McMaster University

MICHAEL G. DEGROOTE SCHOOL OF BUSINESS

RESEARCH AND WORKING PAPER SERIES

CALCULATION OF CONFIDENCE INTERVALS FOR ESTIMATED DISTANCES

By

Halit Üster and **Robert. F. Love**

Michael G. DeGroote School of Business
McMaster University
Hamilton, Ontario

Working Paper # 441

August, 1999

ER UNIVERSITY
Street West
Ontario, Canada L8S 4M4
(905) 525-9140

Abstract

A new method is devised to calculate the confidence intervals for estimated distances. Using this method, the confidence intervals for estimated actual distances are developed for the ℓ_p-norm and ℓ_{bp}-norm. Our empirical study in the seventeen geographical regions indicates that better confidence intervals for the unknown actual distances are obtained with the ℓ_{bp}-norm than the ℓ_p-norm.

1 Introduction

Distance functions are utilized in several applications such as distribution and transportation planning (Eilon et al., 1971; Westwood, 1977), accuracy validation of actual transportation network distance data (Ginsburgh and Hansen, 1974), response-time models for emergency vehicles (Kolesar et al., 1975), construction of Voronoi Diagrams of a region (Klein, 1988), location-allocation problems (Love et al., 1988), and Geographic Information Systems (Star and Estes, 1990). The software packages Roadnet (Roadnet-Technologies, 1993) and TruckStops2 (MicroAnalytics, 1993) utilize distance predicting functions as a substitute to forming large files of distance data. Distance functions appear within the context of larger models such as facilities location and location-allocation problems (Love et al., 1988). Distance functions are also used in cluster analysis. Murray and Estivill-Castro (1998) examine the effects of the distance function choice on forming cluster regions. Additionally, norm-based distance functions can be employed for generating random problem instances in discrete location or transportation models since they satisfy the triangle inequality by nature.

The Weighted Sum of Order p, denoted by $\ell_{bp}(\mathbf{x})$, can be utilized to estimate distances in a transportation network. The $\ell_{bp}(\mathbf{x})$ distance between any two points $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$ in 2-dimensional Euclidean plane is given by

$$\ell_{bp}(\mathbf{u}, \mathbf{v}) = [b_1 \, |u_1 - v_1|^p + b_2 \, |u_2 - v_2|^p]^{1/p}, \quad b_1, b_2 > 0, \;\; p \geq 1. \tag{1}$$

Love and Morris (1972) introduce the concept of axis rotation in their study on the road network in Milwaukee, Wisconsin. Brimberg, Love and Walker (1995) investigate this concept in detail and conclude that a reference axis rotation chosen to align with the underlying pattern of the transportation network improves the accuracy of distance predictions. Huriot and Perreur (1973) also discuss axis rotation and apply it in a study of the nine largest Swiss cities. Incorporating the axis rotation angle θ into the ℓ_{bp}-norm we have

$$\ell_{bp\theta}(\mathbf{u}', \mathbf{v}') = [b_1 \, |u_1{}' - v_1{}'|^p + b_2 \, |u_2{}' - v_2{}'|^p]^{1/p} \tag{2}$$

where

$$\mathbf{u}' = (u'_1, u'_2),\ \ \mathbf{v}' = (v'_1, v'_2),\ \ \theta \in [0, \pi/2], \quad \text{and}$$

$$\begin{pmatrix} u_1{}' & u_2{}' \\ v_1{}' & v_2{}' \end{pmatrix} = \begin{pmatrix} u_1 & u_2 \\ v_1 & v_2 \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{3}$$

The function $\ell_{bp}(\mathbf{x})$ is a norm and is thus a convex function in $\mathbf{x}$ (Üster and Love, 1998). The ℓ_{bp}-norm is a generalization of the well-known weighted ℓ_p-norm. If for a fixed p, the equality $b_1{}^{1/p} = b_2{}^{1/p} = k$ holds, then one obtains the weighted ℓ_p-norm where k represents the weight. Furthermore, if $b_1 = b_2 = 1$, the rectangular and the Euclidean distances can be obtained from the ℓ_{bp}-norm by setting $p = 1$ and $p = 2$, respectively.

With the ℓ_{bp}-norm one introduces unequal weights or non-symmetric distance irregularities along the axis directions. An empirical work on seventeen geographic regions showed that the ℓ_{bp}-norm is better than the weighted ℓ_p-norm in terms of the accuracy of distance estimations (Üster and Love, 1998). Particularly in geographical regions with a predominant direction of nonlinearity (e.g. a mountain range), the gain in the accuracy of distance estimations with the ℓ_{bp}-norm is more pronounced. Furthermore, although the ℓ_{bp}-norm is a three parameter (b_1, b_2 and p) distance function as opposed to a two parameter weighted ℓ_p-norm (k and p), the convexity of the goodness-of-fit criterion function in parameters b_1 and b_2 provides a close performance in speed for distance fitting algorithms of these norms.

Besides its higher accuracy in predicting distances in a geographical region, the weighted sum of order p also seems well-suited to some specific applications. Foulds and Hamacher (1990) presents an example in the field of robotics when the movement is restricted to two directions, generated by one motor for each direction. In such an environment and under the assumption that the motors are not allowed to work simultaneously and their constant speeds are equal, the ℓ_1-norm can effectively be used to determine the minimal time necessary to move from one position in the plane to another. If the assumption of equal speeds of motors is relaxed, then the ℓ_{bp}-norm with $p = 1$, and b_1, b_2 proportional to the speeds in the corresponding directions, is a suitable function to determine the required minimal travel time. Similar examples are harbor cranes used to move containers around and plotters used to plot engineering drawings.

Love and Dowling (1985) examine the effect of "doubling back" by fitting the weighted ℓ_p-norm to a sample of layout patterns which are basically rectangular. They find that while the best value of p stays close to 1 (representing the rectangularity of the layout), the extra travel distance caused by doubling back is captured by the parameter k. With the weighted ℓ_p-norm it is assumed that the effect of doubling back is equal in both horizontal and vertical travel directions. This may not be the case in some types of layouts with dominant directions of travel, such as parallel bays along either horizontal or vertical axis. Suppose that the bays are horizontal and the distance between points $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$, which are located in different bays, is to be measured. Then, because of the doubling back effect, we expect to travel more than $|u_1 - v_1|$ in the horizontal direction. In such a layout, b_1 will be greater than b_2 in the ℓ_{bp} distance. In other words, ℓ_{bp}-norm will capture the directional nonlinearity and provide more accurate representation of the underlying pattern than the rectangular distances.

The optimum parameter values of the distance function are determined so that a criterion value is minimized. The criterion also provides the means to measure the accuracy of a distance function. The criterion usually measures the aggregate amount of error generated by a particular distance function with known parameter values. Three criteria have been used to model the parameters of a distance function to a set of data representing a region of interest: Sum of Absolute Deviations (AD); Sum of Squared Deviations (SD); Sum of Normalized Absolute Deviations (NAD). Let $d(\mathbf{a}_i, \mathbf{a}_j)$ be the predicted distance between points $\mathbf{a}_i$ and $\mathbf{a}_j$. $A(\mathbf{a}_i, \mathbf{a}_j)$ is the actual distance between $\mathbf{a}_i$ and $\mathbf{a}_j$, and n is the number of points in the data set. Then the mathematical expressions for the goodness-of-fit criteria are as follows:

$$\begin{aligned} AD &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} |d(\mathbf{a}_i, \mathbf{a}_j) - A(\mathbf{a}_i, \mathbf{a}_j)| \\ SD &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{(d(\mathbf{a}_i, \mathbf{a}_j) - A(\mathbf{a}_i, \mathbf{a}_j))^2}{A(\mathbf{a}_i, \mathbf{a}_j)} \end{aligned}$$

$$NAD = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{|d(\mathbf{a}_i, \mathbf{a}_j) - A(\mathbf{a}_i, \mathbf{a}_j)|}{A(\mathbf{a}_i, \mathbf{a}_j)}$$

The AD and SD criteria were introduced by Love and Morris (1972), and the NAD criterion by Love and Walker (1991). Several applications of the criteria are found in Berens (1988); Berens and Körling (1985); Brimberg, Dowling and Love (1996); Brimberg, Love and Walker (1995); Love and Morris (1972; 1979); Love, Walker and Tiku (1995); and Ward and Wendell (1980; 1985).

Love and Üster (1996) have conducted a detailed study on the statistical comparison of these criteria by using the weighted ℓ_p-norm with axis rotation. The comparisons were carried out using the estimation error distributions for seventeen geographical regions. The authors considered two estimation error related random variables, $e(\mathbf{a}_i, \mathbf{a}_j)$ and $|e(\mathbf{a}_i, \mathbf{a}_j)|/A(\mathbf{a}_i, \mathbf{a}_j)$, where $e(\mathbf{a}_i, \mathbf{a}_j) = A(\mathbf{a}_i, \mathbf{a}_j) - k\ell_p(\mathbf{a}_i, \mathbf{a}_j)$, to statistically compare the criteria in several aspects such as the homoscedasticity and the expected value of $e(\mathbf{a}_i, \mathbf{a}_j)$, the homoscedasticity of $|e(\mathbf{a}_i, \mathbf{a}_j)|/A(\mathbf{a}_i, \mathbf{a}_j)$, and the accuracy in predicting long and short distances. The only statistically significant difference found is that, for the SD criterion, the $e(\mathbf{a}_i, \mathbf{a}_j)$ have an expected value of zero without any exceptions in all the regions. Therefore, in this study we will consider the distance estimation errors generated by the SD criterion. Üster and Love (1998) used the SD criterion to model the ℓ_{bp} distances in seventeen geographical regions and found that ℓ_{bp}-norm generates lower SD values and thus better distance estimations. The parameter values of the ℓ_{bp} distances are reported in Table 1.

A confidence interval is calculated by using the estimated distance and it provides the analyst a range in which the actual distance lies in with a predetermined level of expectation. Furthermore, the range provides insight about the accuracy of the estimation and the performance of the distance function employed.

Another application of confidence intervals for road distance is found in the verification of road distance data. Ginsburgh and Hansen (1974) describe an ad hoc range into which the estimated distance must fall to be assumed acceptable. However, the authors do not provide any analytical justification for the range.

No.	Region	θ	b_1	b_2	p	SD
1	Australia	86	1.1945	1.2987	1.7996	1117.89
2	BC Province	66	1.3415	1.4859	1.5635	1015.41
3	Canada	45	1.9325	1.3500	2.2317	497.01
4	France	76	1.1529	1.0819	1.8576	78.87
5	Great Britain	0	1.1116	1.3925	2.0352	172.34
6	NY State	86	1.1303	1.1285	1.5841	159.80
7	Pennsylvania	44	1.2250	1.4812	2.3892	95.11
8	United States	0	1.1358	1.1748	1.6956	336.54
9	Brussels	47	1.0696	1.1177	1.7969	3.47
10	London Central	26	1.2366	1.6162	2.2708	15.01
11	London North	14	1.0930	1.2953	1.7901	1.36
12	Los Angeles	0	1.1011	1.3799	1.7750	13.18
13	NY City	6	1.1673	1.2439	1.7539	13.29
14	Paris	40	1.3634	1.1440	2.2734	5.84
15	Sydney	8	1.3675	1.1521	1.5571	1.10
16	Tokyo	57	1.1988	1.1670	1.8260	2.28
17	Toronto	87	1.0581	1.0275	1.2009	5.06

Table 1: Best Parameter Values of the ℓ_{bp}-norm

In the literature there are two papers which discuss the statistical properties of errors and confidence intervals for estimated distances. Love, Walker and Tiku (1995) use the weighted ℓ_p-norm ($k\ell_p$) as the distance predicting function and the sum of Normalized Deviations (NAD) as the goodness-of-fit criterion. The authors investigate two types of prediction errors: $e(\mathbf{a}_i, \mathbf{a}_j)$ and $e(\mathbf{a}_i, \mathbf{a}_j)/A(\mathbf{a}_i, \mathbf{a}_j)$, where $e(\mathbf{a}_i, \mathbf{a}_j) = A(\mathbf{a}_i, \mathbf{a}_j) - k\ell_p(\mathbf{a}_i, \mathbf{a}_j)$, in seventeen geographical regions (nine large geographical regions and eight urban centers). The sample point coordinates and the actual distance data for these seventeen geographical regions are given by Love and Walker (1993). By inspecting the sample Pearson coefficients $\sqrt{b_1}$ (skewness) and b_2 (kurtosis) they conclude that both types of error distributions are non-normal. On the other hand, while the scatter of points for the $e(\mathbf{a}_i, \mathbf{a}_j)$ versus $A(\mathbf{a}_i, \mathbf{a}_j)$ plots for the regions was contained in a diverging funnel that was symmetric about zero, the scatter of points for the $e(\mathbf{a}_i, \mathbf{a}_j)/A(\mathbf{a}_i, \mathbf{a}_j)$ versus $A(\mathbf{a}_i, \mathbf{a}_j)$ plots was contained within a narrow band which was symmetric about zero. This observation suggested that the $e(\mathbf{a}_i, \mathbf{a}_j)$ distributions were heteroscedastic but the $e(\mathbf{a}_i, \mathbf{a}_j)/A(\mathbf{a}_i, \mathbf{a}_j)$ distributions were homoscedastic. Using these empirical results it was assumed that $\mu_e = 0$ and $\sigma_e^2 = \sigma^2 A(\mathbf{a}_i, \mathbf{a}_j)^2$ where μ_e is the mean of $e(\mathbf{a}_i, \mathbf{a}_j)$ distribution, and σ_e^2 and σ^2 are the variances of $e(\mathbf{a}_i, \mathbf{a}_j)$ and $e(\mathbf{a}_i, \mathbf{a}_j)/A(\mathbf{a}_i, \mathbf{a}_j)$ distributions, respectively. Based on these results, authors devise the following confidence interval for an unknown distance $A(\mathbf{a}_i, \mathbf{a}_j)$: $c_1 < A(\mathbf{a}_i, \mathbf{a}_j) < c_2$, where $c_1 = (z_1 s + 1)\ k\ell_p(\mathbf{a}_i, \mathbf{a}_j)$, $c_2 = (z_2 s + 1)\ k\ell_p(\mathbf{a}_i, \mathbf{a}_j)$. The values of z_1 and z_2 are found by using the sample Pearson coefficients of $e(\mathbf{a}_i, \mathbf{a}_j)$ distribution from the tables provided by Johnson, Nixon and Amos (1963), and s^2 is the sample unbiased estimator of σ^2.

In another study, Brimberg, Dowling and Love (1994)use the weighted one-two norm as the distance predicting function. The weighted one-two norm is given by the expression $h(\mathbf{u}, \mathbf{v}; \beta_0, \beta_1) = \beta_o \ell_2(\mathbf{u}', \mathbf{v}') + \beta_1 \ell_2(\mathbf{u}', \mathbf{v}')$, where $\mathbf{u}, \mathbf{v} \in \Re^2$, $\beta_0, \beta_1 \geq 0$ and $\mathbf{u}', \mathbf{v}'$ are as defined in (3). Authors apply the weighted one-two norm model by using the regression R-square value as the goodness-of-fit criterion in regions Toronto and Ontario. They also consider two types of errors: $e(\mathbf{a}_i, \mathbf{a}_j)$ and $e(\mathbf{a}_i, \mathbf{a}_j)/\ell_2(\mathbf{a}_i, \mathbf{a}_j)$, where $e(\mathbf{a}_i, \mathbf{a}_j) = A(\mathbf{a}_i, \mathbf{a}_j) - h(\mathbf{a}_i, \mathbf{a}_j; \beta_0, \beta_1)$. Both error distributions are assumed to have a mean zero. To test the normality and homoscedasticity of these distributions authors use the normal probability plots and non-parametric Smirnov test, respectively. It is found that $e(\mathbf{a}_i, \mathbf{a}_j)$ distributions are non-normal and heteroscedastic in both regions. Additionally, $e(\mathbf{a}_i, \mathbf{a}_j)/\ell_2(\mathbf{a}_i, \mathbf{a}_j)$ distributions are heteroscedastic in both regions and from a normal population in Toronto, but from a non-normal population in Ontario.Next, authors analyze the outliers in $e(\mathbf{a}_i, \mathbf{a}_j)/\ell_2(\mathbf{a}_i, \mathbf{a}_j)$ distributions and find eight unusual observations for Ontario and five unusual observations for Toronto. After excluding the outliers in Ontario and fitting the distance function again, $e(\mathbf{a}_i, \mathbf{a}_j)$ distribution stays non-normal and heteroscedastic but $e(\mathbf{a}_i, \mathbf{a}_j)/\ell_2(\mathbf{a}_i, \mathbf{a}_j)$ distribution becomes normal and homoscedastic. For Toronto, excluding only two outliers provides normality and homoscedasticity for the $e(\mathbf{a}_i, \mathbf{a}_j)/\ell_2(\mathbf{a}_i, \mathbf{a}_j)$ distribution. Based on these results, the confidence intervals are devised by using $e(\mathbf{a}_i, \mathbf{a}_j)/\ell_2(\mathbf{a}_i, \mathbf{a}_j)$ variable where outliers are excluded. The confidence interval is $c_1 < A(\mathbf{a}_i, \mathbf{a}_j) < c_2$, where $c_1 = h(\mathbf{a}_i, \mathbf{a}_j, \beta_0, \beta_1) - z\, s\, \ell_2(\mathbf{a}_i, \mathbf{a}_j)$, $c_2 = h(\mathbf{a}_i, \mathbf{a}_j, \beta_0, \beta_1) + z\, s\, \ell_2(\mathbf{a}_i, \mathbf{a}_j)$. The value of $\mp z$ is found from normal probability tables by using $e(\mathbf{a}_i, \mathbf{a}_j)/\ell_2(\mathbf{a}_i, \mathbf{a}_j)$ distribution.

The rest of this paper is organized as follows: in Section 2, we investigate the statistical properties of the estimation errors generated by the ℓ_{bp}-norm and the SD criterion in the seventeen geographical regions using the data given by Love and Walker (1993). Then, utilizing these properties we develop an error related random variable which is both homoscedastic and normally distributed at the 5% significance level. Since excluding outliers from a region's sample would weaken the sample's representation of the whole region we also aim to obtain this new random variable without excluding the outliers. In Section 3, we develop confidence intervals for unknown actual distances by using the new random variable. In Section 4 we give some example confidence interval calculations, and compare the the confidence interval calculation method developed in Section 3 with the method provided by Love, Walker and Tiku (1995). In Section 5, we provide a comparison of the distance predicting accuracy of the weighted ℓ_p-norm and ℓ_{bp}-norm based on their confidence intervals.

2 Statistical Properties of Estimation Errors

We define the relationship between the fitted distance and the actual distance as

$$A(\mathbf{x}_i, \mathbf{x}_j) = \ell_{bp}(\mathbf{x}_i, \mathbf{x}_j) + \varepsilon(\mathbf{x}_i, \mathbf{x}_j), \tag{4}$$

where $A(\mathbf{x}_i, \mathbf{x}_j)$ is the actual distance between any two points $\mathbf{x}_i$ and $\mathbf{x}_j$, $\ell_{bp}(\mathbf{x}_i, \mathbf{x}_j)$ is the predicted distance, and $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$ is the related prediction error term. From a random sample of points taken from a geographical region, the empirical distance predicting function parameters are calculated. A computational procedure for finding the parameters of the ℓ_{bp}-norm is given by Üster and Love (1998). Substituting these parameters and the point coordinates into the empirical distance predicting function, an estimate of the actual distance, $\ell_{bp}(\mathbf{x}_i, \mathbf{x}_j)$, is obtained. The error $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$ for any pair of points may arise from point coordinate measurements errors, inaccurate instrument calibrations, and road network peculiarities that are not captured by the distance model. Since it is a purely random part of the actual distance $A(x_i, x_j)$, $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$ is a continuous random variable. We assume that the errors for different pairs of points in a region are independent, i.e., the error of $\ell_{bp}(\mathbf{x}_i, \mathbf{x}_j)$ about $A(\mathbf{x}_i, \mathbf{x}_j)$ is not related to the error of $\ell_{bp}(\mathbf{x}_k, \mathbf{x}_l)$ about $A(\mathbf{x}_k, \mathbf{x}_l)$ for any four points $\mathbf{x}_i$, $\mathbf{x}_j$, $\mathbf{x}_k$, $\mathbf{x}_l$ in a geographical region.

In order to empirically examine the statistical properties of the $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$ distributions for seventeen regions we use the same sample set that we used in fitting the parameters of the ℓ_{bp}-norm. To calculate the predicted distances we use the parameter values given Table 1. For our statistical tests we use the statistical analysis package SPSS® (1997).

To test the normality of the $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$ distributions we apply the Kolmogorov-Smirnov test with the Lilliefors correction. The details of this test are given by Lilliefors (1967) and Dallal and Wilkinson (1986). The Lilliefors test is a modification of the Kolmogorov-Smirnov test that examines for normality when means and variances are not known, but must be estimated from the data, and it is based on the largest absolute difference between the observed and the expected cumulative distributions. The p-values of the normality tests for the seventeen regions are reported in Table 2 (Note that the parameter p of a distance norm and the p-value of a statistical test are different. We use italics "p" to refer to a distance norm's parameter, and "p-value" for a statistical test's significance level). The

No.	Region	p-value	No.	Urban Center	p-value
1	Australia	0.000	9	Brussels	0.077
2	BC Province	0.012	10	London City	0.000
3	Canada	0.064	11	London North	0.200
4	France	0.051	12	Los Angeles	0.200
5	Great Britain	0.000	13	NY City	0.000
6	NY State	0.000	14	Paris	0.009
7	Pennsylvania	0.108	15	Sydney	0.200
8	United States	0.038	16	Tokyo	0.036
			17	Toronto	0.025

Table 2: Normality Tests for $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$

p-values indicate that for most of the cases (10 regions) the $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$ distributions are not normally distributed at the 5% significance level. Although the p-values for seven regions are greater than 5%, four of them are not quite convincing. There is significant evidence that $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$ is normally distributed only in regions 11, 12 and 15.

Besides the Kolmogorov-Smirnov test , normal probability plots and histograms are examined. The related graphs are given in Appendices A and B. On normal probability plots, a linear relation is expected between the observed cumulative probabilities and the expected cumulative probabilities for a sample distribution to be from a normally distributed population. The histograms are expected to have a symmetric bell-shaped appearance with no violations at the tails. The normal probability plots and histograms also confirm that in general $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$ is not normally distributed.

To examine the homoscedasticity (Wesolowsky, 1976) of $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$, the sample sets of 105 pairs are divided into three groups (short, medium and long) after they are ordered in their increasing order of predicted distances. In order to clarify what is meant by these three groups, Table 3 is constructed. In Table 3 the means of short, medium and long predicted

No.	Region	Short(S)	Medium(M)	Long(L)	M/S	L/S
1	Australia	1115.74	2116.67	3500.00	1.90	3.14
2	BC Province	298.36	610.78	993.80	2.05	3.33
3	Canada	761.23	2215.23	4262.76	2.91	5.60
4	France	258.16	494.65	724.61	1.92	2.81
5	Great Britain	191.12	389.75	698.02	2.04	3.65
6	NY State	122.83	252.10	413.81	2.05	3.37
7	Pennsylvania	105.01	205.29	367.82	1.95	3.50
8	United States	1080.69	2135.41	3581.48	1.98	3.31
9	Brussels	5.79	10.68	16.37	1.85	2.83
10	London City	4.15	7.75	12.61	1.87	3.04
11	London North	3.55	6.58	11.10	1.85	3.12
12	Los Angeles	9.24	16.82	25.99	1.82	2.81
13	NY State	10.17	18.14	27.74	1.78	2.73
14	Paris	3.26	6.41	9.76	1.97	3.00
15	Sydney	1.22	2.19	3.36	1.80	2.75
16	Tokyo	4.23	7.22	11.22	1.71	2.65
17	Toronto	8.82	15.57	25.90	1.76	2.93

Table 3: The Means of the Predicted Distance Groups

distance distributions, and also the ratio of medium to short and long to short predicted distance means are listed. The ratios are similar for all regions except Canada.

To test the homoscedasticity of $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$ we apply Levene test to three groups of distributions. A p-value less than 5% indicates that there is at least one pair of groups with a significantly different variance. The Levene test is a powerful test when the data come from continuous, but not necessarily normal, distributions (Kotz and Johnson, 1989). The p-values for the tests are reported in Table 4. The p-values, based on a 5% significance level, suggest that $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$ are heteroscedastic in ten of the regions. The homoscedasticity is observed only in the $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$ of seven urban centers. In addition to the Levene test we

No.	Region	p-value	No.	Urban Center	p-value
1	Australia	0.000	9	Brussels	0.030
2	BC Province	0.006	10	London City	0.097
3	Canada	0.000	11	London North	0.021
4	France	0.004	12	Los Angeles	0.746
5	Great Britain	0.006	13	NY City	0.514
6	NY State	0.000	14	Paris	0.059
7	Pennsylvania	0.008	15	Sydney	0.123
8	United States	0.008	16	Tokyo	0.464
			17	Toronto	0.059

Table 4: The Levene Tests for $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$

also inspect the scatter plots of errors for our samples. These scatter plots of $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$ versus $\ell_{bp}(\mathbf{x}_i, \mathbf{x}_j)$ are given in Appendix C. We observe that, as expected, as the predicted distance gets larger the amount of prediction error becomes larger. In other words, the scatter plots of $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$ versus $\ell_{bp}(\mathbf{x}_i, \mathbf{x}_j)$ form a diverging funnel.

To test whether the mean of $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$, μ_ε, is zero we employ the Student t-test. Although this test assumes that the data are normally distributed, it is fairly robust to departures from normality. The p-values for the test are given in Table 5. The p-values suggest that $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$ has zero mean at the 5% significance level in all seventeen regions. This fact is also observed in the scatter plots where points are evenly scattered about zero.

In general, we can say that $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$ is non-normal, heteroscedastic, and its mean is equal to zero at the 5% significance level. In order to develop a confidence interval for an unknown actual distance we need a random variable which is *homoscedastic*. Therefore we need to use a transformation of $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$. For that purpose, we define the following

No.	Region	p-value	No.	Urban Center	p-value
1	Australia	0.498	9	Brussels	0.585
2	BC Province	0.237	10	London City	0.146
3	Canada	0.626	11	London North	0.642
4	France	0.695	12	Los Angeles	0.262
5	Great Britain	0.544	13	NY City	0.418
6	NY State	0.501	14	Paris	0.315
7	Pennsylvania	0.512	15	Sydney	0.485
8	United States	0.706	16	Tokyo	0.564
			17	Toronto	0.570

Table 5: The Student t-test for $\mu_\varepsilon = 0$

transformed random variable:

$$\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j) = \frac{\varepsilon(\mathbf{x}_i, \mathbf{x}_j)}{(\ell_{bp}(\mathbf{x}_i, \mathbf{x}_j))^{1/t}}, \quad t \geq 1, \tag{5}$$

where parameter t is to be determined for a given transportation network so that the corresponding $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ is homoscedastic and normally distributed.

Before proceeding with our analysis for finding the best t value for a transportation network, we will first assume that $t = 1$ and investigate if the normalized error $\varepsilon_1(\mathbf{x}_i, \mathbf{x}_j)$ is homoscedastic. This special form of $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ is very similar to the random variable used by Love, Walker and Tiku (1995). In order to test homoscedasticity we apply the Levene test in the same way that we used it for $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$. In Table 6 we report the p-values. The scatter plots of $\varepsilon_1(\mathbf{x}_i, \mathbf{x}_j)$ with respect to increasing values of the ℓ_{bp} distance are given in Appendix D. The p-values show that, except in Australia, BC Province, Great Britain, New York State and Paris, the $\varepsilon_1(\mathbf{x}_i, \mathbf{x}_j)$ is heteroscedastic at the 5% significance level. Furthermore, for BC Province and Paris the p-values are very close to 5%. Inspecting the scatter plots in Appendix D we see that the scatter of points form a converging funnel, i.e., in general the percentage error is smaller for relatively long distances and larger for relatively short distances.

We now turn our attention to the determination of the best t for a given network. For that purpose we generate samples of $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ for $t = 1, \ldots, 3$ with 0.1 increments. In Table 7 we list the ranges of t in which the $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ are homoscedastic and normally

No.	Region	p-value	No.	Urban Center	p-value
1	Australia	0.5218	9	Brussels	0.0287
2	BC Province	0.0723	10	London City	0.0006
3	Canada	0.0000	11	London North	0.0004
4	France	0.0099	12	Los Angeles	0.0006
5	Great Britain	0.1659	13	NY City	0.0001
6	NY State	0.4847	14	Paris	0.0574
7	Pennsylvania	0.0153	15	Sydney	0.0155
8	United States	0.0026	16	Tokyo	0.0042
			17	Toronto	0.0200

Table 6: The Levene Tests for $\varepsilon_1(\mathbf{x}_i, \mathbf{x}_j)$

distributed. A "✓" represents statistically significant test results at the 5% significance

No.	Region	t Range	t	Homoscedasticity	Normality
1	Australia	1.0–2.3	1.0	✓	X
2	BC Province	1.4–2.6	2.0	✓	✓
3	Canada	1.7–3.0	2.0	✓	✓
4	France	1.5–2.1	2.0	✓	✓
5	Great Britain	1.0–3.0	1.0	✓	X
6	NY State	1.0	1.0	✓	✓
7	Pennsylvania	1.2–3.0	2.0	✓	✓
8	United States	1.4–3.0	2.0	✓	✓
9	Brussels	1.5–3.0	2.0	✓	✓
10	London City	1.4–3.0	2.0	✓	X
11	London North	1.4–3.0	2.0	✓	✓
12	Los Angeles	1.8–3.0	2.0	✓	✓
13	NY City	1.7–3.0	2.0	✓	X
14	Paris	1.3–2.8	2.0	✓	✓
15	Sydney	1.2–3.0	2.0	✓	✓
16	Tokyo	1.6–3.0	2.0	✓	X
17	Toronto	1.2–3.0	2.0	✓	X

Table 7: Homoscedasticity and Normality for Ranges of t

level whereas an "X" represents otherwise. We note that in the United States and London City homoscedasticity could not be obtained for $t \in [1, 3]$. Therefore we exclude the actual distances which correspond to the outliers in the samples of $\varepsilon_1(\mathbf{x}_i, \mathbf{x}_j)$ for these regions and model the ℓ_{bp}-norm with the new reduced sample sets. The new samples provided well-behaved distributions for both regions. The number of outliers in the United States case was five and in the London City case it was nine.

In Table 7 we observe that homoscedasticity is obtained for $t = 2$ in all regions, except in the NY City case. However, if $t = 1$, then both homoscedasticity and normality are obtained for NY City. Based on our samples we devise the following rule of thumb that can be used in obtaining a transformed random variable for confidence interval calculation purposes: $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ with $t = 1$ is tested for homoscedasticity and normality. Table 7 indicates that both conditions are not very likely to occur in a particular region. If not, then $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ with $t = 2$ is tested for homoscedasticity and normality. Based on our samples it is very likely that a well-behaved distribution will be obtained. In the case that neither $t = 1$ nor $t = 2$ provides a homoscedastic and normal distribution then other values of t should be considered. However, if all fails, then the outliers should be excluded and the same checks should be performed with $t = 1$ and $t = 2$. The t values determined by using this rule of thumb for seventeen geographical regions are also reported in Table 7. We note that, for purposes of confidence interval calculations, homoscedasticity is more important than normality. In order to obtain a reliable confidence interval the distribution must have a constant variance. Otherwise, one can easily obtain a confidence interval with a lower or higher level of expectation than the predetermined value, say 95%. Therefore, if only a homoscedastic but non-normal $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ is obtained before excluding the outliers, this distribution can be efficiently used to calculate the confidence interval as shown in the next section. In other words, in order not to weaken the sample's representation of the whole region, excluding outliers should be seen as a last resort in the course of obtaining a well-behaved error distribution.

3 Development of Confidence Intervals

In this section we develop the confidence interval for an estimated distance. For that purpose we use the transformed random variable $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ where $t \geq 1$.

Let $\mu_{\varepsilon t}$ and $\sigma^2_{\varepsilon t}$ represent the mean and the variance of $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$, herein denoted ε_t. Also let (e_{t1}, e_{t2}) be the $100(1-\alpha)\%$ confidence interval for $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ where $0 < \alpha < 1$. We have

$$\Pr(e_{t1} < \varepsilon_t < e_{t2}) = 1 - \alpha, \quad 0 < \alpha < 1. \tag{6}$$

In our development we will assume that the confidence interval is symmetric, i.e., $\Pr(\varepsilon_t < e_{t1}) = \Pr(\varepsilon_t > e_{t2}) = \alpha/2$. The standardized values of e_{t1} and e_{t2} are given by z_{t1} and z_{t2} where

$$z_{t1} = \frac{e_{t1} - \mu_{\varepsilon_t}}{\sigma_{\varepsilon_t}} \quad \text{and} \quad z_{t2} = \frac{e_{t2} - \mu_{\varepsilon_t}}{\sigma_{\varepsilon_t}}. \tag{7}$$

Rearranging (7) and substituting into (6) we have

$$\Pr(\mu_{\varepsilon_t} + z_{t1}\,\sigma_{\varepsilon_t} < \varepsilon_t < \mu_{\varepsilon_t} - z_{t2}\,\sigma_{\varepsilon_t}) = 1 - \alpha, \tag{8}$$

or equivalently

$$\Pr(z_{t1}\,\sigma_{\varepsilon_t} < \varepsilon_t - \mu_{\varepsilon_t} < z_{t2}\,\sigma_{\varepsilon_t}) = 1 - \alpha. \tag{9}$$

We assume that $\mu_{\varepsilon_t} = 0$ and s_t is the sample unbiased estimator of σ_{ε_t}. Then (9) becomes

$$\begin{aligned} &\Pr(z_{t1}\, s_t < \varepsilon_t < z_{t2}\, s_t) = 1 - \alpha \\ &\Pr\left(z_{t1}\, s_t < \frac{A(\mathbf{x}_i, \mathbf{x}_j) - \ell_{bp}(\mathbf{x}_i, \mathbf{x}_j)}{(\ell_{bp}(\mathbf{x}_i, \mathbf{x}_j))^{1/t}} < z_{t2}\, s_t\right) = 1 - \alpha. \end{aligned} \tag{10}$$

By rearranging (10) we find the confidence interval for an unknown actual distance between any two points $\mathbf{x}_i$ and $\mathbf{x}_j$ as

$$\Pr(c_{t1} < A(\mathbf{x}_i, \mathbf{x}_j) < c_{t2}) = 1 - \alpha, \quad 0 < \alpha < 1. \tag{11}$$

where

$$\begin{aligned} c_{t1} &= \ell_{bp}(\mathbf{x}_i, \mathbf{x}_j)\ \left[1 + z_{t1}\, s_t\, (\ell_{bp}(\mathbf{x}_i, \mathbf{x}_j))^{(1/t)-1}\right], \\ c_{t2} &= \ell_{bp}(\mathbf{x}_i, \mathbf{x}_j)\ \left[1 + z_{t2}\, s_t\, (\ell_{bp}(\mathbf{x}_i, \mathbf{x}_j))^{(1/t)-1}\right]. \end{aligned} \tag{12}$$

If $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ is normally distributed, then the standardized values z_{t1} and z_{t2} are found in the standard normal distribution table. Since we assume symmetry, the values of z_{t1} and z_{t2} are of equal magnitude but opposite sign. For example, if a 95% confidence interval is considered, then the standardized values are $z_t = \mp 1.96$. However, if the distribution is not known, then the standardized values z_{t1} and z_{t2} are found in the tables provided by Johnson, Nixon and Amos (1963). These tables utilize the Pearson coefficients (skewness and kurtosis) of the distribution and list the standardized values for several values of $\alpha/2$. For large samples ($n \geq 100$) the sample Pearson coefficients can be used as unbiased estimates of the population Pearson coefficients (Stuart and Ord, 1987).

4 Some Examples

In this section, we demonstrate the use of (12) by constructing the confidence intervals for six unknown actual distances. We use the regions and the point locations used in constructing example confidence intervals with the weighted ℓ_p-norm by Love, Walker and Tiku (1995). The regions, point locations (in centimeters), standard deviation of sample $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ distributions, and the standardized values z_{t1} and z_{t2} are given in Table 8. For four of the regions the $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ are normally distributed (Table 7). Therefore, to

Region	t	$\mathbf{x}_i$	$\mathbf{x}_j$	s_t	z_{t1}	z_{t2}
Australia	1	(15.50,18.65)	(43.00,33,90)	0.0786	-1.8258	2.1689
Canada	2	(106.65,24.50)	(19.10,13.10)	2.1869	-1.9600	1.9600
United States	2	(4.50,28.20)	(41.00,16.20)	1.8343	-1.9600	1.9600
Brussels	2	(19.30,25.55)	(38.30,7.35)	0.1862	-1.9600	1.9600
London North	2	(10.30,2.60)	(38.05,24.60)	0.1145	-1.9600	1.9600
Toronto	2	(54.45,41.85)	(49.05,17.30)	0.2182	-2.1149	1.8739

Table 8: Data for Example Confidence Interval Calculations

construct 95% confidence intervals the values of z_{t1} and z_{t2} are taken as -1.96 and 1.96, respectively. However, for Australia and Toronto, $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ is not normally distributed. Thus, to determine the standardized values we have to use the tables provided by Johnson, Nixon and Amos (1963). For Australia the sample Pearson coefficients are $\sqrt{b_1} = 0.4800$ and $b_2 = 4.7524$, and for Toronto they are $\sqrt{b_1} = -0.4086$ and $b_2 = 5.6495$. Using the interpolation technique as described by Johnson, Nixon and Amos we calculate the standardized values given in Table 8. In Table 9 we report the ℓ_{bp} distances (in kilometers) and the corresponding 95% confidence intervals.,

Region	$\ell_{bp}(\mathbf{x}_i, \mathbf{x}_j)$	(c_{t1}, c_{t2})
Australia	3028.47	(2593.86,3544.75)
Canada	3130.36	(2890.54,3370.17)
United States	3174.07	(2971.56,3376.57)
Brussels	8.80	(7.72,9.88)
London North	6.64	(6.06,7.22)
Toronto	10.92	(9.40,12.27)

Table 9: Example Confidence Intervals

To compare the confidence interval calculation method developed in the previous section with the Love-Walker-Tiku method, we first calculate the confidence intervals (c_1, c_2) for the ℓ_{bp} distances by using the method and the parameters given by Love, Walker and Tiku (1995). Then we compare the range of these confidence intervals with the range of confidence interval calculated by our method. These comparisons are presented in Table 10. We observe that, except in Australia and Brussels, the confidence intervals obtained by the

Region	$\ell_{bp}(\mathbf{x}_i, \mathbf{x}_j)$	(c_1, c_2)	$c_2 - c_1$	$c_{t2} - c_{t1}$
Australia	3028.47	(2565.39,3499.99)	934.60	950.89
Canada	3130.36	(2597.32,3484.85)	887.52	479.63
United States	3174.07	(2942.72,3474.45)	531.73	405.01
Brussels	8.80	(7.85,9.95)	2.10	2.16
London North	6.64	(5.92,7.38)	1.46	1.16
Toronto	10.92	(9.25,12.17)	2.92	2.88

Table 10: Comparison of Example Confidence Intervals

method developed here are smaller than the ones obtained by the Love-Walker-Tiku method. Therefore we next would like to explore this difference.

Let $\mathcal{L}$ be a predicted distance, $\mathcal{I}$ and $\mathcal{I}_t$ be the width of the confidence intervals calculated by the Love-Walker-Tiku method and our method, respectively. Then the expressions for $\mathcal{I}$ and $\mathcal{I}_t$ are

$$\mathcal{I} = \mathcal{L}\,(1 + z_2 s) - \mathcal{L}\,(1 + z_1 s) \tag{13}$$

$$\mathcal{I}_t = \mathcal{L}\,(1 + z_{t2} s_t \mathcal{L}^{(1/t)-1}) - \mathcal{L}\,(1 + z_{t1} s_t \mathcal{L}^{(1/t)-1}) \tag{14}$$

It can easily be verified that $\mathcal{I}$ and $\mathcal{I}_t$ are increasing functions of $\mathcal{L}$ and, for $t > 1$

$$\begin{aligned} \mathcal{I} < \mathcal{I}_t &\quad \text{for} \quad \mathcal{L} < \mathcal{L}_c \\ \mathcal{I} = \mathcal{I}_t &\quad \text{for} \quad \mathcal{L} = \mathcal{L}_c \\ \mathcal{I} > \mathcal{I}_t &\quad \text{for} \quad \mathcal{L} > \mathcal{L}_c \end{aligned} \tag{15}$$

where

$$\mathcal{L}_c = \left(\frac{s_t(z_{t2} - z_{t1})}{s(z_2 - z_1)} \right)^{t/(t-1)}, \tag{16}$$

for $t = 1$

$$\begin{aligned} \mathcal{I} < \mathcal{I}_t &\quad \text{for} \quad \mathcal{R} > 1 \\ \mathcal{I} = \mathcal{I}_t &\quad \text{for} \quad \mathcal{R} = 1 \\ \mathcal{I} > \mathcal{I}_t &\quad \text{for} \quad \mathcal{R} < 1 \end{aligned} \tag{17}$$

where

$$\mathcal{R} = \frac{s_t(z_{t2} - z_{t1})}{s(z_2 - z_1)}. \tag{18}$$

(15) implies that, for $t > 1$, $\mathcal{I}$ and $\mathcal{I}_t$ are equal for a critical $\mathcal{L}$ value given by (16). This fact is observed in the $\mathcal{I}$ and $\mathcal{I}_t$ graphs which are presented in Figure 1 for five example regions (except Australia for which $t = 1$). For values of $\mathcal{L}$ greater than $\mathcal{L}_c$ the range of the confidence interval obtained by the Love-Walker-Tiku method is larger than the one calculated by our method and for values of $\mathcal{L}$ less than $\mathcal{L}_c$ the converse holds. Also note that, the difference between the intervals becomes larger, favouring $\mathcal{I}_t$, as the predicted distance increases. For example in Toronto, for the predicted distance of 10.92 kms. the difference between the two intervals is 0.04 (2.92 vs. 2.88 in Table 10). If the predicted distance was 30 kms., the difference would be 3.31 (8.01 vs. 4.70). On the other hand, (17) implies that, for $t = 1$, $\mathcal{I}$ and $\mathcal{I}_t$ are equal if $\mathcal{R}$ is equal to 1. If $\mathcal{R}$ is less than 1, then the confidence interval calculated by the Love-Walker-Tiku method is larger than the one calculated by our method for any distance $\mathcal{L}$. The converse holds for values of $\mathcal{R}$ greater than 1.

The values of t, ℓ_{bp} distances, $\mathcal{L}_c$ and $\mathcal{R}$ that we use for comparisons are given in Table 11. Observe that for regions with $t = 2$ the predicted distances are longer than their

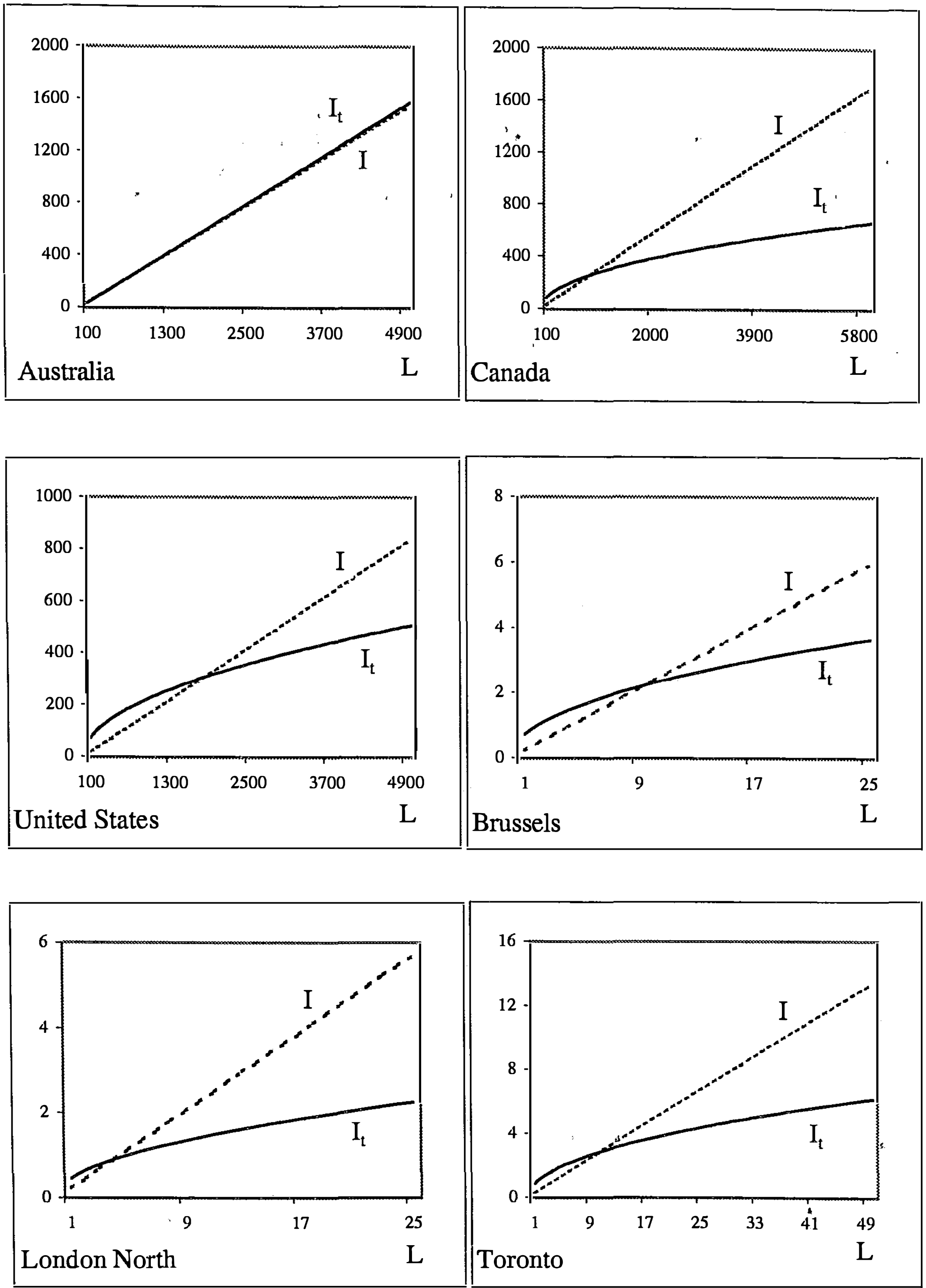


Figure 1: Example Confidence Interval Comparisons

Region	t	$\ell_{bp}(\mathbf{x}_i, \mathbf{x}_j)$	$\mathcal{L}_c$ or $\mathcal{R}$
Australia	1	3028.47	1.017
Canada	2	3130.36	913.54
United States	2	3174.07	1842.33
Brussels	2	8.80	9.44
London North	2	6.64	4.16
Toronto	2	10.92	10.61

Table 11: Example $\mathcal{L}_c$ and $\mathcal{R}$ Values

corresponding critical values $\mathcal{L}_c$, except in Brussels in which case the predicted distance is shorter than $\mathcal{L}_c$. Therefore, as it is already given in Table 10, we obtain smaller confidence intervals for four of the regions, and a larger interval for Brussels. For Australia, where $t = 1$, the $\mathcal{R}$ value is slightly greater than 1, i.e., $\mathcal{I}$ and $\mathcal{I}_t$ values are very close, $\mathcal{I}$ being smaller. This is because for Australia t is equal to 1 and thus $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ is indeed quite similar to the transformed random variable used in the Love-Walker-Tiku method. Furthermore, For Australia the weighted ℓ_{bp}-norm performs similar to the weighted ℓ_p-norm because of the low directional nonlinearity and high rectangularity in the transportation network as discussed by Üster and Love (1998).

In Table 12 the ranges of predicted distances that represent short, medium and long distance groups for example regions are listed. Notice that, except for Australia, $\mathcal{L}_c$ values

Region	Short	Medium	Long
Australia	438–1532	1539–2689	2699–4609
Canada	204–1265	1266–3225	3288–5844
United States	165–1661	1674–2717	2720–4841
Brussels	2.7– 8.4	8.6 –13.2	13.7–22.9
London North	1.5– 5.0	5.1 – 8.1	8.1 –16.7
Toronto	3.9–11.6	11.8–20.2	20.2–43.8

Table 12: Ranges of the Predicted Distance Groups

are close to the cut-off point between short and medium distance groups. This observation suggests that the confidence intervals calculated by the Love-Walker-Tiku method are generally different than what they should be at the 95% level. They are smaller for relatively short distances and larger for relatively long distances. This is indeed an expected result because, when the ℓ_p distances are modelled for these five regions the scatter of points in the $e(\mathbf{a}_i, \mathbf{a}_j)/A(\mathbf{a}_i, \mathbf{a}_j)$ versus $A(\mathbf{a}_i, \mathbf{a}_j)$ plots form a converging funnel (Love and Üster, 1996). In other words, the random variable used in developing confidence intervals is not homoscedastic, but has a higher standard deviation for relatively short distances and a lower standard deviation for relatively long distances. The percentage error generated in predicting long distances is generally lower than the percentage error generated in predicting short distances. We have observed the same phenomenon in predicting distances with the ℓ_{bp}-norm, i.e., $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ with $t = 1$ is generally heteroscedastic. However, the confidence interval calculation method developed in this paper accounts for this phenomenon and provides confidence intervals at the intended level of expectation.

5 Comparison of Distance Functions

In this section we compare the 95% confidence intervals for the weighted ℓ_p and the ℓ_{bp} distances modelled using SD as the goodness-of-fit criterion. This, in turn, provides a comparison of accuracy in distance prediction with the weighted ℓ_p-norm and the ℓ_{bp}-norm. We use the transformed random variable $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ for that purpose, and adopt the notation t' and $\varepsilon_{t'}(\mathbf{x}_i, \mathbf{x}_j)$ for the weighted ℓ_p-norm related distributions. To generate the sample $\varepsilon_{t'}(\mathbf{x}_i, \mathbf{x}_j)$ distributions for each region we use the best parameter values of the weighted ℓ_p-norm, k, p and θ, where $p \in (1, 2)$. These empirical parameter values for seventeen geographical regions are computed by Love and Walker (1994) for the SD criterion and are presented here in Table 13.

We first find the best t' values which provide homoscedasticity of $\varepsilon_{t'}(\mathbf{x}_i, \mathbf{x}_j)$. As in the ℓ_{bp}-norm case we test the homoscedasticity and the normality of sample $\varepsilon_{t'}(\mathbf{x}_i, \mathbf{x}_j)$ distributions for values of t' ranging from 1.0 to 3.0 with increments 0.1. The summary of these test results, including the t' values that are chosen by the rule of thumb devised in Section 2, is given in Table 14. Note that, for $t' \in [1, 3]$, homoscedasticity and normality could not be obtained for the $\varepsilon_{t'}(\mathbf{x}_i, \mathbf{x}_j)$ distributions of the United States and London City. This was also the case for the sample $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ distributions of the same regions. Therefore,

No.	Region	θ	k	p
1	Australia	0	1.1460	1.8585
2	BC Province	68	1.2495	1.5609
3	Canada	83	1.1715	1.4849
4	France	70	1.0609	1.8430
5	Great Britain	40	1.1095	1.7895
6	NY State	86	1.0794	1.5823
7	Pennsylvania	4	1.0611	1.6244
8	United States	0	1.0792	1.6641
9	Brussels	47	1.0549	1.8180
10	London City	72	1.1182	1.9241
11	London North	11	1.0599	1.6456
12	Los Angeles	2	1.0721	1.5734
13	NY City	6	1.1069	1.7340
14	Paris	86	1.0613	1.8189
15	Sydney	8	1.1266	1.4719
16	Tokyo	58	1.0963	1.8252
17	Toronto	88	1.0279	1.1863

Table 13: Best Parameter Values for the Weighted ℓ_p-norm

for comparison purposes, we will not consider these two regions. For the rest of the regions, inspecting t and t' columns in Tables 7 and 14 we see that the chosen values of t and t' are identical, except in regions Canada, Great Britain and Paris. In order to obtain a unified comparison scheme we would like to have equal t and t' values, if possible. Thus, without violating the homoscedasticity requirement of the distributions, we can take $t = 2.1$ for Canada, $t = 2$ for Great Britain, and $t' = 2$ for Paris. The final t and t' values, sample standard deviations, and standardized values for the $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ and $\varepsilon_{t'}(\mathbf{x}_i, \mathbf{x}_j)$ distributions are reported in Table 15. The columns s'_t, z'_{t1} and z'_{t2} represent the sample standard deviation and standardized values related to the sample $\varepsilon_{t'}(\mathbf{x}_i, \mathbf{x}_j)$ distributions. For the error distributions in which normality can not be obtained we calculate the standardized values by interpolation using the sample Pearson coefficients of $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ and $\varepsilon_{t'}(\mathbf{x}_i, \mathbf{x}_j)$, and the biometrika tables of Johnson, Nixon and Amos (1963). We note that the skewness of the sample $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ and $\varepsilon_{t'}(\mathbf{x}_i, \mathbf{x}_j)$ distributions for Tokyo were 2.3707 and 2.4243, respectively. These values are out of the range of skewness values in the biometrika tables where the range is $(0.0, 2.0)$. Thus,

No.	Region	t' Range	t'	Homoscedasticity	Normality
1	Australia	1.0–1.9	1.0	✓	X
2	BC Province	1.2–3.0	2.0	✓	✓
3	Canada	2.1–3.0	2.1	✓	✓
4	France	1.4–2.1	2.0	✓	X
5	Great Britain	1.1–3.0	2.0	✓	X
6	NY State	1.0	1.0	✓	✓
7	Pennsylvania	1.1–3.0	2.0	✓	✓
8	United States	1.0–3.0	-	X	X
9	Brussels	1.3–3.0	2.0	✓	✓
10	London City	1.0–3.0	-	X	X
11	London North	1.3–3.0	2.0	✓	✓
12	Los Angeles	1.7–3.0	2.0	✓	✓
13	NY City	1.6–3.0	2.0	✓	X
14	Paris	1.0–3.0	1.0	✓	✓
15	Sydney	1.7–3.0	2.0	✓	✓
16	Tokyo	1.6–3.0	2.0	✓	X
17	Toronto	1.2–3.0	2.0	✓	X

Table 14: Homoscedasticity and Normality for Ranges of t'

the values of t and t' are taken as 3.3 so that the standardized values can be calculated and both distributions are still homoscedastic.

To compare the two distance functions we will again utilize the range of the confidence intervals. Let $\mathcal{L}$ be any predicted distance as before, and $\mathcal{I}_t$, $\mathcal{I}'_t$ be the ranges of the confidence intervals calculated for an ℓ_{bp} and a weighted ℓ_p distance, respectively. Then using (12) $\mathcal{I}_t$ and $\mathcal{I}'_t$ for the same predicted distance are given by

$$\mathcal{I}_t = \mathcal{L}\,(1 + z_{t2} s_t \mathcal{L}^{(1/t)-1}) - \mathcal{L}\,(1 + z_{t1} s_t \mathcal{L}^{(1/t)-1}) \tag{19}$$

$$\mathcal{I}'_t = \mathcal{L}\,(1 + z'_{t2} s'_t \mathcal{L}^{(1/t')-1}) - \mathcal{L}\,(1 + z'_{t1} s'_t \mathcal{L}^{(1/t')-1}). \tag{20}$$

It readily follows that the intervals $\mathcal{I}_t$ and $\mathcal{I}'_t$ are increasing functions of $\mathcal{L}$ and for $t = t'$ we have

$$\begin{aligned} &\mathcal{I}_t < \mathcal{I}'_t \ \text{ if } \ \mathcal{T} < 1 \\ &\mathcal{I}_t = \mathcal{I}'_t \ \text{ if } \ \mathcal{T} = 1 \\ &\mathcal{I}_t > \mathcal{I}'_t \ \text{ if } \ \mathcal{T} > 1 \end{aligned} \tag{21}$$

No	Region	t', t	s'_t	z'_{t1}	z'_{t2}	s_t	z_{t1}	z_{t2}
1	Australia	1.0	0.0796	-1.7625	2.2002	0.0786	-1.8258	2.1689
2	BC Province	2.0	3.2817	-1.9600	1.9600	3.2441	-1.9600	1.9600
3	Canada	2.1	2.7559	-1.9600	1.9600	2.6074	-1.9600	1.9600
4	France	2.0	0.9215	-1.4853	2.3741	0.8884	-1.9600	1.9600
5	Great Britain	2.0	1.5024	-1.4092	2.4208	1.3483	-1.2548	2.5245
6	NY State	1.0	0.0756	-1.9600	1.9600	0.0756	-1.9600	1.9600
7	Pennsylvania	2.0	1.0135	-1.9600	1.9600	0.9569	-1.9600	1.9600
9	Brussels	2.0	0.1883	-1.9600	1.9600	0.1862	-1.9600	1.9600
11	London North	2.0	0.1315	-1.9600	1.9600	0.1146	-1.9600	1.9600
12	Los Angeles	2.0	0.3676	-1.9600	1.9600	0.3356	-1.9600	1.9600
13	NY City	2.0	0.3818	-1.2668	2.4895	0.3806	-1.1612	2.5634
14	Paris	2.0	0.2433	-1.9600	1.9600	0.2319	-1.9600	1.9600
15	Sydney	2.0	0.1182	-1.9600	1.9600	0.1050	-1.9600	1.9600
16	Tokyo	3.3	0.2236	-1.2693	2.5147	0.2231	-1.2452	2.5274
17	Toronto	2.0	0.2191	-2.1218	1.8648	0.2182	-2.1149	1.8739

Table 15: Data for Confidence Interval Comparisons

where

$$\mathcal{T} = \frac{s_t(z_{t2} - z_{t1})}{s'_t(z'_{t2} - z'_{t1})}. \tag{22}$$

The values of $\mathcal{T}$ for 15 regions used in the comparisons are calculated and reported in Table 16. Observe that, except in NY State, The $\mathcal{T}$ values are always less than 1, i.e., the confidence intervals calculated for the ℓ_{bp} distances are smaller indicating the superiority of the ℓ_{bp}-norm over the weighted ℓ_p-norm. The proximity of the $\mathcal{T}$ value to 1 determines the

No.	Region	$\mathcal{T}$	No.	Urban Center	$\mathcal{T}$
1	Australia	0.9954	9	Brussels	0.9888
2	BC Province	0.9885	11	London North	0.8715
3	Canada	0.9461	12	Los Angeles	0.9129
4	France	0.9792	13	NY City	0.9884
5	Great Britain	0.8856	14	Paris	0.9531
6	NY State	1.0000	15	Sydney	0.8883
7	Pennsylvania	0.9442	16	Tokyo	0.9948
			17	Toronto	0.9964

Table 16: $\mathcal{T}$ Values

level of gained prediction accuracy by using the ℓ_{bp}-norm versus the weighted ℓ_p-norm. The gain in prediction accuracy increases as the $\mathcal{T}$ value becomes smaller. Üster and Love (1998)

No.	Region	SD for ℓ_{bp}	SD for $k\ell_p$	$\Delta(SD)\%$
1	Australia	1117.89	1163.59	3.93
2	BC Province	1015.41	1038.72	2.24
3	Canada	496.88	565.61	12.15
4	France	78.86	84.72	6.92
5	Great Britain	172.34	219.42	21.46
6	NY State	159.80	159.80	0.00
7	Pennsylvania	95.11	107.06	11.16
8	United States	336.53	342.68	1.79
9	Brussels	3.47	3.55	2.25
10	London Central	15.01	16.53	9.20
11	London North	1.36	1.78	23.60
12	Los Angeles	13.17	15.50	15.03
13	NY City	13.29	13.58	2.14
14	Paris	5.84	6.52	10.43
15	Sydney	1.10	1.35	18.52
16	Tokyo	2.28	2.30	0.87
17	Toronto	5.06	5.10	0.78

Table 17: Comparison of ℓ_{bp}-norm and $k\ell_p$-norm

compared the performance of the ℓ_{bp}-norm and the weighted ℓ_p-norm in predicting distances by using the percent decrease in the SD value of the weighted ℓ_p-norm. These results are summarized here in Table 17. Since the ℓ_{bp}-norm models the directional nonlinearity in a transportation network explicitly, we observe that the gain increases as the directional nonlinearity increases. In Table 17 we see that a low gain in prediction accuracy (below 7%) is obtained for regions 1, 2, 4, 6, 9, 13, 16 and 17, a moderate gain $(10-12\%)$ is obtained for regions 3, 7 and 14, and a high gain (above 15%) is obtained for regions 5, 11, 12 and 15. A close inspection of Table 16 reveals that the same groups of regions are formed based on the proximity of $\mathcal{T}$ values to unity.

Acknowledgements : This research was supported by a grant and a postgraduate scholarship from the Natural Sciences and Engineering Research Council of Canada.

References

BERENS, W. 1988. The Suitability of the Weighted ℓ_p-norm in Estimating Actual Road Distances, *European Journal of Operational Research*, **34**, 39–43

BERENS, W. AND KÖRLING, F. J. 1985. Estimating Road Distances by Mathematical Functions, *European Journal of Operational Research*, **21**, 54–56

BRIMBERG, J., DOWLING, P. D., AND LOVE, R. F. 1994. The Weighted One-Two Norm Distance Model: Empirical Validation and Confidence Interval Estimation, *Location Science*, **2(2)**, 91–100

BRIMBERG, J., DOWLING, P. D., AND LOVE, R. F. 1996. Estimating the Parameters of the Weighted ℓ_p-norm by Linear Regression, *IIE Transactions*, **28**, 363–367

BRIMBERG, J., LOVE, R. F., AND WALKER, J. H. 1995. The Effect of Axis Rotation on Distance Estimation, *European Journal of Operational Research*, **80**, 357–364

DALLAL, G. E. AND WILKINSON, L. 1986. An Analytic Approximation to the Distribution of Lilliefors's Test Statistic for Normality, *The American Statistician*, **40(4)**, 294–296

EILON, S., WATSON-GANDY, C. D. T., AND CHRISTOFIDES, N. 1971. *Distribution Management: Mathematical Modeling and Practical Analysis.* Hafner Publishing Company. New York

FOULDS, L. R. AND HAMACHER, H. W. 1990. Optimal Bin Location and Sequencing in Printed Circuit Board Assembly. Technical Report 181. University of Kaiserslauteren

GINSBURGH, V. AND HANSEN, P. 1974. Procedures for the Reduction of Errors in Road Network Data, *Operational Research Quarterly*, **25**, 321–322

HURIOT, J. M. AND PERREUR, J. 1973. On the Weber Problem with Rectangular Distance: A Comment, *Management Science*, **20**, 418–419

Johnson, N. L., Nixon, E., and Amos, D. E. 1963. Tables of Percentage Points of Pearson Curves for Given $\sqrt{\beta_1}$ and β_2 Expressed in Standard Measure, *Biometrika*, **50**, 459–498

Klein, R. 1988. in *Graph-Theoretic Concepts in Computer Science*. 434–441. Amsterdam

Kolesar, P., Walker, W., and Hausner, J. 1975. Determining the Relation Between Fire Engine Travel Times and Travel Distances in New York City, *Operations Research*, **23**, 614–627

Kotz, S. and Johnson, N. L. (eds.) 1989. *Encyclopedia of Statistical Sciences.* Vol. 4. John-Wiley and Sons. New York

Lilliefors, H. W. 1967. On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown, *American Statistical Association Journal*, 399–402

Love, R. F. and Dowling, P. D. 1985. Optimal Weighted ℓ_p-norm Parameters for Facilities Layout Distance Characteristics, *Management Science*, **31**, 200–206

Love, R. F. and Morris, J. G. 1972. Modelling Inter-City Road Distances by Mathematical Models, *Operational Research Quarterly*, **23**, 61–71

Love, R. F. and Morris, J. G. 1979. Mathematical Models of Road Travel Distances, *Management Science*, **25**, 130–139

Love, R. F., Morris, J. G., and Wesolowsky, G. O. 1988. *Facilities Location – Models and Methods.* North-Holland. NewYork

Love, R. F. and Üster, H. 1996. Comparison of the Properties and the Performance of the Criteria Used to Evaluate the Accuracy of Distance Predicting Functions. Technical Report 419. McMaster University, Faculty of Business. Hamilton, Ontario, Canada

Love, R. F. and Walker, J. H. 1991. A New Criterion for Evaluating the Accuracy of Distance Predicting Functions. Technical Report 370. McMaster University, Faculty of Business. Hamilton, Ontario, Canada

Love, R. F. and Walker, J. H. 1993. Distance Data for Eighteen Geographic Regions. Technical Report 383. McMaster University, Faculty of Business. Hamilton, Ontario, Canada

Love, R. F. and Walker, J. H. 1994. An Empirical Comparison of Block and Round Norms for Modeling Actual Distances, *Location Science*, **2(1)**, 21–43

Love, R. F., Walker, J. H., and Tiku, M. L. 1995. Confidence Intervals for $\ell_{k,p,\theta}$ Distances, *Transportation Science*, **29**, 93–100

MicroAnalytics 1993. TruckStops 2. Suite One, 2045 North 15th Street, Arlington, Virginia and 1986 Queen Street East, Toronto, Ontario

Murray, A. T. and Estivill-Castro, V. 1998. Cluster Discovery Techniques for Exploratory Spatial Data Analysis, *International Journal of Geographical Information Science*, **12(5)**, 431–443

Roadnet-Technologies 1993. ROADNET. 10540 York Rd., Huntvalley, Maryland

SPSS 1997. Release 8.0.0.. 444 N.Michigan Ave., Chicago, IL

Star, J. and Estes, J. 1990. *Geographic Information Systems – An Introduction.* Prentice-Hall. New Jersey

Stuart, A. and Ord, J. K. 1987. *Kendall's Advanced Theory of Statistics, Volume I - Distribution Theory.* Oxford University Press. New York

Üster, H. and Love, R. F. 1998. Application of Weighted Sum of Order p to Distance Estimation. Technical Report 427. McMaster University, Faculty of Business. Hamilton, Ontario, Canada

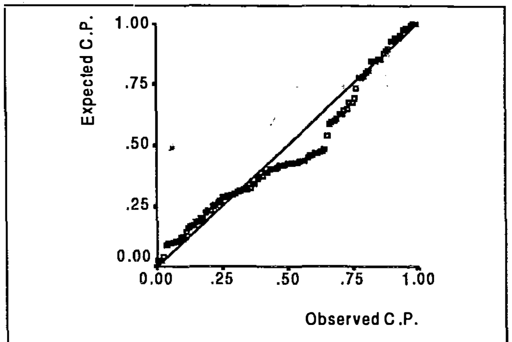
Ward, J. E. and Wendell, R. E. 1980. A New Norm for Measuring Distance Which Yields Linear Location Models, *Operations Research*, **28**, 836–844

Ward, J. E. and Wendell, R. E. 1985. Using Block Norms for Location Modeling, *Operations Research*, **33**, 1074–1090

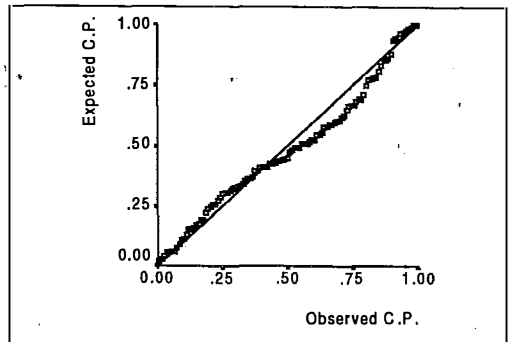
Wesolowsky, G. O. 1976. *Multiple Regression and Analysis of Variance.* John Wiley and Sons. Toronto

Westwood, J. B. 1977. A Transport Planning Model for Primary Distribution, *Interfaces*, **8**, 1–10

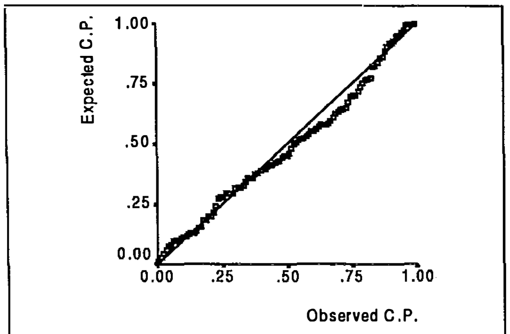
A Normal Probability Plots of $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$



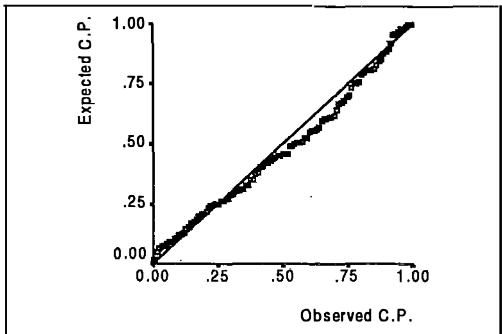
Australia



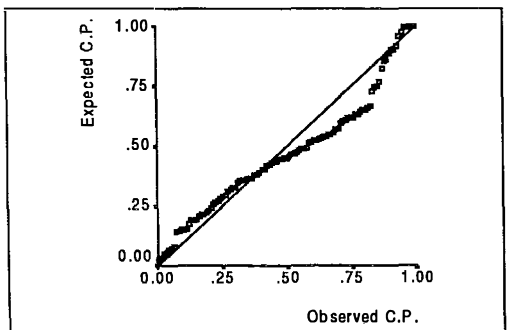
British Columbia



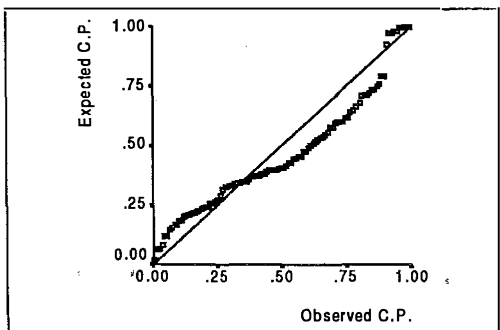
Canada



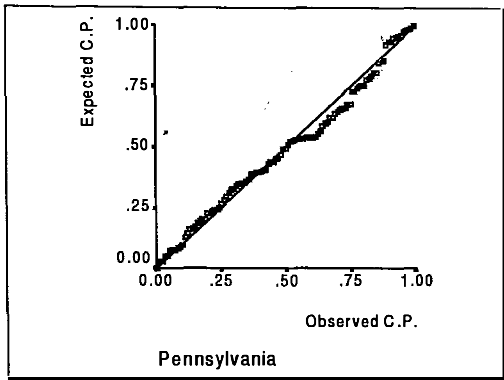
France

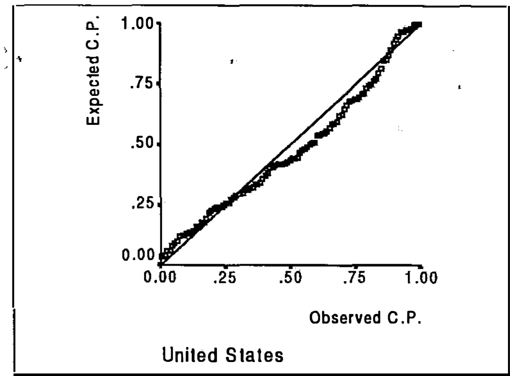


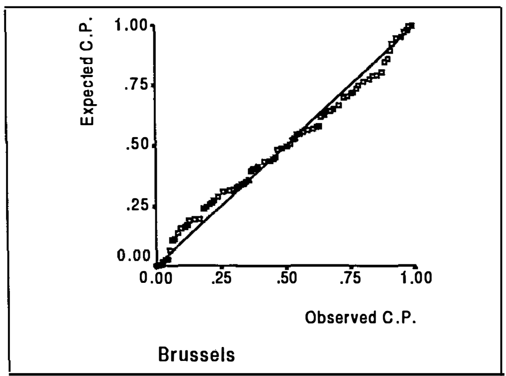
Great Britain

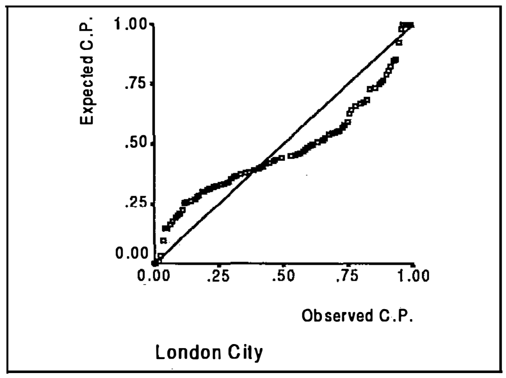


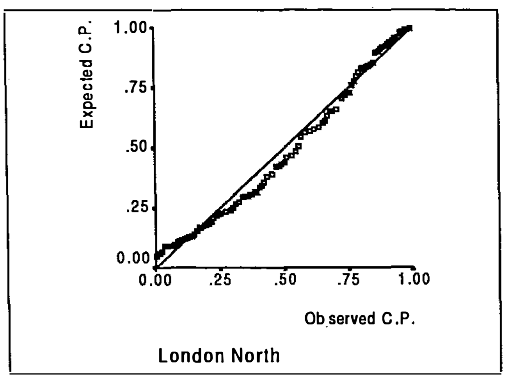
New York State

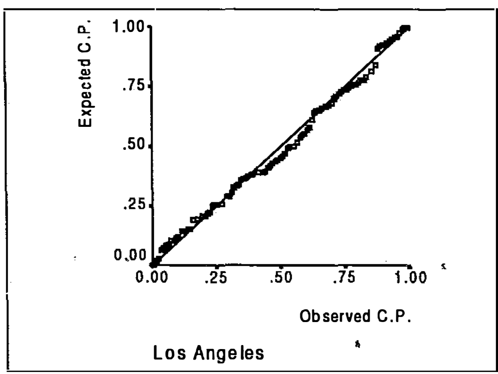

Expected C.P.
1.00
.75
.50
.25
0.00
0.00 .25 .50 .75 1.00
Observed C.P.
Pennsylvania

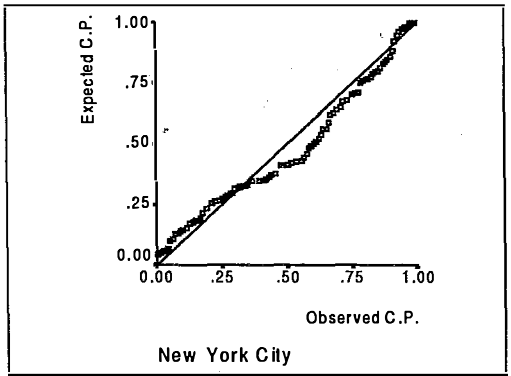

Expected C.P.
1.00
.75
.50
.25
0.00
0.00 .25 .50 .75 1.00
Observed C.P.
United States

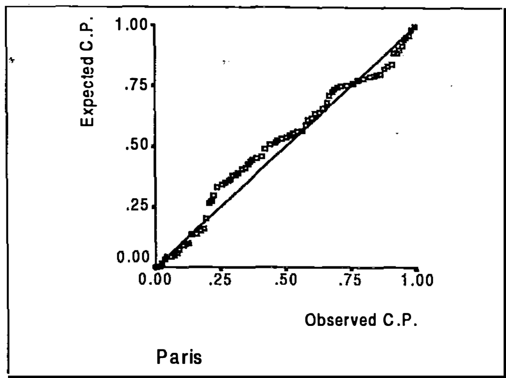

Expected C.P.
1.00
.75
.50
.25
0.00
0.00 .25 .50 .75 1.00
Observed C.P.
Brussels

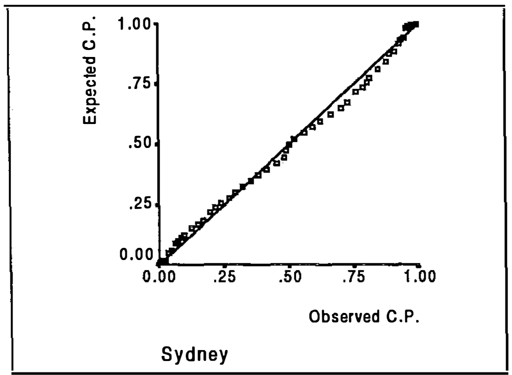

Expected C.P.
1.00
.75
.50
.25
0.00
0.00 .25 .50 .75 1.00
Observed C.P.
London City

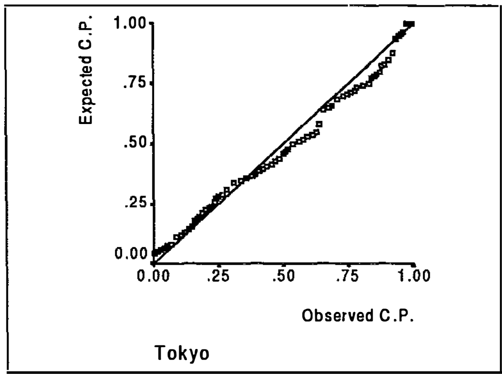

Expected C.P.
1.00
.75
.50
.25
0.00
0.00 .25 .50 .75 1.00
Observed C.P.
London North

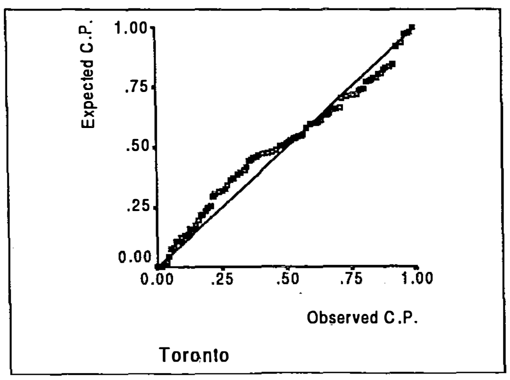

Expected C.P.
1.00
.75
.50
.25
0.00
0.00 .25 .50 .75 1.00
Observed C.P.
Los Angeles


Expected C.P.
1.00
.75
.50
.25
0.00
0.00
.25
.50
.75
1.00
Observed C.P.
New York City

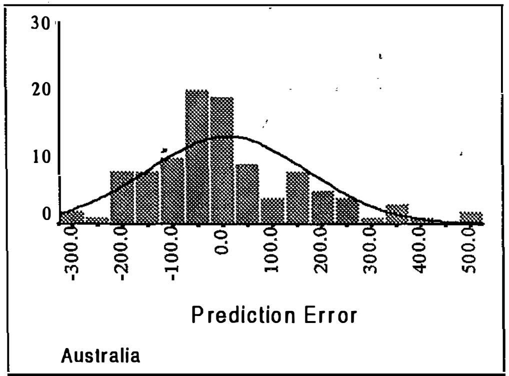

Expected C.P.
1.00
.75
.50
.25
0.00
0.00
.25
.50
.75
1.00
Observed C.P.
Paris

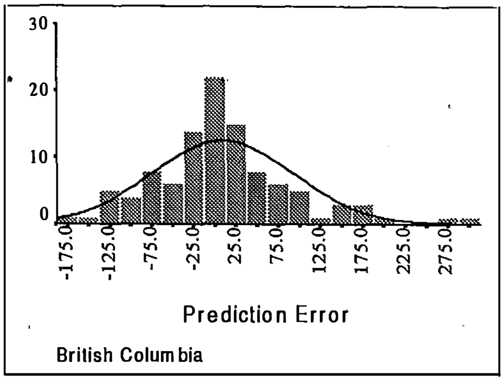

Expected C.P.
1.00
.75
.50
.25
0.00
0.00
.25
.50
.75
1.00
Observed C.P.
Sydney

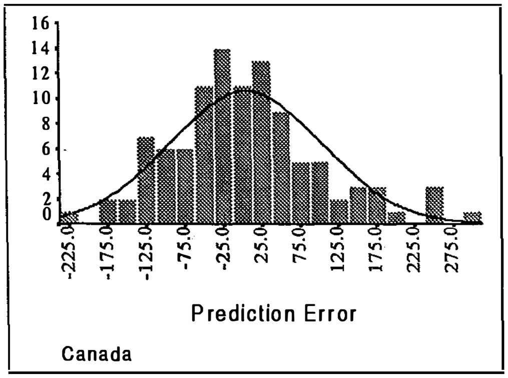

Expected C.P.
1.00
.75
.50
.25
0.00
0.00
.25
.50
.75
1.00
Observed C.P.
Tokyo

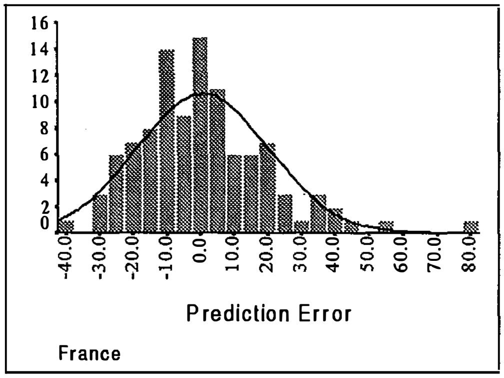

Expected C.P.
1.00
.75
.50
.25
0.00
0.00
.25
.50
.75
1.00
Observed C.P.
Toronto

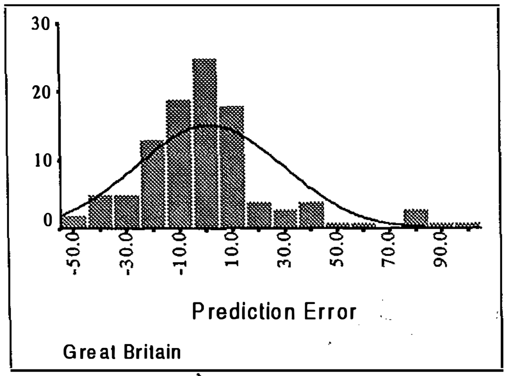
B Histograms of $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$

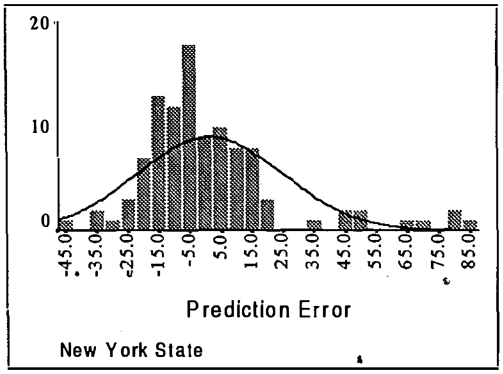


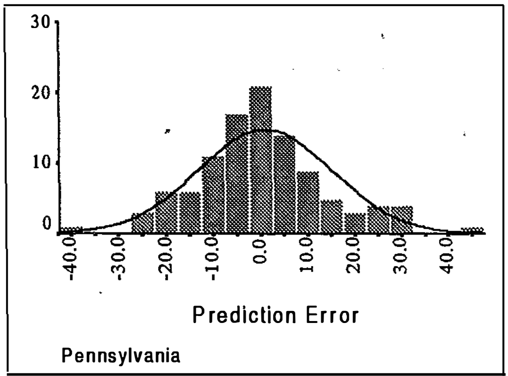


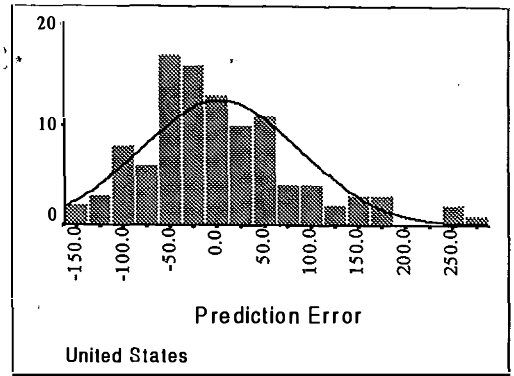


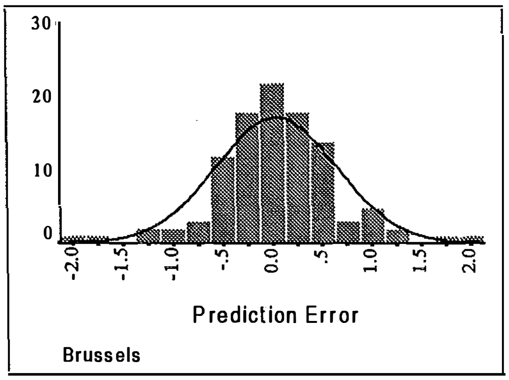


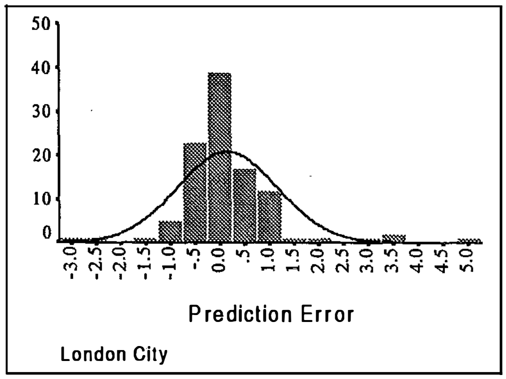


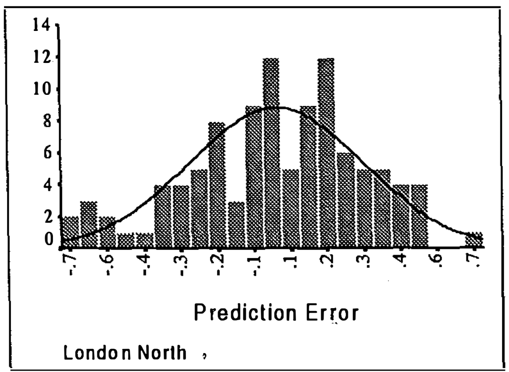


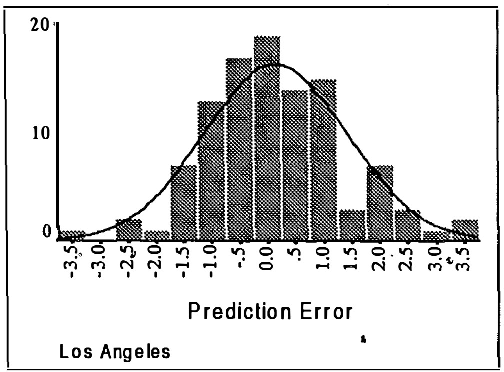

30
20
10
0
-40.0 -30.0 -20.0 -10.0 0.0 10.0 20.0 30.0 40.0
Prediction Error
Pennsylvania

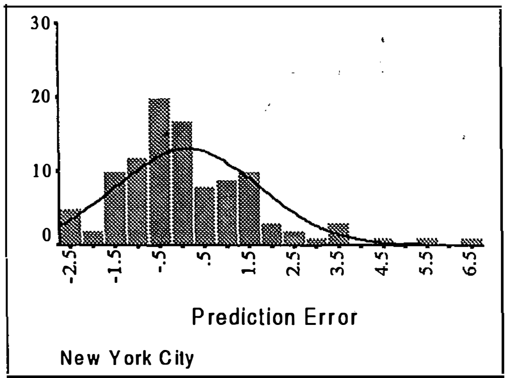

20
10
0
-150.0 -100.0 -50.0 0.0 50.0 100.0 150.0 200.0 250.0
Prediction Error
United States

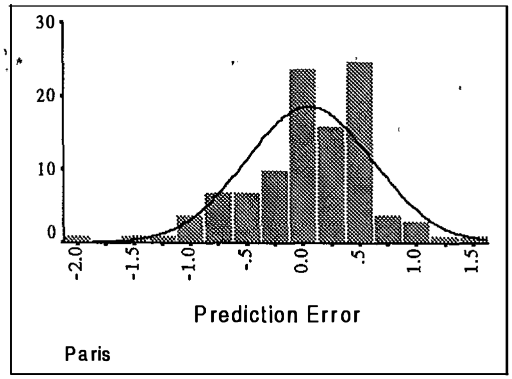

30
20
10
0
-2.0 -1.5 -1.0 -.5 0.0 .5 1.0 1.5 2.0
Prediction Error
Brussels

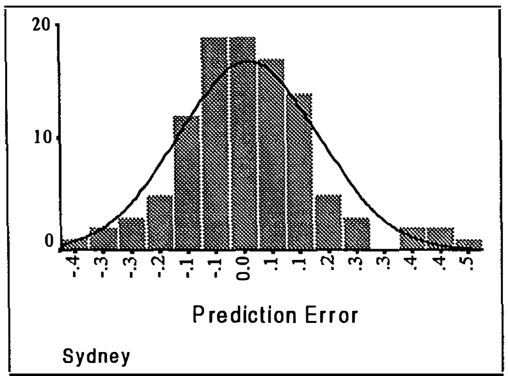

50
40
30
20
10
0
-3.0 -2.5 -2.0 -1.5 -1.0 -.5 0.0 .5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0
Prediction Error
London City

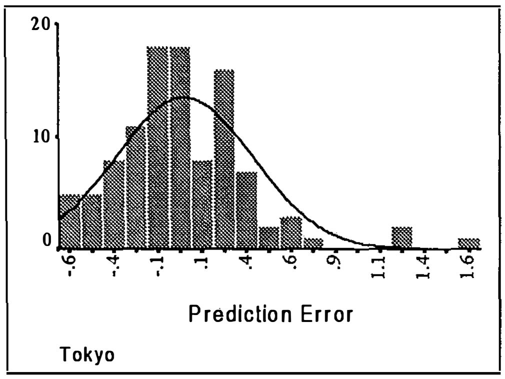

14
12
10
8
6
4
2
0
-.7 -.6 -.4 -.3 -.2 -.1 .1 .2 .3 .4 .6 .7
Prediction Error
London North

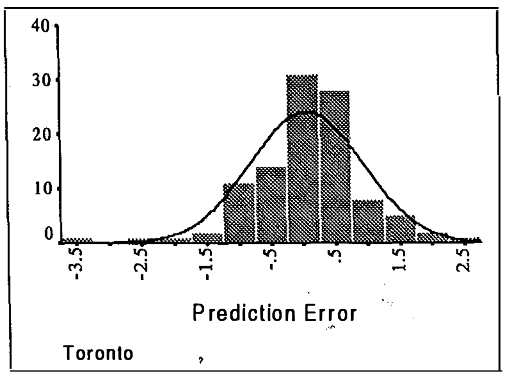

20
10
0
-3.5 -3.0 -2.5 -2.0 -1.5 -1.0 -.5 0.0 .5 1.0 1.5 2.0 2.5 3.0 3.5
Prediction Error
Los Angeles


30
20
10
0
-2.5
-1.5
-.5
.5
1.5
2.5
3.5
4.5
5.5
6.5
Prediction Error
New York City

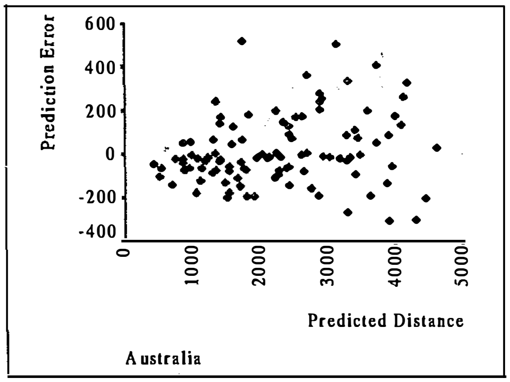

30
20
10
0
-2.0
-1.5
-1.0
-.5
0.0
.5
1.0
1.5
Prediction Error
Paris

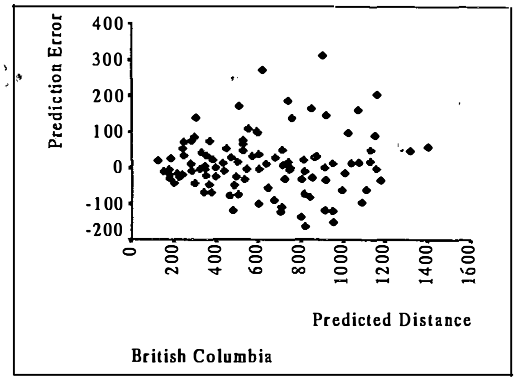

20
10
0
-.4
-.3
-.3
-.2
-.1
-.1
0.0
.1
.1
.2
.3
.3
.4
.4
.5
Prediction Error
Sydney

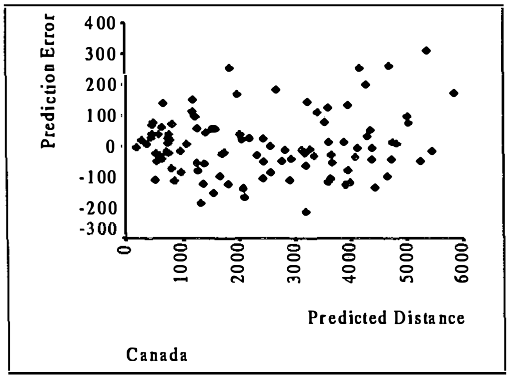

20
10
0
-.6
-.4
-.1
.1
.4
.6
.9
1.1
1.4
1.6
Prediction Error
Tokyo

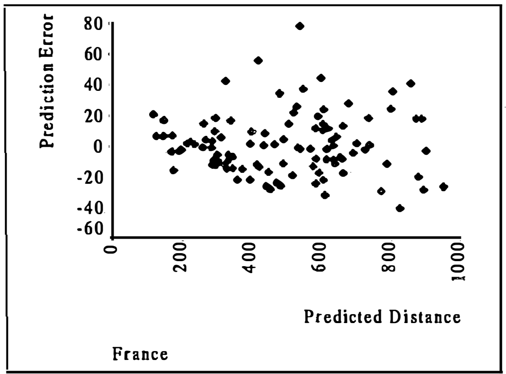

40
30
20
10
0
-3.5
-2.5
-1.5
-.5
.5
1.5
2.5
Prediction Error
Toronto

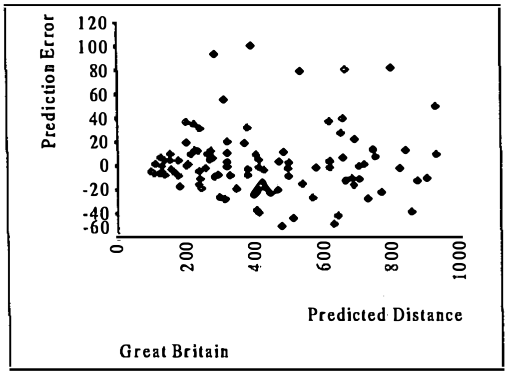
C Scatter Plots of $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$

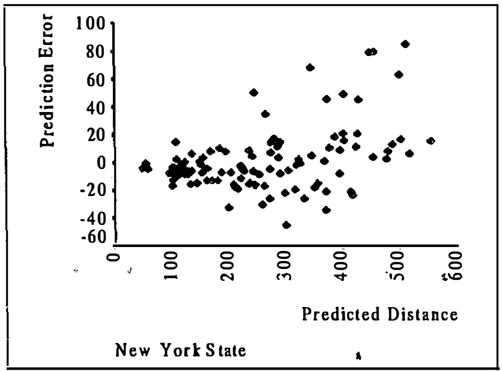


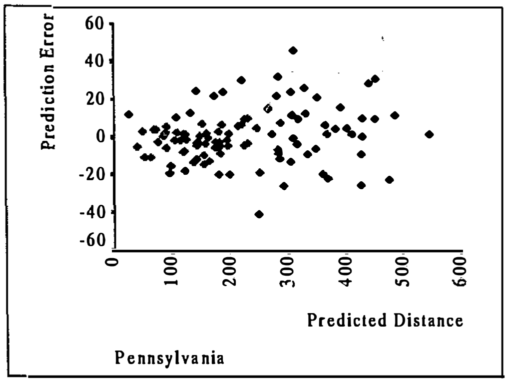


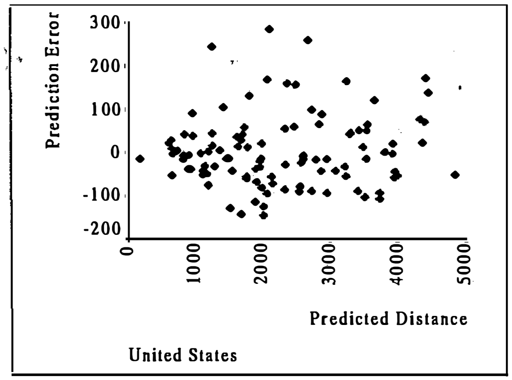


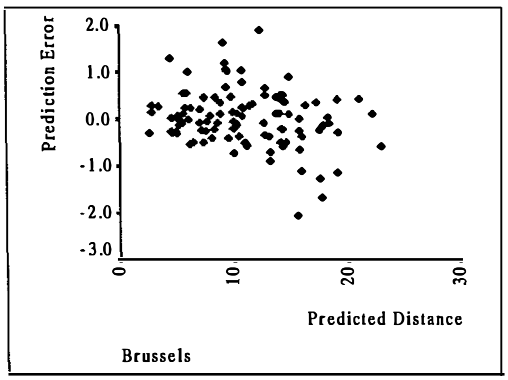


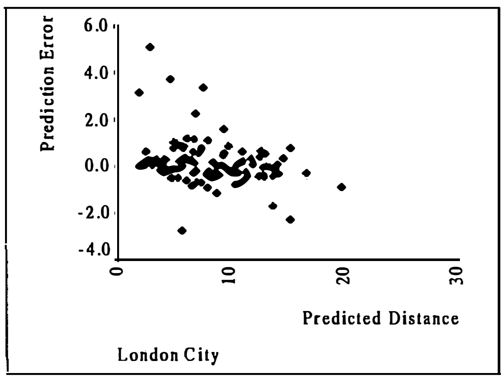


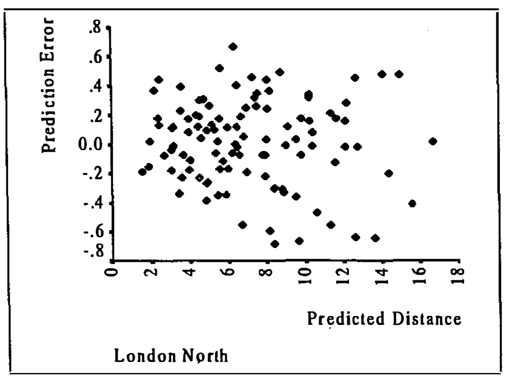


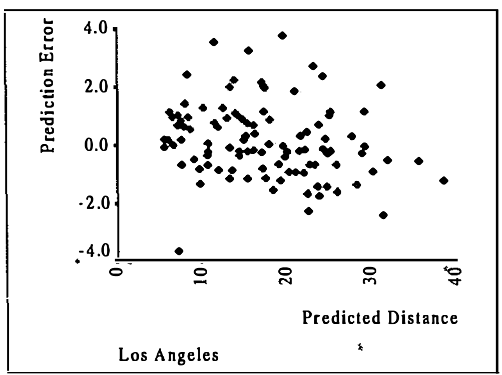

Prediction Error
60
40
20
0
-20
-40
-60
0
100
200
300
400
500
600
Predicted Distance
Pennsylvania

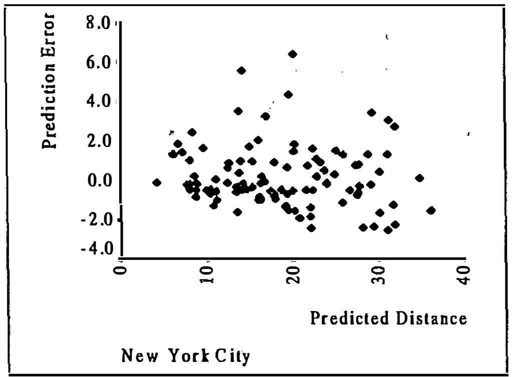

Prediction Error
300
200
100
0
-100
-200
0
1000
2000
3000
4000
5000
Predicted Distance
United States

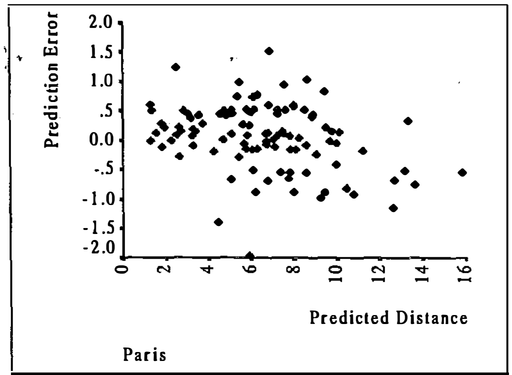

Prediction Error
2.0
1.0
0.0
-1.0
-2.0
-3.0
0
10
20
30
Predicted Distance
Brussels

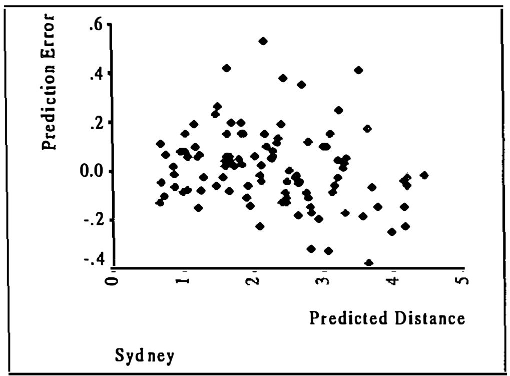

Prediction Error
6.0
4.0
2.0
0.0
-2.0
-4.0
0
10
20
30
Predicted Distance
London City

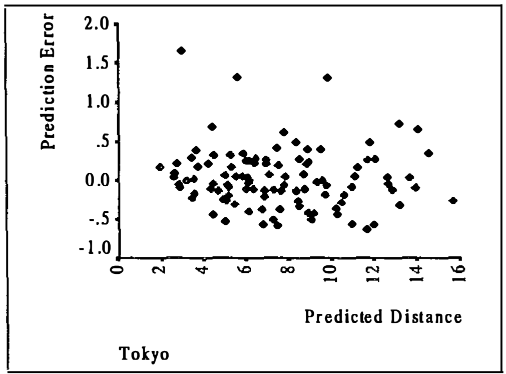

Prediction Error
.8
.6
.4
.2
0.0
-.2
-.4
-.6
-.8
0
2
4
6
8
10
12
14
16
18
Predicted Distance
London North

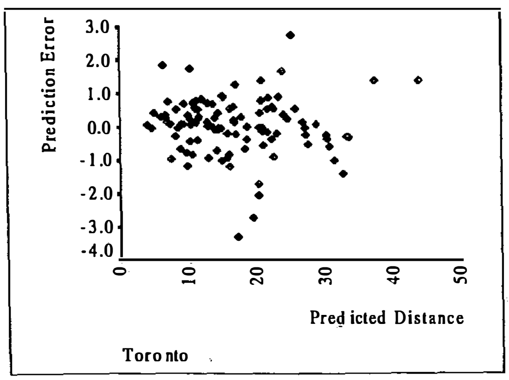

Prediction Error
4.0
2.0
0.0
-2.0
-4.0
0
10
20
30
40
Predicted Distance
Los Angeles


Prediction Error
8.0
6.0
4.0
2.0
0.0
-2.0
-4.0
0
10
20
30
40
Predicted Distance
New York City

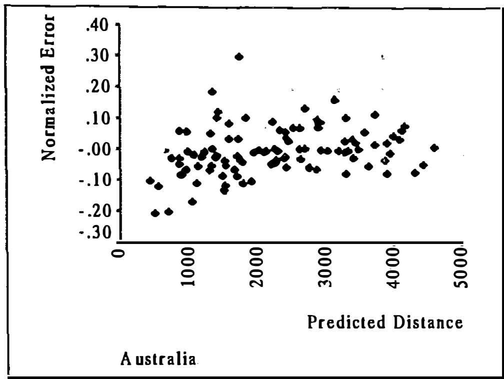

Prediction Error
2.0
1.5
1.0
.5
0.0
-.5
-1.0
-1.5
-2.0
0
2
4
6
8
10
12
14
16
Predicted Distance
Paris

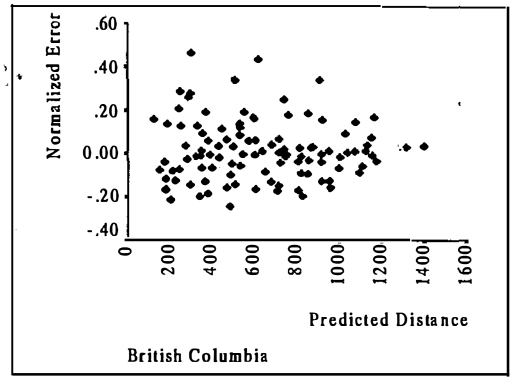

Prediction Error
.6
.4
.2
0.0
-.2
-.4
0
1
2
3
4
5
Predicted Distance
Sydney

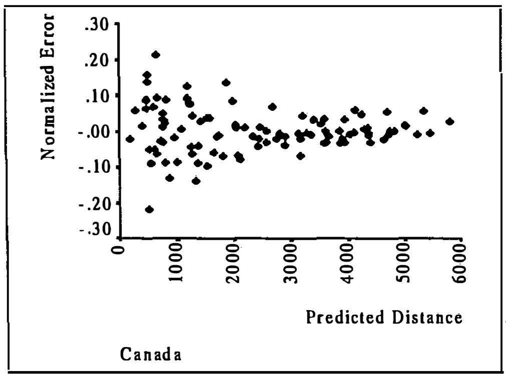

Prediction Error
2.0
1.5
1.0
.5
0.0
-.5
-1.0
0
2
4
6
8
10
12
14
16
Predicted Distance
Tokyo

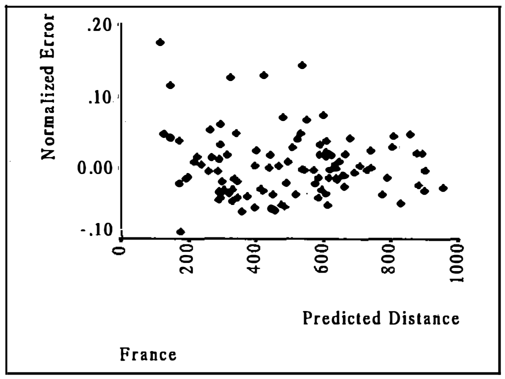

Prediction Error
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
-4.0
0
10
20
30
40
50
Predicted Distance
Toronto

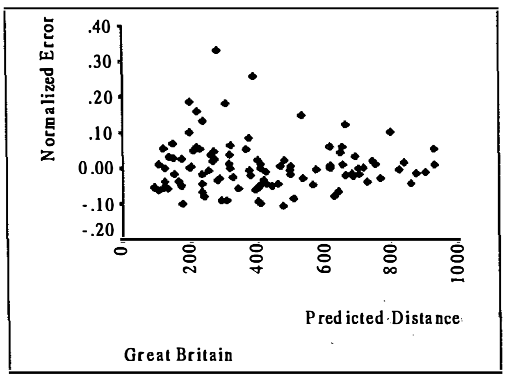
D Scatter Plots of $\varepsilon_1(\mathbf{x}_i, \mathbf{x}_j)$

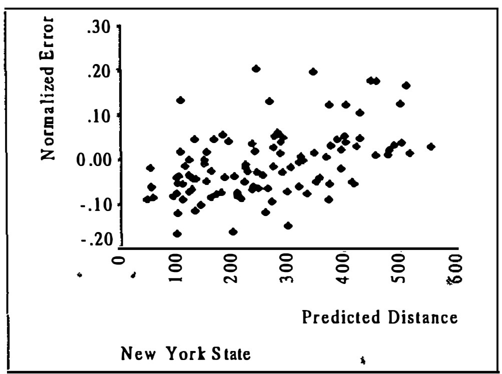


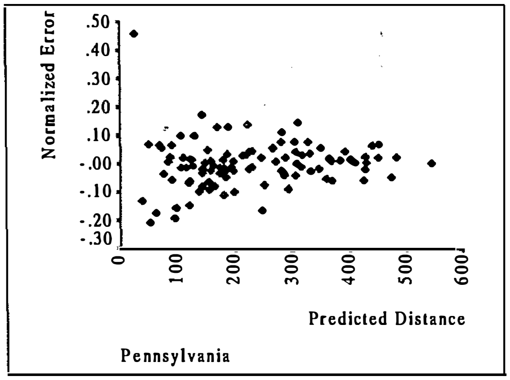


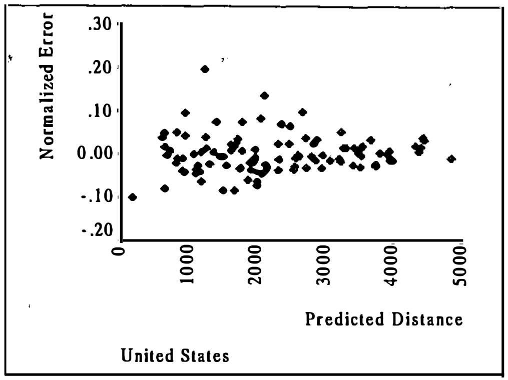


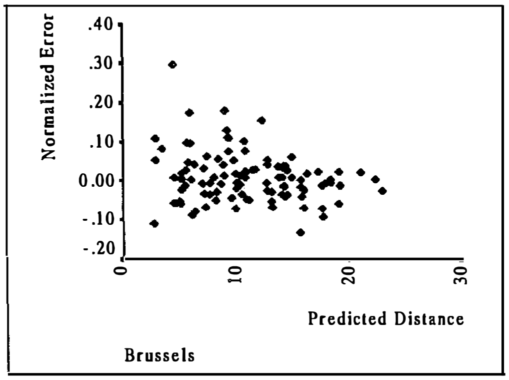


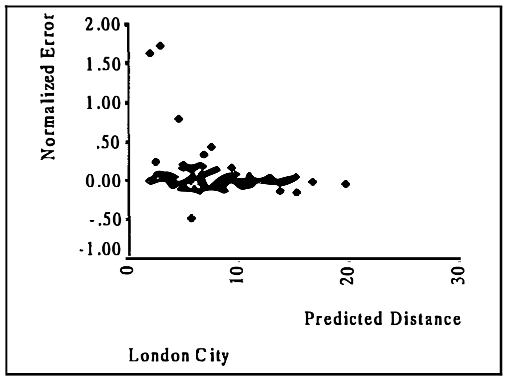


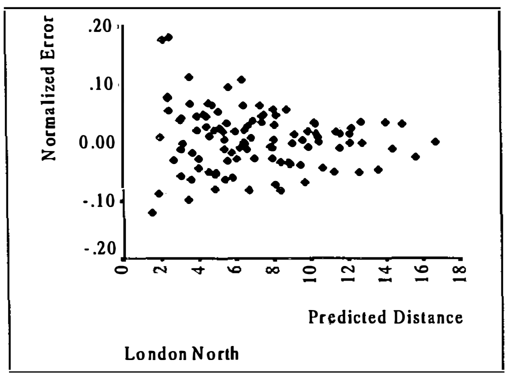


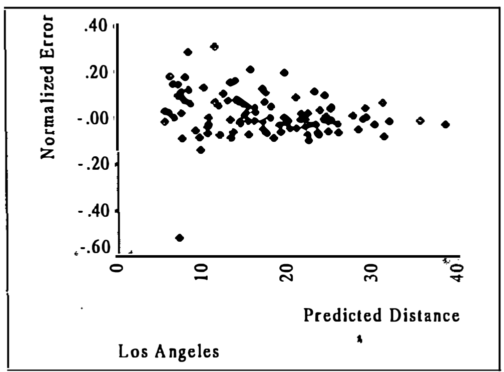

Normalized Error
.50
.40
.30
.20
.10
-.00
-.10
-.20
-.30
0
100
200
300
400
500
600
Predicted Distance
Pennsylvania

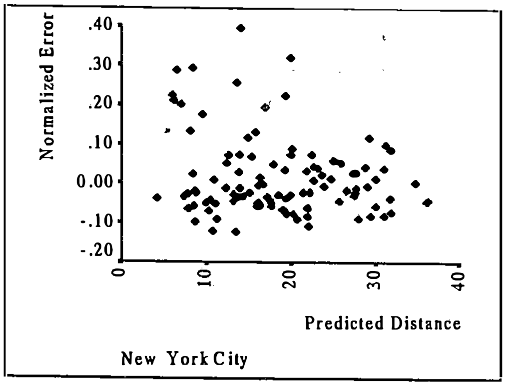

Normalized Error
.30
.20
.10
0.00
-.10
-.20
0
1000
2000
3000
4000
5000
Predicted Distance
United States

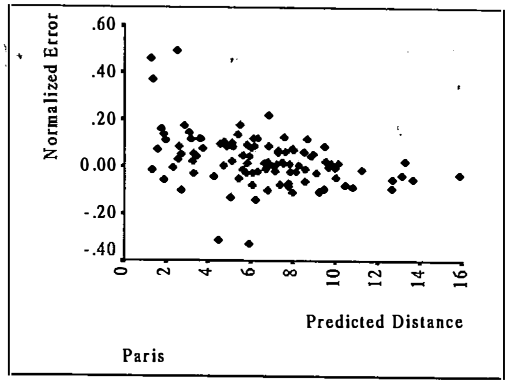

Normalized Error
.40
.30
.20
.10
0.00
-.10
-.20
0
10
20
30
Predicted Distance
Brussels

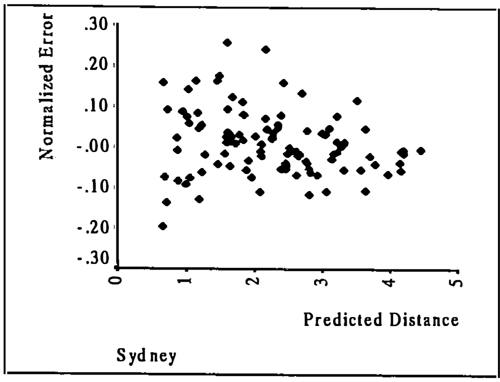

Normalized Error
2.00
1.50
1.00
.50
0.00
-.50
-1.00
0
10
20
30
Predicted Distance
London City

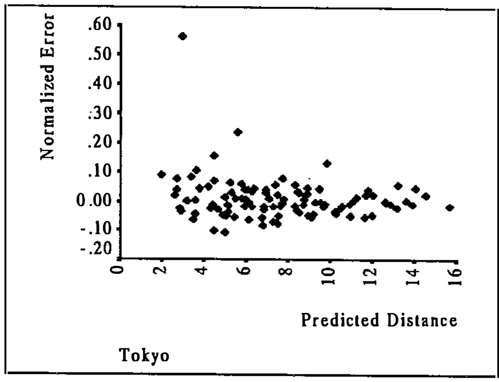

Normalized Error
.20
.10
0.00
-.10
-.20
0
2
4
6
8
10
12
14
16
18
Predicted Distance
London North

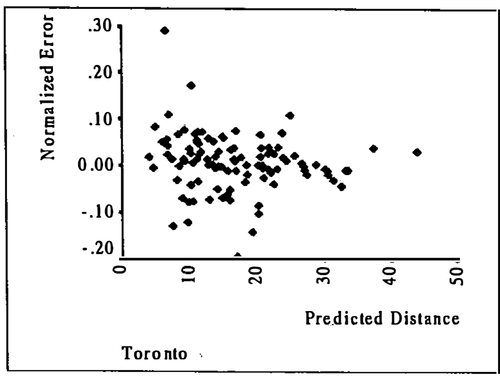

Normalized Error
.40
.20
-.00
-.20
-.40
-.60
0
10
20
30
40
Predicted Distance
Los Angeles


Normalized Error
.40
.30
.20
.10
0.00
-.10
-.20
0
10
20
30
40
Predicted Distance
New York City


Normalized Error
.60
.40
.20
0.00
-.20
-.40
0
2
4
6
8
10
12
14
16
Predicted Distance
Paris


Normalized Error
.30
.20
.10
-.00
-.10
-.20
-.30
0
1
2
3
4
5
Predicted Distance
Sydney


Normalized Error
.60
.50
.40
.30
.20
.10
0.00
-.10
-.20
0
2
4
6
8
10
12
14
16
Predicted Distance
Tokyo


Normalized Error
.30
.20
.10
0.00
-.10
-.20
0
10
20
30
40
50
Predicted Distance
Toronto

Faculty of Business
McMaster University

WORKING PAPERS - RECENT RELEASES

413. J. Brimberg, A. Mehrez and G.O. Wesolowsky, "Allocation of Queueing Facilities Using a Minimax Criterion", January, 1996.

414. Isik Zeytinoglu and Jeanne Norris, "Global Diversity in Employment Relationships: A Typology of Flexible Employment", March, 1996.

415. N. Archer, "Characterizing World Wide Web Search Strategies", April, 1996.

416. J. Rose, "Immediacy and Saliency in Remedying Employer Opposition to Union Organizing Campaigns", July, 1996.

417. Roy J. Adams and Parbudyal Singh, "Worker Rights Under NAFTA: Experience With the North American Agreement on Labor Cooperation", September, 1996.

418. George Steiner and Paul Stephenson, "Subset-Restricted Interchange for Dynamic Min-Max Scheduling Problems", September, 1996.

419. Robert F. Love and Halit Uster, "Comparison of the Properties and the Performance of the Criteria Used to Evaluate the Accuracy of Distance Predicting Functions", November, 1996.

420. Harish C. Jain and Simon Taggar, "The Status of Employment Equity in Canada", December, 1996.

421. Harish C. Jain and Parbudyal Singh, "Beyond The Rhetoric: An Assessment of the Political Arguments and Legal Principles on Strike Replacement Laws in North America", January, 1997.

422. Jason Schwandt, "Electronic Data Interchange: An Overview of Its Origins, Status, and Future", March, 1997.

423. Christopher K. Bart with John C. Tabone, "Mission Statement Rationales and Organizational Alignment in the Not-for-profit Healthcare Sector", November, 1997.

424. Harish C. Jain, Michael Piczak, Işik Urla Zeytinoğlu, "Workplace Substance Testing - An Exploratory Study", November, 1997.

425. S. Suarga, Yufei Yuan, Joseph B. Rose, and Norman Archer, "Web-based Collective Bargaining Support System: A Valid Process Support Tool for Remote Negotiation", January, 1998.

426. Pawan S. Budhwar and Harish C. Jain, "Evaluating Levels of Strategic Integration and Development of Human Resource Management in Britain", March, 1998.

427. Halit Üster and Robert F. Love, "Application of Weighted Sums of Order p to Distance Estimation", April, 1998.

428. Halit Üster and Robert F. Love, "On the Directional Bias of the ℓ_{bp}-norm", April, 1998.

429. Milena Head, Norm Archer, and Yufei Yuan, "MEMOS: A World Wide Web Navigation Aid", October, 1998.

430. Harish C. Jain and Parbudyal Singh, "The Effects of the Use of Strike Replacement Workers on Strike Duration in Canada", February, 1999.

431. Parbudyal Singh and Harish C. Jain, "Strike Replacements in the United States, Canada and Mexico: A Review of the Law and Empirical Research", February, 1999.

432. John W. Medcof and Jeremy Boyko, "Reinforcing, Revising and Reconciling Attributions in the Employment Interview", March, 1999.

433. Norm Archer, "World Wide Web Business Catalogs in Business-to-Business Procurement", March, 1999.

434. Diwakar Gupta and Saifallah Benjaafar, "Make-to-order, Make-to-stock, or Delay Product Differentiation? - A Common Framework for Modeling and Analysis", April, 1999.

435. Harish C. Jain, Parbudyal Singh and Carol Agocs, "Recruitment, Selection and Promotion of Visible Minorities and Aboriginals in Selected Canadian Police Services", April, 1999.

436. Harish C. Jain and Angus Bowmaker-Falconer, "Employment Equity/Affirmative Action Codes of Practice and Best Practices in USA, Britain, Canada and Other Selected Countries", May, 1999.

437. Diwakar Gupta, Yavuz Günalay, and Mandyam M. Srinivasan, "On the Relationship Between Preventive Maintenance and Manufacturing System Performance", June, 1999.

438. Jinliang Cheng, George Steiner, and Paul Stephenson, "A Fast Algorithm to Minimize Makespan for the Two-Machine Flow-Shop Problem with Release Times", June, 1999.

439. Jinliang Cheng, George Steiner, and Paul Stephenson, "A Fast Algorithm to Minimize Maximum Lateness for the Two-Machine Flow-Shop Problem with Release Times, June, 1999.

440. Norm Archer and Yufei Yuan, "Electronic Commerce and the Business-to-Business Customer Relationship Life Cycle", August, 1999.