Non-linear Measure Based Process Monitoring and Fault Diagnosis

12th Annual AIChE Meeting, Reno, NV

[275] Data Driven Approaches to Process Control

4:40 PM, Nov. 6, 2001

Sandeep Rajput

Duane D. Bruns

Department of Chemical Engineering,
The University of Tennessee, Knoxville

C. Stuart Daw

Engineering Technology Division [!?],
Oak Ridge National Laboratory

Outline

1. Introduction and motivation
2. Review of popular approaches
3. Nonlinear measures and symbolization
4. Methodology
5. Results
6. Discussion
7. Future Directions

Introduction

- ✓Complex processes
- ✓Need to know the ‘process state’ for
  - Improved control and monitoring
  - Fault Diagnosis
  - Event Detection/Predicting failures

GOAL

Design Expert System to predict ‘process state’ online

Motivation

- ✓ Most popular methods assume linear structure in data

- ✓ Need to use nonlinear measures for nonlinear and chaotic processes
- ✓ Difference choices for
  - Feature vector formation
  - Measures of similarity

- ✓ **This is a study to investigate the potential of nonlinear measures in helping distinguish different nonlinear processes, with obvious extensions to monitoring and fault diagnosis**

Popular Approaches

- ✓ Form a feature vector that sufficiently characterizes the process state

Choices for feature vectors

- ✓ Power spectral density
- ✓ Wavelet coefficients
- ✓ Vector made by linear or nonlinear model coefficients

Classification Schemes

- ✓ Clustering
- ✓ Artificial Intelligence
- ✓ Multivariate Statistical Methods

Nonlinear measures

Correlation sum-based measures: characterize the overall geometry

$$C^m(X, X, \varepsilon) = \frac{1}{\binom{N}{2}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \Theta(\varepsilon - \| x_i - x_j \|)$$

The superscript m refers to the (embedding) dimension of x_i and x_j.

$x_i, x_j \in R^m$ and $x_i, x_j \subset X$

Θ is a kernel function, usually taken as the heaviside step function, gaussian or a radial basis function.

The vectors x_i and x_j are produced by juxtaposing *m* equally spaced measurements

✓Statistics based on correlation sum or their variants

✓Very Tedious

✓Inaccurate in presence of noise

Nonlinear measures (contd.)

✓ **Measures for characterize the dynamics**

- *Shannon Entropy* as a measure of redundancy of information

$$H_1(X,Y) = -\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} p_{i,j}(X,Y) \ln p_{i,j}(X,Y)$$

- Nonlinear autocorrelation or *mutual information*

$$I_i(X,Y) = H_i(X,X) + H_i(Y,Y) - H_i(X,Y)$$

Measures computed on code series generated by symbolization.

Using fluidized bed data for illustrating the method

Symbolization

- Look for patterns of values (measurements)

Symbolization Analysis

- Symbolization tries to assign a symbol value to every measurement
- Every measurement can be called a 0, 1, 2, …
- Helps reduce the noise
- Preserves the necessary information in less storage space
- Data compression techniques can be used from information theory
- Symbol sequences can be made by juxtaposing symbols

Symbolization

Symbolization Parameters

•Set Size	Number of distinct symbols
•Sequence Length	Number of symbols in the sequence
•Symbolization Interval	Time difference between symbols in the sequence

The optimum values of parameters have to be found by trial and error

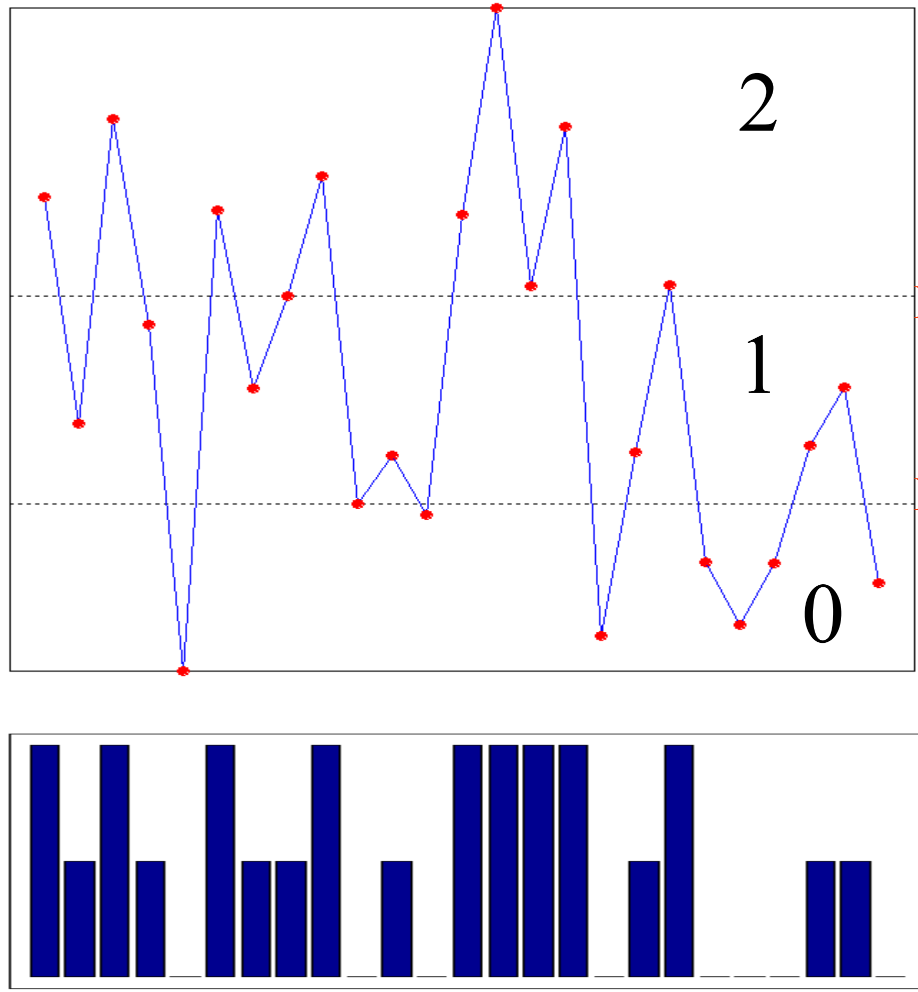
Symbolization parameters used in this study

- •Set Size = 3
- •Sequence Length = 5
- •Symbolization Interval = 1

Illustrating Symbolization

Set Size=3
Seq. Length =4
Sym. Interval=1

Multiplier	27	9	3	1	Code
	2	**1**	**2**	**1**	**70**
	1	**2**	**1**	**0**	**48**
Symbol Sequences	**2**	**1**	**0**	**2**	**65**
	1	0	2	1	35
	0	2	1	1	22
	2	1	1	2	68
	1	1	2	0	42

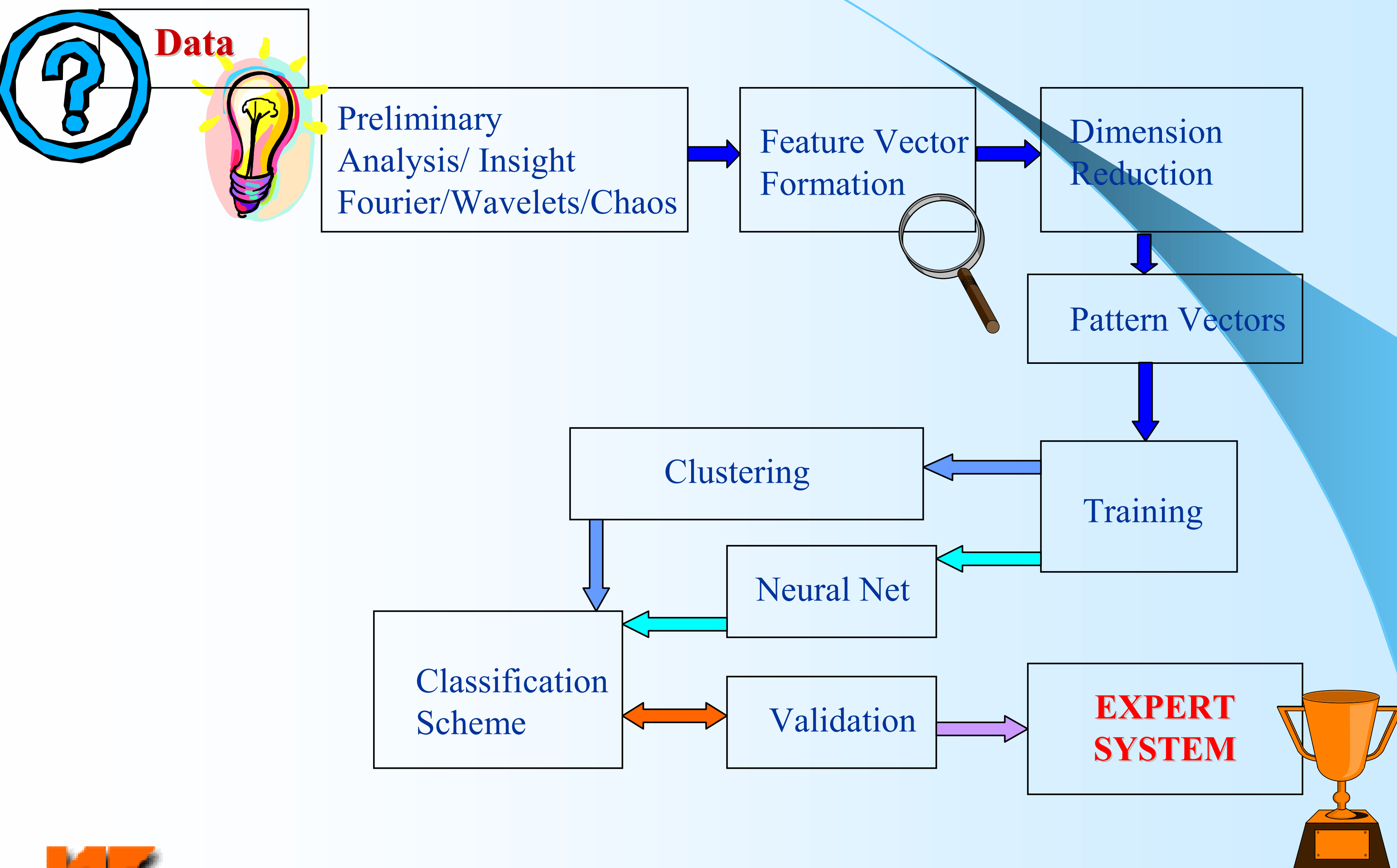


Code Series 2 1 2 1 0 2 1 1 2 0 1 0 2 2 2 2 0 1 2 0 0 0 1 1

70 48 65

Rajput-Bruns-Daw

General Methodology



Rajput-Bruns-Daw

Feature dimension reduction

Fischer's information criterion

Discriminatory Power of each feature

$$J_{f_k}(i,j) = J_{f_k}(j,i) = \frac{\| \mu_{i,f_k} - \mu_{j,f_k} \|^2}{\sigma_{i,f_k}{}^2 + \sigma_{j,f_k}{}^2} \quad for\ k = 1,2,..L$$

Choose features with highest discriminatory power

Expert systems

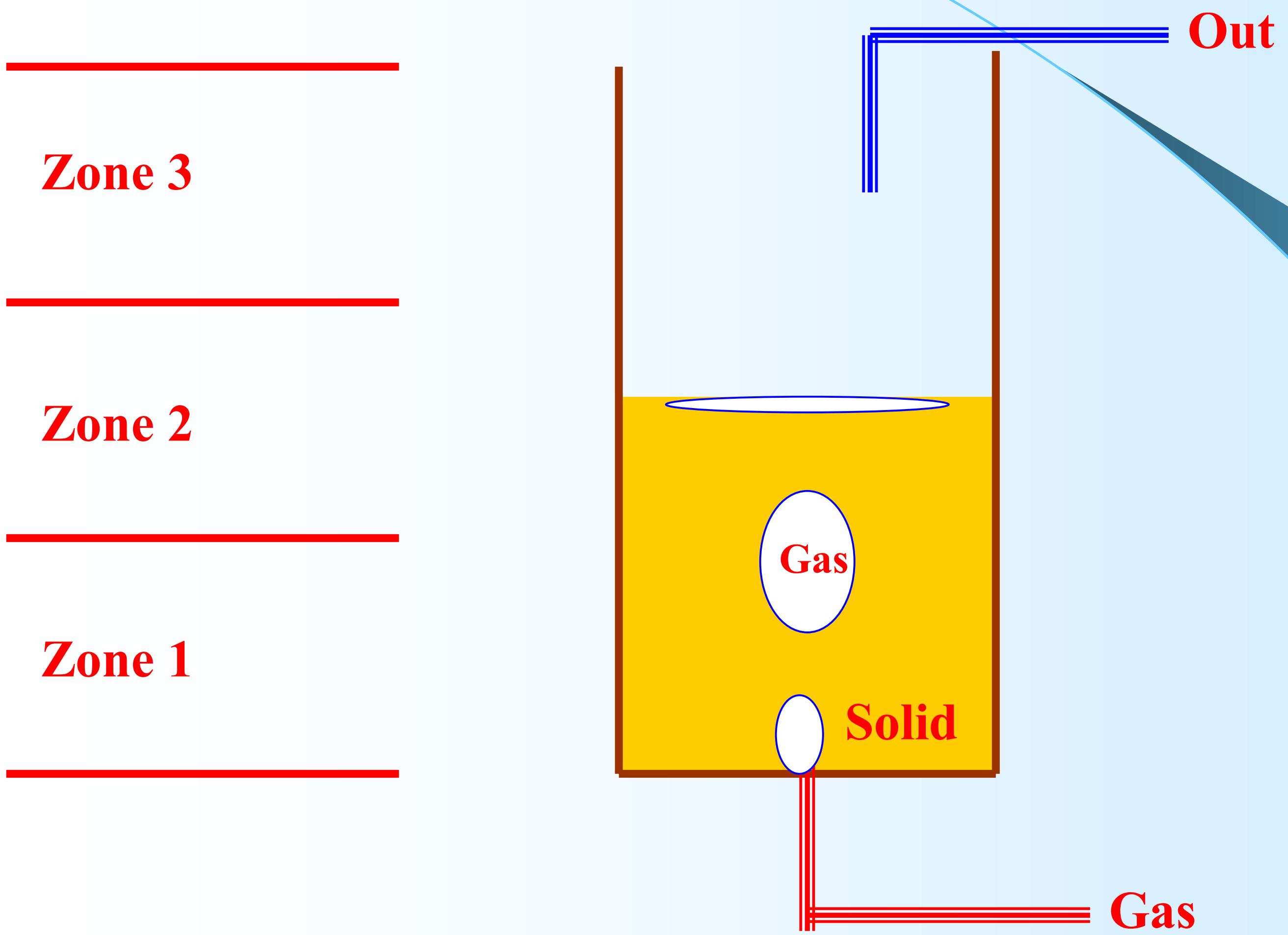
Neural Networks

- ✓Mimic connections between human neurons in the brain
- ✓Learn by back propagating the error
- ✓Can learn non-linear or curved separation boundary
- ✓*A priori* information is required in most cases

Clustering Methods

- ✓Utilizing the ‘similarity’ between two vectors to group them
- ✓Supervised or Unsupervised training
- ✓Can learn non-linear or curved separation boundary
- ✓Clustering can be arbitrary
- ✓*A priori* information not necessary

Fluidized Bed



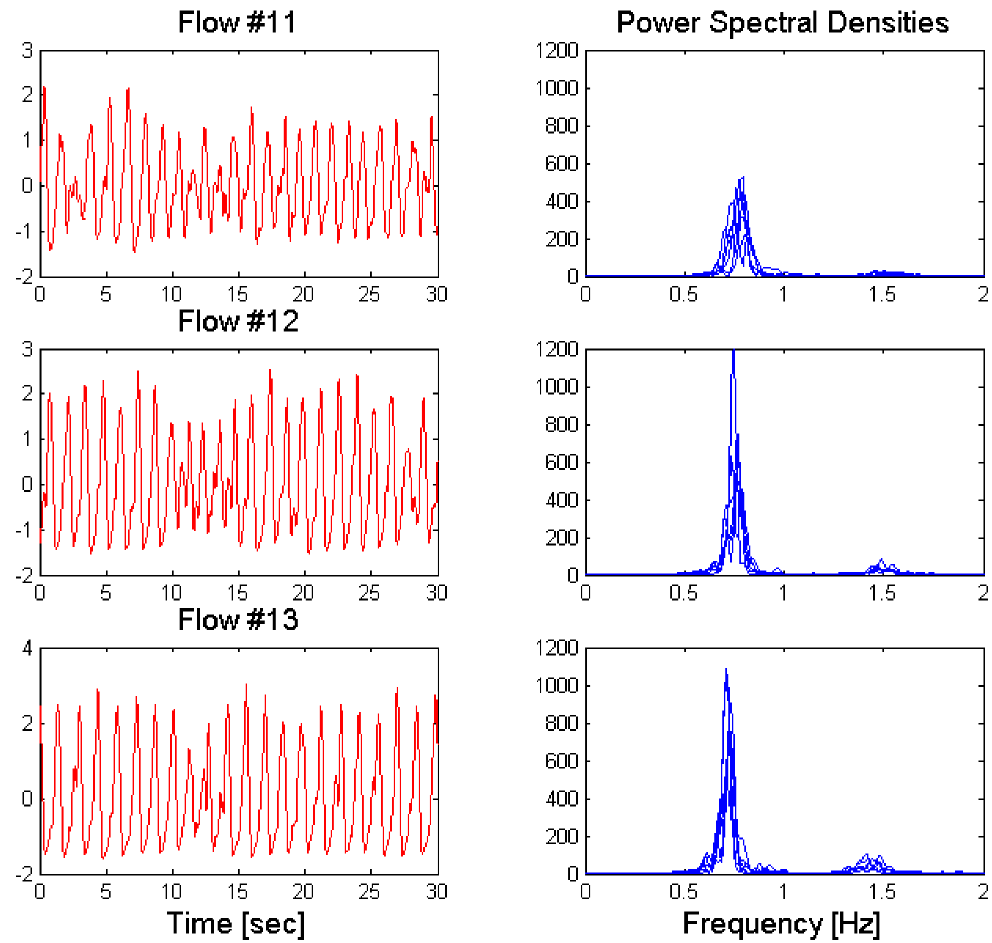
Rajput-Bruns-Daw

Process states available

19 flow rates or 'states' available from an experimental fluidized bed (courtesy Dr. Charles Finney)

Increasing flow rates (1 through 19) cover fluidization behavior from bubbling to extremely turbulent

A sample of time series and power spectral density



Note the varying power in different time series segments

Expert System used in this study

Clustering Methods

✓K-means

✓Algorithm seeks to minimize the sum of squared variance within clusters

✓Need to know the number of clusters required

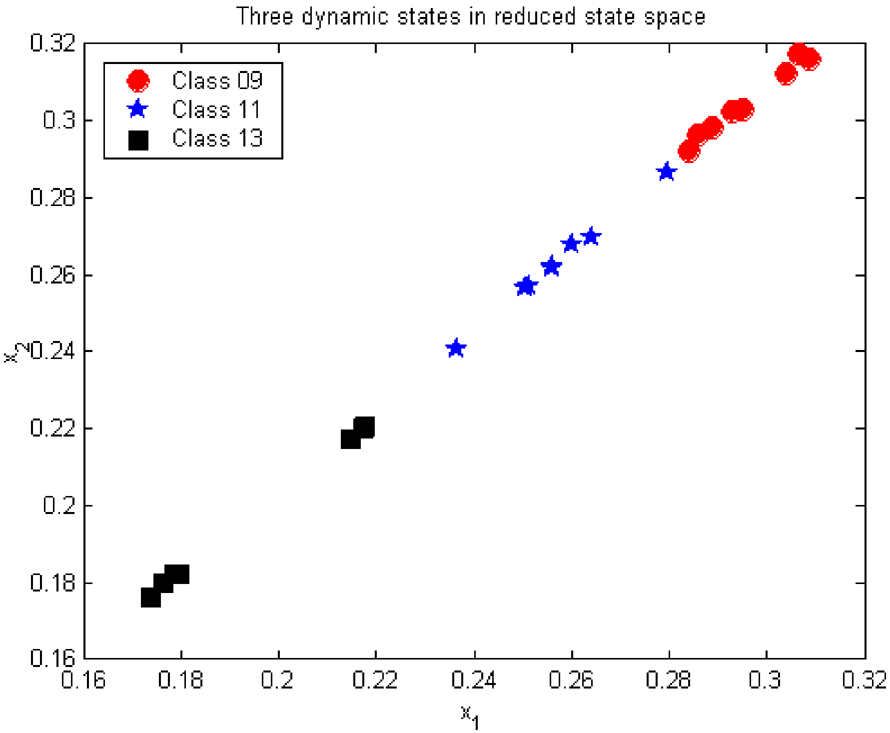
✓Euclidean distance used as the measure of [dis]similarity

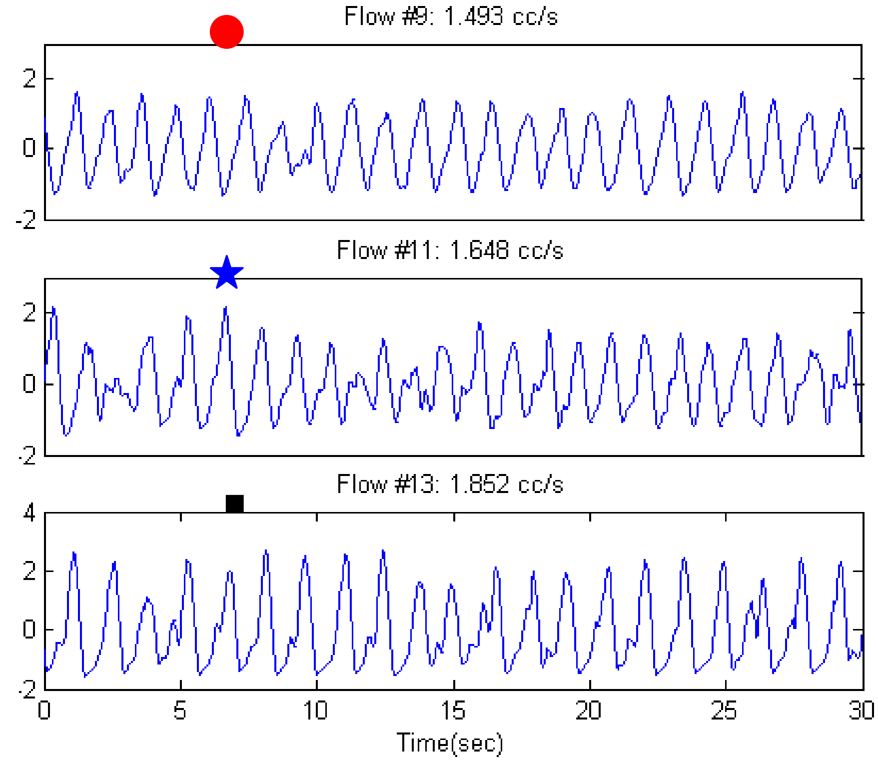
✓Compare clustering efficiency using

- Mutual Information (MIF)
- Power spectral density (PSD)

Comparing similar dynamic states

Features	3	8	14
	Classification Error (%)		
MIF	0%	0%	0%
PSD	22%	34%	33%



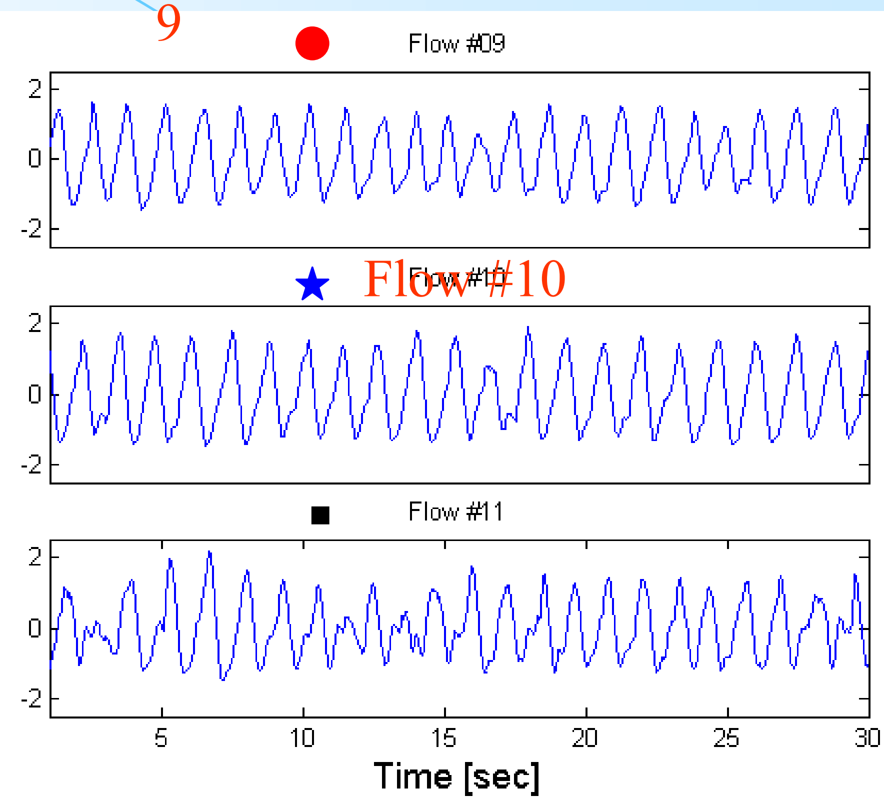


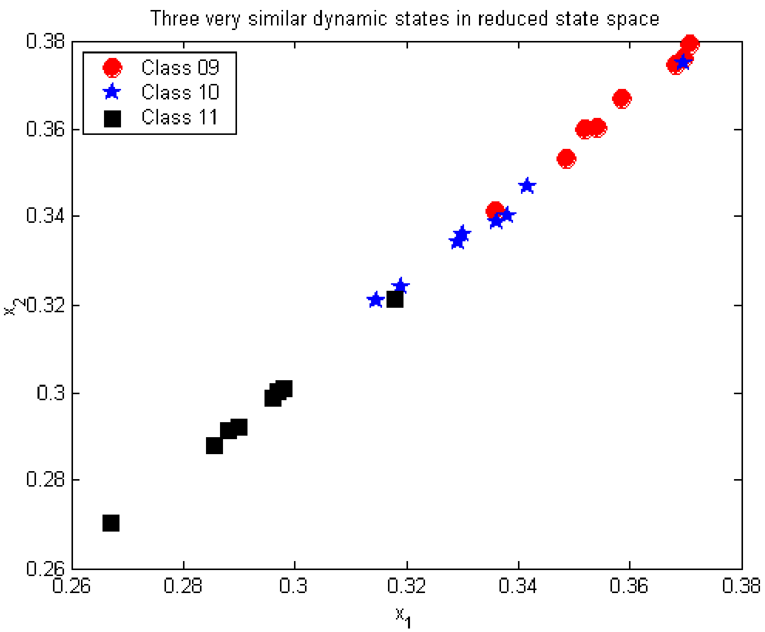
Using a biased estimate of classification error. However it should give general idea about clustering efficiency

Rajput-Bruns-Daw

Comparing very similar dynamic states

Features	3	8	14
	Classification Error		
MIF	14%	10%	9%
PSD	29%	17%	34%





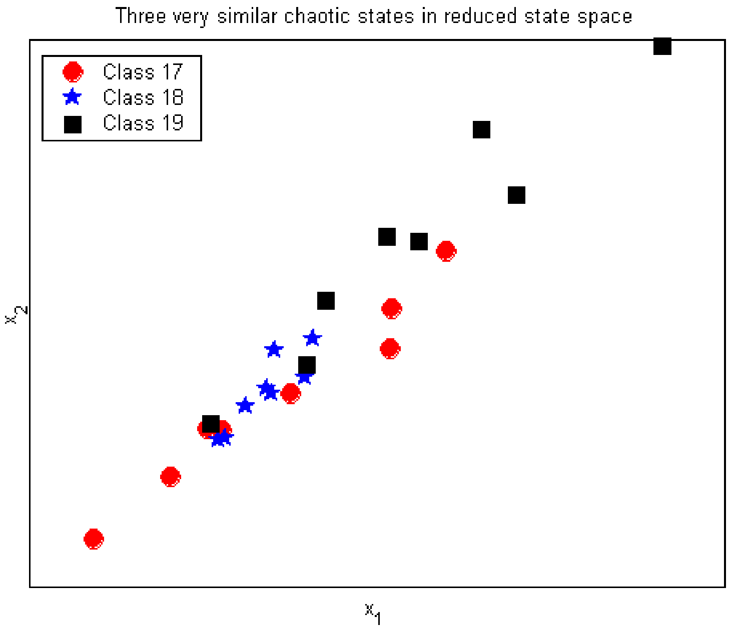
MIF is a better indicator than PSD, but the fact remains that it is hard to distinguish very similar states

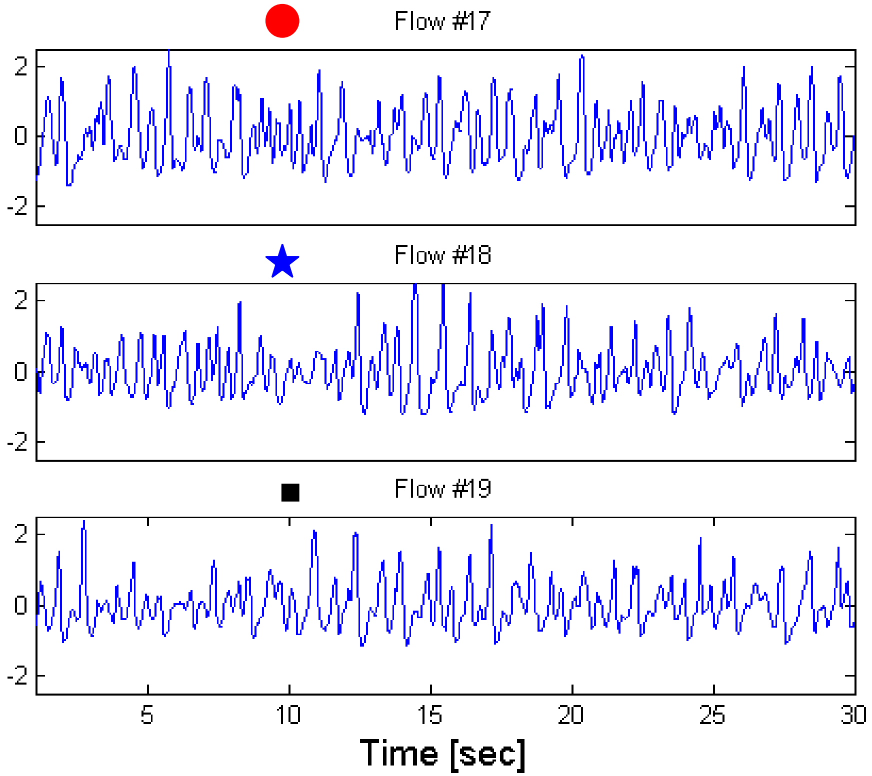
Rajput-Bruns-Daw

Comparing very similar chaotic states

Classification Error

Features	3	8	15
MIF	42%	38%	29%
PSD	39%	48%	54%





To distinguish between very similar chaotic states is very difficult

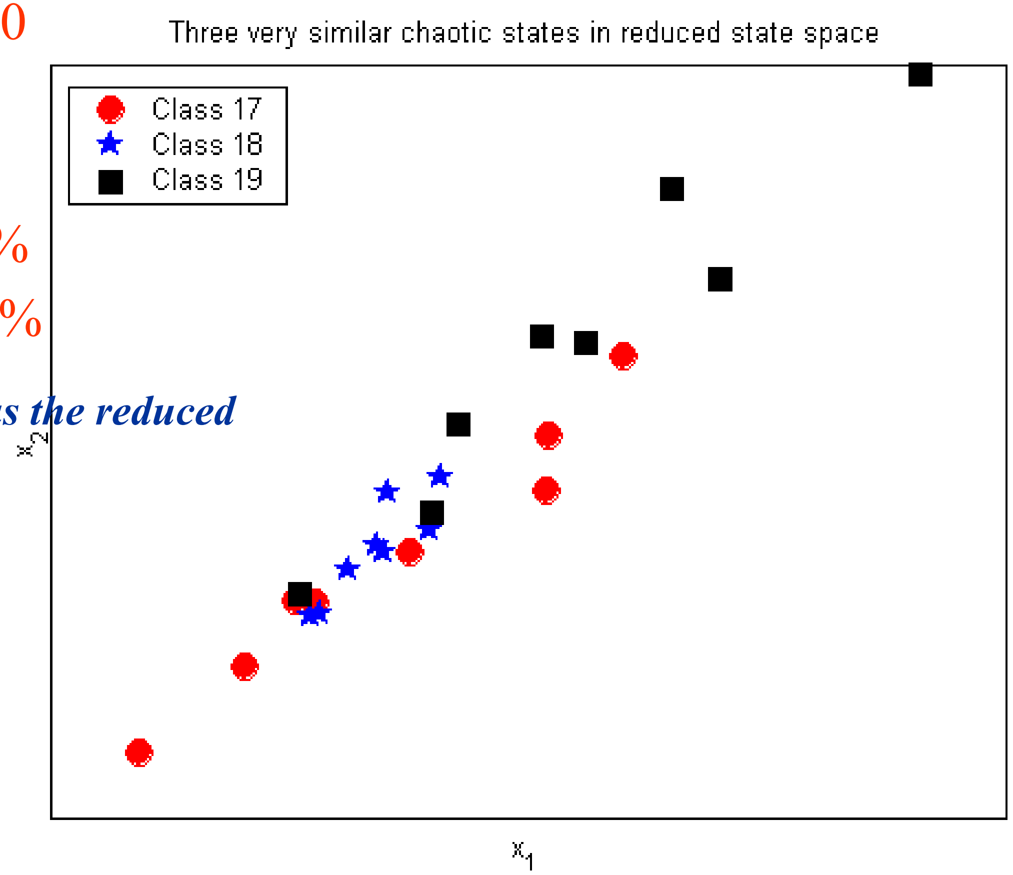
Comparing very similar chaotic states : In Pairs

Class 1	Class 2	Features: 2	5	10
		Classification Error		
17	18	17%	13%	7%
18	19	13%	16%	15%

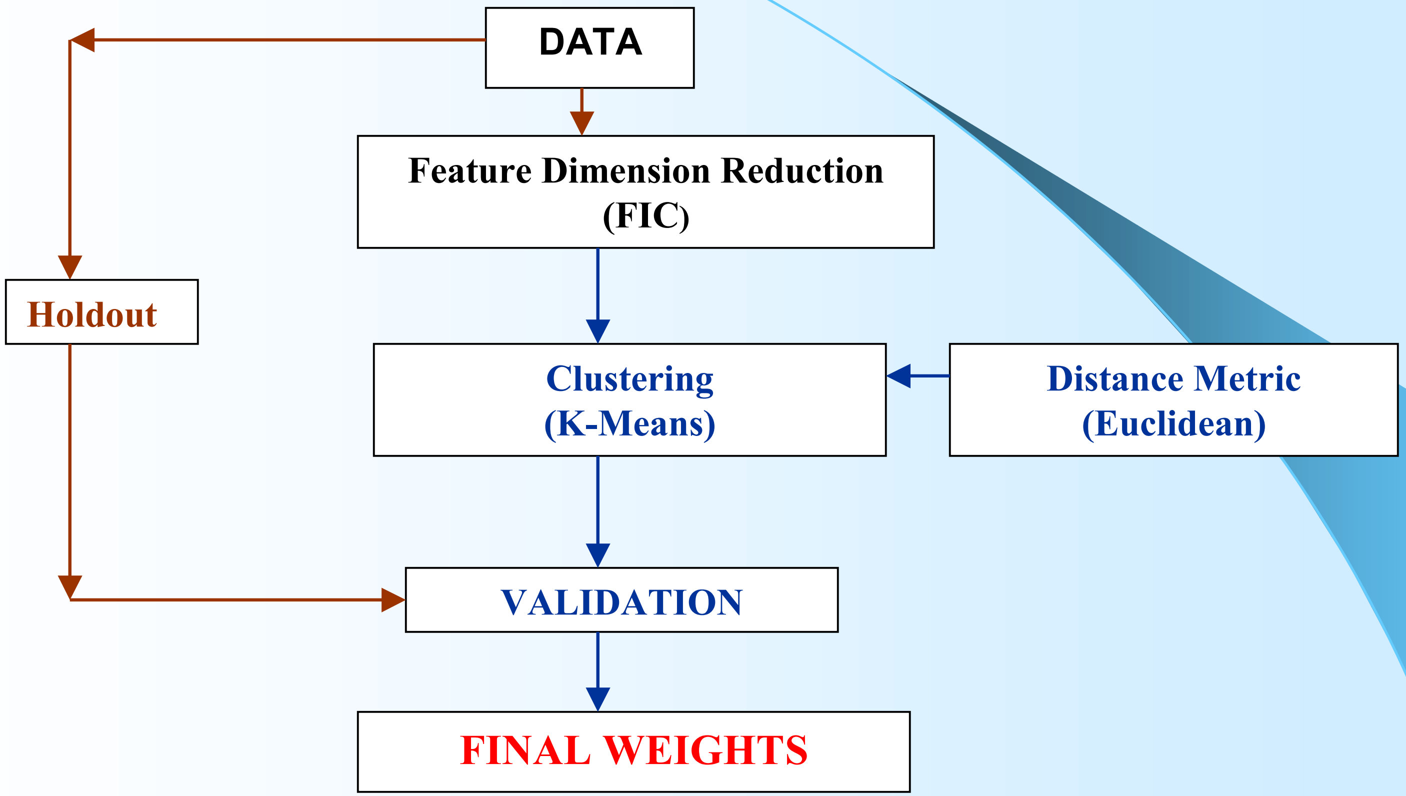
Input to the clustering algorithm was the reduced Mutual Information Vector

Comparing classes in pairs results in less classification errors.

However, the clustering algorithm still can not perfectly distinguish two very close dynamic states



Flow Diagram



Discussion

- ✓ Mutual Information vector can classify similar dynamic states
- ✓ We tested the limits of the scheme by comparing nearly identical dynamic states
- ✓ We demonstrated that mutual information is **superior** to spectral density as a feature vector in state recognition
- ✓ Can use reduced pattern vector for **process monitoring** given the state-clusters

Advantages

- ✓ Computing Mutual Information at a specified number of lags is quick
- ✓ Less storage space required to store symbolized data
- ✓ Compression techniques from Information theory can be used

Future Directions

- ✓Neural networks can be used
- ✓Use different clustering algorithms
- ✓Use other nonlinear measures

A way to deal with slowly changing dynamics

- ✓'Fuzzy' clustering

Thank you!