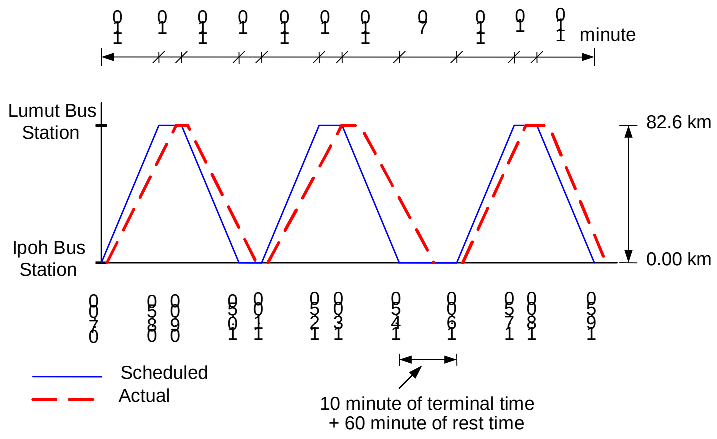


Figure 2 Time-Distance Diagram of Bus Operation

Model Performance Evaluation

Indicators of models performance were measured using the similarity or closeness between predicted and actual values. The statistical closeness was generally stated in term of MARE (Mean Absolute Relative Error) and MAPE (Mean Absolute Percentage Error). MARE and MAPE values represent average difference and percentage difference between the actual or observed and predicted travel time from current bus stop to target bus stop.

Practically, the prediction of bus travel time with the model obtained was evaluated using the MARE and MAPE values. The appropriate model is determined by comparing the MARE and MAPE values among the tested models as indicated by the small value of MARE and MAPE (Jeong et al. 2004).

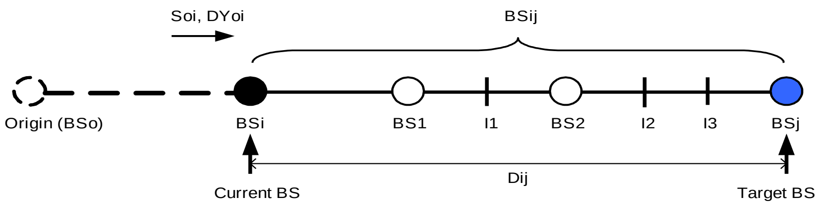
RESULTS AND DISCUSSIONS

The variables used for building model of multiple regressions are described in Figure 3. Data for both directions were completely prepared with accounted independent and dependent variables. The dependent variable is bus travel time from current bus stop to target bus stop in minute (TTij), and the independent variables include:

1. distance from bus stop i-th to j-th in kilometer, Dij
2. average speed from origin (bus stop o) to current bus stop (bus stop i-th) in km/h, Soi
3. number of bus stop from bus stop i-th to j-th, BSij
4. number of intersection from current bus stop to target bus stop, Iij
5. delay from origin bus stop to current bus stop in minute, DYoi

Summary of Models

Table 2 shows the model summary reports for both Ipoh to Lumut and Lumut to Ipoh directions. Clearly, it was shown the list of model profile which was trained and selected by using the STATISTICA neural network tool. Five model profiles were resulted successfully reflecting the network set. Again, as the network's performance improves, the ratio becomes closer to zero. In the model summary report, for both Ipoh to Lumut and Lumut to Ipoh directions, two model profiles obtained are the same. It is clearly revealed that model profile RBF 3:3-4-1:1 is the best model due to the lowest ratio of training error, selection error, and test error. The results show that the network successfully use the information in the input variables, as it is indicated that the ratio much smaller than 1.0. In addition, the RBF 3:3-4-1:1 model profile has the lowest errors due to training, selection, and test steps.



Note:

BS_o : bus stop at origin of trip
BS_i : current bus stop
BS_j : destination or target bus stop
BS_{ij} : number of bus stop from bus stop *i*-th to *j*-th
i,j : 1, 2, 3, etc.
I_{ij} : Number of intersection from current bus stop to target bus stop.
D_{ij} : distance from bus stop *i*-th to *j*-th in kilometer
S_{oi} : average speed from origin to current bus stop (bus stop *i*-th) in km/h
DY_{oi} : delay from origin bus stop to current bus stop in minute
TT_{ij} : travel time from current bus stop to target bus stop in minute

Figure 3 Modeling Illustration

Table 2 Model Summary Report

Model Summary Report (SNN_TravelTime) - Ipoh to Lumut Direction:

Index	Profile	Train Perf.	Select Perf.	Test Perf.	Train Error	Select Error	Test Error	Training/ Members	Inputs	Hidden (1)	Hidden (2)
1	MLP 1:1-1-1:1	0,170484	0,114694	0,125662	0,054068	0,041354	0,045469	BP100,CG20,CG26b	1	1	0
2	Linear 2:2-1:1	0,152120	0,113629	0,126628	0,048238	0,038459	0,045077	PI	2	0	0
3	Linear 3:3-1:1	0,151690	0,110967	0,120777	0,048102	0,038047	0,043956	PI	3	0	0
4	RBF 3:3-2-1:1	0,175244	0,136374	0,167122	0,004919	0,003976	0,005284	KM,KN,PI	3	2	0
5	RBF 3:3-4-1:1	0,155083	0,112281	0,118879	0,004353	0,003417	0,003929	KM,KN,PI	3	4	0

Model Summary Report (SNN_TravelTime) - Lumut to Ipoh Direction:

Index	Profile	Train Perf.	Select Perf.	Test Perf.	Train Error	Select Error	Test Error	Training/Members	Inputs	Hidden (1)	Hidden (2)
1	Linear 1:1-1:1	0,156966	0,199393	0,221850	0,048831	0,069657	0,056863	PI	1	0	0
2	MLP 1:1-4-1:1	0,130859	0,70334	0,162832	0,040710	0,059517	0,041869	BP100,CG20,CG8b	1	4	0
3	MLP 1:1-3-1:1	0,130579	0,170163	0,160940	0,040626	0,059432	0,041422	BP100,CG20,CG7b	1	3	0
4	RBF 3:3-2-1:1	0,135515	0,174622	0,180990	0,003584	0,005231	0,004051	KM,KN,PI	3	2	0
5	RBF 3:3-4-1:1	0,131865	0,173098	0,173334	0,003488	0,005153	0,003889	KM,KN,PI	3	4	0

There are 5 model profiles listed with their network details. Index 1 to 5 indicate that there are 5 retained networks is selected based on the criteria of balance error against diversity

The illustration of structure of selected network is shown in Figure 4. Type of model profile RBF 3:3-4-1:1 indicates a radial basis function network with 3 input variables, 1 output variable, and three layers of neuron (3 input neurons, 4 hidden neurons, and 1 output neuron). The symbols used in the illustration of a neural network are explained as the following: (a) red color for positive activation levels and (b) green color for negative activation levels.

Neurons are represented using one of several shapes:

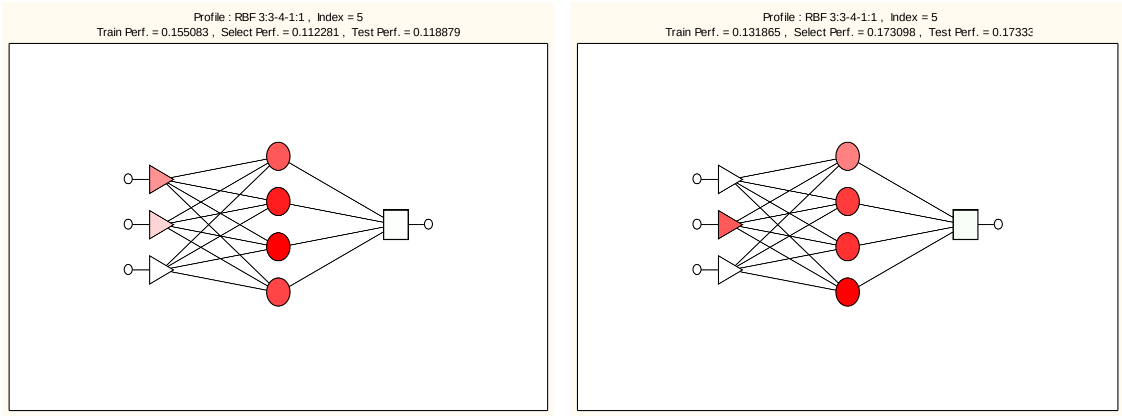
a. Triangles; pointing to the right indicate input neurons, and these neurons perform no processing, and simply introduce the input values to the network.
b. Squares; indicate Dot Product synaptic function
c. Circles; indicate Radial synaptic function units.
d. Small open circles; illustrate input and output variables corresponding to input or output neuron. In some circumstances (nominal variables and time series inputs) a number of neurons are joined to a single input or output variable.

In both selected model profiles, the type of training algorithms used to optimize the neural network were KM, KN, and PI, with:

KM: K-Means (Center Assignment)

KN: K-Nearest Neighbor (Deviation Assigment)

PI: Pseudo-Invert (Linear Least Square Optimization)



(a) Ipoh to Lumut Direction (b) Lumut to Ipoh Direction

Figure 4 The Structure of Selected Network

Sensitivity Analysis

Sensitivity analysis was conducted on each model and the best result was displayed in a spreadsheet (See Table 3). In Ipoh to Lumut direction, the importance (ranking) variables are distance (Dij), number of bus stop (BSij), and average speed before current bus stop (Soi), respectively. For the opposite direction, from Lumut to Ipoh direction, where distance is the most important variable followed by the number of bus stop and the average speed before current bus stop.

Table 3 Sensitivity Analysis

Ipoh to Lumut Direction			
	Dij	Soi	BSij
Ratio.5	3,996254	1,125538	2,946107
Rank.5	1,000000	3,000000	2,000000

Lumut to Ipoh Direction			
	Dpq	Sop	BSpq
Ratio.5	3,319056	1,127048	3,197389
Rank.5	1,000000	3,000000	2,000000

Descriptive Statistics and Residual Analysis

Five regression model profiles were generated successfully from the three network types tested (linear, radial basis function or RBF and three layer perceptron or MLP). Those five regression model profiles were shown in index 1 to 5 i.e. MLP 1:1-1-1:1, Linear 2:2-1:1, Linear 3:3-1:1, RBF 3:3-2-1:1, and RBF 3:3-4-1:1. Table 4 provided descriptive statistic and residual of five model profiles including data mean, data standard deviation, error mean, error standard deviation, absolute error mean, standard deviation ratio, and correlation. Column TTij.5 and TTpq.5 were the descriptive statistic of the selected model profiles for Ipoh to Lumut and Lumut to Ipoh directions, respectively.

Table 4 Descriptive Statistic of the Five Model Profiles

	Regression (1-5) (SNN_TravelTime)					*Regression (1-5) (SNN_TravelTime)*				
Description	Ipoh to Lumut Direction					Lumut to Ipoh Direction				
	TTij.1	TTij.2	TTij.3	TTij.4	TTij.5	TTpq.1	TTpq.2	TTpq.3	TTpq.4	TTpq.5
Data Mean	58,18750	58,18750	58,18750	58,18750	58,18750	61,12500	61,12500	61,12500	61,12500	61,12500
Data S.D.	35,29852	35,29852	35,29852	35,29852	35,29852	37,42444	37,42444	37,42444	37,42444	37,42444
Error Mean	1,02454	0,83131	0,90335	0,90939	0,96912	-0,10082	-0,08933	-0,12840	-0,64011	-0,53987
Error S.D.	5,29755	4,89973	4,83794	5,89379	4,91523	6,74521	5,52025	5,49987	5,79749	5,66007
Abs E. Mean	4,34010	3,80911	3,69528	4,67228	3,80877	5,42377	4,23941	4,22724	4,55913	4,51966
S.D. Ratio	0,15008	0,13881	0,13706	0,16697	0,13925	0,18024	0,14750	0,14696	0,15491	0,15124
Correlation	0,98879	0,99062	0,99083	0,98616	0,99043	0,98363	0,98907	0,98916	0,98795	0,98852

Evaluation of Models and User Defined Cases Prediction

Evaluation of model performance separately was carried out by calculating the MARE and MAPE values. MARE and MAPE value of the same model profile (RBF 3:3-4-1:1) for both directions were shown in Table 5. For the selected model profile (RBF 3:3-4-1:1) for Ipoh to Lumut direction, the MARE value is 3.81 minutes and the MAPE value is 7.56%. Meanwhile, for Lumut to Ipoh direction, both MARE and MAPE values are 4.52 minutes and 8.66%, respectively. The values of MARE and MAPE of both models are small and reasonable. In other word, the error or delay of 3.81 and 4.52 minutes were not significant compared to the average bus travel time of 117 and 123 minutes, respectively.

Table 5 Performance Values of the Model Profiles

	Ipoh to Lumut Direction				
	$TT_{ij.1}$	$TT_{ij.2}$	$TT_{ij.3}$	$TT_{ij.4}$	$TT_{ij.5}$ *)
MARE (minute)	4,34	3,81	3,70	4,67	3,81
MAPE (%)	8,78	7,53	7,11	10,12	7,56
	Lumut to Ipoh Direction				
	$TT_{pq.1}$	$TT_{pq.2}$	$TT_{pq.3}$	$TT_{pq.4}$	$TT_{pq.5}$ *)
MARE (minute)	5,42	4,24	4,23	4,56	4,52
MAPE (%)	11,27	7,99	7,97	8,59	8,66

Note: *) selected model

CONCLUSIONS

The results show that the prediction of bus travel time from the current bus stop to the target bus stop was well identified. In this case, the radial basis function (RBF 3:3-4-1:1) is the best model profile for bus travel time prediction compared to linear and multilayer perceptrons (MLP) model profile based on the smallest regression ratio and error of model selection. A number of related factors, such as distance, average speed, and number of bus stops, are significant in influencing the change of bus travel time. Those input variables selected are meaningful in predicting the bus travel time. The bus travel time increase as the distance and number of bus stops from current bus stop to target bus stop increase. Meanwhile, the higher the average speed approaching current bus stop, the lower the bus travel time from current bus stop to target bus stop.

RBF 3:3-4-1:1 model profile is adequate to model the bus travel time prediction. The MARE and MAPE values of the models are small and reasonable.

ACKNOWLEDGEMENT

The authors would thank the Perak Roadways Sdn. Bhd. for allowing them to conduct on-board surveys for data collection. They also appreciated to anonymous referees for their suggestion on the previous manuscript, and to University Teknologi PETRONAS for the financial support to conduct this research.

REFERENCES

Bishop, C. 1995. *Neural Networks for Pattern Recognition*. Oxford: University Press.

Jeong, R., and Rilett, L. 2004. *The Prediction of Bus Arrival Time using AVL Data*. 83rd Annual General Meeting, Transportation Research Board. Washington, DC: National Research Council.

StatSoft, Inc. 2004. *STATISTICA (data analysis software system), Version 7* (Online), (www.statsoft.com).