An Introduction to Parameterized IFAM Models with Applications in Prediction

Peter Sussner[1] Rodolfo Miyasaki[2] Marcos Eduardo Valle[3]

[1]Department of Applied Mathematics, University of Campinas, Campinas - SP, Brazil

[2]Credit Planning and Management Department, Banco Itaú, São Paulo - SP, Brazil

[3]Department of Mathematics, University of Londrina, Londrina - PR, Brazil

2009 IFSA World Congress
2009 EUSFLAT Conference

Subject of this Talk

Implicative fuzzy associative memories (IFAMs), can be used to implement fuzzy rule-based systems. There are infinitely many IFAMs.

The problem:

Selecting the best IFAM for a given application.

Proposed solution:

We introduce a class of parameterized IFAMs and formulate an optimization problem that yields the IFAM that best fits the given data.

We compare the performance of the resulting Yager IFAM with other techniques in two problems in time series prediction.

Organization of this talk

1. A Brief Introduction to Implicative Fuzzy Associative Memories
2. The Yager Class of Parameterized IFAMs
3. Applications of Yager IFAMs in Prediction
4. Concluding Remarks and Suggestions for Further Research

Fuzzy Associative Memories (FAMs)

Fuzzy Associative Memories (FAMs):

Fuzzy systems that encode associations $(\mathbf{x}^1, \mathbf{y}^1), \ldots, (\mathbf{x}^p, \mathbf{y}^p)$, where $\mathbf{x}^\xi$ and $\mathbf{y}^\xi$ are fuzzy sets.

Max-T FAMs:

Given an fuzzy input pattern $\mathbf{x}$, the output is given by

$$\mathbf{y} = \mathcal{W}(\mathbf{x}) = (W \circ \mathbf{x}) \vee \boldsymbol{\theta}\,.$$

where W is the synaptic weight matrix and $\boldsymbol{\theta}$ is the threshold vector.

Max-T Product:

$$C = A \circ B \quad \Leftrightarrow \quad c_{ij} = \bigvee_{\xi=1}^{k} T(a_{i\xi}, b_{\xi j})\,.$$

Implicative Fuzzy Associative Memories (IFAMs)

IFAMs = Max-T FAM + Implicative Fuzzy Learning

Implicative Fuzzy Learning (IFL):

Given associations $(\mathbf{x}^1, \mathbf{y}^1), \ldots, (\mathbf{x}^p, \mathbf{y}^p)$, define W and $\boldsymbol{\theta}$ as the largest synaptic weight matrix and threshold vector such that $\mathcal{W}(\mathbf{x}^\xi) \leq \mathbf{y}^\xi$:

$$[W, \boldsymbol{\theta}] = \bigvee \left\{ [A, \boldsymbol{\beta}] : (A \circ \mathbf{x}^\xi) \vee \boldsymbol{\beta} \leq \mathbf{y}^\xi, \ \forall \xi = 1, \ldots, p \right\}.$$

If T is a continuous t-norm, then W and $\boldsymbol{\theta}$ can be easily computed as

$$w_{ij} = \bigwedge_{\xi=1}^{p} I_T(x_j^\xi, y_i^\xi) \quad \text{and} \quad \theta_i = \bigwedge_{\xi=1}^{p} y_i^\xi,$$

where I_T is the R-implication associated with T.

Some Examples of IFAMs

Gödel IFAM:

$$T_M(x,y) = x \wedge y \qquad \text{and} \qquad I_M(x,y) = \begin{cases} 1, & x \leq y\,, \\ y, & x > y\,. \end{cases}$$

Lukasiewicz IFAM:

$$T_L(x,y) = 0 \vee (x + y - 1) \qquad \text{and} \qquad I_L(x,y) = 1 \wedge (y - x + 1)\,.$$

Remark:

The drastic t-norm

$$T_D(x,y) = \begin{cases} x \wedge y, & x \vee y = 1\,, \\ 0, & \text{otherwise.} \end{cases}$$

is not left-continous. I_D is the greatest R-implication.

The classes of Yager t-norms and R-implications

Definition (Yager t-norm)

$$T^d(x,y) = 1 - \left\{1 \wedge \left[(1-x)^d + (1-y)^d\right]^{\frac{1}{d}}\right\}, \quad d > 0\,.$$

Definition (Yager R-implication)

$$I^d(x,y) = 1 - \left\{0 \vee \left[(1-y)^d - (1-x)^d\right]\right\}^{\frac{1}{d}}, \quad d > 0\,.$$

Remark:

- For $d = 1$, $T^d = T_L$ (Lukasiewicz t-norm);
- For $d \to 0$, $T_d \to T_D$ (drastic t-norm);
- For $d \to \infty$, $T_d \to T_M$ (minimum).

Yager IFAMs

Definition (Yager IFAM = IFAM based on T^d and I^d)

Consider a set of associations $\{(\mathbf{x}^1, \mathbf{y}^1), \ldots, (\mathbf{x}^p, \mathbf{y}^p)\}$.
Given an input pattern $\mathbf{x}$, the output is given by

$$\mathbf{y} = \mathcal{W}^d(\mathbf{x}) = (W^d \circ^d \mathbf{x}) \vee \boldsymbol{\theta}\,,$$

where the synaptic weight matrix and threshold vectors are

$$w_{ij}^d = \bigwedge_{\xi=1}^{p} I^d(x_j^\xi, y_i^\xi) \qquad \text{and} \qquad \theta_i = \bigwedge_{\xi=1}^{p} y_i^\xi\,.$$

Choosing the best Yager IFAM...

In the prediction problems that we will discuss, we have that

- The input fuzzy set $\mathbf{x}$ is derived from a real-valued vector $\mathbf{v}$;
- The output fuzzy set $\mathbf{y}$ is defuzzified yielding a real-valued output

$$s = \text{defuzz}(\mathbf{y}) \,.$$

Choosing the best Yager IFAM...

Given training patterns in the form $(\mathbf{v}^1, s_1), \ldots, (\mathbf{v}^p, s_p)$, derive fuzzy sets $\mathbf{x}^\xi$ and $\mathbf{y}^\xi$, and determine $d > 0$ that minimize the Euclidean distance between s_ξ and

$$s_\xi^d = \text{defuzz}\Big(\mathcal{W}^d(\mathbf{x}^\xi)\Big) \,,$$

produced by the Yager IFAM for $\xi = 1, \ldots, p$.

The Optimization Problem...

Optimization problem:

Given training patterns in the form $(\mathbf{v}^1, s_1), \ldots, (\mathbf{v}^p, s_p)$, solve

$$\begin{cases} \text{minimize} & \left\|\mathbf{s} - \mathbf{s}^d\right\|_2 \\ \text{subject to} & d > 0 \end{cases}$$

where $\mathbf{s} = [s_1, \ldots, s_p]^T$ and $\mathbf{s}^d = [s_1^d, \ldots, s_p^d]^T$.

Remark:

We used the routine FMINBND of MATLAB's Optimization Toolbox to determine a local minimum of the real-valued objective function

$$f(d) = \|\mathbf{s} - \mathbf{s}^d\|_2 .$$

Prediction of Manpower Requirement

Problem (Choudhury et al., 2002):

Predict the manpower requirement in steel manufacturing industry in the state of West Bengal, India.

We have linguistic values A_i, $i = 1, \ldots, 5$, and fuzzy rules such as

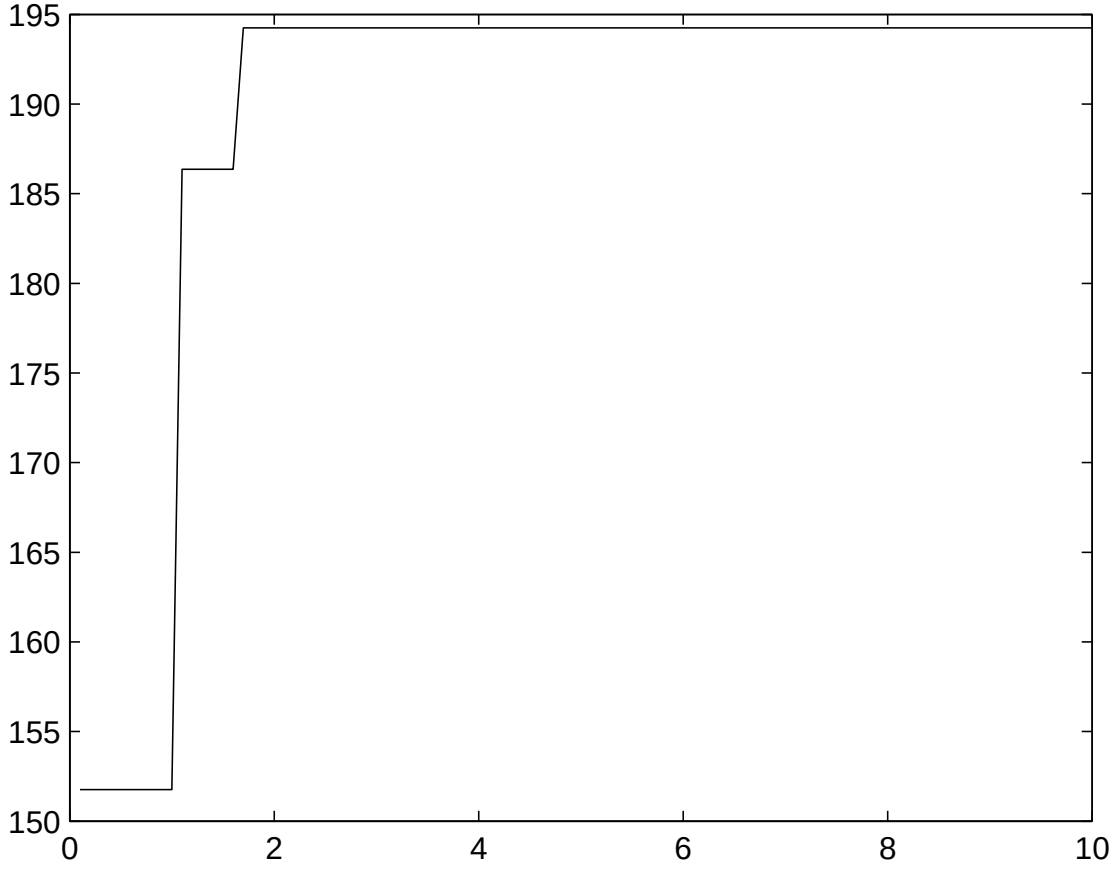
If manpower of year n is A_i, then that of year $n+1$ is A_j.

If $\mathcal{W}$ denotes a max-T FAM, the predicted manpower of year $n+1$ is

$$s_{n+1}^d = \text{defuzz}\left(\mathcal{W}^d(A_n)\right),$$

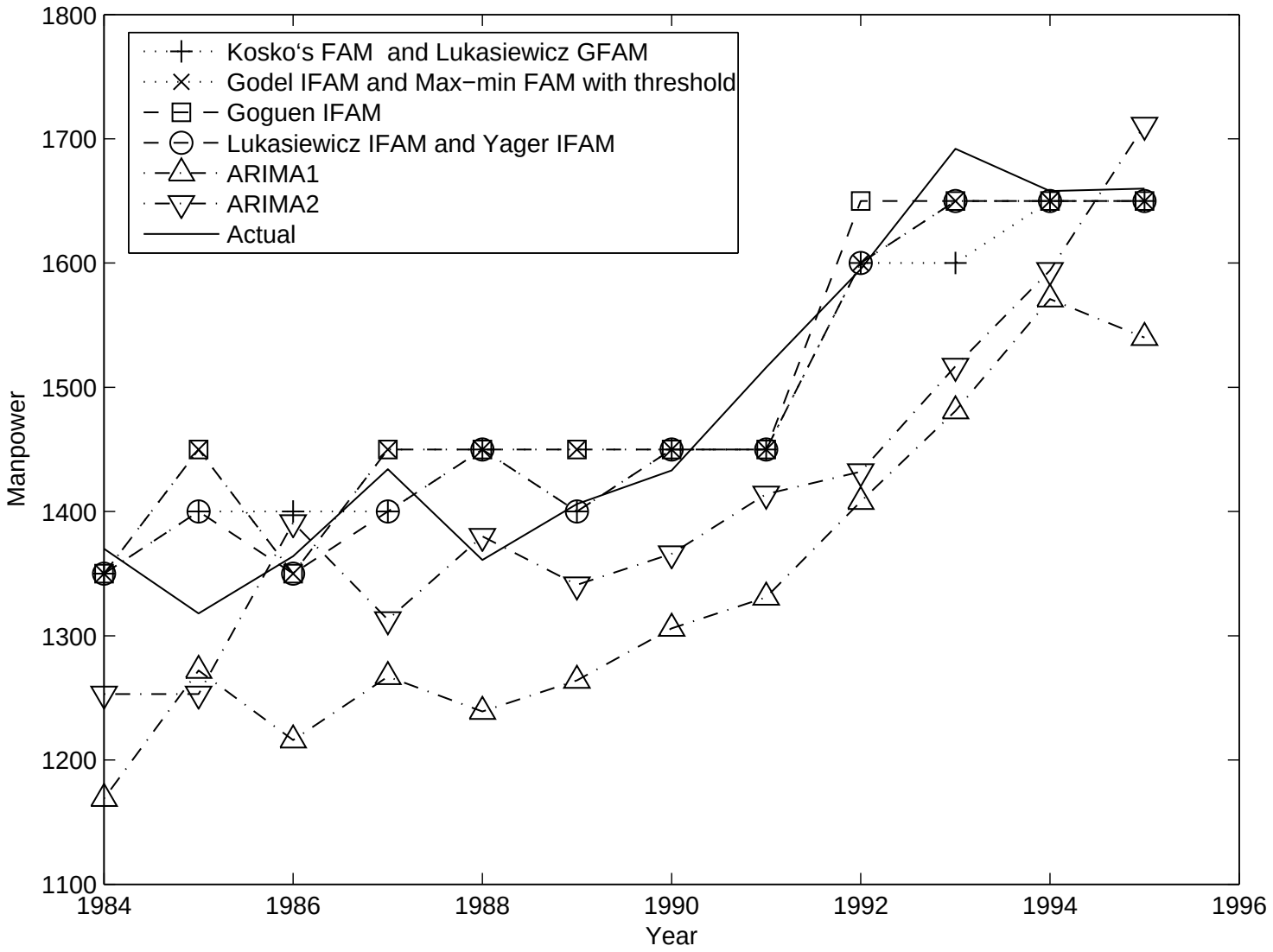
where defuzz$(\cdot)$ is the "mean of maxima" (MOM) and A_n is obtained fuzzifying the manpower requirement of year n.

Plot of the Objective Function for this Problem



The distance $\left\|\mathbf{s} - \mathbf{s}^d\right\|_2$ between the actual demand of manpower and that predicted by the Yager IFAM as a function of d.

Comparison in Forecasting Manpower



Comparison of FAMs and Statistical Models

Method	MSE ($\times 10^5$)	MAE (m^3/s)	MPE (%)
Yager IFAM ($d \in (0,1]$):	1.92	32.75	2.29
Lukasiewicz IFAM:	1.92	32.75	2.29
Kosko's FAM:	2.57	38.75	2.67
Lukasiewicz GFAM:	2.57	38.75	2.67
Gödel IFAM:	2.89	38.58	2.73
Max-min FAM/threshold:	2.89	38.58	2.73
Goguen IFAM:	3.14	42.75	2.99
ARIMA2	9.36	83.55	5.48
ARIMA1	23.26	145.25	9.79

Remarks:

This experiment indicates the utility of IFAMs for prediction problem and the validity of our approach for determining the best Yager IFAM.

Disadvantage:

This prediction problem does not include any test data!

Details can be found in:
J.P. Choudhury, B. Sarkar, and S.K. Mukherjee. Forecasting of Engineering Manpower Through Fuzzy Associative Memory Neural Network with ARIMA: a Comparative Study, *Neurocomputing*, 47:241-257, 2002.

Prediction of Streamflow of a Hydroeletric Plant

Problem (Magalhães et al., 2004):

Forecast the monthly streamflow of a large hydroelectric plant, called Furnas.

The predictions are used for simulation, optimization, and decision-making purposes of the energy system.

A time series prediction problem can be formulated as follows:

Given samples s_μ for $\mu = 1, \ldots, q-1$, obtain an estimate $\hat{s}_q$ for the actual s_q based on a subset of the past values $s_1, \ldots, s_{q-1}$.

Description of the Approach

Our approach uses:

- 12 different models, one for each month of the year, due to the seasonality of the streamflow;
- a fixed number of three antecedents, e.g., the values of January, February, and March to predict the streamflow of April.

How to Predict a Time Series using a FAM Model

FAMs can used as follows to predict a time series:

1. Define fuzzy sets $\mathbf{x}^\xi$ and $\mathbf{y}^\xi$ that comprise some relevant information concerning the past values of the time series.
2. Store the associations $(\mathbf{x}^\xi, \mathbf{y}^\xi)$ in the memory;
3. Given an input $\mathbf{x}^\gamma$ that takes into account some past values, compute the output pattern $\mathbf{y}^\gamma$;
4. A defuzzification of $\mathcal{W}(\mathbf{x}^\gamma)$ yields the estimation $\hat{s}_\gamma$.

How to Predict a Time Series using a FAM Model

These fours steps were performed as follows:

1. We used the subtractive clustering method to determine fuzzy sets $\mathbf{x}^{\xi}$ and $\mathbf{y}^{\xi}$ with Gaussian-type membership functions from streamflow data from 1931 to 1990;
2. We stored the associations in a Yager IFAM $\mathcal{W}^d$.
3. The input pattern $\mathbf{x}^{\gamma}$ corresponds to a "crisp" set;
4. A defuzzification of $\mathcal{W}^d(\mathbf{x}^{\gamma})$ using the centroid method yields $\hat{s}^d_{\gamma}$.

How to Predict a Time Series using a FAM Model

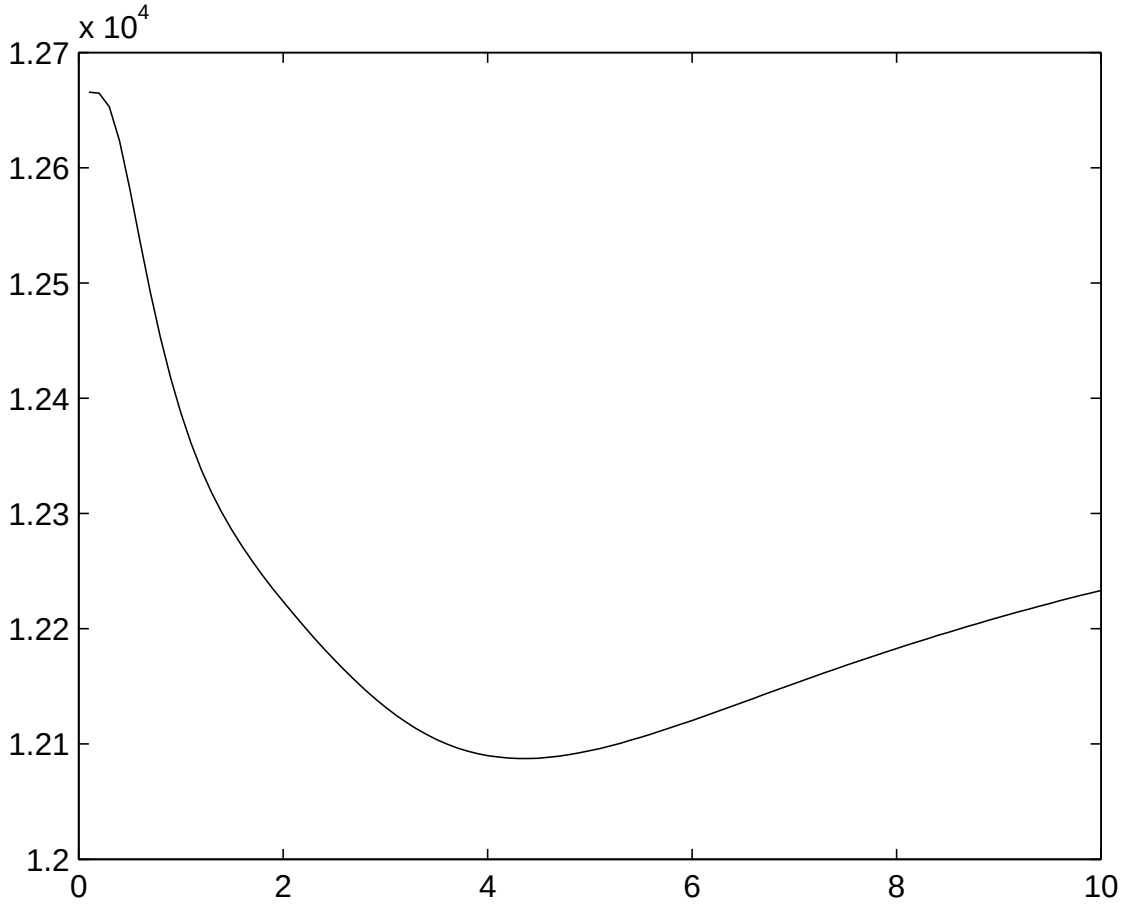
These fours steps were performed as follows:

1. We used the subtractive clustering method to determine fuzzy sets $\mathbf{x}^{\xi}$ and $\mathbf{y}^{\xi}$ with Gaussian-type membership functions from streamflow data from 1931 to 1990;
2. We stored the associations in a Yager IFAM $\mathcal{W}^d$.
3. The input pattern $\mathbf{x}^{\gamma}$ corresponds to a "crisp" set;
4. A defuzzification of $\mathcal{W}^d(\mathbf{x}^{\gamma})$ using the centroid method yields $\hat{s}^d_{\gamma}$.

Choosing the best Yager IFAM:

As before, we generated the vectors $\mathbf{s}$ and $\mathbf{s}^d$ using the training from 1931-1990 to determine the best Yager IFAM for this problem.

Plot of the Objective Function for this Problem



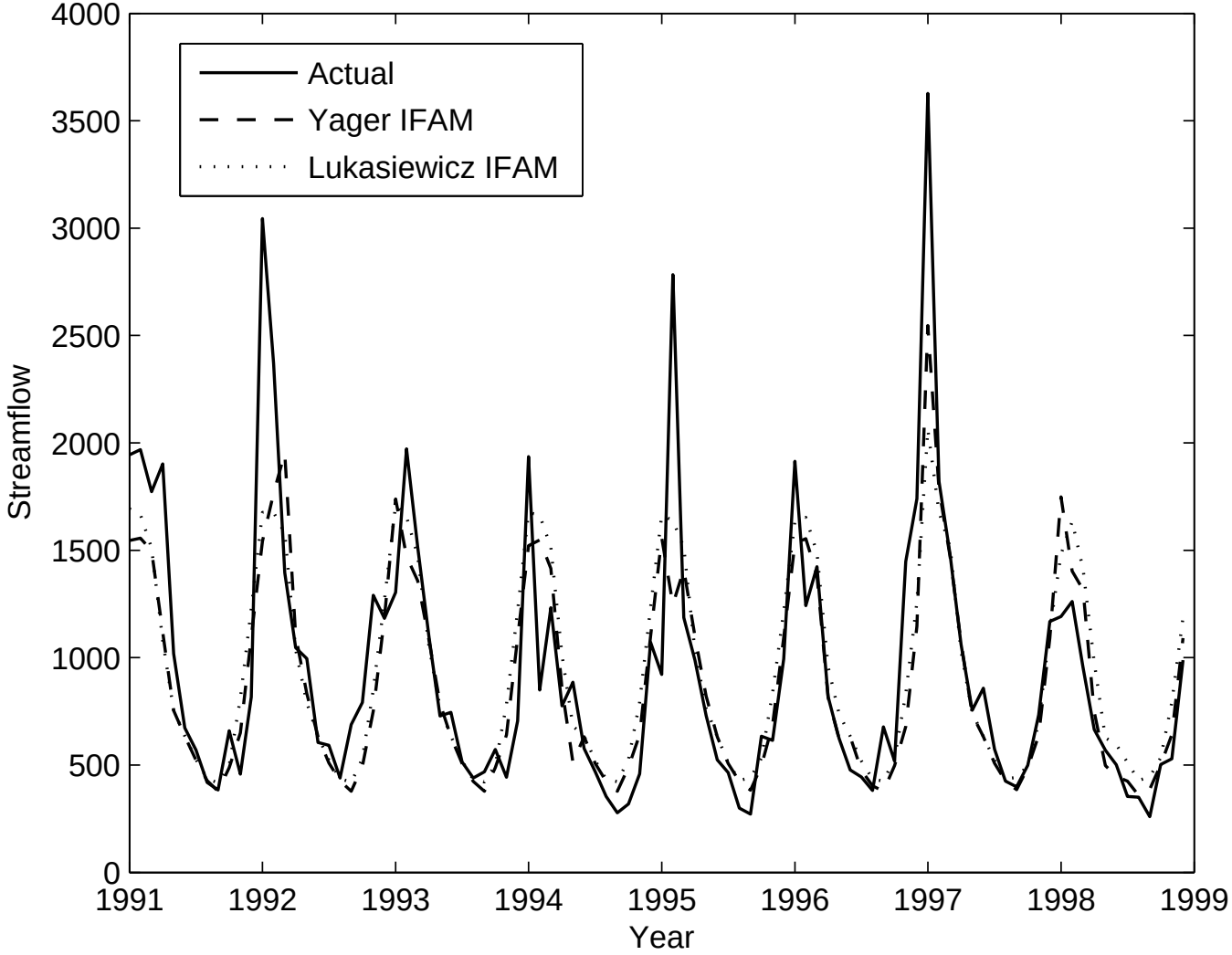
d	f(d)
4.35	1.20×10^4
1	1.23×10^4

Difference ≈ 300.

The distance $\left\|\mathbf{s} - \mathbf{s}^d\right\|_2$ between the actual streamflows and that predicted by the Yager IFAM as a function of d.

The Streamflow Prediction

for the Furnas reservoir from 1991 to 1998.



Comparison of Prediction Models

Method	MSE ($\times 10^5$)	MAE (m^3/s)	MPE (%)
Predictive fuzzy clustering	1.20	200	18
Yager IFAM ($d = 4.3568$)	1.28	216	21
Lukasiewicz IFAM	1.27	229	24
PARMA	1.85	280	28
Multi-layer perceptron	1.82	271	30
Neuro-fuzzy network	1.73	234	20

PARMA - Periodic auto-regressive moving average model

Remarks:

Some of the other methods were initialized by optimizing the number of parameters for each monthly prediction.

The predictive fuzzy clustering method also takes advantages of slopes information.

The Yager IFAM resulting from the minimization of MSE produced very satisfactory predictions.

The Yager IFAM outperformed the Lukasiewicz IFAM in terms of MAE and MPE but produced an slightly higher MSE.

This fact may be due to overfitting to the training data.

Concluding Remarks

This talk described:

First attempt of tackling the problem of selecting the best IFAM for a given application;

Introduction of the the class of Yager parameterized IFAMs;

Formulation of optimization problem for determining the best IFAM;

Application of our approach to two prediction problems from the literature.

Suggestions for Further Research

Incorporate additional parameters into the optimization process;

Use validation and/or regularization techniques to avoid overfitting;

Investigate other classes of parameterized IFAMs.

Thank You!