Fuzzy Sets and Systems 116 (2000) 3–9
www.elsevier.com/locate/fss

A method to defuzzify the fuzzy number: transportation problem application

Shinya Kikuchi *

Department of Civil and Environmental Engineering, University of Delaware, Newark, DE 19716, USA

Received December 1998

Abstract

In many problems of transportation engineering and planning, we encounter situations, in which the observed or derived values of the variables are approximate, yet the variables themselves must satisfy a set of rigid relationships dictated by physical principle. When the observed values do not satisfy the relationships, each value is adjusted until they satisfy the relationship. We propose a simple adjustment method that finds the most appropriate set of crisp numbers. The method assumes that each observed value is an approximate number (or a fuzzy number) and the true value is found in the support of the membership function. For each of many possible sets of values that satisfy the relationships, the lowest membership grade is checked and the set whose lowest membership grade is the highest is chosen as the best set of values for the problem. This process is performed using the fuzzy linear programming method. The paper presents the model, the computational process and applications. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Traffic volume adjustment; Data consistency; Defuzzification; Fuzzy LP

1. Introduction

In many transportation problems we encounter a situation in which the observed set of data needs to be adjusted so that they satisfy one or more relationships that must underlie among them. Consider the following case. The observed traffic volumes on the approaches of a complicated intersection need to be adjusted because they do not satisfy the flow conservation relationship (the total “in” equals to the total “out”). Such an adjustment is normally done by trial and error. A similar problem arises in the process of travel demand forecasting. For example, the trip production and trip attraction of each zone are derived from different models, yet by definition, the sum of the trip productions must equal to that of trip attractions. However, in reality, these sums do not match; hence, they are adjusted by an arbitrary manner.

This paper proposes a method that adjusts the observed (or given) values so that they meet the relationships that underlies among them. The method is suited when information about the original values is minimal and when the values are considered approximate. Given a set of observed values and the basic relationships that they must satisfy, the method yields a set of adjusted numbers that are close to the original numbers and also meets the relationship.

* Tel.: +1-302-831-2657; fax: +1-302-831-3640.
E-mail address: kikuchi@ce.udel.edu (S. Kikuchi).

PII: S0165-0114(99)00033-0

The method assumes that the observed values are fuzzy numbers in the face of no other information, and it employs the fuzzy linear programming approach. The objective is to find the set of values such that the smallest membership grade among them is maximized. This paper presents the approach and the LP formulation, and also, example problems with realistic application in transportation engineering.

2. The approach

Consider the case shown in Fig. 1. At a merging point of a freeway, traffic volumes are observed at cross-sections, A, B and C. The observed values are $A(=100)$, $B(=150)$, and $C(=300)$.These values are considered approximate due to the inaccuracy of the traffic counters. The law of flow conservation is not met by the values observed, i.e. $A+B\neq C$. Let x, y, and z be a set of values that satisfies the relationship and also close to the original observed values. The question is how to find the values of x, y, and z.

The proposed approach is the following:

Consider the observed values (or the values to be adjusted) as fuzzy numbers and assign a membership function to each value. The highest point of the membership function can correspond to the observed value. Based on the knowledge about the accuracy of the measurement, the support of each membership function is assumed. The membership functions assigned are triangular membership functions but need not to be symmetric. Let the membership function for each value as $h_A(x)$, $h_B(y)$ and $h_C(z)$, where A,B,C are the fuzzy set of approximate numbers. Also, let x, y and z be the unknown numbers, but each of them satisfies the relationship.

Formulate the following linear programming model:

Unknowns: x, y, z and h
Objective: Max h
Constraints: The relationship, $x+y=z$
$h\leqslant h_A(x),\ \ h\leqslant h_B(y),\ \ h\leqslant h_C(z),$
$x, y, z, h\geqslant 0,$

where h is the minimum degree of membership that one of the values of x, y or z takes. In the LP model, the value of h is maximized. In other words, for the derived h, h^*, the membership values of x, y, and z are at least greater than h^*:

$$h^* = \text{Max Min}[h_A(x), h_B(y), h_C(z)].$$

The LP approach above follows the concept of fuzzy linear programming originally described by Zimmermann [2]. This concept subscribes to the Bellman–Zadeh principle [1] so that the best solution set is found in the confluence of the goal and the constraint sets at which the degree of satisfaction for the worst case is maximized.

The meaning of the solutions is illustrated in Fig. 1. The three triangles represent the membership functions of A, B, and C. The solution for each variable lies on the support of the membership function. Among all the combinations of the values for x, y and z that satisfy $x+y=z$, the one that maximizes the minimum membership grade (h), $\text{Max}\{h_A(x), h_B(y), h_C(z)\}$, is chosen. If the solution of the LP model cannot be obtained, then it means that the values of x, y and z cannot be found within the assumed support of the membership functions. In this case, the membership function is modified by widening the support to find the values of x, y, z or consider that the observed values are not valid.

Generally, given a set of approximate numbers, $X_1, X_2, \ldots, X_n$, this method is suited for the following situations. Let $x_1, x_2, \ldots, x_n$, be the corresponding crisp numbers.

1. Meeting a target value: $x_1+x_2+x_3+\cdots=Z$ (Z is a predetermined value, either a fuzzy or crisp number).
2. Balancing of set of numbers: $x_1+x_2+x_3\cdots = x_{10}+x_{11}+x_{12}\ldots$.
3. Consistency in the system: $x_1+x_2+\cdots=Z_1$ and $x_{10}+x_{12}+x_{13}\cdots=Z_2$ (additional equations of this type) (Z_1 and Z_2 are predetermined values, either crisp or fuzzy numbers).

3. Formulation of the fuzzy LP model

In Fig. 1, let the left- and right-hand sides of each triangular-shaped membership function $h_{A-}(x)$ and $h_{A+}(x)$, $h_{B-}(y)$ and $h_{B+}(y)$, and $h_{C-}(z)$ and $h_{C+}(z)$, respectively. The triangular membership functions of A, B and C are characterized by $A=(\text{LA},\text{MA},\text{HA})$, $B=(\text{LB},\text{MB},\text{HB})$ and $C=(\text{LC},\text{MC},\text{HC})$, respectively, where (LA, LB, LC) are the left boundaries of the supports; (MA, MB, MC) are the center

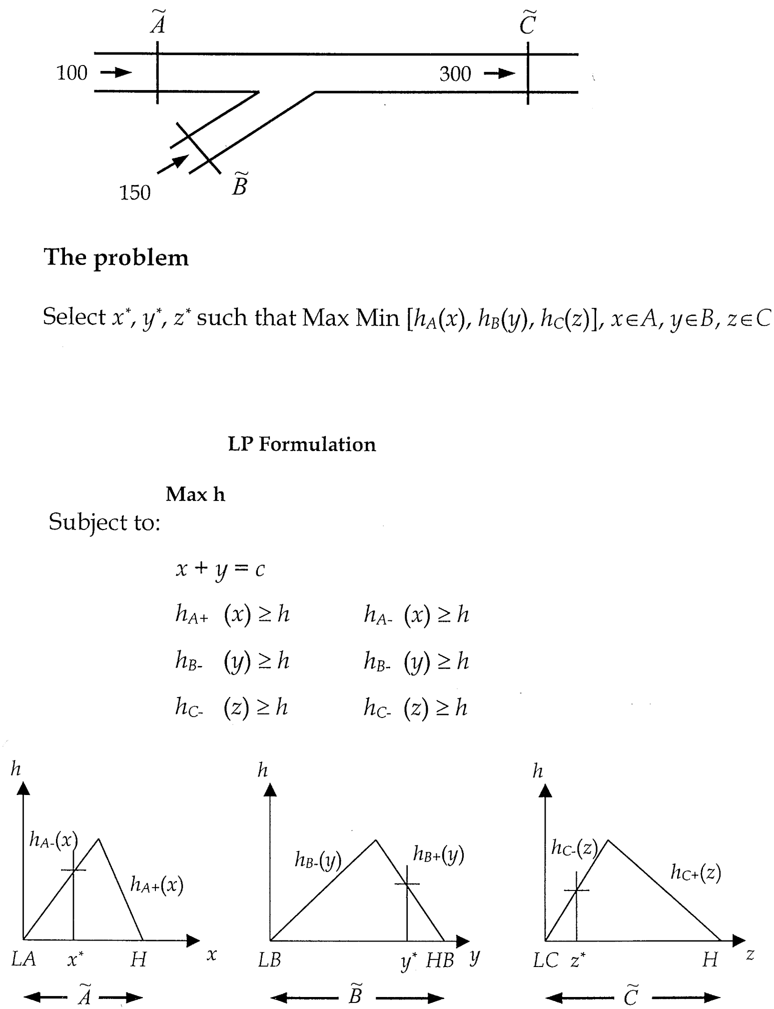


Fig. 1. Problem formulation.

values of the supports; and (HA, HB, HC) are the right boundaries of the supports.

The formulation proposed is formalized in the fuzzy linear programming form

$$
\begin{aligned}
&\text{Maximize} && h \\
&\text{subject to:} && x + y = z, \\
& && h_{A-}(x) \geqslant h, \quad h_{A+}(x) \geqslant h, \\
& && h_{B-}(y) \geqslant h, \quad h_{B+}(y) \geqslant h, \\
& && h_{C-}(z) \geqslant h, \quad h_{C+}(z) \geqslant h, \\
& && x, y, z, h \geqslant 0.
\end{aligned}
$$

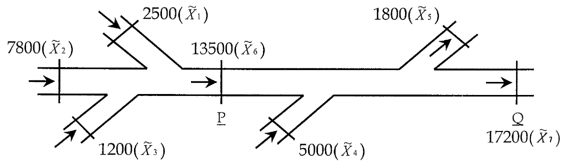
As seen above two constraints (left- and right-hand sides of the triangular membership function) need to be defined for each observed value.

4. Application examples

Two examples are presented. Example 1 is an expanded case of Fig. 1. Example 2 is the case of matrix balancing in which more than one relationship must be satisfied by each variable.

4.1. Example 1: Traffic volume consistency

Let us consider the case shown in Fig. 2, which is slightly more complicated than the one shown in Fig. 1. The number of observed values (traffic volumes) is 7. For each value, a fuzzified range is assumed with the low and high values of the range



$X_1 = (2000, 2500, 3000)$ $X_2 = (6240, 7800, 9360)$

$X_3 = (960, 1200, 1440)$ $X_4 = (4000, 5000, 6000)$

$X_5 = (1440, 1800, 2160)$ $X_6 = (10800, 13500, 16200)$

$X_7 = (13760, 17200, 20640)$

(A, B, C): A: lower bound; B: observed volume; C: upper bound

Fig. 2. Data and supports of the membership functions of Example 1.

as minus and plus 20% of the value. The three values in the parentheses in the figure are the (low, mid and high) values of the support and designated as $(\mathrm{LX}_i, \mathrm{MX}_i, \mathrm{HX}_i)$ for each X_i.

Let $X_1, X_2, \ldots, X_7$, be the obtained values in fuzzy numbers, and let $x_1, x_2, \ldots, x_7$, the unknown values which satisfy the following relationship. The conservation of flow relationship must be satisfied at points, P and Q; therefore, the relationships are

$$\text{at } P: \quad x_1 + x_2 + x_3 = x_6$$

$$\text{at } Q: \quad x_6 + x_4 - x_5 = x_7$$

The LP formulation is as follows:

$$\begin{aligned}
&\text{Max } h\\
&\text{at } P: \quad x_1 + x_2 + x_3 = x_6,\\
&\text{at } Q: \quad x_6 + x_4 - x_5 = x_7,\\
&\text{for } X_1: \quad h_{X_1-}(x_1) \geqslant h, \qquad h_{X_1+}(x_1) \geqslant h,\\
&\text{for } X_2: \quad h_{X_1-}(x_2) \geqslant h, \quad h_{X_1+}(x_2) \geqslant h,\\
&\text{for } X_3: \quad h_{X_1-}(x_3) \geqslant h, \quad h_{X_1+}(x_3) \geqslant h,\\
&\text{for } X_4: \quad h_{X_1-}(x_4) \geqslant h, \quad h_{X_1+}(x_4) \geqslant h,\\
&\text{for } X_5: \quad h_{X_1-}(x_5) \geqslant h, \quad h_{X_1+}(x_5) \geqslant h,\\
&\text{for } X_6: \quad h_{X_1-}(x_6) \geqslant h, \quad h_{X_1+}(x_6) \geqslant h,\\
&\text{for } X_7: \quad h_{X_1-}(x_7) \geqslant h, \quad h_{X_1+}(x_7) \geqslant h,
\end{aligned}$$

where $h_{X_i-}(x_i)$ and $h_{X_i+}(x_i)$ represent the left- and right-hand side of the triangular membership function, respectively. Thus, the expressions for the left- and the right-hand side lines are:

$$h_{X_i-}(x_i) = \{1/(\mathrm{MX}_i - \mathrm{LX}_i)\}(x_i - \mathrm{MX}_i) + 1,$$
$$h_{X_i+}(x_i) = \{1/(\mathrm{HX}_i - \mathrm{MX}_i)\}(x_i - \mathrm{MX}_i) + 1.$$

The optimum values and the membership functions are shown in Fig. 3. The corresponding value of h is 0.648. This indicates that the smallest membership grade among the values is 0.648. In other words, all the values satisfy the membership grade of at least 0.648.

4.2. Example 2: matrix balancing

Let us consider the case shown in Fig. 4 in which a passenger trip *O–D* table is to be constructed for a transit line with n stations. At each station the total number of boarding and alighting passengers are recorded. The recorded number of trips from station i to j is X_{ij}, total number of boardings at station i is A_i, and the total number of alightings at station j is B_j. These values are considered approximate due to possible inaccuracy of the counting device or human error. However, the values must satisfy the flow conservation laws: (1) the sum of the right-hand column total must equal to that of the bottom-row total, (2) for each row, the row total must equal the value in the right-hand column; and (3) for each column, the column total must equal the bottom row value. These relationships are expressed, respectively as

$$\sum_i a_i = \sum_j b_j, \qquad \sum_j x_{ij} = a_i, \qquad \sum_i x_{ij} = b_j,$$

where x_{ij}, a_i, b_j are the values to be derived for X_{ij}, A_i and B_j, the observed approximate values. The objective is to find the values of x_{ij} for all (i, j) and A_i's and B_j's in such a manner that they satisfy these basic relationships:

Let the triangular membership functions of X_{ij} and A_i and B_j defined based on the low, mid and high values be as follows

$$X_{ij} = (\mathrm{LX}_{ij}, \mathrm{MX}_{ij}, \mathrm{HX}_{ij}), \qquad A_i = (\mathrm{LA}_i, \mathrm{MA}_i, \mathrm{HA}_i),$$

$$B_j = (\mathrm{LB}_j, \mathrm{MB}_j, \mathrm{HB}_j).$$

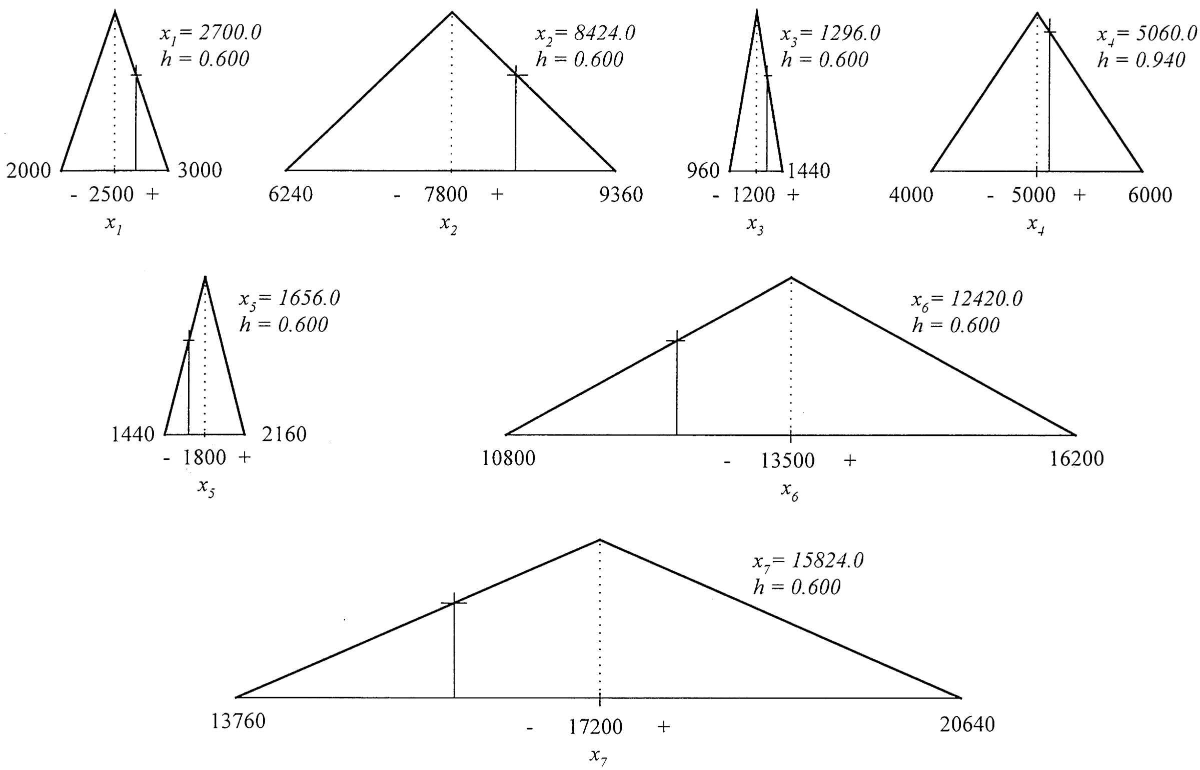


Fig. 3. The optimum values of x_i and the value of h for Example 1.

The fuzzy LP formulation is

Unknowns: x_{ij}, a_i, b_j, h,

Maximize h

Subject to $\sum_i A_i = \sum_j B_j, \quad \sum_j x_{ij} = A_i,$

$$\sum_i x_{ij} = B_j.$$

$h_{X_{ij}-}(x_{ij}) \geqslant h, \quad h_{X_{ij}+}(x_{ij}) \geqslant h \quad$ for all (i,j),

$h_{A_i-}(a_i) \geqslant h, \quad h_{A_i+}(a_i) \geqslant h \quad$ for all i,

$h_{B_j-}(b_j) \geqslant h, \quad h_{B_j+}(b_j) \geqslant h \quad$ for all j,

where

$$h_{X_{ij}-}(x_{ij}) = \{1/(\mathrm{MX}_{ij} - \mathrm{LX}_{ij})\}(x_{ij} - \mathrm{MX}_{ij}) + 1$$
for all (i,j),

$$h_{X_{ij}+}(x_{ij}) = \{1/(\mathrm{HX}_{ij} - \mathrm{MX}_{ij})\}(x_{ij} - \mathrm{MX}_{ij}) + 1$$
for all (i,j),

$$h_{A_i-}(a_i) = \{1/(\mathrm{MA}_i - \mathrm{LA}_i)\}(a_i - \mathrm{MA}_i) + 1$$
for all i,

$$h_{A_i+}(a_i) = \{1/(\mathrm{HA}_i - \mathrm{MA}_i)\}(a_i - \mathrm{MA}_i) + 1$$
for all i,

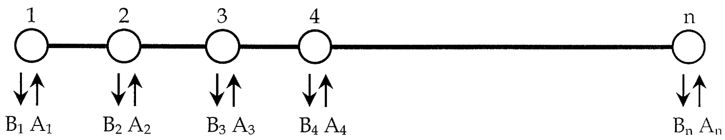
$$h_{B_j-}(b_j) = \{1/(\mathrm{MB}_j - \mathrm{LB}_j)\}(b_j - \mathrm{MB}_j) + 1$$
for all j,

$$h_{B_j+}(b_j) = \{1/(\mathrm{HB}_j - \mathrm{MB}_j)\}(b_j - \mathrm{MB}_j) + 1$$
for all j.

The solution of this formulation, x_{ij}'s, a_i, b_j, represents the set of values that balances the *O–D* matrix.

5. Discussion

In the examples shown above, the constraints can be modified so that they incorporate the problem specific conditions: one, some of the observed values are



X_{11}	X_{12}	X_{13}			A_1
X_{21}	X_{22}	X_{23}			A_2
X_{31}	X_{32}	X_{33}			A_3
X_{n1}	X_{n2}	X_{n3}			A_n

B_1	B_2	B_3		B_n

Observed values: A_i, B_j, X_{ij} *for all values of i, j and (i,j)*

Relationships to be satisfied:

$$\sum_j x_{ij} = a_i\,,\ \sum_i x_{ij} = b_j\,,\ \sum_i a_i = \sum_j b_j$$

Select x_{ij}^*, a_i^*, b_j^* such that

$$\text{Max Min } [h_{Ai}(a_i),\ h_{Bj}(b_j),\ h_{Xij}(x_{ij})] \text{ for all } a_i,\ b_j, \text{ and } x_{ij}$$

Fig. 4. Formulation of problem in Example 2.

crisp definite values; two, some of the observed values are known only in a crisp range. These two cases actually refer to the shape of the membership functions. The former is the case of singleton such that the accuracy of the value is known to be reliable. The latter is the case that information about the value is so little that any value is equally possible and thus, the rectangular shaped function is the appropriate shape of the membership function.

These cases are easily incorporated in the formulation. For the former case, the observed single value is used in the relationship; for example, in Example 2 above, some A_i's and B_j's can be crisp values. For the latter case, the range is expressed by the inequalities such as, $w_1 \leqslant x_{ij} \leqslant w_2$, in place of using the triangular membership function.

While the application of the proposed method relates to a number of adjustment problems in which the observed values are approximate, three typical problem environments can be considered:

(1) to match the results obtained by the microscopic analysis and those by the macroscopic analysis;
(2) to satisfy the principle of conservation of flow, consistency or balancing; and
(3) to find a set of compromised values.

An example of the first type is the case of matching of the demographic data obtained from two different sources: one based on the individual zone level and aggregated, and the other, based on the entire metropolitan area. In this case, the data on individual zones and total area adjusted simultaneously. The former is the microscopic approach and the latter is the macroscopic approach.

Examples of the second type of problem are the ones presented in this paper. In addition, the method can be applied to the schedule preparation process. For example, the times shown on a timetable of a bus route must be crisp; however, the estimated travel times of different segments of the route are known in approximate numbers due to congestion and other unforeseen factors. The total travel time on the route, on the other hand, is a value derived from the combination of the agency's desire, policy or scheduling of the drivers. Given these conditions the timetable must show crisp times satisfying the estimated times of the route segments and the total route travel time. If the solution to the LP method is not available, the timetable is not feasible given the data.

The problems of the third type include the case of adjusting the opinions of two conflicting parties. For example, when planning land use or space allocation, the space desired for individual functions are known approximately and the total land (or space) for allocation is given either approximate or crisp. The planner must determine the exact allocation based on the two requirements: individual's desires and the total resource limitation. Another example is the case of negotiating a contract that involves many tasks, when the total budget is known either exactly or approximately. The contractor proposes the prices for individual tasks. Each price has some room for negotiation, and the total of the prices must fit the budget. In this case, the prices of individual tasks can be determined while satisfying the desire of both parties. The value of h provides the degree of satisfaction. In this problem the power of fuzzy LP is utilized fully, since the desires of individual parties are best formalized by the fuzzy sets.

While the discussion so far has dealt with the case in which the information on the observed value is very little, and as a result, the values are treated as approximate or fuzzy. The proposed method can still be applicable when repeated observations are possible. Let us consider the example in Fig. 1 that many different machines count traffic volumes at A, B and C during the same time period, and the statistical distributions for them are available. Even in this case, the mean values of each distribution for A, B and C, may still not satisfy the flow conservation equation. Hence, the proposed method can be applied to find the values. In this case, the data can be treated in two ways.

One, compute the average of observed volumes for A, B, C, separately, and use them as the approximate input to the model. The membership function should incorporate the statistical distribution of each variable. Alternatively, each observed value for A, B and C are considered fuzzy and construct a membership function for each observed value and apply the method. In this case a membership function is associated with each observed value, the number of constraints in the LP model will increase proportionally to the number of observations.

6. Conclusion

This paper presents a method that can be used to adjust the values of the variables in many practical engineering problems. Normally, in transportation problems, the value of a variable is either the raw data or derived from a model. If a rigid relationship must exist among the variables, then the value of each variable needs to be adjusted to be consistent with the underlying relationship whatever it is. The proposed model provides a mechanism of adjustment, given little information about the nature of the observed values. The method uses the concept of fuzzy LP in such a way so as to find a set of values that respects the original observed values as much as possible; yet, at the same time, they satisfy the rigid relationship that underlies among the variables.

The structure of the model is simple and is flexible to incorporate specific characteristics of the observed values and the relationships encountered in transportation problems. Efforts should be made to expand the model so that it handles a fuzzy relationship among the variables as well as the rigid relationship as shown in this paper. Such a model will be useful for various general data handling problems in which the exact input and output values need to be determined given the approximate data and the approximate causal relationships between input and output.

References

[1] R.E. Bellman, L.A. Zadeh, Decision making in a fuzzy environment, Management Sci. 17(4) (1970) B-141–B-164.

[2] H.-J. Zimmermann, Fuzzy Set Theory and Its Applications, third ed., Kluwer Academic Publishers, Norwell, MA, 1996.