N93-29539

161823

EVALUATION OF FUZZY INFERENCE SYSTEMS USING FUZZY LEAST SQUARES

P. 10

Joseph M. Barone
Loki Software, Inc.
P.O. Box 71
Liberty Corner, NJ 07938   USA

SUMMARY

Efforts to develop evaluation methods for fuzzy inference systems which are not based on crisp, quantitative data or processes (i.e., where the phenomenon the system is built to describe or control is inherently fuzzy) are just beginning. This paper suggests that the method of fuzzy least squares can be used to perform such evaluations. Regressing the desired outputs onto the inferred outputs can provide both global and local measures of success. The global measures have some value in an absolute sense but are particularly useful when competing solutions (e.g., different numbers of rules, different fuzzy input partitions) are being compared. The local measure described here can be used to identify specific areas of poor fit where special measures (e.g. the use of emphatic or suppressive rules) can be applied. Several examples are discussed which illustrate the applicability of the method as an evaluation tool.

INTRODUCTION

Smith and Comer [1] point out that evaluation of the behavior of a fuzzy system can be quite difficult. They also mention (p. 20) that the qualitative knowledge of the controller designer is more suited to accurate specification of the antecedent portions of the control rules that to accurate specification of the consequent portions. This is because (presumably and at least in part) the role of the input variables in system dynamics is more easily understood in general, and also because the input variables are often more directly and more easily expressible in fuzzy (linguistic) terms (e.g., temperature as high, medium, and low). This is perhaps even more true in "softer" areas like psychology and sociology, where "harder" inputs like age and socioeconomic status are used to control (predict) softer outputs like behavior or risk (for interesting comments along these lines in the context of fuzzy classification see [2]). In fact, the very foundations of some methods of analysis and prediction used in these soft areas, especially classical least squares, are predicated upon input variables whose values are assumed to be error-free measurable (see e.g. [3], Section 1.1).

Methods for the evaluation and tuning of fuzzy systems do not really challenge this assumption; they typically assume that the designer has the input distributions about right and then adjust formal "parameters" of the inference mechanism to improve controller performance. Again, this works well in hard areas but should prove difficult to apply in emerging softer applications where there is no aspect of the inference process that can be trusted completely. It becomes important, therefore, in soft applications, to have some way of evaluating the accuracy and effectiveness of a fuzzy inference system which assumes as little as possible about the validity of the rules, and even of their essential characteristics, beyond the linguistic properties they express. Furthermore, there may often be no real way of knowing whether interpolated consequent fuzzy values (values not supplied directly by an expert) are accurate to the point where they can serve to confirm the chosen system and parameters. It should prove useful, therefore, to have available methods which can provide overall evaluation measures given certain assumptions about the structure and regularity of the output (consequent) fuzzy distributions.

Perhaps the most well-characterized and formalized methods for the evaluation and tuning of fuzzy controllers are those based on the concept of cell mapping [1, 4-5]. Nonetheless, the application of cell mapping to evaluation and tuning depends crucially on the existence of sufficient crisp input-output pairs to generate the cell maps (actually, this is a bit of an oversimplification - see [5], pp. 749-750), and also provides no real way to distinguish between competing fuzzifications of the input state space (unless of course the fuzzification is so bad that tuning is impossible). This paper suggests that an evaluation based on fuzzy least squares can indeed distinguish between competing input state space fuzzifications and can be used (quite easily) in cases where neither the input nor the output is readily defuzzifiable.

FUZZY LEAST SQUARES

The method of fuzzy least squares was introduced by Diamond [6] as an approach to the fuzzy regression problem, i.e. as a method for parameterizing the relationship between two sets of fuzzy numbers; its advantage over other techniques (besides computational simplicity), as Diamond points out, is the amenability of the parameterization to evaluation by standard measures, e.g., examination of residuals. For the purpose at hand, it is particularly important that the spatial geometry of the fuzzy least squares method be understood; to accomplish this goal, we turn briefly to crisp models.

Basically, the solutions to linear parameter estimation problems as well as their computational simplicity depend heavily on assumptions regarding which measurements may be considered to be error-free and which measurements may not. If either the independent or the dependent variable measurements are taken to be error-free, then ordinary least squares may reasonably be applied to the data. In such cases, the error (residual) vectors are orthogonal to the axis (or axes) along which the error-free values are measured. If, on the other hand, both dependent and independent variable measurements must be assumed to be made with error, the parameter estimation problem becomes considerably more difficult (even analytically intractable in the general case). In any case, if a solution can be generated, the error vectors will be orthogonal to the fitted line itself (the first chapter of [3] contains an excellent summary and relevant examples).

In extreme cases, especially those in which the data points are contaminated by outliers, the differences in the various solutions may be striking, as is illustrated in the figure below (from [7]). If the x coordinates are assumed to be error-free and a line is fitted by the method suggested in [7] (not ordinary least squares but equivalent for the present purpose), then errors orthogonal to the x axis are minimized by a fitted line which passes through the outlier (the point at 0,0). This is clearly a most undesirable solution. If both the x and y coordinates are assumed to contain errors, on the other hand, (even isotropic ones), the method yields a much more reasonable fitted line (the one parallel to the y axis).
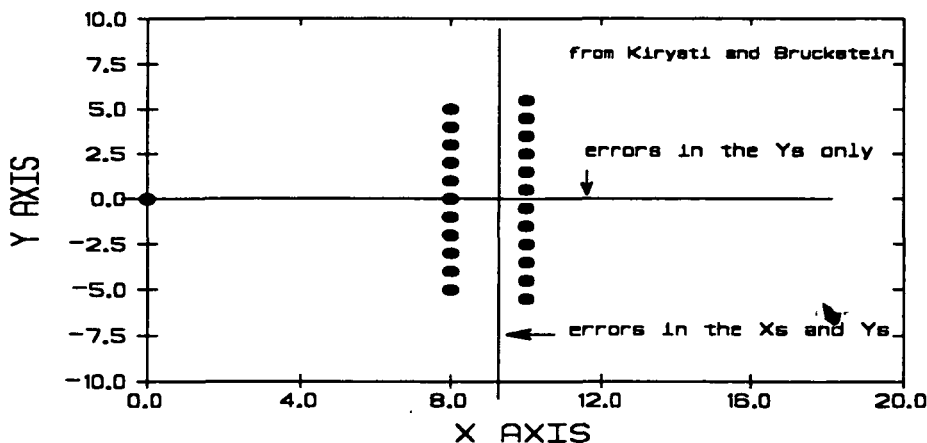
To return to fuzzy considerations, the point is that the method of fuzzy least squares, despite its "ordinary least squares" character, is more closely related (in spirit, as it were) to fitting (regression) approaches in which both dependent and variables are measured with error. It should be emphasized, however, that this is not true from a purely analytic point of view. Once a distance metric is decided upon, and once the hypergeometric characteristics of the set of triangular fuzzy numbers are established, the fuzzy least squares parameter vector is derived by an orthogonal projection of the dependent variable vector onto the "cone" of potential solutions exactly as in ordinary least squares (Diamond's paper [6], pp. 142-146 contains an elegant exposition of these facts, and section 2.3 of [3] contains highly instructive comments and diagrams in a crisp context). Thus, from an analytical point of view, though both the independent and dependent variable vectors are fuzzy, one is assumed to be measured without error while the other is not.

From another point of view, however, fuzzy least squares is more like a "total least squares" approach [3] in which both the dependent and independent vectors (or matrices) are assumed to be measured with errors. This is because the fuzzy least squares method, with its two separate spatial components (mode and spread), permits the search for a solution vector to move about a more complex (and hence more flexible) space (in effect, of course, since the solution is derived analytically). The result of this is that fuzzy least squares can preserve an extremely good fit in fuzziness even if, for some reason, one or more values in the data are outliers relative to mode. Since the fuzziness of the dependent and independent variables, taken together, are a measure of the overall uncertainty of the system, this characteristic has the effect of preserving the degree of overall uncertainty in a manner similar to total least squares methods.

FUZZY LEAST SQUARES AND FUZZY INFERENCE

It would surely be instructive to pursue the analogy between fuzzy least squares and total least squares further and more formally, but that would take us far beyond the scope of this paper. It is worth mentioning, though, by way of leaving the previous topic and beginning the current one, that Diamond's fuzzy least squares minimization condition (1) could conceivably be replaced to good effect by (2), where minimization of the square of the distances between the measured (Y_i) and calculated $(E + bX_i)$ is replaced by minimization of some scalar norm of the "total error" matrix $([.])$ and where X_0 is the unobservable "true" vector of fuzzy predictors (see [3], p.186 and p. 23). If the Frobenius norm were

$$\sum d (E + b X_i, Y_i)^2 \quad (1)$$

$$F (E, b, X_0) = [d (X, X_0); d (E + bX, Y)] \quad (2)$$

used, a solution to (2) would be equivalent to a solution of the "fuzzy total least squares" minimization function (cf. [3], p. 186).

Be all of this as it may, it seems fair to conclude that fuzzy least squares is a relatively "robust" form of regression which is eminently suitable for parameterizing the relationship between two n-dimensional fuzzy vectors with elements of regular shape (at least triangular and trapezoidal [6]). The vectors being compared do not necessarily have to be particularly "linear", though they must at least be "coherent" ([6], pp. 150-151); vectors produced as result of fuzzy inference are as likely to be coherent as not, one would imagine, but the condition is easily tested for [6], so inference systems which do not produce coherent output should simply not be subjected to the evaluation procedures suggested here.

Fuzzy least squares, then, forms the basis for a simple evaluation technique for fuzzy inference systems. Given two possible solutions, regress the known (fuzzy) output (the "correct" values) on the output fuzzy sets generated by the two inference processes. Compare the two solutions via any of many available evaluation methods, and keep the one which evaluates higher. Certain evaluation methods may even suggest ways in which the better solution can be improved. Space does not permit further general discussion, so we conclude by introducing a few evaluation measures and by providing examples of their use. It is worth noting at this point that the calculations needed to perform fuzzy least squares and to compute the evaluation measures are straightforward and can be performed with minimal computational overhead. It is also worth noting that it is may be possible to extend the domain of this method to inference systems which do not produce fuzzy "numerical" output by "fitting" fuzzy numbers over the fuzzy sets by linear interpolation as is done in fuzzy modeling (see, e.g. [8]), but this matter will not be pursued here.

EVALUATION MEASURES

1. GLOBAL MEASURES. The most obvious global measure of success are the least squares residuals. A related value which varies conveniently between 0 (no correlation) and 1 (perfect correlation) is the correlation coefficient. For generality, we define (see [9], p.280) the fuzzy multiple correlation coefficient (MCC) as

$$MCC = (\frac{\sum d(E + bX_i, \overline{Y})^2}{\sum d(Y_i, \overline{Y})^2})^{\frac{1}{2}} \quad (3)$$

where d again is the distance between two fuzzy numbers [6] and where Y is the mean dependent fuzzy value, even though all examples in this paper are univariate and extensions to the multivariate case are non-trivial ([6], p. 156).

Another useful global measure of success is the relative entropy of the fuzzy least squares solution as defined in [10]. This form of relative entropy is a measure of the success of the regression "line" in tracking the fuzziness of the elements of the dependent variable vector. It is defined as (see [10] for a detailed description and rationale):

$$h_\mu = -\left\{\sum_{i=1}^{n}\left[\mu_e(y_i)\ln(\mu_e(y_i)) + (1 - \mu_e(y_i))\ln(1 - \mu_e(y_i))\right]\right\} \quad (4)$$

$$\text{where } \mu_e(y_i) = \max(0.5, \min(\frac{spread(y_i)}{spread(\hat{y}_i)}, \frac{spread(\hat{y}_i)}{spread(y_i)}))$$

and where y_i is the estimated y_i (i.e., $E + bX_i$).

2. LOCAL MEASURES. The only local evaluation method discussed here will be the weighted squared standardized distance [11-12]. In the univariate case, the WSSD can be written as:

$$WSSD_i = \frac{(n - 1)b^2 \cdot d(x_i, \overline{X})^2}{\sum_{j=1}^{n}d(y_j, \overline{Y})^2} \quad (5)$$

$$\text{where } \overline{X} \text{ and } \overline{Y} \text{ are the X and Y means}$$

where b is the regression coefficient, and where d again is the fuzzy distance. In ordinary least squares regression, the magnitude of $WSSD_i$ is used to determine whether or not point i is a "high-leverage point", i.e., a point in a sparse region of the X-space (see, e.g., [12], pp. 94 ff.). We are interested here in the WSSD because a fuzzy inference tends to produce similar or identical output when the inference mechanism operates near the centers of the involved fuzzy sets and to produce rapidly changing output as the inference mechanism operates near areas of overlap (and thus near areas of heavy interpolation). A good inference mechanism should produce transitional areas in its output which correspond to areas of overlap in the output data partition. Thus, the output vector produced by a fuzzy inference should have clusters of similar or identical values which match the reference values near the centers of the elements of the output reference partition, and rapid changes in value which match the reference values in and near the overlap areas of the output reference partition. This phenomenon will produce clusters of points with similar or identical leverage in the regression followed by points with unique leverage values (at the transitional areas). In a good model, then, the clusters and transitions in WSSD values will line up nicely with the centers and overlap areas of the output reference partition respectively.

A NOTE ON "PIECEWISE" APPROXIMATIONS

It is important to note that this paper is not suggesting that fuzzy least squares is to be used to construct an accurate "piecewise" approximation to some unknown "functional relationship" between input and output fuzzy sets. To understand better what is being suggested, consider a fuzzy Lagrangian interpolation polynomial which relates the true output fuzzy sets and the ones generated by the inference (as in [13] with n+1 fuzzy points). As with crisp Lagrangian interpolation, such a polynomial could be used, for instance, to compute error bounds (using contour integrals in the complex plane [14]) if we knew the "true" functional relationship between the actual output fuzzy sets and the generated ones; such a relationship may not exist, of course, in the general case

and in the usual sense of the word functional, but would in any event depend on the accuracy of the inference process as an interpolator and "smoother". The least squares regression line, then, serves in this context as a crude continuous approximation to some presumably nonlinear and possibly unrecoverable "difference" function.
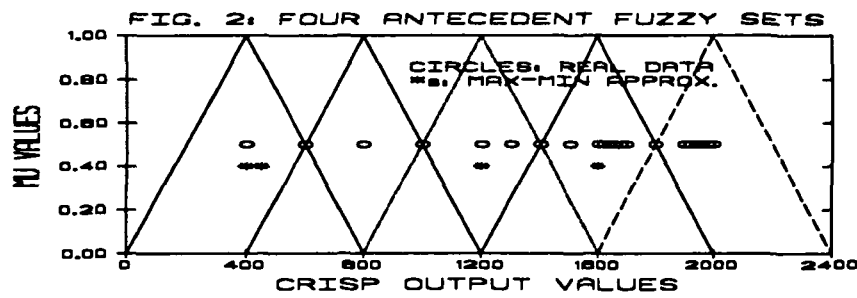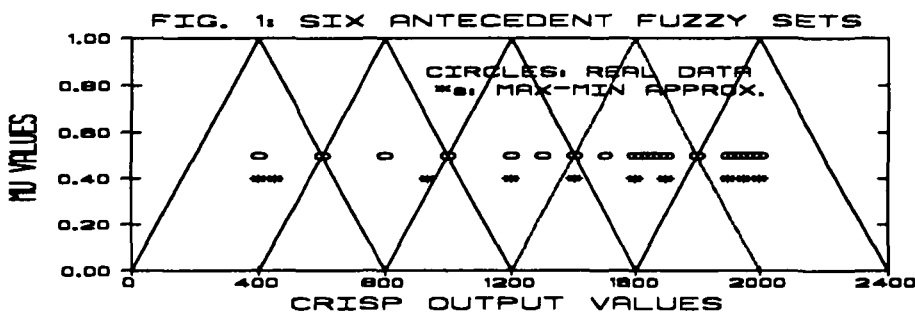
EXAMPLES

All of the examples discussed here are based on examples from Cao and Kandel [15]. Since their examples are based on crisp input and output (rotational speed of a d.c. series motor as a function of input current), the output was "refuzzified" as described below so that fuzzy regression could be applied. The data sets in our examples, therefore, are not "inherently" fuzzy, but they do have the advantage of being associated with thoroughly analyzed models which are easy to evaluate for accuracy (in a crisp sense). Note also that as was mentioned earlier, the notion of "reshaping" the output of a fuzzy inference process so that it can be analyzed by fuzzy least squares is not an unreasonable one, though of course for useful application it would require more elaborate methods than the one used here.
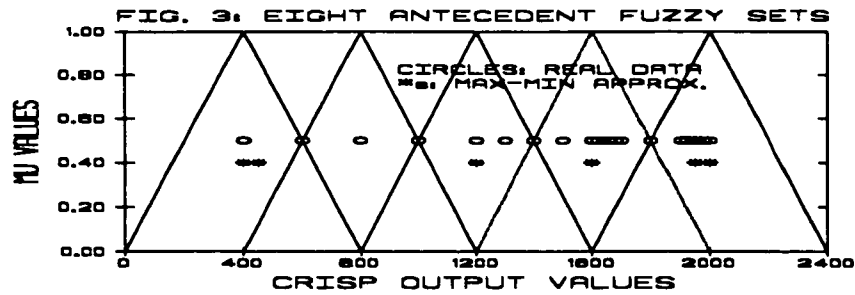
1. The "model" curve of Example 7 in [15] is a connected piecewise linear curve of five segments with overall rising trend. The model curve is "covered" by the five overlapping fuzzy sets shown in Figure 1. Assuming that this consequent set representation is reasonable, the fuzziest areas of coverage (i.e., the areas of maximal overlap) are those around the output values 800, 1400, and 1800 (800 because the rules do not reference the second set (SMALL)). Ideally, the inference system should map the corresponding input values (1.0, 3.0, and 7.0) into these same transition areas. Cao and Kandel cover the input range by six overlapping fuzzy sets; we use the WSSD, MCC, and relative entropy to compare their six input set partition with a four input set partition and an eight input set partition. The rules in the four and eight input set cases are adjusted to conform insofar as is possible with the content of Cao and Kandel's original (six input set) rules. The crisp output data and the crisp inferred output data were fuzzified by using 10%, 15%, or 20% of the mode as the left and right spreads, increasing the percentage as the numbers got larger; in this manner, a reasonably coherent output data set and inferred output data set were constructed. The rules themselves are as follows (in each case the input domain is distributed equally among the component sets):

NULL -> ZERO	NULL -> ZERO	NULL -> ZERO
ZERO-SMALL -> MEDIUM	ZERO -> MEDIUM	ZERO -> MEDIUM
SMALL-MEDIUM -> LARGE	SMALL -> LARGE	ZERO-SMALL -> MEDIUM
LARGE -> VERY LARGE	MEDIUM -> LARGE	SMALL -> LARGE
	LARGE -> VERY LARGE	SMALL-MEDIUM -> LARGE
	VERY LARGE -> VERY LARGE	MEDIUM-LARGE -> VERY LARGE
		LARGE -> VERY LARGE
		VERY-LARGE -> VERY LARGE

As Table 1 shows, good results were obtained when the fuzzified inferred values were regressed on the fuzzified output data (the table shows only the crisp values, i.e., the modes of the fuzzified values). The transition points match nicely, the MCC is high, and the relative entropy is low (of course, the MCC and entropy values are most meaningful when compared with other prospective solutions). When only four antecedent sets are used, however, the results suffer dramatically. The transition points miss the mark by a considerable margin, the MCC is lower, and the relative entropy is higher. With eight antecedent sets results are better but still not as good as with six (it is important to note here that overlap was retained at 50%). If one had started with the four or eight set inference machine, the lack of matchups in the transition areas would have been a clue that the results could be improved upon. It is worth noting that the relative magnitudes of the fuzzy constants are a decent guide to the relative merits of the various models. Figures 1, 2, and 3 show the distributions of the crisp output values relative to the output set and to the covering fuzzy sets (the fuzzy partition) for the consequent portions of the inference rules; note that only the six antecedent set solution produces distinct transition values in the vicinity of the transition regions of the output fuzzy partition, and that this fact is reflected in the WSSD values. For details of the membership functions, input and output data, and the rules themselves refer to [15].

	TABLE 1: RESULTS FOR EXAMPLE 7 OF CAO and KANDEL WITH MAX/MIN INFERENCE ASTERISKS MARK TRANSITION POINTS					
DATA	6 ANT. SETS	WSSD FOR 6	4 ANT. SETS	WSSD FOR 4	8 ANT. SETS	WSSD FOR 8
2000	2000.0	1.1587	1600	1.0106	1950	0.8653
1980	1950.0	0.9439	1600	1.0106	1950	0.8653
1960	1900.0	0.7513	1600	1.0106	2000	1.0614
1940	1950.0	0.9439	1600	1.0106	1950	0.8653
1920	2000.0	1.1587	1600	1.0106	1950	0.8653
1900	1950.0	0.9439	1600	1.0106	2000	1.0614
1800*	1700.0	0.2024*	1200	0.0129	1950	0.8653*
1700	1600.0	0.0608	1200	0.0129	1600	0.0569
1680	1600.0	0.0459	1200	0.0063	1600	0.0432
1660	1600.0	0.0459	1200	0.0063	1600	0.0432
1640	1600.0	0.0459	1200	0.0063	1600	0.0432
1620	1600.0	0.0459	1200	0.0063	1600	0.0432
1600	1600.0	0.0459	1200	0.0063	1600	0.0432
1500	1600.0	0.0459	1200	0.0063	1200	0.3541
1400*	1400.0	0.0446*	1200	0.0063	1200	0.3541
1300	1200.0	0.4275	1200	0.0241	1200	0.3847
1200	1200.0	0.4275	1200	0.0241	1200	0.3847
1000	1200.0	0.4275	450	2.6917	1200	0.3847
800*	937.50	1.4192*	450	2.6917	1200	0.3847
600	450.00	4.8495	400	3.0746	450	4.3971
400	400.00	5.3181	400	3.0746	400	4.8227
MCC		0.9837		0.9411		0.9537
R. ENT		3.7707		5.8724		4.4829
FOR 6 ANTECEDENT SETS Y = (20.4, 3.82, 3.82) + .99X						
FOR 4 ANTECEDENT SETS Y = (248.0, 40.50, 40.50) + 1.08X						
FOR 8 ANTECEDENT SETS Y = (90.6, 15.32, 15.32) + .95X						

FIG. 1: SIX ANTECEDENT FUZZY SETS

CIRCLES: REAL DATA
*s: MAX-MIN APPROX.

MU VALUES

CRISP OUTPUT VALUES

FIG. 2: FOUR ANTECEDENT FUZZY SETS

CIRCLES: REAL DATA
*s: MAX-MIN APPROX.

MU VALUES

CRISP OUTPUT VALUES
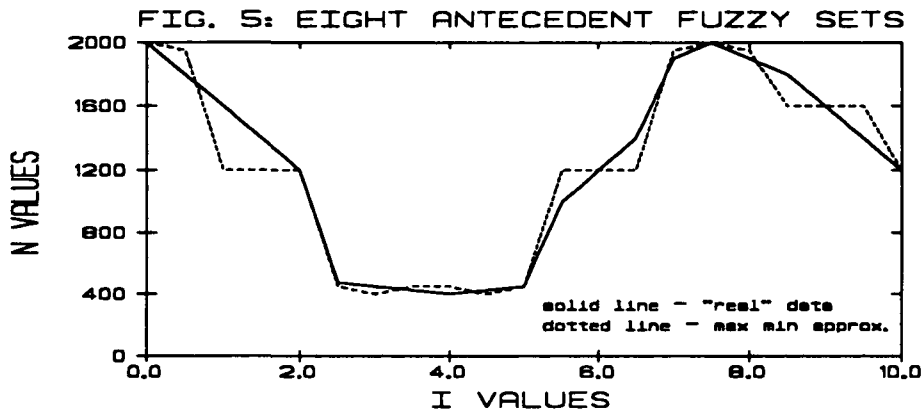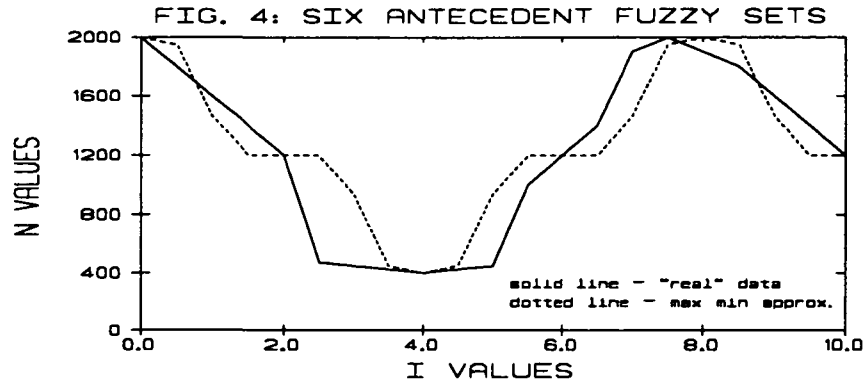
FIG. 3: EIGHT ANTECEDENT FUZZY SETS

2. The model curve of Example 3 in [15] is a piecewise linear curve with two trend shifts. For this example, we flattened the bottom and shifted the second peak left to conform with the output of an eight antecedent fuzzy set approximation. As can be seen from the results below (and as would be expected), the eight-antecedent model yields better statistics. Nevertheless, the six-antecedent model conforms better to the transition points (not shown, but fairly obvious from an inspection of Figures 4 and 5). This suggests that the flattened area might be better approximated by emphasizing the appropriate rule in the rule set [15] and retaining the six antecedent fuzzy sets (note that to do this it is necessary to switch from max-min to product-sum inference - see [16]). As can be seen from the third column of values in Table 2, this hypothesis proves correct - there is little difference between the eight-antecedent results and the six-antecedent results with emphasis, and the six-antecedent version is truer through the transitions. If the second peak is shifted back to its original spot, in fact, the six-antecedent version with emphasis is better on all statistics. Note again that the magnitude of the fuzzy constant is a good indication of the relative merits of the various models. The rules are as follows:
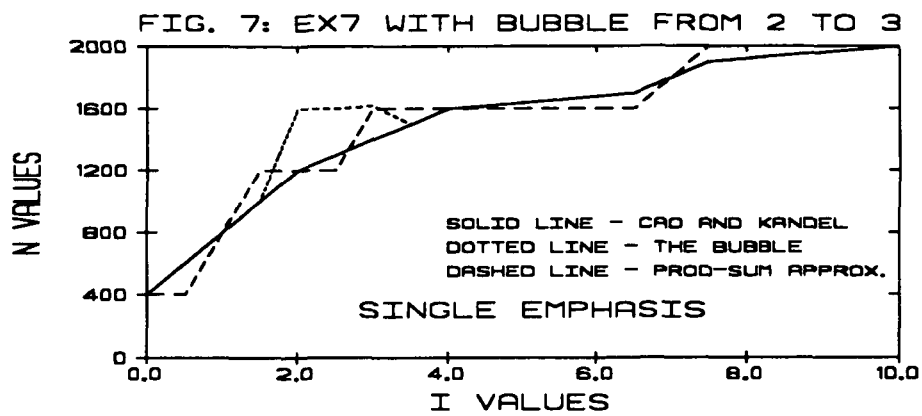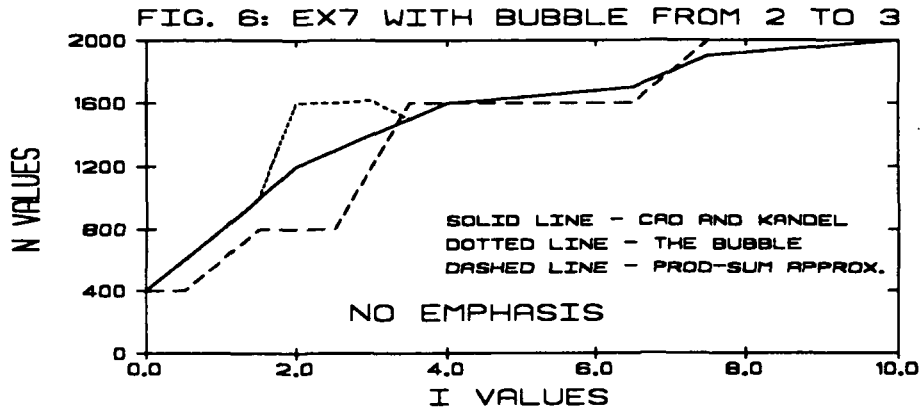
NULL -> VERY LARGE	NULL -> VERY LARGE	NULL -> VERY LARGE
ZERO -> MEDIUM	ZERO -> MEDIUM	ZERO -> MEDIUM
SMALL -> ZERO	ZERO-SMALL -> ZERO	SMALL -> ZERO
MEDIUM -> MEDIUM	SMALL -> ZERO	(repeat above for emphasis)
LARGE -> VERY LARGE	SMALL-MEDIUM -> MEDIUM	(repeat above for emphasis)
VERY LARGE -> MEDIUM	MEDIUM- LARGE -> VERY LARGE	MEDIUM -> MEDIUM
	LARGE -> LARGE	LARGE -> VERY LARGE
	VERY-LARGE -> MEDIUM	VERY-LARGE -> MEDIUM

TABLE 2: RESULTS FOR EX. 3 OF CAO AND KANDEL WITH BOTTOM FLATTENED AND ONE PEAK SHIFTED			
	6 ANT. SETS	8 ANT. SETS	6 ANT. SETS
INFERENCE TYPE	MAX-MIN	MAX-MIN	PROD-SUM
MCC	0.896	0.972	0.969
REL. ENTROPY	7.68	4.97	5.76
FOR 6 ANTECEDENT SETS MM Y = 1.06X - (126.10, 16.29, 16.29)			
FOR 8 ANTECEDENT SETS MM Y = (55.78, 7.85, 7.85) + 0.97X			
FOR 6 ANTECEDENT SETS PS Y = (79.82, 14.16, 14.16) + 0.95X			

211

FIG. 4: SIX ANTECEDENT FUZZY SETS

solid line — "real" data
dotted line — max min approx.

FIG. 5: EIGHT ANTECEDENT FUZZY SETS

solid line — "real" data
dotted line — max min approx.

3. In this example we return to the data of Example 7 in [15], but we add a bubble to the line at input values 2 to 3. As we emphasize the rule which raises the output values in that area (SMALL -> LARGE), first once and then twice, we observe corresponding improvement in the results. This improvement is obvious in the figures below, and is also tracked nicely once again by the statistics. Note that only the "double emphasis" inference creates a transition point in WSSD values in the center of the bubble. Since the effect of emphasis is essentially to shift a transition point toward the emphasized region, this is a sign that the input and output data sets are a good match. As an illustration of the value of the WSSD, we modified the single emphasis inference results so that just the spreads matched better in the bubble. Note that, as one might expect, this improves the overall least squares solution, but note also that this creates a WSSD transition point in the proper place. Since this would not be apparent from an inspection of the modes alone, the value of the WSSD to a detailed evaluation of the inference results is clear.

TABLE 3: RESULTS FOR EXAMPLE 7 OF CAO AND KANDEL WITH BUBBLE ADDED				
STATISTIC	DOUBLE EM	SINGLE EM	NO EMPHAS	SINGL EM+
MCC	0.967	0.9565	0.896	0.9566
ENT	3.90	4.45	6.48	4.27
WSSD 3.5	0.022243	0.026888	0.071660	0.027036
WSSD 3.0	0.022243	0.026888	0.133615	0.027036
WSSD 2.5	0.089423*	0.398910	1.005964	0.378739*
WSSD 2.0	0.438807	0.398910	1.005964	0.400629
WSSD 1.5	0.438807	0.398910	1.005964	0.400629
FOR DOUBLE EMPHASIS Y = (163.27, 24.75, 24.75) + 0.91X				
FOR SINGLE EMPHASIS Y = (199.36, 30.17, 30.17) + 0.89X				
FOR NO EMPHASIS Y = (503.50, 77.23, 77.23) + 0.73X				
FOR SINGLE EMPHAS.+ Y = (198.51, 28.33, 28.33) + 0.89X				
+ DIFFERS FROM SINGLE EMPHASIS ONLY IN FUZZINESS OF VALUES IN BUBBLE (BETTER MATCH)				

FIG. 6: EX7 WITH BUBBLE FROM 2 TO 3

SOLID LINE — CAO AND KANDEL
DOTTED LINE — THE BUBBLE
DASHED LINE — PROD-SUM APPROX.

NO EMPHASIS

FIG. 7: EX7 WITH BUBBLE FROM 2 TO 3

SOLID LINE — CAO AND KANDEL
DOTTED LINE — THE BUBBLE
DASHED LINE — PROD-SUM APPROX.

SINGLE EMPHASIS

FIG. 8: EX7 WITH BUBBLE FROM 2 TO 3

SOLID LINE — CAO AND KANDEL
DOTTED LINE — THE BUBBLE
DASHED LINE — PROD-SUM APPROX.

DOUBLE EMPHASIS

References:

[1] S. Smith and D. Comer, "Automated Calibration of a Fuzzy Logic Controller Using a Cell State Space Algorithm," IEEE Control Systems 11:5 (August, 1991), 18-28.

[2] K. Ostaszewski and W. Karwowski, " Fuzzy Clustering Methods for Assessment of Population Risks in the Insurance Industry, " in P. Ralston, K. Stoll, and T. Ward (eds.), Proceedings of the 1992 International Fuzzy Systems and Intelligent Control Conference, Louisville, Ky., 156-160.

[3] S. Van Huffel and J. Vandewalle, The Total Least Squares Problem: Computational Aspects and Analysis, Philadelphia: SIAM, 1991.

[4] C. S. Hsu, Cell-to-Cell Mapping, New York: Springer-Verlag, 1987.

[5] Y. Chen and T. Tsao, "A Description of the Dynamical Behavior of Fuzzy Systems," IEEE Transactions on Systems, Man, and Cybernetics 19:4 (1989), 745-755.

[6] P. Diamond, "Fuzzy Least Squares," Information Science 46 (1988), 141-157.

[7] N. Kiryati and A, Bruckstein, " What's in a Set of Points," IEEE Transactions on Pattern Analysis and Machine Intelligence 14:4 (1992), 496-500.

[8] M. Sugeno and T. Yasukawa, "Linguistic Modeling Based on Numerical Data, " IFSA '91 Brussels (Computer, Management, and Systems Science Section), 264-267.

[9] R. Johnson and D. Wichern, Applied Multivariate Statistical Analysis, Englewood Cliffs, NJ: Prentice-Hall, 1988.

[10] J. Barone, "Fuzzy Least Squares and Fuzzy Entropy," in P. Ralston, K. Stoll, and T. Ward (eds.), Proceedings of the 1992 International Fuzzy Systems and Intelligent Control Conference, Louisville, Ky., 170-181.

[11] C. Daniel and F. Woods, Fitting Equations to Data: Computer Analysis of Multifactor Data, New York: John Wiley & Sons, 1971.

[12] S. Chatterjee and A. Hadi, Sensitivity Analysis in Linear Regression, New York: John Wiley & Sons, 1988.

[13] O. Kaleva, "Interpolation of Fuzzy Data," IFSA '91 Brussels (Mathematics Section), 99-101.

[14] D. F. Mayers, " Convergence of Polynomial Interpolation," in D. C. Handscomb (ed.), Methods of Numerical Approximation, Oxford: Pergamon Press, 1966, 15-26.

[15] Z Cao and A. Kandel, "Applicability of Some Fuzzy Implication Operators," Fuzzy Sets and Systems 31 (1989), 151-186.

[16] M. Mizumoto, "Fuzzy Controls with Fuzzy Rules of Emphatic and Suppressive Types," in P. Ralston, K. Stoll, and T. Ward (eds.), Proceedings of the 1992 International Fuzzy Systems and Intelligent Control Conference, Louisville, Ky., 134-140.