USC–SIPI REPORT #262

Non-Singleton Fuzzy Logic Systems: Theory and Application

by

George C. Mouzouris and Jerry M. Mendel

June 1994

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Department of Electrical Engineering-Systems
3740 McClintock Avenue, Room 404
Los Angeles, CA 90089-2564 U.S.A.

Non-Singleton Fuzzy Logic Systems: Theory and Application

George C. Mouzouris and Jerry M. Mendel

Signal and Image Processing Institute

Department of Electrical Engineering-Systems

University of Southern California

Los Angeles, CA 90089-2564

Abstract

In this paper we present a formal derivation of general nonsingleton fuzzy logic systems (NSFLSs), and show how they can be efficiently computed. We give examples for special cases of membership functions and inference. We show how a NSFLS can be expressed as a *nonsingleton fuzzy basis function* expansion, and present an analytical comparison of the nonsingleton and singleton fuzzy logic systems formulations. We prove that a NSFLS can uniformly approximate any given continuous function on a compact set, and show that our NSFLS does a much better job of predicting a noisy chaotic time series than does a singleton FLS.

1 Introduction

A fuzzy logic system processes crisp data at the input and produces crisp data at the output; therefore, a *fuzzifier* is used at the front of the system to convert crisp data to fuzzy data, and a *defuzzifier* is used at the output of the system to convert fuzzy data into crisp data. The most widely used fuzzifier is the singleton fuzzifier [11, 28, 30], mainly because of its simplicity and lower computational requirements; however, this kind of fuzzifier may not always be adequate, especially

in cases where noise is present in the training data or in the data which is later processed by the system. A different approach is necessary in order to account for uncertainty in the data, which is why we direct our attention in this paper at the *nonsingleton* fuzzifier and nonsingleton fuzzy logic systems (NSFLSs).

Nonsingleton fuzzifiers have been used successfully in a variety of applications [1, 8, 18, 21, 24, 26]. In neural-fuzzy systems [1, 8], vectors of fuzzy sets are used both to train a fuzzy neural network and as inputs during processing. In [18] an optimizing control method for optimizing the fuel consumption rate of a marine diesel engine utilizes empirical rules that are expressed by fuzzy numbers. Nonsingleton input has also been used in turning process automation [24] to represent a human operator's actions in the fuzzy rule-base, in the design of fuzzy control algorithms [21], and fuzzy information and decision-making [26]. These methods are largely heuristic and provide no closed-form expressions for fuzzy logic systems; hence, their generalizations are very difficult.

In this paper, we develop a quantitative formulation of a NSFLS and its efficient computation. This formulation provides a tool for accounting for uncertainty in either the training data or the input to the system. In Section 2 we give a brief discussion on nonsingleton fuzzification. In Section 3 we derive, from first principles, the continuous and discrete forms of NSFLSs, and show how, and the conditions under which, a NSFLS reduces to a singleton FLS. We also quantify the difference between the output fuzzy sets for nonsingleton and singleton fuzzification; show how our NSFLS can be expressed as a nonsingleton fuzzy basis function (FBF) expansion; and, present examples for product and minimum inference and Gaussian input, antecedent, and consequent fuzzy sets. In Section 4 we show that our NSFLSs uniformly approximate any given continuous function on a compact set. In Section 5 we apply our NSFLS to the prediction of a noisy Mackey-Glass chaotic time series, and show, by means of simulations, that the NSFLS is much more successful in producing accurate forecasts of that series than is a comparable singleton FLS. In Section 6 we

draw our conclusions.

2 The Non-Singleton Fuzzifier

Fuzzy sets have been interpreted as membership functions μ_X [34] that associate with each element x of the universe of discourse, U, a number $\mu_X(x)$ in the interval $[0, 1]$:

$$\mu_X : U \rightarrow [0, 1]. \tag{1}$$

A fuzzifier maps a crisp point $x \in U$ into a fuzzy set $X \in U$.

1. In the case of a *singleton fuzzifier*, the crisp point $x \in U$ is mapped into a fuzzy set X with support x_i, where $\mu_X(x_i) = 1$ for $x_i = x$ and $\mu_X(x_i) = 0$ for $x_i \neq x$, i.e., the *single* point in the support of X with non-zero membership function value is $x_i = x$.

2. In the case of a *nonsingleton fuzzifier*, the point $x \in U$ is mapped into a fuzzy set X with support x_i, where μ_X achieves maximum value at $x_i = x$ and decreases while moving away from $x_i = x$. We assume that fuzzy set X is normalized so that $\mu_X(x) = 1$.

Nonsingleton fuzzification is especially useful in cases where the available training data, or the input data to the fuzzy logic system, are corrupted by noise. Conceptually, the nonsingleton fuzzifier implies that the given input value x is the most likely value to be the correct one from all the values in its immediate neighborhood; however, because the input is corrupted by noise, neighboring points are also likely to be the correct values, but to a lesser degree.

It is up to the system designer to determine the shape of the membership function μ_X, based on an estimate of the kind and quantity of noise present. It would be the logical choice, though, for the membership function to be symmetric about x, since the effect of noise is most likely to be equivalent on all points. Examples of such membership functions are: (1) the Gaussian,

$\mu_X(x_i) = exp[-\frac{(x-x_i)^2}{2\sigma^2}]$, where the variance σ^2 reflects the width (spread) of $\mu_X(x_i)$; (2) triangular, $\mu_X(x_i) = \max(0, 1 - |\frac{x-x_i}{c}|)$; and, (3) $\mu_X(x_i) = \frac{1}{(1+|\frac{x-x_i}{c}|^p)}$ where x and c are respectively the mean and spread of the fuzzy sets. Note that larger values of the spread of the above membership functions imply that more noise is anticipated to exist in the given data.

3 Non-Singleton Fuzzy Logic System Formulation

3.1 General Results

Consider a fuzzy logic system with a rulebase of M rules, and let the lth rule be denoted by R^l. Let each rule have p antecedents and one consequent (as is well known, a rule with q consequents can be decomposed into q rules, each having the same antecedents and one different consequent), i.e., it is of the general form:

$$R^l\text{: IF } u_1 \text{ is } F_1^l \text{ and } u_2 \text{ is } F_2^l \text{ and } \cdots \text{ and } u_p \text{ is } F_p^l \text{ THEN } v \text{ is } G^l,$$

where u_k, $k = 1, \ldots, p$, and v are the input and output linguistic variables, respectively. Each F_k^l and G^l are subsets of possibly different universes of discourse. Let $F_k^l \subset U_k$ and $G^l \subset V$. Each rule can be viewed as a fuzzy relation R^l [35] from a set U to a set V, where U is the Cartesian product space $U = U_1 \times \cdots \times U_p$. R^l itself is a subset of the Cartesian product $U \times V = \{(\underline{x}, y) : \underline{x} \in U, y \in V\}$, where $\underline{x} \equiv (x_1, x_2, \ldots, x_p)$, and x_k and y are the points in the universes of discourse, U_k and V, of u_k and v.

First we present a system formulation for the continuous case, and then for the discrete case. We also demonstrate that in the case of the *height defuzzifier*, the continuous and discrete cases produce identical systems.

R^l is characterized by a continuous multivariate membership function $\mu_{R^l}(\underline{x}, y)$, and can be described by the following:

$$\begin{aligned} R^l &\equiv \int_{U \times V} \mu_{R^l}(\underline{x}, y)/(\underline{x}, y) \\ &= \int_{U_1} \cdots \int_{U_p} \int_V \mu_{F_1^l}(x_1) \star \cdots \star \mu_{F_p^l}(x_p) \star \mu_{G^l}(y)/(\underline{x}, y), \end{aligned} \tag{2}$$

where $\star$ denotes a t-norm. $\int$ denotes the union of individual points of each set in the continuum.

Let the input to R^l be denoted by A, where A is a subset of a p-dimensional Cartesian product space, and is given by:

$$A = \int_{U_1} \cdots \int_{U_p} \mu_{X_1}(x_1) \star \cdots \star \mu_{X_p}(x_p)/(\underline{x}), \tag{3}$$

where $X_k \subset U_k$ $(k = 1, \ldots, p)$ are the fuzzy sets describing the inputs.

Up to this point, the formulation is identical to that of the singleton case. In the singleton case, though, it is assumed that each input fuzzy set X_k has non-zero membership value only at a single point, which reduces A to a set with a single point $\underline{x}_r \in U$. In our treatment here, we do not make this assumption. Each input fuzzy set is represented by the more general nonsingleton form in (3), thereby allowing any uncertainty in the input to be represented in the system.

According to the Compositional Rule of Inference, the fuzzy subset Y^l of V induced by $A \in U$ is given by the composition of A and R^l:

$$Y^l = A \circ R^l = \sup_{\underline{x} \in U}(A \star R^l). \tag{4}$$

Note that all unions (denoted by $\int$) in A and R^l, over U_k, $k = 1, \ldots, p$, are over the same spaces; therefore, we can write Y^l, as

$$Y^l = \sup_{\underline{x} \in U} \left(\int_{U_1} \cdots \int_{U_p} \int_V \mu_{X_1}(x_1) \star \cdots \star \mu_{X_p}(x_p) \star \mu_{F_1^l}(x_1) \star \cdots \star \mu_{F_p^l}(x_p) \star \mu_{G^l}(y) \right) / (y). \tag{5}$$

Since the supremum is only over $\underline{x} \in U$, then by the commutativity and monotonicity properties of a t-norm, we can rewrite Y^l as:

$$Y^l = \int_V \mu_{G^l}(y) \star \sup_{\underline{x} \in U} \left(\int_{U_1} \cdots \int_{U_p} \mu_{X_1}(x_1) \star \cdots \star \mu_{X_p}(x_p) \star \mu_{F_1^l}(x_1) \star \cdots \star \mu_{F_p^l}(x_p) \right) / (y). \tag{6}$$

Next, we recall that, by definition, a t-norm is a two-place function from $[0,1] \times [0,1]$ [36]; thus, we can consider every t-norm in (6) to be acting on a pair of membership functions. The calculation of the t-norm over all the points in the corresponding spaces of the two membership functions is easier to visualize if the membership functions are in the same space; therefore, we rewrite (6) in the following manner:

$$Y^l = \int_V \mu_{G^l}(y) \star \sup_{\underline{x} \in U} \left(\int_{U_1} \cdots \int_{U_p} [\mu_{X_1}(x_1) \star \mu_{F_1^l}(x_1)] \star \cdots \star [\mu_{X_p}(x_p) \star \mu_{F_p^l}(x_p)] \right) / (y). \tag{7}$$

Note that the supremum in (7) is over all points $\underline{x}$ in a p-dimensional Cartesian product space. By the monotonicity property of a t-norm [36], that supremum is attained when each term in brackets attains its supremum. Let

$$\mu_{Q_k^l}(x_k) \equiv \mu_{X_k}(x_k) \star \mu_{F_k^l}(x_k), \tag{8}$$

where $\mu_{Q_k^l}$, μ_{X_k}, $\mu_{F_k^l} \in U_k$. Assuming that $\mu_{Q_k^l}$ produced by (8) is a function whose supremum can be evaluated, let $x_{k,sup}$ denote the point in U_k where that supremum is attained; then, (7)

becomes:

$$Y^l = \int_V \mu_{G^l}(y) \star \mathcal{T}_{k=1}^{p} \mu_{Q_k^l}(x_{k,sup})/(y), \tag{9}$$

where $\mathcal{T}_{k=1}^{p}$ denotes a sequence of p t-norm operations.

Using the same procedure and rationale as in the continuous case, we can derive corresponding expressions for discrete NSFLSs. Each U_i, $(i = 1, \ldots, p)$ and G are finite (and thus countable), and the ith antecedent membership function, $\mu_{F_i^l}$, and consequent membership function, μ_{G^l}, have nonzero membership values at discrete points $x_{i,1}, \ldots, x_{i,n_i}$, and $v_1, \ldots, v_m$, respectively. The lth rule, R^l, is therefore given by:

$$R^l = \sum_{i_1=1}^{n_1} \cdots \sum_{i_p=1}^{n_p} \sum_{j=1}^{m} \mu_{F_1^l}(x_{1,i_1}) \star \cdots \star \mu_{F_p^l}(x_{p,i_p}) \star \mu_{G^l}(y_j)/(x_{1,i_1}, x_{2,i_2}, \ldots, x_{p,i_p}, y_j) \tag{10}$$

where $\sum$ denotes the union of the individual points of each set. The p-dimensional input to R^l is given by $A = \sum_{i_1=1}^{n_1} \cdots \sum_{i_p=1}^{n_p} \mu_{X_1}(x_{1,i_1}) \star \cdots \star \mu_{X_p}(x_{p,i_p})/(x_{1,i_1}, x_{2,i_2}, \ldots, x_{p,i_p})$. Using the Compositional Rule of Inference and the commutativity and monotonicity properties of a t-norm, we can write the discrete output fuzzy set Y^l as:

$$\begin{aligned} Y^l &= \sum_{j=1}^{m} \mu_{G^l}(y_j) \star \\ &\sup_{\underline{x} \in U} \left(\sum_{i_1=1}^{n_1} \cdots \sum_{i_p=1}^{n_p} \mu_{X_1}(x_{1,i_1}) \star \cdots \star \mu_{X_p}(x_{p,i_p}) \star \mu_{F_1^l}(x_{1,i_1}) \star \cdots \star \mu_{F_p^l}(x_{p,i_p}) \right) /(y_j). \end{aligned} \tag{11}$$

From the finiteness of each universe of discourse, U_k, the supremum is the same as the maximum; therefore, we need to calculate the maximum over all $\underline{x} \in U$ of the parenthetical term in (11). That term will be maximized when every term $\mu_{Q_k^l}(x_{k,i_k}) \equiv \mu_{X_k}(x_{k,i_k}) \star \mu_{F_k^l}(x_{k,i_k})$ is maximized. If the

global maximum, $x_{k,max}$, of each $\mu_{Q_k^l}(x_{k,i_k})$ can be evaluated, then,

$$Y^l = \sum_{j=1}^{m} \mu_{G^l}(y_j) \star \mathcal{T}_{k=1}^{p} \mu_{Q_k^l}(x_{k,max})/(y_j), \tag{12}$$

which is the discrete counterpart to (9).

Because our FLS has M rules, the final output fuzzy set, Y, is obtained by t-conorm aggregation of the individual outputs of each rule, i.e.,

$$Y = Y^1 \dot{+} Y^2 \dot{+} \cdots \dot{+} Y^M \tag{13}$$

where $\dot{+}$ denotes t-conorm (s-norm) [15, 36].

Kosko [11] has utilized arithmetic addition as a means of aggregation instead of a t-conorm. In most cases, this produces a central-limit-theorem effect that approximates a unimodal fuzzy set. The smaller the spread of this set, the higher the probability that the defuzzified value is the correct one. This additive structure is encountered in many t-conorms, which impose certain conditions (i.e., bounded sum) or normalization to keep the result in $[0, 1]$ (e.g., Einstein sum, algebraic sum, Hamacher sum [36]).

We will use a special case of an additive fuzzy logic system, one that uses a weighted additive combination of the points where each output fuzzy set becomes a maximum; it is the *modified* height defuzzifier of [9], which was also proposed by Wang [28, 29]. Let $\bar{y}^l$ denote the point in each individual output fuzzy set such that $\mu_{G^l}(\bar{y}^l) = 1$, and δ^l denote the spread of μ_{G^l}. Then, the modified height defuzzifier leads to the following output for our nonsingleton system:

$$y = f_{ns}(\underline{x}) = \frac{\sum_{l=1}^{M} \bar{y}^l \left[\mathcal{T}_{k=1}^{p} \mu_{Q_k^l}(x_{k,sup})\right]/(\delta^l)^2}{\sum_{l=1}^{M} \left[\mathcal{T}_{k=1}^{p} \mu_{Q_k^l}(x_{k,sup})\right]/(\delta^l)^2}. \tag{14}$$

The spread, δ^l, of the output fuzzy set is proportional to the amount of uncertainty that $\bar{y}^l$ is close to the point where $\mu_{G^l}(y^l) = 1$ (e.g., in the case of Gaussian consequent fuzzy sets, δ^l could be chosen to be the standard deviation); thus, if δ^l is very small (high degree of confidence), then the corresponding point $\bar{y}^l$ is more heavily weighted than if δ^l is very large. For the case of the modified height defuzzifier, the discrete nonsingleton fuzzy logic system (with M rules in its rulebase), has the form of (14) with $x_{k,sup}$ replaced by $x_{k,max}$, and, assuming fine enough discretization, will give identical results as the continuous system.

By defining the following nonsingleton fuzzy basis functions (FBFs)

$$p_l(\underline{x}) = \frac{\left[\mathcal{T}_{k=1}^{p} \mu_{Q_k^l}(x_{k,sup})\right] / (\delta^l)^2}{\sum_{l=1}^{M} \left[\mathcal{T}_{k=1}^{p} \mu_{Q_k^l}(x_{k,sup})\right] / (\delta^l)^2}, \tag{15}$$

we can express our nonsingleton system in (14) as the following *fuzzy basis function expansion*:

$$f_{ns}(\underline{x}) = \sum_{l=1}^{M} \bar{y}^l p_l(\underline{x}). \tag{16}$$

The modified height defuzzifier is a simple and fast method to implement a general FLS; however, other defuzzification methods, that have different properties, can also be used, e.g., *Center-Of-Area*, *Center-Of-Sums*, and *First-Of-Maxima* (see [9] for an overview and comparison of different defuzzification methods).

In order to contrast the nonsingleton and singleton discrete cases, and to also show that the singleton case is a special case of (11), we assume, as mentioned before, that the input fuzzy sets are normalized. Consequently, for each input fuzzy set X_k there exists at least one point $x_{k,q_k} \in X_k$ such that $\mu_{X_k}(x_{k,q_k}) = 1$. In addition, we assume that $\bar{x}_k \equiv x_{k,q_k}$ is unique in X_k (extension to multiple such points is easy); then, $\mu_{X_1}(\bar{x}_1) = \ldots = \mu_{X_p}(\bar{x}_p) = 1$, and $\mu_{X_1}(\bar{x}_1) \star \ldots \star \mu_{X_p}(\bar{x}_p) = 1$, which is true for all t-norms. Consequently, the general nonsingleton case in (11) can be decomposed into

contribution of the singleton components and the remaining components with membership values less than unity:

$$
\begin{aligned}
Y^l &= \sum_{j=1}^{m} \mu_{G^l}(y_j) \star \sup_{\underline{x} \in U} \Big[\mu_{F_1^l}(\bar{x}_1) \star \ldots \star \mu_{F_p^l}(\bar{x}_p) + \sum_{i_1 \neq q_1} \sum_{i_2 \neq q_2} \cdots \sum_{i_p \neq q_p} \mu_{X_1}(x_{1,i_1}) \star \\
&\quad \mu_{X_2}(x_{2,i_2}) \star \cdots \star \mu_{X_p}(x_{p,i_p}) \star \mu_{F_1^l}(x_{1,i_1}) \star \mu_{F_2^l}(x_{2,i_2}) \star \cdots \star \mu_{F_p^l}(x_{p,i_p}) \Big] / (y_j) \\
&\equiv \sum_{j=1}^{m} \mu_{G^l}(y_j) \star S_{ns} / (y_j). \qquad (17)
\end{aligned}
$$

where '+' denotes union. If the inputs are indeed singletons, then the term on the right-hand side of '+' becomes zero, and the supremum is no longer necessary, in which case (17) reduces to the familiar singleton case, namely,

$$
\begin{aligned}
Y_s^l &= \sum_{j=1}^{m} \mu_{G^l}(y_j) \star \mu_{F_1^l}(\bar{x}_1) \star \ldots \star \mu_{F_p^l}(\bar{x}_p) / (y_j) \\
&\equiv \sum_{j=1}^{m} \mu_{G^l}(y_j) \star S_s / (y_j). \qquad (18)
\end{aligned}
$$

Equations (17) and (18) show that in the nonsingleton system the scaling of the output fuzzy set for each rule will, in general, be different than in the case of a singleton system.

A general expression for quantifying the scale factor difference, $\delta s = S_{ns} - S_s$, can be found as a direct consequence of (17) and (18), i.e.,

$$
\begin{aligned}
\delta s = S_{ns} - S_s = \sup_{\underline{x} \in U} [& \mu_{F_1^l}(\bar{x}_1) \star \cdots \star \mu_{F_p^l}(\bar{x}_p) + \sum_{i_1 \neq q_1} \cdots \sum_{i_p \neq q_p} \mu_{X_1}(x_{1,i_1}) \star \cdots \\
& \star \mu_{X_p}(x_{p,i_p}) \star \mu_{F_1^l}(x_{1,i_1}) \star \cdots \star \mu_{F_p^l}(x_{p,i_p})] - \mu_{F_1^l}(\bar{x}_1) \star \cdots \star \mu_{F_p^l}(\bar{x}_p). \qquad (19)
\end{aligned}
$$

The most widely used subclasses of FLSs are those with product or minimum inference as a t-norm, and with Gaussian or triangular membership functions. These subclasses are considered

next.

3.2 Product Inference

The kth input fuzzy set and the corresponding k^{th} antecedent are assumed to have the following forms:

$$\mu_{X_k}(x_{k,i_k}) = \exp\left(-\frac{(x_{k,i_k} - m_{X_k})^2}{2\sigma_{X_k}^2}\right), \quad \mu_{F_k^l}(x_{k,i_k}) = \exp\left(-\frac{(x_{k,i_k} - m_{F_k^l})^2}{2\sigma_{F_k^l}^2}\right), \tag{20}$$

and '$\star$' in (11) denotes algebraic product, i.e.,

$$Y^l = \sum_{j=1}^{m} \mu_{G^l}(y_j) \cdot \sup_{\underline{x} \in U} \left(\sum_{i_1=1}^{n_1} \cdots \sum_{i_p=1}^{n_p} [\mu_{X_1}(x_{1,i_1}) \cdot \mu_{F_1^l}(x_{1,i_1})] \cdots [\mu_{X_p}(x_{p,i_p}) \cdot \mu_{F_p^l}(x_{p,i_p})] \right) / (y_j). \tag{21}$$

By maximizing the function

$$\mu_{Q_k^l}(x_{k,i_k}) = \mu_{X_k}(x_{k,i_k}) \cdot \mu_{F_k^l}(x_{k,i_k}), \tag{22}$$

we find that it is maximum at

$$x_{k,max} = \frac{\sigma_{X_k}^2 m_{F_k^l} + \sigma_{F_k^l}^2 m_{X_k}}{\sigma_{X_k}^2 + \sigma_{F_k^l}^2}. \tag{23}$$

An important special case occurs when all input points for each input variable have the same level of uncertainty (after they are fuzzified using the nonsingleton fuzzifier), so that $\sigma_{X_k}^2 = \sigma_X^2$. In this case, $m_{X_k} = x_k$ (see Section 2), and therefore

$$x_{k,max} = \frac{\sigma_X^2 m_{F_k^l} + \sigma_{F_k^l}^2 x_k}{\sigma_X^2 + \sigma_{F_k^l}^2}. \tag{24}$$

Substituting (24) into (21), the latter becomes:

$$Y^l = \sum_{j=1}^{m} \mu_{G^l}(y_j) \prod_{k=1}^{p} \mu_{Q_k^l}(x_{k,max})/(y_j). \tag{25}$$

Since we are using Gaussian consequent fuzzy sets, the point where each output set becomes maximum is unique. Consequently, for the special case of product inference, Gaussian membership functions, and equal uncertainty on all inputs, the nonsingleton FBFs given by (15) become:

$$p_l(\underline{x}) = \frac{\prod_{k=1}^{p} \mu_{Q_k^l}\left(\frac{\sigma_X^2 m_{F_k^l} + \sigma_{F_k^l}^2 x_k}{\sigma_X^2 + \sigma_{F_k^l}^2}\right)/(\delta^l)^2}{\sum_{l=1}^{M} \prod_{k=1}^{p} \mu_{Q_k^l}\left(\frac{\sigma_X^2 m_{F_k^l} + \sigma_{F_k^l}^2 x_k}{\sigma_X^2 + \sigma_{F_k^l}^2}\right)/(\delta^l)^2}. \tag{26}$$

When the uncertainty of the input is zero, $\sigma_X^2 = 0$, and (23) reduces to the singleton case, i.e., $x_{k,max} = m_{X_k} = x_k$. In this case, each $\mu_{X_k}(x_{k,max})$, $k = 1, \ldots, p$, is unity, and therefore $\mu_{Q_k^l}(x_{k,max}) = \mu_{F_k^l}(x_k)$, so that (25) reduces to

$$Y^l = \sum_{j=1}^{m} \mu_{G^l}(y_j) \prod_{k=1}^{p} \mu_{F_k^l}(x_k)/(y_j), \tag{27}$$

and the FBFs in (26) become:

$$p_l^s(\underline{x}) = \frac{\prod_{k=1}^{p} \mu_{F_k^l}(x_k)/(\delta^l)^2}{\sum_{l=1}^{M} \prod_{k=1}^{p} \mu_{F_k^l}(x_k)/(\delta^l)^2}, \tag{28}$$

where p_l^s denotes the singleton FBFs. Additionally [see (14)],

$$f_{ns}(\underline{x}) = \sum_{l=1}^{M} \bar{y}^l p_l(\underline{x}) = \frac{\sum_{l=1}^{M} \bar{y}^l \left[\mathcal{T}_{k=1}^{p} \mu_{Q_k^l}(x_{k,max})\right]/(\delta^l)^2}{\sum_{l=1}^{M} \left[\mathcal{T}_{k=1}^{p} \mu_{Q_k^l}(x_{k,max})\right]/(\delta^l)^2} \tag{29}$$

reduces to

$$f_s(\underline{x}) = \sum_{l=1}^{M} \bar{y}^l p_l^s(\underline{x}) = \frac{\sum_{l=1}^{M} \bar{y}^l \prod_{k=1}^{p} \mu_{F_k^l}(x_k)/(\delta^l)^2}{\sum_{l=1}^{M} \prod_{k=1}^{p} \mu_{F_k^l}(x_k)/(\delta^l)^2}. \tag{30}$$

Comparing (29) and (30), we see that nonsingleton and singleton FLSs are structurally the same.

To illustrate the differences between a NSFLS and a singleton FLS, we present the case of a one rule SISO FLS with Gaussian membership functions and product inference. For the nonsingleton system, the input is also fuzzified using a Gaussian membership function. Figure 1 shows the output fuzzy sets corresponding to each system. Observe the Gaussian nature, in Fig. 1b, of the product function $\mu_{Q_1^1}(x)$, and the location of its maximum value at $x = x_{1,max}$. In this case, the output membership function for the NSFLS is of greater maximum height than that of the singleton FLS.

Figure 2 demonstrates (19) for product inference and Gaussian membership functions. The bimodal curves indicate the scale factor difference between the two systems, as the input sweeps over the entire universe of discourse of the antecedent membership function. The larger amplitude scale factor difference curve corresponds to a nonsingleton input with higher level of uncertainty ($\sigma_X \approx 0.7\sigma_F$), whereas the smaller amplitute difference curve corresponds to nonsingleton input with lower level of uncertainty ($\sigma_X \approx 0.2\sigma_F$). When the input singleton and the mean of the corresponding nonsingleton input are at the mean of the antecedent [i.e., $m_{F_k^l} = m_{X_k}$, so that $x_{k,max} = m_{F_k^l}$ (see (23)], the difference in the scale factor is zero.

Figures 1 and 2 clearly imply that (especially in the cases of more complex systems with more than one rule, and multiple inputs) the defuzzified nonsingleton system output may be significantly different than the corresponding singleton system output.

Figure 3a, which is a direct implementation of (26) (with $p = 1$, $M = 4$, $x_1 \in [1, 100]$, $\delta^l = 1$, $m_{F_1^l} = 20 \cdot l$, and $\sigma_{F_1^l} = 10$, for $l = 1, \ldots, M$), shows the difference between the fuzzy basis functions

in a singleton system versus the fuzzy basis functions in a nonsingleton system. When the degree of uncertainty in the Gaussian input is large, the basis functions of (26) (shown in Fig. 3a by the dash-dotted line) are significantly different than in the singleton case (solid line). Note that, when the uncertainty at the input is small, then the nonsingleton case is essentially identical to the singleton case, and the NS FBFs plotted in Fig. 3a are indistinguishable from the singleton FBFs. For the example presented here, the higher uncertainty input had variance of the order of the variance of the antecedents, whereas the lower uncertainty input had variance about one twentieth of the variance of the antecedents. Figure 3b depicts the fuzzy basis functions for a singleton (solid line) and a nonsingleton FLS with four rules, product inference and triangular membership functions. Observe that the two exterior FBFs saturate at unity.

Greater uncertainty in the input not only "fires" rules at a higher level than a singleton system [Eq. (19), and Figs. 1, and 2], but it also usually fires more rules than a singleton system would. This is due to the fact that a nonsingleton input may have membership (or, nonzero subsethood [10]) in more antecedent fuzzy regions than a singleton input. For example, a vertical line at $x = 40$ in Fig. 3a intersects three singleton FBFs and four nonsingleton FBFs.

3.3 Minimum Inference

Now, we let '$\star$' in (11) be the t-norm 'min', in which case $\mu_{Q_k^l}(x_k)$ in (8) is

$$\mu_{Q_k^l}(x_{k,i_k}) \equiv \min(\mu_{X_k}(x_{k,i_k}), \mu_{F_k^l}(x_{k,i_k})), \tag{31}$$

for finite universes of discourse. As in the case of the algebraic product t-norm, the term in the parenthesis in (11) is maximized when every grouped pair is a maximum, i.e., $\mu_{Q_k^l}(x_{k,i_k})$ must be maximized over all k and all x.

In order for the input fuzzy set X_k to activate the fuzzy rule R^l at F_k^l, some of the points in the

support of X_k must coincide with some points in the support of F_k^l. Let this set of points define the closed interval $[a_k, c_k]$ such that $b_k \in (a_k, c_k)$ and $a_k < b_k < c_k$ (Fig. 4). Without loss of generality, let $\mu_{F_k^l}(x_{k,i_k})$ be strictly increasing, and $\mu_{X_k}(x_{k,i_k})$ be strictly decreasing on the interval $[a_k, c_k]$ and consequently on any closed subintervals $[a_k, b_k]$, $[b_k, c_k]$. [In the case of $\mu_{X_k}(x_{k,i_k})$ being a proper subset of $\mu_{F_k^l}(x_{k,i_k})$, i.e., $\mu_{X_k}(x_{k,i_k}) \subset \mu_{F_k^l}(x_{k,i_k})$, and $\mu_{X_k}(x_{k,i_k}) \neq \mu_{F_k^l}(x_{k,i_k})$ for at least one x_{k,i_k}, or if $\mu_{X_k}(x_{k,i_k}) = \mu_{F_k^l}(x_{k,i_k})$, then obviously $\max_{\forall x} \mu_{Q_k^l}(x_{k,i_k}) \equiv \max_{\forall x} \mu_{X_k}(x_{k,i_k})$]. Let b_k be the point of intersection of $\mu_{F_k^l}$ and μ_{X_k}. From the monotonicity conditions on the membership functions in these intervals, we have:

$$\forall\, x \in [a_k, b_k], \;\; \min(\mu_{X_k}(x_{k,i_k}), \mu_{F_k^l}(x_{k,i_k})) = \mu_{F_k^l}(x_{k,i_k})$$
$$\text{where } \mu_{F_k^l}(x_{k,1}) < \mu_{F_k^l}(x_{k,2}) \text{ if } x_{k,1} < x_{k,2}; \tag{32}$$

thus,

$$\sup_{x \in [a_k, b_k]} \mu_{F_k^l}(x_{k,i_k}) = \mu_{F_k^l}(b_k). \tag{33}$$

Similarly,

$$\forall\, x \in [b_k, c_k], \;\; \min(\mu_{X_k}(x_{k,i_k}), \mu_{F_k^l}(x_{k,i_k})) = \mu_{X_k}(x_{k,i_k})$$
$$\text{where } \mu_{X_k}(x_{k,3}) > \mu_{X_k}(x_{k,4}) \text{ if } x_{k,3} < x_{k,4}, \tag{34}$$

so that

$$\sup_{x \in [b_k, c_k]} \mu_{X_k}(x_{k,i_k}) = \mu_{X_k}(b_k). \tag{35}$$

Consequently, $\mu_{Q_k^l}(x_{k,i_k})$ becomes a maximum at the point of intersection of the two membership functions, which is also intuitively satisfying. For example, for the previously described Gaussian membership functions [Eq. (20)], by equating their exponents and solving a second-order equation, we find the point of intersection to be given by:

$$x_{k,max} = \frac{\sigma_X m_{F_k^l} + \sigma_{F_k^l} x_k}{\sigma_X + \sigma_{F_k^l}}, \tag{36}$$

if $\sigma_{F_k^l} > \sigma_X$ and $m_{F_k^l} > x_k$.

The output fuzzy set Y^l is computed exactly as in (25), after replacing product with *min*, and $x_{k,max}$ with the expression in (36).

The complete solutions for the maximization of $\mu_{Q_k^l}$ for algebraic product and minimum t-norms are summarized in Table 1, along with the corresponding solutions for the case of triangular membership functions. The centers of the antecedent and input triangular membership functions are denoted by $m_{F_k^l}$ and m_{X_k}, respectively, and their corresponding left and right spreads by $l_{F_k^l}$, l_{X_k}, $r_{F_k^l}$, and r_{X_k}. Observe, from Table 1, that $x_{k,max}$ is easiest to calculate in the cases of product inference with Gaussian membership functions, and minimum inference with triangular membership functions.

4 Uniform Approximation Property of Non-Singleton Fuzzy Logic Systems

Wang and Mendel [31] and Kosko [11] have shown that specific subclasses of singleton fuzzy logic systems are universal approximators. In [31], the Stone-Weierstrass theorem (involving algebras of functions) was used, for approximations in the uniform topology on a compact set. This result was proven only for singleton FLSs with product inference and Gaussian membership functions. In

[11], continuity was used to prove uniform approximation capability of singleton additive systems.

Although (14) bears structural resemblence to Kolmogorov's representation theorem [16], the two are quite different. Kolmogorov's theorem (which resolved Hilberts 13th problem) provides a representation of a continuous function of several variables defined on an n-dimensional cube, by sums and superpositions of continuous functions of one variable. In contrast to fuzzy basis functions (15), the nonlinear functions employed in Kolmogorov's theorem, as well as in the subsequent improvements of that theorem by Lorentz [16] and Sprecher [25], are highly nonsmooth and strictly increasing. Even though these nonsmooth functions can be viewed as limits of uniformly converging series of smooth *sigmoidal* type functions [12, 13], the monotonicity requirements are too restrictive for our basis functions.

Here, utilizing concepts from real analysis [23], we present a uniform approximation theorem for nonsingleton fuzzy logic systems, that includes the systems described in [31] and [11] as special cases.

Theorem. Any given continuous function $g : U \rightarrow V$ on a compact set $U \in R^p$ can be uniformly approximated by the nonsingleton fuzzy logic system f_{ns} in (14).

Proof: See Appendix A.

This is an existence theorem, which gives us assurance that there exists an NSFLS capable of uniformly approximating any continuous function on compacta; however, it does not tell us how to choose the parameters of the NSFLS nor does it tell us how many basis functions will be needed to achieve such performance. These topics will be the subject of a later paper.

5 Modeling and Prediction of Complex Systems: Chaotic Time Series

Several well-known algorithms exist [2, 4, 33] for the calculation of dynamical and geometric invariants such as fractal dimension (capacity, correlation, information) and Lyapunov exponents, of an underlying strange attractor of time series. The largest Lyapunov exponent is essentially a measure of how predictable a system can be from observations of the past, whereas the fractal dimension is an indication of the complexity of a system. Unfortunately, data requirements and extreme sensitivity to noisy measurements make these algorithms prohibitively expensive in practical applications. Even if these invariants could be calculated accurately, they still do not provide enough information for the construction of a predictive model.

Other approaches attempt to construct a predictive model based solely on the given time-series data. Several such approaches (e.g., Gabor polynomials, linear prediction) have been compared to prediction using feedforward neural networks, with the best results occuring for the neural networks [14]. In [28] it was shown that a numerical-fuzzy approach based on a singleton FLS can give results equivalent to those or even better than that of a neural network. Here, we show that our nonsingleton FLS can significantly outperform an equivalent singleton FLS in the predictive modeling of a chaotic time series, especially in cases where either, or both, the training data and input data are corrupted by noise.

Chaos is having an impact on many different fields, including physics, biology, chemistry, economics, and medicine [3, 5, 22]. The chaotic time series that we examine is a model for blood production, due to Mackey and Glass [17]. This chaotic time series is described by the following

delay differential equation, that is now known as the "Mackey-Glass equation":

$$\frac{dx(t)}{dt} = \frac{0.2x(t-\tau)}{1+x^{10}(t-\tau)} - 0.1x(t). \tag{37}$$

For τ greater than 17, (37) exhibits chaotic behavior.

Chaotic behavior can be described as bounded fluctuations of the output of a system with high degree of sensitivity to initial conditions [3]; i.e., trajectories with nearly identical initial conditions diverge exponentially (note that exponential divergence of two distinct trajectories does not imply unboundedness. Figure 5 illustrates this kind of behavior). A system, such as the one in (37), exhibiting chaotic dynamics, evolves in a deterministic manner; but, the correlation of observations appears to be limited; thus, prediction of the state of the system is particularly difficult [22].

In order to demonstrate the qualitative nature of the dynamical system given by (37), we display representative portions of the time series for different values of τ in Figs. 6(a) and 6(b). We also depict the amplitude spectrum estimates of these representative sections in Fig. 6(c) and 6(d), and their corresponding two-dimensional phase plots in Fig. 6(e) and 6(f). From these plots we are able to distinguish periodic behavior for small τ's and chaotic behavior for larger τ's. Most commonly, the transition of a system from predictability to chaos occurs via a cascade of bifurcations, which are qualitative changes in the limit sets of a system, where a limit set is the set of points whose every neighborhood B is repeatedly entered by the solution $\phi_t(x)$ of the differential equation as $t \to \infty$. For τ close to 17, which is the onset of chaos for the Mackey-Glass equation, the bifurcations get closer and closer together, until the system breaks down into chaotic behavior.

The largest Lyapunov exponent [5], λ, which quantifies the rate of divergence of the time series and therefore the degree to which the system is predictable in the long term, can be estimated using nonlinear regression [19]. If $\lambda > 0$, then the system will exhibit chaotic behavior (this property

makes it possible to detect chaotic dynamics in a noisy environment just by examining the sign of λ). We estimated λ to be 0.007 for the Mackey-Glass equation with $\tau = 30$, thus confirming that the system is indeed chaotic.

We designed the NSFLS in (16) to be a single-step predictor of the Mackey-Glass chaotic time series. Our design is based on N training pairs that we obtained by simulating (37) for $\tau = 30$, using a finite-dimensional iterated mapping. A higher-order Adams-Bashforth integration algorithm [20] with sufficiently small step size would be a good choice for the simulation of (37) (note that in order for the higher-order algorithms to yield better results than the lower-order ones, the integration step should be chosen to be sufficiently small). In our simulations, we assumed that only noisy measured values of $x(t)$ are available, namely, $\zeta(t) = x(t) + n(t)$, $t = 1, \ldots, N$. The input vector to our fuzzy logic system is $\underline{u}(t-1) = [\zeta(t-1), \zeta(t-2), \ldots, \zeta(t-p)]^T$, so that the estimate of the time series, based on our NSFLS is given by:

$$\hat{x}(t) = f_{ns}(\underline{u}(t-1)). \tag{38}$$

Figure 7 depicts the prediction results, in the absense of noise, for three hundred points of (37) [$x(1001)$ to $x(1300)$, after allowing 1000 points for the transients to die out], produced by our singleton FLS using a numerical-fuzzy algorithm (all fuzzy rules were generated from the training data). Five hundred input-output training pairs of five inputs and a single output were used to construct the system; these were drawn from $x(1)$ to $x(1000)$.The same numerical-fuzzy algorithm and number of training pairs were used to construct a nonsingleton system of the form of (14). All the common parameters between the singleton and nonsingleton FLSs were chosen to be the same. For all simulations, σ_X in the NSFLS case was chosen equal to the standard deviation of the additive noise. In a noise-free environment, for which $\sigma_X = 0$, the performance of the nonsingleton

system was identical to that of the singleton system. As shown in Fig. 7, both systems can closely predict (37) in the given range.

Using the same common parameters for both systems, as in the noise free case, we added zero mean uniform noise to the training and test sets and we retrained both systems. Again, five hundred input-output training pairs of five inputs and a single output were used to construct the system, drawn from $\zeta(1)$ to $\zeta(1000)$. Figure 8 shows typical prediction results for both the singleton and the nonsingleton systems, at signal-to-noise ratios (SNRs) (defined as $20 \cdot log\frac{\sigma_{signal}}{\sigma_{noise}}$, where σ denotes standard deviation) of -5, 0, and 10dB. Figure 9 summarizes the performance of the nonsingleton and singleton systems for these SNRs, averaged over 100 realizations. The average mean-squared error and average standard deviation for each case are given in Table 2. Figure 10a shows the prediction results based on our NSFLS, for twenty different realizations and SNR = 0dB, overlapped on the same figure to give a visual representation of the standard deviation of the predicted estimates. Figure 10b shows the prediction results based on the corresponding singleton FLS, for twenty realizations. It is evident from this figure that the singleton FLS does not appear to be able to deal effectively with the uncertainty during the prediction process, due to the noise in the training data and the input. Our NSFLS is significantly more successful in producing better predicted estimates of the chaotic time series.

Note that no attempt was made to optimize the FLSs used here. Algorithms aimed at obtaining optimum, or near optimum, NSFLSs will be described in a future publication.

6 Conclusions

We have presented the continuous and discrete cases of a general NSFLS, and its efficient computation for the special subclasses of Gaussian and triangular membership functions with product or minimum inference. We have also developed a simple equation that describes the scale factor

difference between singleton and nonsingleton systems; it lets us examine and compare the behavior of singleton and nonsingleton systems for different types of inputs and uncertainty levels. We have also shown that a NSFLS can be expressed as a linear combination of nonsingleton fuzzy basis functions. Although, conceptually, singleton and nonsingleton FLSs are very different, the latter does have essential structural similarities with the former, and they become identical in the absense of noise or uncertainty at the input. We have also shown that the NSFLS (14) can uniformly approximate a continuous function on a compact set, i.e., it is a universal approximator.

The utility of our NSFLS and its superiority over a singleton FLS was demonstrated through our example on the prediction of a chaotic time series from noisy observations. In all cases, our NSFLS significantly outperformed a corresponding singleton FLS. Both were trained in the same manner, with identical common parameters, and were capable of closely approximating the time series in a noise-free environment.

In summary, we have found NSFLSs to be very promising and an effective tool for dealing with uncertainty, totally within the framework of fuzzy logic.

Appendix A. Derivation of Uniform Approximation Theorem

Proof: Uniform approximation requires that

$$\forall \epsilon > 0 \; |f_{ns}(\underline{x}) - g(\underline{x})| < \epsilon \; \forall \; \underline{x} \in U. \tag{A-1}$$

$f_{ns}(\underline{x})$ is given by (14) and represents the crisp output produced by the modified height method defuzzifier of the nonsingleton system. Since U is a compact subset of R^p, continuity of g implies

uniform continuity [23]; thus,

$$\forall\, \epsilon_1 > 0\, \exists\, \delta > 0 \;\ni\; |g(\underline{x}_r) - g(\underline{x}_s)| < \epsilon_1\, \forall\, \underline{x}_r, \underline{x}_s \in U \;\ni\; |\underline{x}_r - \underline{x}_s| < \delta. \tag{A-2}$$

Let $\{B_l\}$ be a family of open balls (open sets) of radius δ that covers U, and B_l' be the set of all limit points of the open set B_l. Since U is compact, any open cover of U contains a finite open subcover:

$$U \subset \bigcup_{l=1}^{q} B_l. \tag{A-3}$$

We construct such a subcover, comprised of open balls $B_{\underline{c}_l}$, centered at $\underline{c}_l$, such that the centers of the nearest neighbors of each open ball, $B_{\underline{c}_l}$, is a limit point of $B_{\underline{c}_l}$, but does not belong to $B_{\underline{c}_l}$, i.e.,

$$\underline{c}_l \in B'_{\underline{c}_l} \text{ and } \underline{c}_l \notin B_{\underline{c}_l}. \tag{A-4}$$

Then, assuming that U is connected, and $B_{\underline{c}_l}$ and $B_{\underline{c}_k}$ are nearest neighbors:

$$I_{lk} \equiv B_{\underline{c}_l} \bigcap B_{\underline{c}_k} \neq \emptyset \text{ for } l, k \;\leq\; q, \tag{A-5}$$

The image of the center of each open ball under g is given by $g(\underline{c}_l)$. Since g is a continuous function, the image of U under g, $g(U)$, in V will also be compact and connected, and

$$g(U) \subset \bigcup_{l=1}^{q} g(B_{\underline{c}_l}). \tag{A-6}$$

Select fuzzy sets $G^l \in V$ centered at $\bar{y}^l = g(\underline{c}_l)$ such that the set of points of support of G^l, S_{G^l}, covers $g(B_{\underline{c}_l} \bigcap U)$, and $\mu_{G^l}(\bar{y}^l) = 1$; hence, $\bigcup_{l=1}^{q} S_{G^l}$ covers $g(U)$. Let $\underline{x}_1 \in B_{\underline{c}_l}$, $\underline{x}_2 \in B_{\underline{c}_k}$ and

$\underline{x}_3 \in I_{lk}$; then, $|\underline{x}_1 - \underline{x}_3| < \delta$ and $|\underline{x}_3 - \underline{x}_2| < \delta$. From (A-2) and the triangular inequality:

$$|g(\underline{x}_1) - g(\underline{x}_2)| \leq |g(\underline{x}_1) - g(\underline{x}_3)| + |g(\underline{x}_3) - g(\underline{x}_2)| < \epsilon_1 + \epsilon_1 = \epsilon_2. \tag{A-7}$$

The distances of $\underline{x}_1$ from $\underline{c}_l$ and $\underline{x}_2$ from $\underline{c}_k$ are both bounded by δ; thus, $\underline{x}_1$ can come arbitrarily close to $\underline{c}_l$, and $\underline{x}_2$ can come arbitrarily close to $\underline{c}_k$. By letting $\underline{x}_1$ and $\underline{x}_2$ approach $\underline{c}_l$ and $\underline{c}_k$ respectively, then (A-7) bounds the difference of their corresponding images: $|g(\underline{c}_l) - g(\underline{c}_k)| = |\bar{y}^l - \bar{y}^k| < \epsilon_2$.

Any point $\underline{x} \in U$ that is not a center of a fuzzy set, belongs to some I_{lk}; therefore, the distance of its image from the image of the center $\underline{c}_l$ is given by (A-7): $|g(\underline{c}_l) - g(\underline{x})| < \epsilon_2$. Let this point $\underline{x}$ be the point $\underline{x}_{k,sup}$ at which the functions $\mu_{Q_k^l}$ in $\mathcal{T}_{k=1}^{p} \mu_{Q_k^l}(\underline{x}_{k,sup})$ in (14) are maximized.

Now suppose that $\underline{x}_{k,sup}$ activates $N \leq M$ rules. Then we have N individual centers, whose final output, according to the modified height method defuzzifier, lies somewhere inbetween, or exactly on, the centers of the sets Y^l, and, consequently, the centers of the sets G^l, by construction. This is ensured by the fact that, for any $\underline{x} \in U$, the sum of the coefficients (15) of each $\bar{y}^l$ in (14) equals 1, i.e., $\sum_{l=1}^{M} p_l(\underline{x}) = 1$. [Note that, from the properties of the most commonly used t-norms (algebraic product, Einstein product, bounded difference, and Hamacher product) and (12), we know that the output fuzzy set Y^l and the consequent fuzzy set of the lth rule, G^l, achieve their maximum at the same point $\bar{y}^l$. In the case of *minimum* as a t-norm, G^l is required to be symmetric]. We have already shown that the distance between centers G^l and G^k is bounded by ϵ_2; thus, the distance of the output $f_{ns}(\underline{x})$ produced by the nonsingleton fuzzy logic system (14) from the centroid of G^l is also bounded by ϵ_2, i.e., $|f_{ns}(\underline{x}) - g(\underline{c}_l)| < \epsilon_2$. Then, by the triangular inequality:

$$|f_{ns}(\underline{x}) - g(\underline{x})| \leq |f_{ns}(\underline{x}) - g(\underline{c}_l)| + |g(\underline{c}_l) - g(\underline{x})| < \epsilon_2 + \epsilon_2 = \epsilon \tag{A-8}$$

Thus, the nonsingleton fuzzy logic system $f_{ns}(\underline{x})$ in (14) uniformly approximates $g(\underline{x})$. □

The above proof contains the systems shown in [31] to be universal approximators as a special case. It can also be easily extended to nonsingleton and singleton additive FLSs [9, 11].

Acknowledgement

This work was performed at the University of Southern California under grant no. MIP-9122018 from the National Science Foundation.

References

[1] Balazinski, M., E. Czogala, and T. Sadowski, "Control of Metal-Cutting Process Using Neural Fuzzy Controller," *Second IEEE International Conference on Fuzzy Systems,* Vol. 1, pp. 161-166, 1993.

[2] Brown, R., P. Bryant, and H.D.I. Abarbanel, "Computing the Lyapunov Spectrum of a Dynamical System from an Observed Time Series," *Physical Review A,* vol. 43, no. 6, pp. 2787-2806, Mar. 1991.

[3] Casdagli, M., "A Dynamical Systems Approach to Modeling Input-Output Systems," *Nonlinear Modeling and Forecasting,* SFI Studies in the Sciences of Complexity Proc. vol. XII, pp. 265-281, Addison-Wesley, NY, 1992.

[4] Eckmann, J.P., S.O. Kamphorst, D. Ruelle, and S. Ciliberto, "Lyapunov Exponents from Time Series," *Physical Review A*, vol. 34, no. 6, pp. 4971-4979, Dec. 1986.

[5] Farmer, J.D., "Chaotic Attractors of an Infinite-Dimensional Dynamical System," *Physica,* 4D, pp. 366-393, 1982.

[6] Farmer, J.D., and J.J. Sidorowich, "Predicting Chaotic Time Series," *Physical Review Letters*, vol. 59, no. 8, pp. 845-848, Aug. 24, 1987.

[7] Filev, D.P., and R.R. Yager, "A Generalized Defuzzification Method via BAD Distributions," *International Journal of Intellligent Systems*, vol.6, pp. 687-697, 1991.

[8] Hayashi, Y., J. J. Buckley, and E. Czogala, "Fuzzy Neural Network with Fuzzy Signals and Weights," *Int. J. of Intelligent Systems*, vol. 8, 527-537, 1993.

[9] Hellendoorn, H., and C. Thomas, "Defuzzification in Fuzzy Controllers," *Journal of Intelligent Systems*, vol. 1, 109-123, 1993.

[10] Kosko, B., *Neural Networks and Fuzzy Systems*, Prentice Hall, Englewood Cliffs, NJ 1992.

[11] Kosko, B., "Fuzzy Systems as Universal Approximators," *IEEE Int. Conf. on Fuzzy Systems*, 1992.

[12] Kůrková, V., "Kolmogorov's Theorem is Relevant," *Neural Computation*, vol. 3, pp. 617-622, 1991.

[13] Kůrková, V., "Kolmogorov's Theorem and Multilayer Neural Networks," *Neural Networks*, vol. 5, pp. 501-506, 1992.

[14] Lapedes, A. and R. Farber, "Nonlinear signal processing using neural networks: prediction and system modeling," *LA-UR-87-2662*, 1987.

[15] Lee, C. C., "Fuzzy Logic in Control Systems: Fuzzy Logic Controller, Part II," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 20, no. 2, pp. 419-435, 1990.

[16] Lorentz, G.G., *Approximation of Functions*, Holt, Reinhart and Winston, New York, 1966.

[17] Mackey, M.C. and L. Glass, "Oscillation and Chaos in Phsiological Control Systems," *Science*, vol. 197, pp. 287-289, 1987.

[18] Muyaram, Y., T. Terano, S. Masui and N. Akiyama, "Optimizing Control of a Diesel Engine," in *Industrial Applications of Fuzzy Control*, (M. Sugeno, ed.) North-Holland, Amsterdam, pp. 63-71, 1992.

[19] Nychka, D.W., S. Ellner, D. McCaffrey and A.R. Gallant, "Finding Chaos in Noisy Systems," *Journal of the Royal Statistical Society*, Ser. B54, pp. 399-426, 1992.

[20] Parker, T.S. and L.O. Chua, "Practical Numerical Algorithms for Chaotic Systems," *Springer-Verlag*, New York, 1989.

[21] Pedrycz, W., "Design of Fuzzy Control Algorithms with the Aid of Fuzzy Models," in *Industrial Applications of Fuzzy Control*, (M. Sugeno, ed.) North-Holland, Amsterdam, pp. 139-151, 1992.

[22] Rasband, S. N., *Chaotic Dynamics of Nonlinear Systems*, Wiley-Interscience, NY 1990.

[23] Rudin, W., *Principles of Mathematical Analysis, Third Edition*, McGraw-Hill, Inc., NY 1976.

[24] Sakai, Y., and K. Ohkusa, "A Fuzzy Controller in Turning Process Automation," *Industrial Applications of Fuzzy Control*, (M. Sugeno, ed.) North-Holland, Amsterdam, pp. 139-151, 1992.

[25] Sprecher, D.A., "On the Strucure of Continuous Functions of Several Variables," *Transactions of the American Mathematical Society*, vol. 115, pp. 340-355.

[26] Tanaka, H., T. Okuda, and K. Asai, "Fuzzy Information and Decision in Statistical Model," in *Advances in Fuzzy Set Theory and Applications*, (M.M. Gupta, R.K. Ragade, and R.R. Yager, eds.) North-Holland, Amsterdam, pp. 303-320, 1979.

[27] Vemuri, V.R., and R.D. Roberts, (*eds.*), *Artificial Neural Networks: Forecasting Time Series*, IEEE Computer Society Press, CA 1994.

[28] Wang, L. X., "Analysis and Design of Fuzzy Systems," *PhD Dissertation*, University of Southern California, Los Angeles, CA, 1992.

[29] Wang, L. X., *Adaptive Fuzzy Systems and Control*, Prentice-Hall, Englewood-Cliffs, NJ, 1994.

[30] Wang, L. X. and J. M. Mendel, "Generating Fuzzy Rules from Numerical Data, with Applications," *USC-SIPI Report* No. 169, 1991. Also in *IEEE Trans. on Systems, Man, and Cybern.*, vol. 22, pp. 1414-1427, 1992.

[31] Wang, L. X. and J. M. Mendel, "Fuzzy Basis Functions, Universal Approximation, and Orthogonal Least Squares Learning," *IEEE Trans. on Neural Networks*, vol. 3, pp. 807-814, Sept. 1992.

[32] Wang, L. X. and J. M. Mendel, "Back-Propagation Fuzzy Systems as Non-Linear Dynamic System Identifiers," *Proc. IEEE International Conf. on Fuzzy Systems*, pp. 1409-1418, San Diego, 1992.

[33] Wolf, A., J.B. Swift, H.L. Swinney, and J.A. Vastano, " Determining Lyapunov Exponents from a Time Series," *Physica*, 16D, pp. 285-317, 1985.

[34] Zadeh, L. A., "Fuzzy Sets," *Information and Control*, vol. 8, pp. 338-353, 1965.

[35] Zadeh, L. A., "Outline of a New Approach to the Analysis of Complex Systems and Decision Processes," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 1, pp. 28-44, Jan. 1973.

[36] Zimmermann, H.J., *Fuzzy Set Theory And Its Applications, Second Edition*, Kluwer Academic Publishers, Boston, Mass., 1991.

Table 1: $x_{k,max}$ for Gaussian and triangular membership functions in the cases of algebraic Product and Minimum t-norms. (Due to space limitations, details of derivations will be given in a future publication.)

t-norms	$x_{k,max}$ for Gaussian membership functions	$x_{k,max}$ for triangular membership functions
Product	$\frac{\sigma^2_{X_k} m_{F^l_k} + \sigma^2_{F^l_k} m_{X_k}}{\sigma^2_{X_k} + \sigma^2_{F^l_k}}$	if $m_{X_k} < m_{F^l_k}$ and $\gamma_1 \in [m_{X_k}, m_{F^l_k}]$, then $x_{k,max} = \gamma_1$ where $\gamma_1 \equiv \frac{m_{X_k} + m_{F^l_k} - l_{F^l_k} + r_{X_k}}{2}$ if $l_{F^l_k} > r_{X_k}$ and $\gamma_1 \notin [m_{X_k}, m_{F^l_k}]$, then $x_{k,max} = m_{X_k}$ if $l_{F^l_k} < r_{X_k}$ and $\gamma_1 \notin [m_{X_k}, m_{F^l_k}]$, then $x_{k,max} = m_{F^l_k}$, if $m_{X_k} > m_{F^l_k}$ and $\gamma_2 \in [m_{X_k}, m_{F^l_k}]$, then $x_{k,max} = \gamma_2$ where $\gamma_2 \equiv \frac{m_{F^l_k} + m_{X_k} - l_{X_k} + r_{F^l_k}}{2}$ if $l_{X_k} > r_{F^l_k}$ and $\gamma_2 \notin [m_{F^l_k}, m_{X_k}]$, then $x_{k,max} = m_{F^l_k}$ if $l_{X_k} < r_{F^l_k}$ and $\gamma_2 \notin [m_{F^l_k}, m_{X_k}]$, then $x_{k,max} = m_{X_k}$ if $m_{X_k} = m_{F^l_k} = m$, then $x_{k,max} = m$
Minimum	$\frac{\sigma_{X_k} m_{F^l_k} + \sigma_{F^l_k} m_{X_k}}{\sigma_{X_k} + \sigma_{F^l_k}}$ if $\sigma_{F^l_k} \geq \sigma_{X_k}$ and $m_{F^l_k} \geq m_{X_k}$, or $\sigma_{F^l_k} < \sigma_{X_k}$ and $m_{F^l_k} < m_{X_k}$ $\frac{\sigma_{F^l_k} m_{X_k} - \sigma_{X_k} m_{F^l_k}}{\sigma_{F^l_k} - \sigma_{X_k}}$ if $\sigma_{F^l_k} > \sigma_{X_k}$ and $m_{F^l_k} < m_{X_k}$, or $\sigma_{F^l_k} < \sigma_{X_k}$ and $m_{F^l_k} > m_{X_k}$	$\frac{l_{F^l_k} m_{X_k} + r_{X_k} m_{F^l_k}}{l_{F^l_k} + r_{X_k}}$ if $m_{X_k} \leq m_{F^l_k}$ $\frac{l_{X_k} m_{F^l_k} + r_{F^l_k} m_{X_k}}{l_{X_k} + r_{F^l_k}}$ if $m_{X_k} \geq m_{F^l_k}$

Table 2: Average standard deviations and average mean-squared errors for different signal-to-noise ratios, for singleton and nonsingleton systems. In the noise-free case, the average standard deviation and mean-squared error were 0.0168 and 2.8329×10^{-4}, respectively, for both systems.

SNR	-5dB	0dB	10dB
Singleton FLS, avg. std	0.5517	0.3308	0.1281
Nonsingleton FLS, avg. std	0.2128	0.1496	0.0765
Singleton FLS avg. mse	0.3066	0.1101	0.0165
Nonsingleton FLS avg. mse	0.0462	0.0228	0.0059

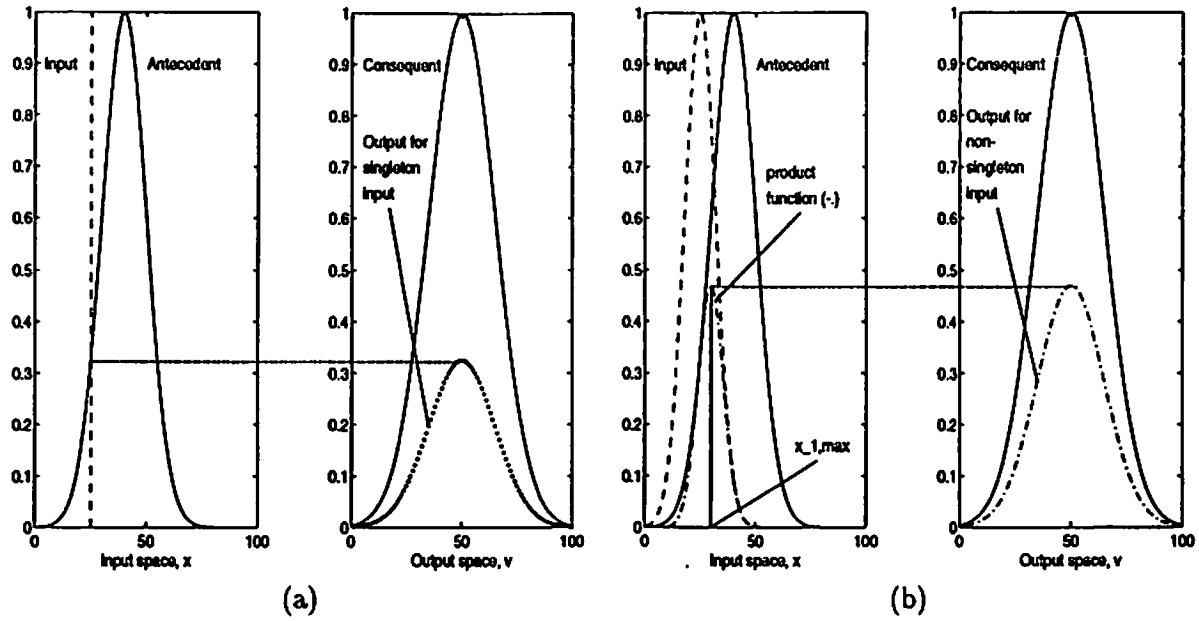


Figure 1: a) SISO FLS: Composition of rule with singleton input, and corresponding output. b) SISO NSFLS: Composition of rule with Gaussian input, and corresponding output. This figure also illustrates Eq. (25) (with M=1, p=1, $m_{F_1^1} = 40$, $\sigma_{F_1^1} = 10$, $x_1 = 25$, $\sigma_X = 7$, $m_{G^1} = 50$, $\sigma_{G^1} = 15$) for Gaussian membership functions. Equation (25) reduces to (18) if the input is singleton, with '$\star$' replaced by product.

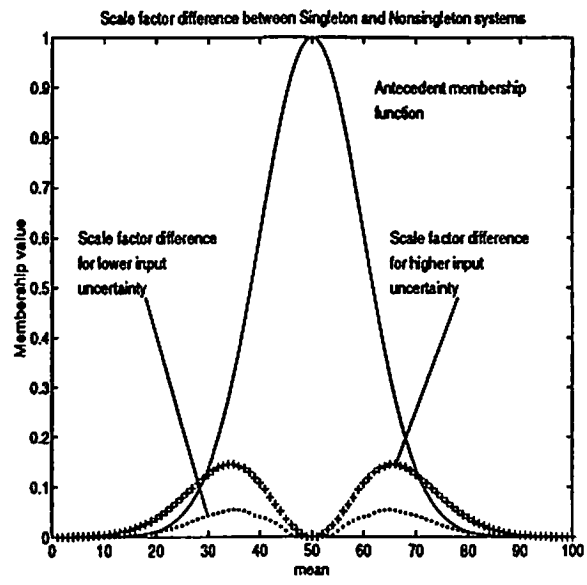


Figure 2: Scale factor difference between SISO singleton and nonsingleton systems, over the entire input range for one antecedent. When the input singleton and the mean of the corresponding nonsingleton are at the mean of the antecedent, the scale factor difference is zero. For higher levels of uncertainty in the input (represented by larger variance in the Gaussian nonsingleton input), the scale factor difference is larger. This difference in scale factors is also evident in Fig. 1 for one particular case of singleton and nonsingleton inputs. The horizontal axis represents the means of the Gaussian input in the nonsingleton case, and the location of the singleton input in the singleton case.

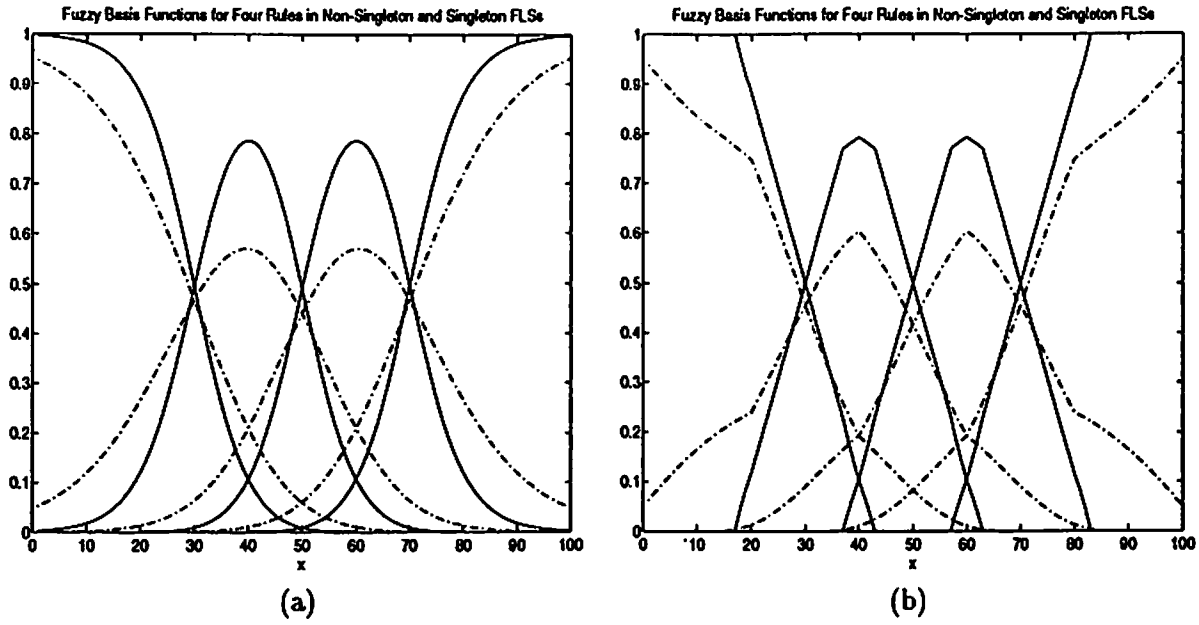


Figure 3: (a) Fuzzy basis functions for nonsingleton (dash-dotted line) and singleton fuzzy logic systems with four rules, for Gaussian membership functions and product inference. (b) Fuzzy basis functions for nonsingleton (dash-dotted line) and singleton fuzzy logic systems with four rules, for triangular membership functions and product inference. The antecedent membership functions were chosen to be the same in the NSFBFs and singleton FBFs.

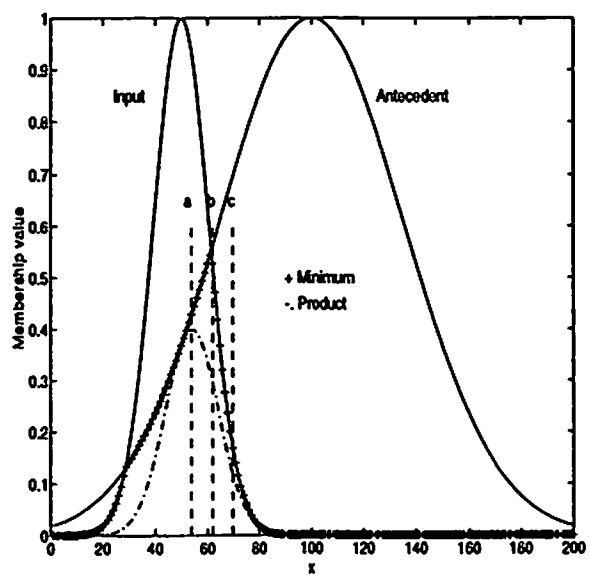


Figure 4: Membership functions of nonsingleton fuzzy input, $\mu_{X_k}(x_{k,i_k})$, and corresponding antecedent, $\mu_{F_k^l}(x_{k,i_k})$, their product and minimum.

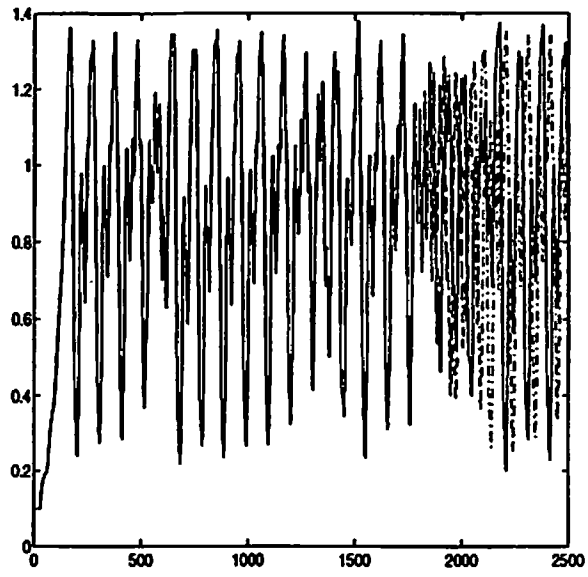


Figure 5: Sample waveforms of the Mackey-Glass chaotic time series ($\tau = 30$) for two different, but very close initial conditions (all values in the initial value interval are set to 0.1 for the first waveform, and to 0.100001 for the second waveform). Observe that after a certain point the two realizations evolve completely differently.

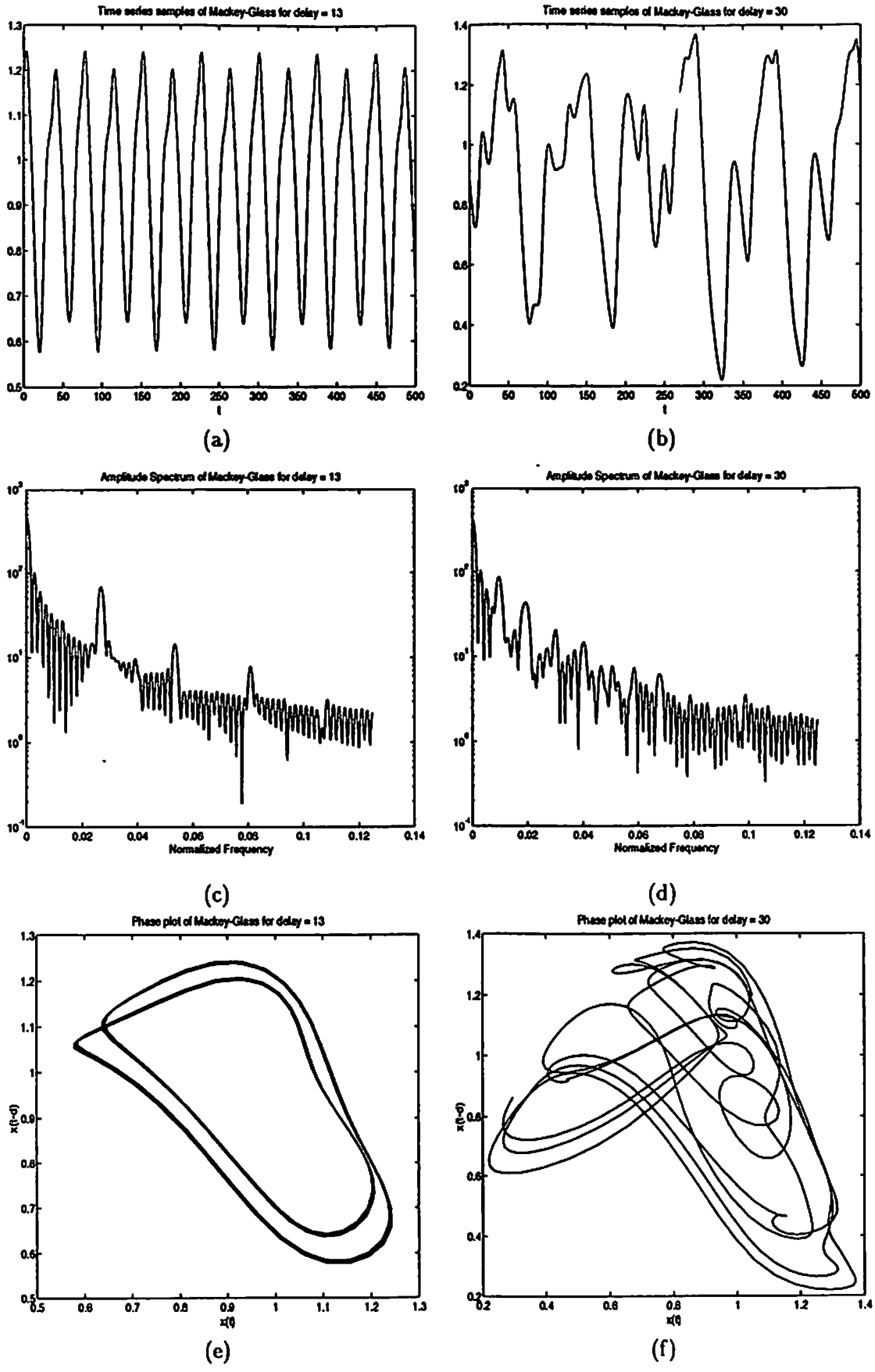


Figure 6: (a)-(b) Representative samples of the Mackey-Glass time series, after letting transients relax. (c)-(d) Amplitude spectra plotted on a semilog scale, for the examples illustrated in (a) and (b). (e)-(f) Corresponding phase plots of the time series segments depicted in (a) and (b). ('d' in x(t-d) on the vertical axis denotes the delay used in the Mackey-Glass equation; it is either 13 or 30).

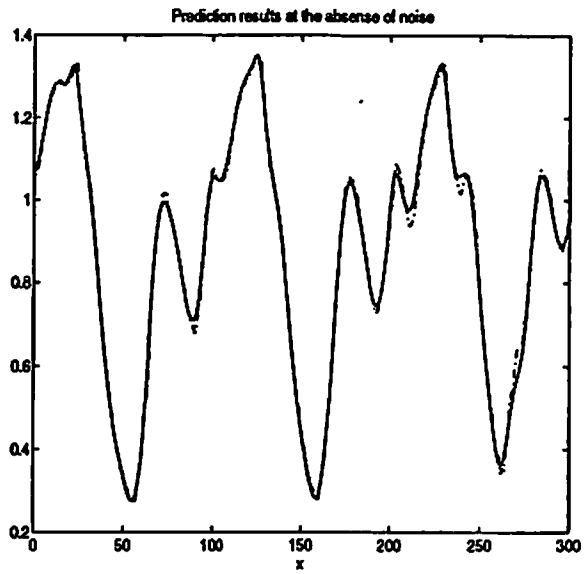


Figure 7: Prediction results in the absense of noise. The solid line is the actual time series, whereas the dash-dotted line shows the predicted estimates. In this case, the NSFLS and the singleton FLS produce identical results. The same parameters of the singleton FLS used here, are also used in the subsequent cases when noise is added. All the common parameters of the NSFLS and the singleton FLS are identical.

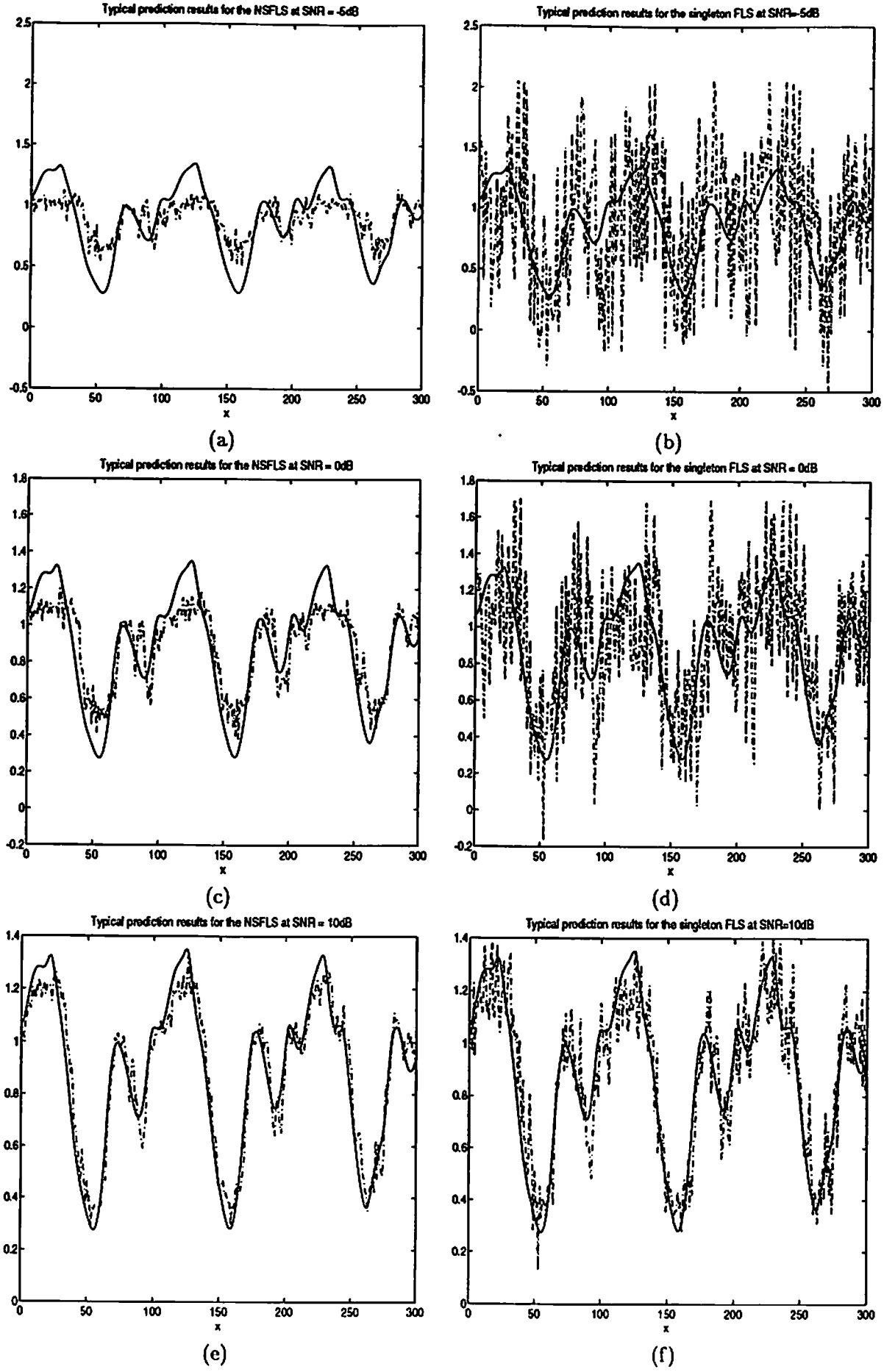


Figure 8: Typical prediction results (single realization) for nonsingleton and singleton fuzzy logic systems, for different signal to noise ratios.

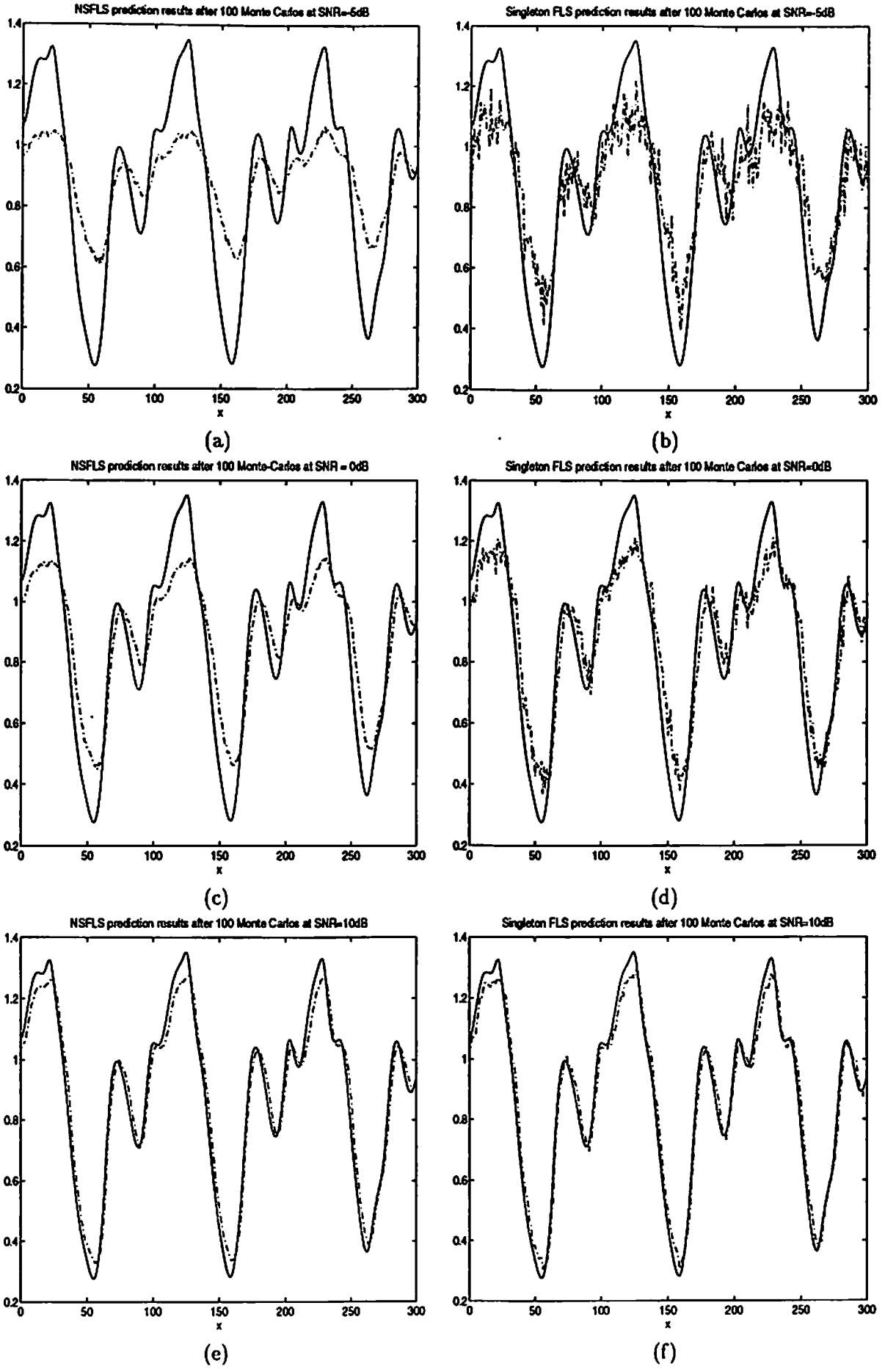
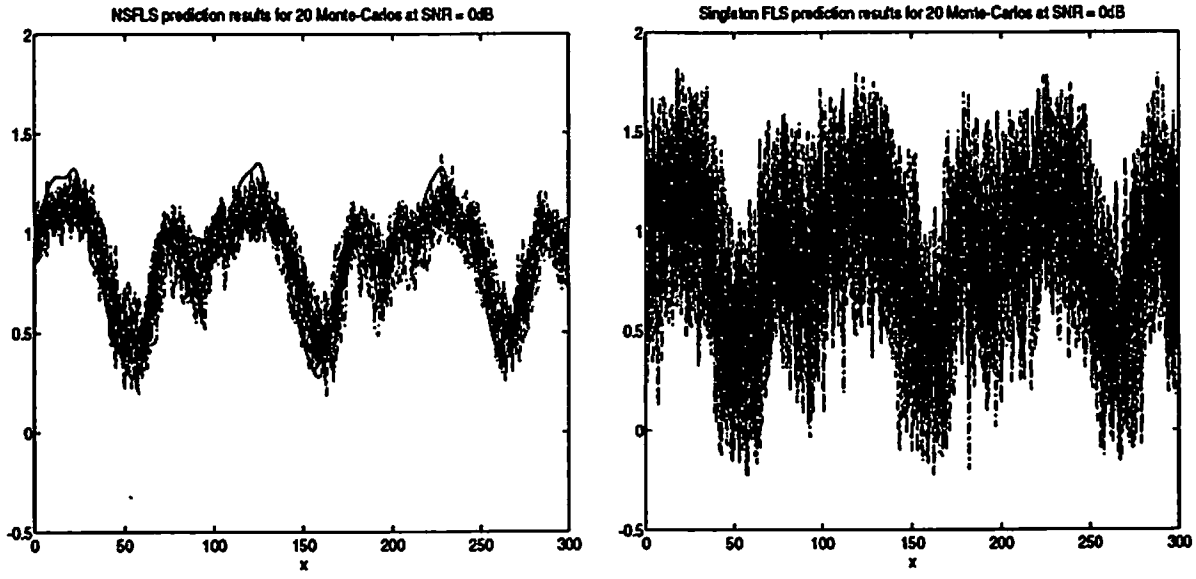


Figure 9: Averaged prediction results for 100 Monte Carlo runs. In all cases, the mean squared errors and the corresponding standard deviations are significantly smaller in the nonsingleton case (see Table 2).



(a) (b)

Figure 10: a) Prediction results for the NSFLS, for twenty different noise realizations, at 0dB. b) Prediction results for the singleton FLS, for twenty different noise realizations, at 0dB. These figures show that the standard deviation of the prediction results is much higher in the singleton case than in the nonsingleton case.