TEMPORAL CLUSTERING OF MULTIVARIATE TIME SERIES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

SİPAN ASLAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
STATISTICS

FEBRUARY 2022

Approval of the thesis:

TEMPORAL CLUSTERING OF MULTIVARIATE TIME SERIES

Submitted by **Sipan Aslan** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in the Department of Statistics, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** ______________

Prof. Dr. Özlem İlk Dağ
Head of Department, **Statistics** ______________

Assoc. Prof. Dr. Ceylan Yozgatlıgil
Supervisor, **Department of Statistics, METU** ______________

Prof. Dr. Cem İyigun
Co-supervisor, **Department of Industrial Eng., METU** ______________

Examining Committee Members:

Prof. Dr. İnci Batmaz
Department of Statistics, METU ______________

Assoc. Prof. Dr. Ceylan Yozgatlıgil
Supervisor, Department of Statistics, METU ______________

Prof. Dr. Uğur Soytaş
Department of Tech., Mgt. and Econ., DTU ______________

Prof. Dr. Esin Firuzan
Department of Statistics, Dokuz Eylül University ______________

Assoc. Prof. Dr. Tülin İnkaya
Department of Industrial Eng., Bursa Uludağ University ______________

Date: ______________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Sipan Aslan

Signature :

ABSTRACT

TEMPORAL CLUSTERING OF MULTIVARIATE TIME SERIES

Aslan, Sipan
Ph.D., Department of Statistics
Supervisor : Assoc. Prof. Dr. Ceylan Yozgatlıgil
Co-Supervisor : Prof. Dr. Cem İyigun

February 2022, 202 pages

Clustering of real-valued time series is a prevalent problem that frequently emerges in various fields and applications. While clustering of univariate time series is very much examined, clustering of multivariate time series has not been extensively addressed. This dissertation considers clustering of real-valued multivariate time series data. When the data analyzed in the clustering task are time series, the time dependencies of the time series and the clusters to be formed should be considered together. In this thesis study, we propose a time series model-based clustering approach that can be used for clustering of univariate and multivariate times series datasets. The proposed approach, rather than searching similar/dissimilar patterns within a given collection of time series, is mainly focused on clustering with respect to approximations to the generating mechanisms of time series by exploring and utilizing its temporal dependency, linear and non-linear behaviors. In addition, the proposed clustering approach is designed to capture time-dependent cluster changes. The efficiency of the proposed approach is demonstrated by using both synthetic and real data. Synthetic datasets are derived under different scenarios and real datasets obtained from several time series classification studies where the memberships of studied time series are already known. The clustering performances of the proposed approach are compared with other proposed clustering methods.

Keywords: Time series, Multivariate, Clustering, Model Based Clustering, Temporal Clustering

ÖZ

ÇOK DEĞİŞKENLİ ZAMAN SERİLERİNİN ZAMANSAL KÜMELEMESİ

Aslan, Sipan
Doktora, İstatistik Bölümü
Tez Yöneticisi : Doç. Dr. Ceylan Yozgatlıgil
Ortak Tez Yöneticisi : Prof. Dr. Cem İyigun

Şubat 2022 , 202 sayfa

Zaman serisi verilerinin kümelenmesi, çeşitli alanlarda ve uygulamalarda sıklıkla ortaya çıkan yaygın bir sorundur. Tek değişkenli zaman serilerinin kümelenmesi çok fazla incelenirken, çok değişkenli zaman serilerinin kümelenmesi kapsamlı bir şekilde ele alınmamıştır. Bu tez gerçel değerli çok değişkenli zaman serileri verilerinin kümelenmesini ele almaktadır. Kümeleme analizinde analiz edilen veriler zaman serileri olduğunda, zaman serilerinin ve oluşturulacak kümelerin zamana olan bağımlılıkları birlikte dikkate alınmalıdır. Bu çalışmada, tek değişkenli ve çok değişkenli zaman serilerinin zamana bağlı kümelenmesi için kullanılabilecek, zaman serisi modeline dayalı bir kümeleme yaklaşımı önermekteyiz. Önerilen yaklaşım, belirli bir zaman serisi veri setinde benzer/benzemez örüntüler aramak yerine, temel olarak zamansal bağımlılık, doğrusal ve doğrusal olmayan davranışları keşfederek ve kullanarak zaman serilerinin üretme mekanizmalarına yaklaşımlar üzerinde kümelenmeye odaklanmaktadır. Ek olarak, önerilen kümelenme yaklaşımı, zamana bağlı küme değişikliklerini yakalamak için tasarlanmıştır. Önerilen yaklaşımın etkinliği hem yapay hem de gerçek veriler kullanılarak gösterilmiştir. Yapay veriler farklı senaryolar altında türetilmiştir ve gerçek veriler, çalışılan zaman serilerinin üyeliklerinin zaten bilindiği çeşitli sınıflandırma çalışmalarından elde edilmiştir. Önerilen yaklaşımın kümelenme performansları, diğer önerilen kümeleme yöntemleri ile karşılaştırılmıştır.

Anahtar Kelimeler: Zaman Serileri, Çok Değişkenli, Kümeleme, Model tabanlı Kümeleme, Zamansal Kümeleme

..."everything separable is distinguishable and everything distinguishable is different. this is the principle of difference/creation."

gilles deleuze

ACKNOWLEDGMENTS

Foremost, I would like to thank my advisor, Ceylan Yozgatlıgil, and co-advisor Cem İyigün, for their guidance, feedback, and unconditional support during my years at METU. I have always felt their encouragement at every stage of the study, and they never stopped to help me whenever I needed it. I am very grateful that they have sincerely shared their experiences with me. Their patience and sincerity towards me have been my most unique chance to accomplish this study. I admiringly respect their principles and ways of working, which will be a lifelong reference for my life.

I would like to acknowledge my appreciation to my thesis committee members Prof. Dr. İnci Batmaz and Prof. Dr. Uğur Soytaş, for guiding my thesis for many years. They always encouraged me to rely on myself and my study. Their valuable comments and insightful suggestions have improved the quality of the study. They have always supported the discussions and made essential contributions in this direction to make the aims of the thesis more visible. My admiration for their manner of academic work will lead me as long as I live and work.

I would also like to acknowledge my most profound appreciation to Prof. Dr. Esin Firuzan, with the support of whom I started my academic journey and Assoc.Prof.Dr. Tülin İnkaya for spending their valuable time reviewing my thesis.

I would like to offer my special thanks to Prof. Dr. Hernando Ombao for giving me the opportunity of studying with him for a year and for granting access to the computational facilities located at the department of statistics at UCI. He has authorized me to use this opportunity for years, assisting me to complete the computational-intensive phase of the study. I would also like to thank him for the eye-opening conversations.

I would like to thank the staff and students in the statistics department at METU; it has been the most valuable experience for my academic journey. I would like to thank all people who have helped and inspired me during my study.

I am indebted to my wife, Özge Çağlar. It's never enough to thank, but I want to acknowledge my wife Özge Çağlar for her constant encouragement and for tolerating my many troubled times. I want to gratefully thank my parents, Bedriye and Huti, and my brothers, Ali and Özgür, for their sincere and neverending support whenever I need it. They were always there for me when I felt hopeless.

Finally, it sounds a bit cliché, but I owe a lot to my daughter Arin Farnah for irreversibly wasting so much time we could have spent together. I would like to thank her for being my muse with her life energy and humorous spirit.

The computations reported in this paper were partially performed at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources). I would like to thank TUBITAK for providing the valuable opportunity of computing facilities.

TABLE OF CONTENTS

ABSTRACT v

ÖZ vii

ACKNOWLEDGMENTS x

TABLE OF CONTENTS xii

LIST OF TABLES xvi

LIST OF FIGURES xxiii

LIST OF ABBREVIATIONS xxvi

CHAPTERS

1 INTRODUCTION 1

1.1 Philosophical Aspects 1

1.2 Time Dependency of Data and Clusters 8

1.3 Objective of this Thesis 10

1.4 Thesis Outline 14

2 TIME SERIES CLUSTERING: A REVIEW 15

2.1 Univariate Time Series Clustering 18

2.2 Multivariate Time Series Clustering 22

2.3 Proposed Contributions 24

3 PROPOSED APPROACH 27

3.1 Motivation 27

3.2 Background on the Double Threshold GARCH Model 31

3.2.1 Threshold Autoregressive Model 31

3.2.2 Self-Exciting Threshold Autoregressive Model . . 33

3.2.3 Double Threshold GARCH Model 35

3.3 Spectral Clustering 39

3.4 Proposed Approach for Time Series Clustering 44

3.4.1 Feature Blocks to be Used in Clustering 45

3.4.2 Multiple Univariate Time Series Clustering 50

3.4.3 Multiple Multivariate Time Series Clustering . . . 56

4 EXPERIMENTAL RESULTS ON UNIVARIATE TIME SERIES . . 59

4.1 Performance Evaluation over Synthetic Data 61

4.1.1 First Set of Experiments 67

4.1.1.1 Non-temporal 10 Cluster 70

4.1.1.2 Temporal 10 Cluster 77

4.2 Performance Evaluation over Real-Life Data 82

4.2.1 Clustering of sEMG Signals 83

4.2.2 Clustering of EEG Signals 88

5 EXPERIMENTAL RESULTS ON MULTIVARIATE TIME SERIES . 93

5.1 Performance evaluation over Synthetic Data 96

5.1.1 First Set of Experiments on Multivariate Time-series Clustering 97

5.1.1.1 Non-temporal 3 Cluster 99

5.1.1.2 Temporal 3 Cluster 103

5.1.1.3 Temporal 9 Cluster 107

5.1.2 Second Set of Experiments on Multivariate Time-series Clustering 112

5.1.2.1 Non-temporal 3 Cluster 115

5.1.2.2 Temporal 6 Cluster 118

5.1.2.3 Temporal 12 Cluster 122

5.2 Performance evaluation over Real Data 127

5.2.1 Clustering of Motion Tracking Data 128

5.2.2 Clustering of Motor Execution and Imagery Data . 133

6 SUMMARY AND DISCUSSION 143

REFERENCES 151

APPENDICES

A UNIVARIATE TIME SERIES CLUSTERING: EXPERIMENTS ON SYNTHETIC DATA 167

A.1 Experiment 1 167

A.1.1 Experiment 1: Case 3 167

A.2 Experiment 2 170

A.2.1 Experiment 2: Case 1 170

A.2.2 Experiment 2: Case 2 172

A.3 Experiment 3 175

A.3.1 Experiment 3: Case 1 175

A.3.2 Experiment 3: Case 2 177

A.4 Experiment 4 180

A.4.1 Experiment 4: Case 1 180

A.4.2 Experiment 4: Case 2 182

B MULTIVARIATE TIME SERIES CLUSTERING: EXPERIMENTS ON SYNTHETIC DATA 185

B.1 Experiment 3 185

B.1.1 Experiment 3: Case 1 185

B.2 Experiment 4 189

B.2.1 Experiment 4: Case 1 189

B.3 Experiment 5 193

B.3.1 Experiment 5: Case 1 193

LIST OF TABLES

TABLES

Table 1.1 Principles to be considered as a guide while developing and assessing clustering approaches. 7

Table 2.1 Model free dissimilarity measures for univariate time series clustering. 19

Table 2.2 Complexity based dissimilarity measures for univariate time series clustering. 21

Table 2.3 Model based dissimilarity measures for univariate time series clustering. 21

Table 2.4 Dissimilarity measures for multivariate time series clustering. 22

Table 3.1 Components of the feature vector for a given time series y_t. 51

Table 4.1 Dissimilarity measures used for comparison in univariate clustering. 62

Table 4.2 The feature vector and fixed hyperparameters in this study. 66

Table 4.3 Utilized specifications of mean processes in the first set of experiments. 67

Table 4.4 Utilized specifications of error structures in the first set of experiments. 68

Table 4.5 Hypothetical DGMs used in Case 1. 70

Table 4.6 Univariate Experiment 1 Case 1: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach. 73

Table 4.7 Estimated number of clusters by the global maximum of the GAP statistics out of 100 repetitions using TSMB-SPCL-UV. 76

Table 4.8 Hypothetical DGMs used in Case 2. 77

Table 4.9 Case 2 average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach. 80

Table 4.10 sEMG Hand Movement dataset information used in the study. 84

Table 4.11 sEMG clustering accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach. 85

Table 4.12 EEG dataset information used in the study. 88

Table 4.13 EEG clustering accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach. 90

Table 5.1 Dissimilarity measures used for comparison in multivariate clustering. 94

Table 5.2 Utilized specifications of SDEs in the first set of experiments. 98

Table 5.3 Hypothetical DGMs used in Case 1. 99

Table 5.4 Multivariate Experiment 1 Case 1: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach. 101

Table 5.5 Hypothetical DGMs used in Case 2. 103

Table 5.6 Multivariate Experiment 1 Case 2: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach. 106

Table 5.7 Hypothetical DGMs used in Case 3. 107

Table 5.8 Multivariate Experiment 1 Case 3: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach. 110

Table 5.9 The Gap statistics peaks out of 100 repetitions using TSMB-SPCL-MV. 111

Table 5.10 Utilized specifications of VECMs in the first set of experiments. . . 112

Table 5.11 Utilized specifications of multivariate error structures in the seconds set of experiments. 113

Table 5.12 Hypothetical DGMs used in Case 1. 115

Table 5.13 Multivariate Experiment 2 Case 1: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach. 117

Table 5.14 Hypothetical DGMs used in Case 2. 118

Table 5.15 Multivariate Experiment 2 Case 2: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach. 121

Table 5.16 Hypothetical DGMs used in Case 3. 122

Table 5.17 Multivariate Experiment 2 Case 3: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach. 125

Table 5.18 The Gap statistics peaks out of 100 repetitions using TSMB-SPCL-MV. 126

Table 5.19 Multivariate Human Motion dataset. 129

Table 5.20 Multivariate Human Motion dataset: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach. 131

Table 5.21 Multivariate EEG Motor Execution and Imagery dataset. 135

Table 5.22 Multivariate EEG Motor Execution and Imagery dataset: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach. 137

Table 5.23 The times required to compute the distance matrices for real data experiments. 140

Table A.1 Utilized specifications of mean processes in the first set of univariate experiments. 167

Table A.2 Utilized specifications of error structures in the first set of univariate experiments. 167

Table A.3 Hypothetical DGMs used in Univariate Experiment 1: Case 3. . . . 168

Table A.4 Univariate Experiment 1: Case 3 average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach. 169

Table A.5 Utilized specifications of mean processes in the second set of univariate experiments. 170

Table A.6 Utilized specifications of error structures in the second set of univariate experiments. 170

Table A.7 Hypothetical DGMs used in Univariate Experiment 2: Case 1. . . . 170

Table A.8 Univariate Experiment 2 Case 1: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach. 171

Table A.9 Utilized specifications of mean processes in the second set of univariate experiments. 172

Table A.10Utilized specifications of error structures in the second set of univariate experiments. 172

Table A.11Hypothetical DGMs used in Univariate Experiment 2: Case 2. . . . 172

Table A.12Univariate Experiment 2: Case 2 average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach. 174

Table A.13Utilized specifications of mean processes in the third set of univariate experiments. 175

Table A.14Utilized specifications of error structures in the third set of univariate experiments. 175

Table A.15Hypothetical DGMs used in Univariate Experiment 3: Case 1. . . . 175

Table A.16Univariate Experiment 3 Case 1: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach. 176

Table A.17Utilized specifications of mean processes in the third set of univariate experiments. 177

Table A.18Utilized specifications of error structures in the third set of univariate experiments. 177

Table A.19Hypothetical DGMs used in Univariate Experiment 3: Case 2. . . . 177

Table A.20Univariate Experiment 3: Case 2 average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach. 179

Table A.21Utilized specifications of mean processes in the fourth set of univariate experiments. 180

Table A.22Hypothetical DGMs used in Univariate Experiment 4: Case 1. . . . 180

Table A.23Univariate Experiment 4 Case 1: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach. 181

Table A.24 Utilized specifications of mean processes in the fourth set of univariate experiments. 182

Table A.25 Hypothetical DGMs used in Univariate Experiment 4: Case 2. . . . 182

Table A.26 Univariate Experiment 4: Case 2 average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach. 184

Table B.1 Utilized specifications of VARs in the third set of experiments. . . . 185

Table B.2 Utilized specifications of multivariate error structures in the third set of experiments. 186

Table B.3 Hypothetical DGMs used in Multivariate Experiment 3: Case 1 (9 cluster). 187

Table B.4 Multivariate Experiment 3 Case 1: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach. 188

Table B.5 Utilized specifications of TVARs in the fourth set of experiments. . . 189

Table B.6 Utilized specifications of multivariate error structures in the fourth set of experiments. 190

Table B.7 Hypothetical DGMs used in Multivariate Experiment 4: Case 1 (9 cluster). 191

Table B.8 Multivariate Experiment 4 Case 1: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach. 192

Table B.9 Utilized specifications of TVECMs in the fifth set of experiments. . 194

Table B.10 Utilized specifications of multivariate error structures in the fifth set of experiments. 195

Table B.11 Hypothetical DGMs used in Multivariate Experiment 5: Case 1 (9 cluster). 196

Table B.12 Multivariate Experiment 4 Case 1: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach. 197

LIST OF FIGURES

FIGURES

Figure 1.1 Number of publications (article, book and book chapter) per year which contains words of **clustering** or **classification** in title from 2000 to 2018. 4

Figure 1.2 Number of publications (article, book and book chapter) per Web of Science categories which contains words of **clustering** or **classification** in title from 2000 to 2018. 5

Figure 2.1 General clustering procedures: (a) raw data based, (b) feature based, (c) model and feature based 16

Figure 3.1 Analogy between the principle of spectroscopy and the feature extraction. 30

Figure 3.2 (a) Sample undirected and edge weighted graph G; (b) Corresponding adjacency matrix of G where $w_{ij} \geq 0; w_{ij} = w_{ji}; i, j = 1, \ldots, 6.$ . . . 40

Figure 3.3 (a) Sample undirected graph G subjected to partitioning; (b) Corresponding $graph\ Laplacian$ of G. 42

Figure 3.4 Steps of the proposed time series clustering approach 56

Figure 4.1 Diagram for testing the temporality of clusters. 65

Figure 4.2 Example of the simulation sample from 10 DGMs listed in Table 4.5. 71

Figure 4.3 Accuracy rates vs. TSMB-SPCL-UV over 100 replicates. 74

Figure 4.4 GAP statistic per cluster number for a replicate using TSMB-SPCL-UV. 75

Figure 4.5 GAP statistics' Box-Plot per cluster number for 100 replicates using TSMB-SPCL-UV. 76

Figure 4.6 The six basic hand movements considered in the study. `https://archive.ics.uci.edu/ml/machine-learning-databases/00313/` 83

Figure 4.7 sEMG samples for basic hand movements. 84

Figure 4.8 GAP statistic per cluster number for the sEMG dataset using TSMB-SPCL-UV. 87

Figure 4.9 Average Silhouette Value per cluster number for the sEMG dataset using TSMB-SPCL-UV. 87

Figure 4.10 EEG samples. 89

Figure 4.11 GAP statistic per cluster number for the EEG dataset using TSMB-SPCL-UV. 92

Figure 4.12 Average Silhouette Value per cluster number for the EEG dataset using TSMB-SPCL-UV. 92

Figure 5.1 Simulation samples from 3 DGMs listed in Table 5.3 100

Figure 5.2 Simulation samples from 3 DGMs listed in Table 5.5 104

Figure 5.3 Simulation samples from 9 DGMs listed in Table 5.7 108

Figure 5.4 GAP statistics' Box-Plots per cluster number using TSMB-SPCL-MV. 111

Figure 5.5 Simulation samples from 3 DGMs listed in Table 5.12 115

Figure 5.6 Simulation samples from 6 DGMs listed in Table 5.14 119

Figure 5.7 Simulation samples from 12 DGMs listed in Table 5.16 123

Figure 5.8 GAP statistics' Box-Plots per cluster number using TSMB-SPCL-MV. 126

Figure 5.9 Samples from multivariate human activity dataset. 130

Figure 5.10 GAP statistic per cluster number for the Human Motion dataset using TSMB-SPCL-MV. 132

Figure 5.11 Average Silhouette Value per cluster number for the Human Motion dataset using TSMB-SPCL-MV. 132

Figure 5.12 A sample of 64-channel EEG placement. 133

Figure 5.13 Samples from multivariate EEG motor execution and imagery dataset.136

Figure 5.14 GAP statistic per cluster number for the Human Motion dataset using TSMB-SPCL-MV. 138

Figure 5.15 Average Silhouette Value per cluster number for the Human Motion dataset using TSMB-SPCL-MV. 139

LIST OF ABBREVIATIONS

ACF	Autocorrelation Function
AIC	Akaike Information Criterion
APARCH	Asymmetric Power ARCH
AR.LPC	AutoRegressive Linear Predictive Coding
AR.MAH	AutoRegressive by Maharaj
AR.PIC	AutoRegressive by Piccolo
AR	AutoRegressive
ARCH	Autoregressive Conditional Heteroscedasticity
ARIMA	AutoRegressive Integrated Moving Average
BCI	Brain Computer Interface
BIC	Bayesian Information Criterion
CDM	Compression based Dissimilarity Measure
CID	Complexit-Invariant Distance
COR	Correlation
CORT	Temporal Correlation
DCC	Dynamic Conditional Correlation
DGM	Data Generating Mechanism
DTGARCH	Double Threshold GARCH
DTW	Dynamic Time Warping
DWT	Discrete Wavelet Transform
ECCC	Extended Constant Conditional Correlation
EEG	Electroencephalography
EGARCH	Exponential GARCH
EUCL	Euclidean

FFT	Fast Fourier Transform
GAK	Global Alignment Kernel
GARCH	Generalized Autoregressive Conditional Heteroscedasticity
GLK	Generalized Likelihood
HQIC	Hannan–Quinn Information Criterion
INT.PER	Integrated Periodogram
ISD	Integrated Squared difference
JD	Jeffreys Divergence
KS	Kolmogorov-Smirnov
LD	Log-Spectral Divergence
LS	Least Square
LSTAR	Logistic Smooth Transition AutoRegressive
ME	Motor Execution
MI	Motor Imagery
MLE	Maximum Likelihood Estimation
MV	Multivariate
NCD	Normalized Compression Distance
PACF	Partial Autocorrelation Function
PAM	Partitioning Around Medoid
PDC	Permutation Distribution Clustering
PER	Periodogram
QMLE	Quasi Maximum Likelihood Estimation
SARIMA	Seasonal AutoRegressive Integrated Moving Average
SAX	Symbolic Aggragate Approximation
SC	Spectral Clustering
SDE	Stochastic Differential Equations
sEMG	Surface Electromyography

SETAR	Self-Exciting TAR
sGARCH	Standart GARCH
SPCL	Spectral Clustering
SPEC.LLR	Local-Linear Estimation of Log-Spectrum
STAR	Smooth Transition AutoRegressive
TAR	Threshold AutoRegressive
TGARCH	Threshold GARCH
TSMB	Time Series Model Based
UV	Univariate
VECM	Vector Error Correction Model
VR	Variance Ratio

CHAPTER 1

INTRODUCTION

1.1 Philosophical Aspects of Clustering, Difference and Dissimilarity

Classification and clustering have been one of the fundamental research problems, from antiquity to modern sciences, in many different areas such as natural, social and information sciences. In classification, class properties or classification rules are investigated/specified at first, and then the conformity of given data to these defined classes is evaluated. However, in clustering, methods are developed or utilized by which the possibilities of creating homogeneous groups/clusters are investigated based on the observed data. The reason why we are trying to solve clustering/classification problems is that there is a constant need for knowledge and interpretable (i.e., manageable and meaningful) information to be retrieved through the organization/order of objects and data heaps observed in nature. For example, Aristotle classified animals and plants [1]; Plato classified such as arts, constitutions, laws, and jobs [2]; Mendeleev classified elements with respect to their chemical properties [3]; Durkheim used classification in sociology [4], and states that "If human societies cannot be classified, they must remain inaccessible to scientific description" [5, p.9]; Astrophysicists classified stars according to their spectral properties [6–8] (see [9–12] for more applications). Moreover, the categorization effort, including clustering and classification, has a vital role in human cognitive development and adaptation to the surrounding environment [13]. Categorization is a constant activity that we have almost practiced since our birth. It is pretty natural that this primary act has become one of the critical topics of science and artificial intelligence researches specifically (Figure 1.2). To understand the complexity we face in nature and be able to make decisions, we have to examine it with the help of its components as is often the case.

The core of the clustering consists of trying to answer the following question: *What is the difference?* Clustering, in general, is the process of distributing objects to an unknown number of homogeneous groups that are usually fewer than the number of objects, depending on some criteria that take into account the differences between the objects. The *difference* between distinguishable is of interest because it is necessary to ask questions about the factors that quantitatively/qualitatively individuate different entities.

We think it would be useful to start with the ideas of some philosophers about the concepts of difference and identity/sameness since it is located at the very center of symbolic logic beginning from G.W. Leibniz's Principle of *the identity of indiscernibles*. The principle states that two entities are identical if and only if they possess exactly the same properties [14, p.308]. Let F be the property (i.e., property function in a broad sense) then the Leibniz's principle can be formulated as follows:

$$(\forall F)(Fx \iff Fy) \to x = y. \tag{1.1}$$

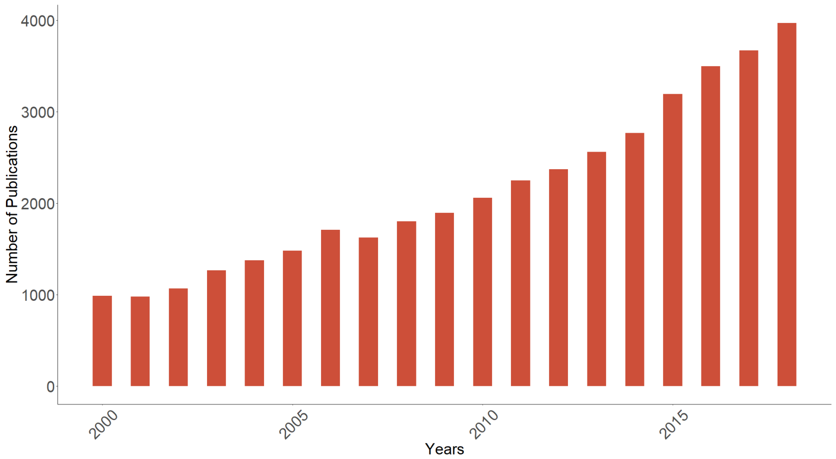
It merely indicates that the two distinct objects are not exactly alike [15]. The equivalent of this principle is *the dissimilarity of the diverse* by McTaggart [16] implies that if x and y are different, then there is at least one property that x and y do not share or vice versa. *The principle of continuity*, which means that nature is continuous and never makes leaps, is another essential principle of Leibniz. Together with *the principle of continuity* and *the identity of indiscernibles*, Leibniz arguments imply that space is filled with unique substances that exist simultaneously but not identical to each other [17, p.58]. Francis Bacon (as cited in [18, p.118]) states that human beings incapable of imagining and figuring out the order in observable nature due to a multitude of unique and different entities though he believes in the existence of the universal order. From this point of view, there is no identical material in nature. When we establish a scientific procedure to find the same things in nature, the only thing we will have at the end is to make their differences clear. Ultimately, this argumentation means that we can find at least one feature that will cause to differentiate the two objects from each other. It is just a matter of scale/resolution. There is some debate about whether this principle will be invalidated when examined in the quantum field (see [19], [20], [21]).

Immanuel Kant gets an interesting idea by attaching the spatial information of the entities to the principles of Leibniz. He argues that it is necessary to distinguish things and their appearances, and rather dwell on conditions of appearance (i.e., emergence/-formation). According to Kant [22, p.325], even-though two objects have the same properties; they are numerically different if their spatial appearances are different at the same time. The emphasis on the notion of spatiotemporality from the 18th century is remarkable. In the philosophy of Gilles Deleuze, the notion of difference reaches a creative/productive level through determinations. The difference, contrary to negation, has a productive, dynamic and positive quality, and it is contingent. Besides, there is some equivalence between concepts of difference and determination. As stated by Deleuze: "Difference is the state in which one can speak of determination as such. The difference between two things is only empirical, and the corresponding determinations are only extrinsic" [23, p.28]. The determination, as it were, implies a difference.

These views on the concept of difference and theoretical premises on the notion that entirely similar objects cannot be found in observable nature are also compatible with the results from empirical clustering analyses (see [24], [25], [26], [27], [28], [29]). In clustering or classification analyses it is not possible to obtain purely homogeneous groups (i.e., true clusters) in which all the members of the group have the same properties. Therefore it is more plausible to use the concept of similarity/dissimilarity instead of identity/difference. In clustering or classification, group or class members can only have a certain degree of similarity/dissimilarity. It would be more accurate to modify the question expressed at the beginning of the section as follows: *What is the similarity/dissimilarity*? In clustering analyses, the answer to this question as of today has to be intentional and extrinsic(i.e., context/aim based). The *difference* is an empirical fact, but a *dissimilarity* can be determined extrinsically in many different ways somewhere around the close neighborhood of the *difference*. In a clustering analysis, it is necessary to ignore some of the infinite numbers of differences, which are most likely to exist among the datasets, via context-based internal consistency. Thus, we can assume that the same dataset can be clustered in many different ways, and to compare datasets with each other and enable clustering, some criteria should be established in which the scope and constraints have been pre-specified.

With the increase in data volume, computer and computing technologies, new clustering/classification methods are being explored in hundreds of published studies which contains words of **clustering** or **classification** in their title each year (see Figures 1.1 and 1.2). Despite all these developments and thousands of publications, however, there is no generalized clustering or classification method in which consensus is established. As we stated before, finding identical things or purely homogeneous groups of things in observable nature is almost impossible, and this makes the cluster definitions somehow/seemingly arbitrary in clustering analyses. Different aims in many clustering applications cause different cluster formations. Therefore, *"homogeneous"* clusters should be formed concerning the context of the application. It seems that the search for a generalized clustering method in response to each clustering problem has been naturally abandoned. This ambiguity leads to debates about whether the methods developed and used are 'objective' or 'subjective', and the findings of clustering studies may become questionable regarding scientific principles.

Because of the reasons stated out before in this section, a generalized clustering procedure has not been constructed yet. However, there are worth mentioning attempts for establishing a scientific framework to assess/guide clustering methodologies and statistical practices in general that still allow us to work within a scientific context.



Source: Web of Science & Clarivate Analytics

Figure 1.1 Number of publications (article, book and book chapter) per year which contains words of **clustering** or **classification** in title from 2000 to 2018.

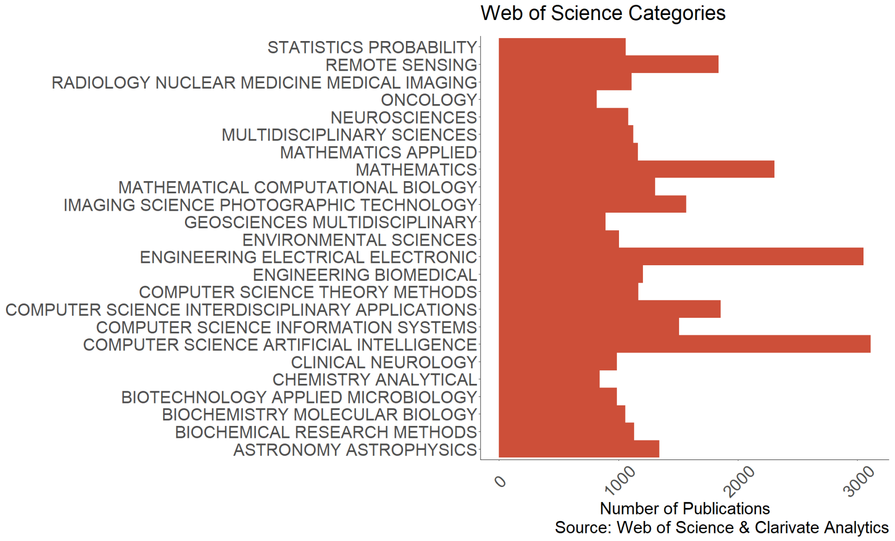


Figure 1.2 Number of publications (article, book and book chapter) per Web of Science categories which contains words of **clustering** or **classification** in title from 2000 to 2018.

It is believed that the scientific activity aims at the some idealized objective truth and avoid subjective decisions as much as possible. Objectivity is seen as an achievable virtue via a scientific method, while subjectivity is thought to be the domain of producing errors and mistakes. The general view in science is that the knowledge is more associated with what is objective than subjective. However, when it is desired to adhere strictly to the principle of objectivity, this may lead to some shortcomings in applied analyses due to the strong contradiction between objective and subjective. While trying to be objective, restrictive and often unverifiable assumptions are made, so that the scope of research may become narrow and remains at the theoretical level. On the other hand, avoiding some of the decisions that are perceived as subjective in the analyses is, in fact, preventing us from taking advantage of available and relevant information about the research questions, and reaching yet undiscovered, unknown innovations and different perspectives [30]. In general, methods that are as user-independent as possible have been evaluated and accepted as more scientific. The truth is that there is no fully automated and researcher-independent scientific activity or method which entirely not depends on decisions that may be regarded as subjective. Gelman and Hennig [30] states that such approaches toward objectivity and subjectivity can be a barrier to good practice in science and statistical analyses. They ground and discuss that though the observer-independent reality is not accessible objectively, there is some reality to be perceived by universal human experience at present time which can be taken as the target of science. This reality is the observer-dependent reality but can not be controlled by the observer instantaneously. According to Chang [31, p.203], this kind of perspective about the reality serves us to be successful in science without even knowing the 'truth'. In [30], they address these arguments comprehensively and suggest the use of some set of attributes instead of objectivity and subjectivity which are used unhelpfully in statistics and science.

Gelman and Hennig [30] propose to replace the term objectivity with a broader set of notions such as *transparency, consensus, impartiality, correspondence to observable reality*; and to replace the term subjectivity with the set of notions such as awareness of *multiple perspectives*, *context dependence*, and investigation of *stability* is suggested as related to both. Unlike the antagonism between objectivity and subjectivity, the attributes proposed in the reformulation of Gelman and Hennig [30] do not oppose each other but complement each other in a positive manner. The key aspect of this argumentation is *transparency*. With the support of these virtues, certain choices/procedures in statistical analyses can be productive and helpful to expand perspectives and space of information if they are transparently and explicitly grounded based on the context/aim of the research. The debate over whether these choices are objective or subjective remains unnecessary, and at the same time, they become the components of the scientific activity. Specifically, in clustering analyses, Hennig [24] emphasis the invalidity of searching true/natural/real clusters or underlying clusters, and instead discusses some characteristics that need to be met by clustering approaches which imperatively depend on the data and the context (i.e., problem specific). As revealed in many studies, different clustering results can be obtained from the same dataset. Although the results seem like contradict each other, such a situation can mean that use of different information, different perspectives and aims may cause this. In such a confusing environment, we need a guide to ensure that we operate on a scientific framework. In order to develop and assess an aim-dependent clustering approach, we summarize several principles which are accordantly deduced from the comprehensive discussions and considerations of Hennig's [24, p.60] and Gelman and Hennig's [30, p.978] (see Table 1.1). The clustering approach proposed in this study takes into account the conformity with the principles stated in Table 1.1.

Table 1.1 Principles to be considered as a guide while developing and assessing clustering approaches.

P1. *Specifying the context and content of the clustering*

- Transparently expression of the clustering context (i.e. aim/goal/intention).
- Definition of the problem specific "true" cluster.
- What quantitative practices and results can be achieved?
- Indication of potential limitations.
- Disclosing links among coverage, procedures, assumptions and limitations.

P2. *Impartiality*

- Consideration of relevant knowledge and researches.
- Transparent comparison of clustering methods.
- Stating the explicit contents of compared methods.
- Eliminating potential sources of biases/arbitrariness as much as possible.

P3. *Correspondence to observable reality*

- Compatibility of methods and concepts to observables.
- Explicit reproducibility and testing conditions.

P4. *Indications of context dependency*

- Explanation of relatedness to context and aims.
- Distinguishing subjective point of views and intentions.

P5. *Indications of context-driven and data-driven decisions*

- Consequences of multiple perspectives (choices, decisions etc.).
- Definition of dissimilarity.
- Clearly detailed procedure and computation.

P6. *Investigation of stability*

- Existing of reliable, consistent and coherent conclusions.
- Transparent discussion on reproducibility of conclusions.

[a] Adapted from "Beyond Subjective and Objective in Statistics", Gelman, A. and Hennig, C., 2017, *J.R. Statist. Soc. A, 180, part 4*, p.978.

[b] Deduced from "What are the True Clusters?", Hennig, C., 2015, *Pattern Recognition Letters, 64*, p.60.

1.2 Time Dependency of Data and Clusters

As we stated before in former section, Kant [22, p.535] emphasis on conditions of appearances (as we might call the what is observed) that are strictly bounded with space and time. Since Kant's thoughts about time are seen as an important milestone in philosophy, we want to pursue some of his ideas from our perspectives briefly.

The being of time (i.e., given a priori acc. Kant) underlies all appearances, and they become actual only in time. This axiom comprises causality which is a category of relation, i.e., an aspect of observable reality, and causality is unknowledgeable without time [32, p.126], [22, p.199]. According to Kant [22, p.92], time is one of two sources of cognition where the other one is the space that we do not want to dwell on it for now. Hence, Kant in [22, p.89] asserts that the time possesses empirical reality and it is objectively valid with regard to appearances. This means that the observations we made in nature are some sort of realizations of underlying mechanisms, not the very being of those mechanisms. All we can able to deduce is the conditions of these representations/experiences/appearances which are time-dependent. This leads us to unpretentiously say that the time dependency possesses the observable reality. By this argumentation, we want to imply that the observable (i.e., empirical) reality cannot be thoroughly accessible, intelligible and expressible without considering time dependency.

In the context of our readings we have made so far, we believe that if any data analysis method contain time-dependent parameters will also be coherent with Kant's time-related propositions. Remarkably, Mittelstaedt in [33, p.848], states that the strategy that physics uses to examine objects follows the conceptual program formulated by Kant, even though the majority of physicists are not aware of it. Specifically Mittelstaedt in [34, p.30], stated that a physical theory should consider the time dependence of empirical events. Hence, approximately identified time-dependent structures/mechanisms of any occurrences will help to intentional comparisons (i.e., inter- or intra-related) and decrease the amount of uncertainty. Bearing in mind that the parameter can not be a result but be a one of the primary cause of anything, Kant seems to imply that time is not a number but a parameter [35, p.60]. However, in statistical models, unlike the time-dependent differential equations, time cannot be included as a parameter on its own, so it is tried to be absorbed by considering time-dependent variables and thereby related parameters in models.

As an intrinsic dimension (i.e., given a priori acc. Kant) to all observations, time and time dependency should somehow be questioned in any data analysis. Sometimes, time dependency is ignored or avoided by using aggregated data, and treated as itself when plausible tools are available. Specifically, in clustering or classification, the issue of time-dependency begins to emerge in the 50's (see [36–42] for example) or maybe even earlier. However, compared to the relatively advanced literature of the time series analysis along that time, time series clustering studies have begun to adapt to the existent literature about time-dependency analysis since the 90's. While the reasons for this is difficult to identify, it can be said that the increase in the storage and processing facilities of data makes it possible to use and interpret the information derived from the time dependency of datasets [43].

The clustering of time series differs from other clustering tasks, depending on how the time-dependency of the data and clusters plays a role in a clustering approach. In addition to the time dependency of the data, it should be taken into account that the clusters that will be created may be time-dependent and may change over time even though it is ignored generally. There are so many different aims can be described for time series clustering (see [43]). Goals of time series clustering can be generalized as follows: grouping the set of time series according to discovery of a hidden motif/pattern/anomaly, clustering time points of a single/univariate time series and clustering the set of univariate or multivariate time series data into groups of similar time series with respect to some sort of similarity perspective. More discussion of the literature about time series clustering will be given in Section 2.

In this study, the research is structured around the general purpose of clustering a set of univariate and multivariate time series data where each time series (i.e., univariate or multivariate) is treated as an object to be clustered. Moreover, the research also addressed the time-dependency of clusters. If available as an outcome or a by-product of the clustering approach, monitoring the formation of clusters across time besides the time series clusters formed for a certain point of time could contribute many information-theoretical benefits about the question of interest.

1.3 Objective of this Thesis

The conventional scientific attitude from antiquity onwards for modern era is organized to thinking about the world through concepts such as unity, identity, analogy, resemblance, and relationship which are also constituents of categorization [44]. This fact is closely related to the ability of categorization which is the primary cognitive act of humans who are trying to understand the nature and mapping the world surrounding them [45,46]. In this sense, categorization being existed before the scientific methodology, and inherently it is the most fundamental stage of scientific understanding [47]. There are principal and vital reasons for performing categorization (i.e., clustering or classification) such as having perception, knowledge, judgment, information retrieval, and even language acquiring. Furthermore, we are grouping things in terms of their similarities and giving them a name or a concept to conceptualize and interiorize those things around us. In this regard, it is not surprising that most of the studies on the classification and clustering made in recent years are in the areas of artificial intelligence (see Figure 1.2).

On the other hand, we tend to categorize things through conventional ways of perceiving that de-emphasize or obliterates forms of differences that might otherwise improve our knowledge [48, p.55]. When categorization is not responsive to some fundamental dissimilarities but instead rely on some shape based features or vague analogies, which should not be considered as authentically distinctive features, we might form groups with things that are not similar in fact. This is the problem of loss of information caused by shortcut assessment. Therefore, we cannot enrich prospects for different applications and advance the inference by ignoring what is essential to distinguishing. Although the usefulness of such habitual categorization has been proven in practice, a method that is focused mostly on the intrinsic(i.e., structural) features, and sensitive to the evolution, dynamism, and change of categories can be devised which would lead to different perspectives and enhance understanding.

This dissertation considers clustering of a set of real-valued univariate or multivariate time series data where each time series is treated as an object to be clustered. In a broad sense, this study aims at distinguishing time series according to characteristics of their sources.

Before specifying the context and content of the proposed clustering approach, some featured remarks accompanying the former sections and supported to the context can be summarized as follows:

- Categorization is the core of cognitive act and the constant need to mapping the world. Consequently, there are many applications in science.
- Clustering is an unsupervised categorization task, i.e., class properties are not pre-defined or structured in clustering.
- The observable nature is composed of unique events or entities. Finding identical things except for mathematical objects is improbable.
- The difference is a measurable phenomenon at every scale, i.e., it is possible to find at least one feature to distinguish between things (e.g., time series objects).
- Aiming at finding 'true' cluster is meaningless unless the content of the 'true' is defined within context.
- Time dependency possesses the observable reality, and observations are realizations of underlying mechanisms at a certain moment or interval in time.
- The observable (i.e., empirical) reality cannot be intelligible without considering time dependency, i.e., conditions of realizations.
- Time series and thus its clusters are time-dependent, and therefore, time dependency structure of time series should play a key role in time series clustering.

Time series clustering is an unsupervised grouping of unlabelled multiple univariate or multivariate time series into homogeneous clusters. The resulting clusters are desired to meet the minimum dissimilarity within-groups and the maximum dissimilarity between-groups. Time series can be considered as multidimensional/high-dimensional objects, and direct usage of high-dimensional raw sequences, which are especially long in length and possessing time-dependent properties, is not feasible in clustering of time series objects. Hence, there will be a problem of high-dimensionality prevents comparisons based on raw time series, and almost all time series clustering studies utilize a clustering strategy such as model or feature based or a mixture of them, which also overcomes a dimension-reduction problem [43, 49]. Many of features can be defined, selected and used for clustering of time series. However, if the features to be used for clustering do not carry information directed to nature, behavior, and generating process of the series, then clusters obtained with the help of these features would not be accurate [50].

One of our main aims is to trace comparable authentic/idiosyncratic properties and footprints from underlying data generating mechanisms (DGMs) of time series. The clustering approach we propose in this study is based on the distance based clustering of feature vectors and matrices composed via the multifaceted time series model estimations. Hereby, the approach we propose for time series clustering should be considered as a mixture of model and feature based clustering strategies. Within the context of the clustering approach proposed in this study, 'true' clusters consists of series that are similar in terms of resemblance of their underlying DGMs.

The important research questions we are addressing here are: how can we get information about underlying DGMs and make appropriate groupings according to this information? One of the common and feasible ways to make inferences about underlying DGMs is to consider time series models. Models of data or time series are guided, adjusted and idealized representational versions, i.e., theoretical form, of empirical reality [51]. Time series modeling is the process of an understanding the behavior of flux of a set of data over time through a theoretical form termed a model. Although there are certain procedures available to find the most accurate model for the data, it is not possible to say that the ultimately selected model is the fundamental structure that produces the data. Considering that the observed data is a limited realization of the underlying data generating process, any selected model can only be an approximation to the underlying DGM. This view shortened by the well-known quote of G.E.P. Box: "All models are wrong, but some are useful" [52, p.2]. Now we made two basic assumptions here. First one is that there is an underlying mechanism that produces each time series. Secondly, this generating mechanism can never be fully understood nor can be exactly represented by a model but can only be approximated.

The approach we propose does not rely on finding the most accurate model for each time series to be clustered. Finding the most appropriate model for each time series could not be useful to compare time series in terms of their sources. For each one of the time series to be clustered, it is most likely to find more than one adequate model, and although these models to be estimated may contain essential information about the underlying DGMs of the series, they will remain as an approximation and could not be clustered, since each estimated model will fully individuate and include structural differences. Instead, it is more feasible to determine a specific multifaceted time series model and search similarities of the series in terms of their amount of characteristics and aspects they do share or not share with this specific time series model. This idea of clustering we propose is similar to the use of a specific ruler to measure things and compare them with respect to what are measured. But in our problem, the ruler should be a multi-purpose ruler that can able to measure numerous features. The peripheral devices of the model selected for model-based comparisons in this study, which is named *double threshold garch model* and detailed in Chapter 3, are comprehensive and inclusive that it contains various time series properties we actually encounter in time series analysis literature. In this regard, the model is rich in terms of its environment and can manifest many essential features about time series underlying generating mechanisms that can be used for clustering. Although the detailed information is given in Chapter 3, it is worth to note here that the nature of revealed features make possible to use of a graph-theoretic clustering (i.e., graph partitioning) method, which is called *spectral clustering*, since the feature vectors to be collected can be considered as graphs.

In the light of ideas stated in Chapter 1 and especially in this Section 1.3, we would like to sum up the objectives of this study as follows:

- The primary objective of this study is to propose a time series clustering approach for a set of univariate or multivariate time series data to be clustered as objects.
- The approach we propose aims to distinguish the time series according to similarities/dissimilarities of the characteristics of the underlying generating mechanisms that are the source of the observed data.
- The clusters created are aimed to bring together time series that resemble each other in terms of their sources, i.e., 'true' clusters consists of series that are similar only when their underlying DGMs are similar
- Underlying data generating mechanisms can be approximated by the multifaceted time series model.
- The context of the clustering approach we propose in this study is establishing a distance-based clustering strategy, which can operate on feature vectors and matrices extracted from estimations of the time series model.

1.4 Thesis Outline

The present study consisted of six chapter, and organized as follows. Philosophy understands, comprehends and communicates the outcomes created by science. Hence, Chapter 1 starts with philosophical aspects of some basic concepts encountered in clustering and also briefly declares the objective of the research. Chapter 2 provides the background on the time series clustering briefly and summarizes a review of numerous studies in the literature. Chapter 3 presents the proposed time series clustering approach for a set of univariate and multivariate time series data and gives detailed information about the motivation, the double threshold garch model (DTGARCH) and the spectral clustering. Chapter 4 and 5 presents and evaluates the performance of the proposed approach by various artificially generated datasets and real-life data. Finally, Chapter 6 summarizes the main conclusions and findings, presents the remaining and new questions, and gives some thoughts about potential future works.

CHAPTER 2

TIME SERIES CLUSTERING: A REVIEW

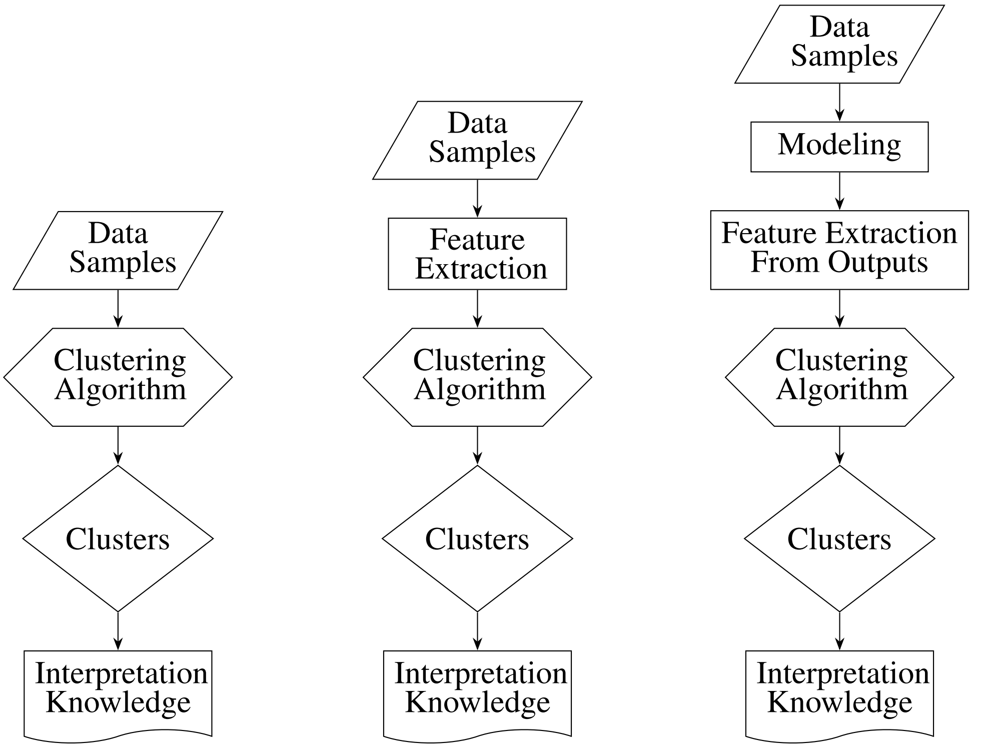
We introduced a brief discussion of the thinking background mediating the development of clustering approaches in Chapter 1. In addition, we appended the consideration based on theoretical and practical arguments on the differentiation of clustering approaches according to specific clustering aims.

The clustering (including categorization and classification), which has a prominent capacity and potential to analyze and understand the nature surrounding us, has led to many scientific types of research and techniques on the subject. Accordingly, with the enormous improvement in computers' data processing and computational capacities, clustering researches has increased considerably. As a result, a vast literature has emerged consisting of diverse and rich approaches to clustering aim.

Since it is impossible to refer to all clustering literature, we prefer to include a limited literature reference in this study. The studies we refer to in this section mainly focus on the clustering of univariate and multivariate time series. Accordingly, this section is designed to assist the reader find where the proposed approach locates within the extensive literature on the subject.

Although the methods developed for clustering vary significantly in terms of their purposes, to see "the big picture" in a way, it would be reasonable to evaluate all methods under three different categories [49]. As shown in Figure 2.1, clustering methods could be regarded basically under three different approaches. Some methods use the raw data directly, while others perform clustering analysis using statistical or basic features derived from the raw data. Extracting the features on a model basis firstly and then clustering those features can also be considered as a third approach.

In brief, since the clustering approach proposed in this study aims to cluster time series using features extracted based on a refined time series model, the proposed clustering approach could be considered under the general category of clustering methods given in Figure 2.1 - (c).



(a) Raw data based

(b) Feature based

(c) Model and Feature based

Figure 2.1 General clustering procedures: (a) raw data based, (b) feature based, (c) model and feature based

Time series datasets are simply encountered in almost every field (e.g., economics, finance, astrophysics, engineering, biomedical sciences, signal processing etc.), and distinguishing groups consisting of similar time series can often be a research problem. As mentioned earlier, grouping unlabeled datasets is called clustering. In the clustering of time series, it is possible to consider time series as objects to be clustered; on the other hand, grouping the observations of a series can also be a research topic (i.e., within-sample clustering). However, this study proposes an approach for clustering time series, in which time series are treated as objects to be clustered (i.e., being objects as realizations originating from unknown data-generating mechanisms).

Thanks to the recent literature surveys of Liao [49], Aghabozorgi et al. [43], Ali et al. [53], it has been possible to reach the extensive literature and capture a general view about the time series clustering. The first two of these surveys received over 2000 citations in total from articles published in peer-reviewed journals.

Eventually, due to the increase in the number of approaches, methods, publications and the extent of their scope, the problem of time series clustering and classification has been recently examined in a volume by Maharaj et al. [54].

The clustering performance of the proposed time series clustering approach in this study is examined on artificially generated data and real data sets as univariate time series clustering in Chapter 4 and multivariate time series clustering in Chapter 5. In the test phase of the proposed approach, we conducted the performance evaluation comparatively with state-of-the-art methods in the literature. These methods essentially offer dissimilarity measures to be used for time series clustering. In this context, we had the opportunity to compare 22 distance measures proposed for univariate time series clustering accompanied by three different clustering procedures (i.e., FUZZY, PAM, and SPECTRAL clustering procedures) in Chapter 4. Again, in Chapter 5, to compare the performance of the multivariate counterpart of the proposed clustering approach, we considered five different distance measures for clustering multivariate time series. Within reach of this study, while citing the studies that we will compare, we acted from the perspective that the applications of these methods should be easily accessible and reproducible on an open-source platform such as **R**. For this reason, we consider that the computational codes of the methods we compared should be on an accessible platform by independent researchers. Fortunately, we could employ the **R** libraries **TSclust** by Montero and Vilar [55], **pdc** by Brandmaier [56], and **dtwclust** by Sarda [57], which also offer an implementable literature. These **R** libraries and associated publications provide an essential starting point for comparative studies. Nevertheless, in this respect, we are aware that there are many clustering methods that we could not include in the comparison study because their computational implementations are not publicly available. However, we want to mention several studies in this section that we could not find the chance to implement and include them in the scope of future studies.

2.1 Univariate Time Series Clustering

The majority of the literature on time-series clustering consists of approaches developed for univariate time-series datasets. The history of the first competent studies in time series clustering goes back to the early 1990s [58, 59]. It could be said that, following the increase in time-series data collecting and processing capacities, time series clustering applications and the methods developed for these applications expand toward considering multivariate time-series datasets.

As mentioned before, time series clustering is usually performed by measuring and comparing the similarity or dissimilarity between specific features extracted from time series. Some approaches, albeit few, rely only on comparing raw data without feature extraction. Feature extraction can be based on a particular function or a model. The information of similarities or dissimilarities (i.e., which is generally provided in the form of a matrix called dissimilarity matrix) between a set of time series is obtained through distance metrics proposed for time series clustering.

Within the scope of this study, we list some of the proposed distance measures we employed for the comparison study in Tables 2.1, 2.2, and 2.3. Tables 2.1 and 2.2 denote model-free methods, while Table 2.3 lists model-based clustering methods.

Distance is a concept that allows us to obtain a numerical measurement of how points or objects are far from each other. In general usage, distance is an estimate of the length between objects based on a physical space. In mathematics, similar to physical distance, distance is a positive real-valued function used to express how the points in a particular space are "close" or "far" from each other. For example, the distance from point X to point Y can be denoted as $D(X, Y)$.

Let's assume that $X_T = \{x_1, \dots, x_T\}$ and $Y_T = \{y_1, \dots, y_T\}$ are time series of length T from unknown data generating processes. The distance between this pair of time series can be denoted as $D(f(X_T), f(Y_T))$, where $f(.)$ can be a specific function or time series model for feature extraction, and it satisfies following conditions:

- $D(f(X_T), f(Y_T)) \geq 0$ & $D(f(X_T), f(Y_T)) = 0$ iff $X_T = Y_T$
- $D(f(X_T), f(Y_T)) = D(f(Y_T), f(X_T))$

For example, among the distance measurements in Table 2.1, only Dynamic Time Warping (DTW) and Euclidean (EUCL) distance measures can use raw data without any preprocessing. Other distance measures require that the series to be clustered should be considered by subjecting them to somewhat data processing (i.e., feature extraction). In the case where the function, $f(.)$, is an identity function (i.e., $f(z) = z$), it can be said that the distance is calculated based on the raw data. An example of this situation can be given as the euclidean distance with the following form:

$$D_{EUCL}(X_T, Y_T) = \left(\sum_{t=1}^{T} (x_t - y_t)^2 \right)^{(1/2)} .$$

Table 2.1 Model free dissimilarity measures for univariate time series clustering.

Dissimilarity Measure	Description	Reference
	Model free approaches	
D_{ACF}	Autocorrelation based distance.	[60]
D_{PACF}	Autocorrelation based distance.	[60]
D_{EUCL}	Euclidean distance.	[55]
D_{CORT}	Dissimilarity index that covers both dissimilarity on raw values and dissimilarity on temporal correlation behaviors.	[61]
D_{COR}	Correlation-based dissimilarity.	[62]
D_{DWT}	Dissimilarity for time series based on wavelet feature extraction.	[63]
D_{PER}	Periodogram based dissimilarity.	[64]
$D_{INT.PER}$	Integrated periodogram based dissimilarity.	[65]
$D_{SPEC.LLR1}$	General spectral dissimilarity measure using local-linear estimation of the log-spectra via least squares estimation.	[66, 67]
$D_{SPEC.LLR2}$	General spectral dissimilarity measure using local-linear estimation of the log-spectra via maximum likelihood estimation.	[66, 67]
D_{GLK}	Dissimilarity based on the generalized likelihood ratio test.	[68, 69]
D_{ISD}	Dissimilarity based on the integrated squared difference between the log-spectra.	[69]
D_{SAX}	Symbolic aggregate approximation related functions.	[70, 71]
D_{VR}	Distance based on variance ratio statistics.	[72]
D_{DTW}	Dynamic time warping distance.	[59, 73]

Such use of Euclidean distance would not include any temporal information for time series clustering, especially when long time series are considered because this distance measurement cannot cover the time dependencies of observations. However, the Euclidean distances between the autocorrelation (ACF) and partial autocorrelation functions (PACF) of the series can provide distinctive information for time series clustering, as shown by Peña and Galeano [60]. Similarly, as Choukakria et al. [61] and Golay et al. [62] have shown, some studies and applications suggest grouping time series using the features extracted from correlation structures of time series to be clustered. It should be noted that DTW [59, 73] and EUCL-like distance measurements can provide useful distances between such features obtained in sequential form. In addition to distance metrics such as D_{DTW}, D_{ACF}, D_{PACF}, D_{CORT}, and D_{COR} which especially focus on the shape similarities and the correlation structure, there are some other useful distance measures such as D_{PER}, $D_{INT.PER}$, $D_{SPEC.LLR1}$, $D_{SPEC.LLR2}$, D_{GLK}, D_{ISD} proposed by Caiado et al. [64], Casado [65], Kakizawa et al. [66], Vilar and Pértega [67], Fan & Zhang [68], and Díaz & Vilar [69] respectively. These distance measures operate on the frequency domain of time series and mainly utilize the estimates of spectral densities of times series to differentiate them. Among the model-free distance measures listed in Table 2.1, D_{DWT} proposed by Zhang et al. [63] measures distance based on wavelet-based features, D_{SAX} proposed by Lin et al. [70] & Keogh et al. [71] proposes distance measurement based on symbolic time series representations, and the distance measure D_{VR} proposed by Bastos & Caiado [72] uses a Variance-Ratio based distance metric for financial time series clustering.

The complexity-based dissimilarity measures based on the probability and information theory such as a distance of permutation distribution clustering (D_{PDC}) by Brandmaier [56], a complexity-invariant dissimilarity measure (D_{CID}) by Batista et al. [74], a compression-based dissimilarity measure (D_{CDM}) by Keogh et al. [75], and a normalized compression distance (D_{NCD}) by Cilibrasi et al. [76] given in Table 2.2 are not model dependent distance measures but also differ from the model-free distance measures that are employing the quantitative descriptive features of time series. These measures mainly attempt to measure the information-theoretic dissimilarities between time series by approximating their level of shared mutual information.

Table 2.2 Complexity based dissimilarity measures for univariate time series clustering.

Dissimilarity Measure	Description	Reference
	Complexity based approaches	
D_{CID}	Complexity-invariant distance measure for time series.	[74]
D_{PDC}	Permutation distribution distance.	[56]
D_{CDM}	Compression-based dissimilarity measure.	[75]
D_{NCD}	Normalized compression distance.	[76]

Model-based dissimilarity measures for time series clustering mainly rely on the knowledge of underlying generating mechanisms. Various model-based distance metrics and model-based clustering methods have been proposed in the literature (see, e.g., Maharaj et al. [54] for more details). Model-based distance measures such as D_{PIC} by Piccolo [58], D_{MAH} by Maharaj [77], and D_{LPC} by Kalpakis et al. [78], which we also employed for comparison objectives in Chapter 4 and have an essential standing in the time series clustering literature, are given in Table 2.3.These methods use the Box-Jenkins' (ARIMA) [79] parametric model family, a well-known model family in the time series analysis literature. They provide good results for clustering time series regarding their autoregressive and autocorrelation-like structures, which are determined mainly by the underlying conditional mean process. However, it can be said that they may be insufficient in time series clustering when the underlying conditional variance processes reveals the presence of possible heteroscedasticity in time series realizations [50].

Table 2.3 Model based dissimilarity measures for univariate time series clustering.

Dissimilarity Measure	Description	Reference
	Model based approaches	
$D_{AR.LPC}$	Dissimilarity based on linear predicitive coding (LPC) cepstral coefficients.	[78]
$D_{AR.MAH}$	Model-based dissimilarity proposed by Maharaj.	[77, 80]
$D_{AR.PIC}$	Model-based dissimilarity measure proposed by Piccolo.	[58]

2.2 Multivariate Time Series Clustering

The number of applied studies on the clustering of multivariate time series is considerably few compared to the studies on the clustering of univariate time series. One reason is that the datasets studied have diversified and extended in volume over the last two decades. Another reason is that the computer technologies that make it possible to perform computations on such datasets have yet to be developed. Of course, although there are few, there have been early studies dealing with multivariate time series clustering (see, e.g., Gersch & Yonemoto [81], Sanderson & Wong [82]), but in a research area that requires such computational processing, the course of research has been primarily determined by the purpose of clustering, and data processing facilities.

The dissimilarity measures used in this study's comparative experimental section of temporal multivariate time series clustering are given in Table 2.4. Some of the dissimilarity measures mentioned here can be applied to both univariate time series clustering and multivariate time series clustering. For example, D_{JD} and D_{LD} are multivariate counterpart of the model-free distance measures $D_{SPEC.LLR1}$ and $D_{SPEC.LLR2}$ proposed by Kakizawa et al. [66]. Another model free dissimilarity measure D_{PDC} proposed by Piccolo [58] and D_{DTW} made available by Giorgino et al. [73] are also applicable to multivariate time series datasets to be clustered. The dissimilarity measure based on Global Alignment Kernels (GAK) proposed by Cuturi [83] is also another model-free distance measure listed in Table 2.4.

Table 2.4 Dissimilarity measures for multivariate time series clustering.

Dissimilarity Measure	Description	Reference
	Model free approaches	
D_{JD}	Multivariate Spectrum based Distance. Also known as symmetrized Kullback-Leibler (KL) Divergence. Also known as J(effreys)-Divergence.	[66]
D_{LD}	Multivariate Spectrum based Distance. Known as Log-Spectral Divergence.	[84]
D_{GAK}	Distance based on Global Alignment Kernels	[83]
D_{DTW}	Dynamic Time Warping Distance.	[59,73]
D_{PDC}	Permutation Distribution Distance.	[56]

The dissimilarity measures D_{JD} and D_{LD} operate on frequency domain while D_{GAK}, D_{DTW} and D_{PDC} compute dissimilarities based on time domain. Of course, there are some application areas where each dissimilarity measure mentioned here performs quite successfully and gives very satisfactory results in terms of clustering. For example, the use of D_{DTW} and D_{GAK} measures to group time series by mainly their shape properties provides good outcomes, especially in short time series [85]. However, the artificially generated time series we examined in the experimental study sections are relatively long time series produced by conditional mean and conditional variance structures. They simultaneously represent several characteristics such as non-stationarity, periodicity, seasonality, aperiodicity, heteroscedasticity, etc. On the other hand, the results of the experimental study in Chapter 5 indicate that the dissimilarity measures D_{JD}, D_{LD}, and D_{PDC} provide reasonable results when the correct grouping of time series is dependent on only comprehending autocorrelation (i.e., revealing conditional mean structures) structures of time series.

When the time series (i.e., univariate or multivariate) clustering literature is examined, it will be seen that model-based approaches proposed for clustering of time-series datasets are much less than data-driven approaches. Studies in recent years have mainly focused on approaches based on data-driven feature extraction instead of proposing algorithms that take into account the statistical structures of underlying mechanisms that produce the time series data. One of the most important reasons for this is the need for data-based algorithms that accumulate momentum following the settled naive understanding of the machine learning concept. However, we want to point out some of the recently published and noteworthy studies that we could not include in comparative experimental sections of the study since their algorithms' computational codes are not readily available. To give examples of these noteworthy works, for example, Roick et al. [86] proposed a finite mixture model-based (i.e., Integer valued Autoregressive (INAR) type models) time series clustering. Fröhwirth-Schnatter & Kauffman [87] also proposed a time series clustering method using finite-mixture models for panel data. Nieto-Barajas & Contreras-Cristán [88] proposed a model-based time series clustering using Poisson-Dirichlet process. Delft & Dette [89] proposed a novel dissimilarity measure considering time dependent spectral densities of non-stationary time series. Dias et al. [90] used Hidden Markov

Models (HMMs) for clustering of financial time series. Alonso & Peña [91] proposed a feature-based distance measure based on linear dependencies (i.e., by considering cross correlations) of time series. D'Urso et al. [92] proposed a feature-based clustering method using cepstral weights revealing the logarithm of the spectral densities. Deb [93] proposed a Vector Autoregressive (VAR) model based clustering method for multivariate time series. Ghassempour et al. [94] proposed model-based clustering approach using HMMs. López & Vilar [95] proposed a feature-based dissimilarity measure operating on cross-spectral densities for multivariate time series clustering. Li & Wei [96] proposed a feature-based multivariate time series clustering approach that utilizes DTW and shape-based distance measures. Vázquez et al. [97] proposed a multivariate time series clustering approach that ensembles several time series representations and distance measures. Li & Liu [98] proposed a multivariate time series clustering method based on complex networks that recasts the time series on grids according to specific mapping methods.

2.3 Proposed Contributions

Almost all of the methods proposed for time series clustering make the conventional clustering methods, which are primarily proposed regarding static datasets, functional by retrieving (i.e., staticizing) the time-dependent properties of the data. It can be said that the approach proposed in this study follows a similar course. Although this conclusion is principally correct, the proposed approach considers extracting the feature vectors/matrices of the time series as time-dependent as possible. Thus, the dissimilarity matrix required for the conventional clustering methods (i.e., k-means, fuzzy c-means, or pam) to function has been obtained iteratively and additively over the entire period of the data. Accumulating the dissimilarity information through iterative computation ensures that the dissimilarity matrix has a time-dependent structure as far as possible.

We expect the approach proposed for time series clustering in this study will contribute to the broad literature on the following points.

The proposed approach;

- essentially, follows a clustering perspective that prioritizes the necessity of grouping the mechanisms that generate the series while dealing with the problem of clustering a dataset consisting of time series.

  (In fact, each time series is the realization of an underlying data generation mechanism that we can not know for sure, and grouping these realizations correctly can only be possible by identifying distinctive features about the mechanisms generating the series. Any time series clustering based only on the properties of realizations will be insufficient to examine time-dependent changes. Therefore, it is intended that the groups to be determined by the proposed approach will also be a grouping of the mechanisms that generate the series.)

- aims to enable clustering using a general framework of competent models, based on the argument that the time series' generating mechanisms can never be determined most accurately.

  (This model, used in the proposed approach, can be used as a competent approximation tool in time series of almost any type and with a particular sample length. The model used in the proposed approach, to put it briefly, acts as a filter and captures distinctive projections of the mechanism generating the series.)

- to capture both linearities an nonlinearities within the series, utilized the double threshold GARCH (DTGARCH) model as the model framework. To the best of our knowledge, this model is used for the first time in the literature for time series clustering.

  (of course, different time series models were also used in the literature.)

- can be used for both univariate and multivariate time series clustering.

  (It should also be noted that the proposed approach can work with higher performance in time series with a certain minimum sample length, i.e., assuming over 350 observations.)

- uses feature vectors or matrices capable of representing series and are obtained using a specific specification of the DTGARCH model. The proposed approach can be evaluated in the category of methods using model-based feature elements regarding the time series clustering literature.

- uses the spectral clustering algorithm proposed by Ng. et. al. [99].

  (Therefore, the proposed approach can be seen as a unique adaptation of the spectral clustering method to time series clustering. Thus, the calculation of the affinity matrix required for spectral clustering is made applicable for time series. In this regard, it makes a contribution to the studies on spectral clustering of time series.)

- takes into account that the clusters to be created for a time series dataset may exhibit time-dependent changes. As such, it adopts a dynamic clustering approach. Time-dependent changes in the mechanisms producing time series would also affect the cluster formations to be obtained. In order to show that the DTGARCH model can work in such situations, various simulation studies have been conducted in Chapter 4 and 5.

CHAPTER 3

PROPOSED APPROACH

This chapter is intended to introduce the novel clustering approach we propose for a set of univariate or multivariate time series data. In short, the proposed approach combines the double threshold garch (DTGARCH) and the AR models (i.e., linear time series model) based feature extraction and the spectral clustering method. The motivation of the proposed approach and related remarks, which are previously conceptualized in Chapter 1, highlighted in Section 3.1. The approach we propose for time series clustering mainly relies on clustering of features extracted from estimations of the DTGARCH model. Section 3.2 presents the details and theoretical background of the DTGARCH model. The extracted features hold time-dependent, idiosyncratic properties about underlying (i.e., latent) DGMs. Thus, after goal-oriented feature extraction phase, the task of clustering of a set of time series becomes a task of clustering of a set of feature vectors that represents time series to be clustered. The distances between those features can be considered as connected graphs in which some path connects vertices/points. The spectral clustering operates well for the solution of a graph-partitioning problem. Section 3.3 gives basics about the graph-theoretic spectral clustering. The details of the proposed time series clustering approach for a set of univariate and multivariate time series data, which incorporates two state-of-the-art methods, are given in Section 3.4.

3.1 Motivation

In time series clustering studies, resulting clusters are desired to form as 'homogeneous clusters' as possible within the context of a similarity perspective. By and large, regarding the aims of clustering, time series clustering studies can be generalized into three categories as follows: (1) clustering the set of time series relating to discovery of a latent pattern; (2) clustering time points of a univariate time series; and (3) clustering a set of univariate or multivariate time series data where each time series considered as objects to be clustered. More specifically, it can be said that each study includes at least one perspective or aim differs from the others. As previously discussed in Chapter 1, individuated aims and perspectives of studies may expose different information that leads to different cluster formation for even the same set of data. Although it is not possible to find a generalized method for each specific time series clustering problem, the transparently expressed context and content of an any developed clustering method can allow us to produce meaningful and comparable results at a scientific level.

In this study, we propose a novel time series clustering approach for a set of univariate and multivariate time series data where each time series data treated as objects to be clustered. Clustering task of a set of time series data has its applications in a wide variety of fields such as environmental, computer, engineering, astronomy, economic, and neuro sciences (see Figure 1.2). If time series data is considered as an object to be clustered, derived information about underlying DGM of time series should essentially be taken into account rather than consideration of transient features such as shape, motif, and pattern. Succinctly, the context of the proposed approach is finding out comparable idiosyncratic properties of underlying DGMs of time series to be clustered. Thus, within the context of the proposed approach, the 'true' cluster is a cluster that groups time series regarding the similarities of their underlying DGMs.

Two of the basic research questions to be considered here are as follows: (1) how can we get information about underlying DGMs; and (2) how to make appropriate groupings based on these information?

(1) How can we make inferences about the underlying DGM of a time series that is intended to be clustered as an object?

Time series are limited realizations of their underlying DGMs, and time series models are plausible tools to make inferences about their underlying generating processes. If it were possible to find the very structure of the time series generating mechanism exactly, this could have enabled us to make an absolute or a perfect classification. Nevertheless, the fact that a time series is only a limited realization of its underlying dynamics prevents us from obtaining certain and complete knowledge about the mechanisms that produce time series. Naturally, finding out the most accurate model for a time series includes meaningful information about its underlying generating dynamics, but in the end, this is no more than an approximation. Considering that the aim is to discover an appropriate way of approximation for the purpose of clustering, finding out the most accurate models for time series to be clustered would not be useful. To make underlying DGMs' approximations comparable for a clustering purpose, common/shared ground/basis must be provided for measuring similarities/dissimilarities. To this end, we utilize a multifaceted time series model named as DTGARCH besides the autoregressive (AR) time series model.

(2) How to make appropriate clustering according to extracted information related to underlying DGMs?

Utilizing the specific AR and DTGARCH model approximations to extract features about time series objects to be clustered is a crucial phase of the proposed approach called feature extraction. Thus, goal-oriented feature extraction in the proposed approach is accomplished by considering parametric time series model based outputs of estimations. The DTGARCH model is rich in its environment/background and can reproduce/represent numerous properties defined thus far in time series analysis literature such as stationarity, non-stationarity, chaotic behavior, linearity, non-linearity, seasonality, abrupt changes, and regime switching. Parameter estimates of the specific AR and DTGARCH models and autocorrelation structures of the residuals and squared-residuals are used to form feature vectors to represent time series. These feature vectors contain/hold several time-dependent properties of time series up to a certain level that is sufficiently useful for comparisons. Differences (i.e., dissimilarities) between the properties extracted from approximations to underlying DGMs by DTGARCH and AR models estimates creates connected graphs. Clustering of features by using dissimilarity matrix computed on those features can be considered as a graph partitioning. For this purpose, we employ the spectral clustering method in the proposed approach, which works satisfactorily for the graph partitioning problems.

Another important motivation of the proposed cluster approach is to make it capable of capturing the possible changes in the number of clusters and clusters across time. As we encounter in time series data analyses, the mechanisms that produce time series may change over time, or time-dependent characteristics of data may vary in time. Therefore, when newly observed data are added to the available set of data, time series model based feature extraction step of the proposed approach also makes the clustering approach data-adaptive to examine temporal dependencies.

We would like to complete this sub-section with an analogy, which illustrates the motivation of the proposed approach. Feature extraction phase of the proposed approach resembles the feature extraction principle of the spectroscopy (see Figure 3.1). Spectroscopy, in general, is the investigation of the physical properties of elements, compounds, stars or galaxies by analyzing their electromagnetic radiation (i.e., spectrum). Electromagnetic radiation includes spectrum of visible light and invisible lights such as microwave, infrared, ultraviolet, x-ray and gamma waves. Visible light from each object has its unique source of energy composition. Any object shines out the energy in the form of light which contain all colors. In this sense, the light sources to be examined in their raw form are indistinguishable or non-informative.

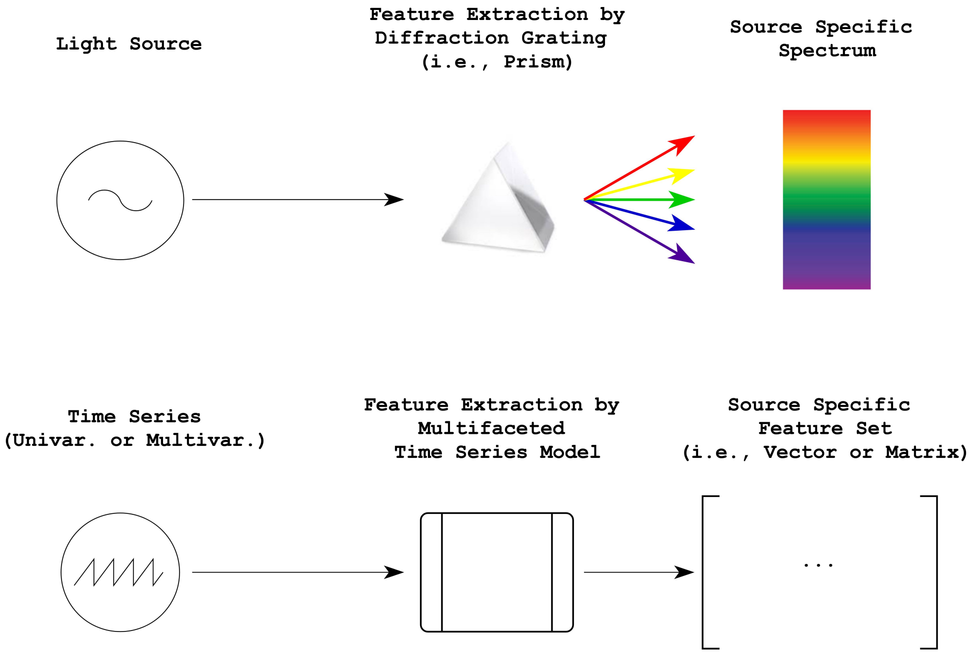

Light Source
Feature Extraction by Diffraction Grating (i.e., Prism)
Source Specific Spectrum
Time Series (Univar. or Multivar.)
Feature Extraction by Multifaceted Time Series Model
Source Specific Feature Set (i.e., Vector or Matrix)
...

Figure 3.1 Analogy between the principle of spectroscopy and the feature extraction.

In order to distinguish or identify the sources of lights, the method of spectroscopy can be used which is rely on to artificially splitting the light into its building blocks or constituents (i.e., colors). The splitting can be accomplished by using a particular device known as a diffraction grating (e.g., prism). There are individual repeating patterns in each element's, compound's or star's spectrum known as dark absorption lines (i.e., spectral features). Chemists, physicists or astrophysicists can identify which combinations of those patterns belong to which chemical element. The light of its matter (e.g., star) contains such fingerprints on its own spectrum and exhibits which chemical elements make up that matter.

3.2 Background on the Double Threshold GARCH Model

In this sub-section, the DTGARCH model is explained and then a procedure is generated for fitting the DTGARCH model to each time series object to be clustered. In particular, we designate a procedure that uses the specific number of regimes, data-adaptive threshold values, and the specific orders of the DTGARCH models for each of the regimes. Moreover, DTGARCH model is a unification of a threshold autoregressive (TAR) model and a GARCH component. Accordingly, the TAR model is mentioned briefly and then DTGARCH model is illustrated.

3.2.1 Threshold Autoregressive Model

The concept of regime alteration controlled by specific threshold parameters was firstly introduced by Quandt [100] in the context of linear regression models. Thereafter, the function of the regime switching, which is regulated by the threshold parameters, was extended to time series models [101] and [102]. Since its initiation, the TAR model has been widely applied to different type of time series data. As addressed in [103], TAR models are useful because of their ability to generate and capture non-linear dynamics, limit cycles, severe jumps, and asymmetries. The purpose of TAR is to define regimes through thresholds and get local approximation over the regimes. The thresholding is effective because it decomposes a complex stochastic system into simpler subsystems, and thus allowing expeditious implementation and more straightforward interpretation. TAR models can represent data that manifest abrupt changes of large amplitude at arbitrary time points which are usually exhibited by numerous types of time series data (see [104, 105] for a review of the TAR models via a wide variety of applications).

A general TAR model with k regimes is represented as

$$y_t = \sum_{j=1}^{k} \left[(\phi_0^{(j)} + \sum_{i=1}^{p_j} \phi_i^{(j)} y_{t-i} \ + \ \sigma^{(j)} e_t^{(j)}) \ I_{(r_{j-1}, r_j)}(z_{t-d}) \right] \tag{3.1}$$

where y_t is the observed time series at time t; $j = 1, \ldots, k$ denotes the regime and k is the number of regimes; p_j is for j^{th} regime autoregressive (AR) lag order; $\phi_0^{(j)}$ is intercept and $\{\phi_i^{(j)} | i = 1, 2, \ldots, p_j\}$ are AR terms coefficients for j^{th} regime; and $I_{\mathcal{A}}(z)$ is the indicator function that takes the value of 1 when $z \in \mathcal{A}$ and 0 otherwise; $r = (r_1, \ldots, r_{k-1})$, satisfying $r_1 < \ldots < r_{k-1}$, are the threshold values; $r_0 = -\infty$ and $r_k = \infty$; z_{t-d} is the threshold variable with delay parameter d; for each j, $\{\varepsilon_t^{(j)} = \sigma^{(j)} e_t^{(j)}\}$ is a sequence of martingale differences satisfying $\mathbb{E}(\varepsilon_t^{(j)} | F_{t-1}) = 0$, $sup_t \mathbb{E}(|\varepsilon_t^{(j)}|^{\delta} | F_{t-1}) < \infty$ *almost surely* for some $\delta > 2$, where F_{t-1} is the σ-field generated by $\{\varepsilon_{t-i}^{(j)} | i = 1, 2 \ldots; j = 1, \ldots, k\}$; e_t is *i.i.d.* with $\mathbb{E} e_t = 0$ and $\mathbb{V}\mathrm{ar}\, e_t = 1$ and $\sigma^{(j)} > 0$.

For example, if at time t, $z_{t-d} = a \in (r_{\ell-1}, r_\ell)$ then the active regime at that time is characterized by the AR model as follows:

$$y_t = \phi_0^{(\ell)} + \sum_{i=1}^{p_\ell} \phi_i^{(\ell)} y_{t-i} \ + \ \sigma^{(\ell)} e_t^{(\ell)}.$$

Thus, describing different regimes within time series by threshold variable and values allow us to obtain a temporal behavior of a dynamics. The unknown parameters for the above model are collected in the vector

$$\Omega = \left[(\phi_0^1, \phi_{i:p_1}^1), \ldots, (\phi_0^k, \phi_{i:p_k}^k), r = (r_1, \ldots, r_{k-1}), k, d \right],$$

and can be estimated by least-squares (LS) estimation (which is equivalent to the maximum likelihood estimation under the assumption that e_t is i.i.d.$N(0, 1)$). The LS estimator of the parameter vector Ω can be obtained by the minimization problem

$$\widehat{\Omega} = \text{arg min}_{\Omega} \sum_{t=1}^{T} \left[y_t - (\phi_0^{(j)} + \sum_{i=1}^{p_j} \phi_i^{(j)} y_{t-i}) I_{(r_{j-1}, r_j)}(z_{t-d}) \right]^2.$$

The asymptotic properties of the LS estimator for the TAR parameter are developed in [106]; statistical inference of the TAR models is discussed in [107]. Some applications on Bayesian inference for TAR models are developed in [108, 109]; and [110]. Furthermore, there is a strong interest for testing linearity versus threshold non-linearity, see [111] and [112].

3.2.2 Self-Exciting Threshold Autoregressive Model

The model in Equation (3.1) is called *self-exciting* threshold autoregressive (SETAR) model if the threshold variable is the same as the actual time series, i.e., $z_{t-d} = y_{t-d}$. In this case, the regime of the time series y_t would be determined by its own past value, y_{t-d}.

A general SETAR(k) model where k denotes the number of regimes can be represented as

$$y_t = \sum_{j=1}^{k} \left[(\phi_0^{(j)} + \sum_{i=1}^{p_j} \phi_i^{(j)} y_{t-i} \; + \; \sigma^{(j)} e_t^{(j)}) \; I_{(r_{j-1}, r_j)}(y_{t-d}) \right]. \qquad (3.2)$$

Similar to TAR model, the unknown parameters for the model in (3.2) are

$$\Omega = \left[(\phi_0^1, \phi_{i:p_1}^1), \ldots, (\phi_0^k, \phi_{i:p_k}^k), r = (r_1, \ldots, r_{k-1}) \in [min\{y\}, max\{y\}], k, d \right],$$

and can be estimated by a least-squares (LS) estimation under the assumption that $\varepsilon_t^{(j)}$ is $i.i.d.\ N(0, \sigma^2)$. The minimization of the sum of squared residuals yields the LS estimators:

$$\widehat{\Omega} = \text{arg min}_{\Omega} \sum_{t=1}^{T} \left[y_t - \sum_{j=1}^{k} \left[(\phi_0^{(j)} + \sum_{i=1}^{p_j} \phi_i^{(j)} y_{t-i}) I(r_{j-1} < y_{t-d} \leq r_j) \right] \right]^2, \quad (3.3)$$

where, T denotes the sample size in a time series. The minimization problem of (3.3) can be handled by a grid search over all combinations of possible values of parameters $\{j, p_j, r, d | j = 1, \ldots, k\}$. Consequently, the grid search requires a number of nearly $T^{k-1}/2 \times d \times p_{(1)} \times \ldots \times p_{(k)}$ arranged auto-regressions. Alternatively, the unknown thresholds and parameters can be estimated from a model selection perspective through minimization of a specific criterion (e.g., AIC, BIC or HQIC). Nevertheless, for estimating all threshold parameters one at a time and to reduce the computational cost considerably, Gonzalo and Pitarakis [113] suggested a conditional model selection procedure under an unknown number of thresholds. The procedure starts with deciding between a linear and a one threshold (i.e., two regime) TAR specification. If the existence of a threshold cannot be rejected then the sample can be organized into two subsamples by the threshold. To search for the presence of another threshold is renewed over sub-samples that are conditionally formed on the threshold estimation in the previous step. The iteration ends when the model selection procedure cannot confirm the presence of another threshold over sub-samples. Then, the demanded number of arranged auto-regressions to be estimated moderately decreased to $(T \times d \times p_{(1)} \times \ldots \times p_{(k)} + (k-1) \times T)$.

More formally, let S_T denotes the sum of squared residuals from a p_{ar} order AR(p_{ar}) fit and $S_T(r_1, \ldots, r_{k-1})$ denotes the sum of squared residuals defined in (3.3), the model selection procedure would be based on the optimization of the following objective functions

$$IC_T(p_{ar}) = logS_T + \lambda/T,$$

$$IC_T(p_{(1)}, \ldots, p_{(k)}) = logS_T(r_1, \ldots, r_{k-1}) + (\lambda/T) \times (k-1),$$

where λ denotes a model selection criterion, such as AIC or BIC. The model selection procedure would lead to the choice of a linear AR model if

$$\arg\min_{p_{ar}}(IC_T(p_{ar})) < \arg\min_{p_{(1)}, p_{(2)}, d, r_1}(IC_T(p_{(1)}, p_{(2)}))$$

with $1 \leq p_{ar} \leq p_{(1)} \leq p_{(2)} \leq p_{max}$ and $d \leq min(p_{(1)}, p_{(2)})$. If the above inequality cannot be satisfied, then the optimum threshold number can be found by minimizing the objective function $IC_T(p_{(1)}, \ldots, p_{(k)})$ where the optimal threshold number

$$\hat{k} = \arg\min_{0 \leq k \leq K} logS_T(r_1, \ldots, r_{k-1}) + (\lambda/T) \times (k-1).$$

It is important to note that the threshold parameter estimates, $\hat{\Omega}$, are obtained as a by-product of the number of optimal regime determination procedure. For more detailed information and the statistical properties of the SETAR model, we refer the interested reader to [106], [107] and [114].

3.2.3 Double Threshold GARCH Model

For a given time series $y_t,\ t = 1 \dots T$, the global structure of a time series model consists of two main components known as conditional mean and conditional variance and can be represented as follows:

$$y_t = \underbrace{g(\Psi_{t-1})}_{\mu_t} + \underbrace{\underbrace{\sqrt{\eta(\Psi_{t-1})}}_{\sigma_t} .\ e_t}_{\varepsilon_t}, \qquad e_t \overset{i.i.d}{\sim} D(0,1),$$

where Ψ_{t-1} is the set of information available at time $t-1$; ε_t is a series of innovations (i.e., errors) that is independent of past of the time series, y_t; and e_t is an $i.i.d.$ sequence with distribution D.

In time series analysis, conditional mean, $\mu_t = g(\Psi_{t-1}) = E(y_t|\Psi_{t-1})$, and conditional variance, $\sigma_t^2 = \eta(\Psi_{t-1}) = V(y_t - \mu_t|\Psi_{t-1}) = V(\varepsilon_t|\Psi_{t-1})$, are both can be time dependent and be modeled by numerous parametric [115] or non-parametric [116] time series models. For example, the TAR and SETAR models mentioned in subsections 3.2.1 and 3.2.2 are assuming constant variance in innovations, ε_t, and provide modeling perspective for conditional mean only via autoregressive terms with regime switching mechanism.

In the scope of this study, to extract distinguishable features from a time series in relating to their unknown underlying DGMs, we propose the use of a rich environment in the form of a time series model which incorporates the time dependencies of conditional mean and variance. For this purpose, we utilize one of a complex time series models known as the double threshold generalized ARCH (DTGARCH) model. In this model, conditional mean, μ_t, and conditional variance, ${\sigma_t}^2$, are both regime specific:

$$y_t = \sum_{j=1}^{k} [{\mu_t}^{(j)} + {\sigma_t}^{(j)} {e_t}^{(j)}] I_{(r_{j-1}, r_j)}(z_{t-d}), \qquad {e_t}^{(j)} \overset{i.i.d}{\sim} D^{(j)}(0,1),$$

where y_t is the observed time series; $j = 1, \dots, k$ where k is the total number of regimes; and $I_{\mathcal{A}}(z)$ is the indicator function that takes the value of 1 when $z \in \mathcal{A}$ and 0 otherwise; $r = (r_1, \dots, r_{k-1})$ with $r_1 < \dots < r_{k-1}$ are the threshold values, $r_0 = -\infty$ and $r_k = \infty$; z_{t-d} is the threshold variable with delay parameter d.

Following the seminal work of Tong [101] on the different regime dynamics, Li and Li [117] implemented the ARCH models for the conditional variance part giving rise to the double threshold ARCH (DTARCH) model. Brooks [118] extended it to DTGARCH model to analyze volatility in the exchange rate. Due to its ability to capture complex dynamics, we consider the DTGARCH model for feature extraction of a given time series to be clustered as object. The general representation of the DTGARCH($k, ar_1, q_1, p_1, \ldots, ar_k, q_k, p_k$) model is

$$y_t = \sum_{j=1}^{k} [(\phi_0^{(j)} + \sum_{i=1}^{ar_j} \phi_i^{(j)} y_{t-i} + \varepsilon_t^{(j)}) I_{(r_{j-1}, r_j)}(z_{t-d})]$$

$$\varepsilon_t^{(j)} = \sqrt{\eta_t^{(j)}} e_t^{(j)} = \sigma_t^{(j)} e_t^{(j)}, \quad e_t^{(j)} \overset{i.i.d}{\sim} D_{\nu^{(j)}}(0,1) \tag{3.4}$$

$$\eta_t^{(j)} = [(\alpha_0^{(j)} + \sum_{l=1}^{q_j} \alpha_l^{(j)} \varepsilon_{t-l}^{2\,(j)} + \sum_{m=1}^{p_j} \beta_m^{(j)} \eta_{t-m}^{(j)})]$$

where y_t is the observed time series; $j = 1, \ldots, k$ where k is the total number of regimes; and $I_{\mathcal{A}}(z)$ is the indicator function that takes the value of 1 when $z \in \mathcal{A}$ and 0 otherwise; $r = (r_1, \ldots, r_{k-1})$ with $r_1 < \ldots < r_{k-1}$ are the threshold values, $r_0 = -\infty$ and $r_k = \infty$; z_{t-d} is the threshold variable with delay parameter d; $D_{\nu^{(j)}}(0,1)$ is the density function of the standardized error for the j-th regime with additional distribution parameters ν which describes the skewness and shape of the density. To ensure the non-negativity of the conditional variance, it is sufficient to impose the following conditions on the parameters:

$$\alpha_0^{(j)} > 0; \quad \alpha_l^{(j)}, \beta_m^{(j)} \geq 0; \quad \Big[\sum_{l=1}^{q_j} \alpha_l^{(j)} + \sum_{m=1}^{p_j} \beta_m^{(j)}\Big] < 1. \tag{3.5}$$

Estimation of the DTGARCH model parameters.

Simultaneous estimation of both the DTGARCH model parameters and the threshold values is not feasible. The likelihood of the model cannot be tractable with respect to the thresholds because of the discontinuity of the underlying functional relationship between the variables at the thresholds. However, given the form of the conditional mean and variance, one can use the quasi-maximum likelihood estimation (QMLE) approach to estimate the parameters of the DTGARCH model under predetermined threshold values. The log-likelihood of the model maximized with respect to different regimes that determined by fixed threshold values. Thus, log-likelihood of the DTGARCH model in general is

$$\mathcal{L}_T(\Theta) = \sum_{j=1}^{k} \Big[log \prod_{t \in R_j} D_{\nu^{(j)}}(y_t, E(y_t|\Psi_{t-1})^{(j)}, \eta_t^{(j)}) \Big] I_{(r_{j-1},r_j)}(z_{t-d}), \qquad (3.6)$$

where $\Theta = [\theta^{(1)'}, \ldots, \theta^{(k)'}]$; $R_j = (r_{j-1}, r_j]$; $\theta^{(j)} = [\phi_0^{(j)}, \phi_i^{(j)}, \alpha_0^{(j)}, \alpha_l^{(j)}, \beta_m^{(j)}]$ and since the equation (3.5) splits the likelihoods into different regimes according to threshold values, QML estimator of regime specific $\theta^{(j)}$ can be shown as

$$\widehat{\theta^{(j)}} = arg \max_{\theta^{(j)}} \Big[log \prod_{t \in R_j} D_{\nu^{(j)}}(y_t^{(j)}, E(y_t|\Psi_{t-1})^{(j)}, \eta_t^{(j)}) \Big],$$

where $j = 1, \ldots, k$ and k is the number of regimes; Ψ_{t-1} is set of available information set at time $t-1$ and $t = 1, \ldots, T$.

Results on the asymptotic distribution of the parameters.

Under the assumption that y_t is stationary and ergodic and the threshold parameters are known, Li and Li [117] proved that the MLE of the DTARCH model is consistent and asymptotically normal.

When the regimes are prespecified (i.e., threshold parameters are known), the DT-GARCH model can be considered as a combination of distinct GARCH models and the estimation procedure can be facilitated via using the asymptotic results for the QMLE in GARCH models by [119]. Here, strict stationarity of ε_t and e_t are assumed for consistency of the QMLE. It does not require the e_t to be $i.i.d.$ but the assumption $E\big[|e_t|^{4\xi}\big] < \infty$, for some $\xi > 0$, is required for asymptotic normality.

The issue of asymmetry in GARCH models.

One of the major disadvantage of the standard ARCH/GARCH (sGARCH) model is that it does not allow the asymmetric response of volatility to the past values of innovations. Thus, under the sGARCH model, the conditional variance depends only on the magnitude of past of innovations. That is, the conditional variance treats past positive and negative values equally (see [120]). This oversimplifying constraint is constitute the limitation of the standart GARCH structure and should be adapted for allowing asymmetric response of volatility to the positive and negative past values of innovations.

To overcome this drawback, we consider the exponential GARCH (EGARCH), the threshold-GARCH (TGARCH) and the asymmetric power GARCH (APARCH) models.

The EGARCH process, which uses the natural logarithm of the conditional variance, within the DTGARCH context can be shown as

$$\log \eta_t^{(j)} = \alpha_0^{(j)} + \sum_{l=1}^{q_j} (\alpha_l^{(j)} e_{t-l}^{(j)} + \gamma_l^{(j)} (|e_{t-l}^{(j)}| + E|e_{t-l}^{(j)}|)) + \sum_{m=1}^{p_j} \beta_m^{(j)} \log \eta_{t-m}^{(j)}, \quad (3.7)$$

where the coefficient α_l captures the sign effect and γ_l is the size effect (see [121]).

The APARCH model of Ding et al. [122] can be shown as

$$\eta_t^{\delta_{(j)}/2} = \alpha_0^{(j)} + \sum_{l=1}^{q_j} \alpha_l^{(j)} (|\varepsilon_{t-l}^{(j)}| + \gamma_l^{(j)} \varepsilon_{t-l}^{(j)})^{\delta_{(j)}} + \sum_{m=1}^{p_j} \beta_m^{(j)} \eta_{t-m}^{\delta_{(j)}/2}, \quad (3.8)$$

where $\delta \in \mathbb{R}^+$ being a Box-Cox transformation of $\sqrt{\eta_t}$ and $-1 < \gamma < 1$ is the leverage effect. The APARCH model contains several ARCH/GARCH models as special cases and the introduction of the power δ increases the flexibility of GARCH-type models, and allows the a priori selection of an arbitrary power to be avoided [115].

TGARCH model of Zakoian [123] is a special case of the APARCH model and it arise from the process in Equation 3.8 when $\delta = 1$.

Therefore, in addition to the sGARCH model within DTGARCH context, we consider EGARCH, TGARCH and APARCH models throughout the study.

Almost all of the applications in the field of the DTGARCH models are dedicated to analyze the differences in dynamics in economic and financial time series data. For example, Brooks [118] analyzed the exchange rate via DTGARCH model and stated the improvement of the out of sample forecast and highlighted the importance of considering both types of regime shift (i.e. thresholds in variance as well as in mean) when analyzing the financial time series. Audrino et al. [124] proposed a general double tree structured AR-GARCH model for the analysis of global equity index returns and found strong evidence for more than one multivariate threshold in conditional means and variances of global equity index returns. Suardi [125] evaluated the intervention of Bank of Japan and the Federal Reserve that are more effective in changing the direction of the exchange rate movements via DTGARCH model.

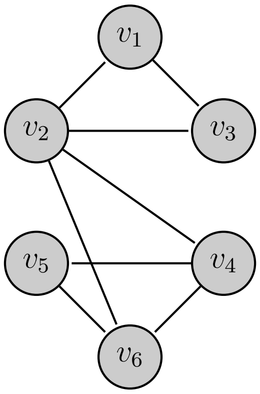
To the best of our knowledge, our work illustrates the first application of DTGARCH model to time series clustering. In time series clustering approach that we propose, in line with the potential of the DTGARCH specification to distinguish multiple univariate and multivariate time series, the DTGARCH specification is used for observing nonlinear associations and the AR specification is used for observing linear associations.

3.3 Spectral Clustering

Spectral clustering, in the most general sense, is a distance-based clustering algorithm with no assumptions on the shape of clusters. Specifically, the clustering algorithm operates on a spectrum (i.e., eigenpairs) of a similarity/affinity matrix derived from the distance matrix. Spectral clustering works well in conditions where connectedness (i.e., interconnectivity) of points is more important than distances between points (i.e., objects to be clustered), such as for intertwined spirals, nested structures or non-convex shapes [99]. Theoretical background of the spectral clustering relies on the graph theory in mathematics, and since the affinity/similarity matrix used in the spectral clustering algorithm is a matrix representation of a given graph, the spectral clustering can be considered as a counterpart of graph partitioning [126].

A graph G is consists of set of vertices, V and set of edges (i.e, nodes), E. Let demonstrate the elements of a graph by an example of an arbitrary simple undirected and edge weighted graph illustrated in Figure 3.2. Edges between each pair of vertices can be denoted by $w_{ij} = w_{ji}; w_{ij} \geq 0;\ \ i, j = 1, \ldots, 6$, which shows the amplitude of adjacencies/affinities between each pair of vertices.

$G = \big\{V, E\big\}$,
$V = \big\{v_1, v_2, v_3, v_4, v_5, v_6\big\}$,
$E = \big\{\{v_1, v_2\}, \{v_1, v_3\}, \{v_2, v_3\}, \{v_2, v_6\}, \{v_4, v_5\}, \{v_4, v_6\}, \{v_5, v_6\}\big\}$
or showing edges via weights where $w_{ij} \geq 0; w_{ij} = w_{ji};\ \ i, j = 1, \ldots, 6$,
$E = \big\{w_{12}, w_{13}, w_{23}, w_{24}, w_{26}, w_{45}, w_{46}, w_{56}\big\}$.



(a)

$$A_G = \begin{array}{c|cccccc|} & v_1 & v_2 & v_3 & v_4 & v_5 & v_6 \\ v_1 & 0 & w_{12} & w_{13} & 0 & 0 & 0 \\ v_2 & w_{21} & 0 & w_{23} & w_{24} & 0 & w_{26} \\ v_3 & w_{31} & w_{32} & 0 & 0 & 0 & 0 \\ v_4 & 0 & w_{42} & 0 & 0 & w_{45} & w_{46} \\ v_5 & 0 & 0 & 0 & w_{54} & 0 & w_{56} \\ v_6 & 0 & w_{62} & 0 & w_{64} & w_{65} & 0 \end{array}$$

(b)

Figure 3.2 (a) Sample undirected and edge weighted graph G; (b) Corresponding adjacency matrix of G where $w_{ij} \geq 0; w_{ij} = w_{ji}; i, j = 1, \ldots, 6$.

If the edges are not weighted then each edge takes values of 1. There are many fields in which graph theory can be applied and the solution of many problems can be explored by presenting the problem through the graph-theoretic problem [127, 128].

In graph theory, matrix representation of a graph can be given by $Laplacian$ matrix and the spectrum of $graph\ Laplacian$, L_G, is useful to explore various properties of a given graph especially when the goal is described as graph partitioning of a graph $G = \big\{V, E\big\}$ into several partitions where the number of cut edges (i.e., zero edges) should be minimized and the vertices in the same partition are more close and similar to each other compared to other vertices in other partitions. Spectrum of $graph\ Laplacian$, L_G, contains information about connectivity of the graph.

From a given graph $G = \big\{V, E\big\}$ in Figure 3.2 one can construct its unnormalized $graph\ Laplacian$, L_G, as:

$$L_G = D_G - A_G,$$

where A_G is adjacency (i.e., affinity) matrix of graph G, D_G is the degree matrix of graph G and defined as the diagonal matrix where the degrees $d_{11}, \ldots, d_{66}$ on the diagonal. The degree d_{ii} of vertex v_i is defined as:

$$d_{ii} = \textstyle\sum_{j=1}^{6} w_{ij}.$$

Therefore, for the sample graph $G = \{V, E\}$ given in Figure 3.2, the matrix L_G can be shown as:

$$L_G = \begin{bmatrix} d_{11} & -w_{12} & -w_{13} & 0 & 0 & 0 \\ -w_{21} & d_{22} & -w_{23} & -w_{24} & 0 & -w_{26} \\ -w_{31} & -w_{32} & d_{33} & 0 & 0 & 0 \\ 0 & -w_{42} & 0 & d_{44} & -w_{45} & -w_{46} \\ 0 & 0 & 0 & -w_{54} & d_{55} & -w_{56} \\ 0 & -w_{62} & 0 & -w_{64} & -w_{65} & d_{66} \end{bmatrix},$$

where

$$d_{11} = w_{12} + w_{13}$$

$$d_{22} = w_{21} + w_{23} + w_{24} + w_{26}$$

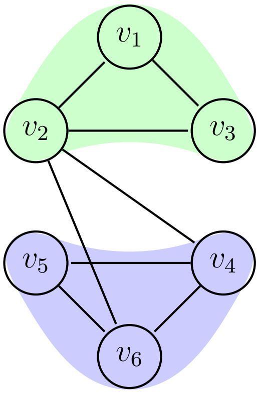
$$d_{33} = w_{31} + w_{32}$$

$$d_{44} = w_{42} + w_{23} + w_{45} + w_{46}$$

$$d_{55} = w_{54} + w_{56}$$

$$d_{66} = w_{62} + w_{64} + w_{65}.$$

Several ways of constructing the matrix L_G such as unnormalized, normalized and random walk normalized $graph\ Laplacian$ are available in the literature (see, e.g., [126, 129, 130]). Spectrum of $graph\ Laplacian$ can be utilized to investigate properties of a graph. Let the graph shown in Figure 3.2 is now to be subjected to partitioning. A very simple example of a graph partitioning is illustrated in Figure 3.3. If the edges between vertices 2 and 4, and between 2 and 6 are negligibly close to zero or relatively small than other weighted edges in magnitude then zero edges minimized bipartition of the graph can be achieved by investigating the spectrum of L_G. The block diagonal form of the L_G will clearly manifest the algebraic connectivity of vertices in the eigenvector through signs of vector elements corresponding to the second smallest eigenvalue (i.e., Fiedler eigenvalue thanks to Fiedler [131]). For more details about graph partitioning and graph theory please see [132–134].



(a)

$$L_G = \begin{bmatrix} d_{11} & -w_{12} & -w_{13} & 0 & 0 & 0 \\ -w_{21} & d_{22} & -w_{23} & -w_{24} & 0 & -w_{26} \\ -w_{31} & -w_{32} & d_{33} & 0 & 0 & 0 \\ 0 & -w_{42} & 0 & d_{44} & -w_{45} & -w_{46} \\ 0 & 0 & 0 & -w_{54} & d_{55} & -w_{56} \\ 0 & -w_{62} & 0 & -w_{64} & -w_{65} & d_{66} \end{bmatrix}$$

(b) $graph\ Laplacian$ of G

Figure 3.3 (a) Sample undirected graph G subjected to partitioning; (b) Corresponding $graph\ Laplacian$ of G.

In this context, any time series clustering task can be considered as a graph partitioning problem if the vertices to be partitioned are regarded as time series objects. The case here now becomes how to find the informative edges (i.e., association, similarity) between time series to help with clustering. Typically, edges can be understood as a manifestation of similarities between each pair of time series. Thus, plausible construction of adjacency (i.e., affinity, similarity) matrix will present a graph that can be partitioned by considering the properties of spectrum of $graph\ Laplacian$. However, as previously mentioned in Section 3.1, time series are not feasible to use them in their raw forms for clustering. Most of the important time-dependent properties of the time series are not directly noticeable or detectable from their raw forms. Instead, plausible and informative feature sets extracted from time series can be employed as objects to be clustered.

Spectral clustering (SC) mainly relies on graph theory. As stated earlier in this section, SC accomplishes well in nested structures where connectedness of points is more important than distances between points [99, 126]. Set of time series to be clustered through model-based features can be regarded as a graph and each vertice can be represented by the feature set instead of raw time series.

Therefore, this study invokes the spectral clustering algorithm proposed by Ng et al. [99] and adapts for time series clustering task. Implementation of SC in our proposed approach, as shown in Section 4 and 5, operates well over the model-based extracted feature sets from time series. Slightly modified spectral clustering algorithm and implementation of the proposed clustering approach is detailed in following Section 3.4.

The SC algorithm proposed by Ng et al. [99] can be summarized in 6 steps as follows:

Given a set of v-dimensional points $F = \{f_1, f_2, f_3, \ldots, f_n\}$ in $\mathbb{R}^v$ can be clustered into cl subsets:

1. Constructing the $n \times n$ Affinity (i.e., Adjacency) matrix A which is defined by
$$A_{ij} = \frac{exp(- \parallel f_i - f_j \parallel^2)}{(2\varrho^2)}$$
if $i \neq j$, and $A_{ii} = 0$.

2. Define matrix D to be diagonal matrix whose (i, i) element is the sum of A' s i^{th} row, and compute the normalized $graph\ Laplacian$ matrix by $L = D^{-1/2}AD^{-1/2}$.

3. Construct the eigenvector matrix $X_{n\times cl}$ from cl largest eigenvectors of L that are stacked in columns of X.

4. Compute the matrix U from re-normalized rows of the matrix X.

5. Cluster each row of U as a point in $\mathbb{R}^{cl}$ into cl clusters by any clustering algorithm (e.g., We consider the fuzzy C-means clustering throughout the study).

6. Assign the original point f_i to cluster number j iff row i of the re-normalized eigenvector matrix U is assigned to cluster number j.

Here, value of distance scaling parameter, ϱ, can be chosen over candidate values where the tightest clusters observed. For more information about spectral clustering please see [99] and [126].

3.4 Proposed Approach for Time Series Clustering

In this study, we propose a novel time series clustering approach for multiple univariate and multivariate time series where time series data, taking into account all the time span they have, is considered as objects to be clustered. Clustering perspective of the proposed approach is clustering set of time series in terms of their underlying DGMs. Although underlying (i.e., true) DGMs cannot be known for sure but only approximated, the revealed clusters are expected to discover time series that are similar in terms of DGMs. In this regard, it can be said that the aim is to assign a property set to the centers of the clusters to be formed, which is the composition of approximations of underlying DGMs. It is also important to note that the mechanism that produces the time series data may vary depending on time. Accordingly, it is equivalently important to distribute time series into clusters that are as homogeneous as possible in terms of the similarity of time-varying generating mechanisms.

As highlighted in previous Sections 3.1 and 3.3, within the scope of the clustering perspective, time series are not expedient to utilize them directly for clustering. The question of how we can obtain distinguishing and characteristic information about the mechanism that produces time series has been put forward as an important question in Section 3.1. This research question together with the infeasibility of using raw time series in clustering leads us to represent each time series with a specific feature vector that can be employed as objects to be clustered instead of time series. To obtain a idiosyncratic feature vector with a large enough capacity to represent a time series, we primarily utilize the linear autoregressive and the multi-faceted nonlinear time series model based estimation outputs.

Feature vectors are aimed to summarize associations of time series with the rich environment represented by aforementioned (i.e., AR and DTGARCH) models. This is because, although we cannot know the exact mechanisms in which time series are produced, there will be a certain similarity in the resulting estimation outputs when time series with similar generating mechanisms are examined in terms of a specific model. In terms of the clustering approach we propose, it has appeared an important principle that the specific time series model to be chosen for time series clustering should contain and generate many time series properties as in the DTGARCH model.

Thus, even though we will not know the exact mechanisms in which the time series are generated, we can group the time series in terms of underlying DGMs by comparing the time series' connections with a specific time series model that is complex enough and rich in environment.

To enhance idiosyncraticity of feature vectors, we append some frequency-domain based properties of time series from the frequency spectrum of raw time series, squared residuals and conditional variance estimates. In this respect, feature vectors combine both time series model-based and model-free properties. Features to be used in clustering, and configuration of feature vectors and matrices using extracted feature blocks are given in the following subsection 3.4.1.

Feature vectors formed for clustering can be considered as vertices of a graph, while the distances between them can be regarded as edges of a graph. Correspondingly, the clustering task can be seen as a graph partitioning problem. We utilize spectral clustering algorithm by Ng et al. [99] with implementing slightly modified version of Affinity (i.e., Adjacency) matrix calculation adapted to time series clustering. Details of the proposed clustering approach for univariate and multivariate set of time series given in subsections 3.4.2 and 3.4.3.

3.4.1 Feature Blocks to be Used in Clustering

We mainly consider linear AR and DTGARCH models based estimation outputs as feature blocks for the formation of feature vectors/matrices. These outputs are consist of estimates of coefficients; autocorrelation and partial autocorrelation functions of residuals and squared residuals; estimates of conditional variances. In addition to these model-based outputs over time domain, we consider some model-free properties from frequency-domain of time series. Thus, we append the complex modulus of Fourier coefficients (i.e., frequency spectrum) from the Fourier transform of raw time series, squared residuals and conditional variance estimates. Here we note the time series models, serial correlation functions and the notion of frequency spectrum in brief from which the components used for the establishment of the feature vectors/matrices are obtained. Extracted features from these tools considered as blocks of features which constitutes feature vectors/matrices.

Autoregressive Model: AR model of order p_{ar}, $AR(p_{ar})$, can be expressed as

$$y_t = \phi_0^{(ar)} + \phi_1^{(ar)} y_{t-1} + \phi_2^{(ar)} y_{t-2} + \ldots \phi_p^{(ar)} y_{t-p_{ar}} + a_t,$$

where a_t is a white noise sequence with variance σ_a^2.

To provide comparable outputs for time series to be clustered, the AR order, p_{ar}, is set to 12. Therefore, a part of linear associations that can be measured by coefficient estimates, $\widehat{\phi}_{1:12}^{(ar)}$, and residuals, $\widehat{a}_t$, and remaining conditional heteroscedasticity preserved by squared residuals, $\widehat{a_t^2}$, are accepted as linear time series model-based components for construction of feature vectors.

Double Threshold GARCH Model: General representation of DTGARCH model with order parameters, $k, ar_1, q_1, p_1, \ldots, ar_k, q_k, p_k$, is shown in equation 3.4. However, since we develop a clustering algorithm to be bound to a specific model within the scope of this study, we have already fixed the order parameters (i.e., considered as hyperparameters) and used the model given below because of its computational convenience and applicability. At the same time, the choice of this model allows us to compare time series with each other. The number of the regimes, k, is set to 3, the order of autoregressive components for each regime is set to 4, and standard GARCH(1,1) is considered for each regime's conditional variance.

DTGARCH ($k = 3, ar_1 = 4, q_1 = 1, p_1 = 1, ar_2 = 4, q_2 = 1, p_2 = 1, ar_3 = 4, q_3 = 1, p_3 = 1$) model can be shown as

$$y_t = \begin{cases} \phi_1^{(1)} y_{t-1} + \phi_2^{(1)} y_{t-2} + \phi_3^{(1)} y_{t-3} + \phi_4^{(1)} y_{t-4} + \varepsilon_t^{(1)}, & y_{t-1} \leq r_1 \\ \phi_1^{(2)} y_{t-1} + \phi_2^{(2)} y_{t-2} + \phi_3^{(2)} y_{t-3} + \phi_4^{(2)} y_{t-4} + \varepsilon_t^{(2)}, & r_1 < y_{t-1} \leq r_2 \\ \phi_1^{(3)} y_{t-1} + \phi_2^{(3)} y_{t-2} + \phi_3^{(3)} y_{t-3} + \phi_4^{(3)} y_{t-4} + \varepsilon_t^{(3)}, & y_{t-1} \geq r_2 \end{cases}$$

$$\varepsilon_t^{(j)} = \sqrt{\eta_t^{(j)}} e_t^{(j)} = \sigma_t^{(j)} e_t^{(j)}, \;\; e_t^{(j)} \overset{i.i.d}{\sim} D_{\nu^{(j)}}(0,1), j = 1, 2, 3$$

$$\eta_t = \begin{cases} \alpha_0^{(1)} + \alpha_1^{(1)} \varepsilon_{t-1}^2{}^{(1)} + \beta_1^{(1)} \eta_{t-1}{}^{(1)} \\ \alpha_0^{(2)} + \alpha_1^{(2)} \varepsilon_{t-1}^2{}^{(2)} + \beta_1^{(2)} \eta_{t-1}{}^{(2)} \\ \alpha_0^{(3)} + \alpha_1^{(3)} \varepsilon_{t-1}^2{}^{(3)} + \beta_1^{(3)} \eta_{t-1}{}^{(3)}. \end{cases}$$

Various combinations of the DTGARCH model specification given here covers and exhibits many time series features such as stationarity, non-stationarity, chaotic behavior, linearity, non-linearity, seasonality, abrupt changes, and regime switching. The success of the specified model in the experimental study supports our clustering perspective. This means that the selected model's environment is rich enough to cover numerous type of behavior in time series. In this sense, the selection of a specific DTGARCH model is not an unique choice. It should be noted that models with a certain degree of diversity in the background will have similar performance. However, finding the optimal degree of background diversity for the selection of the model is beyond the scope of this study and left as future work.

Consequently, a part of non-linear associations that can be measured by coefficient estimates, $\widehat{\phi}_{1:4}^{(1:3)}$, $\widehat{\alpha}_0^{(1:3)}$, $\widehat{\alpha}_1^{(1:3)}$, $\widehat{\beta}_1^{(1:3)}$, conditional variance estimates, $\widehat{\eta}_t^{(1:3)}$, residuals, $\widehat{\varepsilon}_t^{(1:3)}$, and remaining conditional heteroscedasticity preserved by squared residuals, $\widehat{\varepsilon_t^2}^{(1:3)}$, are acknowledged as non-linear time series model-based components for construction of feature vectors.

Sample Autocorrelation and Sample Partial Autocorrelation Functions: Autocorrelation function (ACF) and Partial Autocorrelation function (PACF) carry information about linear temporal dependence of stationary time series. Therefore, a part of linear associations that can be measured by sample ACF and sample PACF of raw time series and residuals are considered as a component of feature vectors.

For a given observed time series $z_t, \;\; t = 1 : T$, the sample auto-covariance function $\widehat{\gamma}(h)$, is defined as follows:

$$\widehat{cov}(z_t, z_{t+h}) = \widehat{\gamma_z}(h) = \frac{1}{T} \textstyle\sum_{t=1}^{T-h} (z_t - \bar{z})(z_{t+h} - \bar{z}).$$

Sample ACF estimates the linear correlation among the lagged values of a time series. For a given observed time series $z_t, \;\; t = 1 : T$, the sample ACF can be defined as follows:

$$\widehat{cor}(z_t, z_{t+h}) = \widehat{\rho_z}(h) = \frac{\widehat{\gamma_z}(h)}{\widehat{\gamma_z}(0)} = \frac{\sum_{t=1}^{T-h}(z_t - \bar{z})(z_{t+h} - \bar{z})}{\sum_{t=1}^{T}(z_t - \bar{z})^2}, \qquad h = 0, 1, 2, \ldots.$$

In addition to the estimation of sample ACF between z_t and z_{t+h}, there is another tool for measuring the linear association between z_t and z_{t+h} which is based on removing their mutual linear dependency on the lie in between variables $z_{t+1}, z_{t+2}, \ldots, z_{t+h-1}$ [135].

The conditional correlation estimation

$$\widehat{cor}(z_t, z_{t+h}|z_{t+1}, \ldots, z_{t+h-1}) = \widehat{\pi_z}(h, h)$$

is called as the sample PACF in time series analysis.

For a given observed time series $z_t, \quad t = 1 : T$, Durbin's [136] recursive formula starting with $\widehat{\pi_z}(1, 1) = \widehat{\rho_z}(1)$ for computing the sample PACF, $\widehat{\pi_z}(h, h)$, can be defined as follows:

$$\widehat{\pi_z}(h+1, j) = \widehat{\pi_z}(h, j) - \widehat{\pi_z}(h+1, h+1)\widehat{\pi_z}(h, h+1-j) \qquad j = 1, 2, \ldots, h$$

$$\widehat{cor}(z_t, z_{t+h+1}|z_{t+1}, \ldots, z_{t+h}) = \widehat{\pi_z}(h+1, h+1) = \frac{\widehat{\rho_z}(h+1) - \sum_{j=1}^{h}\widehat{\pi_z}(h, j)\widehat{\rho_z}(h+1-j)}{1 - \sum_{j=1}^{h}\widehat{\pi_z}(h, j)\widehat{\rho_z}(j)}.$$

ACFs and PACFs of raw time series (series stationarized where it is necessary) y_t, i.e., $\widehat{\rho_y}(h)$ and $\widehat{\pi_y}(h, h)$; ACFs and PACFs of residuals and squared residuals from linear model estimations, i.e., $\widehat{\rho_a}(h)$, $\widehat{\rho_{a^2}}(h)$, $\widehat{\pi_a}(h, h)$ and $\widehat{\pi_{a^2}}(h, h)$; ACFs and PACFs of residuals and squared residuals from non-linear model estimations, i.e., $\widehat{\rho_\varepsilon}(h)$, $\widehat{\rho_{\varepsilon^2}}(h)$, $\widehat{\pi_\varepsilon}(h, h)$ and $\widehat{\pi_{\varepsilon^2}}(h, h)$ are features which constitutes blocks of feature vectors/matrices. To provide comparable resolution among serial correlations number of delay lags, h, for up to 18 is considered within the settings of experimental study.

Frequency Spectrum: Fourier representation of a given time series $z_t, \quad t = 1 : T$, and for $\kappa = 0, \ldots, T/2$ can be given as

$$z_t = \sum_{\kappa=0}^{T/2} [A_\kappa \cos(\tfrac{2\pi\kappa t}{T}) + B_\kappa \sin(\tfrac{2\pi\kappa t}{T})]$$

$$= \begin{cases} \sum_{\kappa=-(T-1)/2}^{(T-1)/2} C_\kappa e^{(\frac{2\pi i\kappa t}{T})}, & if\ T\ is\ odd, \\ \sum_{\kappa=-(T/2)+1}^{T/2} C_\kappa e^{(\frac{2\pi i\kappa t}{T})}, & if\ T\ is\ even. \end{cases}$$

The A_κ, B_κ and C_κ are called the Fourier coefficients or the complex valued spectrum of z at the corresponding Fourier frequency $\omega_\kappa = \frac{2\pi\kappa}{T}$.

Considering the following Euler relation

$$e^{i\omega} = \cos\ \omega + i \times \sin\ \omega$$

and the identities

$$\sin\ \omega = \frac{e^{i\omega} - e^{-i\omega}}{2i},$$

$$\cos\ \omega = \frac{e^{i\omega} + e^{-i\omega}}{2i},$$

the Fourier coefficients A_κ, B_κ and B_κ are related as

$$\begin{cases} C_0 = A_0;\ C_{T/2} = A_{T/2},\ if\ T\ is\ even \\ C_{+\kappa} = \frac{A_\kappa - iB_\kappa}{2}, \\ C_{-\kappa} = \frac{A_\kappa + iB_\kappa}{2}. \end{cases}$$

Above mentioned statements implies that any given observed time series can be represented by a linear combinations of sine-cosine waves or the complex exponentials [135]. Therefore, for a given time series $z_t, \quad t = 1 : T$ and assuming that the T is even, the Fourier coefficients can be computed via the Fourier transformation as follows:

$$C_\kappa(z) = \sum_{t=1}^{T} z_t e^{-i2\pi\kappa t/T}, \qquad \kappa = -T/2 + 1, \ldots, 0, \ldots, T/2.$$

Consequently, the frequency spectrum of the time series, z_t, can be produced by the computation of the complex modulus, $|C_\kappa(z)|$, versus the Fourier frequencies $\omega_\kappa =$

$\frac{2\pi\kappa}{T}$ for $\kappa = 0, \ldots, T/2$. For the large number of T, the computation of the Fourier transforms requires great numbers of calculations and become impractical. However, using the faster computation algorithm known as the fast Fourier transform (FFT) one can quickly and efficiently accomplish the Fourier transforms. Please see [135, 137, 138] for more details about the FFT.

The frequency-domain properties are included in feature vectors via the computation of complex modulus of the Fourier coefficients (i.e., frequency spectrum) of raw time series, y_t, squared residuals and conditional variance estimates . Thus, $|C(y_t)|$, $|C(\widehat{\varepsilon_t^2})|$ and $|C(\widehat{\eta}_t^{(1:3)})|$ are corresponding components of feature vectors/matrices.

3.4.2 Multiple Univariate Time Series Clustering

The proposed approach for multiple univariate time series clustering is designated to accomplish the purpose of clustering a set of univariate time series $\boldsymbol{y} = \{y_t^1, \ldots, y_t^n\}$ each with size of T into cl subsets at time point T regarding all the time span they have.

The main clustering perspective of the proposed clustering approach is to cluster time series with respect to similarities of their underlying DGMs. In this context, it is aimed to appoint latent DGMs to cluster centers. To this end, each time series represented by a feature vector in which components are obtained from the specific time series models estimations and non-model-based properties. By doing this we treated time series as objects to be clustered. Here, time series models can be considered as multi faceted prisms. After filtering of time series with practical and comprehensive tools, the resultants provide reduced and refined information to enable us to evaluate the similarities of time series in terms of their underlying DGMs.

Since the components of the feature vectors are provided from different feature spaces, it is more convenient to express and treat these vectors in block layered form. The tools for feature extraction are briefly mentioned in the previous subsection 3.4.1. Consequently, components of the feature vector for a given time series y_t shown in Table 3.1.

Table 3.1 Components of the feature vector for a given time series y_t.

(a) model-based components

Feature	Source
$\widehat{\phi}^{(ar)}_{1:p_{ar}}$	AR
$\widehat{\phi}^{(1:k)}_{1:ar_j}$	DTGARCH
$\widehat{\alpha}^{(1:k)}_0$, $\widehat{\alpha}^{(1:k)}_1$, $\widehat{\beta}^{(1:k)}_1$	DTGARCH
$\widehat{\rho}_a(h)$, $\widehat{\pi}_a(h,h)$	AR
$\widehat{\rho_{a^2}}(h)$, $\widehat{\pi_{a^2}}(h,h)$	AR
$\widehat{\rho}_\varepsilon(h)$, $\widehat{\pi}_\varepsilon(h,h)$	DTGARCH
$\widehat{\rho_{\varepsilon^2}}(h)$, $\widehat{\pi_{\varepsilon^2}}(h,h)$	DTGARCH
$\|C_\kappa(\widehat{\varepsilon^2})\|$	DTGARCH
$\|C_\kappa(\widehat{\eta}^{(1:k)})\|$	DTGARCH

(b) model-free components

Feature	Source
$\widehat{\rho}_y(h)$, $\widehat{\pi}_y(h,h)$	Raw time series
$\|C_\kappa(y)\|$	Raw time series

p_{ar} denotes the order of the AR model.

k denotes the number of regimes in DTGARCH model.

ar_j denotes the regime specific AR order where $j = 1, \ldots, k$.

h denotes the lag where the serial correlations are calculated up to.

κ denotes the spectrum length where $\kappa = 0, \ldots, T/2$ and T is the length of a given time series y.

The number of feature block considered as 11 in the feature vector which contains the linear AR and the non-linear DTGARCH models based features, and frequency domain properties of raw series and DTGARCH estimation results.

Forming a feature vector, f_s, from a univariate time series, y^s_t, can be illustrated as follows:

$$
y_t^s = \begin{bmatrix} y_1^s \\ y_2^s \\ \vdots \\ \vdots \\ \vdots \\ y_T^s \end{bmatrix}_{T\times 1} =====\Rightarrow \quad f_s^{\in 1:T} = \left[\begin{array}{c}
\widehat{\phi}_{1:p_{ar}}^{(ar)} \\ \hdashline
\widehat{\phi}_{1:ar_j}^{(1:k)} \\ \hdashline
\widehat{\alpha}_{0,1}^{(1:k)}, \widehat{\beta}_1^{(1:k)} \\ \hdashline
\widehat{\rho}_y(h)_{h\times 1} \\
\widehat{\pi}_y(h,h)_{h\times 1} \\ \hdashline
\widehat{\rho}_a(h)_{h\times 1} \\
\widehat{\pi}_a(h,h)_{h\times 1} \\ \hdashline
\widehat{\rho_{a^2}}(h)_{h\times 1} \\
\widehat{\pi_{a^2}}(h,h)_{h\times 1} \\ \hdashline
\widehat{\rho}_\varepsilon(h)_{h\times 1} \\
\widehat{\pi}_\varepsilon(h,h)_{h\times 1} \\ \hdashline
\widehat{\rho_{\varepsilon^2}}(h)_{h\times 1} \\
\widehat{\rho_{\pi^2}}(h,h)_{h\times 1} \\ \hdashline
|C_\kappa(\widehat{\varepsilon^2})|_{T/2\times 1} \\ \hdashline
|C_\kappa(\widehat{\eta})|_{T/2\times 1} \\ \hdashline
|C_\kappa(y)|_{(T/2)\times 1}
\end{array}\right]_M
\begin{array}{l}
\rightarrow 1^{st} \text{ feature block} \\
\rightarrow 2^{nd} \text{ feature block} \\
\rightarrow 3^{rd} \text{ feature block} \\
\rightarrow 4^{th} \text{ feature block} \\
\rightarrow 5^{th} \text{ feature block} \\
\rightarrow 6^{th} \text{ feature block} \\
\rightarrow 7^{th} \text{ feature block} \\
\rightarrow 8^{th} \text{ feature block} \\
\rightarrow 9^{th} \text{ feature block} \\
\rightarrow 10^{th} \text{ feature block} \\
\rightarrow 11^{th} \text{ feature block}
\end{array}
$$

where $M = (p_{ar} + k * ar_j + k * 3 + 10h + 3T/2) \times 1$ is the dimension of the vector, f_s.

Therefore, a set of univariate time series $\boldsymbol{y} = \{y_t^1, \ldots, y_t^s, \ldots, y_t^n\}$ each with length of T can be represented via feature matrix, $|\boldsymbol{F}(\boldsymbol{y})|_{M\times n}$, which constitutes of feature vectors as $|\boldsymbol{F}(\boldsymbol{y})|_{M\times n} = \{f_1, \ldots, f_s, \ldots, f_n\}$. Thereafter, clustering task can be performed over feature vectors instead of raw time series.

Accordingly, by considering feature vectors, distance-based clustering approaches become a more practical solution for clustering. However, since the intra-connectedness and the inter-connectedness of each feature block is a fact, the task of clustering of these feature vectors leads us to efficiently utilize the spectral clustering algorithm. As we noted previously, SC outperforms conventional clustering algorithms whenever the connectivity of points to be clustered is non-ignorable [99].

To adapt the SC algorithm given in Section 3.3 for time series clustering via block layered feature vectors, the affinity matrix, A, over given set of feature vectors $\{f_1, \dots, f_n\}$ in $\mathbb{R}^M$ where $M = (p_{ar} + k * ar_j + k * 3 + 10h + 3T/2)$ is modified and defined by

$$A_{ij} = \frac{exp\Big(- \sum_{wn=1}^{T^*} \big(\sum_{bl=1}^{8} \parallel f_{i,bl}^{\in[wn,T]} - f_{j,bl}^{\in[wn,T]} \parallel^2 + \sum_{bl=9}^{11} D_{KS}(f_{i,bl}^{\in[wn,T]}, f_{j,bl}^{\in[wn,T]})\big)\Big)}{(2\varrho^2)}, \tag{3.9}$$

in the proposed approach. To provide legitimate similarities toward clustering, the distance calculations in the affinity matrix are made compatible with the block layered structure of the feature vector.

Additionally, to evaluate the time dependency of the time series (i.e., clusters) in clustering, the cumulative sum of distances is provided to the affinity matrix calculation, which is the cumulative sum of feature distances over all temporal resolutions, i.e., the procedure of calculating the matrix A utilize the iterative feature extraction and the accumulated sum of feature distances obtained through all available time windows such as, $D(f_i, f_j) = D(f_i, f_j)_{\in[1,T]} + D(f_i, f_j)_{\in[2,T]} + \dots + D(f_i, f_j)_{\in[wn,T]}$ where $wn = 1, \dots, T^*$, and $T^* \approx T - 350$ is the smallest window of a time series observations needed for DTGARCH estimation. Therefore, the persistent change in the underlying generating mechanism of time series will be evaluated in the cumulative distance calculation. Likewise, if there is no change, the effect of the cumulative sum of the distance to the clustering via the affinity matrix will be the same as the effect of distance that is based on the overall time span.

Various distance measures can be used for the first eight blocks such as $L^p - norms$, $Hausdorff$, $Minkowski$, $Manhattan$, $Dynamic Time Warping (DTW)$ and many others. Defining an optimal distance metric for the proposed approach is left as a future work since it is partly irrelevant to the scope of the study. However, we utilize $L^2 - norm$, which is known as Euclidean distance, because it performs better than those compared (e.g., Hausdorff distance metric, DTW) in Aslan et al. [50]. Euclidean distance, D_{EUCL} can be given as follows:

$$D_{EUCL}(f_i, f_j) = \parallel f_i - f_j \parallel^2 = \sqrt{\sum_{\tau=1}^{M} \Big(f_i(\tau) - f_j(\tau) \Big)^2}.$$

The last three blocks of the feature vector consists of the frequency spectrum of raw series, squared residuals and conditional variances. For this reason, we propose to use of Kolmogorov-Smirnov distance, D_{KS}, in the proposed approach. D_{KS} is a metric that is based on the non-parametric Kolmogorov-Smirnov test, which is the maximal distance between the cumulative distribution functions [138] (i.e., the spectra of the cumulative amplitudes of the Fourier transforms within the scope of this study). D_{KS} can be shown as follows:

$$D_{KS}(f_i, f_j) = \max_{1 \leq l \leq M} \left| \sum_{\tau=1}^{l} f_i(\tau) - \sum_{\tau=1}^{l} f_j(\tau) \right|, \qquad D_{KS} \in [0, 1].$$

The proposed univariate time series clustering approach, which we name *The Linear and the Non-linear Time Series Model Based Spectral Clustering* (TSMB-SPCL-UV), intends to provide a feasible and useful grouping of time series variables concerning their underlying DGMs by using idiosyncratic feature vectors.

Determining the appropriate number of clusters: The problem of finding the appropriate number of clusters is itself a research area and is closely related to definition of the problem-specific "true" cluster and clustering. Our aim is to group the series in terms of the data generating structures they belong to so that the resulting sets are as homogeneous as possible. Since the proposed approach mainly operates on distances and connections of feature vectors, we use the slightly modified version of the GAP statistics developed by Tibshirani [139] which evaluates the within-cluster distances (i.e., within-cluster dispersion) around the cluster means.

The sum of within-cluster pairwise distances between all points, n_{cl}, for a given cluster CL_{cl}, where cl refer to the cluster number, can be given as

$$D_{cl} = \sum_{i,i' \in CL_{cl}} d_{ii'},$$

where d is euclidean(L^2-norm) distance. Therefore, within-cluster dispersion (i.e. cluster compactness), W_{cl}, can be set as

$$W_{cl} = \sum_{cl=1}^{cl_{max}} \frac{D_{cl}}{2n_{cl}}.$$

Then the GAP statistic, $GAP(cl)$, proposed by [139], is defined as

$$GAP_n(cl) = E_n^* \left\{ logW_{cl} \right\} - logW_{cl},$$

where n is the sample size and E_n^* denotes the expectation determined by bootstrapping, i.e, simulating from a reference distribution. Eventually, the estimate of cluster number, $\widehat{cl}$, is the value that maximizing $GAP_n(cl)$.

In order to get a smoother curve from the GAP statistics and to avoid local maxima, we propose a generic function which multiplies the GAP statistics by a monotonically increasing function given as follows:

$$f_g(GAP, cl) = GAP(cl) \times \left(\frac{cl - 1}{cl} \right)^2.$$

Thus, the estimate of the optimal number of clusters, $\widehat{cl}$, can be selected at the global maximum of the generic function, $f_g(GAP, cl)$, where $cl \geq 1$. Henceforth, we call this generic function as the modified GAP statistics.

Steps of the proposed time series clustering approach are described in Figure 3.4.

Step 1. Extract features based on non-linear associations:
Define the specific DTGARCH($k, ar_1, q_1, p_1, \ldots, ar_k, q_k, p_k$) model to extract non-linear based features, where k denotes the number of regimes.

- **Step 1.1.** Extract the statistically significant coefficient estimates,$\widehat{\phi}^{(1:k)}_{1:ar_j}$, from each estimation, where $j = 1, \ldots, k$.
- **Step 1.2.** Extract the residuals, $\widehat{\varepsilon}$, and the conditional variance estimates, $\widehat{\eta}$.
  1. Calculate $\widehat{\rho_{\varepsilon}}(h)$ and $\widehat{\pi_{\varepsilon}}(h,h)$ of the residuals up to lag h.
  2. Calculate $\widehat{\rho_{\varepsilon^2}}(h)$ and $\widehat{\pi_{\varepsilon^2}}(h,h)$ of the squared residuals up to lag h.
  3. Calculate the frequency spectrum of $\widehat{\varepsilon^2}$, $|C_{\kappa}(\widehat{\varepsilon^2})|$.
  4. Calculate the frequency spectrum of $\widehat{\eta}$, $|C_{\kappa}(\widehat{\eta})|$.

Step 2. Extract features based on linear associations:
Define the specific AR(p_{ar}) model to extract linear based features, where p_{ar} denotes the number of AR order.

- **Step 2.1.** Extract the statistically significant coefficient estimates,$\widehat{\phi}_{1:p_{ar}}$, from each estimation.
- **Step 2.2.** Extract the residuals, $\widehat{a}$.
  1. Calculate $\widehat{\rho_{a}}(h)$ and $\widehat{\pi_{a}}(h,h)$ of the residuals up to lag h.
  2. Calculate $\widehat{\rho_{a^2}}(h)$ and $\widehat{\pi_{a^2}}(h,h)$ of the squared residuals up to lag h.

Step 3. Extract features based on raw time series:
Calculate $\widehat{\rho_{y}}(h)$ and $\widehat{\pi_{y}}(h,h)$ of the raw series up to lag h, and the frequency spectrum of raw series, $|C_{\kappa}(y)|$.

Step 4. Form the feature vectors toward spectral clustering:
Form the M -dimensional feature vectors by stacking feature blocks.

Step 5. Temporal (i.e., time-dependent) Clustering:
Run the spectral clustering algorithm on the constructed feature matrix using proposed affinity matrix calculation given in Equation 3.9. Time dependency of the time series in clustering is provided by the distance in the Affinity matrix calculation, which is the cumulative sum of feature distances over all temporal resolutions via repeating the feature extraction phase all the way from **Step 1** to **Step 4**. Number of clusters determined at, or near, the global maximum of the modified GAP statistics.

Step 6. Obtain overall cluster for given data period.

Figure 3.4 Steps of the proposed time series clustering approach

3.4.3 Multiple Multivariate Time Series Clustering

The proposed approach for clustering of set of multiple multivariate time series clustering is almost similar to the univariate time series clustering approach that we proposed in previous subsection 3.4.2. However, in multivariate case the feature vectors are replaced by feature matrices, and thereby, proposed clustering approach can be performed over set of feature matrices instead of set of multivariate time series. Similar to the univariate case of the proposed clustering approach, each multivariate set of time series treated as objects to be clustered.

The aim in this setting is clustering a set of multivariate time series $\boldsymbol{Y} = (\boldsymbol{Y_t^1}, \boldsymbol{\ldots}, \boldsymbol{Y_t^n})$ each with dimension of $T \times m$ into cl subsets at time point T regarding all the time span they have.

The feature matrix, $\boldsymbol{f_s}$, for a multivariate time series, $\boldsymbol{Y_t^s}$, which contains m time series variables can be shown as follows:

$$\boldsymbol{Y_t^s} = \left[\begin{array}{c:c:c:c} y_{t1}^s & y_{t2}^s & \ldots & y_{tm}^s \end{array}\right]_{T\times m},$$

$$\begin{array}{cccc} \downarrow & \downarrow & \cdots & \downarrow \end{array}$$

$$\boldsymbol{f_s^{\in 1:T}} = \left[\begin{array}{c:c:c:c} f_1^s & f_2^s & \ldots & f_m^s \end{array}\right]_{M\times m},$$

$$\begin{array}{ccc} f_1^s & \cdots & f_m^s \\ \downarrow & & \downarrow \end{array}$$

$$\boldsymbol{f_s^{\in 1:T}} = \left[\begin{array}{ccc}
{}^{(1)}\widehat{\phi}^{(ar)}_{1:p_{ar}} & \cdots & {}^{(m)}\widehat{\phi}^{(ar)}_{1:p_{ar}} \\ \hdashline
{}^{(1)}\widehat{\phi}^{(1:k)}_{1:ar_j} & \cdots & {}^{(m)}\widehat{\phi}^{(1:k)}_{1:ar_j} \\ \hdashline
\vdots & \cdots & \vdots \\
\vdots & \cdots & \vdots \\ \hdashline
{}^{(1)}\widehat{\rho_{\varepsilon^2}}(h)_{h\times 1} & \cdots & {}^{(m)}\widehat{\rho_{\varepsilon^2}}(h)_{h\times 1} \\
{}^{(1)}\widehat{\rho_{\pi^2}}(h,h)_{h\times 1} & \cdots & {}^{(m)}\widehat{\rho_{\pi^2}}(h,h)_{h\times 1} \\ \hdashline
{}^{(1)}|C_\kappa(\widehat{\varepsilon^2})|_{(T/2)\times 1} & \cdots & {}^{(m)}|C_\kappa(\widehat{\varepsilon^2})|_{(T/2)\times 1} \\ \hdashline
{}^{(1)}|C_\kappa(\widehat{\eta})|_{(T/2)\times 1} & \cdots & {}^{(m)}|C_\kappa(\widehat{\eta})|_{(T/2)\times 1} \\ \hdashline
{}^{(1)}|C_\kappa(y)|_{(T/2)\times 1} & \cdots & {}^{(m)}|C_\kappa(y)|_{(T/2)\times 1} \\ \hdashline
\end{array}\right]
\begin{array}{l}
\leftarrow \text{block matrix of } 1st \text{ feature} \\
\leftarrow \text{block matrix of } 2nd \text{ feature} \\
\\
\\
\left.\begin{array}{l} \\ \\ \end{array}\right\} \leftarrow \text{block matrix of } \ 8th \text{ feature} \\
\leftarrow \text{block matrix of } \ 9th \text{ feature} \\
\leftarrow \text{block matrix of } 10th \text{ feature} \\
\leftarrow \text{block matrix of } 11th \text{ feature} \\
M \times m,
\end{array}$$

where $M = (p_{ar} + k * ar_j + k * 3 + 10h + 3T/2)$. Thus, feature extraction from a set of multivariate time series data follows a similar procedure as in subsection 3.4.2 and accomplished by examining each univariate time series variable separately.

Consequently, the dimensionality of the multivariate time series is taken into account for the calculation of the affinity matrix, A, over corresponding set of feature matrices $\{\boldsymbol{f_1}, \ldots, \boldsymbol{f_s}, \ldots, \boldsymbol{f_n}\}$ in $\mathbb{R}^{M \times m}$. Since it is necessary to use the matrix norm to calculate the differences between the blocks of feature matrices in the calculation of the affinity matrix, we have used the Frobenius norm for this purpose.

The Frobenius norm is a matrix norm and also known as the Euclidean norm of a matrix. It can be used to calculate the distance between any two matrices of the same size. For example, suppose we have two matrices of the same size, $\mathbf{X}_{k \times m}$ and $\mathbf{Y}_{k \times m}$. The distance between $\mathbf{X}_{k \times m}$ and $\mathbf{Y}_{k \times m}$ can be calculated by the Frobenius norm as:

$$\| \mathbf{X} - \mathbf{Y} \|_F = \| \mathbf{Y} - \mathbf{X} \|_F = \sqrt{tr[(\mathbf{X} - \mathbf{Y})(\mathbf{X} - \mathbf{Y})']}.$$

Therefore, the affinity matrix, A, in multivariate case is modified and defined by

$$A_{ij} = \frac{exp\Bigg(- \sum_{wn=1}^{T^*} \Big(\sum_{bl=1}^{8} \| \boldsymbol{f}_{i,bl}^{\in[wn,T]} - \boldsymbol{f}_{j,bl}^{\in[wn,T]} \|_F + \sum_{m^*=1}^{m} \sum_{bl=9}^{11} D_{KS}(f_{i,m^*,bl}^{\in[wn,T]}, f_{j,m^*,bl}^{\in[wn,T]})\Big)\Bigg)}{(2\varrho^2)}, \tag{3.10}$$

in the proposed multivariate time series clustering approach, which we abbreviate as TSMB-SPCL-MV. The clustering procedure summarized in Figure 3.4 is also the same for multivariate case except for the calculation of the affinity matrix, A.

CHAPTER 4

EXPERIMENTAL RESULTS: UNIVARIATE TIME SERIES CLUSTERING

This chapter is devoted to the assessment of the clustering performance of the proposed approach, which is given in Section 3.4, for a set of univariate time series data where each univariate time series is desired to be clustered as an object.

What we mean by clustering and ’true cluster’?

As mentioned in Sections 1.3, 3.1 and 3.4, the clustering perspective of this study and the aim of the proposed clustering approach is to distinguish and group a set of time series in terms of their similarities of the sources from which they are produced/emerged. In this sense, ’true’ clusters of time series refer to clusters that contain similar objects (i.e., time series) in terms of their underlying generating mechanisms. Accordingly, the clustering algorithm proposed in the study is structured by regarding the definition of the ’true’ cluster given here.

On the relations between the artificially created data generating procedures and authentic time series data.

Taking into account that the underlying (i.e., latent) DGMs of observed time series can not be fully identified but can only be approximated, the clustering approach we propose mainly relies on time-dependent features of the time series that are derived from raw data and approximations to their underlying data generating processes.

The underlying data generation mechanism of an observed time series variable may be time-dependent because of alternating states over time. For example, neuronal activities of a human brain that can be observed via the electroencephalography (EEG) would exhibit significant differences among sleep and awake times. Being asleep or awake are two completely different physical states and the effect of these situations on neuronal activity cannot be ignored. In such a case, cluster formations (i.e., memberships) of neuronal responses during a day may heavily depend on time. But for most of the case these types of phase differences or alternating states are not so apparent. Due to such circumstances, like in time series variables analyzing, it is necessary to problematize that time series clustering can be time-dependent. The main motivation of the study is developing the clustering approach as sensitive to possible temporal cluster formations. Hence, the proposed clustering approach utilizes multifaceted time series models that can approximate the time-dependent structure of generating mechanisms. In this regard, the datasets and corresponding clusters we generated to test our approach have been produced with similar concerns.

On the extent of the experiments.

We artificially generated sets (i.e., clusters) of time series data regarding 5 sets of experiments employed well-known and heterogeneously complex time series DGMs. For each experiment set and its DGMs (i.e., models), we considered two main cases that can be named as non-temporal and temporal formations. First, we evaluated the performances over fixed clusters during a period of time. Then we extended the evaluation for each case by composing changes of DGMs over a longer period of time and so clusters. We want to state here that each generated cluster is characterized by a bunch of time series produced by a specific DGM or by a composition of diversified DGMs depending on time. In this sense, one can consider that the centers of generated clusters are composed of underlying DGMs. That is to say, if a DGM specifies a cluster has diversified into several new DGMs during a time, the cluster structure will also change accordingly. Thus, in this chapter, the clustering performances of the proposed time series clustering approach and the state-of-the-art time series clustering methods are evaluated and compared over synthetically generated clusters.

Generated sets of time series are desired to have and exhibit various properties of time series seen in real life such as stationarity, non-stationarity, seasonality, non-linearity, explicit and smooth regime switching structures, and with heteroscedasticity in conditional variance.

To evaluate the responsiveness of the proposed approach over similar data-generating processes, considered DGMs that produce synthetic data are determined considering their similarity in terms of dynamics and structure, such as lag formation, coefficient strings, and volatility compositions. Besides, each considered DGM contains the same type of the component (i.e., conditional mean/variance) present in at least one of the other DGMs. It should be noted that this perspective of synthetic data generation makes the clustering task of time series more challenging and realistic. Nonetheless, as there may be an infinite number of scenario assumptions that can be considered to test the proposed approach, it should also be noted that we limit our work to the research done in this section.

In this chapter, a total of 66 different clustering plans comprised of the use of 22 different dissimilarity measures on 3 different clustering methods (i.e., Fuzzy c-means, Partitioning Around Medoids (PAM), Spectral) are compared with the proposed time series clustering approach. The state-of-the-art dissimilarity measures proposed by recent literature are presented in Table 4.1. The dissimilarity measures used in the comparison study have been comprehensively discussed in [55], and the corresponding tools have been made available in the **R** package named as **`TSclust`**.

Model-Free dissimilarity measures focus on the pattern or shape related similarities and properties of time series. Complexity Based dissimilarities consider the use of the information-theoretical properties of time series for measuring the dissimilarities between time series. Model-Based dissimilarity measures operate on the use of the properties of probability or time series models to get similarities between time series. On the other hand, as explained in Chapter 3, the proposed clustering approach can be considered as a distance-based time series clustering algorithm that operates on a mixture of a model and non-model based features derived from the time series dataset to be clustered.

4.1 Performance Evaluation over Synthetic Data

Four different experiments on synthetic datasets are conducted within the scope of this chapter. Each included evaluations considering different types of synthetic datasets/-clusters produced based on the specific time series models and combinations of these models. Thus, a total of 9 distinct scenarios are examined. However, for readability and ease of follow-up, the results of only two of these experiments are given in a subsection, while the remaining experiments are given in Appendix A.

Table 4.1 Dissimilarity measures used for comparison in univariate clustering.

Dissimilarity Measure	Description	Reference
	Model free approaches	
D_{ACF}	Autocorrelation based distance.	[60]
D_{PACF}	Autocorrelation based distance.	[60]
D_{EUCL}	Euclidean distance.	[55]
D_{CORT}	Dissimilarity index that covers both dissimilarity on raw values and dissimilarity on temporal correlation behaviors.	[61]
D_{COR}	Correlation-based dissimilarity.	[62]
D_{DWT}	Dissimilarity for time series based on wavelet feature extraction.	[63]
D_{PER}	Periodogram based dissimilarity.	[64]
$D_{INT.PER}$	Integrated periodogram based dissimilarity.	[65]
$D_{SPEC.LLR1}$	General spectral dissimilarity measure using local-linear estimation of the log-spectra via least squares estimation.	[66, 67]
$D_{SPEC.LLR2}$	General spectral dissimilarity measure using local-linear estimation of the log-spectra via maximum likelihood estimation.	[66, 67]
D_{GLK}	Dissimilarity based on the generalized likelihood ratio test.	[68, 69]
D_{ISD}	Dissimilarity based on the integrated squared difference between the log-spectra.	[69]
D_{SAX}	Symbolic aggregate approximation related functions.	[70, 71]
D_{VR}	Distance based on variance ratio statistics.	[72]
D_{DTW}	Dynamic time warping distance.	[59, 73]
	Model based approaches	
D_{AR-LPC}	Dissimilarity based on linear predicitive coding (LPC) cepstral coefficients.	[78]
D_{AR-MAH}	Model-based dissimilarity proposed by Maharaj.	[77, 80]
D_{AR-PIC}	Model-based dissimilarity measure proposed by piccolo.	[58]
	Complexity based approaches	
D_{CID}	Complexity-invariant distance measure for time series.	[74]
D_{PDC}	Permutation distribution distance.	[56]
D_{CDM}	Compression-based dissimilarity measure.	[75]
D_{NCD}	Normalized compression distance.	[76]

As stated before, artificially generated data sets are created by hypothetical DGMs. It should be noted again that the bunch of time series generated from each DGM also forms an artificial cluster subject to clustering experiments for performance evaluation of the proposed clustering approach. Each DGM is composed of specific time series models that we denominate the definite components as the stems. In some scenarios, a stem with an error component may directly form a DGM, and in some scenarios, combinations of stems form a DGM. Therefore, under each subsection, the stems and then the DGMs consists of these stems are introduced, and how the clusters corresponding to these DGMs are formed is described.

For an experiment in this section, the general procedure of generating samples, clusters, and then evaluations of clustering performances over the generated (i.e., synthetic) dataset can be summarized as follows:

* **nos** $\Leftarrow the\ number\ of\ sources$: Defines the number of hypothetical DGMs, nos, be used in the experiment.

* **noc** $\Leftarrow the\ number\ of\ clusters$: The number of hypothetical DGMs, nos, also defines the number of true clusters, noc (i.e., $nos = noc$).

* **ss** $\Leftarrow sample\ size$: Define the time span, length or number of observations, ss, for each generated series.

* **cs** $\Leftarrow cluster\ size$: Define the cluster size, cs, which determines the number of series/members in each cluster.

* **rn** $\Leftarrow repetition\ number$: Defines the number of repetitions of the experiment.

After deciding values for ***nos***, ***noc***, ***ss***, ***cs*** and ***rn***, the synthetic dataset generation, clustering, and comparison phases of an experiment can be summarized in 4 steps:

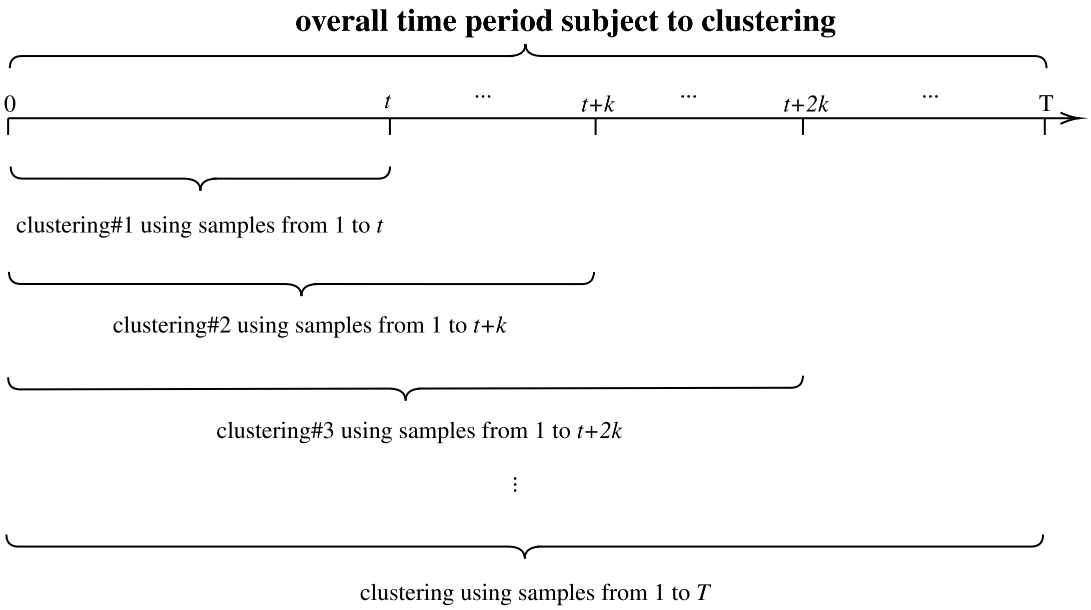
1. Each DGM used to generate ***cs*** number of time series with the length of ***ss*** to form a corresponding cluster.
2. Clustering by proposed approach and other considered clustering methods for comparison are performed over ***nos*** $\times$ ***cs*** $\times$ ***ss*** dimensional dataset and clustering accuracies are saved.
3. Steps from **1** to **2** repeated **rn** times to get average accuracy indexes of clustering methods and the proposed approach.
4. The average accuracy indexes of all methods are compared.

In this context, the number of clusters in an artificial dataset is equal to the number of DGMs used to create it. The number of elements in each cluster is equal to the number of time series generated from each corresponding DGM. Hence, while generating the dataset, we also know the true cluster formation in it. Accordingly, by clustering the generated dataset with the proposed clustering approach, we can assess its performance and compare the clustering results with the true cluster formation.

In addition to performance evaluation over non-temporal clusters, we essentially desired to evaluate the performance of the proposed clustering approach over changing cluster formations through time. To do so, in some experiments, we ensured that hypothetical DGMs consists of different generating mechanisms in different time windows.

In this type of scenario, different from generating a dataset for non-temporal cluster formations, first, we need to generate a dataset consisting of a different number of clusters and cluster sizes for varying time intervals provided that they are within the overall time period subject to clustering. Then, this artificial dataset needs to be clustered in a consecutive and additive manner to evaluate the performance of the proposed approach. In other words, the number of clusters and cluster sizes may change after a time point, and clustering sometime after that point should result in a different cluster formation due to the differentiation in data generation mechanisms. Accordingly, we require to examine the clustering performance of the proposed approach against such likely circumstances. As we state in Section 3.4.2, the proposed clustering approach is designed to utilize the cumulative sum of distances to discover the changes of clusters over time if there is any. In this type of experiment, to examine the clustering approach's aforementioned capacity, the time intervals in an overall time period that is available are enlarged step by step during clustering. Thus, evaluation can be made by comparing cluster formations obtained at different time intervals with actual cluster formations. This is also because we intend to test how soon the proposed approach can respond to probable variations in clusters that may occur when incoming (i.e., streaming) data in real-life extend a dataset to be clustered.

Let's try to simplify these explanations with a fictitious example, and depict the corresponding time-code of a generated dataset as a diagram in Figure 4.1. Suppose that by using several hypothetical DGMs we have created a synthetic dataset of sample sizes T, and also assume that some of time series generating mechanisms have been subjected to the changes after the time point, t, which will also effect the cluster formations after that time point. This means that the structures that generate some series in the dataset do not remain unchanged over time. Therefore, after the time point t, such series should not be considered the members of the clusters they belonged to before. In brief, the cluster formation on the overall time period from 1 to T is different than the cluster formation on samples from 1 to t. To understand whether we can identify these reformations of clusters via the proposed clustering approach, we can evaluate the performance by comparing the results of the clusterings, made at the time intervals like shown in the diagram in Figure 4.1, with the cluster formations already present in the dataset. Such artificial datasets, designed to contain different cluster formations through time, can enable us to evaluate the performance of the proposed clustering approach on checking whether temporal/dynamic cluster formations in the dataset.


overall time period subject to clustering
0
t
...
t+k
...
t+2k
...
T
clustering#1 using samples from 1 to t
clustering#2 using samples from 1 to t+k
clustering#3 using samples from 1 to t+2k
⋮
clustering using samples from 1 to T

*k is the increment value of samples subjected to clustering.

Figure 4.1 Diagram for testing the temporality of clusters.

However, in real-life problems, the aim is to determine the clusters by using all available observations for clustering at time T. And, more importantly, the ultimate goal is able to detect probable changes in clusters when clustering the updated dataset after new incoming data.

On the Implementation of the Proposed Approach and Hyperparameters

Before presenting the experiments and their results, we would like to state the specifications (i.e., hyperparameters) we employed to implement the proposed approach. Let us remind that the performance evaluations presented in this section are limited to predetermined assignations of hyperparameters.

In Section 3.4.2, while explaining the proposed approach, we stated that each time series subject to clustering should be represented by a competent feature vector. The elements of the feature vector consisting of 11 parts and fixed values of hyperparameters corresponding to them are given in Table 4.2. The implementation steps of the proposed approach are also given in Figure 3.4.

Table 4.2 The feature vector and fixed hyperparameters in this study.

Feature	Hyperparameter Value	Source
		model-based elements
$\widehat{\phi}^{(ar)}_{1:p_{ar}}$	$p_{ar} = 12$	AR
$\widehat{\phi}^{(1:k)}_{1:ar_j}$	$k = 3, ar_1 = 4, ar_2 = 4, ar_3 = 4$	DTGARCH
$\widehat{\alpha}^{(1:k)}_0$, $\widehat{\alpha}^{(1:k)}_1$, $\widehat{\beta}^{(1:k)}_1$	$k = 3$	DTGARCH
$\widehat{\rho_a}(h)$, $\widehat{\pi_a}(h,h)$	$h = 18$	AR
$\widehat{\rho_{a^2}}(h)$, $\widehat{\pi_{a^2}}(h,h)$	$h = 18$	AR
$\widehat{\rho_\varepsilon}(h)$, $\widehat{\pi_\varepsilon}(h,h)$	$h = 18$	DTGARCH
$\widehat{\rho_{\varepsilon^2}}(h)$, $\widehat{\pi_{\varepsilon^2}}(h,h)$	$h = 18$	DTGARCH
$\lvert C_\kappa(\widehat{\varepsilon^2})\rvert$	$\kappa = 0, \ldots, T/2$	DTGARCH
$\lvert C_\kappa(\widehat{\eta}^{(1:k)})\rvert$	$\kappa = 0, \ldots, T/2$	DTGARCH
		model-free elements
$\widehat{\rho_y}(h)$, $\widehat{\pi_y}(h,h)$	$h = 18$	raw time series
$\lvert C_\kappa(y)\rvert$	$\kappa = 0, \ldots, T/2$	raw time series

p_{ar} denotes the order of the AR model.
k denotes the number of regimes in DTGARCH model.
ar_j denotes the regime specific AR order of DTGARCH model where $j : 1, 2, 3$.
$\widehat{a}$ denotes residuals obtained from the estimation of the AR model.
$\widehat{\varepsilon}$ denotes residuals obtained from the estimation of the DTGARCH model.
h denotes the lag where the serial correlations are calculated up to.
$\widehat{\eta}$ denotes conditional variance from the estimation of the DTGARCH model.
κ denotes the spectrum length and T is the length of a time series.

Fixation of hyper-parameter values may seem discretional, however these values are just proper for the purposes of the study. It should be noted that the aim is not to find the most suitable model for the time series. Here, for our clustering aims, we mainly focus on the operability of the results obtained from the associations of the time series with the determined forms of the mentioned models. Therefore while aiming the construct competent feature vectors, the practical approach for determining the hyper-parameter values is whether the models we employ are easily accessible and whether the combinations of the estimation outputs could eliminate the overfitting problems.

4.1.1 First Set of Experiments

In this set of experiments, we used 5 specific stems given in Table 4.3 which are originating from well-known time series models such as Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Moving Average (SARMA), Logistic Smooth Transition Autoregressive (LSTAR), and Self Exciting Threshold Autoregressive (SETAR) models.

Table 4.3 Utilized specifications of mean processes in the first set of experiments.

STEM_1: $Y_t = 0.80Y_{t-12} + \varepsilon_t + 0.70\varepsilon_{t-12}$

STEM_2: $Y_t = 0.90Y_{t-1} - 0.50Y_{t-2} + 0.15Y_{t-3} + \varepsilon_t - 0.20\varepsilon_{t-1} + 0.25\varepsilon_{t-2}$

STEM_3: $Y_t = 2.55Y_{t-1} - 2.30Y_{t-2} + 0.75Y_{t-3} + \varepsilon_t + 0.80\varepsilon_{t-1} + 0.50\varepsilon_{t-2}$

STEM_4: $Y_t = 0.80Y_{t-1} - 0.80Y_{t-1}(1 + e^{-10Y_{t-1}})^{-1} + \varepsilon_t$

STEM_5: $Y_t =$

$$\begin{cases} -0.15 + 0.85Y_{t-1} - 0.15Y_{t-2} + 0.30Y_{t-3} - 0.40Y_{t-4} + \varepsilon_t^{(1)}, & Y_{t-1} < -1.2 \\ 2.20 + 0.20Y_{t-1} - 1.70Y_{t-2} + 0.25Y_{t-3} + \varepsilon_t^{(2)}, & -1.2 < Y_{t-1} \leq 1.2 \\ 1.00 + 0.50Y_{t-1} - 1.15Y_{t-2} - 0.60Y_{t-4} + \varepsilon_t^{(3)}, & Y_{t-1} \geq 1.2 \end{cases}$$

Additionally, error components listed in Table 4.4 are picked from Generalized Autoregressive Conditional Heteroskedasticity (GARCH) and Asymmetric Power Autoregressive Conditional Heteroskedasticity (APARCH) types of conditional variance models.

Table 4.4 Utilized specifications of error structures in the first set of experiments.

$$\mathbf{APARCH^{(a)}}\text{: } h_t^{1.5/2} = 0.50 + 0.55(|\varepsilon_{t-1}| - 0.10\varepsilon_{t-1})^{1.5/2} + 0.40h_{t-1}^{1.5/2}$$

$$\mathbf{APARCH^{(b)}}\text{: } h_t^{1.5/2} = 0.50 + 0.20(|\varepsilon_{t-2}| + 0.30\varepsilon_{t-2})^{1.5/2} + 0.75h_{t-1}^{1.5/2}$$

$$\mathbf{GARCH^{(a_1,a_2,a_3)}} =$$

$$\begin{cases} h_t = 0.10 + 0.15\varepsilon_{t-1}^2 + 0.10\varepsilon_{t-2}^2 + 0.25h_{t-1} + 0.10h_{t-2} + 0.20h_{t-3} \\ h_t = 0.15 + 0.25\varepsilon_{t-1}^2 + 0.10\varepsilon_{t-2}^2 + 0.10h_{t-1} + 0.05h_{t-2} + 0.15h_{t-3} \\ h_t = 0.15 + 0.10\varepsilon_{t-1}^2 + 0.25\varepsilon_{t-2}^2 + 0.25h_{t-1} + 0.10h_{t-2} + 0.15h_{t-3} \end{cases}$$

$$\mathbf{GARCH^{(b_1,b_2,b_3)}} =$$

$$\begin{cases} h_t = 0.30 + 0.05\varepsilon_{t-1}^2 + 0.10\varepsilon_{t-2}^2 + 0.05h_{t-1} + 0.25h_{t-2} + 0.10h_{t-3} \\ h_t = 0.35 + 0.15\varepsilon_{t-1}^2 + 0.15\varepsilon_{t-2}^2 + 0.20h_{t-1} + 0.25h_{t-2} + 0.05h_{t-3} \\ h_t = 0.05 + 0.35\varepsilon_{t-1}^2 + 0.05\varepsilon_{t-2}^2 + 0.05h_{t-1} + 0.15h_{t-2} + 0.25h_{t-3} \end{cases}$$

where $\varepsilon_t = \sqrt{h_t}e_t,$ and $e_t \overset{i.i.d}{\sim} D_{sged}(0,1).$

Here in this list of specific time series processes, the $\mathbf{STEM\_1}$ can be viewed as an example of the model known as SARMA which can generate seasonal characteristics found in observed seasonal time series. The process $\mathbf{STEM\_2}$ can produce stationary time series and it is an example of a linear time series model named ARMA. The ARMA type models are widely used in the literature of time series analysis and have extensive application areas. The process $\mathbf{STEM\_3}$ is an example of the ARMA models with an $Integration$ order means non-stationarity exists in the mean process and it is also a famous modeling concept widely used for time series analysis. The $\mathbf{STEM\_3}$ can generate non-stationarity that is one of the most seen characteristics in real time series data. The process $\mathbf{STEM\_4}$ is an example from the LSTAR model which is a non-linear time series model used for explaining non-linearities in time series. The process $\mathbf{STEM\_4}$ can generate non-stationary non-linear time series. The process $\mathbf{STEM\_5}$ is an example of the SETAR model described by two threshold values. It is also a non-linear model and an extension of the linear time series models. By employing the $\mathbf{STEM\_5}$, we can produce time series data with several salient properties such as non-stationary, non-linear, and abrupt behavioral properties.

The LSTAR and the SETAR type models can represent or generate many features such as complexity, abruptness, non-linearity, and non-stationarity encountered in observed time series.

For a long time, observed time series data were analyzed with linearity and stationarity assumptions. Efforts to estimate and explain the features such as non-linearity, non-stationarity, and complexity in the observed data are now considered to be a more realistic approach. The study of the real world with the premise of nonlinearity has become quite frequent in the literature [140].

Instead of following the normal distribution or pure white noise assumption for the error structures in hypothetical DGMs, we decided to utilize the heteroscedasticity phenomenon that is frequently encountered in observed time series data. The APARCH and GARCH are well-known models for modeling time varying variance of the errors. The APARCH model nests most other volatility models, as well as the GARCH model. The processes $\mathbf{GARCH^{(a_1,a_2,a_3)}}$ and $\mathbf{GARCH^{(b_1,b_2,b_3)}}$ are alternatively used for the 3 regime error terms of the $\mathbf{STEM\_5}$ (e.g., $\varepsilon_t^{(1)}$, $\varepsilon_t^{(2)}$, $\varepsilon_t^{(3)}$). We use the beforementioned error structures in hypothetical DGMs to compare the sensitivity of the proposed approach versus other clustering methods regarding heteroscedasticity.

Given the context above, for instance, connecting $\mathbf{STEM\_1}$ with the $\mathbf{APARCH^{(a)}}$ process for the error term can be considered as an example of a hypothetical data generating mechanism. By generating a bunch of time series from this source, we can create an artificial cluster of time series. It is possible to produce a data set, as comprehensive as it's required, from hypothetical DGMs which are made by the combinations of the stems with error processes given above. Hence, we can state that there are a number of clusters present in the dataset equal to the number of hypothetical DGMs used to create it.

The results given in the following parts demonstrate to what extent the clusters that are assumed to be present in the created data set can be correctly determined.

4.1.1.1 Case 1 : Non-temporal 10 Cluster

Essentially, in this scenario or case, the proposed approach and the other clustering algorithms given in Table 4.1 are compared over 10 artificial clusters of time series created by generating data from 10 hypothetical DGM. Hypothetical DGMs given in Table 4.5 made from specific processes given in the Tables 4.3 and 4.4.

Table 4.5 Hypothetical DGMs used in Case 1.

$\mathbf{DGM01} \coloneqq \mathbf{STEM\_1} \mid \{\varepsilon_t\} \sim \mathbf{APARCH^{(a)}}$

$\mathbf{DGM02} \coloneqq \mathbf{STEM\_1} \mid \{\varepsilon_t\} \sim \mathbf{APARCH^{(b)}}$

$\mathbf{DGM03} \coloneqq \mathbf{STEM\_2} \mid \{\varepsilon_t\} \sim \mathbf{APARCH^{(a)}}$

$\mathbf{DGM04} \coloneqq \mathbf{STEM\_2} \mid \{\varepsilon_t\} \sim \mathbf{APARCH^{(b)}}$

$\mathbf{DGM05} \coloneqq \mathbf{STEM\_3} \mid \{\varepsilon_t\} \sim \mathbf{APARCH^{(a)}}$

$\mathbf{DGM06} \coloneqq \mathbf{STEM\_3} \mid \{\varepsilon_t\} \sim \mathbf{APARCH^{(b)}}$

$\mathbf{DGM07} \coloneqq \mathbf{STEM\_4} \mid \{\varepsilon_t\} \sim \mathbf{APARCH^{(a)}}$

$\mathbf{DGM08} \coloneqq \mathbf{STEM\_4} \mid \{\varepsilon_t\} \sim \mathbf{APARCH^{(b)}}$

$\mathbf{DGM09} \coloneqq \mathbf{STEM\_5} \mid \{\varepsilon_t^{(1)},\, \varepsilon_t^{(2)},\, \varepsilon_t^{(3)}\} \sim \mathbf{GARCH^{(a_1,a_2,a_3)}}$

$\mathbf{DGM10} \coloneqq \mathbf{STEM\_5} \mid \{\varepsilon_t^{(1)},\, \varepsilon_t^{(2)},\, \varepsilon_t^{(3)}\} \sim \mathbf{GARCH^{(b_1,b_2,b_3)}}$

The hypothetical DGMs used in Case 1 are designated to cover different types of aspects of time series, such as stationarity in mean, non-stationarity, seasonality, nonlinearity, and regime-switching structure. To evaluate the success of the proposed clustering approach on similar hypothetical DGMs, generating mechanisms are chosen regarding their similarity with respect to dynamics/structure, such as orders, lag delays, coefficients, and error components. Furthermore, each DGM is possessing the identical STEM component (i.e. conditional mean/variance), which is also being in at least one of the other hypothetical DGMs.

Visual exemplars of the DGMs' motives are shown in Figure 4.2. Please observe that DGMs sharing the same STEMs produce similar traces.

Therefore, to compare clustering methods through challenging experiments, artificial clusters are formed by the data generation mechanisms containing similar structures.

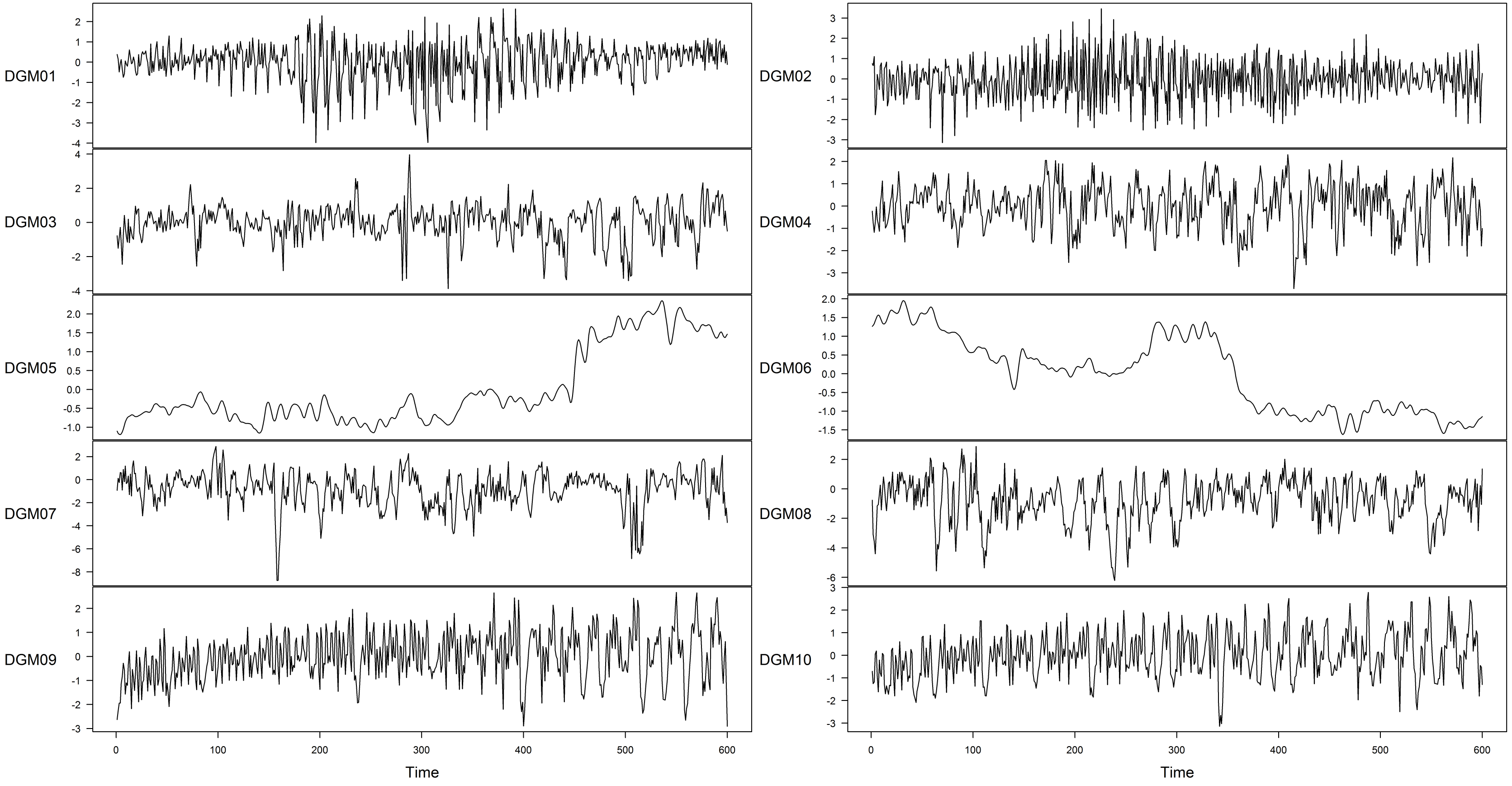


Figure 4.2 Example of the simulation sample from 10 DGMs listed in Table 4.5.

Hypothetical DGMs numbered from 1 to 6 in Table 4.5 can be named as linear autoregressive processes in the conditional mean, and DGMs numbered from 7 to 10 are non-linear processes. Especially: the 1st and 2nd are seasonal processes and have the same seasonal properties in the conditional mean but only varies w.r.t error processes which are also very similar components; the 3rd and 4th ones are stationary linear autoregressive representations that holds the same conditional mean but with slightly different error components; the 5th and 6th ones are non-stationary linear autoregressive processes partake the same conditional mean; the 7th and 8th processes are non-linear non-stationary autoregressive processes called as logistic smooth transition regressions that using the same conditional mean but with different errors. The last two models in Table 4.5 are three regime threshold non-linear processes with different conditional variance processes in each regime.

The generating samples, clusters, and then evaluations of clustering performances for Case 1 can be summarized as follows:

* **nos = 10** (i.e., the number of hypothetical DGMs used in the experiment.)

* **noc = 10** (i.e., the number of true clusters.)

* **cs = 10** (i.e., the number of series/members in each cluster.)

* **ss = 600** (i.e., the number of observations for each generated series.)

* **rn = 100** (i.e., the number of repetition of the experiment.)

1. Each DGM in Table 4.5 is used to generate **10** (i.e., **cs**) number of time series with the length of **600** (i.e., **ss**) to form a corresponding cluster.
2. Clustering done by the proposed approach given in Section 3.4.2 and through implementation steps given in Figure 3.4, and **22** dissimilarity measures given in Table 4.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed over **10** $\times$ **10** $\times$ **600** dimensional dataset and clustering accuracies are saved.
3. Steps from **1** to **2** repeated **100** (i.e., **rn**) times to get average accuracies of **66** (i.e., **3** $*$ **22**) clustering schemes and the proposed approach.
4. The average accuracy indexes of all methods considered in the study are displayed for comparison in Table 4.6.

Clustering results for Case 1 given in Table 4.6. Results for compared clustering methods are obtained using samples with a length of 600 observations. However, 3 different sample sizes (i.e., $ss = 350$, $ss = 475$, $ss = 600$) are considered to see the effects of the lengths of the series on the performance of the proposed approach. Please remind that at least 350 observations are needed to estimate the parameters of the DTGARCH model.

Table 4.6 Univariate Experiment 1 Case 1: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach.

Dissimilarity measure	Clustering method		
	FUZZY	PAM	SPECTRAL
Model free			
D_{ACF}	0.545	0.575	0.579
D_{PACF}	0.438	0.602	0.596
D_{EUCL}	0.262	0.357	0.452
D_{CORT}	0.268	0.387	0.360
D_{COR}	0.264	0.389	0.327
D_{DWT}	0.410	0.447	0.404
D_{PER}	0.309	0.549	0.572
$D_{INT.PER}$	0.591	0.605	0.531
$D_{SPEC.LLR1}$	0.607	0.621	0.601
$D_{SPEC.LLR2}$	0.615	0.615	0.606
D_{GLK}	0.605	0.613	0.441
D_{ISD}	0.595	0.609	0.503
D_{SAX}	0.377	0.309	0.327
D_{VR}	0.548	0.556	0.360
D_{DTW}	0.282	0.510	0.291
Model based			
D_{AR-LPC}	0.573	0.596	0.585
D_{AR-MAH}	0.620	0.618	0.607
D_{AR-PIC}	0.632	0.611	0.598
Complexity based			
D_{CID}	0.371	0.496	0.405
D_{PDC}	0.554	0.614	0.506
D_{CDM}	0.298	0.394	0.266
D_{NCD}	0.315	0.395	0.269

	Proposed Clustering Approach		
	TSMB-SPCL-UV		
	$ss = 350$	$ss = 475$	$ss = 600$
L^2 (Euclidean) Distance	0.796	0.823	0.848

Under Case 1, the outcomes revealed that 8 of the compared clustering approaches produce clustering accuracies less than 0.50. The overall best result among compared methods is 0.63 in average, achieved by the AR models based dissimilarity measures.

Furthermore, as can be seen from the Table 4.6, the accuracy indexes of the proposed approach for three different sample sizes are considerably high compared to other methods. It is seen that the increase in sample length has a positive effect on the performance of the proposed approach. According to the results obtained, we can say that, an average of 84% of the series generated in this scenario will be correctly grouped with similar ones by the proposed clustering approach. The most important reason for this is the ability of the proposed clustering approach to detect heteroscedasticity in the series.

When the number of true clusters is set to 10, the classification accuracies of the time series using the proposed clustering approach are given in Figure 4.3. Classification accuracies over 100 replicates have varied between 0.73 and 0.94, and the exact grouping accuracies are mostly dispersed around 0.84.

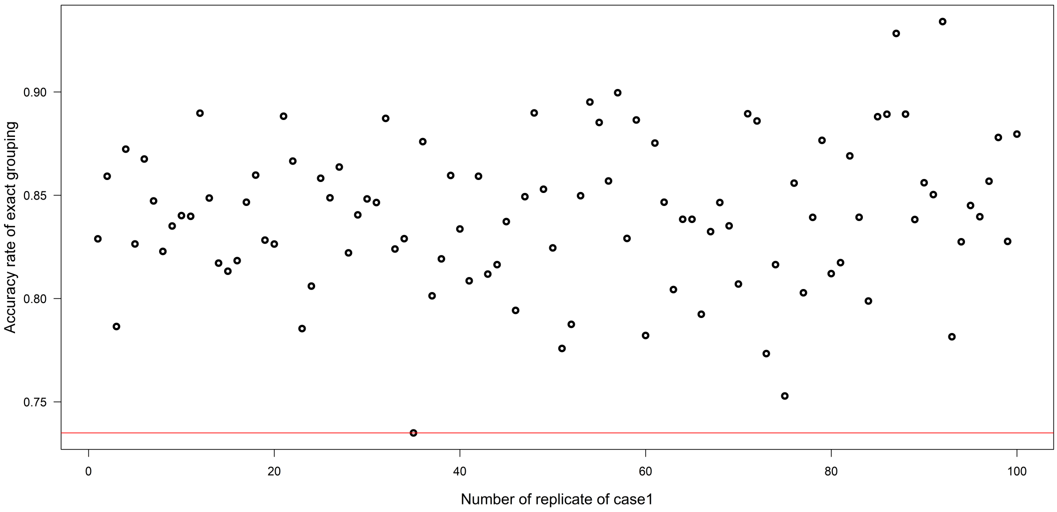


Figure 4.3 Accuracy rates vs. TSMB-SPCL-UV over 100 replicates.

The accuracy indexes in Table 4.6 are the outcomes obtained employing the true number of clusters 10 while clustering the dataset. We stated in previous chapter that in cases where the true number of clusters in the dataset is unknown, the GAP statistic proposed by Tibshirani et al. [139] described in Section 3.4.2 could be used to determine the possible number of clusters. In the graph given in Figure 4.4, we see the values of the GAP statistic over the possible number of clusters for a repetition of Case 1. Thus, in cases where the number of clusters is unknown, the number of clusters for which the GAP statistic approaches its maximum value can be used as a take-off. In the sample graph depicted in Figure 4.4, we observe that the GAP statistic luckily peaks at the actual number of clusters in the data set.

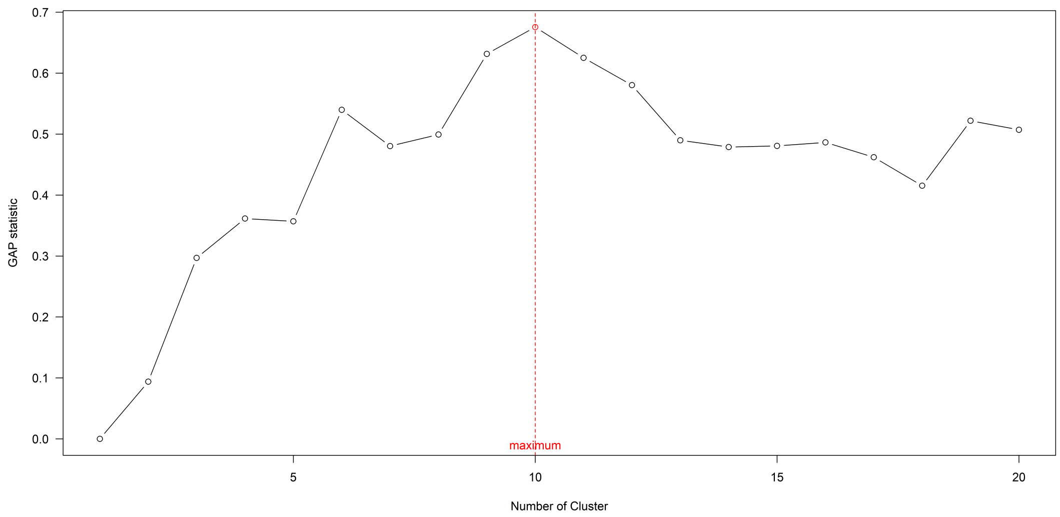


Figure 4.4 GAP statistic per cluster number for a replicate using TSMB-SPCL-UV.

The GAP statistic is a promising tool in inferring the optimum/plausible number of clusters when the number of clusters is unknown. Sampling distribution of the GAP statistics could give a view regarding its potential. In Figure 4.5, the sampling distribution of the GAP statistics per possible cluster value obtained from 100 trials is represented with the help of box plots.

Notice that box-plots obtained at cluster numbers 6, 9, 10, and 11 in Figure 4.5 are possible cluster numbers can be offered by the GAP statistics. This indicates that the GAP statistics may offer local maximums that we should aware of while we utilizing these statistics. For example, the sudden jumps and subsequent declines of the GAP statistics obtained at cluster number 6 give us a clue that the possible number of clusters in the data set may be considered as 6. Similarly, in some repetitions, the GAP statistic may reach its maximum within the close vicinity of the actual number of clusters.

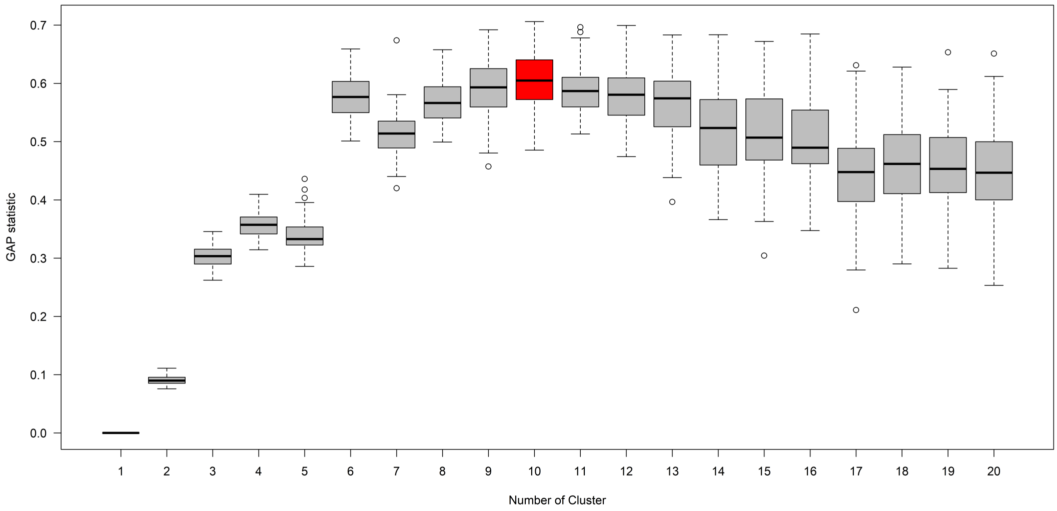


Figure 4.5 GAP statistics' Box-Plot per cluster number for 100 replicates using TSMB-SPCL-UV.

In the experiments included in case 1, the actual number of clusters of 10 is determined correctly in 30 out of 100 repetitions. Apart from this, the optimum number of clusters is suggested as 9 or 11 in 32 repetitions. The distribution of possible cluster numbers in which the GAP statistic reaches its maximum throughout 100 replicates is given in Table 4.7.

Table 4.7 Estimated number of clusters by the global maximum of the GAP statistics out of 100 repetitions using TSMB-SPCL-UV.

cluster number, $\widehat{cl}$	6	7	8	9	10	11	12	13	14	15	16
counts of peaks	3	1	7	20	30	12	7	8	4	4	4

Although the GAP statistics obtained over possible cluster numbers do not give an exact solution for determining the number of clusters, they still can provide some clues for the initial stage of a clustering task. Additionally, finding the optimum number of classes when the possible number of clusters in any data set is not known is not one of the primary questions that this study explores to answer, so we leave more detailed investigations on the subject to further studies.

4.1.1.2 Case 2 : Temporal 10 Cluster

This subsection aims to evaluate the performance of the proposed approach and the compared methods to detect the time-varying cluster formations. For this aim, in Case 2 of the first set of experiments, the datasets are generated to test the proposed clustering approach's capacity to detect time-varying cluster compositions. These datasets are obtained from 10 different DGMs, and thus there are 10 clusters in total. In other word, the dataset generation scheme in this scenario consists of 10 cluster sources when considering the entire sample sizes of time series of length 1000. However, some of the sources of time series generation mechanisms are desired to have slight changes after a certain time point. Therefore, our artificial datasets have 2 different cluster formations throughout the whole sample sizes.

Hypothetical DGMs used to create the dataset subjected to clustering in this subsection are given in Table 4.8. These configurations are made from specific processes given in the Tables 4.3 and 4.4.

Table 4.8 Hypothetical DGMs used in Case 2.

	used to generate time series of length 1000		
	sources of samples from 1 to 600		sources of samples from 601 to 1000
DGM01 $\coloneqq$	**STEM_1** $\mid \{\varepsilon_t\} \sim$ **APARCH**$^{(\mathbf{a})}$	▷	**STEM_1** $\mid \{\varepsilon_t\} \sim$ **APARCH**$^{(\mathbf{a})}$
DGM02 $\coloneqq$	**STEM_1** $\mid \{\varepsilon_t\} \sim$ **APARCH**$^{(\mathbf{a})}$	▶	**STEM_1** $\mid \{\varepsilon_t\} \sim$ **APARCH**$^{(\mathbf{b})}$
DGM03 $\coloneqq$	**STEM_2** $\mid \{\varepsilon_t\} \sim$ **APARCH**$^{(\mathbf{a})}$	▷	**STEM_2** $\mid \{\varepsilon_t\} \sim$ **APARCH**$^{(\mathbf{a})}$
DGM04 $\coloneqq$	**STEM_2** $\mid \{\varepsilon_t\} \sim$ **APARCH**$^{(\mathbf{a})}$	▶	**STEM_2** $\mid \{\varepsilon_t\} \sim$ **APARCH**$^{(\mathbf{b})}$
DGM05 $\coloneqq$	**STEM_3** $\mid \{\varepsilon_t\} \sim$ **APARCH**$^{(\mathbf{a})}$	▷	**STEM_3** $\mid \{\varepsilon_t\} \sim$ **APARCH**$^{(\mathbf{a})}$
DGM06 $\coloneqq$	**STEM_3** $\mid \{\varepsilon_t\} \sim$ **APARCH**$^{(\mathbf{a})}$	▶	**STEM_3** $\mid \{\varepsilon_t\} \sim$ **APARCH**$^{(\mathbf{b})}$
DGM07 $\coloneqq$	**STEM_4** $\mid \{\varepsilon_t\} \sim$ **APARCH**$^{(\mathbf{a})}$	▷	**STEM_4** $\mid \{\varepsilon_t\} \sim$ **APARCH**$^{(\mathbf{a})}$
DGM08 $\coloneqq$	**STEM_4** $\mid \{\varepsilon_t\} \sim$ **APARCH**$^{(\mathbf{a})}$	▶	**STEM_4** $\mid \{\varepsilon_t\} \sim$ **APARCH**$^{(\mathbf{b})}$
DGM09 $\coloneqq$	**STEM_5** $\mid \{\varepsilon_t\} \sim$ **GARCH**$^{(\mathbf{a_1,a_2,a_3})}$	▷	**STEM_5** $\mid \{\varepsilon_t\} \sim$ **GARCH**$^{(\mathbf{a_1,a_2,a_3})}$
DGM10 $\coloneqq$	**STEM_5** $\mid \{\varepsilon_t\} \sim$ **GARCH**$^{(\mathbf{a_1,a_2,a_3})}$	▶	**STEM_5** $\mid \{\varepsilon_t\} \sim$ **GARCH**$^{(\mathbf{b_1,b_2,b_3})}$

′ ▷′ indicates where the data generation mechanisms remain unchanged.
′ ▶′ indicates where the data generating mechanisms are changed slightly after 600th observation.

Some information about the properties of the processes in Table 4.8 is given in the preceding subsection. The data generation mechanisms in this table are used to generate series with a length of 1000 observations. As seen from the table, the structural variations between the DGMs are created only in error components (i.e., conditional variances). Conditional mean processes have remained unchanged. Thus, for example, the source that constitutes DGM01 does not change in this scenario. In contrast, the source that creates DGM02 changes when it reaches the 600th observation, and the last 400 observations are produced under the change in conditional variance only. The reason for this setup in Case 2 is to evaluate the responding ability of our proposed approach to changes in heteroscedasticity and small but determinant perturbations in sources of time series. It should be noted that, when examining the first 600 data-length samples of a dataset generated by these mechanisms, the dataset has 5 clusters. However, from the 601st time point, this data generating scheme changes the cluster formations of a dataset into a composition with 10 clusters. Therefore, clustering algorithms are performed at the 600th time point and the 1000th time point of the dataset to measure the success of the proposed approach and the compared methods on the time-varying clusters.

It should be noted that when 10 time series generation is planned from each DGM within the specified framework above, the number of clusters be 5 and the number of cluster members be 20 in the slice containing the first 600 observations of the dataset. Thus, the number of clusters and cluster members to be determined in the dataset to be examined in two theoretical time periods are emphasized. It should also be noted that any change in a data generation mechanism effects on clustering formations can be noticed by the proposed clustering approach approximately after it leaves a trace of at least 350 observations to the history of the time series. Among other reasons, this is more related to the minimum number of observations required for the estimations of the time series model used in the proposed approach.

The changes placed in the data generation mechanisms in this scenario do not create dramatic variations in the trace of time series. The reason for the comparatively small changes created in DGMs is to see the responsiveness of the proposed clustering approach in likely similar situations. Clustering accuracies obtained for Case 2 are given in Table 4.9.

The generating samples, clusters, and then evaluations of clustering performances for Case 2 can be summarized as follows:

* **nos_oa = 10** (i.e., the number of hypothetical DGMs used in the experiment.)
* **noc_oa = 10** (i.e., the number of true clusters on the overall dataset.)
* **cs_oa = 10** (i.e., the number of cluster members when the overall length of samples are clustered.)
* **ss_oa = 1000** (i.e., the number of observations for each generated series.)
* **nos_sub = 5** (i.e., the number of DGMs on the sub-samples of length 600.)
* **noc_sub = 5** (i.e., the number of clusters on the sub-samples of length 600.)
* **cs_sub = 20** (i.e., the number of cluster members when the sub-samples are considered for clustering.)
* **ss_sub = 600** (i.e., the number of observations for the sub-samples from the 1st time point.)
* **rn = 100** (i.e., the number of repetition of the experiment.)

1. Each DGM in Table 4.8 is used to generate **10** (i.e., **cs_oa**) number of time series with the length of **1000** (i.e., **ss_oa**).

$\xrightarrow{\text{ss\_sub: 1 to 600}}$

2. The slice of the dataset consisting of the first 600 time points is filtered out.

3. Clustering done on this slice by the proposed approach given in Section 3.4.2 and through implementation steps given in Figure 3.4, and **22** dissimilarity measures given in Table 4.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed over **5** × **20** × **600** dimensional dataset and clustering accuracies are saved.

4. Steps from **1** to **3** repeated **100** (i.e., **rn**) times to get average accuracies of **66** (i.e., **3** $*$ **22**) clustering schemes and the proposed approach.

5. The average accuracy indexes of all methods considered in the study are displayed for comparison in Table 4.9.

$\xrightarrow{\text{ss\_oa: 1 to 1000}}$

2. The dataset consisting of time series length of 1000 is taken for clustering.

3. Clustering done on the dataset by the proposed approach given in Section 3.4.2 and through implementation steps given in Figure 3.4, and **22** dissimilarity measures given in Table 4.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed over **10** × **10** × **1000** dimensional dataset and clustering accuracies are saved.

4. Steps from **1** to **2** repeated **100** (i.e., **rn**) times to get average accuracies of **66** (i.e., **3** $*$ **22**) clustering schemes and the proposed approach.

5. The average accuracy indexes of all methods considered in the study are displayed for comparison in Table 4.9.

Results for compared clustering methods are obtained using samples with a length of 600 and 1000 observations. However, 3 different sample sizes (i.e., $ss = 600$,

Table 4.9 Case 2 average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach.

	Clustering method					
	FUZZY		PAM		SPECTRAL	
	sample sizes (ss)		sample sizes (ss)		sample sizes (ss)	
Dissimilarity Measure	1 *to* 600 $(5\ cluster)$	1 *to* 1000 $(10\ cluster)$	1 *to* 600 $(5\ cluster)$	1 *to* 1000 $(10\ cluster)$	1 *to* 600 $(5\ cluster)$	1 *to* 1000 $(10\ cluster)$
Model Free						
D_{ACF}	0.920	0.614	0.857	0.595	0.874	0.585
D_{PACF}	0.663	0.552	0.917	0.608	0.891	0.589
D_{EUCL}	0.425	0.302	0.496	0.357	0.341	0.336
D_{CORT}	0.427	0.281	0.515	0.381	0.338	0.427
D_{COR}	0.435	0.280	0.465	0.380	0.579	0.320
D_{DWT}	0.574	0.434	0.629	0.434	0.519	0.416
D_{PER}	0.515	0.322	0.705	0.552	0.888	0.578
$D_{INT.PER}$	0.925	0.581	0.905	0.597	0.684	0.412
$D_{SPEC.LLR1}$	0.959	0.590	0.955	0.610	0.956	0.581
$D_{SPEC.LLR2}$	0.961	0.600	0.919	0.612	0.964	0.583
D_{GLK}	0.957	0.598	0.889	0.611	0.566	0.372
D_{ISD}	0.960	0.603	0.913	0.612	0.956	0.591
D_{SAX}	0.511	0.386	0.421	0.317	0.527	0.401
D_{VR}	0.904	0.592	0.834	0.590	0.461	0.339
D_{DTW}	0.467	0.289	0.678	0.517	0.558	0.291
Model Based						
D_{AR-LPC}	0.810	0.554	0.778	0.586	0.785	0.565
D_{AR-PIC}	0.956	0.631	0.948	0.603	0.906	0.584
D_{AR-MAH}	0.983	0.611	0.939	0.616	0.898	0.597
Complexity based						
D_{CID}	0.691	0.353	0.637	0.480	0.422	0.324
D_{PDC}	0.863	0.640	0.926	0.618	0.992	0.505
D_{CDM}	0.428	0.324	0.538	0.420	0.684	0.291
D_{NCD}	0.455	0.308	0.402	0.422	0.417	0.292

	Proposed Clustering Approach		
	TSMB-SPCL-UV		
	$ss \rightarrow 1$ *to* 600 $(5\ cluster)$	$ss \rightarrow 1$ *to* 800 $(10\ cluster)$	$ss \rightarrow 1$ *to* 1000 $(10\ cluster)$
L^2 (Euclidean) Dist.	0.991	0.659	0.810

$ss = 800$, $ss = 1000$) are considered to see the effects of the lengths of the series on the performance of the proposed approach. As can be seen from Table 4.9, some of the compared clustering methods performed well, especially when there is no cluster change in the dataset, that is, when the first 600 observations of the data set are considered for clustering. In the first considered part of the dataset, there are 5 clusters with 20 members each. Data generation mechanisms differ from each other only in terms of conditional mean processes. Thus, we can say that the clustering performances of almost half of the compared clustering methods are good and promising, particularly when the composition of the generating mechanisms of time series subject to clustering depends only on the conditional mean processes. Complexity-based dissimilarity measure PDC with Spectral Clustering, model-based dissimilarity measure AR-MAH with FUZZY or K-means clustering, dissimilarity measures based on time series spectrums such as SPEC.LLR, GLK, ISD with FUZZY, or K-means clustering can be considered the best methods among compared methods. However, when the set of time series desired to be clustered differ from each other in terms of heteroscedasticity, the clustering accuracy rates of the aforementioned approaches decrease considerably.

The clustering performance of the proposed approach on the first part of the dataset was ranked as the 2nd best. When the entire dataset was clustered, the accuracy indexes of the compared methods did not exceed 0.64, while the proposed clustering approach clustered the time series with 80% accuracy.

4.2 Performance Evaluation over Real-Life Data

This section assesses the clustering performances of the proposed clustering approach and the other state-of-the-art clustering methods on 2 datasets obtained from real-life time series classification experiments. These datasets are designed for time series classification studies, and accordingly, the class labels and sizes are known beforehand. Therefore, the datasets are used here to assess the ability of the proposed approach to identify underlying classes/clusters correctly.

In the literature, there are many time series classification and clustering studies regarding very different goals. Nonetheless, there are also publicly available many benchmarking aimed datasets constructed for time series classification studies [141]. Proposed methods in clustering studies can be different in terms of being suitable for the datasets desired to be clustered. Therefore, it is difficult to assume that any proposed clustering approach can be suitable for all time series datasets to be clustered. Although the proposed clustering approach in this study is not suitable for time series with a small number of observations (i.e., ≤ 300), it appears to be more useful in time series datasets where noise and heteroscedasticity are evident. In this respect, we consider that the proposed approach can provide useful outcomes in clustering datasets containing complex and volatile time series (i.e., > 300 obs.) from fields such as signal processing, finance, econometrics, and climate.

The number of classes and the data generating mechanisms that make up these classes in the real-life datasets we examined in this section are especially self-evident. Both datasets we used here consist of long signals obtained from processes with different dynamics. The first of these datasets consists of signals of 6 different hand movements observed by electromyogram, while the other dataset consists of EEG signals from 5 different neuronal activities observed by electroencephalogram. The reason we are used real-life data from the signal processing field is that it is almost infeasible to find real-life time series datasets with explicit class formations with labels in the areas such as economics, finance, and climate where information of underlying class formations is not exactly accessible due to uncertainty inherent to the dynamics of observations from these fields.

4.2.1 Clustering of sEMG Signals for Basic Hand Movements

This dataset consists of signals (i.e., time series) called surface Electromyography (sEMG) of 6 basic hand movements and was first designed and used for proposing an approach to classify hand movements in Sapsanis et al. [142]. The dataset is publicly available at the UCI Machine Learning Repository: `http://archive.ics.uci.edu/ml/datasets.php` [141]. The dataset named as **sEMG for Basic Hand Movements** on this repository.

Of the whole dataset, in this study, we used the observed signals in time domain of length 1000 from 1 healthy female subject who conducted the 6 different hand movements 30 times. Figure 4.6 shows 6 different hand movements. The signals were obtained at a sampling rate of 500 Hz (i.e., 500 obs. for a second) from 2-sensors placed on the subject's hand.

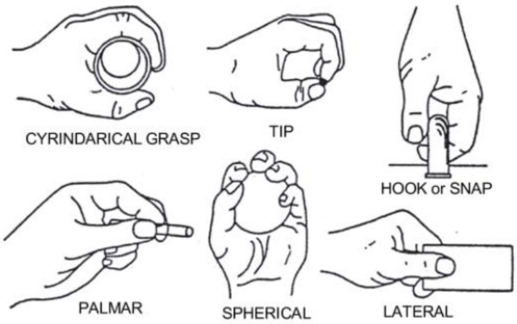


Figure 4.6 The six basic hand movements considered in the study. `https://archive.ics.uci.edu/ml/machine-learning-databases/00313/`

The subject performed and repeated each hand movement for 6 seconds of duration. We have used the 2-second length of signals (i.e., 2 × 500 Hz) in the middle of these 6-second length signals. Thus, we are picked the $6 \times 30 \times 1000 \times 2$ dimensional dataset to be clustered. Notice that each basic hand gesture can be considered as a distinct data generation mechanism when observed and measured. Hence, it can be stated that there are 6 different classes/categories in the dataset to be obtained from such an experiment. Accordingly, this section aims to evaluate the ability of grouping similar/dissimilar time series in this dataset, of which we know the true class formation, with the proposed approach and the other clustering methods.

Information about the dataset used in this subsection given in Table 4.10. This dataset consists of 6 classes and 30 samples measured through two channels for each cluster.

Table 4.10 sEMG Hand Movement dataset information used in the study.

6 Basic Hand Movements		**Number of Repetition**	**Sample Length**	
(i.e., Number of Sources or Number of Classes)			channel1 (sensor1)	channel2 (sensor2)
class1:	Cylindrical Grasp	30	1000	1000
class2:	Tip	30	1000	1000
class3:	Hook or Snap	30	1000	1000
class4:	Palmar Grasp	30	1000	1000
class5:	Spherical Grasp	30	1000	1000
class6:	Lateral Grasp	30	1000	1000

- The dataset's dimension is 6×30×1000×2

Sample of signals' traces are given in Figure 4.7. The plots in the left column show examples of signals (i.e., time series) obtained from the 1st sensor, while the plots in the right column show examples of signals obtained from the 2nd sensor.

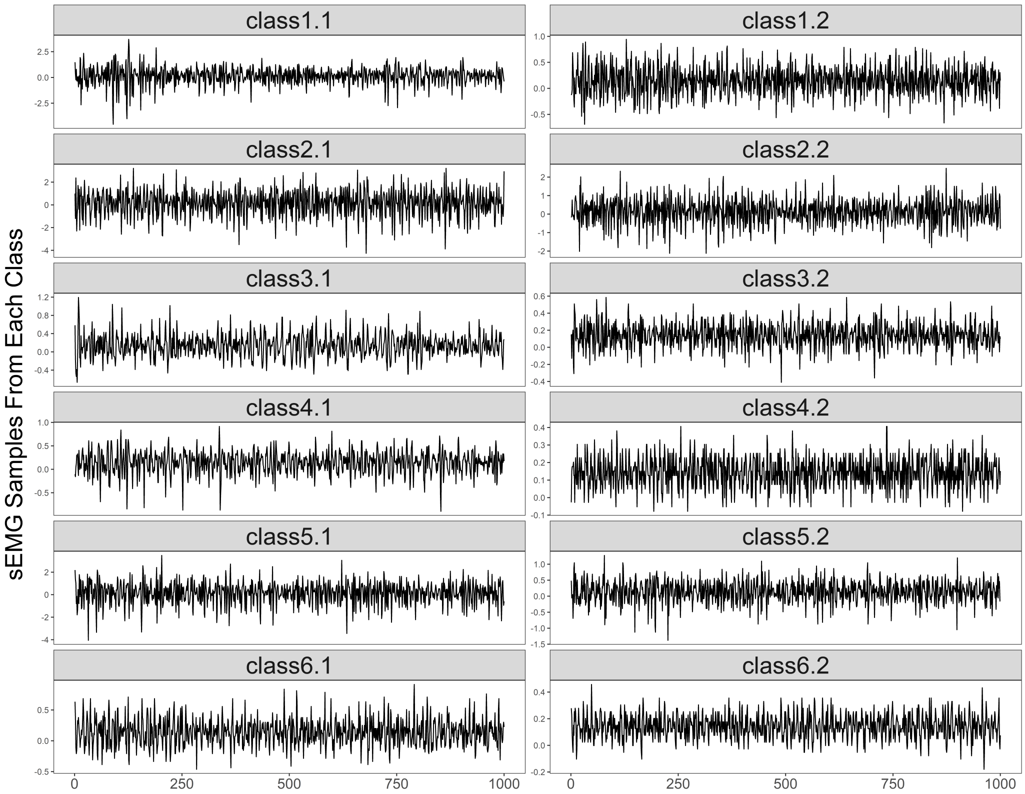


Figure 4.7 sEMG samples for basic hand movements.

Clustering of the dataset done by the proposed approach given in Section 3.4.2. Implementation steps are given in Figure 3.4 and hyper-parameter values are given in Figure 4.2. **22** dissimilarity measures given in Table 4.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed on this dataset and overall clustering accuracies are displayed in Table 4.11.

Table 4.11 sEMG clustering accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach.

Dissimilarity Measure	Clustering method		
	FUZZY	PAM	SPECTRAL
Model free			
D_{ACF}	0.516	0.693	0.779
D_{PACF}	0.497	0.665	0.855
D_{EUCL}	0.500	0.284	0.524
D_{CORT}	0.500	0.286	0.466
D_{COR}	0.282	0.283	0.323
D_{DWT}	0.500	0.288	0.464
D_{PER}	0.594	0.596	0.867
$D_{INT.PER}$	0.566	0.754	0.798
$D_{SPEC.LLR1}$	0.889	0.792	0.910
$D_{SPEC.LLR2}$	0.878	0.878	0.905
D_{GLK}	0.876	0.860	0.802
D_{ISD}	0.878	0.878	0.909
D_{SAX}	0.285	0.284	0.631
D_{VR}	0.823	0.806	0.703
D_{DTW}	0.498	0.483	0.631
Model based			
D_{AR-LPC}	0.285	0.608	0.715
D_{AR-MAH}	0.590	0.847	0.845
D_{AR-PIC}	0.528	0.564	0.621
Complexity based			
D_{CID}	0.565	0.646	0.330
D_{PDC}	0.428	0.726	0.709
D_{CDM}	0.525	0.773	0.831
D_{NCD}	0.500	0.766	0.836

	Proposed Clustering Approach
	TSMB-SPCL-UV
L^2 (Euclidean) Distance	0.910

As can be seen from Table 4.11, some of the dissimilarity measures provide very high accuracy results with 2 or all 3 of the clustering methods to cluster the sEMG dataset. Those that produce the most accurate results are the dissimilarity measures denoted by the abbreviations SPEC.LLR proposed by Kakizawa et al. [66], ISD proposed by Diaz and Vilar [69], GLK proposed by Fan and Zhang [68], AR-MAH proposed by Maharaj [77], and VR proposed by Bastos and Caiado [72]. Another prominent point in the results displayed in Table 4.11 is that higher accuracies are obtained when the distance matrices computed by usage of some dissimilarity measures (i.e., PACF, PER, AR-LPC, CDM, NCD) are employed with the spectral clustering method given in Section 3.3.

In the experiment of sEMG dataset clustering, the results showed that the proposed clustering approach had been one of the best approaches that can assign signals (i.e., time series) to the correct groups with the highest accuracy. With the proposed clustering approach, approximately 91% of the series in the sEMG dataset are correctly clustered. As we mentioned in the previous subsections, if the series to be clustered differ from each other only in terms of their conditional mean processes, the proposed approach and some other dissimilarity measures (i.e., SPEC.LLR, GLK, AR-MAH, etc.) can make accurate clustering with very high performance.

Determining the optimum (i.e., exact or true) number of clusters in a dataset is another problematic issue related to clustering when there is no information. Although our main purpose in this study is not to propose a method for determining the number of clusters in the dataset, we wanted to show and evaluate the GAP statistics' potential about finding cluster numbers in a dataset. Hence, it is necessary to note the success of this statistic in some cases, which might serve as a reference for further studies.

There are several alternative methods in literature can be considered to find the optimum number of groups in the data (for example, see [143], [144], [145]). However, the other most generally known tool is determining plausible cluster number through looking at the average silhouette value. The silhouette value takes a value between -1 and +1 and is used to measure the fitness of objects to their assigned cluster and firstly proposed by Rousseeuw [146]. Please see Kaufman and Rousseeuw [147] for more detailed information about the silhouette method.

The GAP statistic we practiced for the real-life data in this subsection reached the peak value at 6, which is the true number of clusters in the data. Figure 4.8 displays the GAP statistics vs. the possible number of clusters. Likewise, as shown in Figure 4.9, the average silhouette value also reaches its highest value at 6.

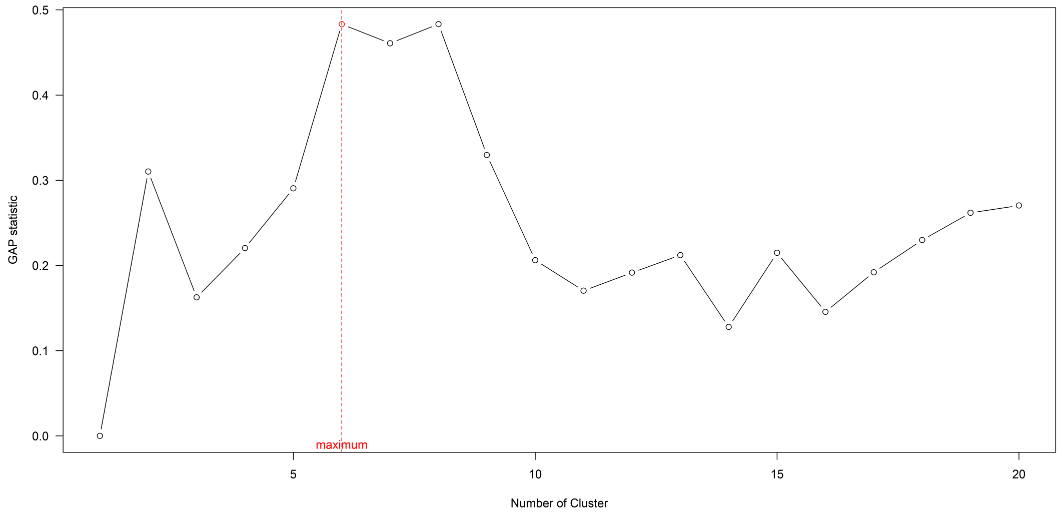


Figure 4.8 GAP statistic per cluster number for the sEMG dataset using TSMB-SPCL-UV.

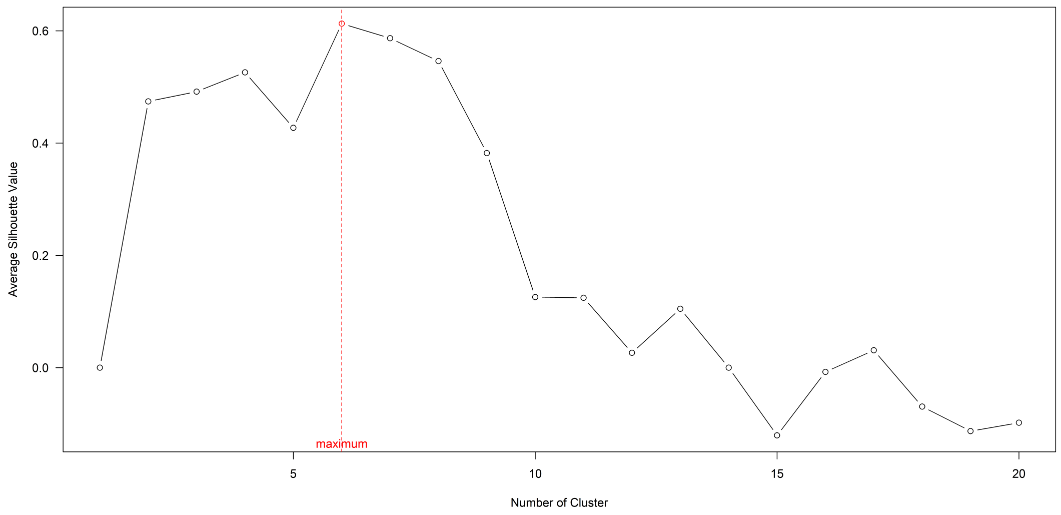


Figure 4.9 Average Silhouette Value per cluster number for the sEMG dataset using TSMB-SPCL-UV.

4.2.2 Clustering of EEG Signals

The other dataset we discussed in Chapter 4 for real-life data application consists of EEG time series. This dataset is one of the well-known and publicly available data for classification purposes. We used this dataset for testing the clustering performance of the proposed clustering approach. It is known as the University of Bonn EEG dataset and was firstly obtained and employed by Andrzejak et al. [148]. The data can be downloaded from `http://epileptologie-bonn.de/cms/upload/workgroup/lehnertz/eegdata.html` (last accessed 6/30/2021).

The dataset contains 5 classes/categories. Each class includes 100 EEG recordings with a length of 4097 observations. The signals were recorded at a sampling rate of 173.61 Hz (i.e., approx. 173 obs. for a second) from electrodes placed on the subjects' scalp's surface and hippocampus (i.e., intra-cranial). Please note that the observations in the first two classes are taken from the scalp surface, and signal samples in the last 3 classes are intra-cranial. The total observation time for each recording is 23.6 seconds. Thus, the dimension of the dataset is 5 (classes)×100 (samples for each class)×4097 (length of each sample observed during 23.6 seconds). The description of the dataset is summarized in Table 4.12 .

Table 4.12 EEG dataset information used in the study.

5 Different Neuronal States (i.e., Number of Classes)		**Number of Samples**	**Sample Length**
class1:	EEG signals from healthy subjects Recordings during Eyes Open (EO)	100	4097
class2:	EEG signals from healthy subjects Recordings during Eyes Closed (EC)	100	4097
class3:	EEG signals from epileptic patients Seizure free recordings at non-epileptogenic zone	100	4097
class4:	EEG signals from epileptic patients Seizure free recordings at epileptogenic zone	100	4097
class5:	EEG signals from epileptic patients Seizure activity recordings at epileptogenic zone	100	4097

- The dataset's dimension is 5×100×4097

We assumed each neuronal state in this experiment is a different data generation mechanism than the others. Each produces different signals as they have different dynamics and should be considered different categories since they evidently resulted in different physical states. Therefore, it should be noted that we theoretically accept the actual number of clusters in the entire dataset as 5.

Time series graphs of a sample EEG from each class are shown in Figure 4.10.

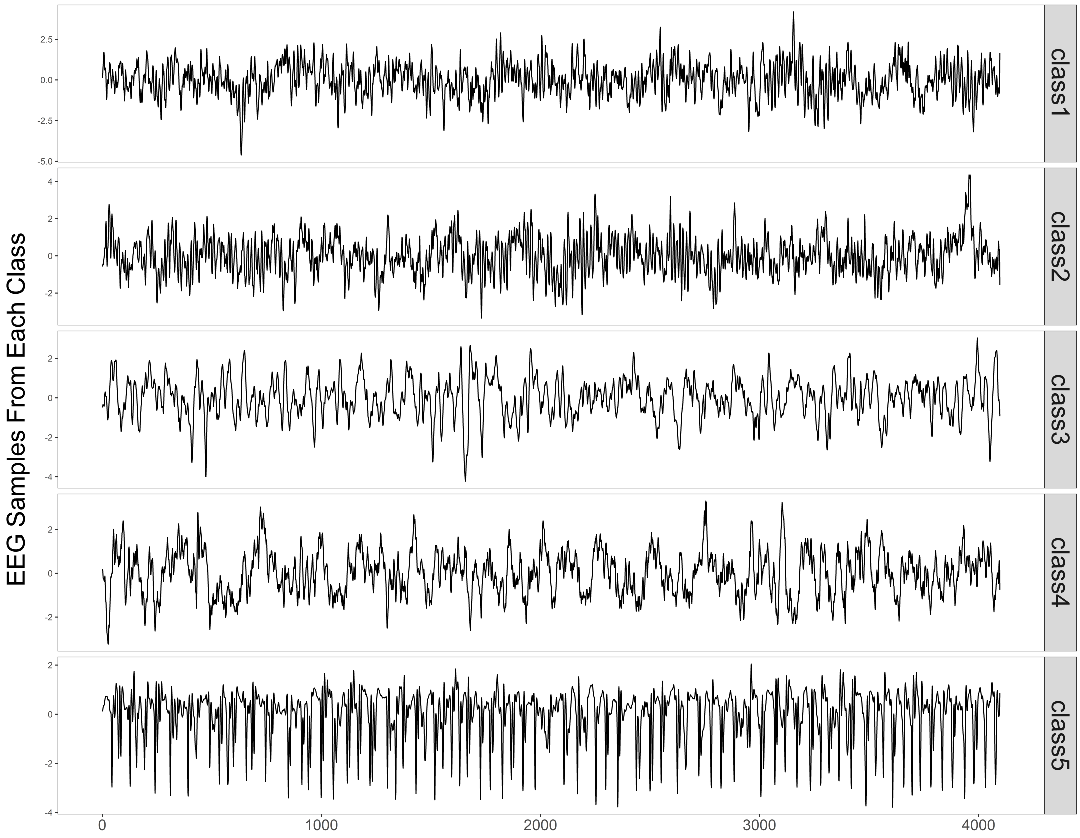


Figure 4.10 EEG samples.

There are many studies on the nonlinear nature of EEG signals (for example, see [149–151]). In this context, it can be stated that the time series in this dataset has more complexity than the dataset examined in the previous subsection, which makes clustering without any prior knowledge is more challenging in this case.

The complexity and possible heteroscedasticity in this dataset is also reflected in the clustering performances, and the clustering accuracies of many methods remained below 60%.

Clustering of the dataset done by the proposed approach given in Section 3.4.2. Implementation steps are given in Figure 3.4 and hyper-parameter values are given in Figure 4.2. **22** dissimilarity measures given in Table 4.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed on this dataset and overall clustering accuracies are displayed in Table 4.13.

Table 4.13 EEG clustering accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach.

Dissimilarity Measure	Clustering method		
	FUZZY	PAM	SPECTRAL
Model free			
D_{ACF}	0.477	0.554	0.491
D_{PACF}	0.419	0.608	0.595
D_{EUCL}	0.333	0.282	0.332
D_{CORT}	0.333	0.268	0.332
D_{COR}	0.333	0.282	0.307
D_{DWT}	0.333	0.316	0.332
D_{PER}	0.332	0.438	0.515
$D_{INT.PER}$	0.523	0.607	0.325
$D_{SPEC.LLR1}$	0.522	0.511	0.656
$D_{SPEC.LLR2}$	0.525	0.530	0.618
D_{GLK}	0.485	0.505	0.320
D_{ISD}	0.527	0.551	0.539
D_{SAX}	0.404	0.314	0.270
D_{VR}	0.492	0.492	0.331
D_{DTW}	0.495	0.541	0.809
Model based			
D_{AR-LPC}	0.441	0.443	0.546
D_{AR-MAH}	0.673	0.693	0.706
D_{AR-PIC}	0.484	0.583	0.589
Complexity based			
D_{CID}	0.496	0.484	0.332
D_{PDC}	0.551	0.535	0.572
D_{CDM}	0.454	0.355	0.515
D_{NCD}	0.333	0.438	0.533

	Proposed Clustering Approach
	TSMB-SPCL-UV
L^2 (Euclidean) Distance	0.790

As shown in Table 4.13, dissimilarity measure performances of clustering methods are decreased as the complexity or nonlinearity in the dataset is more evident. We consider that this is compatible with the results obtained from the experiments with synthetic datasets.

Results also show that almost all of the compared methods in this real-data clustering task produced poor accuracies. However, the performance of the dynamic time warping (DTW) based dissimilarity measure employed with spectral clustering is considerably remarkable. Clustering with the highest accuracy index (i.e., 0.81) among all methods is obtained with this dissimilarity measure. Therefore, this performance in favor of spectral clustering and DTW should be taken into account. Thus, it can be said that the fact that DTW did not produce remarkable results in other datasets is not a failure, but because the purpose and appropriate dataset match have not been achieved. It should be noted that this is also a valid argument for other methods and dissimilarity measures. This result is a finding in favor of one of the main arguments we attempted to emphasize throughout the study. Ascertaining the appropriate method for the dataset to be clustered and structuring the pre-analysis processes will improve the clustering performance.

The proposed clustering approach in this study provided the second best clustering result with an accuracy index of 0.79, which implies that 79% of 500 signals are accurately clustered. However, classification accuracies can be considerably high (i.e., almost 100%) in classification studies related to this dataset (for example, see [152]). In the experiments we conducted in this study, no pre-learning is carried out using the part of the dataset reserved for training, as is generally made in classification studies. In this sense, clustering brings more risks than classification. This contrast stems from the purposes of both approaches. To increase the clustering performance of any proposed method for datasets with high complexity, it would be appropriate to conduct further experiments and comparisons.

This complexity, which we assumed defines the dataset, is also affected the GAP statistics and average silhouette values we calculated. As can be seen in Figures 4.11 and 4.12, both measures could not peak at the correct cluster number of 5.

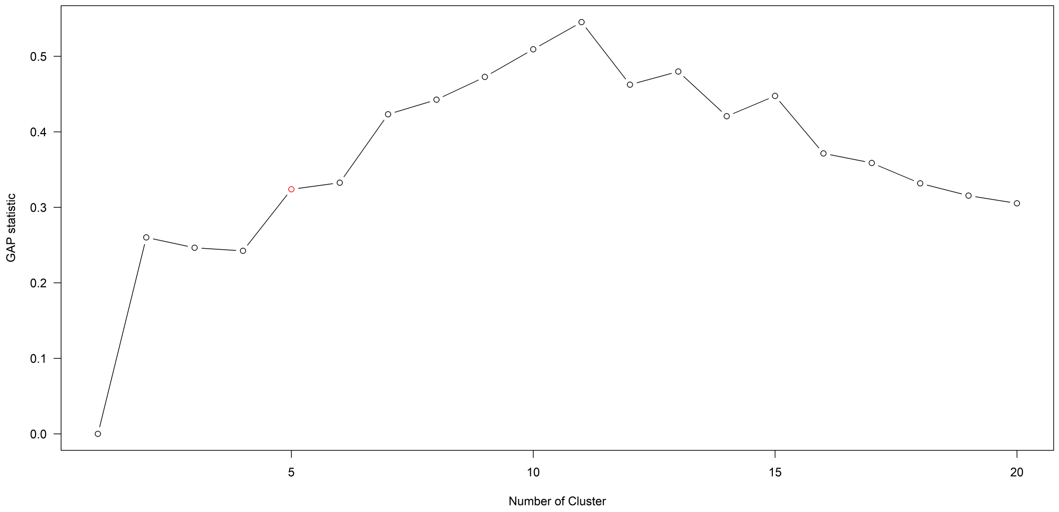


Figure 4.11 GAP statistic per cluster number for the EEG dataset using TSMB-SPCL-UV.

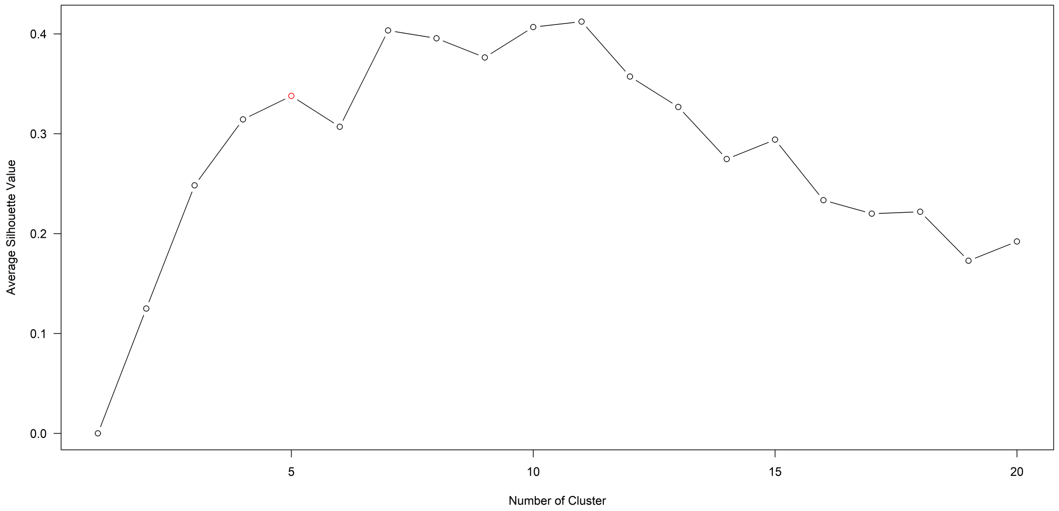


Figure 4.12 Average Silhouette Value per cluster number for the EEG dataset using TSMB-SPCL-UV.

CHAPTER 5

EXPERIMENTAL RESULTS: MULTIVARIATE TIME SERIES CLUSTERING

The primary motivation of the study is to propose a clustering approach for multivariate time series datasets. For this purpose, this chapter is devoted to evaluate the proposed approach for clustering multivariate time series. While clustering a univariate time series dataset, we proposed the approach in which each time series are represented with an assigned feature vector. Thus, instead of using raw data, the proposed approach distinguishes the time series by measuring the distances of the feature vectors from each other. The proposed approach has a very similar procedure in clustering both univariate and multivariate time series datasets. Like in the clustering of a univariate time series dataset, each multivariate time series is considered as an object to be clustered. However, unlike the univariate case, the objects we cluster here has more than one variable. So, in this case, each multivariate time series has to be represented by a feature matrix where the column dimension of a matrix corresponds to the variable dimension of the object to be clustered. Consequently, the proposed approach employs the distances of the feature matrices from each other in a multivariate case. In other words, the proposed approach differs only in the affinity matrix calculation if the data to be clustered are univariate or multivariate. The rest of the proposed clustering procedure is identical. The details of the proposed approach for both univariate and multivariate time series clustering are addressed in Section 3.4. Besides, the statements, explanations and findings given in Chapter 4 on the implementation of the proposed approach are also largely valid for this section. Therefore, it would be useful to consider the explanations and findings expressed in Chapter 4 while examining Chapter 5.

In this chapter, a total of 15 different clustering plans comprised of the use of 5 different dissimilarity measures on 3 different clustering methods (i.e., Fuzzy c-means, Partitioning Around Medoids (PAM), Spectral) are compared with the proposed time series clustering approach. The state-of-the-art distance measures proposed by literature are presented in Table 5.1. The dissimilarity measures used in the comparison study have been discussed in [137], [66], [56], [84], [153], [83] and some of them have been made available in the **R** packages named as **TSclust** [55], **dtwclust** [57] and **pdc** [56].

Table 5.1 Dissimilarity measures used for comparison in multivariate clustering.

Dissimilarity Measure	Description	Reference
	Model free approaches	
D_{JD}	Multivariate Spectrum based Distance. Also known as symmetrized Kullback-Leibler (KL) Divergence. Also known as J(effreys)-Divergence.	[66]
D_{LD}	Multivariate Spectrum based Distance. Known as Log-Spectral Divergence.	[84]
D_{GAK}	Distance based on Global Alignment Kernels	[83]
D_{DTW}	Dynamic Time Warping Distance.	[59, 73]
D_{PDC}	Permutation Distribution Distance.	[56]

The distance measures listed for comparison purposes in Chapter 5 have been developed especially for clustering multivariate time series. Most of the distance measures we used for comparison in the previous Chapter 4 listed in Table 4.1 are not suitable for multivariate cases. They require investigation whether to be used for multivariate clustering. However, we can assume that the distance measures, which appeared not to show high performance when clustering artificially generated univariate time series datasets in the previous Chapter, would not also help to provide remarkable accuracies for the multivariate clustering. Because the artificial datasets we purposed to cluster in both chapters are generated in a heteroscedastic composition, and it is evident that these distance measures are not developed for clustering of heteroscedastic or nonlinear time series. For example, the J-Divergence and LD we used in this chapter are the multivariate equivalent of distance measures developed for clustering time series by comparing their spectrum estimates (i.e., clustering time series on the frequency domain).

Similarly, we see that DTW and PDC are used in both univariate and multivariate clustering. Of these, DTW makes a shape-based distance measurement between series, while PDC makes a complexity-based distance measurement. The GAK distance measure, similar to DTW, which we have included in this chapter, tries to distinguish the series from each other based on their shape (i.e., traces, trajectories, patterns) properties. We should state here that these distance measures could provide very successful results for clustering of speech signals [154]. Such distance measures are distance measures developed and useful for clustering of time series where randomness is negligible but shape-based features are prominent.

On the Implementation of the Proposed Approach in Multivariate Case

In Section 3.4.3, while representing the proposed approach, we stated that each multivariate time series subject to clustering should be represented by a feature matrix. The columns of the feature matrices consist of feature vectors corresponding to each time series variable in a multivariate time series object. In other words, the only difference here in multivariate case is to obtain a feature matrix in which the vectors corresponding to each time series in the multivariate time series are placed in the columns of this matrix. And, the elements of the feature vectors consist of 11 parts as shown in Table 3.1. Predetermined values of hyperparameters are given in Table 4.2. The implementation steps of the proposed approach are displayed in Figure 3.4. In this regard, the procedure we follow for clustering of a set of multivariate time series is quite similar to the clustering of a group of univariate time series.

The procedure we propose for feature matrix extraction in multivariate cases may be questioned in a way that why we consider each series one by one instead of estimating the properties of the series with the use of a multivariate model and obtaining the feature matrix directly. As we have mentioned before, our principal aim here is not to perform the modeling process most accurately. Instead, it is more important to examine that the DTGARCH and AR model estimations we use for feature extractions serve our purpose. However, the feature matrix in our proposed clustering procedure can also be obtained using multivariate model estimations (i.e., multivariate type GARCH models, etc.). We should note that we leave such approaches to develop through further researches.

5.1 Performance evaluation over Synthetic Data

As in the univariate case, we conducted some experiments on the artificially generated multivariate time-series datasets.

In each experiment in this section, the standard procedure of generating multivariate time series, clusters, and then assessment of clustering accuracies on the synthetic dataset can be given as follows:

* **nos** $\Leftarrow the\ number\ of\ sources$: Defines the number of hypothetical DGMs, nos, be used in the experiment.

* **noc** $\Leftarrow the\ number\ of\ clusters$: The number of hypothetical DGMs, nos, also defines the number of true clusters, noc (i.e., $nos = noc$).

* **ss** $\Leftarrow sample\ size$: Define the time span, length or number of observations, ss, for each generated series.

* **cs** $\Leftarrow cluster\ size$: Define the cluster size, cs, which determines the number of series/members in each cluster.

* **rn** $\Leftarrow repetition\ number$: Defines the number of repetitions of the experiment.

After deciding values for ***nos***, ***noc***, ***ss***, ***cs*** and ***rn***, the synthetic dataset generation, clustering, and comparison phases of an experiment can be summarized in 4 steps:

1. Each DGM used to generate ***cs*** number of time series with the length of ***ss*** to form a corresponding cluster.
2. Clustering by proposed approach and other considered clustering methods for comparison are performed over ***nos*** $\times$ ***cs*** $\times$ ***ss*** dimensional dataset and clustering accuracies are saved.
3. Steps from **1** to **2** repeated **rn** times to get average accuracy indexes of clustering methods and the proposed approach.
4. The average accuracy indexes of all methods are compared.

Synthetic multivariate datasets are created by hypothetical DGMs. The collection of multivariate time series generated from each DGM also forms a cluster subject to clustering experiments for performance evaluation of the proposed clustering approach. Each DGM is composed of specific multivariate time series models that we name the body components as the STEMs. In some scenarios, a STEM with an error part may directly form a DGM, and in some situations, combinations of STEMs form a DGM. Consequently, under each subsection, the STEMs and then the DGMs consisting of these STEMs are presented, and the performances of the proposed approach for determining the clusters corresponding to these DGMs are provided.

The multivariate time series datasets that we are considering here in some cases are generated to have different cluster formations for different subsequent periods within the overall length of samples. Thus, we assessed the performance of the proposed approach in determining time-dependent cluster changes and suggested the minimum sample length required to capture these changes.

In this section, a total of 9 distinct scenarios are considered. 2 of them are based on non-temporal clusters (i.e., clusters are not changing over time). 7 of them are based on temporal clusters (i.e., clusters are changing over time). However, for readability and ease of follow-up, the results of only 6 of these experiments are given in subsections, while the remaining experiments are given in Appendix B.

5.1.1 First Set of Experiments on Multivariate Time-series Clustering

In the first experiment, to evaluate the accuracy performance of the proposed clustering approach for a set of multiple multivariate time series, we employed a 3-dimensional Stochastic Differential Equations (SDE) for multivariate time series data generation, which is not one of the parametric time series models such as Vector Autoregressive (VAR) models or Vector Error Correction Models (VECM).

A general form of a 1-Dimensional SDE can be given as follow:

$dX_t = f\left(X_t, t\right)\ dt + g\left(X_t, t\right)\ dW_t,$

where the functions $f(.)$ and $g(.)$ are called as the coefficients of drift and diffusion, respectively. And, W referred as standard Wiener process (i.e., Brownian motion).

The drift term defines the instantaneous expectation of the process X_t at time t while the diffusion term establishes the instantaneous volatility (i.e., standard deviation) of the process X_t at time t. For further details about SDEs please see [155–157].

The DGMs used to generate the artificial datasets clustered in this section are composed of STEMs given in Table 5.2. The drift terms of the SDEs used in this section as the data generation mechanism are selected from dynamical systems, also known as systems of differential equations. For example, the drift term of the STEM_1 is an example of a biological model known as a biomass transfer model [158]. The drift term of the STEM_2 is an example of virus dynamics modeling in living organisms known as a Nowak-May HIV model [159], and the drift term of the STEM_3 is an example for electrical circuit model through differential equations [160]. The diffusion terms determining the variance component of STEM_2 and STEM_3 are constant coefficients, while the variance component of STEM_1 is not given in a specific structure. This is because, in the following subsection, we intended to generate a dataset with the same drift process but according to different diffusion (variance) components. Hence, it is possible to evaluate the sensitivity of the proposed clustering approach to minor/significant perturbations in the data generation mechanisms. Please note that, the Wiener processes' diffusion coefficient, $C_{...}$, in STEM_1 can be described by any constant, function, process or distribution.

Table 5.2 Utilized specifications of SDEs in the first set of experiments.

$$\textbf{STEM\_1}: \begin{cases} dX_t &= (-X_t + 3\,Y_t)\ dt + C_{...}\ dW_{1,t} \\ dY_t &= (-3\,Y_t + 5\,Z_t)\ dt + C_{...}\ dW_{2,t} \\ dZ_t &= (-5\,Z_t)\,dt + C_{...}\ dW_{3,t} \end{cases}$$

$$\textbf{STEM\_2}: \begin{cases} dX_t &= (15 - 0.01\,X_t - 1\,X_t\,Z_t)\,dt + 4.5\,dW_{1,t} \\ dY_t &= (1\,X_t\,Z_t - 1\,Y_t)\,dt + 4.5\,dW_{2,t} \\ dZ_t &= (6\,Y_t - 12\,Z_t)\,dt + 4.5\,dW_{3,t} \end{cases}$$

$$\textbf{STEM\_3}: \begin{cases} dX_t &= \left(-\left(\tfrac{50}{70}\right)\,Y_t - \left(\tfrac{45}{70}\right)\,Z_t\right)\ dt + 1\,dW_{1,t} \\ dY_t &= \left(-\left(\tfrac{50}{70} + \tfrac{50}{75}\right)\,Y_t + \left(\tfrac{35}{70} - \tfrac{45}{75}\right)\,Z_t\right)\ dt + 1\,dW_{2,t} \\ dZ_t &= \left(\left(\tfrac{35}{55} - \tfrac{50}{70}\right)\,Y_t - \left(\tfrac{35}{55} + \tfrac{45}{70} + \tfrac{45}{55}\right)\,Z_t\right)\ dt + 1\,dW_{3,t} \end{cases}$$

$W_{1,t},\ W_{2,t},\ W_{3,t}$ are standard Wiener processes (Brownian Motions).

Please note that the factor of the Wiener processes, $C_{...}$, in **STEM_1** is a coefficient that can be described by any constant, function, process or distribution.

5.1.1.1 Case 1 : Non-temporal 3 Cluster

The artificial multivariate time series datasets to be clustered in this subsection is generated using 3 different DGMs shown in Table 5.3. These DGMs in this subsection mainly consist of STEM_1 in Table 5.2. However, they differ in terms of specific diffusion terms of each STEM's variance component. For example, the variance of DGM01 is obtained by multiplying the constant variance of the Wiener process by 1.75, and this variance is constant along time. Therefore, in this particular data generation mechanism, the Wiener process performs like a normally distributed white noise process. On the other hand, the diffusion coefficient, C_2, of DGM02 is distributed as GARCH(1,1) with specific parameter values given in Table 5.3 and unlike the other data generation mechanisms in this section, the DGM02 causes a time-varying variance component in the data it generates. Similarly, the diffusion term is distributed as the ARCH(1) process in DGM03 with specific parameter values given in Table 5.3. It should be noted that the processes determining the diffusion coefficients of DGM02 and DGM03 are very close to each other, which makes it very difficult to distinguish the multivariate time series data generated from these mechanisms.

Table 5.3 Hypothetical DGMs used in Case 1.

$$\mathbf{DGM01} := \mathbf{STEM\_1} \mid C_1$$

$$\mathbf{DGM02} := \mathbf{STEM\_1} \mid C_2$$

$$\mathbf{DGM03} := \mathbf{STEM\_1} \mid C_3$$

$C_1 = 1.75$

$C_2 \sim GARCH(1,1)|h_t = 0.15 + 0.25\varepsilon_{t-1}^2 + 0.45h_{t-1}$ where $\varepsilon_t = \sqrt{h_t}e_t,$ and $e_t \overset{i.i.d}{\sim} D_{norm}(0,1)$

$C_3 \sim GARCH(1,0)|h_t = 0.15 + 0.65\varepsilon_{t-1}^2$ where $\varepsilon_t = \sqrt{h_t}e_t,$ and $e_t \overset{i.i.d}{\sim} D_{norm}(0,1)$

Although the bodies of hypothetical DGMs share mainly the same structure, these DGMs differ in the terms that define the variance component. Besides, since we used 3 DGMs to generate multivariate time series data in this experiment's Case 1, the actual number of clusters in the data to be created should also be considered as 3. In this regard, it would not be wrong to assume that the hypothetical DGMs we use are at the center of each cluster. The traces/paths (i.e., time series plots) of the multivariate time-series generated from each DGM (i.e., class) is illustrated in Figure 5.1.

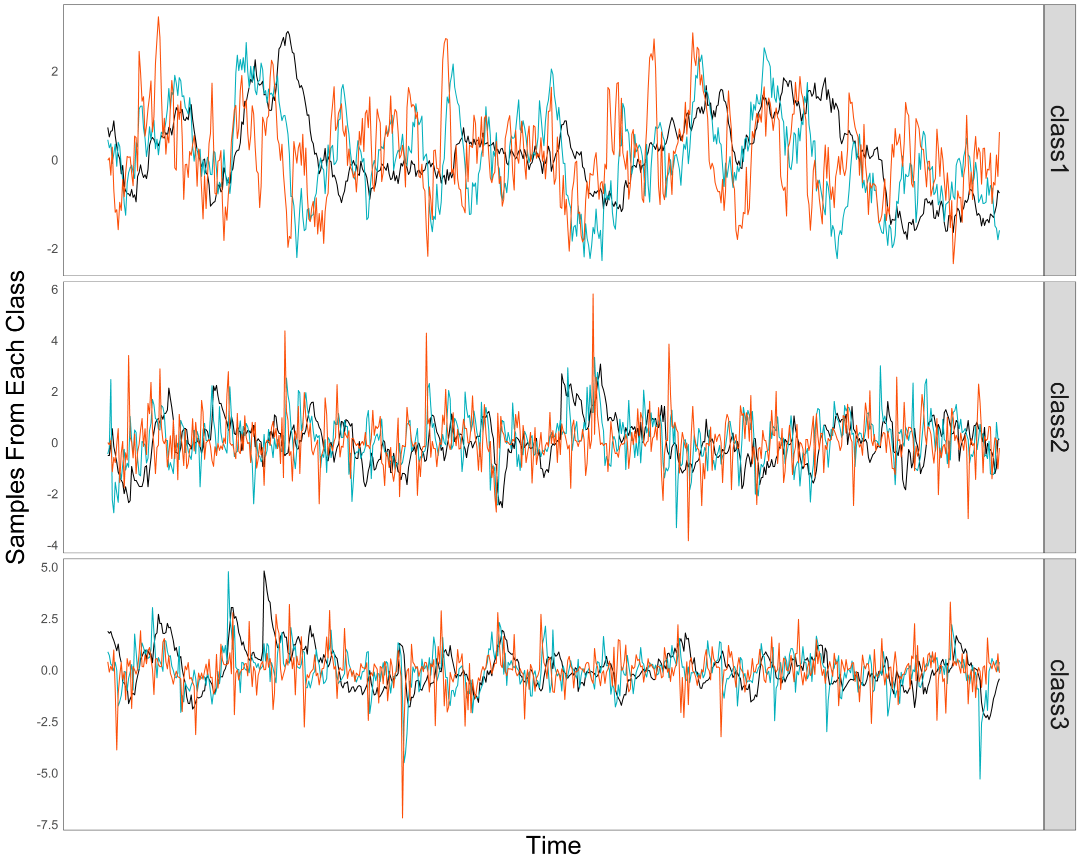


Figure 5.1 Simulation samples from 3 DGMs listed in Table 5.3

Each DGM given in Table 5.3 is used to generate a 3-variable multivariate time series, and the linear correlations of these variables with each other are pretty low. Since it is desired to evaluate the clustering capability of the proposed approach under low correlated series, the simulations are designed to generate uncorrelated multivariate time series in this case. The sample lengths of the multivariate time series created in this section are determined as 600. 30 samples are generated from each DGM. Therefore, for example, the derived data set from 1 repetition of the data generation step consists of 3 clusters containing 30 objects (i.e., multivariate time series).

The sample graphs of each DGM given in Figure 5.1 are exhibited that there are no formal or shape-based features that exist to distinguish multivariate time series from each other.

Path simulations or data generation from SDEs given in this section are performed by the use of **R** package **`Sim.DiffProc`** [155].

The generating samples, clusters, and then evaluations of clustering performances for Case 1 can be summarized as follows:

* **nos** **= 3** (i.e., the number of hypothetical DGMs used in the experiment.)

* **noc** **= 3** (i.e., the number of true clusters.)

* **cs** **= 30** (i.e., the number of series/members in each cluster.)

* **ss** **= 600** (i.e., the number of observations for each generated multivariate series.)

* **rn** **= 100** (i.e., the number of repetition of the experiment.)

1. Each DGM in Table 5.3 is used to generate **30** (i.e., **cs**) number of 3-dimensional multivariate time series with the length of **600** (i.e., **ss**) to form a corresponding cluster.
2. Clustering done by the proposed approach given in Section 3.4 and through implementation steps given in Figure 3.4, and **5** dissimilarity measures given in Table 5.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed over **3** $\times$ **30** $\times$ **600** dimensional dataset and clustering accuracies are saved.
3. Steps from **1** to **2** repeated **100** (i.e., **rn**) times to get average accuracies of **15** (i.e., **3** $*$ **5**) clustering schemes and the proposed approach.
4. The average accuracy indexes of all methods considered in the study are displayed for comparison in Table 5.4.

Table 5.4 Multivariate Experiment 1 Case 1: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach.

Dissimilarity measure	Clustering method		
	FUZZY	PAM	SPECTRAL
Model free			
D_{JD}	0.719	0.716	0.717
D_{LD}	0.711	0.483	0.687
D_{GAK}	0.469	0.426	0.424
D_{DTW}	0.592	0.534	0.551
D_{PDC}	0.769	0.547	0.711

	Proposed Clustering Approach		
	TSMB-SPCL-MV		
	$ss = 350$	$ss = 475$	$ss = 600$
L^2 (Euclidean) Distance	0.738	0.747	0.756

In the artificial datasets we examined in this section, each cluster has 30 members, and the actual number of clusters is 3. The success rates of the proposed approach over three different sample sizes and the other distance measures' accuracy results are given in Table 5.4. The accuracy rates are obtained as a result of 100 repetitions of dataset generation and clustering. The fact that the series in the 2nd and 3rd clusters are obtained from data generation mechanisms that are very close to each other in structure complicates the clustering task. The proposed approach yielded clustering performance similar to clustering based on PDC, JD, and LD distance measurements in this experiment. The highest clustering accuracy is obtained by the PDC distance measure based fuzzy clustering. The proposed approach provided the 2nd best result, with very close to the highest accuracy.

However, in this case, the fuzzy clustering algorithm based on each distance measure produced unstable outcomes. According to the fuzzy algorithm, an object to be clustered should be assigned to a cluster by specific probabilities. That is, each object has a probability of belonging to the cluster to which it is assigned. But, in the scope of this experiment, the objects to be distributed into 3 clusters are always assigned to a cluster with a probability of 1/3 as if the events (i.e., clustering of objects) are equally likely. Thanks to the **R** package **TSclust** [55] and the `fanny` function for fuzzy clustering, this problem could be tried to overcome by changing/decreasing the hyperparameter value called the degree of fuzziness used in the fuzzy clustering algorithm. Nevertheless, the changes made in desired fuzziness degree of cluster memberships (i.e., please see [161] for an assessment on one of the hyper-parameters in fuzzy c-means clustering) are not enough to make the fuzzy clustering algorithm stable and reliable. For example, when the fuzziness degree value desired for fuzzy C-means clustering is decreased from 2 to 1 by 0.1 step increments, the clustering results obtained for PDC distance vary between 0.67 and 0.77. For instance, if the degree of fuzziness is decided as 1.7, the clustering accuracy rate over 100 repetitions is obtained as 0.778, if the degree of fuzziness is decided as 1.6, the clustering accuracy rate over 100 repetitions is obtained as 0.679, if the degree of fuzziness is decided as 1.5, the clustering accuracy rate over 100 repetitions is obtained as 0.716. Similar results are valid for other distance measures. Hence, in order to make the comparisons appropriately, the hyper-parameter value has been fixed to 1.8.

Fortunately, spectral clustering, which constitutes one of the main axis of the proposed clustering approach, provides more stable results regardless of the degree of fuzziness to be determined for fuzzy clustering. It should also be noted that although we use fuzzy clustering in the final step of spectral clustering (please refer Section 3.3), it is not dramatically affected by the degree of fuzziness chosen for clustering. Accordingly, the spectral clustering results for the distance measures given in the last column of Table 5.4 can be considered as more robustly obtained accuracy rates. The results of this experiment also support our findings that the series lengths should be as long as possible for the proposed clustering approach, and sample sizes should not be less than 400 observations to produce clustering results most efficiently.

5.1.1.2 Case 2 : Temporal 3 Cluster

This section addresses one of the experiments and its results on evaluating the time-varying clustering performance of the proposed approach and the compared clustering methods. In this section, we employed the hypothetical data generation mechanisms given in Table 5.5 to generate the multivariate time series datasets that are to be clustered. These DGMs consist of the bodies presented in Table 5.2 and their combinations.

Table 5.5 Hypothetical DGMs used in Case 2.

	used to generate time series of length 1200		
	sources of samples from 1 to 600		sources of samples from 601 to 1200
DGM01 :=	**STEM_1** $\mid C_1$	▷	**STEM_1** $\mid C_1$
DGM02 :=	**STEM_1** $\mid C_2$	▶	**STEM_2**
DGM03 :=	**STEM_1** $\mid C_3$	▶	**STEM_3**

$'\triangleright'$ indicates where the data generation mechanisms remain unchanged.

$'\blacktriangleright'$ indicates where the data generating mechanisms are changed after 600th observation.

$C_1 = 1.75$

$C_2 \sim GARCH(1,1)|h_t = 0.15 + 0.25\varepsilon_{t-1}^2 + 0.45h_{t-1}$ where $\varepsilon_t = \sqrt{h_t}e_t,$ and $e_t \overset{i.i.d}{\sim} D_{norm}(0,1)$

$C_3 \sim GARCH(1,0)|h_t = 0.15 + 0.65\varepsilon_{t-1}^2$ where $\varepsilon_t = \sqrt{h_t}e_t,$ and $e_t \overset{i.i.d}{\sim} D_{norm}(0,1)$

As can be seen in Table 5.5, while the components of DGM01 do not change over time, DGM02 and DGM03 have time-varying components. However, all three DGMs have a data generation component with the same structure over a certain period (i.e., for the first 600-length observations). It should also be noted that the diffusion terms of the SDE components that we name STEM_2 and STEM_3 are determined by constant coefficients. In other words, the variances of these STEMs are determined by the normally distributed white noise processes (i.e., standard Wiener processes).

The time series plots of sample multivariate time series generated from each DGM are exhibited in Figure 5.2. Since such visualizations comprehend visual information about the dataset we are examining for clustering, we regard it significant to present such plots.

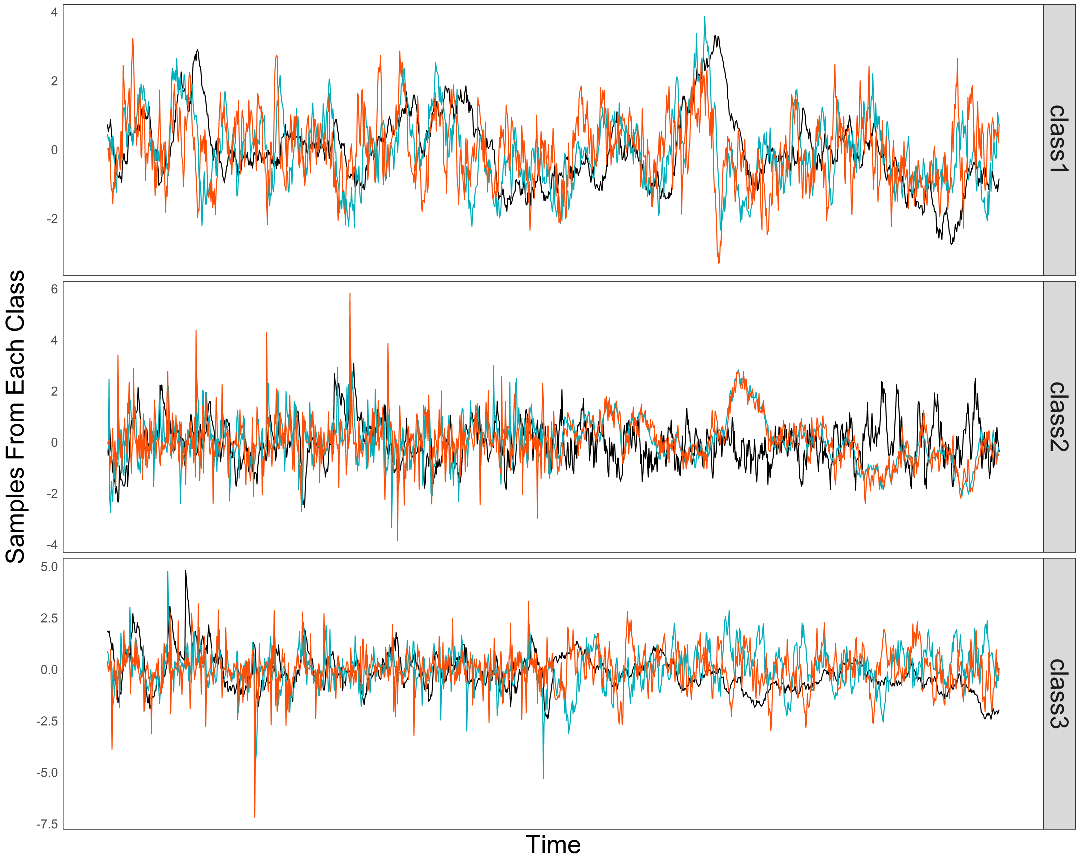


Figure 5.2 Simulation samples from 3 DGMs listed in Table 5.5

The lengths of the time series to be generated are determined as 1200, and the actual number of clusters in the dataset is equal to 3, which is the same as the number of

DGMs used to generate the time series. In this case, like in the previous case, the cluster sizes are set to 30. In other words, the number of members of each cluster in the dataset is determined as 30.

When Figure 5.2 is carefully examined, it can be noticed from the graphs of multivariate time series that the mechanisms generating these series may change over time, except for the first 600 observations of class2 and class3.

In the previous subsection, we mentioned that the multivariate time series we considered in the fist experiment set are generated in a way that they do not linearly correlate with each other. However, there is an exception to this specification for the two variables in STEM_2 for practical reasons. The structural correlation exists between Y_t and Z_t variables due to the SDE model in STEM_2 does not allow to generation of uncorrelated time series from this STEM. Accordingly, the essential correlation relationship between Y_t and Z_t in SDE is also manifested in the generated data.

The generating samples, clusters, and then evaluations of clustering performances for Case 2 can be summarized as follows:

* **nos** **= 3** (i.e., the number of hypothetical DGMs used in the experiment.)

* **noc** **= 3** (i.e., the number of true clusters.)

* **cs** **= 30** (i.e., the number of series/members in each cluster.)

* **ss** **= 1200** (i.e., the number of observations for each generated multivariate series.)

* **rn** **= 100** (i.e., the number of repetition of the experiment.)

1. Each DGM in Table 5.5 is used to generate **30** (i.e., **cs**) number of 3-dimensional multivariate time series with the length of **1200** (i.e., **ss**) to form a corresponding cluster.
2. Clustering done by the proposed approach given in Section 3.4 and through implementation steps given in Figure 3.4, and **5** dissimilarity measures given in Table 5.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed over **3** $\times$ **30** $\times$ **1200** dimensional dataset and clustering accuracies are saved.
3. Steps from **1** to **2** repeated **100** (i.e., **rn**) times to get average accuracies of **15** (i.e., **3** $*$ **5**) clustering schemes and the proposed approach.
4. The average accuracy indexes of all methods considered in the study are displayed for comparison in Table 5.6.

Table 5.6 Multivariate Experiment 1 Case 2: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach.

Dissimilarity measure	Clustering method		
	FUZZY	PAM	SPECTRAL
Model free			
D_{JD}	0.995	0.996	0.998
D_{LD}	0.687	0.569	0.989
D_{GAK}	0.478	0.457	0.498
D_{DTW}	0.559	0.598	0.729
D_{PDC}	0.688	0.434	0.882
	Proposed Clustering Approach		
	TSMB-SPCL-MV		
	$ss = 800$	$ss = 1000$	$ss = 1200$
L^2 (Euclidean) Distance	0.999	1 (%100)	1(%100)

The average of the results, which are given in Table 5.6, is obtained after the clustering of 100 artificial datasets created by the 3 DGMs in Table 5.5. As can be seen, in this experiment, clustering based on JD, LD and PDC distance measures in several clustering algorithms have a very high accuracy level. And the clustering performance of the proposed approach has achieved the highest accuracy percentage. Accurate and almost perfect clustering results are closely related to which extent the differences are evident between the mechanisms that generate the datasets. Another reason for achieving high clustering performances is that the actual number of clusters (i.e., number of true classes is 3) is much fewer than objects to be clustered in the synthetic datasets. Although in each DGM, the SDEs for generating the first 600 points of the series are similar, the differences of the combinations creating the overall data from the others become quite obvious in terms of the realizations of these data-generating mechanisms. These results show that in a clustering problem where clusters can be inferred beforehand, the proposed approach could also accurately cluster data with high performance, just like the other methods we have compared. Distance measures such as JD, LD, and PDC are efficient at clustering such datasets if the structures that produce the mean of time series are more substantial and easily distinguishable. On the other hand, since such distance measures cannot designed to detect possible heteroscedasticity inclusions in time series datasets, it can be said that clustering performance may decrease in clustering such datasets with the help of these type of distance measures.

Within the scope of this experiment, the proposed approach revealed similar and higher accuracy clustering results to the other high-accuracy clustering rates providing methods. In addition, significant findings have been obtained that the spectral clustering method can be an alternative to fuzzy and pam clustering algorithms for other distance measures.

5.1.1.3 Case 3 : Temporal 9 Cluster

The last dataset in this section that we generated from SDEs has time-varying cluster structures. Each DGM consists of a combination of 3 STEMs given in Table 5.2. Utilized hypothetical DGMs for multivariate time series generation are shown in Table 5.7. In addition, the dataset creation scenario here also contains the data generation structure that we used in the previous two experiments.

Table 5.7 Hypothetical DGMs used in Case 3.

	used to generate time series of length 1800				
	sources of samples from 1 to 600		sources of samples from 601 to 1200		sources of samples from 1201 to 1800
DGM01 :=	STEM_1 $\mid C_1$	▷	STEM_1 $\mid C_1$	▷	STEM_1 $\mid C_1$
DGM02 :=	STEM_1 $\mid C_1$	▷	STEM_1 $\mid C_1$	▶	STEM_2
DGM03 :=	STEM_1 $\mid C_1$	▷	STEM_1 $\mid C_1$	▶	STEM_3
DGM04 :=	STEM_1 $\mid C_1$	▶	STEM_2	▶	STEM_1 $\mid C_1$
DGM05 :=	STEM_1 $\mid C_1$	▶	STEM_2	▷	STEM_2
DGM06 :=	STEM_1 $\mid C_1$	▶	STEM_2	▶	STEM_3
DGM07 :=	STEM_1 $\mid C_1$	▶	STEM_3	▶	STEM_1 $\mid C_1$
DGM08 :=	STEM_1 $\mid C_1$	▶	STEM_3	▶	STEM_2
DGM09 :=	STEM_1 $\mid C_1$	▶	STEM_3	▷	STEM_3

′▷′ indicates where the data generation mechanisms remain unchanged.
′▶′ indicates where the data generating mechanisms are changed at that point.
$C_1 = 1.75$

The number of hypothetical DGMs is 9, which means that the actual number of clusters in the artificially generated datasets is also 9, considering the entire dataset length. The multivariate time series are generated to contain 1800 points in total. Notice that the actual number of clusters is three for the first 1200 realizations of the scenario. However, the number of clusters in the scheme increases to 9 after the 1200th time point due to the changes in DGMs. The time series plots of sample multivariate time series generated from each DGM are exhibited in Figure 5.3. Through this experiment, we aimed to evaluate the ability of the proposed clustering approach to recognize new cluster formations when the data generating sources that form clusters are changed over time. Through this scenario in this section, it is aimed to assess the capability of the clustering approach whether it can recognize probable changes of cluster formations when a dataset is updated with new coming data (e.g., streaming data). For this purpose, the proposed approach aims to cluster dynamically structured datasets with a dynamic clustering approach.

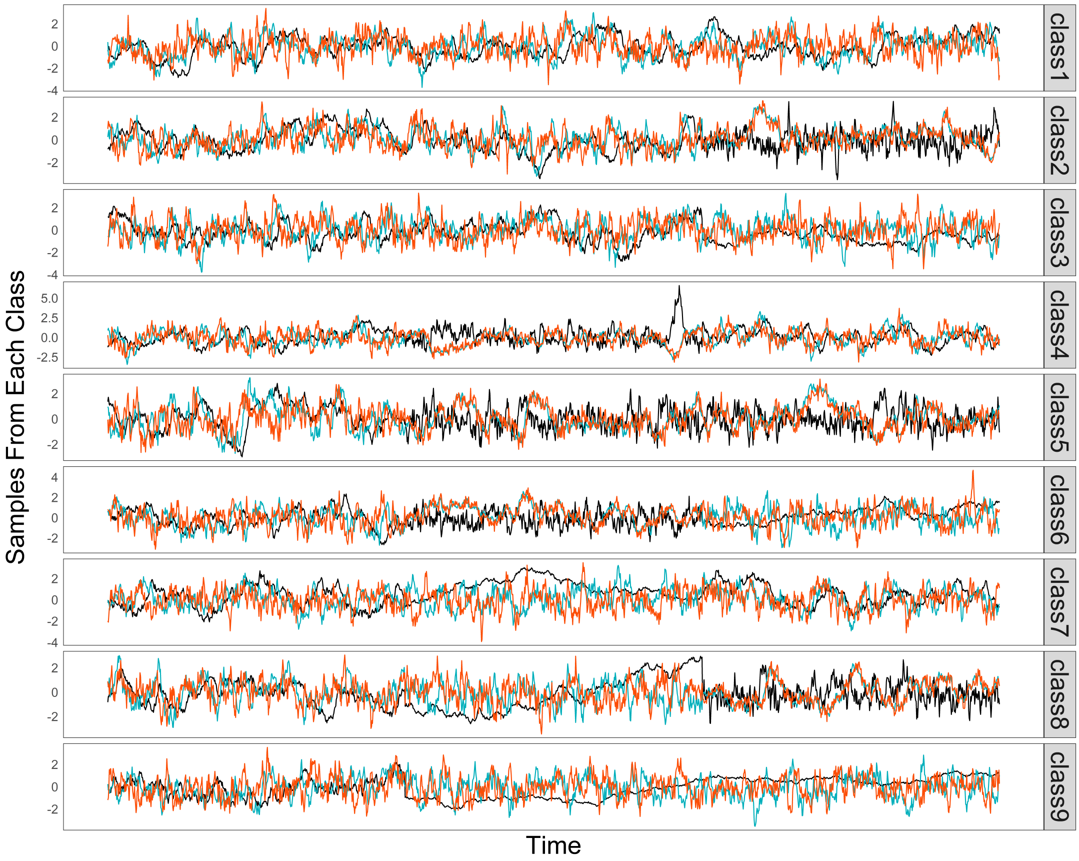


Figure 5.3 Simulation samples from 9 DGMs listed in Table 5.7

The generating samples, clusters, and then evaluations of clustering performances for Case 3 can be summarized as follows:

* **nos = 9** (i.e., the number of hypothetical DGMs used in the experiment.)

* **noc = 9** (i.e., the number of true clusters.)

* **cs = 10** (i.e., the number of series/members in each cluster.)

* **ss = 1800** (i.e., the number of observations for each generated multivariate series.)

* **rn = 100** (i.e., the number of repetition of the experiment.)

1. Each DGM in Table 5.7 is used to generate **10** (i.e., **cs**) number of 3-dimensional multivariate time series with the length of **1800** (i.e., **ss**) to form a corresponding cluster.
2. Clustering done by the proposed approach given in Section 3.4 and through implementation steps given in Figure 3.4, and **5** dissimilarity measures given in Table 5.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed over **9** $\times$ **10** $\times$ **1800** dimensional dataset and clustering accuracies are saved.
3. Steps from **1** to **2** repeated **100** (i.e., **rn**) times to get average accuracies of **15** (i.e., **3** $*$ **5**) clustering schemes and the proposed approach.
4. The average accuracy indexes of all methods considered in the study are displayed for comparison in Table 5.8.

Please note that, for all dataset lengths, the first two source components of DGMs from 1 to 3, 4 to 6, and from 7 to 9 are the same. The third source components, which are responsible for generating the last 600 realizations, partaking at the end of these DGM structures, provide each DGM to be labeled as different from one another. Consequently, there would be a total of 9 clusters in the dataset consisting of 1800-length multivariate time series generated from these DGMs.

The first 1200 realizations of the dataset generated from the DGMs in Table 5.7 are discussed in the previous experiment in subsection 5.1.1.2. The clustering results for this part of the dataset given in Table 5.6. The number of clusters for the mentioned part of the dataset is three, and the number of members of each cluster is 30. After the 1200th point, the number of clusters increased to 9 due to the changes in DGMs, and the number of members in each cluster decreased to 10 accordingly.

The first 600 observations of the dataset are all generated from the same SDE structure, STEM_1$|C_1$, given in Table 5.2. The number of clusters for this part of the data is equal to 1. Therefore, while the number of clusters in the dataset was one initially, it increased to 3 after the 600th time point and remained the same till the 1200th point. And, then after the 1200th time point of the samples the number of clusters increased to 9.

The performance outcomes of the clusterings made by considering the whole dataset length, that is, assuming that we are at the 1800th time point as location, are given in Table 5.8. In addition, the proposed clustering approach is repeated over three different sample sizes (i.e., lengths) to make suggestions about the optimum data length required to identify the number of clusters that change over time, and the results are given in Table 5.8.

Table 5.8 Multivariate Experiment 1 Case 3: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach.

Dissimilarity measure	Clustering method		
	FUZZY	PAM	SPECTRAL
Model free			
D_{JD}	0.548	0.757	0.775
D_{LD}	0.345	0.276	0.593
D_{GAK}	0.328	0.325	0.257
D_{DTW}	0.278	0.407	0.301
D_{PDC}	0.331	0.261	0.273
	Proposed Clustering Approach		
	TSMB-SPCL-MV		
	$ss = 1400$	$ss = 1600$	$ss = 1800$
L^2 (Euclidean) Distance	0.886	0.970	0.970

The accuracy index of the proposed clustering approach to correctly recognize clusters in the dataset is quite high. However, the sample size required to distinguish clusters with the highest accuracy is about 1600 observations. In other words, the optimum data length to be examined in order to determine the changing cluster structure after the 1200th point correctly is around 350 - 400 observations. These findings support the results obtained from previous experiments.

In the scope of this experiment, on the other hand, among the measures we compared, only the JD distance measure has clustered the datasets at a level of good accuracy that can be mentioned. Since JD and similar distance measures operate only based on the frequency domain, the time-dependent properties of time series do not seem to affect distance computations. The clustering performance of such distance measures would be riskier if the time series data to be clustered contains heteroscedasticity and nonlinearity.

The results of the GAP statistics are also presented for this experiment. The Figure 5.4 and Table 5.9 shows the distribution of GAP statistics over 100 repetitions for possible cluster numbers from 2 to 20. In 70 out of 100 repetitions, the GAP statistic peaked at cluster number 9. The correctness of the GAP statistics is directly linked to the success of the clustering method. The more successful the clustering method is, the higher the use-value and accuracy of the GAP statistic.

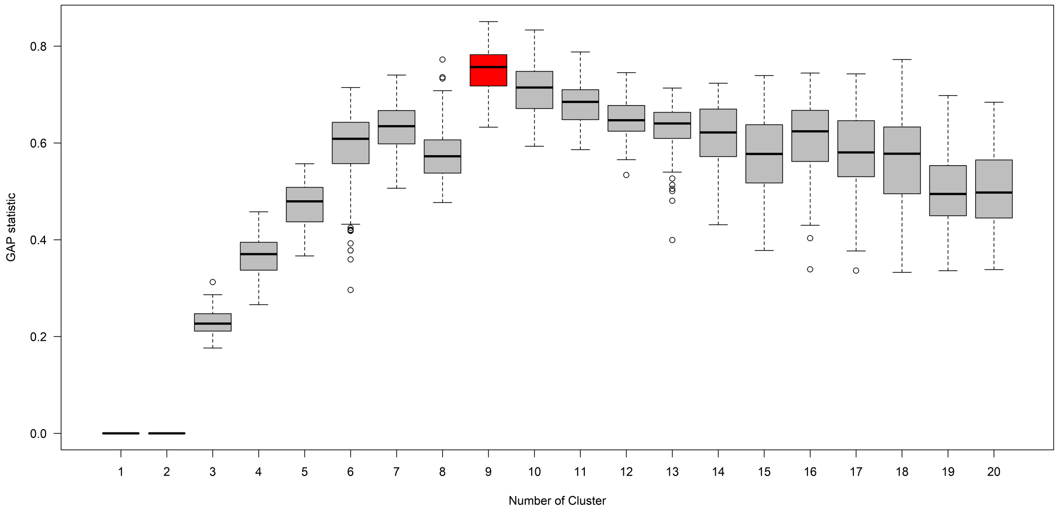


Figure 5.4 GAP statistics' Box-Plots per cluster number using TSMB-SPCL-MV.

Table 5.9 The Gap statistics peaks out of 100 repetitions using TSMB-SPCL-MV.

cluster number, $\widehat{cl}$	9	10	11	>12
counts of peaks	70	15	10	5

5.1.2 Second Set of Experiments on Multivariate Time-series Clustering

In this section, to assess the performance of the clustering approach proposed for multivariate time series, we generated datasets based on Vector Error Correction Model (VECM) structures developed especially for modeling co-integrated multivariate time series. Three distinct and three-variate VECM(1) specifications (i.e., VECM of lag order 1) are used in this experiment for composing DGMs to generate datasets to be clustered. These specifications are labeled as STEM_1, STEM_2 and STEM_3, and are presented in Table 5.10. Since the STEMs we employed are only take on the role of the sources that generates the datasets to be clustered, detailed information about the VECM are not provided. Besides, during the datasets generation we adopted heteroscedastic error term specifications which violates the standard VECM assumptions. For detailed information about VECM and similar multivariate time series models, please see Lüetkepohl and Kraetzig [162] and Pfaff [163].

Table 5.10 Utilized specifications of VECMs in the first set of experiments.

STEM_1 :

$$\begin{bmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \\ \Delta y_{3,t} \end{bmatrix} = \begin{bmatrix} -0.10 & 0.10 & -0.15 \\ 0.07 & -0.07 & 0.10 \\ 0.02 & -0.02 & 0.03 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + \begin{bmatrix} -0.05 & -0.55 & 1.05 \\ -0.20 & -0.25 & 0.30 \\ -0.15 & -0.75 & 0.10 \end{bmatrix} \begin{bmatrix} \Delta y_{1,t-1} \\ \Delta y_{2,t-1} \\ \Delta y_{3,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}$$

STEM_2 :

$$\begin{bmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \\ \Delta y_{3,t} \end{bmatrix} = \begin{bmatrix} -0.02 & 0.01 & -0.02 \\ 0.04 & -0.03 & 0.04 \\ -0.06 & 0.04 & -0.07 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + \begin{bmatrix} -0.85 & 0.08 & -0.85 \\ -0.15 & 0.55 & -0.80 \\ 0.35 & 0.05 & -0.65 \end{bmatrix} \begin{bmatrix} \Delta y_{1,t-1} \\ \Delta y_{2,t-1} \\ \Delta y_{3,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}$$

STEM_3 :

$$\begin{bmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \\ \Delta y_{3,t} \end{bmatrix} = \begin{bmatrix} 0.05 & -0.06 & 0.02 \\ 0.07 & -0.09 & 0.03 \\ 0.00 & -0.01 & 0.00 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + \begin{bmatrix} 0.10 & 0.45 & -0.20 \\ -0.15 & 1.00 & 0.00 \\ -0.20 & 0.15 & 0.90 \end{bmatrix} \begin{bmatrix} \Delta y_{1,t-1} \\ \Delta y_{2,t-1} \\ \Delta y_{3,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}$$

$\Delta y_t = y_t - y_{t-1}$

$\varepsilon_{1,t},\ \varepsilon_{2,t},\ \varepsilon_{3,t}$ are error terms.

Please note that the error terms can be described by any process or distribution.

For realizations of the error terms, $\overrightarrow{\varepsilon_t}$, of the three-variate VECM(1) specifications the Extended Constant Conditional Correlation (ECCC) GARCH models are employed during dataset generation phase. By doing so, we ensured that the generated multivariate time series contains heteroscedasticity. ECCC-GARCH models are useful tools for modeling multivariate GARCH processes. Please see Silvennoinen and Teräsvirta [164], He and Teräsvirta [165] or Francq and Zakoïan [166] for information about multivariate GARCH, ECCC-GARCH and its various applications. In this experiment, three distinct GARCH(1,1) specifications are adopted for error terms. These processes are labeled as $\mathbf{C_1}$, $\mathbf{C_2}$ and $\mathbf{C_3}$, and are given in Table 5.11.

Table 5.11 Utilized specifications of multivariate error structures in the seconds set of experiments.

$$\mathbf{C_1} : \mathbf{ECCC - GARCH(1,1)} =$$

$$\begin{cases} h_{1,t} = 0.003 + 0.15\varepsilon_{1,t-1}^2 + 0.45h_{1,t-1} \\ h_{2,t} = 0.005 + 0.30\varepsilon_{2,t-1}^2 + 0.35h_{2,t-1} \\ h_{3,t} = 0.001 + 0.45\varepsilon_{3,t-1}^2 + 0.25h_{3,t-1} \end{cases}$$

$$\mathbf{C_2} : \mathbf{ECCC - GARCH(1,1)} =$$

$$\begin{cases} h_{1,t} = 0.001 + 0.20\varepsilon_{1,t-1}^2 + 0.05\varepsilon_{2,t-1}^2 + 0.20\varepsilon_{3,t-1}^2 + 0.10h_{1,t-1} + 0.15h_{2,t-1} + 0.15h_{3,t-1} \\ h_{2,t} = 0.001 + 0.25\varepsilon_{1,t-1}^2 + 0.30\varepsilon_{2,t-1}^2 + 0.15\varepsilon_{3,t-1}^2 + 0.01h_{1,t-1} + 0.25h_{2,t-1} + 0.05h_{3,t-1} \\ h_{3,t} = 0.002 + 0.05\varepsilon_{1,t-1}^2 + 0.01\varepsilon_{2,t-1}^2 + 0.15\varepsilon_{3,t-1}^2 + 0.35h_{1,t-1} + 0.05h_{2,t-1} + 0.10h_{3,t-1} \end{cases}$$

$$\mathbf{C_3} : \mathbf{ECCC - GARCH(1,1)} =$$

$$\begin{cases} h_{1,t} = 0.000 + 0.05\varepsilon_{1,t-1}^2 + 0.15\varepsilon_{2,t-1}^2 + 0.10\varepsilon_{3,t-1}^2 + 0.04h_{1,t-1} + 0.05h_{2,t-1} + 0.04h_{3,t-1} \\ h_{2,t} = 0.000 + 0.05\varepsilon_{1,t-1}^2 + 0.10\varepsilon_{2,t-1}^2 + 0.05\varepsilon_{3,t-1}^2 + 0.00h_{1,t-1} + 0.08h_{2,t-1} + 0.01h_{3,t-1} \\ h_{3,t} = 0.000 + 0.01\varepsilon_{1,t-1}^2 + 0.00\varepsilon_{2,t-1}^2 + 0.05\varepsilon_{3,t-1}^2 + 0.15h_{1,t-1} + 0.01h_{2,t-1} + 0.02h_{3,t-1} \end{cases}$$

where $\varepsilon_{i,t} = \sqrt{h_{i,t}}e_{i,t},$ and $e_{i,t} \overset{i.i.d}{\sim} D_{snorm}(0,1)$, $i = 1, 2, 3$.

The constant conditional correlation matrix, ρ, assumed as

$$\rho = \begin{array}{cc} \begin{array}{ccc} h_{1,t} & h_{2,t} & h_{3,t} \end{array} & \\ \begin{bmatrix} 1.00 & 0.10 & 0.45 \\ 0.10 & 1.00 & 0.85 \\ 0.45 & 0.85 & 1.00 \end{bmatrix} & \begin{array}{c} h_{1,t} \\ h_{2,t} \\ h_{3,t} \end{array} \end{array}$$ for all three multivariate GARCH specification.

Thanks to the root structures given in Table 5.10 and Table 5.11, it is possible to generate 3-variate co-integrated time series with heteroscedasticity. Please note that, the generated series throughout the experiment are individually non-stationary time series. Also, the linear and nonlinear correlation structure within and between multivariate objects is unrestricted for conditional mean processes. This means that the correlation structures of a multivariate time series created by any DGM in this experiment can include every possible correlation combination. In this way, it is aimed to illustrate that there is no significant effect of the correlatedness or uncorrelatedness of the objects to be clustered on the clustering performance. In the appendix of this study, the clustering performances on the datasets derived from both correlated and uncorrelated structures are also provided. It can be said that, within the scope of experiments done in this study, there is no evidence that the correlation structure of time series to be clustered seriously affects the clustering performance of the proposed approach.

Please note that, the ECCC-GARCH model necessitates constant conditional correlation by its constitution. However, we considered the same fixed constant conditional correlation structure for all three distinct conditional variance processes throughout the experiment. The assumed correlation matrix, ρ, is indicated in Table 5.11. Therefore, the conditional variances generated from all three GARCH(1,1) models would have the same correlation structure. Nonetheless, the constant correlation structure of the conditional variance neither transfers over to the conditional mean process nor does it have any positive or negative effect on the clustering performance of the proposed approach. According to the findings in this section, we can discuss that the performance of the proposed clustering approach is independent of the correlation structures in the conditional mean or variance processes. Another important feature of the ECCC-GARCH models is their capability to simulate or modeling the volatility spillover phenomena in the field of economics. For example, it can be said that processes $\mathbf{C_2}$ and $\mathbf{C_3}$ can create spillover effects within multivariate objects. In future research, it may be considered to consolidate a model similar to ECCC-GARCH instead of the DTHGARCH model used in the proposed clustering approach.

Simulations of VECMs and multivariate GARCH processes given in this experiment are performed by the use of **R** packages **tsDyn** [167] and **ccgarch** [168].

5.1.2.1 Case 1 : Non-temporal 3 Cluster

In Case 1 of the second experiment sets, the datasets to be clustered are generated with the help of DGMs given in Table 5.12.

Table 5.12 Hypothetical DGMs used in Case 1.

$$\textbf{DGM01} \coloneqq \textbf{STEM\_1} \mid \varepsilon_t \sim C_1$$

$$\textbf{DGM02} \coloneqq \textbf{STEM\_1} \mid \varepsilon_t \sim C_2$$

$$\textbf{DGM03} \coloneqq \textbf{STEM\_1} \mid \varepsilon_t \sim C_3$$

C_1, C_2, and C_3 are three distinct ECCC-GARCH(1,1) specifications given in Table 5.11.

The time series plots of the sample multivariate time-series generated from each DGM (i.e., class) is illustrated in Figure 5.5.

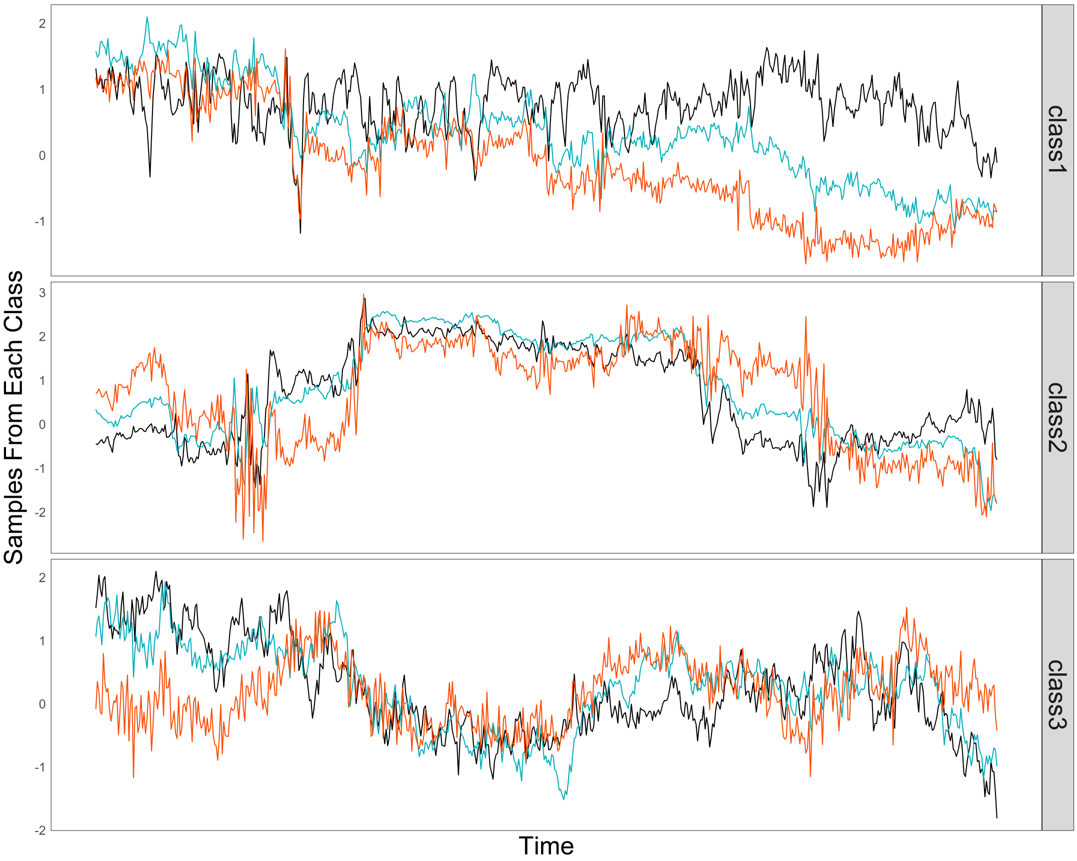


Figure 5.5 Simulation samples from 3 DGMs listed in Table 5.12

While all three DGMs include the same STEM_1 structure, they differ only in conditional variance processes.

Set of the multivariate time series generated from each DGM given in Table 5.12 forms a cluster. Thus, it can be said that there are 3 clusters in the datasets created. In this regard, according to similar logic followed in previous cases, the hypothetical DGMs we utilized are latently underlying the center of each cluster.

Set of the multivariate time series length of 600 are created in this scenario. In previous experiments, the sizes of the clusters, that is, the number of elements in each cluster, are set equal to each other (i.e., balanced). In the experiments in this section, however, it is desired to evaluate the performance of the proposed clustering approach in determining such unevenly distributed cluster sizes throughout the dataset to be clustered. For this reason, a total of 120 multivariate time series (i.e., objects) are produced for each dataset to be clustered, with 60 objects belongs to the first of the 3 clusters and the remaining 2 clusters have 30 objects each. Therefore, the created dataset from 1 repetition consists of 3 clusters corresponding to the number of DGMs used where class1 contains 60 objects, class2 and class3 contains 30 objects each.

The generating samples, clusters, and then evaluations of clustering performances for second set of multivariate experiments' Case 1 can be summarized as follows:

* **nos** = **3** (i.e., the number of hypothetical DGMs used in the experiment.)

* **noc** = **3** (i.e., the number of true clusters.)

* **cs** = **60, 30, 30** (i.e., the number of series/members in each cluster.)

* **ss** = **600** (i.e., the number of observations for each generated multivariate series.)

* **rn** = **100** (i.e., the number of repetition of the experiment.)

1. DGM01 in Table 5.12 is used to generate **60**, DGM02 is used to generate **30** and DGM03 is used to generate **30** number of 3-dimensional multivariate time series with the length of **600** (i.e., **ss**) to form a dataset.
2. Clustering done by the proposed approach given in Section 3.4 and through implementation steps given in Figure 3.4, and **5** dissimilarity measures given in Table 5.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed over **120** $\times$ **600** dimensional dataset and clustering accuracies are saved.
3. Steps from **1** to **2** repeated **100** (i.e., **rn**) times to get average accuracies of **15** (i.e., **3** $*$ **5**) clustering schemes and the proposed approach.
4. The average accuracy indexes of all methods considered in the study are displayed for comparison in Table 5.13.

Table 5.13 Multivariate Experiment 2 Case 1: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach.

Dissimilarity measure	Clustering method		
	FUZZY	PAM	SPECTRAL
Model free			
D_{JD}	0.444	0.429	0.431
D_{LD}	0.477	0.434	0.673
D_{GAK}	0.427	0.394	0.388
D_{DTW}	0.425	0.392	0.388
D_{PDC}	0.583	0.400	0.485

	Proposed Clustering Approach		
	TSMB-SPCL-MV		
	$ss = 350$	$ss = 475$	$ss = 600$
L^2 (Euclidean) Distance	0.649	0.739	0.803

According to Table 5.13 containing the clustering methods' accuracies, it can be said that almost all clustering methods cannot truly cluster the examined datasets. However, the proposed approach has the highest success rate and correctly groups 90 objects out of 120, on average.

The main reason for low accuracy clustering is that the datasets to be clustered here are generated by almost identical DGMs. If the series desired to be clustered differ from each other especially in terms of conditional variance characteristics, the clustering performances of the methods decreases considerably. It should be noted that most of the methods compared here are not designed for such a purpose. Although we have designed the proposed clustering approach to also distinguish between GARCH-like volatilities, there is a limit to the accuracy rate of the proposed approach. In addition, in cases where detection and extraction of heteroscedasticity as a property is decisive for exact grouping, the data length required by the proposed approach for clustering in the most efficient way exceeds 450.

5.1.2.2 Case 2 : Temporal 6 Cluster

In Case 2 of the second experiment sets, the time-varying clustering performance of the proposed approach and the compared clustering methods are assessed via another temporal clusters scenario. In this case, we used the hypothetical data generation mechanisms given in Table 5.14 to generate the multivariate time series datasets that are to be clustered. These DGMs consist of the combinations of the STEMs and multivariate GARCH processes presented in Table 5.10 and Table5.11, respectively.

Table 5.14 Hypothetical DGMs used in Case 2.

	used to generate time series of length 1200		
	sources of samples from 1 to 600		sources of samples from 601 to 1200
DGM01 :=	$\mathbf{STEM\_1} \mid \varepsilon_t \sim C_1$	$\triangleright$	$\mathbf{STEM\_1} \mid \varepsilon_t \sim C_1$
DGM02 :=	$\mathbf{STEM\_1} \mid \varepsilon_t \sim C_1$	$\blacktriangleright$	$\mathbf{STEM\_2} \mid \varepsilon_t \sim C_2$
DGM03 :=	$\mathbf{STEM\_1} \mid \varepsilon_t \sim C_2$	$\blacktriangleright$	$\mathbf{STEM\_3} \mid \varepsilon_t \sim C_3$
DGM04 :=	$\mathbf{STEM\_1} \mid \varepsilon_t \sim C_3$	$\triangleright$	$\mathbf{STEM\_1} \mid \varepsilon_t \sim C_3$
DGM05 :=	$\mathbf{STEM\_1} \mid \varepsilon_t \sim C_3$	$\blacktriangleright$	$\mathbf{STEM\_2} \mid \varepsilon_t \sim C_1$
DGM06 :=	$\mathbf{STEM\_1} \mid \varepsilon_t \sim C_3$	$\blacktriangleright$	$\mathbf{STEM\_3} \mid \varepsilon_t \sim C_2$

$'\triangleright'$ indicates where the data generation mechanisms remain unchanged.
$'\blacktriangleright'$ indicates where the data generating mechanisms are changed at that point.
C_1, C_2, and C_3 are three distinct ECCC-GARCH(1,1) specifications given in Table 5.11.

In this experiment, datasets created through simulated multivariate time series from 6 different DGMs are clustered, and clustering performance results are presented. Each dataset to be clustered contains 120 objects (i.e., multivariate time series).

Although the correct number of clusters in the datasets is 6, each cluster is constructed not to include an equal number of objects. In other words, cluster sizes or the number of members in each cluster are unbalanced. Therefore, while 30 multivariate time series are generated from DGM01, DGM02, DGM03, 10 multivariate series are generated from DGM04, DGM05, and DGM06.

The time series plots of sample multivariate time series generated from each DGM are exhibited in Figure 5.6.

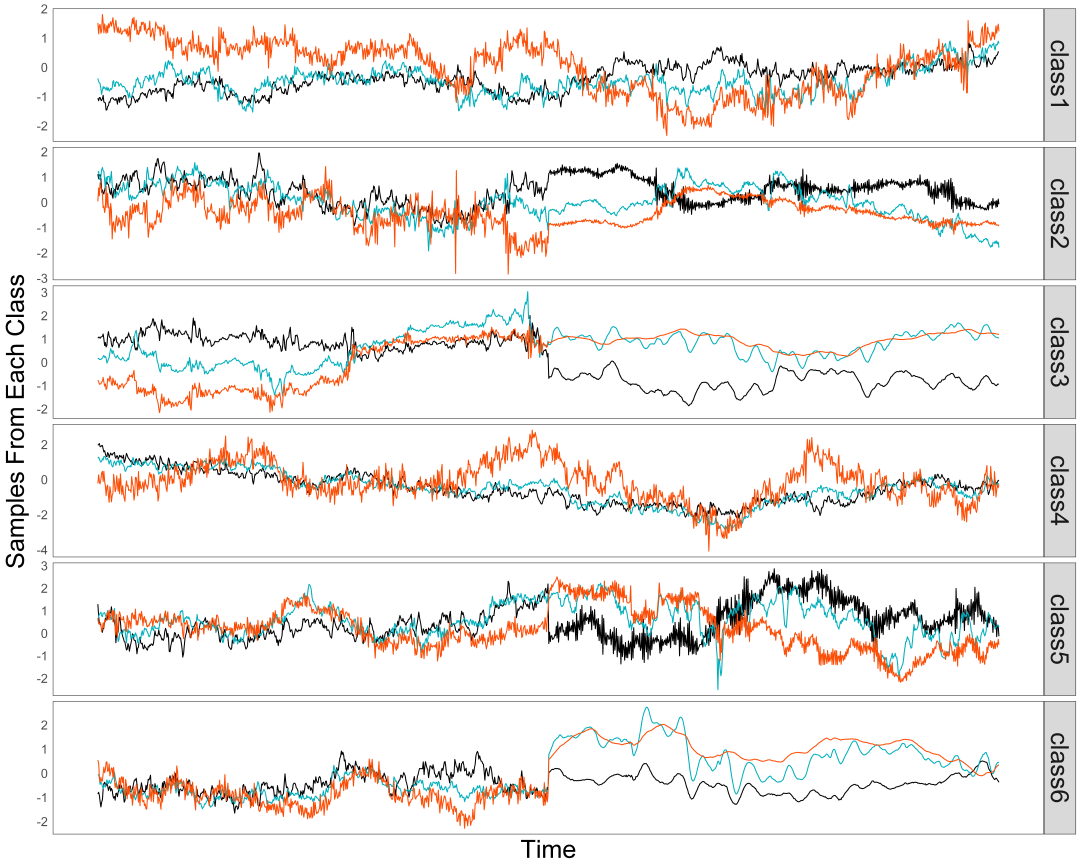


Figure 5.6 Simulation samples from 6 DGMs listed in Table 5.14

As can be seen from the time series plots, it can be visually noticed that the traces left by STEM_1, STEM_2 and STEM_3 over time are different from each other. However, although some part of time series graphs are seemingly identical, they are ultimately different from each other because of the conditional variance structures they include. For example, the last 600-length portions of series belonged to class3 and class6 are generated from the same mean processes but different variance structures. Accordingly, it is important to detect the heteroscedasticity phenomenon, which is not visually noticeable, and to use it as a distinctive feature in cluster analysis. In this experiment, let's note that while the number of clusters is 3 for the first 600-length segments of the datasets, the number of clusters in the datasets changed to 6 starting from the 600th time point.

The clustering results for the first 600-length parts of the datasets are given in the previous experiment. Hence, this case includes the same dataset generation for the first 600-length structure as the previous experiment and is a continuation of the 1st case's scenario.

As stated before, the generated multivariate time series within the scope of the experiment are individually non-stationary time series. Moreover, the linear and nonlinear correlation structure within and between three-variate time series is unrestricted, especially for mean processes. This means that the correlation structures of a three-variate time series produced by any DGM throughout the experiment can accommodate all possible correlation combinations. By doing this, it is aimed to show that whether the objects to be clustered are correlated or uncorrelated, stationary or non-stationary do not have a significant effect on clustering performance.

The generating samples, clusters, and then evaluations of clustering performances for second set of multivariate experiments' Case 2 can be summarized as follows:

* **nos = 6** (i.e., the number of hypothetical DGMs used in the experiment.)

* **noc = 6** (i.e., the number of true clusters.)

* **cs = 30, 30, 30, 10, 10, 10** (i.e., the number of members in each cluster.)

* **ss = 1200** (i.e., the number of observations for each generated multivariate series.)

* **rn = 100** (i.e., the number of repetition of the experiment.)

1. DGM01 in Table 5.14 is used to generate **30**, DGM02 is used to generate **30**, DGM03 is used to generate **30**, DGM04 is used to generate **10**, DGM05 is used to generate **10** and DGM06 is used to generate **10** number of 3-variate time series with the length of **1200** (i.e., **ss**) to form a dataset.
2. Clustering done by the proposed approach given in Section 3.4 and through implementation steps given in Figure 3.4, and **5** dissimilarity measures given in Table 5.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed over **120** $\times$ **1200** dimensional dataset and clustering accuracies are saved.
3. Steps from **1** to **2** repeated **100** (i.e., **rn**) times to get average accuracies of **15** (i.e., **3** $*$ **5**) clustering schemes and the proposed approach.
4. The average accuracy indexes of all methods considered in the study are displayed for comparison in Table 5.15.

Table 5.15 Multivariate Experiment 2 Case 2: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach.

Dissimilarity measure	Clustering method		
	FUZZY	PAM	SPECTRAL
Model free			
D_{JD}	0.453	0.468	0.500
D_{LD}	0.382	0.494	0.498
D_{GAK}	0.277	0.266	0.274
D_{DTW}	0.278	0.266	0.274
D_{PDC}	0.459	0.561	0.570
	Proposed Clustering Approach		
	TSMB-SPCL-MV		
	$ss = 800$	$ss = 1000$	$ss = 1200$
L^2 (Euclidean) Distance	0.730	0.779	0.855

The clustering performances obtained for Case 2 are very close to the results obtained in the previous case. Although the success rate does not decrease despite the cluster structure that changes over time, the proposed approach cannot cluster 15% of the three-variate time series correctly, on average. These results support our findings that the performance of the proposed clustering approach is limited in clustering datasets in which heteroscedasticity features are dominant than mean structures. Again, as in the previous case, where the definite extraction of conditional variance features is essential for correct clustering, the data length required by the proposed approach for the most efficient clustering is around 500.

The clustering performances of the distance measures compared are also quite limited. These distance measures and the proposed clustering approach can group time series with high accuracy when they entirely depend on conditional mean processes. However, the methods mentioned in this study demand further improvement to accurately cluster all sorts of heteroscedasticity and various complex qualities of real-life time series.

5.1.2.3 Case 3 : Temporal 12 Cluster

The last scenario about the temporal cluster under the second set of experiment on multivariate synthetic data is consist of 12 DGM/cluster. Each DGM consists of sequences of STEMs and multivariate GARCH processes shown in Table 5.10 and Table 5.11, respectively. The hypothetical DGMs employed in this scenario are given in Table 5.16. Furthermore, the dataset formation scenario here completely includes the data generation arrangement we practiced in the previous two experiments.

Table 5.16 Hypothetical DGMs used in Case 3.

	used to generate time series of length 1800				
	sources of samples from 1 to 600		sources of samples from 601 to 1200		sources of samples from 1201 to 1800
$\mathbf{DGM01} :=$	$\mathbf{STEM\_1} \mid C_1$	$\triangleright$	$\mathbf{STEM\_1} \mid C_1$	$\triangleright$	$\mathbf{STEM\_1} \mid C_1$
$\mathbf{DGM02} :=$	$\mathbf{STEM\_1} \mid C_1$	$\triangleright$	$\mathbf{STEM\_1} \mid C_1$	$\blacktriangleright$	$\mathbf{STEM\_2} \mid C_2$
$\mathbf{DGM03} :=$	$\mathbf{STEM\_1} \mid C_1$	$\triangleright$	$\mathbf{STEM\_1} \mid C_1$	$\blacktriangleright$	$\mathbf{STEM\_3} \mid C_3$
$\mathbf{DGM04} :=$	$\mathbf{STEM\_1} \mid C_1$	$\blacktriangleright$	$\mathbf{STEM\_2} \mid C_2$	$\blacktriangleright$	$\mathbf{STEM\_1} \mid C_1$
$\mathbf{DGM05} :=$	$\mathbf{STEM\_1} \mid C_1$	$\blacktriangleright$	$\mathbf{STEM\_2} \mid C_2$	$\triangleright$	$\mathbf{STEM\_2} \mid C_2$
$\mathbf{DGM06} :=$	$\mathbf{STEM\_1} \mid C_1$	$\blacktriangleright$	$\mathbf{STEM\_2} \mid C_2$	$\blacktriangleright$	$\mathbf{STEM\_3} \mid C_3$
$\mathbf{DGM07} :=$	$\mathbf{STEM\_1} \mid C_2$	$\blacktriangleright$	$\mathbf{STEM\_3} \mid C_3$	$\blacktriangleright$	$\mathbf{STEM\_1} \mid C_1$
$\mathbf{DGM08} :=$	$\mathbf{STEM\_1} \mid C_2$	$\blacktriangleright$	$\mathbf{STEM\_3} \mid C_3$	$\blacktriangleright$	$\mathbf{STEM\_2} \mid C_2$
$\mathbf{DGM09} :=$	$\mathbf{STEM\_1} \mid C_2$	$\blacktriangleright$	$\mathbf{STEM\_3} \mid C_3$	$\triangleright$	$\mathbf{STEM\_3} \mid C_3$
$\mathbf{DGM10} :=$	$\mathbf{STEM\_1} \mid C_3$	$\triangleright$	$\mathbf{STEM\_1} \mid C_3$	$\triangleright$	$\mathbf{STEM\_1} \mid C_3$
$\mathbf{DGM11} :=$	$\mathbf{STEM\_1} \mid C_3$	$\blacktriangleright$	$\mathbf{STEM\_2} \mid C_1$	$\triangleright$	$\mathbf{STEM\_2} \mid C_1$
$\mathbf{DGM12} :=$	$\mathbf{STEM\_1} \mid C_3$	$\blacktriangleright$	$\mathbf{STEM\_3} \mid C_2$	$\triangleright$	$\mathbf{STEM\_3} \mid C_2$

$'\triangleright'$ indicates where the data generation mechanisms remain unchanged.
$'\blacktriangleright'$ indicates where the data generating mechanisms are changed at that point.
C_1, C_2, and C_3 are three distinct ECCC-GARCH(1,1) specifications given in Table 5.11.

The number of hypothetical DGMs is 12 in this case, which means that the actual number of clusters in the datasets is also 12, considering the whole dataset length. The three-variate time series are generated to include 1800 points in total.

Please note that the actual number of clusters is six for the first 1200-length simulation of the scenario. Nevertheless, the number of clusters in the plan increases to 12 following the 1200th time point due to the changes in DGMs. The sample three-variate time series plots from each DGM/class are shown in Figure 5.7.

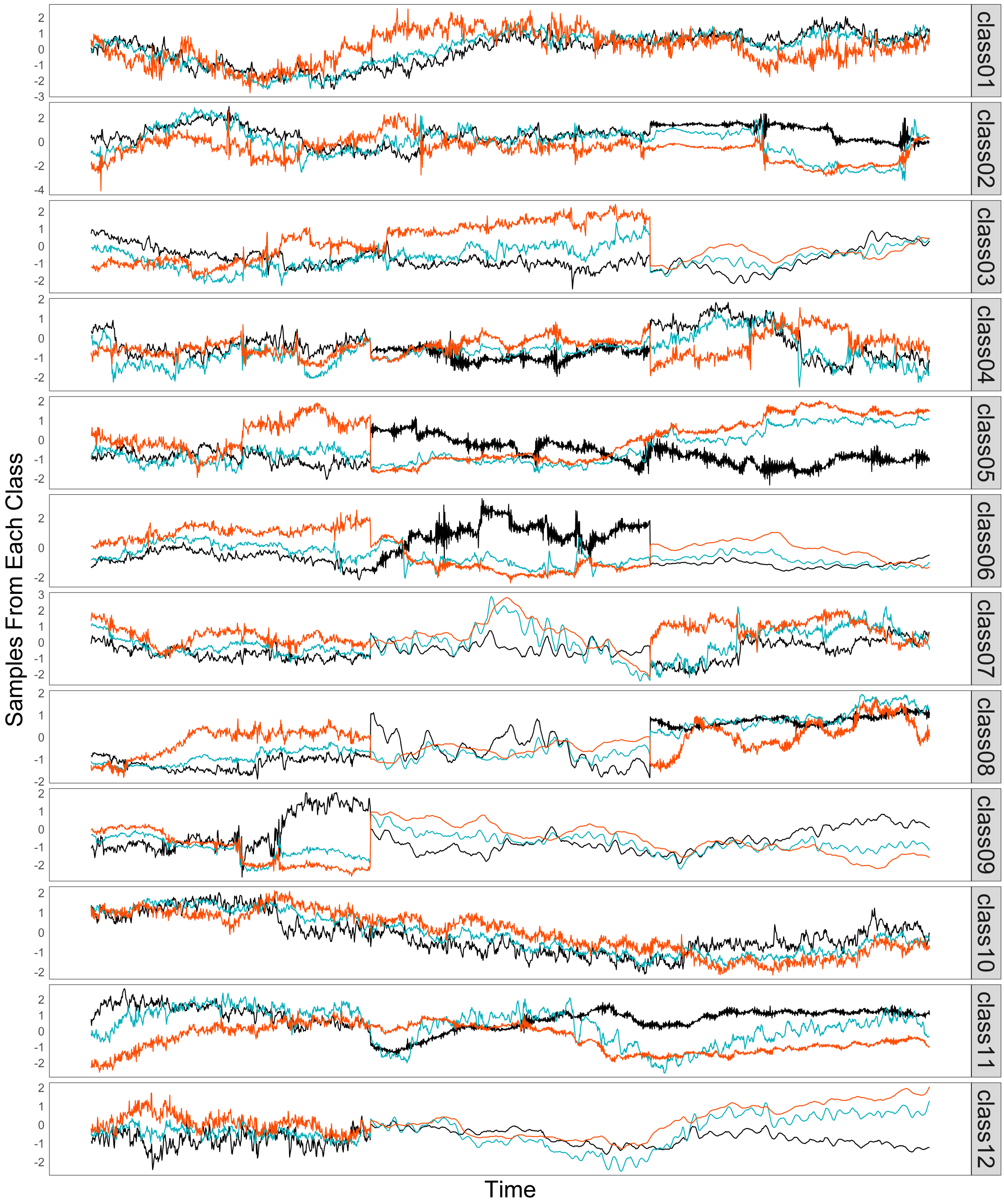


Figure 5.7 Simulation samples from 12 DGMs listed in Table 5.16

Hence, this scenario is a continuation of the 1st and 2nd cases.

Through this experiment, we intended to assess the capability of the proposed clustering approach to recognize changing cluster formations when the data generating sources that form clusters are modified over time. That is to say, it aims to assess the ability of the clustering approach to recognize changes of cluster formations when a dataset is updated with new coming data (e.g., streaming data). For this goal, the proposed procedure aims to cluster dynamically structured datasets with a dynamic clustering approach.

The generating samples, clusters, and then evaluations of clustering performances for Case 3 can be summarized as follows:

* **nos** **= 12** (i.e., the number of hypothetical DGMs used in the experiment.)

* **noc** **= 12** (i.e., the number of true clusters.)

* **cs** **= 10** (i.e., the number of series/members in each cluster.)

* **ss** **= 1800** (i.e., the number of observations for each generated multivariate series.)

* **rn** **= 100** (i.e., the number of repetition of the experiment.)

1. Each DGM in Table 5.16 is used to generate **10** (i.e., **cs**) number of 3-dimensional multivariate time series with the length of **1800** (i.e., **ss**) to form a corresponding cluster.
2. Clustering done by the proposed approach given in Section 3.4 and through implementation steps given in Figure 3.4, and **5** dissimilarity measures given in Table 5.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed over **12** $\times$ **10** $\times$ **1800** dimensional dataset and clustering accuracies are saved.
3. Steps from **1** to **2** repeated **100** (i.e., **rn**) times to get average accuracies of **15** (i.e., **3** $*$ **5**) clustering schemes and the proposed approach.
4. The average accuracy indexes of all methods considered in the study are displayed for comparison in Table 5.17.

The source components, which are responsible for generating the last 600 realizations, participating in the DGM structures at the end, made each DGM to be identified as different from one another. As a result, there is a total of 12 clusters in the dataset consisting of 1800-length three-variate time series generated from these DGMs.

The first 1200-length simulations of the DGMs listed in Table 5.16 are addressed in the previous case in subsection 5.1.2.2. The clustering results for this part of the dataset are given in Table 5.15. The number of clusters for the mentioned part of the dataset is six, and the number of members of each class is unbalanced and described as 30, 30, 30, 10, 10, and 10, from 1 to 6 respectively. Following the 1200th point, the number of clusters expanded to 12 due to the changes in DGMs, and the number of members in each cluster set at 10.

Accordingly, while the number of clusters in the dataset was 3 initially, it increased to 6 after the 600th time point and remained the same up to the 1200th point. And, then following the 1200th time point of the samples the number of clusters extended to 12. The clustering results for the first 600-length parts of the datasets are given in subsection 5.1.2.1. The results of the clusterings obtained over the entire length of the dataset, that is, pretending that we are at the 1800th time location, are given in Table 5.17.

Table 5.17 Multivariate Experiment 2 Case 3: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach.

Dissimilarity measure	Clustering method		
	FUZZY	PAM	SPECTRAL
Model free			
D_{JD}	0.365	0.465	0.480
D_{LD}	0.271	0.380	0.388
D_{GAK}	0.198	0.260	0.229
D_{DTW}	0.194	0.259	0.223
D_{PDC}	0.283	0.410	0.634

	Proposed Clustering Approach		
	TSMB-SPCL-MV		
	$ss = 1400$	$ss = 1600$	$ss = 1800$
L^2 (Euclidean) Distance	0.880	0.971	0.963

Furthermore, the proposed clustering approach is repeated over three different sample sizes (i.e., lengths) to make suggestions about the optimum data length needed to recognize the number of clusters that change over time, and the results are given in Table 5.17.

The accuracy index of the proposed clustering approach to correctly recognize clusters in the dataset is quite high. One reason for this is that the time series increase in length and the STEM combinations that create the dataset changing over time make the series more distinguishable in terms of conditional mean processes rather than heteroscedasticity features. However, the sample size required to distinguish clusters with the highest accuracy is about 1600 observations. In other words, the optimum data length to be examined in order to determine the changing cluster structure after the 1200th point correctly is around 400 observations.

The results of the GAP statistics are also given for this experiment. The Figure 5.8 and Table 5.18 shows the distribution of GAP statistics over 100 repetitions for possible cluster numbers from 2 to 20. In this case, in 54 out of 100 repetitions, the GAP statistic peaked at cluster number 12. As we stated before, the precision of the GAP statistics is directly linked to the capability of the clustering method. The more successful the clustering method is, the higher the use-value and accuracy of the GAP statistic.

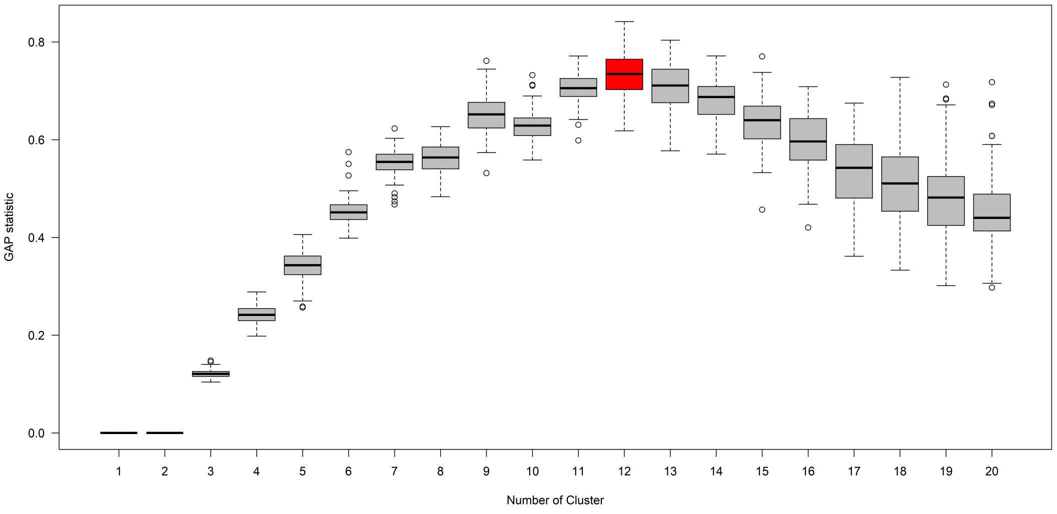


Figure 5.8 GAP statistics' Box-Plots per cluster number using TSMB-SPCL-MV.

Table 5.18 The Gap statistics peaks out of 100 repetitions using TSMB-SPCL-MV.

cluster number, $\widehat{cl}$	11	12	13	14	>15
counts of peaks	17	54	16	9	4

5.2 Performance evaluation over Real Data

In this section, the clustering performances of the proposed clustering approach and 15 other clustering methods based on five distance measures are compared on two datasets consisting of real-time multivariate time series. These datasets from real-life experiments are created for multivariate time series classification investigations, and accordingly, the class labels and sizes are known beforehand. Hence, the datasets are studied here to evaluate the capability of the proposed clustering approach for multivariate time series datasets whether it could correctly recognize underlying classes/-clusters.

There are many time series classification and clustering studies for different purposes in the literature. In addition, many datasets generated for time series classification studies are publicly available [141] for reproducibility of the results and comparison purposes. Therefore, the proposed methods for clustering tasks may vary in terms of their appropriateness for the datasets desired to be clustered and from the point of the objectives of the studies. Accordingly, assuming that any proposed clustering method could be proper for all sorts of time series datasets to be clustered is misleading. Although the clustering approach proposed in this study is not suitable for short-length time-series (i.e., ≤ 300) datasets, it appears useful in datasets where noise, non-linearity, non-stationarity, and heteroscedasticity are leading features of time series. In this context, we believe that the proposed approach can contribute and produce valuable outcomes in clustering datasets containing complex and heteroscedastic time series (i.e., $>$300 obs.) from fields such as signal processing, finance, econometrics, and computer sciences.

The number of dynamics that generate the multivariate time series datasets we study and the number of corresponding classes to these mechanisms are simply self-evident in this section. The datasets of multivariate time series we considered here consist of long signals collected from various situations with different dynamics. The first of these datasets consists of measurements of 8 different bodily movements observed by 9 sensors spread out to the whole body. The second one is consists of multivariate (i.e., 64 channel) EEG signals assembled from 7 subjects' neuronal activities.

The main reason we are practiced real-life data from the sensor-based measurements is that it is almost infeasible to find real-life time-series datasets consisting of explicit class formations and corresponding labels in the fields such as economics, finance, and climate where knowledge of underlying class formations is not accurately accessible due to uncertainty intrinsic to the dynamics of observations from these fields.

5.2.1 Clustering of Human Motion Tracking Data

The first real-life dataset examined in this section consists of observations of an experiment designed to classify several human motions (i.e., routine or sports activities). The experiment consists of recorded measurements of various bodily motions performed by 8 participants for 5 minutes. These observations are recorded by magnetic sensors placed on the subjects' bodies at five different locations (i.e., torso, right and left arms, right and left and legs). For detailed information about the experiment and the dataset, please see Altun et al. [169] and Barshan [170]. The dataset is shared with the scientific community by the researchers who designed the experiment and is publicly available from the UCI Machine Learning Repository [171].

The dataset we examined in this section for clustering consists of approximately 1-minute segments taken from 5-minute observations, each made for eight different activities. These 1-minute observations are extracted from the dataset to correspond to the middle segment of the total experimental time. In other words, the signals used from each sensor in the dataset consist of 1400 observations. Sensors are received the data at a 25 Hz sampling frequency (i.e., 25 measurements per second). Each activity (i.e., eight different activities) is recorded with nine magnetic sensors placed on the subjects' bodies at five positions. Therefore, the size of each measurement is consists of 45(9×5)-variate multivariate time series of lengths 1400 each. Accordingly, the size of the dataset to be clustered is 8 (subjects or repetition) × 8 (activities) × 1400 (length) × 45 (sensors).

In this scenario, we assumed that each bodily activity is a data generation mechanism that makes up the dataset, and therefore forms a class. The dataset, in this context, is consists of 8 classes in terms of different activities.

The structure of the dataset is presented in Table 5.19 .

Table 5.19 Multivariate Human Motion dataset.

8 Bodily Activity Number of Sources or Number of Classes		Number of Subjects or Number of Repetition	Number of sensors	Sample Length
class1:	Sitting	8	45	1400
class2:	Standing	8	45	1400
class3:	Lying	16	45	1400
class4:	As/Des-cending	16	45	1400
class5:	Walking	8	45	1400
class6:	Running	8	45	1400
class7:	Rowing	8	45	1400
class8:	Jumping	8	45	1400

- The dataset's dimension is 80×1400×45

Each participant repeated each bodily activity as a data generation mechanism eight times, except for Lying and As/Descending activities. Activities that form class1 and class2 are performed in two different forms, such as lying on the back or the right side and ascending or descending stairs. These activities are repeated 16 times. Accordingly, the dataset includes 80 multivariate time series with 45 variables each.

The subjects are asked to perform bodily movements in their way, and they are not dictated to how they should perform certain activities. Accordingly, there are some paricipant-specific differences during the performance of the movements. Here, whether the true number of classes corresponds to 8 different activities or eight different subjects can be questioned. However, since the magnetic sensors make directional measurements (i.e., on a 3-D axis), the determining factors for these observations are not the subjects' bodily structures or styles but the direction the movements themselves require to occur. Hence, participant-specific variations can be considered as processes that generate a certain amount of noise.

Please note here that the very steady motions of sitting and lying are very similar in terms of data-generating mechanisms. This situation causes that these two activities cannot be wholly recognized during the clustering phase.

Exemplars of multivariate time series belonging to each class/activity are illustrated in Figure 5.9. Although each multivariate time series consists of 45 variables, for the sake of clarity, only the time traces of 6 variables are given as an example in the graphs.

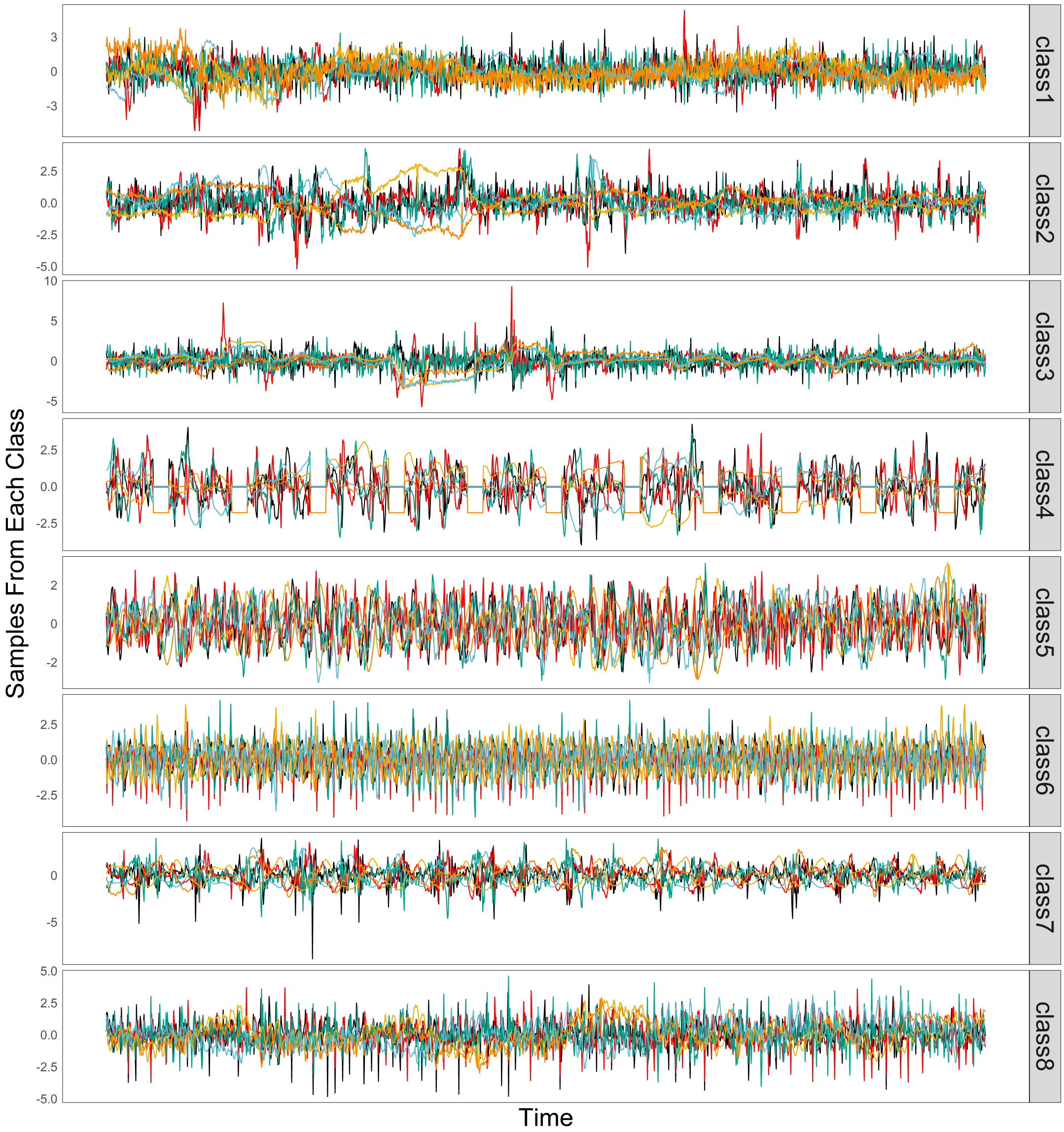


Figure 5.9 Samples from multivariate human activity dataset.

Clustering of the dataset is performed by the proposed approach given in Section 3.4.2. Implementation steps are given in Figure 3.4 and hyper-parameters values are given in Figure 4.2. **5** dissimilarity measures given in Table 5.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed on this dataset and overall clustering accuracies are displayed in Table 5.20.

Table 5.20 Multivariate Human Motion dataset: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach.

Dissimilarity measure	Clustering method		
	FUZZY	PAM	SPECTRAL
Model free			
D_{JD}	0.367	0.737	0.657
D_{LD}	0.478	0.398	0.639
D_{GAK}	0.318	0.232	0.282
D_{DTW}	0.438	0.609	0.615
D_{PDC}	0.388	0.478	0.755

	Proposed Clustering Approach
	TSMB-SPCL-MV
	$ss = 1400$
L^2 (Euclidean) Distance	0.885

Contrary to learning the classes with the supervised training in classification studies, clustering methods try to identify the underlying clusters in the dataset without any supervision. In this aspect, the results of the unsupervised completion of the proposed clustering approach in assigning the objects(i.e., multivariate time series) in the dataset to the actual clusters are presented in Table 5.20. According to these results, the proposed approach can group the objects into the correct number of clusters with an accuracy of 88$\%$. Besides, the spectral clustering application of PDC distance provided the second best result. The performance of the JD distance measure via the PAM clustering method is also obtained as the third-best result. PDC, JD, and LD distance measures are seen to provide good outcomes throughout the study, but these performances are noticed to vary from case to case.

As mentioned earlier, the proposed approach has difficulties assigning objects belonging to class1 and class3 without supervision. Another important point is that the time series objects to be clustered consist of many layers or variables, such as 45. While this may facilitate clustering, however, in some cases, it may make clustering quite tricky. Moreover, no detailed pre-processing and data manipulation is performed on the data before clustering. In fact, with such preliminary preparations, dataset clustering can be completed more carefully. Here, however, we wanted to measure and present the performance of the proposed approach in its most unrefined form.

The GAP statistic we practiced for the real-life data in this subsection reached the peak value at 8, which is the true number of clusters in the data. Figure 5.10 displays the GAP statistics vs. the possible number of clusters. Likewise, as shown in Figure 5.11, the average silhouette value also reaches its highest value at 8.

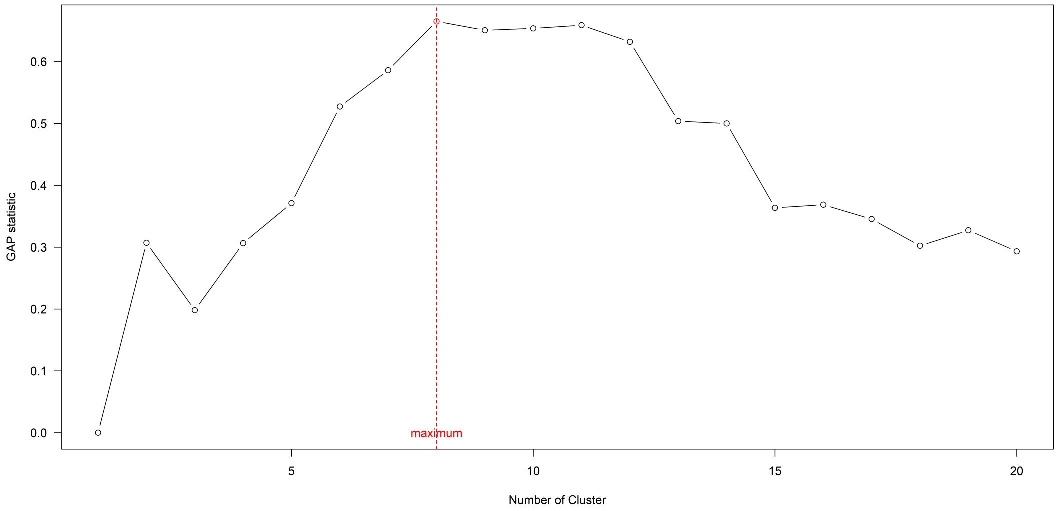


Figure 5.10 GAP statistic per cluster number for the Human Motion dataset using TSMB-SPCL-MV.

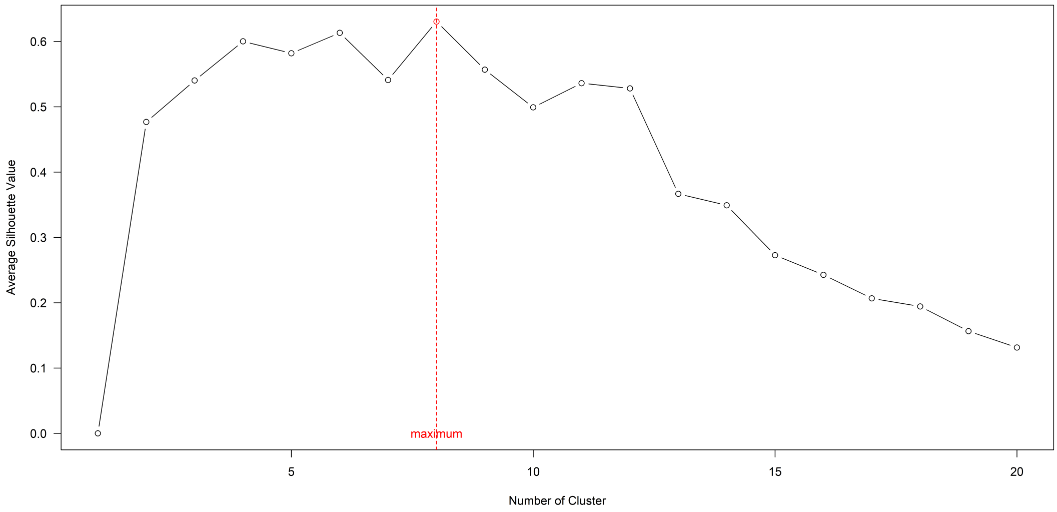


Figure 5.11 Average Silhouette Value per cluster number for the Human Motion dataset using TSMB-SPCL-MV.

5.2.2 Clustering of Multivariate EEG Motor Execution and Imagery Dataset

The second real-life dataset considered in this section is taken from the experiment conducted to contribute and developing brain-computer interface (BCI) systems. The entire dataset consists of recorded neuronal activities of 109 participants while performing a predetermined sequence of motor execution (ME) or motor imagery (MI) tasks. These neuronal activities are recorded as EEG signals and measured with the help of the BCI2000 system (i.e., 64-channel EEG). A sample scalp surface placement of 64-electrode per the international 10-10 EEG is shown in Figure 5.12. The experiment and the creation of the dataset are completed by Gerwin Schalk and his coworkers. For more detailed information about the experiment and the dataset, please see Schalk et al. [172] and visit `https://doi.org/10.13026/C28G6P`. The whole dataset is publicly available by `physionet.org`, which is an online archive that contains a huge data library of various recordings of physiologic time series for use in various research [173].

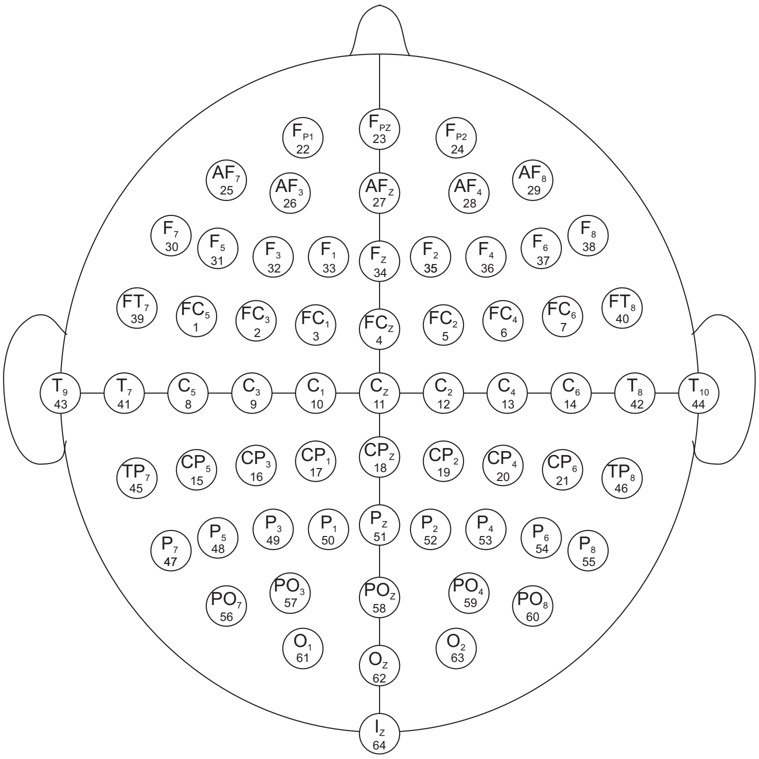


Figure 5.12 A sample of 64-channel EEG placement.

In this section, the dataset we employ to evaluate the clustering performance of the proposed approach consists of a part extracted from the entire dataset mentioned above. The extracted dataset includes EEG recordings of 7 randomly selected participants from 109 healthy participants. The experiment consists of 14 consecutive runs that each participant is required to do. These runs are of a consecutive order of MI and ME types activities. The order of an experiment runs for a participant can be given as follows:

1. Baseline, eyes open
2. Baseline, eyes closed
3. Task 1 (ME) (open and close left or right fist)
4. Task 2 (MI) (imagine opening and closing left or right fist)
5. Task 3 (ME) (open and close both fists or both feet)
6. Task 4 (MI) (imagine opening and closing both fists or both feet)
7. Task 1
8. Task 2
9. Task 3
10. Task 4
11. Task 1
12. Task 2
13. Task 3
14. Task 4

For the clustering scenario we have practiced here, a dataset consisting of 6-second segments, which is selected from the middle section and corresponds to 1024 observations in length, is taken from the neurophysiological activities measured during the tasks performed by the participants for 2 minutes. Therefore, the dataset we worked on includes 14 multivariate time series (i.e., objects to be clustered) for each 7 participants. These objects are multivariate time series with 64 variables (i.e., channels), each consisting of 1024 observations. These 64 variables are the number of electrodes used for EEG measurement, and an exemplary placement of electrodes is presented in Figure 5.12. The framework of the dataset we considered for clustering in this section is given in Table 5.21.

Table 5.21 Multivariate EEG Motor Execution and Imagery dataset.

7 Participant Number of Sources or Number of Classes		Number of Obj. in each class	Number of electrodes	Sample Length
class1/DGM1:	Subject1's neurophys.	14	64	1024
class2/DGM2:	Subject2's neurophys.	14	64	1024
class3/DGM3:	Subject3's neurophys.	14	64	1024
class4/DGM4:	Subject4's neurophys.	14	64	1024
class5/DGM5:	Subject5's neurophys.	14	64	1024
class6/DGM6:	Subject6's neurophys.	14	64	1024
class7/DGM7:	Subject7's neurophys.	14	64	1024

- The dataset's dimension is 98(7*14)×1024×64

In this framework, we assume that the actual number of clusters in the dataset we extracted from the entire dataset is equal to 7, which is also the number of participants. That is, the neurophysiology of each individual in the experiment is considered as a distinct data generation mechanism. As in the previous real dataset application, it can be questioned whether it is more accurate to cluster the dataset in terms of the number of motor tasks performed, that is, to cluster according to 6 different tasks (i.e., eyes open, eyes closed, task1, task2, task3, task4). In fact, within the scope of this framework, the actual number of clusters in the dataset is equal to the number of mechanisms (i.e., the individual neurophysiology of each participant) that generate the dataset even if each subject has done the same type of experiments. This assumption is in line with studies that have found inter-subject variations in the neurophysiologies of MI and ME type activities based on EEG signals. Saha and Baumert [174] can be seen for a detailed literature review on the intra- and inter-subject variability in brain topography.

The multivariate time series plots of the neuronal activities of each participant recorded while they performing ME task 3, only samples from 6 selected frontal electrodes' observations are visualized for the sake of clarity, are shown in Figure 5.13. As can be seen from the figure, the neuronal activities of the participants recorded during the same task differ visibly from each other.

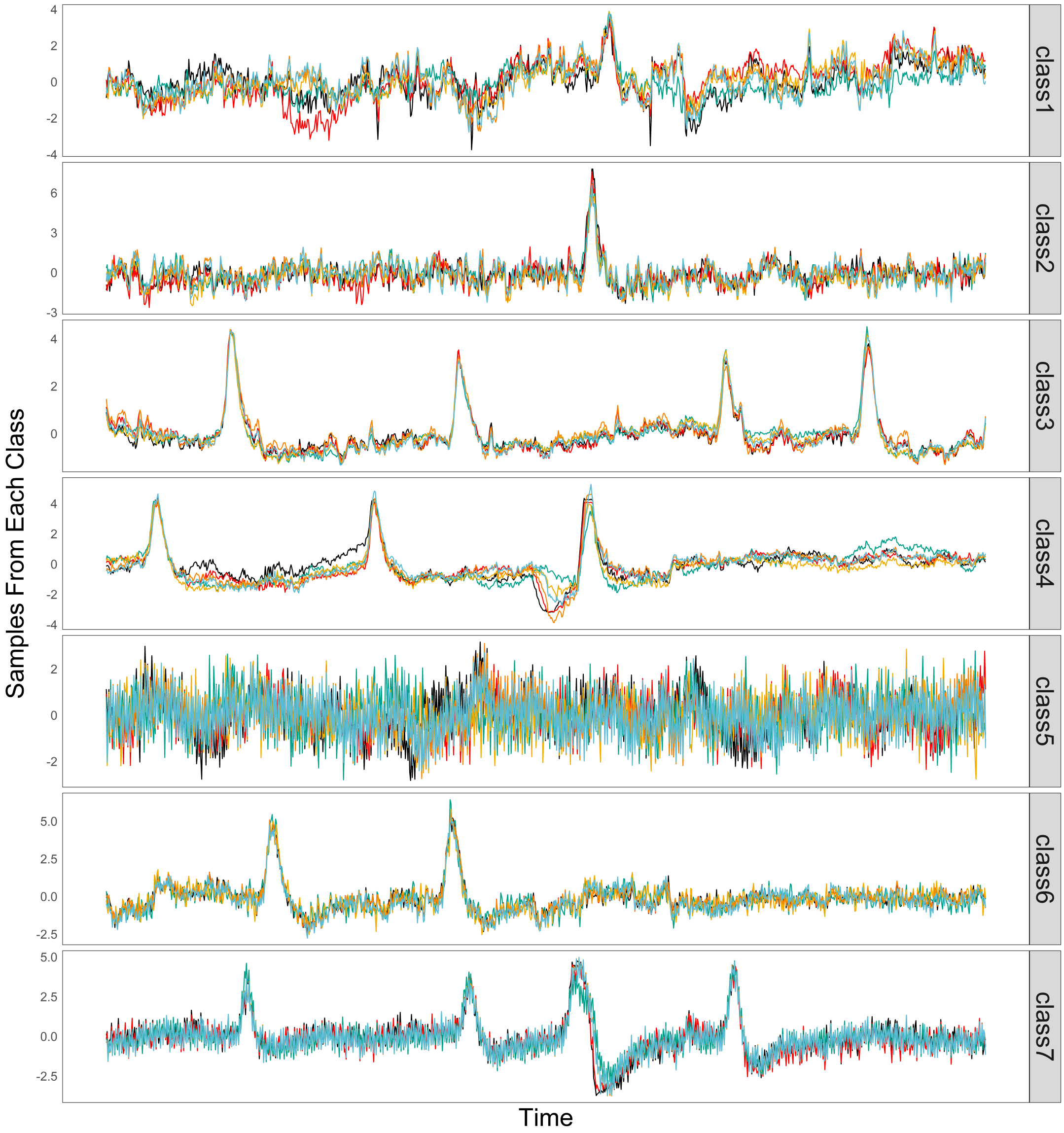


Figure 5.13 Samples from multivariate EEG motor execution and imagery dataset.

However, in such a data set, it is quite reasonable and important task to state the goal of clustering as grouping or classifying the dataset in terms of the different tasks performed. In such a case, it would be a more correct approach to apply supervised learning for classification by examining the tasks in more detail with the experimental protocol that constitutes the dataset. Because if the dataset is required to be grouped in terms of the tasks, it is essential to use the meta-information about the stages of experimental protocol and the time intervals in which each task is performed. The clustering operation we apply here is to evaluate the capacity of the proposed approach to recognize cluster structures that are already known to exist.

Clustering of the dataset is performed by the proposed approach given in Section 3.4.2. Implementation steps are given in Figure 3.4 and hyper-parameters values are given in Figure 4.2. **5** dissimilarity measures given in Table 5.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed on this dataset and overall clustering accuracies are displayed in Table 5.22.

Table 5.22 Multivariate EEG Motor Execution and Imagery dataset: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach.

Dissimilarity measure	Clustering method		
	FUZZY	PAM	SPECTRAL
Model free			
D_{JD}	0.449	0.265	0.435
D_{LD}	0.552	0.606	0.999
D_{GAK}	0.250	0.315	0.276
D_{DTW}	0.428	0.553	0.644
D_{PDC}	0.428	0.624	0.969

	Proposed Clustering Approach
	TSMB-SPCL-MV
	$ss = 1024$
L^2 (Euclidean) Distance	0.989

According to the results given in Table 5.22, the spectral clustering with the use of LD distance perfectly clustered the dataset. Besides, the performance of the PDC distance measure via the spectral clustering method is also obtained as the third-best result. The proposed approach is also clustered the dataset with very high accuracy. The proposed approach can group the 64-variate time series into the correct number of clusters with an accuracy of 98$\%$.

The results obtained in this subsection are in accordance with the findings obtained throughout the study. The proposed approach provided more stable and sometimes high-accuracy clustering results in both synthetic and real datasets. On the other hand, JD and LD distance measures operating on the frequency domain and PDC operating based on complexity information produced very high accuracy clustering results in some cases. However, the wide spectrum of performance levels obtained for the mentioned distance measures are not consistent enough to highlight these measures for every sort of dataset analyzed in the study.

The performance of these dissimilarity measures is limited and unsteady, especially in datasets with heteroscedasticity or similar complexity. Nevertheless, the results indicate that these distance measures can be quite competent if adjusted for different scenarios. On the other hand, when these distance measures are used together with spectral clustering, the high accuracies they produced are important findings in favor of the spectral clustering. This finding can also be evaluated that the spectral clustering using model-based feature extraction approach proposed in the study can be a useful alternative for univariate and multivariate time series clustering.

The GAP statistic we tested for the real-life data in this subsection reached the peak value at the cluster number 8. However, the actual number of clusters in the dataset is 7. Figure 5.14 displays the GAP statistics vs. the possible number of clusters.

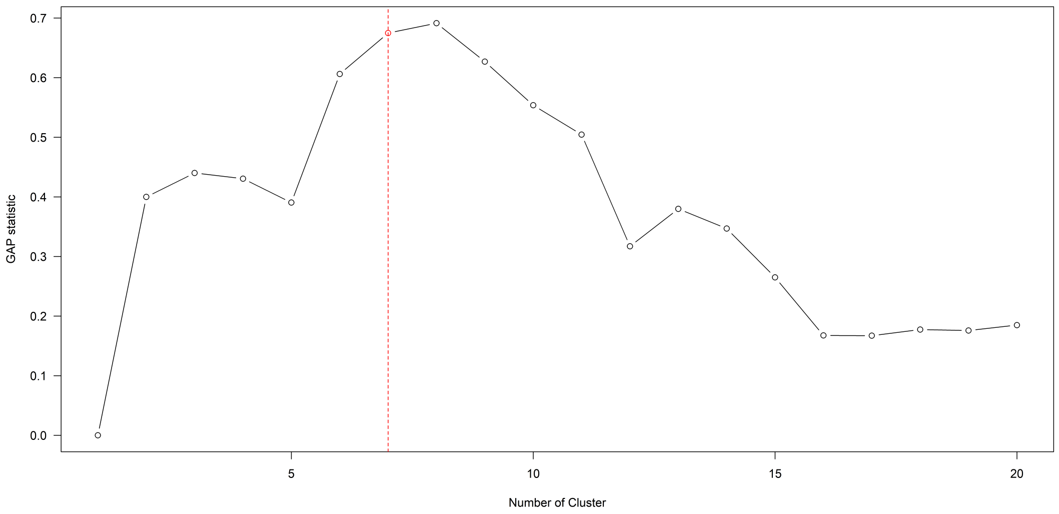


Figure 5.14 GAP statistic per cluster number for the Human Motion dataset using TSMB-SPCL-MV.

Besides, as shown in Figure 5.15, the average silhouette value reaches its highest value at 6. Both methods seems incorrectly estimated the correct number of clusters with a margin of error.

Although the main purpose of this study is not to propose a method to determine the actual number of clusters in the dataset, the GAP statistic, which we consider can be used for this purpose, has been mentioned in the previous sections.

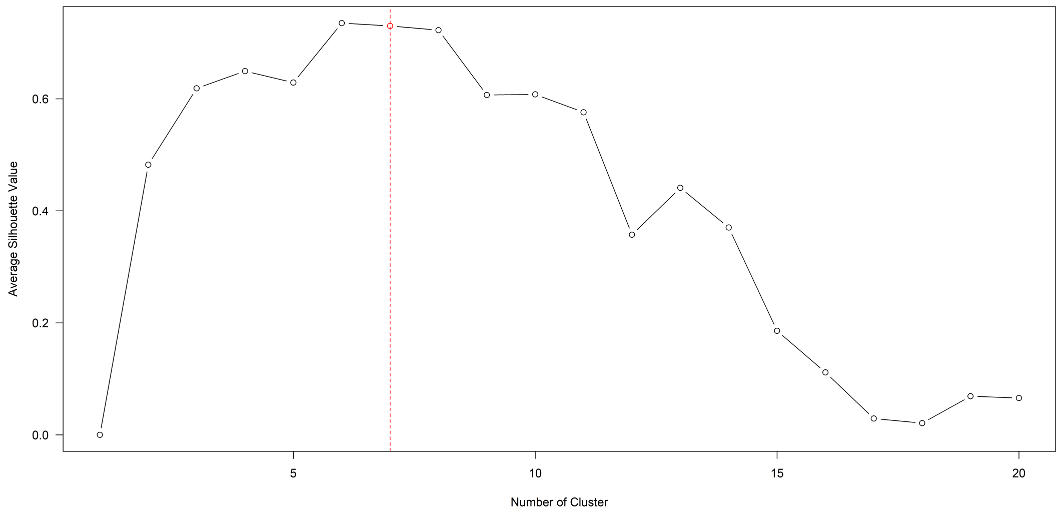


Figure 5.15 Average Silhouette Value per cluster number for the Human Motion dataset using TSMB-SPCL-MV.

There are cases where the GAP statistic is able to determine the correct number of clusters, but it also gave incorrect estimates in many trials, including synthetic datasets. As we mentioned previously, the accuracy of the GAP statistics is immediately connected to the ability of the clustering method. The more successful the clustering method is, the higher the use-value and accuracy of the GAP statistic.

However, both the GAP statistic and the Silhouette values can be used as an auxiliary indicator when the number of clusters is unknown. More importantly, as it is previously discussed that the meaning of true cluster in any dataset desired to be clustered can be established concerning the goal of the study and the number of true cluster is closely related to this. Therefore, it would be a more accurate approach to address about reasonable cluster numbers rather than the correct number of clusters in any dataset.

On the Required Computational Time for the Proposed Approach

One of the most important things for a data scientist is the time cost of a computational application. It is always aimed to design computationally inexpensive routines in statistical computing. Therefore, researchers are expected to foreknow the time they will spend on computational work as much as possible. On the other hand, algorithms used in computational-intensive jobs are desired to be suitable for existing data processing capacities.

The proposed approach use R routines. The computational algorithm of the proposed approach is self-adaptive to the configuration of the computer hardware. And it is under development for faster computation. Thanks to many R developers, the proposed approach uses a parallel computation algorithm in R. For example, clustering on a computer with 24 cores will be faster than on a device with 8 cores. However, despite this, it can be considered that the computational speed of our proposed approach is still slow.

In Table 5.23 below, the required times are given for computing the distance matrices used to cluster the real data sets in Section 5.2. In the first real data experiment given in 5.2.1, 80 multivariate time series of 1400 lengths, each with 45 variables, are clustered. In the second real data experiment given in 5.2.2, 98 multivariate time series with 1024 lengths and 64 variables each are clustered.

Table 5.23 The times required to compute the distance matrices for real data experiments.

	Real Data Experiment 5.2.1	Real Data Experiment 5.2.2
Dissimilarity measure	Human Motion Tracking Data Data size is 80×1400×45	Multivariate EEG MI-ME Data Data size is 98×1024×64
Model free		
D_{JD}	$\sim$ 35 min	$\sim$ 90 min
D_{LD}	$\sim$ 140 min	$\sim$ 360 min
D_{GAK}	$\sim$ 0.3 min	$\sim$ 0.6 min
D_{DTW}	$\sim$ 0.1 min	$\sim$ 0.4 min
D_{PDC}	$\sim$ 1 min	$\sim$ 2 min
$Proposed$ $Approach$	$\sim$ 40 min	$\sim$ 25 min

The required time for numerical calculations reported in Table 5.23 are obtained at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources). These results are obtained by the computing facility having 24 cores. As can be seen from the Table 5.23, the proposed approach requires more computation time regarding the individual time series length rather than the total data size. In particular, the R codes used for DTGARCH model estimation is still slow, even though it is designed for parallel computing. On the other hand, it is seen that the computations of distance matrices by GAK, DTW and PDC dissimilarity measures can be done very quickly. This is more about the codes used for calculations than the simplicity of distance measures. Similarly, the slowness seen in JD and LD mostly depends on the algorithm used in the calculation. Calculation of these distance measurements could also be done fast once the necessary adjustments are assembled.

One of the most remarkable results here is that high-performance results are obtained quickly, especially when the distance matrices obtained by PDC dissimilarity measures are employed in spectral clustering. The distance matrix is quickly calculated on the data sets, which can be said to be big in dimensions, and clustering could be done with an accuracy of over 75% in the first real data experiment and over 95% in the second real data experiment. One reason for this is the C language routines (known to be faster than R) that the PDC uses embedded in R code scripts. Different functions can be found or produced in programming to accomplish each operation faster. For example, one can utilize many techniques in calculations to get the Singular Value Decomposition or multivariate spectrum estimations. Being able to use these facilities highly depends on the developer's skills.

In addition to code development, significant time reductions can be achieved provided that parallel computing facilities are used. When the number of processors that can be used simultaneously is increased (e.g., in cloud computing), the time required for distance matrix calculation of the proposed approach will decrease considerably. Viewing that even today's laptops' configurations can use 64 cores, significant reductions in the computation time of statistical procedures will be performed efficiently.

CHAPTER 6

SUMMARY AND DISCUSSION

In this study, a novel approach is proposed for clustering datasets consisting of univariate or multivariate time series. The clustering performances of the proposed approach and state-of-the-art clustering methods are evaluated and compared with the datasets formed from both artificial and real-life experiments.

The clustering approach proposed in this study is for clustering problems in which time series subject to clustering are treated as objects to be clustered. On the other hand, these objects are desired to be clustered based on the similarities/dissimilarities of the mechanisms that generate these objects. For this reason, with the notion of "homogeneous/correct cluster" stated in the study, it is meant that the underlying mechanisms that create time series should be homogeneously grouped. In this context, in the introduction part of the study, some basic concepts specific to the field such as clustering, classification, similarity, true cluster, the homogeneous group are addressed with their philosophical backgrounds and historicities. The benefits of defining the true cluster concept according to the purpose and expectations of the study, and the discussions and scientific procedures developed in this context are reviewed. Therefore, the hypotheses used to develop the proposed clustering approach are addressed and attempted to be grounded in the introduction and the following two chapters.

In this study, it is believed that the idea followed for proposing a time series clustering approach is developed in accordance with the most fundamental statistical inference principle. This approach relies on clustering by distinguishing or identifying the features of the underlying data-generating mechanisms (i.e., distributions, processes, etc.). The statistical approach does not treat the observed data as isolated, spontaneous, and once-obtained observations due to uncertainty and randomness.

Statistical inferences are made by developing approaches to the mechanisms that generate the data with the help of probability distributions. Of course, the main tools for this are statistical models. When the data is time dependent, the tools that can be used are time series models. Accordingly, it is aimed to contribute to the clustering of time series by proposing the approach which can be considered in the category of model-based approaches.

Methods developed for time series clustering are briefly reviewed in Chapter 2. Some of these methods attempt to cluster time series through computational approaches, while others use model-based approaches. As stated where we mentioned motivation in the first and third chapters of the study, instead of clustering the raw observations (i.e., by evaluating the features supposed to be specific to the data), it is aimed to group the DGMs that created time series. At this point, it is necessary to note the statistical determination that the underlying DGMs that generate the data cannot be entirely determined, but only approximations can be made. Nonetheless, if the aim of clustering is stated toward this perspective, it would be pointless to find the most proper time series model for each time series and compare them with each other. Because, in the dataset to be clustered, as many suitable models can be found as the number of objects to be grouped, the clustering will become meaningless.

If the set of multiple time series objects is desired to be clustered in terms of the mechanisms that generate these series, and since the underlying DGMs cannot be fully identified, comparable outputs should be collected from the approximations to these hypothetical DGMs that produce the datasets. Therefore, a suitable tool/ruler (e.g., time series model) is required to obtain comparable approximations. It is necessary to determine the specifications of the model beforehand that are to be used for the approximation and then extract comparable information from each series through this structure specified. In this context, we attempted to illustrate the clustering perspective of the study with an analogical example from another field (i.e., spectroscopy) given in Chapter 3.1.

This perspective can be briefly summarized as follows: Since we can't determine the underlying mechanisms that generate the series with their physical structure, let's alternatively filter the realizations (i.e., observations) received from these underlying

DGMs through a model so that the products coming out of this filter are as specific as possible to the source that generates the series.

The DTGARCH model, which constitutes an essential pillar of the proposed approach and acts as a ruler or filter in the approach, is introduced in the 3rd chapter. The DTGARCH model is chosen as a suitable tool for the purpose of the proposed approach, as it has the capacity to generate and model many features encountered in real-life time series. The feature extractions using outputs obtained with the help of the DTGARCH model and made with other auxiliary devices for clustering time series are addressed in the 3rd chapter of the study.

Each series subject to clustering is represented by feature vector if series is univariate or feature matrices if multivariate. Therefore, these feature vectors/matrices are treated as objects to be clustered instead of time series. The extracted feature vectors/-matrices are expected to represent the series to be clustered in many different respects (i.e., linearity, nonlinearity, nonstationarity, autocorrelation structure, heteroscedasticity, etc.) as competently as possible. These feature vectors/matrices are consist of specific DTGARCH and AR model estimation outputs, correlation structures in raw data, and frequency-based spectrum estimations. Feature vectors/matrices, extracted through filtering using a specific model selected from the DTGARCH universe, expected to represent series can be considered as projections of the underlying mechanisms that generate the series. Subsequently, by evaluating the distances between these projections, the clustering phase can be performed. Thus, if these projections are definitely projections of the mechanisms creating the series, as a result of the accurate grouping of these projections, series close to each other in terms of underlying DGMs would be inferred.

The distance matrix, which is obtained by comparing the feature vectors/matrices, is also a representation of a graph that evocates graph theory and partitioning. If there is a connection between the series in terms of DGMs based on feature vectors/matrices, clustering can be performed throughout the algebraic examination of this distance matrix. Therefore, the proposed method for clustering the distance matrix is the spectral clustering method proposed by Ng et al. [99].

Spectral clustering provides very successful results in clustering nested connected features. Thus, we have integrated spectral clustering into time series clustering with a novel approach.

We can compile as a list what we have stated so far and our suggestions for solutions to the problems within the scope of this study, as follows,

Problems/Tasks:

1. Clustering any set of univariate or multivariate time series.
2. To consider the time-dependent variability of the clusters while clustering the time series. That is, to be able to identify if there are cluster formations that change over time.

Solutions pathway:

i Determining in which category of clustering we developed the approach and framing the notion of "true cluster" (i.e., model based feature extraction).

ii Clustering the time series in terms of the similarities of the sources from which they are created (i.e., source seperation).

iii Finding and extracting the features/footprints resulting from the underlying mechanisms by which the series are generated (i.e., several comparable feature components).

iv For this, using a model's or models' estimation outputs which have a competent peripheral (i.e., DTGARCH model).

v Accurately clustering comparable interconnected features (i.e., spectral clustering).

The experiments conducted to evaluate the performance of the proposed approach and similar methods are presented in the 4th and 5th chapters. While the studies on the clustering problem of univariate time series are given in the 4th chapter, the experiments on the clustering problem of the multivariate series are given in the 5th chapter. When we evaluate the artificial and real-life experiments' datasets and their corresponding clustering results in this study, we can state that the proposed approach provided more steady and more reliable results than other methods.

In particular, there is a definite advantage in clustering datasets containing heteroscedasticity-like phenomena. There is also substantial evidence that spectral clustering is an advantageous alternative for other distance measures. Furthermore, in these sections, the performance of the proposed approach in capturing time-varying clusters is evaluated by designing the datasets to include time-varying cluster formations.

Ultimately, the proposed clustering approach is based on a template consisting of 2 components, which we can describe as model-based feature extraction and clustering of these features with spectral clustering. A particular procedure to this template has been tested by developing it with the DTGARCH model. In this context, the proposed approach is a procedural/operational clustering approach that allows new models or combinations of models to be used in the feature extraction phase. The results obtained from the experiments support the findings for the use of model-based feature extraction approaches in clustering problems.

Although it is not directly tested whether the proposed approach is stuck in the curse of dimensionality, it can be stated that the two main procedures embedded in the proposed approach can be considered as dimension reduction facilities. Thus, the high dimensionality that prevents true clustering is relieved. First of all, specific time series model-based feature vector extraction is a form of dimension reduction. By doing this, the clustering task is accomplished in lower-dimensional feature spaces instead of using all the time-series observations assumed to be high-dimensional. That is, autocorrelation structures and residuals spectrums are not employed together while calculating dissimilarity measurements. For instance, the distances between autocorrelation structures are added to the distance matrix separately, just like other feature vectors do. Another dimension reduction is performed by the spectral clustering method, in which the distance matrix obtained from the feature vectors is decomposed into spectral features, and instead of using the multidimensional square matrix, a reduced-dimensional matrix consisting of eigenvectors is used for clustering. However, different experiments are still required to evaluate the effects of dimensionality on the proposed clustering approach. On the other hand, experiments should be arranged to determine how much loss of information will be incurred by dimension reduction.

We can list the future research spaces that can be done to improve the approach proposed here, as follows:

- To improve the response reflex to time-varying cluster formations and to improve the dynamic clustering capacity of the approach.
- To approximate underlying DGMs by using more than one time-series model at the same time. Then, to identify features with high distinctiveness, and to use these components by weighting them according to their separation performance.
- To test the proposed approach with augmented trials on real-life observations from different fields.
- Integrating computationally facilitating algorithms into the approach for very large volumes and high dimensional data.
- To evaluate cluster formation forecasting performances through time series forecasting.
- Robustification of the proposed approach should be investigated by designing different experiments to observe how the proposed approach will respond to different kinds of outliers. It can be noted that the proposed approach can eradicate the effects of outliers (i.e., abrupt shifts, structural breaks, spikes, etc.) occurring within samples due to the regime-switching structure model used in the estimation. However, cases where the time series themselves (i.e., as an object to be clustered) occur as outliers should be tested.
- Although the proposed approach does not offer any procedure for determining the model specifications, such as the number of regimes, lag orders, etc., for feature extraction, some hyperparameter values are assigned for practical purposes throughout the study. The effect of different hyperparameters on clustering performance should be evaluated. Besides, a general approach can be explored for hyperparameter preferences, although it cannot be fully automated.
- The application of the proposed approach to classification problems can be investigated.

- Different routines can be explored to increase the computational speed of the proposed approach.
- The proposed approach will not perform well on short time series due to the long data needed for estimations of the DTGARCH model used. For such problems, data augmentation alternatives can be explored and tested.
- In a publication [50] about the proposed approach, besides the Euclidean distance, we reported different distance measures' performances such as Hausdorff and DTW. Due to the reasonable performance of Euclidean distance in the proposed approach, we utilized the Euclidean distance throughout this study. However, the effects of different distance measures on the clustering performance of the proposed approach can be elaborated.

Throughout the study, we stated that the proposed approach and other methods can produce different solutions to different problems and that each approach can perform differently for context-based targets.

We hope that the clustering approach we propose establishes clustering paths/procedures based on time series model based clustering that can be developed and further improved. In this direction, we also hope that the proposed approach will contribute to time series clustering problems from different fields and enrich the research and discussions in clustering.

REFERENCES

[1] P. Pellegrin and A. Preus, *Aristotle's Classification of Animals: Biology and the Conceptual Unity of the Aristotelian Corpus*. University of California Press, 1986.

[2] E. Hamilton, H. Cairns, and L. Cooper, *The Collected Dialogues of Plato*. Bollingen Series (General), Princeton University Press, 1961.

[3] D. Mendeleev, *The Principles of Chemistry, Vol. I (first English translation from the Russian 5th edition, translated by Kemensky, G.)*. Collier: New York, 1891.

[4] E. Durkheim and M. Mauss, *Primitive Classification (Routledge Revivals)*. Routledge, 2009.

[5] E. Durkheim, *Montesquieu and Rousseau: Forerunners of Sociology, trans*. Ralph Manheim, foreword Henri Peyre (Ann Arbor, 1960), 1960.

[6] A. Secchi, "Analyse spectrale de la lumière de quelques étoiles, et nouvelles observations sur les taches solaires," *Comptes Rendus des Sances de l'Acadmie des Sciences*, vol. 63, pp. 364–368, 1866.

[7] E. C. Pickering, *The Draper Catalogue of stellar spectra photographed with the 8-inch Bache telescope as a part of the Henry Draper memorial*, vol. 27. Annals of Harvard College Observatory, 1890.

[8] W. W. Morgan, P. C. Keenan, and E. Kellman, *An atlas of stellar spectra, with an outline of spectral classification*. Chicago, Ill., The University of Chicago press [1943], 1943.

[9] P. Arabie, L. Hubert, and G. de Soete, *Clustering and Classification*. World Scientific, 1996.

[10] D. Banks, L. House, F. McMorris, P. Arabie, and W. Gaul, *Classification, Clustering, and Data Mining Applications: Proceedings of the Meeting of the International Federation of Classification Societies (IFCS), Illinois Institute of Technology, Chicago, 15–18 July 2004*. Studies in Classification, Data Analysis, and Knowledge Organization, Springer Berlin Heidelberg, 2011.

[11] V. Batagelj, H. Bock, A. Ferligoj, and A. Žiberna, *Data Science and Classification*. Studies in Classification, Data Analysis, and Knowledge Organization, Springer Berlin Heidelberg, 2006.

[12] S. Halgamuge and L. Wang, *Classification and Clustering for Knowledge Discovery*. Studies in Computational Intelligence, Springer Berlin Heidelberg, 2005.

[13] I. Van Mechelen, J. Hampton, R. S. Michalski, and P. Theuns, *Categories and concepts: Theoretical views and inductive data analysis*. Academic Press New York, 1993.

[14] L. E. Loemker, *Gottfried Wilhelm Leibniz. Philosophical Papers and Letters*. Kluwer Academic Publishers, 1989.

[15] P. Forrest, "The identity of indiscernibles," in *The Stanford encyclopedia of philosophy* (E. N. Zalta, ed.), Center for the Study of Language and Information (CSLI), Stanford University, 2008. urlhttps://plato.stanford.edu/archives/fall2008/entries/identity-indiscernible/.

[16] C. Broad, "Mctaggart's principle of the dissimilarity of the diverse," in *Proceedings of the Aristotelian Society*, vol. 32, pp. 41–52, JSTOR, 1931.

[17] D. A. Anapolitanos, *Leibniz: Representation, continuity and the spatiotemporal*, vol. 7. Springer Science & Business Media, 1999.

[18] J. L. Harrison, "Bacon's view of rhetoric, poetry, and the imagination," *Huntington Library Quarterly*, vol. 20, no. 2, pp. 107–125, 1957.

[19] S. French and M. Redhead, "Quantum physics and the identity of indiscernibles," *The British Journal for the Philosophy of Science*, vol. 39, no. 2, pp. 233–246, 1988.

[20] S. French, “Why the principle of the identity of indiscernibles is not contingently true either,” *Synthese*, vol. 78, no. 2, pp. 141–166, 1989.

[21] P. Teller, *An interpretive introduction to quantum field theory*. Princeton University Press, 1997.

[22] I. Kant, *Critique of pure reason: Unified edition (WS Pluhar, Trans.)*. Indianapolis, IN: Hackett.(Original work published 1781), 1996.

[23] G. Deleuze, *Difference and repetition*. Columbia University Press, 1994.

[24] C. Hennig, “What are the true clusters?,” *Pattern Recognition Letters*, vol. 64, pp. 53 – 62, 2015. Philosophical Aspects of Pattern Recognition.

[25] C. Hennig, “Mathematical models and reality: A constructivist perspective,” *Foundations of Science*, vol. 15, no. 1, pp. 29–48, 2010.

[26] C. Hennig and T. F. Liao, “How to find an appropriate clustering for mixed-type variables with application to socio-economic stratification,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 62, no. 3, pp. 309–369, 2013.

[27] S. Ben-David, U. von Luxburg, and D. Pál, “A sober look at clustering stability,” in *Learning Theory* (G. Lugosi and H. U. Simon, eds.), (Berlin, Heidelberg), pp. 5–19, Springer Berlin Heidelberg, 2006.

[28] U. Von Luxburg, R. C. Williamson, and I. Guyon, “Clustering: Science or art?,” in *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, pp. 65–79, 2012.

[29] P. Berkhin, *A Survey of Clustering Data Mining Techniques*, pp. 25–71. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006.

[30] A. Gelman and C. Hennig, “Beyond subjective and objective in statistics,” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 180, no. 4, pp. 967–1033, 2017.

[31] H. Chang, *Is water H2O?: Evidence, realism and pluralism*, vol. 293. Springer Science & Business Media, 2012.

[32] A. Bardon and H. Dyke, *A Companion to the Philosophy of Time*, vol. 154. John Wiley & Sons, Ltd., Publication, 2013.

[33] P. Mittelstaedt, "Cognition versus constitution of objects: From kant to modern physics," *Foundations of Physics*, vol. 39, no. 7, pp. 847–859, 2009. Cited By :3.

[34] P. Mittelstaedt and W. Riemer, *Philosophical problems of modern physics*, vol. 95. Springer, 1976.

[35] G. Deleuze, F. aslından çeviren: U. Baker, and A. Kovanlıkaya, *Kant üzerine dört ders*. Kabalcı Yayınevi, 2007.

[36] G. H. Moore, "Classification of series according to conformity and timing," in *Statistical Indicators of Cyclical Revivals and Recessions*, pp. 31–45, NBER, 1950.

[37] C. R. Rao, "Statisical inference applied to classificatory problems: Part iii: The discriminant function approach in the classification of time series," *SANKHYA: Indian Journal of Statistics*, vol. 11, no. 3-4, pp. 252–272, 1951.

[38] B. King, "Step-wise clustering procedures," *Journal of the American Statistical Association*, vol. 62, no. 317, pp. 86–101, 1967.

[39] T. Harley, "Sequential classification of nonstationary time series," *Philco-Ford Corporation, Blue Bell, Pa*, 1967.

[40] K. Ma, G. Celesia, and W. Birkemeier, "Cluster analysis and spike detection in eeg," in *Epileptology*, pp. 386–396, Thieme, Stuttgart, 1976.

[41] W. Gersch, "Kullback leibler number cluster/classification analysis of stationary time series," in *Proc. 4th Int. Joint Conf. on Pattern Recognition*, 1978.

[42] S. Watanabe, *Methodologies of pattern recognition*. Academic Press, NY, USA, 1969.

[43] S. Aghabozorgi, A. S. Shirkhorshidi, and T. Y. Wah, "Time-series clustering–a decade review," *Information Systems*, vol. 53, pp. 16–38, 2015.

[44] B. Mirkin, *Mathematical Classification and Clustering, Nonconvex Optimization and Its Applications, Volume 11.* Kluwer Academic Publishers, The Netherlands, 1996.

[45] E. Rosch and B. B. Lloyd, *Cognition and categorization.* Lawrence Erlbaum Associates Hillsdale, NJ, 1978.

[46] T. J. van Gelder, *Is Cognition Categorization?*, vol. 29 of *Psychology of Learning and Motivation - Advances in Research and Theory.* 1993.

[47] E. W. Beth, "Science and classification," *Synthese*, vol. 11, no. 3, pp. 231–244, 1959.

[48] J. Williams, *Gilles Deleuze's Difference and Repetition: A Critical Introduction and Guide: A Critical Introduction and Guide.* Edinburgh University Press, 2013.

[49] T. W. Liao, "Clustering of time series data—a survey," *Pattern recognition*, vol. 38, no. 11, pp. 1857–1874, 2005.

[50] S. Aslan, C. Yozgatligil, and C. Iyigun, "Temporal clustering of time series via threshold autoregressive models: application to commodity prices," *Annals of Operations Research*, vol. 260, no. 1-2, pp. 51–77, 2018.

[51] R. Frigg and S. Hartmann, "Models in science," in *The Stanford Encyclopedia of Philosophy* (E. N. Zalta, ed.), Metaphysics Research Lab, Stanford University, summer 2018 ed., 2018.

[52] G. E. Box, "Robustness in the strategy of scientific model building," in *Robustness in statistics*, pp. 201–236, Elsevier, 1979.

[53] M. Ali, A. Alqahtani, M. W. Jones, and X. Xie, "Clustering and classification for time series data in visual analytics: A survey," *IEEE Access*, vol. 7, pp. 181314–181338, 2019.

[54] E. A. Maharaj, P. D'Urso, and J. Caiado, *Time series clustering and classification.* Chapman and Hall/CRC, 2019.

[55] P. Montero and J. A. Vilar, "Tsclust: An r package for time series clustering," *Journal of Statistical Software,*, vol. 62, no. 1, pp. 1–43., 2014.

[56] A. M. Brandmaier, “pdc: An r package for complexity-based clustering of time series,” *Journal of Statistical Software,*, vol. 67, no. 5, pp. 1–23., 2015.

[57] A. Sardá-Espinosa, “Time-series clustering in r using the dtwclust package,” *The R Journal*, 2019.

[58] D. Piccolo, “A distance measure for classifying arima models,” *Journal of Time Series Analysis,*, vol. 11, no. 2, pp. 153–164., 1990.

[59] D. J. Berndt and J. Clifford, “Using dynamic time warping to find patterns in time series.,” in *KDD workshop*, vol. 10, pp. 359–370., Seattle, WA, 1994.

[60] D. Peña, P. Galeano, *et al.*, “Multivariate analysis in vector time series,” tech. rep., Universidad Carlos III de Madrid. Departamento de Estadística., 2001.

[61] A. D. Chouakria and P. N. Nagabhushan, “Adaptive dissimilarity index for measuring time series proximity,” *Advances in Data Analysis and Classification,*, vol. 1, no. 1, pp. 5–21., 2007.

[62] X. Golay, S. Kollias, G. Stoll, D. Meier, A. Valavanis, and P. Boesiger, “A new correlation-based fuzzy logic clustering algorithm for fmri,” *Magnetic Resonance in Medicine,*, vol. 40, no. 2, pp. 249–260., 1998.

[63] H. Zhang, T. B. Ho, Y. Zhang, and M. S. Lin, “Unsupervised feature extraction for time series clustering using orthogonal wavelet transform,” *Informatica,*, vol. 30, no. 3, pp. 305–319., 2006.

[64] J. Caiado, N. Crato, and D. Peña, “A periodogram-based metric for time series classification,” *Computational Statistics & Data Analysis,*, vol. 50, no. 10, pp. 2668–2684., 2006.

[65] D. Casado, *Classification techniques for time series and functional data*. PhD thesis, Universidad Carlos III de Madrid., 2010.

[66] Y. Kakizawa, R. H. Shumway, and M. Taniguchi, “Discrimination and clustering for multivariate time series,” *Journal of the American Statistical Association,*, vol. 93, no. 441, pp. 328–340., 1998.

[67] J. A. Vilar and S. Pértega, "Discriminant and cluster analysis for gaussian stationary processes: Local linear fitting approach," *Journal of Nonparametric Statistics,*, vol. 16, no. 3-4, pp. 443–462., 2004.

[68] J. Fan and W. Zhang, "Generalised likelihood ratio tests for spectral density," *Biometrika,*, vol. 91, no. 1, pp. 195–209., 2004.

[69] S. P. Díaz and J. A. Vilar, "Comparing several parametric and nonparametric approaches to time series clustering: a simulation study," *Journal of classification,*, vol. 27, no. 3, pp. 333–362., 2010.

[70] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, pp. 2–11., ACM, 2003.

[71] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, "Dimensionality reduction for fast similarity search in large time series databases," *Knowledge and Information Systems,*, vol. 3, no. 3, pp. 263–286., 2001.

[72] J. A. Bastos and J. Caiado, "Clustering financial time series with variance ratio statistics," *Quantitative Finance,*, vol. 14, no. 12, pp. 2121–2133., 2014.

[73] T. Giorgino *et al.*, "Computing and visualizing dynamic time warping alignments in r: the dtw package," *Journal of statistical Software,*, vol. 31, no. 7, pp. 1–24., 2009.

[74] G. E. Batista, X. Wang, and E. J. Keogh, "A complexity-invariant distance measure for time series," in *Proceedings of the 2011 SIAM International Conference on Data Mining*, pp. 699–710., SIAM, 2011.

[75] E. Keogh, S. Lonardi, and C. A. Ratanamahatana, "Towards parameter-free data mining," in *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data mining*, pp. 206–215., ACM, 2004.

[76] R. Cilibrasi and P. M. Vitányi, "Clustering by compression," *IEEE Transactions on Information theory,*, vol. 51, no. 4, pp. 1523–1545., 2005.

[77] E. A. Maharaj, “A significance test for classifying arma models,” *Journal of Statistical Computation and Simulation,*, vol. 54, no. 4, pp. 305–331., 1996.

[78] K. Kalpakis, D. Gada, and V. Puttagunta, “Distance measures for effective clustering of arima time-series,” in *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, pp. 273–280., IEEE, 2001.

[79] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control.* John Wiley & Sons, 2015.

[80] E. A. Maharaj, “Cluster of time series,” *Journal of Classification,*, vol. 17, no. 2, pp. 297–314., 2000.

[81] W. Gersch and J. Yonemoto, “Parametric time series models for multivariate eeg analysis,” *Computers and Biomedical Research*, vol. 10, no. 2, pp. 113–125, 1977.

[82] A. C. Sanderson and A. K. Wong, “Pattern trajectory analysis of nonstationary multivariate data,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 10, no. 7, pp. 384–392, 1980.

[83] M. Cuturi, “Fast global alignment kernels,” in *Proceedings of the 28th international conference on machine learning (ICML-11)*, pp. 929–936, 2011.

[84] A. Walden and L. Zhuang, “Constructing brain connectivity group graphs from eeg time series,” *Journal of Applied Statistics*, vol. 46, no. 6, pp. 1107–1128, 2019.

[85] A. Sardá-Espinosa, “Comparing time-series clustering algorithms in r using the dtwclust package,” *R package vignette*, vol. 12, p. 41, 2017.

[86] T. Roick, D. Karlis, and P. D. McNicholas, “Clustering discrete-valued time series,” *Advances in Data Analysis and Classification*, vol. 15, no. 1, pp. 209–229, 2021.

[87] S. Fröhwirth-Schnatter and S. Kaufmann, “Model-based clustering of multiple time series,” *Journal of Business & Economic Statistics*, vol. 26, no. 1, pp. 78–89, 2008.

[88] L. E. Nieto-Barajas and A. Contreras-Cristán, “A bayesian nonparametric approach for time series clustering,” *Bayesian Analysis*, vol. 9, no. 1, pp. 147–170, 2014.

[89] A. van Delft and H. Dette, “A similarity measure for second order properties of non-stationary functional time series with applications to clustering and testing,” *Bernoulli*, vol. 27, no. 1, pp. 469–501, 2021.

[90] J. G. Dias, J. K. Vermunt, and S. Ramos, “Clustering financial time series: New insights from an extended hidden markov model,” *European Journal of Operational Research*, vol. 243, no. 3, pp. 852–864, 2015.

[91] A. M. Alonso and D. Peña, “Clustering time series by linear dependency,” *Statistics and Computing*, vol. 29, no. 4, pp. 655–676, 2019.

[92] P. D’Urso, L. De Giovanni, R. Massari, R. L. D’Ecclesia, and E. A. Maharaj, “Cepstral-based clustering of financial time series,” *Expert Systems with Applications*, vol. 161, p. 113705, 2020.

[93] S. Deb, “Var model based clustering method for multivariate time series data,” *Journal of Mathematical Sciences*, vol. 237, no. 6, pp. 754–765, 2019.

[94] S. Ghassempour, F. Girosi, and A. Maeder, “Clustering multivariate time series using hidden markov models,” *International journal of environmental research and public health*, vol. 11, no. 3, pp. 2741–2763, 2014.

[95] Á. López-Oriona and J. A. Vilar, “Quantile cross-spectral density: A novel and effective tool for clustering multivariate time series,” *Expert Systems with Applications*, vol. 185, p. 115677, 2021.

[96] H. Li and M. Wei, “Fuzzy clustering based on feature weights for multivariate time series,” *Knowledge-Based Systems*, vol. 197, p. 105907, 2020.

[97] I. Vázquez, J. R. Villar, J. Sedano, S. Simić, and E. de la Cal, “An ensemble solution for multivariate time series clustering,” *Neurocomputing*, vol. 457, pp. 182–192, 2021.

[98] H. Li and Z. Liu, “Multivariate time series clustering based on complex network,” *Pattern Recognition*, vol. 115, p. 107919, 2021.

[99] A. Y. Ng, M. I. Jordan, and Y. Weiss, “On spectral clustering: Analysis and an algorithm,” in *Advances in neural information processing systems*, pp. 849–856, 2002.

[100] R. E. Quandt, “The estimation of the parameters of a linear regression system obeying two separate regimes,” *Journal of the american statistical association*, vol. 53, no. 284, pp. 873–880, 1958.

[101] H. Tong, *On a threshold model.* No. 29, Sijthoff & Noordhoff, 1978.

[102] H. Tong and K. S. Lim, “Threshold autoregression, limit cycles and cyclical data,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 245–292, 1980.

[103] H. Tong, *Non-linear time series: a dynamical system approach.* Oxford University Press, 1990.

[104] B. E. Hansen, “Threshold autoregression in economics,” *Statistics and its Interface*, vol. 4, no. 2, pp. 123–127, 2011.

[105] C. W. Chen, M. K. So, and F.-C. Liu, “A review of threshold time series models in finance,” *Statistics and Its Interface*, vol. 4, no. 2, p. 167, 2011.

[106] K.-S. Chan, “Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model,” *The annals of statistics*, pp. 520–533, 1993.

[107] B. E. Hansen, “Inference in tar models,” *Studies in nonlinear dynamics & econometrics*, vol. 2, no. 1, 1997.

[108] C. W. Chen and J. C. Lee, “Bayesian inference of threshold autoregressive models,” *Journal of Time Series Analysis*, vol. 16, no. 5, pp. 483–492, 1995.

[109] C. W. Chen, “A bayesian analysis of generalized threshold autoregressive models,” *Statistics & probability letters*, vol. 40, no. 1, pp. 15–22, 1998.

[110] C. W. Chen, T. C. Chiang, and M. K. So, “Asymmetrical reaction to us stock-return news: evidence from major stock markets based on a double-threshold model,” *Journal of Economics and Business*, vol. 55, no. 5, pp. 487–502, 2003.

[111] R. S. Tsay, “Testing and modeling threshold autoregressive processes,” *Journal of the American Statistical Association*, vol. 84, no. 405, pp. 231–240, 1989.

[112] B. E. Hansen, “Threshold effects in non-dynamic panels: Estimation, testing, and inference,” *Journal of econometrics*, vol. 93, no. 2, pp. 345–368, 1999.

[113] J. Gonzalo and J.-Y. Pitarakis, “Estimation and model selection based inference in single and multiple threshold models,” *Journal of Econometrics*, vol. 110, no. 2, pp. 319–352, 2002.

[114] A. Amendola, M. Niglio, and C. Vitale, “Statistical properties of threshold models,” *Communications in Statistics—Theory and Methods*, vol. 38, no. 15, pp. 2479–2497, 2009.

[115] C. Francq and J.-M. Zakoian, *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons, 2011.

[116] W. Härdle, H. Lütkepohl, and R. Chen, “A review of nonparametric time series analysis,” *International Statistical Review*, vol. 65, no. 1, pp. 49–72, 1997.

[117] C. Li and W. Li, “On a double-threshold autoregressive heteroscedastic time series model,” *Journal of applied econometrics*, vol. 11, no. 3, pp. 253–274, 1996.

[118] C. Brooks, “A double-threshold garch model for the french franc/deutschmark exchange rate,” *Journal of Forecasting*, vol. 20, no. 2, pp. 135–143, 2001.

[119] J. C. Escanciano, “Quasi-maximum likelihood estimation of semi-strong garch models,” *Econometric Theory*, vol. 25, no. 02, pp. 561–570, 2009.

[120] L. Bauwens, C. M. Hafner, and S. Laurent, *Handbook of volatility models and their applications*, vol. 3. John Wiley & Sons, 2012.

[121] D. B. Nelson, “Conditional heteroskedasticity in asset returns: A new approach,” *Econometrica: Journal of the Econometric Society*, pp. 347–370, 1991.

[122] Z. Ding, C. W. Granger, and R. F. Engle, “A long memory property of stock market returns and a new model,” *Journal of empirical finance*, vol. 1, no. 1, pp. 83–106, 1993.

[123] J.-M. Zakoian, “Threshold heteroskedastic models,” *Journal of Economic Dynamics and control*, vol. 18, no. 5, pp. 931–955, 1994.

[124] F. Audrino and F. Trojani, “Estimating and predicting multivariate volatility thresholds in global stock markets,” *Journal of Applied Econometrics*, vol. 21, no. 3, pp. 345–369, 2006.

[125] S. Suardi, “Central bank intervention, threshold effects and asymmetric volatility: Evidence from the japanese yen–us dollar foreign exchange market,” *Economic Modelling*, vol. 25, no. 4, pp. 628–642, 2008.

[126] U. Von Luxburg, “A tutorial on spectral clustering,” *Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.

[127] H. Walther, *Ten applications of graph theory*, vol. 7. Springer Science & Business Media, 2012.

[128] L. R. Foulds, *Graph theory applications*. Springer Science & Business Media, 2012.

[129] F. R. Chung, *Spectral graph theory (CBMS regional conference series in mathematics, No. 92)*. American Mathematical Society, 1996.

[130] S. Kurras, *Variants of the graph Laplacian with applications in machine learning*. PhD thesis, Universität Hamburg, 2017.

[131] M. Fiedler, “Algebraic connectivity of graphs,” *Czechoslovak mathematical journal*, vol. 23, no. 2, pp. 298–305, 1973.

[132] J. L. Gross, J. Yellen, and P. Zhang, *Handbook of graph theory*. Chapman and Hall/CRC, 2013.

[133] C.-E. Bichot and P. Siarry, *Graph partitioning*. John Wiley & Sons, 2013.

[134] A. Buluç, H. Meyerhenke, I. Safro, P. Sanders, and C. Schulz, “Recent advances in graph partitioning,” in *Algorithm Engineering*, pp. 117–158, Springer, 2016.

[135] W. Wei, *Time Series Analysis Univariate and Multivariate Methods*. Pearson Modern Classics for Advanced Statistics Series, Pearson Education, 2006.

[136] J. Durbin, "The fitting of time-series models," *Revue de l'Institut International de Statistique / Review of the International Statistical Institute*, vol. 28, no. 3, pp. 233–244, 1960.

[137] R. H. Shumway and D. S. Stoffer, *Time series analysis and its applications*. Springer, 2016.

[138] J. Sueur, *Sound Analysis and Synthesis with R*. Springer, 2018.

[139] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *Journal of the Royal Statistical Society: Series B,*, vol. 63, no. 2, pp. 411–423., 2001.

[140] H. Tong, "Discussion of 'an analysis of global warming in the alpine region based on nonlinear nonstationary time series models' by battaglia and protopapas," *Statistical methods & applications*, vol. 21, no. 3, pp. 335–339, 2012.

[141] D. Dua and C. Graff, "UCI machine learning repository," 2017.

[142] C. Sapsanis, G. Georgoulas, A. Tzes, and D. Lymberopoulos, "Improving emg based classification of basic hand movements using emd," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5754–5757, IEEE, 2013.

[143] A. Hardy, "On the number of clusters," *Computational Statistics & Data Analysis*, vol. 23, no. 1, pp. 83–96, 1996.

[144] C. A. Sugar and G. M. James, "Finding the number of clusters in a dataset: An information-theoretic approach," *Journal of the American Statistical Association*, vol. 98, no. 463, pp. 750–763, 2003.

[145] B. Mirkin, "Choosing the number of clusters," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 3, pp. 252–260, 2011.

[146] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.

[147] L. Kaufman and P. J. Rousseeuw, *Finding groups in data: an introduction to cluster analysis*, vol. 344. John Wiley & Sons, 2009.

[148] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Physical Review E*, vol. 64, no. 6, p. 061907, 2001.

[149] C. Elger, G. Widman, R. Andrzejak, J. Arnhold, P. David, and K. Lehnertz, "Nonlinear eeg analysis and its potential role in epileptology," *Epilepsia*, vol. 41, pp. S34–S38, 2000.

[150] K. Lehnertz, F. Mormann, T. Kreuz, R. G. Andrzejak, C. Rieke, P. David, and C. E. Elger, "Seizure prediction by nonlinear eeg analysis," *IEEE Engineering in Medicine and Biology Magazine*, vol. 22, no. 1, pp. 57–63, 2003.

[151] C. J. Stam, "Nonlinear dynamical analysis of eeg and meg: review of an emerging field," *Clinical neurophysiology*, vol. 116, no. 10, pp. 2266–2301, 2005.

[152] A. Brenner, E. Kutafina, and S. M. Jonas, "Automatic recognition of epileptiform eeg abnormalities," in *Building Continents of Knowledge in Oceans of Data: The Future of Co-Created eHealth*, pp. 171–175, IOS Press, 2018.

[153] X. Jiang, L. Ning, and T. T. Georgiou, "Distances and riemannian metrics for multivariate spectral densities," *IEEE Transactions on Automatic Control*, vol. 57, no. 7, pp. 1723–1735, 2012.

[154] M. Cuturi, J.-P. Vert, O. Birkenes, and T. Matsui, "A kernel for time series based on global alignments," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, vol. 2, pp. II–413, IEEE, 2007.

[155] A. C. Guidoum and K. Boukhetala, "Performing parallel monte carlo and moment equations methods for itô and stratonovich stochastic differential systems: R package Sim.DiffProc," *Journal of Statistical Software*, vol. 96, no. 2, pp. 1–82, 2020.

[156] E. Allen, *Modeling with Itô stochastic differential equations*, vol. 22. Springer Science & Business Media, 2007.

[157] M. Carletti, K. Burrage, and P. M. Burrage, "Numerical simulation of stochastic ordinary differential equations in biomathematical modelling," *Mathematics and Computers in Simulation*, vol. 64, no. 2, pp. 271–277, 2004.

[158] N. Berwal, D. Panchal, and C. Parihar, "Solving system of linear differential equations using haar wavelet," *Applied Mathematics and Computational Intelligence*, vol. 2, no. 2, pp. 183–193, 2013.

[159] M. Nowak and R. M. May, *Virus dynamics: mathematical principles of immunology and virology: mathematical principles of immunology and virology*. Oxford University Press, UK, 2000.

[160] W. H. Hayt Jr, J. E. Kemmerly, and S. M. Durbin, "Engineering circuit analysis (eigth edition)," 2006.

[161] D. J. Bora and A. K. Gupta, "Impact of exponent parameter value for the partition matrix on the performance of fuzzy c means algorithm," *arXiv preprint arXiv:1406.4007*, 2014.

[162] H. LUETKEPOHL and M. KRAETZIG, *Applied time series econometrics*. Cambridge University Press, 2004.

[163] B. Pfaff, *Analysis of integrated and cointegrated time series with R*. Springer Science & Business Media, 2008.

[164] A. Silvennoinen and T. Teräsvirta, "Multivariate garch models," in *Handbook of financial time series*, pp. 201–229, Springer, 2009.

[165] C. He and T. Teräsvirta, "An extended constant conditional correlation garch model and its fourth-moment structure," *Econometric Theory*, vol. 20, no. 5, pp. 904–926, 2004.

[166] C. Francq and J.-M. Zakoïan, "Estimating multivariate volatility models equation by equation," *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, pp. 613–635, 2016.

[167] Antonio, F. D. Narzo, J. L. Aznarte, and M. Stigler, *tsDyn: Time series analysis based on dynamical systems theory*, 2009. R package.

[168] T. Nakatani, *ccgarch: An R Package for Modelling Multivariate GARCH Models with Conditional Correlations*, 2014. R package version 0.2.3, URL <http://CRAN.r-project/package=ccgarch>.

[169] K. Altun, B. Barshan, and O. Tunçel, "Comparative study on classifying human activities with miniature inertial and magnetic sensors," *Pattern Recognition*, vol. 43, no. 10, pp. 3605–3620, 2010.

[170] B. Barshan and M. C. Yüksek, "Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units," *The Computer Journal*, vol. 57, no. 11, pp. 1649–1667, 2014.

[171] "Daily and Sports Activities." UCI Machine Learning Repository, 2013.

[172] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "Bci2000: a general-purpose brain-computer interface (bci) system," *IEEE Transactions on biomedical engineering*, vol. 51, no. 6, pp. 1034–1043, 2004.

[173] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[174] S. Saha and M. Baumert, "Intra-and inter-subject variability in eeg-based sensorimotor brain computer interface: a review," *Frontiers in computational neuroscience*, vol. 13, p. 87, 2020.

APPENDIX A

UNIVARIATE TIME SERIES CLUSTERING: EXPERIMENTS ON SYNTHETIC DATA

A.1 Experiment 1

A.1.1 Experiment 1: Case 3

Table A.1 Utilized specifications of mean processes in the first set of univariate experiments.

$\mathbf{STEM\_1}$: $Y_t = 0.80Y_{t-12} + \varepsilon_t + 0.70\varepsilon_{t-12}$

$\mathbf{STEM\_2}$: $Y_t = 0.90Y_{t-1} - 0.50Y_{t-2} + 0.15Y_{t-3} + \varepsilon_t - 0.20\varepsilon_{t-1} + 0.25\varepsilon_{t-2}$

$\mathbf{STEM\_3}$: $Y_t = 2.55Y_{t-1} - 2.30Y_{t-2} + 0.75Y_{t-3} + \varepsilon_t + 0.80\varepsilon_{t-1} + 0.50\varepsilon_{t-2}$

$\mathbf{STEM\_4}$: $Y_t = 0.80Y_{t-1} - 0.80Y_{t-1}(1 + e^{-10Y_{t-1}})^{-1} + \varepsilon_t$

$\mathbf{STEM\_5}$: $Y_t =$

$$\begin{cases} -0.15 + 0.85Y_{t-1} - 0.15Y_{t-2} + 0.30Y_{t-3} - 0.40Y_{t-4} + \varepsilon_t^{(1)}, & Y_{t-1} < -1.2 \\ 2.20 + 0.20Y_{t-1} - 1.70Y_{t-2} + 0.25Y_{t-3} + \varepsilon_t^{(2)}, & -1.2 < Y_{t-1} \leq 1.2 \\ 1.00 + 0.50Y_{t-1} - 1.15Y_{t-2} - 0.60Y_{t-4} + \varepsilon_t^{(3)}, & Y_{t-1} \geq 1.2 \end{cases}$$

Table A.2 Utilized specifications of error structures in the first set of univariate experiments.

$\mathbf{APARCH^{(a)}}$: $h_t^{1.5/2} = 0.50 + 0.55(|\varepsilon_{t-1}| - 0.10\varepsilon_{t-1})^{1.5/2} + 0.40h_{t-1}^{1.5/2}$

$\mathbf{APARCH^{(b)}}$: $h_t^{1.5/2} = 0.50 + 0.20(|\varepsilon_{t-2}| + 0.30\varepsilon_{t-2})^{1.5/2} + 0.75h_{t-1}^{1.5/2}$

$\mathbf{GARCH^{(a_1,a_2,a_3)}} =$

$$\begin{cases} h_t = 0.10 + 0.15\varepsilon_{t-1}^2 + 0.10\varepsilon_{t-2}^2 + 0.25h_{t-1} + 0.10h_{t-2} + 0.20h_{t-3} \\ h_t = 0.15 + 0.25\varepsilon_{t-1}^2 + 0.10\varepsilon_{t-2}^2 + 0.10h_{t-1} + 0.05h_{t-2} + 0.15h_{t-3} \\ h_t = 0.15 + 0.10\varepsilon_{t-1}^2 + 0.25\varepsilon_{t-2}^2 + 0.25h_{t-1} + 0.10h_{t-2} + 0.15h_{t-3} \end{cases}$$

$\mathbf{GARCH^{(b_1,b_2,b_3)}} =$

$$\begin{cases} h_t = 0.30 + 0.05\varepsilon_{t-1}^2 + 0.10\varepsilon_{t-2}^2 + 0.05h_{t-1} + 0.25h_{t-2} + 0.10h_{t-3} \\ h_t = 0.35 + 0.15\varepsilon_{t-1}^2 + 0.15\varepsilon_{t-2}^2 + 0.20h_{t-1} + 0.25h_{t-2} + 0.05h_{t-3} \\ h_t = 0.05 + 0.35\varepsilon_{t-1}^2 + 0.05\varepsilon_{t-2}^2 + 0.05h_{t-1} + 0.15h_{t-2} + 0.25h_{t-3} \end{cases}$$

where $\varepsilon_t = \sqrt{h_t}e_t,$ and $e_t \overset{i.i.d}{\sim} D_{sged}(0, 1)$.

Table A.3 Hypothetical DGMs used in Univariate Experiment 1: Case 3.

	used to generate time series of length 1000		
	sources of samples from 1 to 600		sources of samples from 601 to 1000
DGM01 :=	STEM_1 | $\{\varepsilon_t\} \sim$ APARCH$^{(a)}$	▶	STEM_4 | $\{\varepsilon_t\} \sim$ APARCH$^{(b)}$
DGM02 :=	STEM_1 | $\{\varepsilon_t\} \sim$ APARCH$^{(a)}$	▶	STEM_2 | $\{\varepsilon_t\} \sim$ APARCH$^{(b)}$
DGM03 :=	STEM_2 | $\{\varepsilon_t\} \sim$ APARCH$^{(a)}$	▶	STEM_3 | $\{\varepsilon_t\} \sim$ APARCH$^{(b)}$
DGM04 :=	STEM_2 | $\{\varepsilon_t\} \sim$ APARCH$^{(a)}$	▶	STEM_5 | $\{\varepsilon_t\} \sim$ GARCH$^{(b_1,b_2,b_3)}$
DGM05 :=	STEM_3 | $\{\varepsilon_t\} \sim$ APARCH$^{(a)}$	▶	STEM_1 | $\{\varepsilon_t\} \sim$ APARCH$^{(b)}$
DGM06 :=	STEM_3 | $\{\varepsilon_t\} \sim$ APARCH$^{(a)}$	▶	STEM_4 | $\{\varepsilon_t\} \sim$ APARCH$^{(b)}$
DGM07 :=	STEM_4 | $\{\varepsilon_t\} \sim$ APARCH$^{(a)}$	▶	STEM_3 | $\{\varepsilon_t\} \sim$ APARCH$^{(b)}$
DGM08 :=	STEM_4 | $\{\varepsilon_t\} \sim$ APARCH$^{(a)}$	▶	STEM_5 | $\{\varepsilon_t\} \sim$ GARCH$^{(b_1,b_2,b_3)}$
DGM09 :=	STEM_5 | $\{\varepsilon_t\} \sim$ GARCH$^{(a_1,a_2,a_3)}$	▶	STEM_1 | $\{\varepsilon_t\} \sim$ APARCH$^{(b)}$
DGM10 :=	STEM_5 | $\{\varepsilon_t\} \sim$ GARCH$^{(a_1,a_2,a_3)}$	▶	STEM_2 | $\{\varepsilon_t\} \sim$ APARCH$^{(b)}$

′▷′ indicates where the data generation mechanisms remain unchanged.
′▶′ indicates where the data generating mechanisms are changed slightly after 600th observation.

1. Each DGM in Table A.3 is used to generate **10** number of time series with the length of **1000**.

ss_sub: 1 to 600 →

2. The slice of the dataset consisting of the first 600 time points is filtered out.

3. Clustering done on this slice by the proposed approach given in Section 3.4.2 and through implementation steps given in Figure 3.4, and **12** dissimilarity measures given in Table 4.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed over **5** × **20** × **600** dimensional dataset and clustering accuracies are saved.

4. Steps from **1** to **3** repeated **100** (i.e., **rn**) times to get average accuracies of **36** (i.e., **3**∗**12**) clustering schemes and the proposed approach.

5. The average accuracy indexes of all methods considered in the study are displayed in Table A.4.

ss_oa: 1 to 1000 →

2. The dataset consisting of time series length of 1000 is taken for clustering.

3. Clustering done on the dataset by the proposed approach given in Section 3.4.2 and through implementation steps given in Figure 3.4, and **12** dissimilarity measures given in Table 4.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed over **10** × **10** × **1000** dimensional dataset and clustering accuracies are saved.

4. Steps from **1** to **3** repeated **100** (i.e., **rn**) times to get average accuracies of **36** (i.e., **3**∗**12**) clustering schemes and the proposed approach.

5. The average accuracy indexes of all methods considered in the study are displayed in Table A.4.

Table A.4 Univariate Experiment 1: Case 3 average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach.

	Clustering method					
	FUZZY		PAM		SPECTRAL	
	sample sizes (ss)		sample sizes (ss)		sample sizes (ss)	
Dissimilarity Measure	$1\ to\ 600$ $(5\ cluster)$	$1\ to\ 1000$ $(10\ cluster)$	$1\ to\ 600$ $(5\ cluster)$	$1\ to\ 1000$ $(10\ cluster)$	$1\ to\ 600$ $(5\ cluster)$	$1\ to\ 1000$ $(10\ cluster)$
Model Free						
D_{ACF}	0.920	0.443	0.857	0.761	0.874	0.796
D_{PACF}	0.663	0.331	0.917	0.724	0.891	0.825
D_{PER}	0.515	0.295	0.705	0.458	0.888	0.620
$D_{INT.PER}$	0.925	0.603	0.905	0.632	0.684	0.329
$D_{SPEC.LLR1}$	0.959	0.865	0.955	0.801	0.956	0.570
$D_{SPEC.LLR2}$	0.961	0.853	0.919	0.800	0.964	0.650
D_{GLK}	0.957	0.822	0.889	0.763	0.566	0.302
D_{VR}	0.904	0.712	0.834	0.742	0.461	0.335
Model Based						
D_{AR-LPC}	0.810	0.344	0.778	0.574	0.785	0.598
D_{AR-PIC}	0.956	0.391	0.948	0.707	0.906	0.751
D_{AR-MAH}	0.983	0.812	0.939	0.765	0.898	0.837
Complexity B.						
D_{PDC}	0.863	0.481	0.926	0.745	0.992	0.399

	Proposed Clustering Approach	
	TSMB-SPCL-UV	
	$ss\ \rightarrow\ 1\ to\ 600$ $(5\ cluster)$	$ss\ \rightarrow\ 1\ to\ 1000$ $(10\ cluster)$
L^2 Dist.	0.991	0.910

A.2 Experiment 2

A.2.1 Experiment 2: Case 1

Table A.5 Utilized specifications of mean processes in the second set of univariate experiments.

$\mathbf{STEM\_1}$: $Y_t = \quad 0.9Y_{t-1} + \varepsilon_t$

$\mathbf{STEM\_2}$: $Y_t = \quad 0.3Y_{t-1} + \varepsilon_t$

$\mathbf{STEM\_3}$: $Y_t = -0.3Y_{t-1} + \varepsilon_t$

$\mathbf{STEM\_4}$: $Y_t = -0.9Y_{t-1} + \varepsilon_t$

Table A.6 Utilized specifications of error structures in the second set of univariate experiments.

$\mathbf{GARCH^{(a)}}$: $h_t = 0.70 + 0.45\varepsilon_{t-1}^2 + 0.20h_{t-1} + 0.20h_{t-2}$

$\mathbf{GARCH^{(b)}}$: $h_t = 0.40 + 0.15\varepsilon_{t-1}^2 + 0.20h_{t-1} + 0.40h_{t-2}$

$\mathbf{GARCH^{(c)}}$: $h_t = 0.10 + 0.10\varepsilon_{t-1}^2 + 0.15\varepsilon_{t-2}^2 + 0.10h_{t-1} + 0.50h_{t-2}$

where $\varepsilon_t = \sqrt{h_t}e_t,$ and $e_t \overset{i.i.d}{\sim} D_{sged}(0,1).$

Table A.7 Hypothetical DGMs used in Univariate Experiment 2: Case 1.

$\mathbf{DGM01} \coloneqq \mathbf{STEM\_1} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(a)}}$

$\mathbf{DGM02} \coloneqq \mathbf{STEM\_2} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(a)}}$

$\mathbf{DGM03} \coloneqq \mathbf{STEM\_3} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(a)}}$

$\mathbf{DGM04} \coloneqq \mathbf{STEM\_4} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(a)}}$

$\mathbf{DGM05} \coloneqq \mathbf{STEM\_1} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(b)}}$

$\mathbf{DGM06} \coloneqq \mathbf{STEM\_2} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(b)}}$

$\mathbf{DGM07} \coloneqq \mathbf{STEM\_3} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(b)}}$

$\mathbf{DGM08} \coloneqq \mathbf{STEM\_4} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(b)}}$

$\mathbf{DGM09} \coloneqq \mathbf{STEM\_1} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(c)}}$

$\mathbf{DGM10} \coloneqq \mathbf{STEM\_2} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(c)}}$

$\mathbf{DGM11} \coloneqq \mathbf{STEM\_3} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(c)}}$

$\mathbf{DGM12} \coloneqq \mathbf{STEM\_4} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(c)}}$

1. Each DGM in Table A.7 is used to generate **10** number of time series with the length of **600** to form a corresponding cluster.
2. Clustering done by the proposed approach given in Section 3.4.2 and through implementation steps given in Figure 3.4, and **12** dissimilarity measures given in Table 4.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed over **10** $\times$ **10** $\times$ **600** dimensional dataset and clustering accuracies are saved.
3. Steps from **1** to **2** repeated **100** (i.e., **rn**) times to get average accuracies of **36** (i.e., **3** $*$ **12**) clustering schemes and the proposed approach.
4. The average accuracy indexes of all methods considered in the study are displayed for comparison in Table A.8.

Table A.8 Univariate Experiment 2 Case 1: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach.

Dissimilarity measure	Clustering method		
	FUZZY	PAM	SPECTRAL
Model free			
D_{ACF}	0.475	0.469	0.443
D_{PACF}	0.492	0.470	0.300
D_{PER}	0.276	0.430	0.394
$D_{INT.PER}$	0.473	0.471	0.452
$D_{SPEC.LLR1}$	0.482	0.483	0.473
$D_{SPEC.LLR2}$	0.479	0.485	0.461
D_{GLK}	0.472	0.475	0.453
D_{VR}	0.499	0.486	0.474
Model based			
D_{AR-LPC}	0.469	0.485	0.464
D_{AR-MAH}	0.472	0.489	0.465
D_{AR-PIC}	0.488	0.481	0.466
Complexity based			
D_{PDC}	0.496	0.478	0.300

	Proposed Clustering Approach
	TSMB-SPCL-UV
	$ss = 600$
L^2 (Euclidean) Distance	0.760

A.2.2 Experiment 2: Case 2

Table A.9 Utilized specifications of mean processes in the second set of univariate experiments.

$\mathbf{STEM\_1}$: $Y_t = 0.9Y_{t-1} + \varepsilon_t$

$\mathbf{STEM\_2}$: $Y_t = 0.3Y_{t-1} + \varepsilon_t$

$\mathbf{STEM\_3}$: $Y_t = -0.3Y_{t-1} + \varepsilon_t$

$\mathbf{STEM\_4}$: $Y_t = -0.9Y_{t-1} + \varepsilon_t$

Table A.10 Utilized specifications of error structures in the second set of univariate experiments.

$\mathbf{GARCH^{(a)}}$: $h_t = 0.70 + 0.45\varepsilon_{t-1}^2 + 0.20h_{t-1} + 0.20h_{t-2}$

$\mathbf{GARCH^{(b)}}$: $h_t = 0.40 + 0.15\varepsilon_{t-1}^2 + 0.20h_{t-1} + 0.40h_{t-2}$

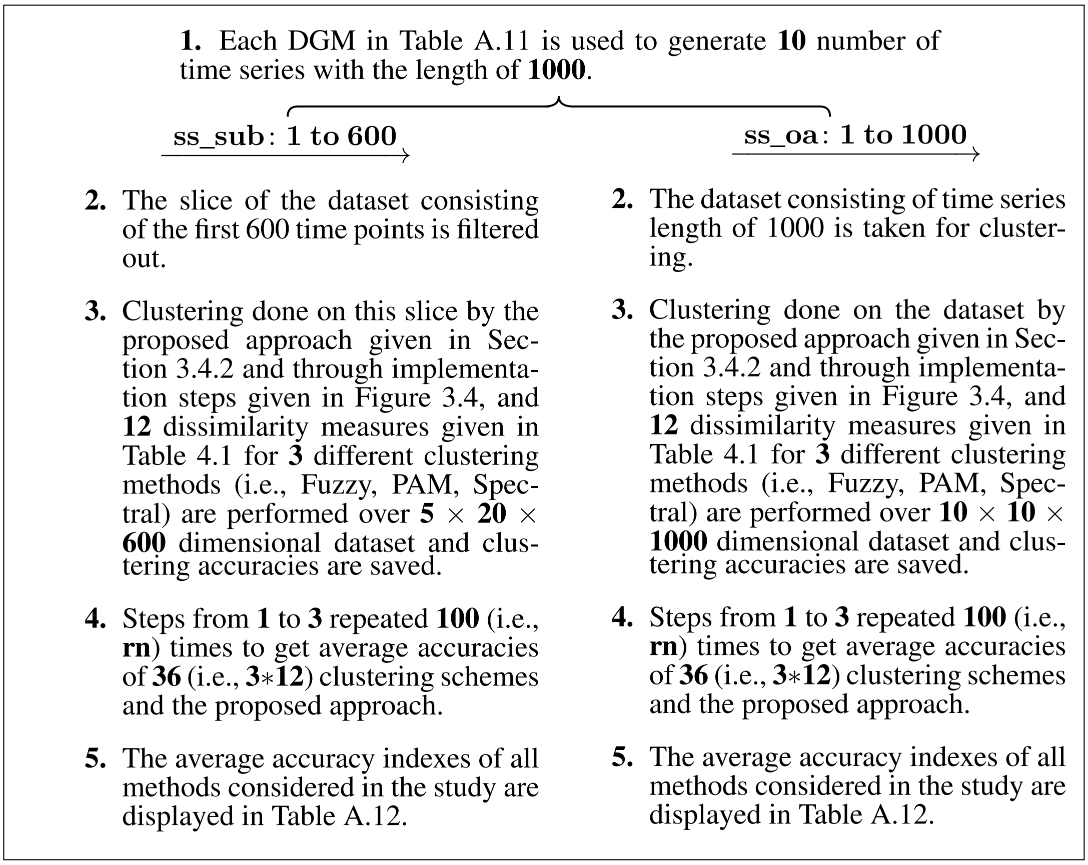
$\mathbf{GARCH^{(c)}}$: $h_t = 0.10 + 0.10\varepsilon_{t-1}^2 + 0.15\varepsilon_{t-2}^2 + 0.10h_{t-1} + 0.50h_{t-2}$

where $\varepsilon_t = \sqrt{h_t}e_t,$ and $e_t \overset{i.i.d}{\sim} D_{sged}(0, 1).$

Table A.11 Hypothetical DGMs used in Univariate Experiment 2: Case 2.

	used to generate time series of length 1000		
	sources of samples from 1 to 600		sources of samples from 601 to 1000
$\mathbf{DGM01} \coloneqq$	$\mathbf{STEM\_1} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(a)}}$	$\triangleright$	$\mathbf{STEM\_1} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(a)}}$
$\mathbf{DGM02} \coloneqq$	$\mathbf{STEM\_1} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(a)}}$	$\blacktriangleright$	$\mathbf{STEM\_2} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(a)}}$
$\mathbf{DGM03} \coloneqq$	$\mathbf{STEM\_3} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(a)}}$	$\triangleright$	$\mathbf{STEM\_3} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(a)}}$
$\mathbf{DGM04} \coloneqq$	$\mathbf{STEM\_3} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(a)}}$	$\blacktriangleright$	$\mathbf{STEM\_4} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(a)}}$
$\mathbf{DGM05} \coloneqq$	$\mathbf{STEM\_1} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(b)}}$	$\triangleright$	$\mathbf{STEM\_1} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(b)}}$
$\mathbf{DGM06} \coloneqq$	$\mathbf{STEM\_1} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(b)}}$	$\blacktriangleright$	$\mathbf{STEM\_2} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(b)}}$
$\mathbf{DGM07} \coloneqq$	$\mathbf{STEM\_3} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(b)}}$	$\triangleright$	$\mathbf{STEM\_3} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(b)}}$
$\mathbf{DGM08} \coloneqq$	$\mathbf{STEM\_3} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(b)}}$	$\blacktriangleright$	$\mathbf{STEM\_4} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(b)}}$
$\mathbf{DGM09} \coloneqq$	$\mathbf{STEM\_1} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(c)}}$	$\triangleright$	$\mathbf{STEM\_1} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(c)}}$
$\mathbf{DGM10} \coloneqq$	$\mathbf{STEM\_1} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(c)}}$	$\blacktriangleright$	$\mathbf{STEM\_2} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(c)}}$
$\mathbf{DGM11} \coloneqq$	$\mathbf{STEM\_3} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(c)}}$	$\triangleright$	$\mathbf{STEM\_3} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(c)}}$
$\mathbf{DGM12} \coloneqq$	$\mathbf{STEM\_3} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(c)}}$	$\blacktriangleright$	$\mathbf{STEM\_4} \mid \{\varepsilon_t\} \sim \mathbf{GARCH^{(c)}}$

$'\triangleright'$ indicates where the data generation mechanisms remain unchanged.
$'\blacktriangleright'$ indicates where the data generating mechanisms are changed slightly after 600th observation.



1. Each DGM in Table A.11 is used to generate **10** number of time series with the length of **1000**.

$\xrightarrow{\textbf{ss\_sub: 1 to 600}}$

2. The slice of the dataset consisting of the first 600 time points is filtered out.

3. Clustering done on this slice by the proposed approach given in Section 3.4.2 and through implementation steps given in Figure 3.4, and **12** dissimilarity measures given in Table 4.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed over **5** × **20** × **600** dimensional dataset and clustering accuracies are saved.

4. Steps from **1** to **3** repeated **100** (i.e., **rn**) times to get average accuracies of **36** (i.e., **3**∗**12**) clustering schemes and the proposed approach.

5. The average accuracy indexes of all methods considered in the study are displayed in Table A.12.

$\xrightarrow{\textbf{ss\_oa: 1 to 1000}}$

2. The dataset consisting of time series length of 1000 is taken for clustering.

3. Clustering done on the dataset by the proposed approach given in Section 3.4.2 and through implementation steps given in Figure 3.4, and **12** dissimilarity measures given in Table 4.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed over **10** × **10** × **1000** dimensional dataset and clustering accuracies are saved.

4. Steps from **1** to **3** repeated **100** (i.e., **rn**) times to get average accuracies of **36** (i.e., **3**∗**12**) clustering schemes and the proposed approach.

5. The average accuracy indexes of all methods considered in the study are displayed in Table A.12.

Table A.12 Univariate Experiment 2: Case 2 average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach.

Dissimilarity Measure	Clustering method					
	FUZZY		PAM		SPECTRAL	
	sample sizes (ss)		sample sizes (ss)		sample sizes (ss)	
	$1\ to\ 600$ $(6\ cluster)$	$1\ to\ 1000$ $(12\ cluster)$	$1\ to\ 600$ $(6\ cluster)$	$1\ to\ 1000$ $(12\ cluster)$	$1\ to\ 600$ $(6\ cluster)$	$1\ to\ 1000$ $(12\ cluster)$
Model Free						
D_{ACF}	0.475	0.470	0.441	0.458	0.438	0.405
D_{PACF}	0.487	0.499	0.430	0.463	0.425	0.276
D_{PER}	0.499	0.285	0.444	0.390	0.438	0.381
$D_{INT.PER}$	0.439	0.472	0.439	0.469	0.436	0.375
$D_{SPEC.LLR1}$	0.439	0.481	0.445	0.475	0.428	0.397
$D_{SPEC.LLR2}$	0.447	0.481	0.448	0.476	0.428	0.378
D_{GLK}	0.432	0.470	0.440	0.473	0.440	0.431
D_{VR}	0.462	0.500	0.461	0.502	0.463	0.422
Model Based						
D_{AR-LPC}	0.464	0.471	0.482	0.447	0.442	0.456
D_{AR-PIC}	0.438	0.480	0.472	0.481	0.440	0.431
D_{AR-MAH}	0.470	0.492	0.458	0.481	0.464	0.396
Complexity B.						
D_{PDC}	0.491	0.392	0.440	0.454	0.452	0.279

	Proposed Clustering Approach	
	TSMB-SPCL-UV	
	$ss \rightarrow 1\ to\ 600$ $(6\ cluster)$	$ss \rightarrow 1\ to\ 1000$ $(12\ cluster)$
L^2 Dist.	0.730	0.778

A.3 Experiment 3

A.3.1 Experiment 3: Case 1

Table A.13 Utilized specifications of mean processes in the third set of univariate experiments.

$\textbf{STEM\_1}$: $Y_t = 0.8Y_{t-1} + \varepsilon_t$

$\textbf{STEM\_2}$: $Y_t = 0.4Y_{t-1} + \varepsilon_t$

$\textbf{STEM\_3}$: $Y_t = 0.8Y_{t-1} - 0.8Y_{t-1}(1 + e^{-10Y_{t-1}})^{-1} + \varepsilon_t$

$\textbf{STEM\_4}$: $Y_t = 0.8Y_{t-1} - 0.4Y_{t-1}(1 + e^{-10Y_{t-1}})^{-1} + \varepsilon_t$

Table A.14 Utilized specifications of error structures in the third set of univariate experiments.

$\textbf{APARCH}^{(\mathbf{a})}$:
$h_t^{3/2} = 0.50 + 0.05(|\varepsilon_{t-1}| - 0.10\varepsilon_{t-1})^{3/2} + 0.55(|\varepsilon_{t-2}| + 0.10\varepsilon_{t-2})^{3/2} + 0.35h_{t-1}^{3/2}$

$\textbf{APARCH}^{(\mathbf{b})}$:
$h_t^{3/2} = 0.50 + 0.55(|\varepsilon_{t-1}| - 0.10\varepsilon_{t-1})^{3/2} + 0.05(|\varepsilon_{t-2}| + 0.10\varepsilon_{t-2})^{3/2} + 0.35h_{t-1}^{3/2}$

where $\varepsilon_t = \sqrt{h_t}e_t$, and $e_t \overset{i.i.d}{\sim} D_{sged}(0, 1)$.

Table A.15 Hypothetical DGMs used in Univariate Experiment 3: Case 1.

$\textbf{DGM01} \coloneqq \textbf{STEM\_1} \mid \{\varepsilon_t\} \sim \textbf{APARCH}^{(\mathbf{a})}$

$\textbf{DGM02} \coloneqq \textbf{STEM\_1} \mid \{\varepsilon_t\} \sim \textbf{APARCH}^{(\mathbf{b})}$

$\textbf{DGM03} \coloneqq \textbf{STEM\_3} \mid \{\varepsilon_t\} \sim \textbf{APARCH}^{(\mathbf{b})}$

$\textbf{DGM04} \coloneqq \textbf{STEM\_3} \mid \{\varepsilon_t\} \sim \textbf{APARCH}^{(\mathbf{a})}$

1. **Each** DGM in Table A.15 is used to generate **20** number of time series with the length of **600** to form a corresponding cluster.
2. **Clustering** done by the proposed approach given in Section 3.4.2 and through implementation steps given in Figure 3.4, and **12** dissimilarity measures given in Table 4.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed over **4** × **20** × **600** dimensional dataset and clustering accuracies are saved.
3. **Steps** from **1** to **2** repeated **100** times to get average accuracies of **36** (i.e., **3** $*$ **12**) clustering schemes and the proposed approach.
4. **The** average accuracy indexes of all methods considered in the study are displayed for comparison in Table A.16.

Table A.16 Univariate Experiment 3 Case 1: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach.

Dissimilarity measure	Clustering method		
	FUZZY	PAM	SPECTRAL
Model free			
D_{ACF}	0.486	0.441	0.450
D_{PACF}	0.535	0.436	0.479
D_{PER}	0.469	0.400	0.430
$D_{INT.PER}$	0.507	0.506	0.499
$D_{SPEC.LLR1}$	0.502	0.516	0.489
$D_{SPEC.LLR2}$	0.509	0.516	0.489
D_{GLK}	0.502	0.507	0.460
D_{VR}	0.544	0.559	0.390
Model based			
D_{AR-LPC}	0.475	0.496	0.504
D_{AR-MAH}	0.511	0.500	0.454
D_{AR-PIC}	0.542	0.545	0.497
Complexity based			
D_{PDC}	0.624	0.527	0.718

	Proposed Clustering Approach
	TSMB-SPCL-UV
	$ss = 600$
L^2 (Euclidean) Distance	0.800

A.3.2 Experiment 3: Case 2

Table A.17 Utilized specifications of mean processes in the third set of univariate experiments.

STEM_1: $Y_t = 0.8Y_{t-1} + \varepsilon_t$

STEM_2: $Y_t = 0.4Y_{t-1} + \varepsilon_t$

STEM_3: $Y_t = 0.8Y_{t-1} - 0.8Y_{t-1}(1 + e^{-10Y_{t-1}})^{-1} + \varepsilon_t$

STEM_4: $Y_t = 0.8Y_{t-1} - 0.4Y_{t-1}(1 + e^{-10Y_{t-1}})^{-1} + \varepsilon_t$

Table A.18 Utilized specifications of error structures in the third set of univariate experiments.

$\mathbf{APARCH}^{(\mathbf{a})}$:

$$h_t^{3/2} = 0.50 + 0.05(|\varepsilon_{t-1}| - 0.10\varepsilon_{t-1})^{3/2} + 0.55(|\varepsilon_{t-2}| + 0.10\varepsilon_{t-2})^{3/2} + 0.35h_{t-1}^{3/2}$$

$\mathbf{APARCH}^{(\mathbf{b})}$:

$$h_t^{3/2} = 0.50 + 0.55(|\varepsilon_{t-1}| - 0.10\varepsilon_{t-1})^{3/2} + 0.05(|\varepsilon_{t-2}| + 0.10\varepsilon_{t-2})^{3/2} + 0.35h_{t-1}^{3/2}$$

where $\varepsilon_t = \sqrt{h_t}e_t,$ and $e_t \overset{i.i.d}{\sim} D_{sged}(0,1).$

Table A.19 Hypothetical DGMs used in Univariate Experiment 3: Case 2.

	used to generate time series of length 1000		
	sources of samples from 1 to 600		sources of samples from 601 to 1000
$\mathbf{DGM01} \coloneqq$	$\mathbf{STEM\_1} \mid \{\varepsilon_t\} \sim \mathbf{APARCH}^{(\mathbf{a})}$	$\triangleright$	$\mathbf{STEM\_1} \mid \{\varepsilon_t\} \sim \mathbf{APARCH}^{(\mathbf{a})}$
$\mathbf{DGM02} \coloneqq$	$\mathbf{STEM\_1} \mid \{\varepsilon_t\} \sim \mathbf{APARCH}^{(\mathbf{a})}$	$\blacktriangleright$	$\mathbf{STEM\_2} \mid \{\varepsilon_t\} \sim \mathbf{APARCH}^{(\mathbf{b})}$
$\mathbf{DGM03} \coloneqq$	$\mathbf{STEM\_1} \mid \{\varepsilon_t\} \sim \mathbf{APARCH}^{(\mathbf{b})}$	$\triangleright$	$\mathbf{STEM\_1} \mid \{\varepsilon_t\} \sim \mathbf{APARCH}^{(\mathbf{b})}$
$\mathbf{DGM04} \coloneqq$	$\mathbf{STEM\_1} \mid \{\varepsilon_t\} \sim \mathbf{APARCH}^{(\mathbf{b})}$	$\blacktriangleright$	$\mathbf{STEM\_2} \mid \{\varepsilon_t\} \sim \mathbf{APARCH}^{(\mathbf{a})}$
$\mathbf{DGM05} \coloneqq$	$\mathbf{STEM\_3} \mid \{\varepsilon_t\} \sim \mathbf{APARCH}^{(\mathbf{a})}$	$\triangleright$	$\mathbf{STEM\_3} \mid \{\varepsilon_t\} \sim \mathbf{APARCH}^{(\mathbf{a})}$
$\mathbf{DGM06} \coloneqq$	$\mathbf{STEM\_3} \mid \{\varepsilon_t\} \sim \mathbf{APARCH}^{(\mathbf{a})}$	$\blacktriangleright$	$\mathbf{STEM\_4} \mid \{\varepsilon_t\} \sim \mathbf{APARCH}^{(\mathbf{b})}$
$\mathbf{DGM07} \coloneqq$	$\mathbf{STEM\_3} \mid \{\varepsilon_t\} \sim \mathbf{APARCH}^{(\mathbf{b})}$	$\triangleright$	$\mathbf{STEM\_3} \mid \{\varepsilon_t\} \sim \mathbf{APARCH}^{(\mathbf{b})}$
$\mathbf{DGM08} \coloneqq$	$\mathbf{STEM\_3} \mid \{\varepsilon_t\} \sim \mathbf{APARCH}^{(\mathbf{b})}$	$\blacktriangleright$	$\mathbf{STEM\_4} \mid \{\varepsilon_t\} \sim \mathbf{APARCH}^{(\mathbf{a})}$

$'\triangleright'$ indicates where the data generation mechanisms remain unchanged.
$'\blacktriangleright'$ indicates where the data generating mechanisms are changed slightly after 600th observation.

1. Each DGM in Table A.19 is used to generate **10** number of time series with the length of **1000**.

$\xrightarrow{\mathbf{ss\_sub}:\ \mathbf{1\ to\ 600}}$

2. The slice of the dataset consisting of the first 600 time points is filtered out.

3. Clustering done on this slice by the proposed approach given in Section 3.4.2 and through implementation steps given in Figure 3.4, and **12** dissimilarity measures given in Table 4.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed over **4** × **20** × **600** dimensional dataset and clustering accuracies are saved.

4. Steps from **1** to **3** repeated **100** (i.e., **rn**) times to get average accuracies of **36** (i.e., **3**∗**12**) clustering schemes and the proposed approach.

5. The average accuracy indexes of all methods considered in the study are displayed in Table A.20.

$\xrightarrow{\mathbf{ss\_oa}:\ \mathbf{1\ to\ 1000}}$

2. The dataset consisting of time series length of 1000 is taken for clustering.

3. Clustering done on the dataset by the proposed approach given in Section 3.4.2 and through implementation steps given in Figure 3.4, and **12** dissimilarity measures given in Table 4.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed over **8** × **10** × **1000** dimensional dataset and clustering accuracies are saved.

4. Steps from **1** to **3** repeated **100** (i.e., **rn**) times to get average accuracies of **36** (i.e., **3**∗**12**) clustering schemes and the proposed approach.

5. The average accuracy indexes of all methods considered in the study are displayed in Table A.20.

Table A.20 Univariate Experiment 3: Case 2 average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach.

	Clustering method					
	FUZZY		PAM		SPECTRAL	
	sample sizes (ss)		sample sizes (ss)		sample sizes (ss)	
Dissimilarity Measure	1 *to* 600 $(4\ cluster)$	1 *to* 1000 $(8\ cluster)$	1 *to* 600 $(4\ cluster)$	1 *to* 1000 $(8\ cluster)$	1 *to* 600 $(4\ cluster)$	1 *to* 1000 $(8\ cluster)$
Model Free						
D_{ACF}	0.486	0.310	0.441	0.341	0.450	0.292
D_{PACF}	0.535	0.340	0.436	0.337	0.479	0.310
D_{PER}	0.469	0.305	0.400	0.276	0.430	0.284
$D_{INT.PER}$	0.507	0.377	0.506	0.380	0.499	0.329
$D_{SPEC.LLR1}$	0.502	0.374	0.516	0.409	0.489	0.309
$D_{SPEC.LLR2}$	0.509	0.369	0.516	0.406	0.489	0.312
D_{GLK}	0.502	0.369	0.507	0.383	0.460	0.326
D_{VR}	0.544	0.390	0.559	0.414	0.390	0.285
Model Based						
D_{AR-LPC}	0.475	0.324	0.496	0.337	0.504	0.353
D_{AR-PIC}	0.511	0.319	0.500	0.359	0.454	0.323
D_{AR-MAH}	0.542	0.384	0.545	0.379	0.497	0.329
Complexity B.						
D_{PDC}	0.624	0.374	0.527	0.404	0.718	0.322

	Proposed Clustering Approach	
	TSMB-SPCL-UV	
	$ss\ \rightarrow\ 1$ *to* 600 $(4\ cluster)$	$ss\ \rightarrow\ 1$ *to* 1000 $(8\ cluster)$
L^2 Dist.	0.800	0.584

A.4 Experiment 4

A.4.1 Experiment 4: Case 1

Table A.21 Utilized specifications of mean processes in the fourth set of univariate experiments.

$\mathbf{STEM\_01}$: $Y_t = 0.9Y_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim D_{snorm}(0, 1)$.

$\mathbf{STEM\_02}$: $Y_t = 0.7Y_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim D_{snorm}(0, 1)$.

$\mathbf{STEM\_03}$: $Y_t = 0.5Y_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim D_{snorm}(0, 1)$.

$\mathbf{STEM\_04}$: $Y_t = 0.3Y_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim D_{snorm}(0, 1)$.

$\mathbf{STEM\_05}$: $Y_t = 0.1Y_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim D_{snorm}(0, 1)$.

$\mathbf{STEM\_06}$: $Y_t = -0.1Y_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim D_{snorm}(0, 1)$.

$\mathbf{STEM\_07}$: $Y_t = -0.3Y_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim D_{snorm}(0, 1)$.

$\mathbf{STEM\_08}$: $Y_t = -0.5Y_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim D_{snorm}(0, 1)$.

$\mathbf{STEM\_09}$: $Y_t = -0.7Y_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim D_{snorm}(0, 1)$.

$\mathbf{STEM\_10}$: $Y_t = -0.9Y_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim D_{snorm}(0, 1)$.

Table A.22 Hypothetical DGMs used in Univariate Experiment 4: Case 1.

$\mathbf{DGM01} \coloneqq \mathbf{STEM\_01} \mid \{\varepsilon_t\} \sim \mathbf{D_{snorm}(0, 1)}$

$\mathbf{DGM02} \coloneqq \mathbf{STEM\_03} \mid \{\varepsilon_t\} \sim \mathbf{D_{snorm}(0, 1)}$

$\mathbf{DGM03} \coloneqq \mathbf{STEM\_05} \mid \{\varepsilon_t\} \sim \mathbf{D_{snorm}(0, 1)}$

$\mathbf{DGM04} \coloneqq \mathbf{STEM\_07} \mid \{\varepsilon_t\} \sim \mathbf{D_{snorm}(0, 1)}$

$\mathbf{DGM05} \coloneqq \mathbf{STEM\_09} \mid \{\varepsilon_t\} \sim \mathbf{D_{snorm}(0, 1)}$

1. Each DGM in Table A.22 is used to generate **20** number of time series with the length of **600** to form a corresponding cluster.
2. Clustering done by the proposed approach given in Section 3.4.2 and through implementation steps given in Figure 3.4, and **12** dissimilarity measures given in Table 4.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed over $\mathbf{5} \times \mathbf{20} \times \mathbf{600}$ dimensional dataset and clustering accuracies are saved.
3. Steps from **1** to **2** repeated **100** times to get average accuracies of **36** (i.e., $\mathbf{3} * \mathbf{12}$) clustering schemes and the proposed approach.
4. The average accuracy indexes of all methods considered in the study are displayed for comparison in Table A.23.

Table A.23 Univariate Experiment 4 Case 1: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach.

Dissimilarity measure	Clustering method		
	FUZZY	PAM	SPECTRAL
Model free			
D_{ACF}	0.734	0.947	0.983
D_{PACF}	0.960	0.999	1.000
D_{PER}	0.686	0.446	0.959
$D_{INT.PER}$	0.999	0.999	0.893
$D_{SPEC.LLR1}$	0.999	0.999	0.899
$D_{SPEC.LLR2}$	0.999	1.000	0.895
D_{GLK}	0.999	0.998	0.988
D_{VR}	0.999	0.999	0.873
Model based			
D_{AR-LPC}	0.982	0.982	0.982
D_{AR-MAH}	0.999	0.999	1.000
D_{AR-PIC}	1.000	1.000	1.000
Complexity based			
D_{PDC}	0.969	0.956	0.870

	Proposed Clustering Approach
	TSMB-SPCL-UV
	$ss = 600$
L^2 (Euclidean) Distance	1.000

A.4.2 Experiment 4: Case 2

Table A.24 Utilized specifications of mean processes in the fourth set of univariate experiments.

STEM_01: $Y_t = 0.9Y_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim D_{snorm}(0, 1)$.

STEM_02: $Y_t = 0.7Y_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim D_{snorm}(0, 1)$.

STEM_03: $Y_t = 0.5Y_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim D_{snorm}(0, 1)$.

STEM_04: $Y_t = 0.3Y_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim D_{snorm}(0, 1)$.

STEM_05: $Y_t = 0.1Y_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim D_{snorm}(0, 1)$.

STEM_06: $Y_t = -0.1Y_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim D_{snorm}(0, 1)$.

STEM_07: $Y_t = -0.3Y_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim D_{snorm}(0, 1)$.

STEM_08: $Y_t = -0.5Y_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim D_{snorm}(0, 1)$.

STEM_09: $Y_t = -0.7Y_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim D_{snorm}(0, 1)$.

STEM_10: $Y_t = -0.9Y_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim D_{snorm}(0, 1)$.

Table A.25 Hypothetical DGMs used in Univariate Experiment 4: Case 2.

	used to generate time series of length 1000		
	sources of samples from 1 to 600		sources of samples from 601 to 1000
DGM01 :=	**STEM_01** $\mid \{\varepsilon_t\} \sim \mathbf{D_{snorm}(0, 1)}$	$\triangleright$	**STEM_01** $\mid \{\varepsilon_t\} \sim \mathbf{D_{snorm}(0, 1)}$
DGM02 :=	**STEM_01** $\mid \{\varepsilon_t\} \sim \mathbf{D_{snorm}(0, 1)}$	$\blacktriangleright$	**STEM_02** $\mid \{\varepsilon_t\} \sim \mathbf{D_{snorm}(0, 1)}$
DGM03 :=	**STEM_03** $\mid \{\varepsilon_t\} \sim \mathbf{D_{snorm}(0, 1)}$	$\triangleright$	**STEM_03** $\mid \{\varepsilon_t\} \sim \mathbf{D_{snorm}(0, 1)}$
DGM04 :=	**STEM_03** $\mid \{\varepsilon_t\} \sim \mathbf{D_{snorm}(0, 1)}$	$\blacktriangleright$	**STEM_04** $\mid \{\varepsilon_t\} \sim \mathbf{D_{snorm}(0, 1)}$
DGM05 :=	**STEM_05** $\mid \{\varepsilon_t\} \sim \mathbf{D_{snorm}(0, 1)}$	$\triangleright$	**STEM_05** $\mid \{\varepsilon_t\} \sim \mathbf{D_{snorm}(0, 1)}$
DGM06 :=	**STEM_05** $\mid \{\varepsilon_t\} \sim \mathbf{D_{snorm}(0, 1)}$	$\blacktriangleright$	**STEM_06** $\mid \{\varepsilon_t\} \sim \mathbf{D_{snorm}(0, 1)}$
DGM07 :=	**STEM_07** $\mid \{\varepsilon_t\} \sim \mathbf{D_{snorm}(0, 1)}$	$\triangleright$	**STEM_07** $\mid \{\varepsilon_t\} \sim \mathbf{D_{snorm}(0, 1)}$
DGM08 :=	**STEM_07** $\mid \{\varepsilon_t\} \sim \mathbf{D_{snorm}(0, 1)}$	$\blacktriangleright$	**STEM_08** $\mid \{\varepsilon_t\} \sim \mathbf{D_{snorm}(0, 1)}$
DGM09 :=	**STEM_09** $\mid \{\varepsilon_t\} \sim \mathbf{D_{snorm}(0, 1)}$	$\triangleright$	**STEM_09** $\mid \{\varepsilon_t\} \sim \mathbf{D_{snorm}(0, 1)}$
DGM10 :=	**STEM_09** $\mid \{\varepsilon_t\} \sim \mathbf{D_{snorm}(0, 1)}$	$\blacktriangleright$	**STEM_10** $\mid \{\varepsilon_t\} \sim \mathbf{D_{snorm}(0, 1)}$

$'\triangleright'$ indicates where the data generation mechanisms remain unchanged.
$'\blacktriangleright'$ indicates where the data generating mechanisms are changed slightly after 600th observation.

1. Each DGM in Table A.25 is used to generate **10** number of time series with the length of **1000**.

ss_sub: 1 to 600

2. The slice of the dataset consisting of the first 600 time points is filtered out.
3. Clustering done on this slice by the proposed approach given in Section 3.4.2 and through implementation steps given in Figure 3.4, and **12** dissimilarity measures given in Table 4.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed over **4** × **20** × **600** dimensional dataset and clustering accuracies are saved.
4. Steps from **1** to **3** repeated **100** (i.e., **rn**) times to get average accuracies of **36** (i.e., **3**∗**12**) clustering schemes and the proposed approach.
5. The average accuracy indexes of all methods considered in the study are displayed in Table A.26.

ss_oa: 1 to 1000

2. The dataset consisting of time series length of 1000 is taken for clustering.
3. Clustering done on the dataset by the proposed approach given in Section 3.4.2 and through implementation steps given in Figure 3.4, and **12** dissimilarity measures given in Table 4.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed over **8** × **10** × **1000** dimensional dataset and clustering accuracies are saved.
4. Steps from **1** to **3** repeated **100** (i.e., **rn**) times to get average accuracies of **36** (i.e., **3**∗**12**) clustering schemes and the proposed approach.
5. The average accuracy indexes of all methods considered in the study are displayed in Table A.26.

Table A.26 Univariate Experiment 4: Case 2 average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach.

	Clustering method					
	FUZZY		PAM		SPECTRAL	
	sample sizes (ss)		sample sizes (ss)		sample sizes (ss)	
Dissimilarity Measure	1 *to* 600 $(5\ cluster)$	1 *to* 1000 $(10\ cluster)$	1 *to* 600 $(5\ cluster)$	1 *to* 1000 $(10\ cluster)$	1 *to* 600 $(5\ cluster)$	1 *to* 1000 $(10\ cluster)$
Model Free						
D_{ACF}	0.734	0.691	0.947	0.692	0.983	0.855
D_{PACF}	0.960	0.814	0.999	0.736	1.000	0.709
D_{PER}	0.686	0.469	0.445	0.362	0.959	0.821
$D_{INT.PER}$	0.999	0.955	0.999	0.948	0.893	0.473
$D_{SPEC.LLR1}$	0.999	0.929	0.999	0.857	0.899	0.723
$D_{SPEC.LLR2}$	0.999	0.924	1.000	0.858	0.895	0.724
D_{GLK}	1.000	0.886	0.998	0.801	0.988	0.761
D_{VR}	0.999	0.931	0.999	0.843	0.873	0.746
Model Based						
D_{AR-LPC}	0.982	0.940	0.982	0.858	0.982	0.947
D_{AR-PIC}	0.999	0.968	0.999	0.940	1.000	0.969
D_{AR-MAH}	1.000	0.947	1.000	0.934	1.000	0.944
Complexity B.						
D_{PDC}	0.969	0.678	0.956	0.689	0.870	0.512

	Proposed Clustering Approach	
	TSMB-SPCL-UV	
	$ss \rightarrow 1$ *to* 600 $(5\ cluster)$	$ss \rightarrow 1$ *to* 1000 $(10\ cluster)$
L^2 Dist.	1.000	0.952

APPENDIX B

MULTIVARIATE TIME SERIES CLUSTERING: EXPERIMENTS ON SYNTHETIC DATA

B.1 Experiment 3

B.1.1 Experiment 3: Case 1

Table B.1 Utilized specifications of VARs in the third set of experiments.

$$\textbf{STEM\_1}: \quad \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} = \begin{bmatrix} 0.20 & 0.50 & 0.70 \\ -0.60 & -0.50 & 0.50 \\ 0.30 & -0.80 & -0.30 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}$$

$$\textbf{STEM\_2}: \quad \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} = \begin{bmatrix} -0.75 & -0.15 & -0.30 \\ 0.10 & -0.65 & -0.65 \\ -0.30 & 0.90 & 0.50 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}$$

$$\textbf{STEM\_3}: \quad \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} = \begin{bmatrix} -0.15 & -0.07 & -0.21 \\ -0.06 & -0.30 & -0.13 \\ -0.09 & -0.70 & -0.15 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}$$

$\varepsilon_{1,t},\ \varepsilon_{2,t},\ \varepsilon_{3,t}$ are error terms.
Please note that the error terms can be described by any process or distribution.

Table B.2 Utilized specifications of multivariate error structures in the third set of experiments.

$$\mathbf{C_1} : \mathbf{ECCC} - \mathbf{GARCH(1,1)} =$$

$$\begin{cases} h_{1,t} = 0.0015 + 0.20\varepsilon_{1,t-1}^2 + 0.05\varepsilon_{2,t-1}^2 + 0.20\varepsilon_{3,t-1}^2 + 0.10h_{1,t-1} + 0.15h_{2,t-1} + 0.15h_{3,t-1} \\ h_{2,t} = 0.0015 + 0.25\varepsilon_{1,t-1}^2 + 0.30\varepsilon_{2,t-1}^2 + 0.15\varepsilon_{3,t-1}^2 + 0.01h_{1,t-1} + 0.25h_{2,t-1} + 0.05h_{3,t-1} \\ h_{3,t} = 0.0025 + 0.05\varepsilon_{1,t-1}^2 + 0.01\varepsilon_{2,t-1}^2 + 0.15\varepsilon_{3,t-1}^2 + 0.35h_{1,t-1} + 0.05h_{2,t-1} + 0.10h_{3,t-1} \end{cases}$$

$$\mathbf{C_2} : \mathbf{ECCC} - \mathbf{GARCH(1,1)} =$$

$$\begin{cases} h_{1,t} = 0.0035 + 0.15\varepsilon_{1,t-1}^2 + 0.00\varepsilon_{2,t-1}^2 + 0.00\varepsilon_{3,t-1}^2 + 0.45h_{1,t-1} + 0.00h_{2,t-1} + 0.00h_{3,t-1} \\ h_{2,t} = 0.0055 + 0.00\varepsilon_{1,t-1}^2 + 0.30\varepsilon_{2,t-1}^2 + 0.00\varepsilon_{3,t-1}^2 + 0.00h_{1,t-1} + 0.35h_{2,t-1} + 0.00h_{3,t-1} \\ h_{3,t} = 0.0015 + 0.00\varepsilon_{1,t-1}^2 + 0.00\varepsilon_{2,t-1}^2 + 0.45\varepsilon_{3,t-1}^2 + 0.00h_{1,t-1} + 0.00h_{2,t-1} + 0.25h_{3,t-1} \end{cases}$$

$$\mathbf{C_3} : \mathbf{ECCC} - \mathbf{GARCH(1,1)} =$$

$$\begin{cases} h_{1,t} = 0.1000 + 0.09\varepsilon_{1,t-1}^2 + 0.05\varepsilon_{2,t-1}^2 + 0.05\varepsilon_{3,t-1}^2 + 0.06h_{1,t-1} + 0.05h_{2,t-1} + 0.06h_{3,t-1} \\ h_{2,t} = 0.2500 + 0.08\varepsilon_{1,t-1}^2 + 0.12\varepsilon_{2,t-1}^2 + 0.06\varepsilon_{3,t-1}^2 + 0.06h_{1,t-1} + 0.07h_{2,t-1} + 0.04h_{3,t-1} \\ h_{3,t} = 0.7000 + 0.05\varepsilon_{1,t-1}^2 + 0.01\varepsilon_{2,t-1}^2 + 0.02\varepsilon_{3,t-1}^2 + 0.08h_{1,t-1} + 0.03h_{2,t-1} + 0.09h_{3,t-1} \end{cases}$$

where $\varepsilon_{i,t} = \sqrt{h_{i,t}}e_{i,t},$ and $e_{i,t} \overset{i.i.d}{\sim} D_{snorm}(0,1), i = 1,2,3.$

The constant conditional correlation matrix, ρ, assumed as

$$\rho = \begin{array}{cccc} h_{1,t} & h_{2,t} & h_{3,t} & \\ \left[\begin{matrix} 1.00 \\ 0.15 \\ 0.25 \end{matrix}\right. & \begin{matrix} 0.15 \\ 1.00 \\ 0.20 \end{matrix} & \left.\begin{matrix} 0.25 \\ 0.20 \\ 1.00 \end{matrix}\right] & \begin{matrix} h_{1,t} \\ h_{2,t} \\ h_{3,t} \end{matrix} \end{array}$$ for all three multivariate GARCH specification.

Table B.3 Hypothetical DGMs used in Multivariate Experiment 3: Case 1 (9 cluster).

	used to generate time series of length 1800				
	sources of samples from 1 to 600		sources of samples from 601 to 1200		sources of samples from 1201 to 1800
DGM01 :=	STEM_1 | C_1	▷	STEM_1 | C_1	▷	STEM_1 | C_1
DGM02 :=	STEM_1 | C_1	▷	STEM_1 | C_1	▶	STEM_2 | C_2
DGM03 :=	STEM_1 | C_1	▷	STEM_1 | C_1	▶	STEM_3 | C_3
DGM04 :=	STEM_1 | C_1	▶	STEM_2 | C_2	▶	STEM_1 | C_1
DGM05 :=	STEM_1 | C_1	▶	STEM_2 | C_2	▷	STEM_2 | C_2
DGM06 :=	STEM_1 | C_1	▶	STEM_2 | C_2	▶	STEM_3 | C_3
DGM07 :=	STEM_1 | C_1	▶	STEM_3 | C_3	▶	STEM_1 | C_1
DGM08 :=	STEM_1 | C_1	▶	STEM_3 | C_3	▶	STEM_2 | C_2
DGM09 :=	STEM_1 | C_1	▶	STEM_3 | C_3	▷	STEM_3 | C_3

′ ▷′ indicates where the data generation mechanisms remain unchanged.
′ ▶′ indicates where the data generating mechanisms are changed at that point.
C_1, C_2, and C_3 are three distinct ECCC-GARCH(1,1) specifications given in Table B.2.

1. Each DGM in Table B.3 is used to generate **10**number of 3-dimensional multivariate time series with the length of **1800** to form a corresponding cluster.
2. Clustering done by the proposed approach given in Section 3.4 and through implementation steps given in Figure 3.4, and **5** dissimilarity measures given in Table 5.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed over **9** $\times$ **10** $\times$ **1800** dimensional dataset and clustering accuracies are saved.
3. Steps from **1** to **2** repeated **100** times to get average accuracies of **15** (i.e., **3** $*$ **5**) clustering schemes and the proposed approach.
4. The average accuracy indexes of all methods considered in the study are displayed for comparison in Table B.4.

Table B.4 Multivariate Experiment 3 Case 1: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach.

Dissimilarity measure	Clustering method		
	FUZZY	PAM	SPECTRAL
Model free			
D_{JD}	0.577	0.866	0.949
D_{LD}	0.349	0.545	0.719
D_{GAK}	0.384	0.233	0.285
D_{DTW}	0.756	0.921	0.441
D_{PDC}	0.386	0.455	0.598

	Proposed Clustering Approach
	TSMB-SPCL-MV
	$ss = 1800$
L^2 (Euclidean) Distance	0.954

B.2 Experiment 4

B.2.1 Experiment 4: Case 1

General representation of 3 dimensional two regime threshold VAR (TVAR) process can be represented as follows:

$$\mathbf{y_t} = \begin{cases} \phi_\mathbf{0}^{(\mathbf{1})} + \phi_\mathbf{1}^{(\mathbf{1})}\mathbf{y_{t-1}} + \varepsilon_\mathbf{t}^{(\mathbf{1})}, & s_{t-d} \leq r \\ \phi_\mathbf{0}^{(\mathbf{2})} + \phi_\mathbf{1}^{(\mathbf{2})}\mathbf{y_{t-1}} + \varepsilon_\mathbf{t}^{(\mathbf{2})}, & r < s_{t-d} \end{cases}$$

where $\mathbf{y_t} = \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix}$, and $\varepsilon_\mathbf{t} = \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}$.

Table B.5 Utilized specifications of TVARs in the fourth set of experiments.

$\mathbf{STEM\_1}$:

$$\phi_\mathbf{0}^{(\mathbf{1})} = \begin{bmatrix} 0.35 \\ -0.45 \\ -0.25 \end{bmatrix}, \phi_\mathbf{1}^{(\mathbf{1})} = \begin{bmatrix} 0.25 & -0.30 & -0.50 \\ 0.95 & 0.10 & -0.10 \\ -0.70 & 0.70 & -0.35 \end{bmatrix}, \phi_\mathbf{0}^{(\mathbf{2})} = \begin{bmatrix} -0.15 \\ -0.15 \\ 0.20 \end{bmatrix}, \phi_\mathbf{1}^{(\mathbf{2})} = \begin{bmatrix} 0.25 & -0.65 & 0.20 \\ -0.65 & -0.30 & 0.70 \\ -0.60 & 0.15 & -0.30 \end{bmatrix}$$

$\mathbf{STEM\_2}$:

$$\phi_\mathbf{0}^{(\mathbf{1})} = \begin{bmatrix} 0.45 \\ 0.40 \\ -0.20 \end{bmatrix}, \phi_\mathbf{1}^{(\mathbf{1})} = \begin{bmatrix} -0.75 & -0.05 & 0.25 \\ -0.35 & 0.75 & -0.15 \\ 0.75 & -0.05 & 0.70 \end{bmatrix}, \phi_\mathbf{0}^{(\mathbf{2})} = \begin{bmatrix} 0.65 \\ 0.35 \\ -0.75 \end{bmatrix}, \phi_\mathbf{1}^{(\mathbf{2})} = \begin{bmatrix} -0.01 & 0.35 & -0.85 \\ -0.04 & -0.60 & 0.15 \\ -0.40 & 0.05 & -0.75 \end{bmatrix}$$

$\mathbf{STEM\_3}$:

$$\phi_\mathbf{0}^{(\mathbf{1})} = \begin{bmatrix} 0.45 \\ -0.10 \\ -0.42 \end{bmatrix}, \phi_\mathbf{1}^{(\mathbf{1})} = \begin{bmatrix} 0.20 & 0.40 & 0.55 \\ -0.35 & 0.50 & -0.28 \\ 0.40 & 0.80 & -0.12 \end{bmatrix}, \phi_\mathbf{0}^{(\mathbf{2})} = \begin{bmatrix} 0.32 \\ -0.60 \\ 0.40 \end{bmatrix}, \phi_\mathbf{1}^{(\mathbf{2})} = \begin{bmatrix} 0.20 & 0.30 & -0.42 \\ 0.15 & 0.32 & 0.57 \\ -0.07 & -0.62 & -0.16 \end{bmatrix}$$

$\varepsilon_{1,t},\ \varepsilon_{2,t},\ \varepsilon_{3,t}$ are error terms.
Please note that the error terms can be described by any process or distribution.

Table B.6 Utilized specifications of multivariate error structures in the fourth set of experiments.

$$\mathbf{C_1} : \mathbf{ECCC - GARCH(1,1)} =$$

$$\begin{cases} h_{1,t} = 0.0015 + 0.20\varepsilon^2_{1,t-1} + 0.05\varepsilon^2_{2,t-1} + 0.20\varepsilon^2_{3,t-1} + 0.10h_{1,t-1} + 0.15h_{2,t-1} + 0.15h_{3,t-1} \\ h_{2,t} = 0.0015 + 0.25\varepsilon^2_{1,t-1} + 0.30\varepsilon^2_{2,t-1} + 0.15\varepsilon^2_{3,t-1} + 0.01h_{1,t-1} + 0.25h_{2,t-1} + 0.05h_{3,t-1} \\ h_{3,t} = 0.0025 + 0.05\varepsilon^2_{1,t-1} + 0.01\varepsilon^2_{2,t-1} + 0.15\varepsilon^2_{3,t-1} + 0.35h_{1,t-1} + 0.05h_{2,t-1} + 0.10h_{3,t-1} \end{cases}$$

$$\mathbf{C_2} : \mathbf{ECCC - GARCH(1,1)} =$$

$$\begin{cases} h_{1,t} = 0.0035 + 0.15\varepsilon^2_{1,t-1} + 0.00\varepsilon^2_{2,t-1} + 0.00\varepsilon^2_{3,t-1} + 0.45h_{1,t-1} + 0.00h_{2,t-1} + 0.00h_{3,t-1} \\ h_{2,t} = 0.0055 + 0.00\varepsilon^2_{1,t-1} + 0.30\varepsilon^2_{2,t-1} + 0.00\varepsilon^2_{3,t-1} + 0.00h_{1,t-1} + 0.35h_{2,t-1} + 0.00h_{3,t-1} \\ h_{3,t} = 0.0015 + 0.00\varepsilon^2_{1,t-1} + 0.00\varepsilon^2_{2,t-1} + 0.45\varepsilon^2_{3,t-1} + 0.00h_{1,t-1} + 0.00h_{2,t-1} + 0.25h_{3,t-1} \end{cases}$$

$$\mathbf{C_3} : \mathbf{ECCC - GARCH(1,1)} =$$

$$\begin{cases} h_{1,t} = 0.0000 + 0.03\varepsilon^2_{1,t-1} + 0.01\varepsilon^2_{2,t-1} + 0.09\varepsilon^2_{3,t-1} + 0.04h_{1,t-1} + 0.05h_{2,t-1} + 0.03h_{3,t-1} \\ h_{2,t} = 0.0000 + 0.03\varepsilon^2_{1,t-1} + 0.09\varepsilon^2_{2,t-1} + 0.06\varepsilon^2_{3,t-1} + 0.00h_{1,t-1} + 0.08h_{2,t-1} + 0.01h_{3,t-1} \\ h_{3,t} = 0.0000 + 0.00\varepsilon^2_{1,t-1} + 0.00\varepsilon^2_{2,t-1} + 0.06\varepsilon^2_{3,t-1} + 0.01h_{1,t-1} + 0.01h_{2,t-1} + 0.02h_{3,t-1} \end{cases}$$

where $\varepsilon_{i,t} = \sqrt{h_{i,t}}e_{i,t},$ and $e_{i,t} \overset{i.i.d}{\sim} D_{snorm}(0,1), i = 1,2,3.$

The constant conditional correlation matrix, ρ, assumed as

$$\rho = \begin{array}{cccc} h_{1,t} & h_{2,t} & h_{3,t} & \\ \left[\begin{matrix} 1.00 \\ 0.20 \\ 0.25 \end{matrix}\right. & \begin{matrix} 0.20 \\ 1.00 \\ 0.15 \end{matrix} & \left.\begin{matrix} 0.25 \\ 0.15 \\ 1.00 \end{matrix}\right] & \begin{matrix} h_{1,t} \\ h_{2,t} \\ h_{3,t} \end{matrix} \end{array}$$ for all three multivariate GARCH specification.

Table B.7 Hypothetical DGMs used in Multivariate Experiment 4: Case 1 (9 cluster).

	used to generate time series of length 1800				
	sources of samples from 1 to 600		sources of samples from 601 to 1200		sources of samples from 1201 to 1800
DGM01 :=	STEM_1 $\mid C_1$	▷	STEM_1 $\mid C_1$	▷	STEM_1 $\mid C_1$
DGM02 :=	STEM_1 $\mid C_1$	▷	STEM_1 $\mid C_1$	▶	STEM_2 $\mid C_2$
DGM03 :=	STEM_1 $\mid C_1$	▷	STEM_1 $\mid C_1$	▶	STEM_3 $\mid C_3$
DGM04 :=	STEM_1 $\mid C_1$	▶	STEM_2 $\mid C_2$	▶	STEM_1 $\mid C_1$
DGM05 :=	STEM_1 $\mid C_1$	▶	STEM_2 $\mid C_2$	▷	STEM_2 $\mid C_2$
DGM06 :=	STEM_1 $\mid C_1$	▶	STEM_2 $\mid C_2$	▶	STEM_3 $\mid C_3$
DGM07 :=	STEM_1 $\mid C_1$	▶	STEM_3 $\mid C_3$	▶	STEM_1 $\mid C_1$
DGM08 :=	STEM_1 $\mid C_1$	▶	STEM_3 $\mid C_3$	▶	STEM_2 $\mid C_2$
DGM09 :=	STEM_1 $\mid C_1$	▶	STEM_3 $\mid C_3$	▷	STEM_3 $\mid C_3$

$'\triangleright'$ indicates where the data generation mechanisms remain unchanged.
$'\blacktriangleright'$ indicates where the data generating mechanisms are changed at that point.
C_1, C_2, and C_3 are three distinct ECCC-GARCH(1,1) specifications given in Table B.6.

1. Each DGM in Table B.7 is used to generate **10**number of 3-dimensional multivariate time series with the length of **1800** to form a corresponding cluster.
2. Clustering done by the proposed approach given in Section 3.4 and through implementation steps given in Figure 3.4, and **5** dissimilarity measures given in Table 5.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed over **9** $\times$ **10** $\times$ **1800** dimensional dataset and clustering accuracies are saved.
3. Steps from **1** to **2** repeated **100** times to get average accuracies of **15** (i.e., **3** $*$ **5**) clustering schemes and the proposed approach.
4. The average accuracy indexes of all methods considered in the study are displayed for comparison in Table B.8.

Table B.8 Multivariate Experiment 4 Case 1: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach.

Dissimilarity measure	Clustering method		
	FUZZY	PAM	SPECTRAL
Model free			
D_{JD}	0.589	0.723	0.794
D_{LD}	0.361	0.578	0.725
D_{GAK}	0.326	0.208	0.410
D_{DTW}	0.426	0.209	0.862
D_{PDC}	0.383	0.438	0.709

	Proposed Clustering Approach
	TSMB-SPCL-MV
	$ss = 1800$
L^2 (Euclidean) Distance	0.916

B.3 Experiment 5

B.3.1 Experiment 5: Case 1

General representation of 3 dimensional two regime threshold VECM (TVECM) process can be represented as follows:

$$\mathbf{\Delta y_t} = \begin{cases} \mathbf{\Pi^{(1)} y_{t-1}} + \phi_{\mathbf{1}}^{\mathbf{(1)}} \mathbf{\Delta y_{t-1}} + \varepsilon_{\mathbf{t}}^{\mathbf{(1)}}, & s_{t-d} \leq r \\ \mathbf{\Pi^{(2)} y_{t-1}} + \phi_{\mathbf{1}}^{\mathbf{(2)}} \mathbf{\Delta y_{t-1}} + \varepsilon_{\mathbf{t}}^{\mathbf{(2)}}, & r < s_{t-d} \end{cases}$$

where $\mathbf{\Delta y_t} = \begin{bmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \\ \Delta y_{3,t} \end{bmatrix} = \begin{bmatrix} y_{1,t} - y_{1,t-1} \\ y_{2,t} - y_{2,t-1} \\ y_{3,t} - y_{3,t-1} \end{bmatrix}$, and $\varepsilon_{\mathbf{t}} = \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}$.

Table B.9 Utilized specifications of TVECMs in the fifth set of experiments.

STEM_1 :

$$\boldsymbol{\Pi}^{(\mathbf{1})} = \begin{bmatrix} -0.02 & 0.17 & 0.04 \\ 0.04 & -0.36 & -0.08 \\ -0.01 & 0.00 & 0.00 \end{bmatrix}, \boldsymbol{\Pi}^{(\mathbf{2})} = \begin{bmatrix} -0.95 & 0.15 & 0.10 \\ 0.00 & -0.04 & 0.05 \\ 0.10 & -0.40 & 0.05 \end{bmatrix},$$

$$\phi_{\mathbf{1}}^{(\mathbf{1})} = \begin{bmatrix} -0.95 & 0.15 & 0.10 \\ 0.00 & -0.04 & 0.05 \\ 0.10 & -0.40 & 0.06 \end{bmatrix}, \ \phi_{\mathbf{1}}^{(\mathbf{2})} = \begin{bmatrix} 1.30 & -0.65 & 0.15 \\ -0.45 & 0.15 & -0.25 \\ 0.06 & -0.03 & 0.03 \end{bmatrix}$$

STEM_2 :

$$\boldsymbol{\Pi}^{(\mathbf{1})} = \begin{bmatrix} -0.27 & 0.00 & 0.32 \\ 0.57 & 0.00 & -0.68 \\ -0.00 & 0.00 & 0.00 \end{bmatrix}, \boldsymbol{\Pi}^{(\mathbf{2})} = \begin{bmatrix} -0.07 & 0.22 & 0.37 \\ 0.15 & -0.47 & -0.78 \\ -0.00 & 0.00 & 0.00 \end{bmatrix},$$

$$\phi_{\mathbf{1}}^{(\mathbf{1})} = \begin{bmatrix} 0.15 & 0.25 & 0.75 \\ -0.95 & -0.75 & 0.00 \\ 0.01 & -0.15 & -0.05 \end{bmatrix}, \ \phi_{\mathbf{1}}^{(\mathbf{2})} = \begin{bmatrix} 0.05 & 0.01 & 0.45 \\ 0.45 & 0.02 & 0.05 \\ -0.01 & -0.95 & -0.40 \end{bmatrix}$$

STEM_3 :

$$\boldsymbol{\Pi}^{(\mathbf{1})} = \begin{bmatrix} -0.04 & 0.14 & 0.17 \\ 0.08 & -0.29 & -0.35 \\ -0.00 & 0.00 & 0.00 \end{bmatrix}, \boldsymbol{\Pi}^{(\mathbf{2})} = \begin{bmatrix} 0.02 & -0.14 & -0.24 \\ -0.05 & 0.29 & 0.51 \\ 0.00 & -0.00 & -0.00 \end{bmatrix},$$

$$\phi_{\mathbf{1}}^{(\mathbf{1})} = \begin{bmatrix} -0.24 & -0.21 & -0.13 \\ -0.23 & -0.25 & -0.05 \\ -0.42 & -0.41 & 0.08 \end{bmatrix}, \ \phi_{\mathbf{1}}^{(\mathbf{2})} = \begin{bmatrix} 0.09 & 0.02 & -0.06 \\ 0.10 & -0.38 & -0.17 \\ 0.21 & -0.61 & -0.16 \end{bmatrix}$$

$\varepsilon_{1,t},\ \varepsilon_{2,t},\ \varepsilon_{3,t}$ are error terms.

Please note that the error terms can be described by any process or distribution.

Table B.10 Utilized specifications of multivariate error structures in the fifth set of experiments.

$$\mathbf{C_1} : \mathbf{ECCC-GARCH(1,1)} =$$

$$\begin{cases} h_{1,t} = 0.0015 + 0.20\varepsilon^2_{1,t-1} + 0.05\varepsilon^2_{2,t-1} + 0.20\varepsilon^2_{3,t-1} + 0.10h_{1,t-1} + 0.15h_{2,t-1} + 0.15h_{3,t-1} \\ h_{2,t} = 0.0015 + 0.25\varepsilon^2_{1,t-1} + 0.30\varepsilon^2_{2,t-1} + 0.15\varepsilon^2_{3,t-1} + 0.01h_{1,t-1} + 0.25h_{2,t-1} + 0.05h_{3,t-1} \\ h_{3,t} = 0.0025 + 0.05\varepsilon^2_{1,t-1} + 0.01\varepsilon^2_{2,t-1} + 0.15\varepsilon^2_{3,t-1} + 0.35h_{1,t-1} + 0.05h_{2,t-1} + 0.10h_{3,t-1} \end{cases}$$

$$\mathbf{C_2} : \mathbf{ECCC-GARCH(1,1)} =$$

$$\begin{cases} h_{1,t} = 0.0035 + 0.15\varepsilon^2_{1,t-1} + 0.00\varepsilon^2_{2,t-1} + 0.00\varepsilon^2_{3,t-1} + 0.45h_{1,t-1} + 0.00h_{2,t-1} + 0.00h_{3,t-1} \\ h_{2,t} = 0.0055 + 0.00\varepsilon^2_{1,t-1} + 0.30\varepsilon^2_{2,t-1} + 0.00\varepsilon^2_{3,t-1} + 0.00h_{1,t-1} + 0.35h_{2,t-1} + 0.00h_{3,t-1} \\ h_{3,t} = 0.0015 + 0.00\varepsilon^2_{1,t-1} + 0.00\varepsilon^2_{2,t-1} + 0.45\varepsilon^2_{3,t-1} + 0.00h_{1,t-1} + 0.00h_{2,t-1} + 0.25h_{3,t-1} \end{cases}$$

$$\mathbf{C_3} : \mathbf{ECCC-GARCH(1,1)} =$$

$$\begin{cases} h_{1,t} = 0.0000 + 0.03\varepsilon^2_{1,t-1} + 0.01\varepsilon^2_{2,t-1} + 0.09\varepsilon^2_{3,t-1} + 0.04h_{1,t-1} + 0.05h_{2,t-1} + 0.03h_{3,t-1} \\ h_{2,t} = 0.0000 + 0.03\varepsilon^2_{1,t-1} + 0.09\varepsilon^2_{2,t-1} + 0.06\varepsilon^2_{3,t-1} + 0.00h_{1,t-1} + 0.08h_{2,t-1} + 0.01h_{3,t-1} \\ h_{3,t} = 0.0000 + 0.00\varepsilon^2_{1,t-1} + 0.00\varepsilon^2_{2,t-1} + 0.06\varepsilon^2_{3,t-1} + 0.01h_{1,t-1} + 0.01h_{2,t-1} + 0.02h_{3,t-1} \end{cases}$$

where $\varepsilon_{i,t} = \sqrt{h_{i,t}}e_{i,t},$ and $e_{i,t} \overset{i.i.d}{\sim} D_{snorm}(0,1), i = 1,2,3.$

The constant conditional correlation matrix, ρ, assumed as

$$\rho = \begin{array}{cc} \begin{array}{ccc} h_{1,t} & h_{2,t} & h_{3,t} \end{array} & \\ \begin{bmatrix} 1.00 & 0.20 & 0.25 \\ 0.20 & 1.00 & 0.15 \\ 0.25 & 0.15 & 1.00 \end{bmatrix} & \begin{array}{c} h_{1,t} \\ h_{2,t} \\ h_{3,t} \end{array} \end{array}$$

for all three multivariate GARCH specification.

Table B.11 Hypothetical DGMs used in Multivariate Experiment 5: Case 1 (9 cluster).

	used to generate time series of length 1800				
	sources of samples from 1 to 600		sources of samples from 601 to 1200		sources of samples from 1201 to 1800
DGM01 :=	**STEM_1** $\mid C_1$	▷	**STEM_1** $\mid C_1$	▷	**STEM_1** $\mid C_1$
DGM02 :=	**STEM_1** $\mid C_1$	▷	**STEM_1** $\mid C_1$	▶	**STEM_2** $\mid C_2$
DGM03 :=	**STEM_1** $\mid C_1$	▷	**STEM_1** $\mid C_1$	▶	**STEM_3** $\mid C_3$
DGM04 :=	**STEM_1** $\mid C_1$	▶	**STEM_2** $\mid C_2$	▶	**STEM_1** $\mid C_1$
DGM05 :=	**STEM_1** $\mid C_1$	▶	**STEM_2** $\mid C_2$	▷	**STEM_2** $\mid C_2$
DGM06 :=	**STEM_1** $\mid C_1$	▶	**STEM_2** $\mid C_2$	▶	**STEM_3** $\mid C_3$
DGM07 :=	**STEM_1** $\mid C_1$	▶	**STEM_3** $\mid C_3$	▶	**STEM_1** $\mid C_1$
DGM08 :=	**STEM_1** $\mid C_1$	▶	**STEM_3** $\mid C_3$	▶	**STEM_2** $\mid C_2$
DGM09 :=	**STEM_1** $\mid C_1$	▶	**STEM_3** $\mid C_3$	▷	**STEM_3** $\mid C_3$

′ ▷′ indicates where the data generation mechanisms remain unchanged.
′ ▶′ indicates where the data generating mechanisms are changed at that point.
C_1, C_2, and C_3 are three distinct ECCC-GARCH(1,1) specifications given in Table B.10.

1. Each DGM in Table B.11 is used to generate **10**number of 3-dimensional multivariate time series with the length of **1800** to form a corresponding cluster.
2. Clustering done by the proposed approach given in Section 3.4 and through implementation steps given in Figure 3.4, and **5** dissimilarity measures given in Table 5.1 for **3** different clustering methods (i.e., Fuzzy, PAM, Spectral) are performed over **9** × **10** × **1800** dimensional dataset and clustering accuracies are saved.
3. Steps from **1** to **2** repeated **100** times to get average accuracies of **15** (i.e., **3** ∗ **5**) clustering schemes and the proposed approach.
4. The average accuracy indexes of all methods considered in the study are displayed for comparison in Table B.12.

Table B.12 Multivariate Experiment 4 Case 1: average accuracy indexes (i.e., approx. correct clustering percentages) of compared time series clustering methods and the Proposed Clustering Approach.

Dissimilarity measure	Clustering method		
	FUZZY	PAM	SPECTRAL
Model free			
D_{JD}	0.442	0.619	0.648
D_{LD}	0.336	0.441	0.522
D_{GAK}	0.248	0.288	0.254
D_{DTW}	0.239	0.291	0.243
D_{PDC}	0.361	0.533	0.692

	Proposed Clustering Approach
	TSMB-SPCL-MV
	$ss = 1800$
L^2 (Euclidean) Distance	0.902

CURRICULUM VITAE

PERSONAL INFORMATION

Place & Date of born	*Mardin*, Turkey / 15.01.1981
Contact Info.	Department of Econometrics, *Van 100th Year University*, İİBF, Ekonometri Bölümü. TUŞBA *Van*/TURKEY
Phone & fax	+90(0)432 4445065 – 27647
E-mail	sipanaslan@yyu.edu.tr / aslan.sipan@gmail.com
Marital Status	Married

CURRENT POSITION

Research Assistant, Department of Econometrics, Van Yuzuncu Yil University (2017-Present)

EDUCATION

Ph.D. Department of Statistics
Middle East Technical University, Ankara, Turkey, Sept. 2010-Present
Advisor: Assoc.Prof.Dr. Ceylan Yozgatligil
Co-Advisor: Assoc.Prof.Dr. Cem Iyigun
Thesis Title: Temporal Clustering of Multivariate Time Series

M.Sc. Department of Statistics
Middle East Technical University, Ankara, Turkey, August 2010
Advisor: Assoc.Prof.Dr. Ceylan Yozgatligil
Thesis Title: Comparison of Missing Value Imputation Methods for Meteorological Time Series Data

B.Sc. Department of Statistics
Ege University, Izmir, Turkey, June 2004
Advisor: Prof.Dr. Saim Kendir
Final Project Title: Chaos Theory and Nonlinear Time Series Analysis

RESEARCH INTERESTS

Time Series Clustering
Time Series Analysis and Panel Data Econometrics
Modeling and Forecasting
Missing Data Handling

RESEARCH & TEACHING EXPERIENCE

RA and TA mem.	*Department of Statistics, METU, Fall 2007 - July 2016*
Research Group mem.	*NINLIL, Department of Statistics, METU, Fall 2008-2016* http://stat.metu.edu.tr/ninlil-research-group
Research Scholarship	*Visiting Student at University of California Irvine (2013-2014).* *Visiting Student at Space-Time Modelling Group (2013-2014).*
Project mem.	*Project title: Redefining Climate Regions of Turkey via Data Mining Methods and Developing Precipitation Forecasting Models* *Supported by: Center of Scientific Research Projects of METU* *Project Title: Generating Commodity Price Index via Time Series Model based Clustering* *Supported by: Center of Scientific Research Projects of METU*
Assisted Courses	*Introduction to Probability and Statistics I Fall 2007, Fall 2009* *Introduction to Probability and Statistics II Spring 2008* *Probability I Fall 2008* *Probability II Spring 2009* *Survey Sampling Techniques Fall 2010* *Multivariate Analysis I Fall 2011* *Multivariate Analysis II Spring 2008, Spring 2012* *Statistical Decision Analysis Spring 2014* *Time Series Analysis 2007, 2008, 2009, 2010, 2012, 2014, 2015*
IT Assistant	*Department of Statistics, METU, Fall 2008 - Fall 2012* *Dokuz Eylul University Graduate School of Natural and Applied Sciences IT Dept., Fall 2005- Spring 2006*

TECHNICAL SKILLS

Programming in	R, SAS, MATLAB, RATS
Experience with:	E-Views, JMP, SPSS, S-Plus[Spotfire], J-Multi, Arc-GIS

PUBLICATIONS

Papers in International Journals (Peer-reviewed)

Temporal Clustering of Time Series via Threshold Autoregressive Models: Application to Commodity Prices. Aslan, Sipan., Yozgatlıgil, Ceylan., Iyigun, Cem. (2018) Annals of Operations Research, 260(1), 51-77.

Has Climate Been Changing in Turkey? A Comparative Statistical Analysis of Two Consecutive Time Periods: 1950-1980 and 1981-2010. Türkeş, M., Yozgatlıgil, C., İyigün, C., Kartal-Koç, E., Fahmi, F., Aslan, S., and İ. Batmaz. (2016) Climate Research, vol.70, pp.77-93.

Comparison of Missing Value Imputation Methods in Time Series: The Case of Turkish Meteorological Data. Yozgatligil, C., Aslan, S., İyigün, C. and İ. Batmaz. (2013) Theoretical and Applied Climatology, 112(1-2), 143-167.

Papers in National Journals (Peer-reviewed)

Comparison of Missing Data Imputation Methods for Meteorological Time Series Data via Correlation Dimension Technique(in Turkish). Aslan, S., Yozgatlıgil, C., İyigün, C., Batmaz, İ., and H. Tatlı. (2011) Journal of Statistical Research, 8 (2), 55-67.

Papers in Preperation

Modeling the Non-Linear Dynamics in Epileptic Seizure EEG Data. Aslan, S., Ombao, H.

International Conference

Comparison of Missing Value Imputation Methods for Turkish Monthly Total Precipitation Data. 9th International Conference Computer Data Analysis and Modeling: Complex Stochastic Data and Systems. Aslan, S., C. Yozgatligil, C. İyigün, İ. Batmaz, M. Türkeş, H. Tatlı. Minsk, Belarus. September 7-11, 2010. Vol: 2, pp. 137-140. (Paper in Proceeding)

Spectral Clustering on Time Series Data. 25th Conference of European Chapter on Combinatorial Optimization. Aslan, S., C. İyigün., C. Yozgatligil., İ. Batmaz., Antalya, Turkey. April 26-28, 2012.

Analyzing the Long-Term Projections of Precipitation Totals via Linear and Nonlinear Time Series Models. International Conference on Applied and Computational Mathematics. Aslan, S., Yozgatligil, C., Iyigun, C., Batmaz, I. Ankara, Turkey. October 3 – 6, 2012.

Precipitation Forecasting: Assessment via Nonlinear Autoregressive Neural Network and Vector Autoregressive models. International Institute of Industrial Engineering Conference. Aslan, S., İyigün, C., Yozgatlıgil, C., Batmaz, İ. Istanbul, Turkey. June 26-28, 2013.

Double Threshold GARCH Model with Applications to EEG Data. Joint Statistical Meetings. Aslan, S. Ombao, H. Boston, USA. 2-7 August, 2014.
Temporal Clustering of Commodity Prices via Threshold Autoregressive Models. 55th Meeting of the EURO Working Group for Commodities and Financial Modelling. Aslan, S., Yozgatlıgil, C., İyigün, C. Ankara, Turkey. May 14-16, 2015.

National Conference and Symposium

Doğrusal Olmayan Dinamik Zaman Serileri Analizi ve Korelasyon Boyutunun Kayıp Veri Atama Yöntemlerinde Başarım Ölçütü Olarak Kullanılması/[Nonlinear Time Series Analysis and Using Correlation Dimension as a Performance Measure for Missing Value Imputation Methods] . Aslan, S., Yozgatlıgil, C., İyigün, C., Batmaz, İ., Tatlı, H. İGS'2010: 7. İstatistik Günleri Sempozyumu. Ankara, Türkiye. 28-30 Haziran, 2010. 100-102.

1950-2006 Yılları Arasındaki Türkiye Yağış Verilerinin Tanımlayıcı Veri Madenciliği Yöntemleri ile Analizi /[Handling and Analysis of Turkish Precipitation Data for the Period of 1950-2006 Using Descriptive Data Mining Techniques]. Asar, Ö., Kartal-Koç, E., Aslan, S., Öztürk, M. Z., Yozgatlıgil, C., Çınar, İ., Batmaz, İ., Köksal, G., Türkeş, M., Tatlı, H. İGS'2010: 7. İstatistik Günleri Sempozyumu. Ankara, Türkiye. 28-30 Haziran, 2010. 77-79.

Zaman Serisinde Kayıp Veri Tamamlama Yöntemlerinin Karşılaştırılması: Türkiye İklim Verileri Üzerine Bir Uygulama/[Comparison of Missing Value Imputation Methods in Time Series: an Application for Turkish Climate Data] . Yozgatlıgil, C., Aslan, S., İyigün, C., Batmaz, İ., Türkeş, M., Tatlı, H. YAEM'2010: Yöneylem Araştırması ve Endüstri Mühendisliği 30. Ulusal Kongresi. İstanbul, Türkiye. 30 Haziran-2 Temmuz, 2010. 127.

Determining Different Regimes Using Nonlinear Time Series with a Threshold Value Model: An Application on Turkey Precipitation Data (in Turkish). Aslan, S., Yozgatlıgil, C., İyigün, C. and İ. Batmaz. YAEM: 35th National Congress of Operations Research and Industrial Engineering, Ankara, Turkey, September 9-11, 2015.