Bilevel fuzzy optimization to pre-process traffic data to satisfy the law of flow conservation

By: Juan de Oña, Penélope Gómez and Enrique Mérida-Casermeiro

Bilevel fuzzy optimization to pre-process traffic data to satisfy the law of flow conservation

J. de Oña[a], P. Gómez[b] and E. Mérida-Casermeiro[c]

[a] Corresponding author. TRYSE Research Group. Department of Civil Engineering, University of Granada, ETSI Caminos, Canales y Puertos, c/ Severo Ochoa, s/n, 18071 Granada (Spain), Phone: +34 958 24 99 79, Fax: +34 958 24 61 38, jdona@ugr.es
[b] Department of Civil Engineering, University of Granada, ETSI Caminos, Canales y Puertos, c/ Severo Ochoa, s/n, 18071 Granada (Spain), penelopegj@ugr.es
[c] Applied Mathematics Department, University of Malaga, E.T.S. Computer Science Engineering, Boulevard Louis Pasteur s/n, 29071 Málaga (Spain), merida@ctima.uma.es

Abstract

Traffic data obtained in the field usually have some errors. For instance, traffic volume data on the various links of a network must be consistent and satisfy flow conservation, but this rarely occurs. This paper presents a method for using fuzzy optimization to adjust observed values so they meet flow conservation equations and any consistency requirements. The novelty lies in the possibility of obtaining the best combination of adjusted values, thereby preserving data integrity as much as possible. The proposed method allows analysts to manage field data reliability by assigning different ranges to each observed value. The paper is divided into two sections: The first section explains the theory through a simple example of a case in which the data is equally reliable and a case in which the observed data comes from more or less reliable sources, and the second one is an actual application of the method in a freeway network in southern Spain where data were available but some data were missing.

Keywords: traffic counts, fuzzy logic, transport planning, optimization, data consistency, subjective analyst knowledge

2

1. Introduction

The study of any transport system requires enormous quantities of data and an understanding of their dependence on each other. Arguably, volume is the most important traffic datum of them all. Field data is generally inconsistent, and therefore they need to be processed in a way that will make them consistent before they can be used in algorithms for prediction, monitoring and decision-making purposes. The methods used to estimate Origin-Destination (O/D) matrices are based on the hypothetical availability of precise traffic volume data and reliable preliminary O/D data. The input data for most traffic networks, however, are either unavailable or contain measurement errors, as in the case of traffic counts and sensor speed measurements. In fact, some studies (Zhong et al., 2004) demonstrate that 50% of the Permanent Traffic Counts (PTCs) set up on highways contain lost data, making it difficult to ignore measurement errors when processing data used to plan, design, control and manage traffic (Sharma et al., 1996). The existence of errors makes data obtained in the field difficult to manage and to analyze.

In the past, certain methods were applied to adjust the observed values so they would comply with flow conservation laws at each network node, aside from other requirements that values need to meet before they can be used as input data in traffic planning algorithms. The methods used were manual value adjustment, least square adjustment and the maximum likelihood method (Kikuchi et al., 2000). Recently, new methods of value adjustment based on fuzzy logic have been developed to preserve data integrity as much as possible. The methods are: fuzzy regression, fuzzy optimization and necessity-interval-regression method (Kikuchi et al., 2000). A number of important publications on fuzzy logic have been submitted over the past twenty years, although most of them are based on the fields of deduction and control in situations of complex

behavior. Papis and Mamdani (1977) were the first to apply fuzzy logic to transport; specifically, to traffic signal controllers.

Lost data processing is another frequent issue. When available input data exist at all, they often contain errors due to the sensors´ operating faults (Kwon et al., 2008). From a formal viewpoint, the problem of debugging input data in order to avoid inconsistency and of assigning values to missing data has generally been analyzed by an area of Statistics (Data Editing and Imputation). Most efforts have focused on processing 'Missing values', and on detecting and debugging. Inconsistencies have been avoided by using redundant or related information. Some classical techniques are: imputation by mean, median, regression or hot-deck (Chambers, 2001; Laaksonen, 1999). Recently, some new techniques based on Artificial Intelligence and on neural networks, in particular, are being developed (Silva-Ramírez, 2007; Tussel, 2002). Certain authors (Kaczmarek, 2005; Marzano et al., 2008; Rudy et al., 2008) have submitted methods based on the characteristics of erroneous traffic data in urban networks, supplemented with the latest data imputation models (Lee et al., 1998; Geng and Wu, 2008). Other methods based on weighted least squares regression also exist, such as the methods submitted by Kwon et al. (2008).

The aim of this article is to submit a method whereby field data could be pre-processed to make them consistent while preserving their integrity as much as possible, and which would include their reliability as perceived subjectively by the analyst. The method is based on fuzzy logic and is intended to optimize the solution obtained. The result would be a reliable solution that comes close to the observed values, thereby resolving measurement errors in traffic counts. The method also allows field data to be processed when there are lost values.

(Place Fig. 1 about here)

4

2. Description of the problem

A simple freeway network is used to explain the method. Consider the situation shown in Fig. 1, in which real consistent data are available (Table 1, column 2). The data are used to simulate a scenario with non consistent data: traffic counts from the database are randomized within ±25% of their values at all intersections to simulate a case in which data is not consistent (Table 1, column 3). Next, the randomly obtained data in the database are considered to be field data; i.e., the observed values (OV).

(Place Table. 1 about here)

Theoretically, in any transport network such as the one shown in Fig. 1, the total "incoming volumes" should be equal to the total "outgoing volumes" at any node in the network and in any flow direction in such a way that the law of conservation of flow is satisfied. In the simulated scenario, (Table 1, column 3), however, it is found that:

$$x_3+x_5 \neq y_5+y_6$$

$$x_1+x_2 \neq y_1+y_4$$

$$y_1+y_2 \neq z_1+z_4$$

$$y_3+y_5 \neq z_5+z_6$$

$$y_2+y_6 \neq w_{10}+w_{11}+w_{12}$$

$$y_3+y_4 \neq w_1+w_5+w_7$$

Actually, this is usually the case, particularly when the network is large. Pentrice (1987) stated that data inconsistency is inevitable even in a well-controlled survey, but volume count consistency at different links is critical to ensuring the integrity of the results of any of the ITS-related algorithms.

When the network becomes larger, the possibility of inconsistency in traffic volume counts increases, so flow conservation is more difficult. The concern in this paper is how to adjust the individual observed volumes to a set of new values that satisfy the

flow conservation principle at any point in the network. Furthermore, the adjustment should be such that the integrity of the observed values is preserved as much as possible. To this end, a fuzzy optimization method is used to obtain adjusted values that comply with the law of flow conservation and that resemble consistent real data as closely as possible. In this example, the integrity of the results obtained can be verified with the available real consistent data.

3. The bilevel fuzzy optimization method

The search for the "best" set of adjusted values is an optimization process that aims to find a set of values close to the observed ones that verifies the conservation of flow principle.

The proposed method is based on the following concept: Each observed value is considered an approximate value represented by a fuzzy number, defined by a membership function. If the value is x, it is interpreted as "approximately x". The true value is considered to lie near x. The method attempts to find an adjusted value as close to the observed value as possible while satisfying the conservation of flow at every point in the network. This is accomplished by applying the concept of fuzzy optimization developed in fuzzy set theory.

Given a set of observed values, there are an infinite number of combinations of adjusted values, each of which satisfies the set of flow conservation equations. For a given combination, the membership grade $h_{x_i}(x'_i)$ of each adjusted value (x'_i) in the corresponding fuzzy set (x_i) is calculated. Three methods of optimization could be used:

a. by Maximizing the Minimum $h_{x_i}(x'_i)$ for all i,

b. by Maximizing the Sum of $h_{x_i}(x'_i)$, and,

6

c. by maximizing the Minimum $h_{x_i}(x'_i)$ for all i at one level and, after this has been achieved, by applying a second level of optimization by Maximizing the Sum of $h_{x_i}(x'_i)$. Thus, the combination with the highest sum is selected from among all the combinations that could maximize the lowest membership grade. The value with the least membership grade is taken into consideration, and also all the other observed data.

In case (a) (MM method), the lowest membership grade for the combination is recorded. By comparing the lowest membership grades among all the combinations of traffic volumes, the one that has the highest value is chosen as the best combination of a set of adjusted values. This method was already introduced by Kikuchi and Miljkovic (1999).

On the other hand, in the objective function sum of $h_{x_i}(x'_i)$'s (case (b)) (MS method) for a given combination, the membership grade of each adjusted value in the corresponding fuzzy set is calculated. The sum of the membership grades among all the combinations of traffic volumes is recorded, and the one that has the highest sum of membership grades is chosen as the best combination of a set of adjusted values.

The third possibility is a two step way of optimization or Bilevel Optimization method (BO method). In step one, case a), the lowest membership grade is maximized. In step two, the membership grades that would produce the largest possible max(min(h_i)) and that would seek to increase the value of all of the h_i at the same time (which would achieve the sum of both) are summed up and maximized.

The MM method can attend to a set of data which its minimum membership grade is maximized but the problem is that an infinite number of combinations could satisfy this condition and the MM method randomly chooses one of them. The BO method chooses a set that while it satisfies that condition; it optimizes the rest of the values, maximizing

7

the membership grade of all the data, so the BO method uses both ways of optimization in order to improve the solution.

The mathematical steps involved in addressing the optimization problem are:

1. Use fuzzy numbers to represent observed values

2. Formulate the objective and constraints

3. Solve as a mixed linear programming problem

The process is explained step by step by using the simple highway network shown in Fig. 1.

3.1. Using fuzzy numbers to represent observed values

The observed values are "fuzzified" and are considered a fuzzy set with a triangular membership function.

(Place Fig. 2 about here)

Fig. 2 shows the shape of the membership function with the centre value x_i and a range $[x_i-\alpha x_i, x_i+\alpha x_i]$, where α is a constant higher than 0. The triangular membership function is not a prerequisite but, in the absence of any other information, this is a reasonable assumption, and such assumption is often used in fuzzy set theory (Zimmermann, 2001).

The selection of the constant α depends on the judgement of the analyst with respect to the adjusted value's acceptable deviation from the observed value. This value allows the analyst to enter the reliability of each datum (i.e. the more reliable data will have a lower value of α than if they were less reliable). If only one value of α is used for all data, the scope of the range has little effect on the final adjusted values, once it is broad enough for a feasible set of solutions to be found.

8

The membership function is defined for the left- and right-hand sides of the triangle. For an observed value of x_i and the assumed range $[x_i-\alpha x_i, x_i+\alpha x_i]$, the general expression of the membership functions is:

$$h_{x_i}(x'_i) = \begin{cases} h_{x_i}^L = \dfrac{x'_i - (x_i - \alpha x_i)}{\alpha x_i} & \text{if } x_i - \alpha x_i < x'_i \leq x_i \\[2mm] h_{x_i}^R = \dfrac{x'_i - (x_i + \alpha x_i)}{-\alpha x_i} & \text{if } x_i < x'_i \leq x_i + \alpha x_i \end{cases} \tag{1}$$

In this formula $-\infty < x_i-\alpha x_i \leq x_i \leq x_i+\alpha x_i < \infty$, the triangular fuzzy number x_i is presented by $(x_i-\alpha x_i, x_i, x_i+\alpha x_i)$.

For the sake of simplicity, a symmetric triangle is used in this paper for the membership function. However, the left and right-hand limits can be set separately. To solve this example problem, it is assumed that the value x_i is the observed value and that the value of $\alpha > 0$. So the value of α is the spread of triangular fuzzy number x_i. The narrower the spread area, the less fuzzy the evaluation data will be, hence more precise. To the contrary, fuzziness is higher and thus more vague and ambiguous when the spread area increases (Tzeu-Chen Han, 2008).

Some authors have researched calibration of the membership function extensively. The classical approach to calibration has been the intuitive trial and error process, in which the analyst modifies the shapes of the membership functions little by little until the predicted output approximately fits the output data obtained from the real world (Chakroborty and Kikuchi, 2003). However, this process is time consuming. Other authors have developed a systematic way of carrying out the trial and error process (Wang and Mendel, 1992a, 1992b, 1992c; Homaifar and McCormick, 1995). The purpose of calibration is to modify the membership functions of the Fuzzy Inference

9

System (FIS) so that the outcome predicted by the model is equal (or nearly equal) to the outcome obtained in the real world. Therefore, Chakroborty and Kikuchi (2003) presented a method in which a representation framework allows the FIS parameters to be modified in relation to the bases. FIS outputs are dictated by the parameters that define the membership functions of the fuzzy sets appearing in the antecedents and the consequents of the rules and the algebraic operators used for the logical connectives and to determine the final inferred value. They have developed a procedure that calibrates the membership function of the fuzzy sets by transforming the inference system into an Artificial Neural Network format. They have applied this procedure to the complex control task of car-following, but this procedure has not been applied yet to an urban transport system or a large-scale civil infrastructure system.

3.2. Formulating the objective function and its constraints

In a fuzzy number representation of observed values, fuzzy optimization techniques would be used to search for the adjusted values. The mathematical formulation of the three proposed methods used to solve the problem would be as follows:

A. <u>MM method</u>:

$$\text{Max(h) where h is min(h}_i) \tag{2}$$

Subject to

o Constraints related to the membership functions:

$$h^L_{x_i}(x'_i) \geq h \qquad h^R_{x_i}(x'_i) \geq h \qquad h_i \geq h \qquad \text{for i=1, k} \tag{3}$$

which means there are 2k+k constraints (where k is the number of control points)

o Constraints related to the conservation of flow at each control point. The constraints are defined by reviewing the flow pattern at each node in Fig. 1 as follows:

$$x'_3 + x'_5 = y'_5 + y'_6$$

$$x'_1 + x'_2 = y'_1 + y'_4$$

10

$$y'_1 + y'_2 = z'_1 + z'_4 \tag{4}$$

$$y'_3 + y'_5 = z'_5 + z'_6$$

$$y'_2 + y'_6 = w'_{10} + w'_{11} + w'_{12}$$

$$y'_3 + y'_4 = w'_1 + w'_5 + w'_7$$

$$x'_i, y'_i, z'_i, w'_i \geq 0 \text{ for all } i$$

Where

x'_i, y'_i, z'_i, w'_i integer unknown adjusted values

x_i, y_i, z_i, w_i fuzzy set corresponding to the observed value x_i

$h_{x_i}(x'_i)$ membership grade of x'_i in the fuzzy set x_i, the same treatment

for y_i, z_i and w_i

h   an operational parameter that represents the smallest membership grade among all $h_{x_i}(x'_i)$'s. Where $h^L_{x_i}(x'_i) \geq h$ and $h^R_{x_i}(x'_i) \geq h$, respectively, show the expressions for the left- and right-hand sides of the triangle.

B. <u>MS method</u>:

$$\text{Max}(g) \text{ where } g \text{ is sum}(h_i) \tag{5}$$

Subject to the same constraints as in the MM method, with regard to the membership functions (eq. 3) and to the conservation of flow at each control point (eq. 4).

C. <u>BO method</u>:

Step 1: The problem is solved using the MM method (eq. 2), and we obtain a value of h=h*.

Step 2: The problem is solved using the MS method (eq. 5) subject to the same constraints with regard to the conservation of flow at each control point (eq. 4) as in the MM or MS method, and to the following constraints related to the membership functions:

11

$$h_{x_i}^L(x_i') \ge h* \qquad h_{x_i}^R(x_i') \ge h* \qquad h_i \ge h* \qquad \text{for } i=1, k \qquad (6)$$

The total number of unknowns in Step 2 is reduced by one compared to Step 1.

If only Max(h) is performed (case A), there may be several imputations for the observed data that produce the same value for h (Tussel, 2002; Silva-Ramírez, 2007). Therefore, they would be the same from the objective function point of view, whereas, in fact, some are better than others. The combination (0.9, 0.9, 0.9), for instance, would have the same value as (0.9, 1, 1), whereas the latter is better than the former. On the other hand, if the objective function were just Max(g) (case B), some values would show a h_i=0.00, despite the fact that almost all the rest are 1.00, which is of no interest. The bilevel optimization process (case C) allows the combination where the remaining membership degrees are the highest ones to be chosen from among all the combinations where the lowest value of h is maximized.

3.3. Solving as a mixed linear programming problem

Since every x'_i must be an integer number and h_i are real numbers, this is a mixed linear programming formulation. A mixed linear programming algorithm is formulated for the problem to maximize the membership grade of the adjusted values.

In Fig. 1, the mixed linear programming algorithm consists of 90 (3x30 observed volumes) inequality constraints related to membership functions and six equations related to flow conservation.

3.4. Introduction of data reliability

The selection of the value of α depends on the judgement of the analyst with respect to the adjusted value's acceptable deviation from the observed value.

In a complex transport network, there may be permanent traffic count stations where count data are fairly reliable, and other nodes where counting is sporadic, as well as points where traffic volumes have not been measured. Therefore, to define the α

parameter coherently, the method must allow the analyst to assign different values to the α parameter in order to define the membership functions of each observed value. The values will depend on whether the parameter belongs to a set of data that are highly reliable (permanent traffic count station), averagely reliable (sporadic count) or highly unreliable (lost data).

3.5. Example network

As shown in Fig.1, the example consists in analysing a network of 4 intersections, of which three have 6 movements and one has twelve.

In this example, the real consistent data are known (RV) (Table 1 column 2). The data are used to simulate a scenario with non consistent data. The simulated data are considered the OV (Table 1 column 3).

In this example, it is considered that traffic count station W is a permanent station, so the values have maximum reliability and their α parameter is the lowest, $\alpha=0.3$. The reliability of stations Y and Z is lower so α takes a value of 0.5 (sporadic count stations) and, finally, the data from traffic count station X is supposed to be the least reliable one, so α is assigned a value of 0.65.

3.6. Results

In this case, since real data were available, three indicators could be used to verify the goodness of the adjustment of each one of the three optimization methods used (MM, MS and BO methods):

- The first indicator is the lowest value of h, which indicates the membership grade of the worst adjusted value (the degree of compatibility between the adjusted value and the observed value). If the value of h is near zero, then the adjusted value is close to the right or left end of the base of the membership function; if the value of h is near 1, then the adjusted value is close to the

13

observed value. Therefore, the solution where the lowest value of h is maximum is chosen as the best solution from the point of view of this parameter.

- The second indicator is the sum of h_i. The bestsolution is where the sum of h_i is maximum, because the adjusted values are closer to the observed values and integrity is more preserved.

- The last indicator, for which the consistent real data are available, is the average of the differences between the real consistent values and the adjusted values.

The results for the three methods are given in Table 1, where the adjusted values (AV) and the value of the membership grade (h) for each observed value are shown. The membership grade of the individual AV is computed by entering the adjusted value (x'_i) in the respective membership function, $h_{x_i}(x'_i)$. The table also shows the effect of using different α values, depending on the reliability of the observed volumes at each intersection.

Column 1 of Table 1 shows each movement in nodes W, X, Y and Z. Column 2 shows the consistent RV used to obtain the OV that show inconsistencies by randomizing the values within ±25%.

Columns 4, 6 and 8 in Table 1 shows the AV, the corresponding values of h (h_i) and the difference (Δ) between RV and AV in absolute value, using the MM method, MS method and BO method respectively for an α parameter of 0.4 in all cases:

o MM method's results are shown in column 4. The lowest value of h in column 4 (h=0.56) indicates the membership grade of the worst adjusted value. In this case Σh_i is 24.04.

o In column 6, MS method's results show that whereas most of the adjusted values get h=1.00, other values show lower h and h could even be 0.00, in order to manage the highest Σh_i. The lowest value of h is reached for y_1 (h=0.13). This situation,

14

therefore, is not desirable either, since it allows a set of values with some h very close to 0.00 to be considered, providing the sum is the maximum. In this case, Σh_i is 27.93.

o   The BO method's results are shown in column 8. If columns 4 and 8 are compared, it can be seen that the minimum value of h remains the same (h=0.56). However, there has been an increase in Σh_i, which has gone from 24.04 (MM method) to 25.89 (BO method). Thus, this new method allows a combination where the remaining membership degrees are the highest ones to be chosen from among all the combinations with the lowest value of h.

As explained above, introducing the analyst's knowledge of the different precisions of the data he is working with improves the results of the adjustment. This is shown in columns 5, 7 and 9 in Table 1 where the AV, h_i and Δ are calculated, using the three methods for different α parameters depending on the reliability of the data. The α values used in this example have been 0.65 for "X", 0.5 for "Y" and "Z" and 0.3 for "W".

As in the case of the same α for every observed value, for any α parameter, the MM and the BO methods obtain the same and a higher value of h minimum (h=0.61) than the MS method (h=0.25). However, the latter method obtains a higher value of Σh_i (28.26 versus 24.92 for the MM method and 25.98 for the BO method).

The results shown in column 5 are better than those shown in column 4. This is because the minimum value of h and Σh_i were higher and the value of average Δ was lower. Similar results are obtained by comparing columns 6-7 and 8-9 for the MS and the BO methods. This confirms the advantage of distinguishing between reliable data and less reliable data or, in other words, of introducing the subjective perception of the analyst.

The last row in Table 1 shows the average of Δ for each of the three methods used. It can be seen that the lowest value (6.97) is obtained for the BO method with different values of α, in comparison to the values of the MM method (7.13) and the MS method (10.70) with different values of α. This shows that the AV obtained with the BO method are closer to the real values than with the other two methods, so this is the method that best preserves the integrity of data.

4. Real intersections in Andalucía motorway´s network

Next, the three methods are used to adjust the traffic volumes of a series of adjacent intersections in Andalusia's freeway network (see Fig. 3 and Fig. 4) for which real and therefore inconsistent data are available. In this example, the parameter Δ is omitted, and only two parameters have been used to verify the goodness of the adjustment: the lowest value of h and the sum of h_i.

(Place Fig. 3 about here)

(Place Fig. 4 about here)

The network has five intersections, of which three have six movements (intersections V, X and Z), while the other two have twelve potential movements (intersections W and Y). Data is available for all potential movements except for v_1, v_2, v_3, v_4, y_6, y_7, and y_8, whose values were lost. A special membership function with h=1.00 always (α→∞) was assigned to the lost values so that any adjusted value that met the boundary conditions would always have a membership grade of 1.00 (Fig. 5). Table 2 shows that for movements v_1, v_2, v_3, v_4, y_6, y_7 and y_8, the value of h associated to the AV is always 1.00 for the three methods studied and for the hypothesis of equal or different α.

(Place Fig. 5 about here)

(Place Table 2 about here)

Columns 3, 5 and 7 in Table 2 show the AV and h_i using the three methods for α=0.1.

16

On the other hand, columns 4, 6 and 8 in Table 2 show the AV and h_i using the three methods for different α parameters depending on the reliability of the data. The α values used in this example were 0.2 for "W", 0.3 for "Z", and 0.1 for the rest.

As in the previous example, for any α parameter, the MM and the BO methods obtain the same and a higher value of minimum h (h=0.29) than the MS method (h=0.00). However, MS method obtains a higher value of Σh_i (40.14 versus 26.16 for the MM method and 35.97 for the BO method). Thus, the results demonstrate that the BO method, while keeping the highest minimum of h, attains the best sum of h_i, so the best solution is chosen from among all the possibilities that satisfy the condition of maximizing the minimum h. Furthermore, introducing the analyst's knowledge of the different precisions of the data he is working with improves the results of the adjustment.

5. Summary and conclusions

The consistency of the observed traffic data is a concern because in nearly all cases traffic data contain some errors. The degree to which consistency must be satisfied depends on the purpose of the analysis. Processing observed data for consistency is crucial in an analysis where data interrelationships are important.

This paper proposes another step forward in using fuzzy logic optimization to obtain adjusted values. Two examples are given to present and explain the theoretical formulation and computational procedure. The proposed approach is robust enough to deal with other typical data discrepancies in transport situations. It preserves the integrity of observed data as much as possible, and allows the analyst to distinguish between reliable and less reliable data.

The approach is able to:

17

o Preserve the integrity of the observed data as much as possible. There are increasing concerns about data imputation and Base Data Integrity. The principle of Base Data Integrity is an important theme discussed by the American Society for Testing and Materials (ASTM, 1991) and the American Association of State Highway and Transportation Officials (AASHTO, 1992). The principle says that traffic measurements must be retained without modification and adjustment. Missing values should not be imputed in the base data. However, this does not prohibit imputing data at the analysis stage. In some cases, traffic counts with missing values could be the only data available for certain purposes and data imputation is necessary for further analysis. In accordance with the principle of Truth-in Data, AASHTO Guidelines (AASHTO, 1992) also recommends highway agencies to document the procedures for editing traffic data. For traffic counts with missing values, highway agencies usually either retake the counts or estimate the missing values. Estimating missing values is known as data imputation.

o Ensure flow consistency at any point in the network; the final estimate satisfies the law of flow conservation.

o Handle a large complicated network of any size and shape. The aim is to be able to solve any real problem, as shown in example 2.

o Handle data reliability; traffic-responsive control systems require reliable real-time information on the prevailing traffic counts to make sensible control decisions. This requisite is met by using the α parameter to define a different range for the membership function associated to each observed value.

o Limit the adjusted value within a tolerable deviation from the observed value, but allowing one tolerance for each value to be defined; this is achieved by using fuzzy logic and the definition of the α parameter.

    o Be solved in a short computation time. The triangular membership function allows solving the problem using mixed linear programming.

The method is flexible so that it can handle cases in which data are questionable, some of the observed values are known and fixed ($\alpha=0$), and there are considerable discrepancies in the observed data. The base of the membership function within which a feasible set of solutions is searched should be established according to the acceptable difference between adjusted and observed values.

Finally, the method is applicable to many other transportation problems in which consistency is important.

Acknowledgements

References

1. AASHTO Guidelines for Traffic Data Programs, 1992 American Association of State Highway and Transportation Officials.

2. American Society for Testing and Materials, ASTM, 1991. Standard Practice E1442, Highway Traffic Monitoring Standards, Philadelphia, PA.

3. Chambers, R., 2001. Evaluation Criteria for Statistical Editing and Imputation. National Statistics Methodological Series, pp. 28- 41.

4. Chakroborty, P., Kikuchi, S., 2003. Calibrating the membership functions of the fuzzy inference system: instantiated by car-following data. Transportation Research Part C 11(2), 91-119.

5. Dijker, T., Hoogendoorn, S.P., 1998. Perspective of Fuzzy Logic in Traffic Engineering: Reviews and Annotated Bibliography. Ministry of Transport, Public Works and Water Management, The Netherlands.

19

6. Geng, Y., Wu, X., 2008. The erroneous data imputation models for Beijing´s urban traffic flow data by time series and correlation analysis. Transportation Research Board 87th Annual Meeting, Washington, D.C., January 13-17, 2008, on CD-ROM.

7. Homaifar, A., McCormick, E., 1995. Simultaneous design of membership functions and rule sets for fuzzy controllers using genetic algorithms. IEEE Transactions on Fuzzy Systems 3 (1), 129-139.

8. Hoogendoorn, S.P., Copinga, G., Kaymak, U., 1998. Perspective of Fuzzy Logic in Traffic Engineering: Main Document. Ministry of Transport, Public Works and Water Management, The Netherlands.

9. Kaczmarek, M., 2005. Fuzzy group model of traffic flow in street networks. Transportation Research Part C 13(2), 93-105.

10. Kikuchi, S., Miljkovic, D., 1999. Method to Preprocess Observed Traffic Data for Consistency: Application of Fuzzy Optimization Concept. Transportation Research Record 1679, pp. 73-80.

11. Kikuchi, S., Miljkovic, D., Van Zuylen, H.J., 2000. Examination of Methods that Adjust Observed Traffic Volumes on a Network. Transportation Research Record 1717, pp. 109-119.

12. Kwon, J., Petty, K., Shieh, E., Kopelias, P., Papandreou, K., 2008. An automatic Method for Imputing and Balancing Link Traffic Counts. Transportation Research Board 87th Annual Meeting, Washington, D.C., January 13-17, 2008, on CD-ROM.

13. Laaksonen, S., 1999. How to find the best Imputation Technique? Test with various methods. International Conference on Survey Nonresponse. Portland, Oregon, pp. 28-31 October 1999.

14. Lee, S., Krammes, R.A., Yen, J., 1998. Fuzzy Logic-based incident detection for signalized diamond interchanges. Transportation Research Part C 6(5-6), 359-377.

15. Marzano, V., Papola, A., Simonelli, F., 2008. Investigating the effectiveness of the o-d matrix correction procedure using traffic counts. Transportation Research Board 87th Annual Meeting, Washington, D.C., January 13-17, 2008, on CD-ROM.

16. Pappis, C., Mamdani, W., 1977. A fuzzy logic controller for a traffic junction. IEEE Transactions on Systems, Man, and Cybernetics SMC-7, pp. 707-717.

17. Pentrice, G., 1987. Problems of Present Data Collection and Analysis. Proc. Conference of the Institution of Civil Engineers, Institution of Civil Engineers, London.

18. Rudy, K., Wang, H., Ni, D., 2008. Modeling and Optimization of Link Traffic Flow. Transportation Research Board 87th Annual Meeting, Washington, D.C., January 13-17, 2008, on CD-ROM.

19. Sharma, S.C., Kilburn, P., Wu, Y.Q., 1996. The precision of AADT volumes estimates from seasonal traffic counts. Alberta example. Canadian Journal of Civil Engineering 23 (1), 302-304.

20. Silva-Ramírez, E.L., 2007. Redes de Neuronas Artificiales en la edicion e imputacion de datos. Ph.D Thesis. University of Sevilla, (in Spanish).

21. Teodorovic, D., Vukadinovic, K., 1998. Traffic Control and Transport Planning: A Fuzzy Sets and Neural Network Approach. Kluwer Academic Publishers, Boston.

22. Tussel, F., 2002. Neural Networks and Predictive Matching for Flexible Imputation. DataClean 2002 Conference. Jyväskylä (Finland), pp. 29-31.

23. Tzeu-Chen Han, 2008. Application of Fuzzy Regression on Air Cargo Volume Forecast. Transportation Research Board 87th Annual Meeting, Washington, D.C., January 13-17, 2008, on CD-ROM.

24. Wang, L.-X., Mendel, J., 1992a. Back propagation of fuzzy systems as nonlinear dynamic system identifiers. Proceedings IEEE International Conference on Fuzzy Systems, San Diego, pp. 807-813.

25. Wang, L.-X., Mendel, J., 1992b. Fuzzy basis functions, universal approximation, and orthogonal least squares learning. IEEE Transactions on Neural Networks 3, pp.807-813.

26. Wang, L.-X., Mendel, J., 1992c. Generating fuzzy rules by learning from examples. IEEE Transactions on systems, Man and Cybernetics 22, pp. 1414-1427.

27. Zhong, M., Lingras, P., Sharma, S., 2004. Estimation of missing counts using factor, genetic, neural, and regression techniques. Transportation Research Part C 12(2), 139-166.

28. Zimmermann, H. J., 2001. Fuzzy Set Theory and Its Applications, 4th ed. Springer, Berlin.

Figures legends

Fig. 1. Simple freeway network used to explain the method

Fig. 2. Triangular membership function

Fig. 3. Real intersections in Andalusia's freeway network (South of Spain)

Fig. 4. Movements in every node of the real network

Fig. 5. Missing values' membership function

Tables legends

Table 1. Example 1: base data, randomized inconsistent data, adjusted data, and results

for different α ranges

Table 2. Real intersection in the South of Spain: real base data with missing values,

adjusted data, and results for different α ranges

Figure

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

Fig. 1. Simple freeway network used to explain the method

31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
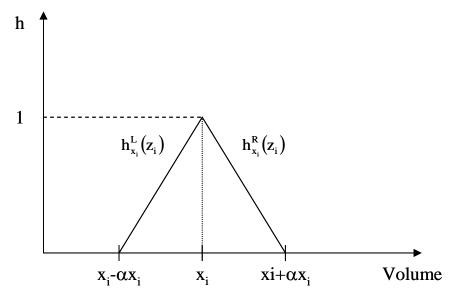62
63
64
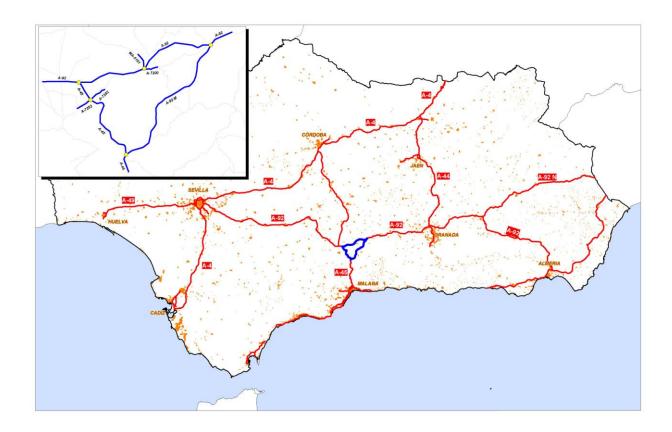65

Fig. 2. Triangular membership function

Fig. 3. Real intersections in Andalusia's freeway network (South of Spain)
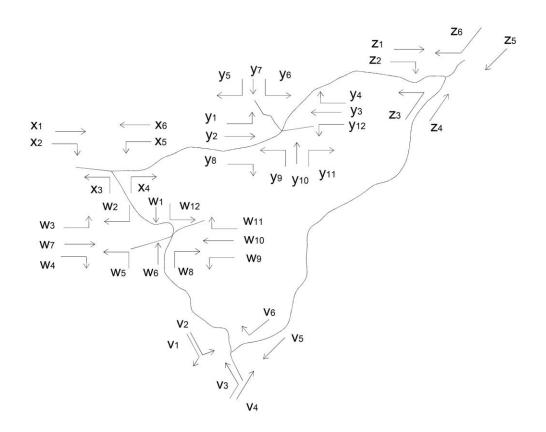
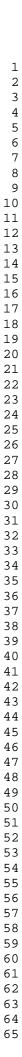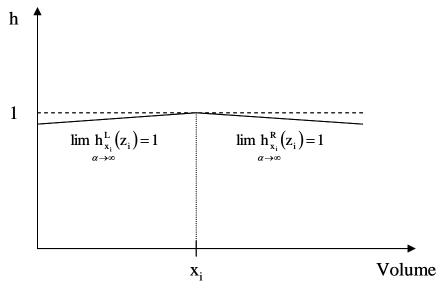Fig. 4. Movements in every node of the real network

Fig. 5. Missing values' membership function

(1)	(2) RV	(3) OV	(4) MM method α=0.4			(5) MM method α*			(6) MS method α=0.4			(7) MS method α*			(8) BO method α=0.4			(9) BO method α*		
			AV	h_i	Δ	AV	h_i	Δ	AV	h_i	Δ	AV	h_i	Δ	AV	h_i	Δ	AV	h_i	Δ
w1	135	113	128	0.68	7	126	0.63	9	113	1.00	22	113	1.00	22	128	0.68	7	126	0.63	9
w2	30	37	39	0.87	9	40	0.91	10	37	1.00	7	37	1.00	7	37	1.00	7	37	1.00	7
w3	43	42	42	1.00	1	44	0.92	1	42	1.00	1	42	1.00	1	42	1.00	1	42	1.00	1
w4	104	96	95	0.97	9	97	0.97	7	96	1.00	8	96	1.00	8	95	0.97	9	97	0.97	7
w5	148	134	152	0.67	4	150	0.61	2	134	1.00	14	134	1.00	14	152	0.67	4	150	0.61	2
w6	19	18	21	0.59	2	20	0.82	1	18	1.00	1	18	1.00	1	18	1.00	1	18	1.00	1
w7	28	27	30	0.73	2	30	0.64	2	27	1.00	1	27	1.00	1	30	0.73	2	30	0.64	2
w8	35	37	40	0.80	5	40	0.91	5	37	1.00	2	37	1.00	2	37	1.00	2	37	1.00	2
w9	22	18	21	0.59	1	20	0.82	2	18	1.00	4	18	1.00	4	18	1.00	4	18	1.00	4
w10	102	78	77	0.97	25	79	0.96	23	78	1.00	24	78	1.00	24	77	0.97	25	79	0.96	23
w11	175	171	172	0.99	3	169	0.96	6	171	1.00	4	171	1.00	4	172	0.99	3	170	0.98	5
w12	3	4	4	1.00	1	4	1.00	1	4	1.00	1	4	1.00	1	4	1.00	1	4	1.00	1
x1	265	215	253	0.57	12	255	0.71	10	220	0.94	45	220	0.96	45	253	0.57	12	253	0.73	12
x2	54	53	62	0.59	8	61	0.77	7	53	1.00	1	53	1.00	1	62	0.59	8	61	0.77	7
x3	105	116	109	0.85	4	111	0.93	6	116	1.00	11	116	1.00	11	109	0.85	4	111	0.93	6
x4	110	132	130	0.96	20	133	0.99	23	132	1.00	22	132	1.00	22	130	0.96	20	133	0.99	23
x5	200	177	168	0.88	32	168	0.92	32	161	0.78	39	161	0.86	39	168	0.88	32	168	0.92	32
x6	58	51	48	0.86	10	52	0.97	6	51	1.00	7	51	1.00	7	50	0.95	8	51	1.00	7
y1	26	31	26	0.61	0	28	0.79	2	20	0.13	6	20	0.25	6	26	0.61	0	26	0.66	0
y2	20	17	15	0.71	5	14	0.62	6	17	1.00	3	17	1.00	3	15	0.71	5	15	0.75	5
y3	18	21	21	1.00	3	18	0.70	0	21	1.00	3	21	1.00	3	21	1.00	3	18	0.70	0
y4	293	353	289	0.56	4	288	0.61	5	253	0.31	40	253	0.40	40	289	0.56	4	288	0.61	5
y5	45	39	39	1.00	6	41	0.89	4	41	0.87	4	41	0.89	4	39	1.00	6	41	0.89	4
y6	260	226	238	0.87	22	238	0.89	22	236	0.89	24	236	0.91	24	238	0.87	22	238	0.89	22
z1	33	26	29	0.72	4	29	0.75	4	26	1.00	7	26	1.00	7	29	0.72	4	29	0.75	4
z2	22	17	20	0.57	2	20	0.87	2	17	1.00	5	17	1.00	5	17	1.00	5	17	1.00	5
z3	25	27	29	0.82	4	31	0.92	6	27	1.00	2	27	1.00	2	27	1.00	2	27	1.00	2
z4	13	11	12	0.78	1	13	0.61	0	11	1.00	2	11	1.00	2	12	0.78	1	12	0.81	1
z5	28	33	32	0.93	4	31	0.87	3	33	1.00	5	33	1.00	5	32	0.93	4	31	0.87	3
z6	35	29	28	0.92	7	28	0.93	7	29	1.00	6	29	1.00	6	28	0.92	7	28	0.93	7
g=sum h_i			24.04			24.92			27.93			28.26			25.89			25.98		
h=min h_i			0.56			0.61			0.13			0.25			0.56			0.61		
Average Δ			7.23			7.13			10.70			10.70			7.10			6.97		

Note: RV (Real Value); OV (Observed Value); AV (Adjusted Value); Δ (difference between RV and AV in absolute value)

* α=0.65 for x_i; α=0.5 for y_i and z_i; α=0.3 for w_i

Table 1. Example 1: base data, randomized inconsistent data, adjusted data, and results for different α ranges

(1)	(2)	(3)		(4)		(5)		(6)		(7)		(8)	
		MM method α=0.1		MM method α*		MS method α=0.1		MS method α*		BO method α=0.1		BO method α*	
	OV	AV	h_i	AV	h_i	AV	h_i	AV	h_i	AV	h_i	AV	h_i
v1[#]	-	11091	1.00	10819	1.00	10703	1.00	10704	1.00	10951	1.00	10893	1.00
v2[#]	-	1764	1.00	1758	1.00	1743	1.00	1743	1.00	1497	1.00	1555	1.00
v3[#]	-	11085	1.00	11014	1.00	11240	1.00	11240	1.00	10994	1.00	10847	1.00
v4[#]	-	10288	1.00	10307	1.00	10329	1.00	10329	1.00	10575	1.00	10517	1.00
v5	10865	10727	0.87	10809	0.95	10865	1.00	10865	1.00	10618	0.77	10865	1.00
v6	1207	1174	0.73	1201	0.95	1207	1.00	1207	1.00	1207	1.00	1207	1.00
w1	5427	5669	0.56	5491	0.94	5427	1.00	5427	1.00	5445	0.97	5445	0.98
w2	2714	2905	0.30	2719	0.99	2714	1.00	2714	1.00	2714	1.00	2714	1.00
w3	3135	2914	0.30	3023	0.82	3135	1.00	3135	1.00	3135	1.00	3135	1.00
w4	3123	3258	0.57	3179	0.91	3123	1.00	3124	1.00	3123	1.00	3123	1.00
w5	3735	3764	0.92	3773	0.95	3735	1.00	3735	1.00	3735	1.00	3735	1.00
w6	5601	5313	0.49	5311	0.74	5600	1.00	5600	1.00	5354	0.56	5207	0.65
w7	695	744	0.30	695	1.00	695	1.00	695	1.00	695	1.00	695	1.00
w8	3112	3182	0.78	3131	0.97	3112	1.00	3112	1.00	3112	1.00	3112	1.00
w9	3880	3928	0.88	3907	0.97	3896	0.96	3896	0.98	3880	1.00	3880	1.00
w10	505	470	0.31	502	0.97	505	1.00	505	1.00	505	1.00	505	1.00
w11	310	289	0.32	307	0.95	310	1.00	310	1.00	310	1.00	310	1.00
w12	904	841	0.30	913	0.97	904	1.00	904	1.00	904	1.00	904	1.00
x1	4935	4588	0.30	4711	0.54	4812	0.75	4812	0.75	4800	0.73	4709	0.54
x2	4725	5021	0.38	4788	0.87	4848	0.74	4848	0.74	4719	0.99	4715	0.98
x3	7236	6739	0.31	6905	0.54	7236	1.00	7236	1.00	6990	0.66	6905	0.54
x4	1809	1777	0.82	1736	0.60	1809	1.00	1809	1.00	1809	1.00	1747	0.66
x5	4197	4394	0.53	4335	0.67	4197	1.00	4197	1.00	4344	0.65	4348	0.64
x6	3350	3306	0.87	3197	0.54	3350	1.00	3350	1.00	3351	1.00	3197	0.54
y1	1230	1176	0.56	1236	0.97	1230	1.00	1230	1.00	1230	1.00	1230	1.00
y2	3700	3662	0.90	3717	0.95	3700	1.00	3700	1.00	3700	1.00	3700	1.00
y3	4255	4555	0.29	4307	0.88	4257	1.00	4257	1.00	4555	0.29	4255	1.00
y4	1410	1509	0.30	1441	0.78	1410	1.00	1410	1.00	1509	0.30	1410	1.00
y5	2140	2076	0.70	2094	0.79	2140	1.00	2140	1.00	1990	0.30	2140	1.00
y6[#]	-	2320	1.00	2332	1.00	2369	1.00	2369	1.00	2369	1.00	2369	1.00
y7[#]	-	658	1.00	611	1.00	521	1.00	521	1.00	671	1.00	521	1.00
y8[#]	-	1527	1.00	1494	1.00	1691	1.00	1691	1.00	1679	1.00	1526	1.00
y9	1150	1069	0.30	1131	0.83	1150	1.00	1150	1.00	1150	1.00	1150	1.00
y10	310	289	0.32	324	0.65	310	1.00	310	1.00	310	1.00	310	1.00
y11	1013	1044	0.69	1023	0.91	1013	1.00	1013	1.00	1013	1.00	1013	1.00
y12	1410	1509	0.30	1442	0.77	1410	1.00	1410	1.00	1509	0.30	1410	1.00
z1	4960	4975	0.97	4968	0.99	4962	1.00	4962	1.00	4962	1.00	4962	1.00
z2	2120	2051	0.67	2104	0.97	2120	1.00	2120	1.00	2120	1.00	2120	1.00
z3	1207	1122	0.30	1092	0.68	1207	1.00	1207	1.00	1122	0.30	1087	0.67
z4	10865	10930	0.94	10973	0.97	10865	1.00	10865	1.00	10950	0.92	10985	0.96
z5	9660	9850	0.80	9906	0.92	9952	0.70	9952	0.90	9705	0.95	9952	0.90
z6	6940	6451	0.30	6098	0.60	5870	0.00	5870	0.49	6451	0.30	5988	0.54
g=sum h_i		26.16		36.50		40.14		40.85		35.97		38.61	
h=min h_i		0.29		0.54		0.00		0.49		0.29		0.54	

Note: OV (Observed Value); AV (Adjusted Value)

* α=0.2 for w_i; α=0.3 for z_i; α=0.1 for rest of cases

[#] v_1, v_2, v_3, v_4, y_6, y_7 and y_8 are missing values

Table 2. Real intersection in the South of Spain: real base data with missing values, adjusted data, and results for different α ranges