

**PREDICTIVE MODELLING FROM MONTHLY
RAINFALL PATTERNS USING IMPUTATION
APPROACHES COMBINED WITH
MULTIVARIATE ANALYSIS**

MUHAMAD AFDAL BIN AHMAD BASRI

SULTAN IDRIS EDUCATION UNIVERSITY

2024



05-4506832



pustaka.upsi.edu.my



Perpustakaan Tuanku Bainun
Kampus Sultan Abdul Jalil Shah



PustakaTBainun



ptbupsi

**PREDICTIVE MODELLING FROM MONTHLY RAINFALL PATTERNS
USING IMPUTATION APPROACHES COMBINED WITH
MULTIVARIATE ANALYSIS**

MUHAMAD AFDAL BIN AHMAD BASRI



05-4506832



pustaka.upsi.edu.my



Perpustakaan Tuanku Bainun
Kampus Sultan Abdul Jalil Shah



PustakaTBainun



ptbupsi

**DISERTASI DIKEMUKAKAN BAGI MEMENUHI SYARAT UNTUK
MEMPEROLEHI IJAZAH SARJANA SAINS
(MOD PENYELIDIKAN)**

**FAKULTI SAINS DAN MATEMATIK
UNIVERSITI PENDIDIKAN SULTAN IDRIS**

2024



05-4506832



pustaka.upsi.edu.my



Perpustakaan Tuanku Bainun
Kampus Sultan Abdul Jalil Shah



PustakaTBainun



ptbupsi



- Sila tanda (✓)
- Kertas Projek
 - Sarjana Penyelidikan
 - Sarjana Penyelidikan dan Kerja Kursus
 - Doktor Falsafah

✓

INSTITUT PENGAJIAN SISWAZAH
PERAKUAN KEASLIAN PENULISAN

Perakuan ini telah dibuat pada ...14.....(hari bulan).....05..... (bulan) 20.....24

i. Perakuan pelajar :

Saya, MUHAMAD AFDAL BIN AHMAD BASRI, M20201000214, SAINS DAN MATEMATIK (SILA NYATAKAN NAMA PELAJAR, NO. MATRIK DAN FAKULTI) dengan ini mengaku bahawa disertasi/tesis yang bertajuk PREDICTIVE MODELLING FROM MONTHLY RAINFALL PATTERNS USING IMPUTATIONS APPROACHES COMBINED WITH MULTIVARIATE ANALYSIS.

adalah hasil kerja saya sendiri. Saya tidak memplagiat dan apa-apa penggunaan mana-mana hasil kerja yang mengandungi hak cipta telah dilakukan secara urusan yang wajar dan bagi maksud yang dibenarkan dan apa-apa petikan, ekstrak, rujukan atau pengeluaran semula daripada atau kepada mana-mana hasil kerja yang mengandungi hak cipta telah dinyatakan dengan sejelasnya dan secukupnya

afdal
Tandatangan pelajar

ii. Perakuan Penyelia:

Saya, DR SHAZLYN MILLEANA SHAHARUDIN (NAMA PENYELIA) dengan ini mengesahkan bahawa hasil kerja pelajar yang bertajuk PREDICTIVE MODELLING FROM MONTHLY RAINFALL PATTERNS USING IMPUTATIONS APPROACHES COMBINED WITH MULTIVARIATE.

(TAJUK) dihasilkan oleh pelajar seperti nama di atas, dan telah diserahkan kepada Institut Pengajian Siswazah bagi memenuhi sebahagian/sepenuhnya syarat untuk memperoleh Ijazah _____ (SILA NYATAKAN NAMA IJAZAH).

14/5/2024
Tarikh

Shazlyn
Tandatangan Penyelia



**INSTITUT PENGAJIAN SISWAZAH /
INSTITUTE OF GRADUATE STUDIES**

**BORANG PENGESAHAN PENYERAHAN TESIS/DISERTASI/LAPORAN KERTAS PROJEK
DECLARATION OF THESIS/DISSERTATION/PROJECT PAPER FORM**

Tajuk / Title: PREDICTIVE MODELLING FROM MONTHLY RAINFALL PATTERNS USING
IMPUTATIONS APPROACHES COMBINED WITH MULTIVARIATE ANALYSIS.

No. Matrik / Matric's No.: M20201000214

Saya / I : MUHAMAD AFDAL BIN AHMAD BASRI

(Nama pelajar / Student's Name)

mengaku membenarkan Tesis/Disertasi/Laporan Kertas Projek (Kedoktoran/Sarjana)* ini disimpan di Universiti Pendidikan Sultan Idris (Perpustakaan Tuanku Bainun) dengan syarat-syarat kegunaan seperti berikut:-

acknowledged that Universiti Pendidikan Sultan Idris (Tuanku Bainun Library) reserves the right as follows:-

1. Tesis/Disertasi/Laporan Kertas Projek ini adalah hak milik UPSI.
The thesis is the property of Universiti Pendidikan Sultan Idris
2. Perpustakaan Tuanku Bainun dibenarkan membuat salinan untuk tujuan rujukan dan penyelidikan.
Tuanku Bainun Library has the right to make copies for the purpose of reference and research.
3. Perpustakaan dibenarkan membuat salinan Tesis/Disertasi ini sebagai bahan pertukaran antara Institusi Pengajian Tinggi.
The Library has the right to make copies of the thesis for academic exchange.
4. Sila tandakan (✓) bagi pilihan kategori di bawah / *Please tick (✓) for category below:-*

SULIT/CONFIDENTIAL

Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub dalam Akta Rahsia Rasmi 1972. / *Contains confidential information under the Official Secret Act 1972*

TERHAD/RESTRICTED

Mengandungi maklumat terhad yang telah ditentukan oleh organisasi/badan di mana penyelidikan ini dijalankan. / *Contains restricted information as specified by the organization where research was done.*

TIDAK TERHAD / OPEN ACCESS

Dr. SHAZLYN MULLIYANA SHAHARUDIN
Senior Lecturer
Department of Mathematics
Faculty of Science and Mathematics
Universiti Pendidikan Sultan Idris
35900 Tanjung Malim, Perak

afdal
(Tandatangan Pelajar/ Signature)

Dr. SHAZLYN MULLIYANA SHAHARUDIN
(Tandatangan Penyelia / Signature of Supervisor)
& (Nama & Cop Rasmi / Name & Official Stamp)

Tarikh: 14/5/2024

Catatan: Jika Tesis/Disertasi ini **SULIT @ TERHAD**, sila lampirkan surat daripada pihak berkuasa/organisasi berkenaan dengan menyatakan sekali sebab dan tempoh laporan ini perlu dikelaskan sebagai **SULIT** dan **TERHAD**.

Notes: If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentiality or restriction.



ACKNOWLEDGEMENT

First of all, I am beyond grateful to Allah S.W.T. for the strength that I've been blessed with throughout this journey full of challenges and filled with tears. My deepest appreciation and millions of thanks to my supervisor, Dr. Shazlyn Milleana Shahrudin for all the knowledge shared, time spent, and energy contributed to help me complete this study. All her support and sacrifices are surely unrequited. I would also like to thank my mother and my father, Fatimah Abustan and Basri Daud for always faithfully listening to the ups and downs in finishing this journey. Not forgetting also for my wife and friends who are endlessly giving me words of encouragement and praying for this success. I believe I could not have reached this point without the prayed and support from all of them. I would also like to thank the Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris for the chance and facilities provided, and also a special thanks to the Meteorological, Environment, and Geophysical Agency, Indonesia for the data and information provided for this study.





ABSTRACT

This study identifies torrential rainfall patterns in Yogyakarta, Indonesia using multivariate and univariate approaches to propose a statistical model for solving associated issues. First, addressing its long-gap missing rainfall data (approximately 52.8%) is crucial. Therefore, classical imputation methods were enhanced by combining them with the bootstrap algorithm. The hybrid of Random Forest (RF) and the bootstrap algorithm, with the lowest Root Mean Square Error (RMSE) of 7.96 and Mean Absolute Error (MAE) of 0.29, is the best statistical method for imputing Yogyakarta rainfall data missing values. Cluster analysis then classified stations into different rainfall regimes; hierarchical clustering analysis (HCA) recognised four distinct, homogenous regions. The multivariate approach, principal component analysis (PCA), homogenised the rainfall series and optimally reduced long-term rainfall data to validate the HCA analysis results. From the 75% cumulative variation, 14 factors for the Dry and the Rainy seasons and 12 factors for the Intermonsoon season were extracted using varimax rotation. Subsequently, the rainfall pattern forecasting was done using a univariate approach, Singular Spectrum Analysis (SSA). Recurrent Forecasting and Vector Forecasting-Singular Spectrum Analysis (RF-SSA and VF-SSA) were proposed by establishing length (L) and eigentriple (ET) parameters. This forecasting model effectively discriminated noise in a time series trend, producing significant forecasting results. Overall, the best performances are from $L = 6$ and $L = 5$ for the Rainy and Dry seasons and the Intermonsoon season, respectively, based on the lowest MAE and Mean Forecast Error (MFE). For the dry season, $L = 6$ has the lowest MAE (2.7116); the MFE tends to over forecast monthly rainfall data by 0.0126%. For the Intermonsoon season, $L = 5$ has the lowest MAE (3.3940); the MFE tends to under forecast by -0.0621%. All results utilised the RF-SSA algorithm. Concisely, the hybrid of the bootstrap algorithm and SSA improves the forecasting results. Different data classifications may advance climate warning systems.





PEMODELAN DESKRIPTIF BERDASARKAN CORAK HUJAN BULANAN MENGUNAKAN GABUNGAN PENDEKATAN IMPUTASI DENGAN ANALISIS MULTIVARIAT

ABSTRAK

Kajian ini mengenal pasti corak hujan lebat di Yogyakarta, Indonesia menggunakan pendekatan multivariat dan univariat untuk mencadangkan model statistik bagi menyelesaikan isu berkaitan. Pada asasnya, penting untuk menangani isu data hujan dengan jurang yang besar (kira-kira 52.8%). Oleh itu, kaedah imputasi klasik telah dipertingkatkan dengan menggabungkannya dengan algoritma *bootstrap*. Hibrid *Random Forest* (RF) dan algoritma *bootstrap* merupakan kaedah statistik terbaik untuk mengira data hilang dalam data hujan Yogyakarta yang menunjukkan Ralat Punca Min Kuasa Dua (RMSE) terendah iaitu 7.96 dan Min Ralat Mutlak (MAE) terendah iaitu 0.29. Analisis kluster kemudian mengelaskan stesen-stesen hujan ke dalam rejim yang berbeza; analisis pengelompokan hierarki (HCA) mengiktiraf empat kawasan homogen yang berbeza. Pendekatan multivariat, iaitu analisis komponen utama (PCA), menyeragamkan siri hujan dan mengurangkan data hujan jangka panjang secara optimum untuk mengesahkan keputusan analisis HCA. Daripada 75% variasi kumulatif, 14 faktor untuk musim Kering dan Hujan dan 12 faktor untuk musim Antara Monsun telah diekstrak menggunakan putaran varimaks. Tambahan, peramalan corak hujan dilakukan menggunakan pendekatan univariat iaitu Analisis Spektrum Tunggal (SSA). Ramalan Berulang dan Ramalan Vektor-Analisis Spektrum Tunggal (RF-SSA dan VF-SSA) telah dicadangkan dengan mewujudkan parameter panjang (L) dan *eigen triple* (ET). Model ramalan ini mendiskriminasi bunyi secara berkesan dalam aliran siri masa dan menghasilkan hasil ramalan signifikan. Secara keseluruhan, prestasi terbaik adalah daripada $L = 6$ dan $L = 5$ untuk musim Hujan dan Kering dan musim Antara Monsun, masing-masing berdasarkan MAE dan Ralat Ramalan Min (MFE) terendah. Bagi musim kemarau, $L = 6$ mencatatkan MAE terendah (2.7116) manakala MFE cenderung meramal data hujan bulanan secara lebih sebanyak 0.0126%. Bagi musim Antara Monsun, $L = 5$ menunjukkan MAE terendah (3.3940) manakala MFE cenderung meramal secara kurang sebanyak -0.0621%. Semua keputusan menggunakan algoritma RF-SSA. Secara ringkas, gabungan algoritma *bootstrap* dan SSA meningkatkan hasil ramalan. Klasifikasi data yang berbeza dikira berupaya memajukan sistem amaran iklim.



CONTENTS

	Page
DECLARATION OF ORIGINAL WORK	ii
DECLARATION OF DISSERTATION	iii
ACKNOWLEDGEMENT	iv
ABSTRACT	v
ABSTRAK	vi
CONTENTS	vii
LIST OF TABLES	xi
LIST OF FIGURES	xii
LIST OF ABBREVIATIONS	xv
LIST OF SYMBOLS	xvii
CHAPTER 1 INTRODUCTION	1
1.1 Background of Study	1
1.2 Study Area	6
1.3 Problem Statement	8
1.4 Research Questions	12
1.5 Research Objectives	12
1.6 Significance of Study	13
1.7 Limitations of Study	14
1.8 Thesis Organisation	15
CHAPTER 2 LITERATURE REVIEW	16
2.1 Introduction	16

2.2	Introduction to Missing Values	16
2.2.1	Missing Value Imputation Experimental Design	20
2.2.2	Dealing with Missing Data	24
2.2.2.1	Deletion Methods	25
2.2.2.2	Single Imputation Methods	27
2.2.2.3	Multiple Imputation Methods	27
2.2.3	Imputation Methods for Missing Data	28
2.2.4	Advancing Bootstrap Algorithm for Improved Missing Data Imputation	32
2.2.5	Performance Indicator for Imputation Methods	32
2.3	Introduction to Dimensional Reduction Methods	33
2.3.1	Identifying Rainfall Patterns Using Dimension Reduction Approach	42
2.3.2	Hierarchical Clustering Analysis	46
2.3.3	Singular Spectrum Analysis	50
2.4	Forecasting Model	51
2.5	Summary	53

CHAPTER 3 METHODOLOGY

3.1	Introduction	55
3.2	Research Methodology	57
3.2.1	Flow Chart of Research Methodology	57
3.2.2	Framework of Research Methodology	60
3.3	Imputation Methods	62
3.3.1	Replace by Mean	62

3.3.2	Nearest Neighbor	63
3.3.3	Markov Chain Monte Carlo (MCMC)	64
3.3.4	Non-linear Interactive Partial Least-Square (NIPALS)	66
3.3.5	Random Forest	67
3.3.6	Bagging Bootstrap Method	71
3.4	Research Novelty 1: Hybrid of Classical Imputations and Bagging Bootstrap Methods	73
3.5	Performance Indicator for Imputation Methods	74
3.5.1	Root Mean Square Error (RMSE)	74
3.5.2	Mean Absolute Error (MAE)	75
3.6	Cluster Analysis	75
3.6.1	Ward's Methods	77
3.7	Research Novelty 2: Principal Component Analysis (PCA)	78
3.7.1	Method of Selecting Numbers of Principal Component	81
3.8	Proposed Approach of Enhancement of Classical Imputations Combined with Multivariate Approaches	84
3.9	Singular Spectrum Analysis	85
3.10	Enhancement of SSA Forecasting Model	93
3.11	Summary	94
CHAPTER 4 RESULTS AND DISCUSSION		95
4.1	Introduction	95
4.2	Classification of Missing Data Mechanism	97

4.3	Clustering of Rainfall Stations	102
4.4	Statistical Properties of Rainfall Clusters	103
4.5	Cumulative Percentage of Principal Component Analysis (PCA)	107
4.6	Principal Component Analysis (PCA)	109
4.7	Forecasting Rainfall Data: A Case Study in Station Ledoknongko	117
4.7.1	Assessing Separability in Time Series Data	117
4.7.2	Comparison of Forecasting Methods	123
4.7.3	Forecasting Monthly Rainfall Data Using RF-SSA	134
4.8	Summary	137

CHAPTER 5 CONCLUSION 140

5.1	Introduction	140
5.2	Summary	140
5.3	Future Research	145

REFERENCES

APPENDICES

**LIST OF TABLES**

Table No.		Page
1.1	Geographical coordinates of 24 rainfall stations in Yogyakarta, Indonesia	8
2.1	Missing data rules according to Widaman (2006)	19
4.1	Statistical measure as indicator in selecting the best classical imputation method for monthly rainfall data in Yogyakarta, Indonesia	100
4.2	Statistical measure as indicator in selecting the best hybrid of classical imputation method and bootstrap algorithm for monthly rainfall data in Yogyakarta, Indonesia	101
4.3	Statistics for cluster-wise mean rainfall for the seasonal series	104
4.4	The seasonal rainfall amount within four clusters in a region	105
4.5	Summary of the PCA results for eigenvalues, percentage of variation and number of components obtained for each season in Yogyakarta, Indonesia	108
4.6	Loadings for the first 14 PCs of the Dry Season in Yogyakarta, Indonesia	110
4.7	Loadings for the first 14 PCs of the rainy season in Yogyakarta, Indonesia	113
4.8	Loadings for the first 12 PCs of the intermonsoon season in Yogyakarta, Indonesia.	115
4.9(a)	Window Length, L of Station 22 for dry and rainy seasons	117
4.9(b)	Window Length, L of Station 22 for intermonsoon	117
4.10	Comparison of Singular Spectrum Analysis Prediction Performance for Several Window Length (L).	124
4.11	Hybrid Random Forest-Bootstrap & RF-SSA Prediction Performance Several Window Lengths (L).	135



LIST OF FIGURES

No. Figures		Page
1.1	Locations of 24 rainfall stations in Yogyakarta, Indonesia	7
2.1	The experimental design procedure for missing value imputation	21
2.2	Spatial distributions of varimax rotated PC scores using annual precipitation data, a) PC-1, b) PC-2, c) PC-3, and d) PC-4.	45
3.1	Flow Chart of Research Methodology	59
3.2	Framework of Research Methodology	61
3.3	Algorithm of Random Forest	69
3.4	Procedures of PCA employed in the monthly rainfall data of Yogyakarta, Indonesia	81
3.5	Methodological flow of identifying regionalisation rainfall regimes using hybrid classical imputation methods coupled with multivariate approaches	84
3.6	Framework of Singular Spectrum Analysis (SSA)	92
3.7	Two enhancement procedures for developing SSA Forecasting Model	93
4.1	Missing percentage of data set	98
4.2	Correlation visualisation of missing and non-missing data for Yogyakarta, Indonesia	99
4.3	Clustered rainfall stations classified into four groups based on (a) dry season; (b) rainy season; (c) intermonsoon season.	103
4.4	Results of HCA for rainfall stations classified into 4 group based on (a) dry season, (b) rainy season, and (c) Intermonsoon in Special Region Yogyakarta, Indonesia.	106
4.5	Spatial distribution of monthly rainfall that resulted from principal component analysis in the dry season	112



with (a) PC1; (b) PC2; (c) PC3; (d) PC4; (e) PC5; (f) PC6; (g) PC7; (h) PC8; (i) PC9; (j) PC 10; (k) PC11; (l) PC12; (m) PC13; (n) PC14, respectively.

- | | | |
|------------|---|-----|
| 4.6 | Spatial distribution of monthly rainfall that resulted from principal component analysis in the rainy season with (a) PC1; (b) PC2; (c) PC3; (d) PC4; (e) PC5; (f) PC6; (g) PC7; (h) PC8; (i) PC9; (j) PC 10; (k) PC11; (l) PC12; (m) PC13; (n) PC14, respectively. | 114 |
| 4.7 | Spatial distribution of monthly rainfall that resulted from principal component analysis in intermonsoon season with (a) PC1; (b) PC2; (c) PC3; (d) PC4; (e) PC5; (f) PC6; (g) PC7; (h) PC8; (i) PC9; (j) PC 10; (k) PC11; (l) PC12; respectively. | 116 |
| 4.8(a–h) | W-correlation plot of dry season using SSA with varied windows length (a) $L = 6$, (b) $L = 7$, (c) $L = 8$ (d), $L = 10$ (e), $L = 13$, (f) $L = 25$, (g), $L = 50$ (h), and $L = 125$. | 120 |
| 4.9. (a–h) | W-correlation plot of rainy season using SSA with varied windows length (a) $L = 6$, (b) $L = 7$, (c) $L = 8$, (d) $L = 10$, (e) $L = 13$, (f) $L = 25$, (g) $L = 50$, and (h) $L = 125$. | 122 |
| 4.10(a–d) | W-correlation plot of intermonsoon using SSA with varied windows length (a) $L = 5$, (b) $L = 10$, (c) $L = 20$, and (d) $L = 50$. | 123 |
| 4.11(a-c) | Eigenvectors plot using Singular Spectrum Analysis: (a) dry season, (b) rainy season, and (c) intermonsoon. | 126 |
| 4.12(a-c) | First stage: Elementary Reconstructed Series (a) Dry Season ($L=6$), (b) Rainy Season ($L=6$), and (c) Intermonsoon ($L=5$). | 128 |
| 4.13 | (a) Data of rainfall reconstructed components from extracted trends using SSA at $L = 6$, (b) Logarithms of six eigenvalues, (c) Plots of eigenvectors (EV) pairs: 1-EV and 2-EV, 2-EV and 3-EV, 3-EV and 4-EV, 4-EV and 5-EV, 5-EV and 6-EV for Dry Season. | 130 |





- 4.14 (a) Data of rainfall reconstructed components from extracted trends using SSA at $L = 6$, (b) Logarithms of six eigenvalues, (c) Plots of eigenvectors (EV) pairs: 1-EV and 2-EV, 2-EV and 3-EV, 3-EV and 4-EV, 4-EV and 5-EV, 5-EV and 6-EV for rainy season. 132
- 4.15 (a) Data of rainfall reconstructed components from extracted trends using SSA at $L = 5$, (b) Logarithms of five eigenvalues, (c) Plots of eigenvectors (EV) pairs: 1-EV and 2-EV, 2-EV and 3-EV, 3-EV and 4-EV, as well as 4-EV and 5-EV for intermonsoon rainfall data. 134
- 4.16 (a–c) Predicted RF-SSA and confirmed rainfall data in Yogyakarta for different seasons. 137





LIST OF ABBREVIATIONS

AR	Autoregressive
BMKG	Regulation of Head of Meteorology, Climatology, and Geophysics Agency
BNPB	The National Agency for Disaster Management
CA	Cluster analysis
EM	Expectation-maximisation
EM-MCMC	Monte Carlo Markov chain method-based expectation-maximization
ET	Eigentriple
EV	Eigenvectors
FA	Factor analysis
GDP	Gross domestic product
HCA	Hierarchical Clustering Analysis
IDWA	Inverse Distance Weighting Algorithm
IPCC	The Intergovernmental Panel of Climate Change
KNN	K-nearest neighbor
LRF	Linear recurrent formula
LRRs	Linear recurrence relations
MAE	Mean Absolute Error
MAR	Missing at Random
MCAR	Missing Completely at Random
MCMC	Markov Chain Monte Carlo
MI	Multiple imputations
MNAR	Missing Not at Random





MVI	Missing value imputation
NIPALS	Nonlinear Iterative Partial Least Squares
NN	Nearest Neighbor
NR	Normal Ratio
PC	Principal components
PCA	Principal component analysis
PD	Pairwise deletion
RF	Random Forest
RF-B	Random Forest-Bootstrap
RF-SSA	Recurrent Forecasting-SSA
RSME	Root Mean Square Error
SI	Single imputations
SSA	Singular Spectrum Analysis
SVD	Singular Value Decomposition
VF-SSA	Vector Forecasting-SSA





LIST OF SYMBOLS

O	Matrix of dummies
X	Data matrix
a	Centroid of clusters A
b	Centroids of clusters B
A	Cluster with centres a
B	Cluster with centres b
$B(t)$	Number of bootstrap replications
C	Number of iterations for convergence.
C_{tt}	Correlation matrix
d	Variables
$d(A, B)$	Ward's distance between two clusters A and B
$d_q(x_i, x_j)$	Computation formula of the distance
E_t	Matrix with columns of eigenvectors
g	Components
$h_{d,g}$	Each entry of matrix of component loadings
H_e	$n \times n$ matrix of component loadings
i	Index
j	Index
K	Number of clusters
l_k	Variance of principal component
m	Number of components
n	Number of rainfall days





n_A	Frequency of clusters A
n_B	Frequency of clusters B
NA	Number of the missing values in the categorical variables
o_{is}	Corresponding matrix of dummies
p	Rainfall stations
p_h	Principal components
P_i	Initial observed rainfall data
P_x	Observed rainfall data
Q	Diagonal matrix of eigenvalues of correlation matrix
\mathfrak{R}	Coefficient vector
t_h	Principal components
t_m	Cumulative percentage of the variance of m
T	Total number of bootstrap replications
u_i	Left singular vectors
U	$L \times L$ orthogonal matrix
U_j	Eigenvector
v	Value in the domain of the target feature
v_i	Right singular vectors
V	$K \times K$ orthogonal matrix
x	Rainfall days
x_i	Measured value of rainfall
$x_{i,j}$	Elements in the input matrix, i.e. rainfall amount
x_{is}	The i th observation of the s th variable





X'	Spatial T-mode matrix
\mathbf{X}_i	Elementary matrices
X_s	Arbitrary variable
\tilde{X}_i	Reconstructed columns of trajectory matrix
X_{new}^{imp}	Newly imputed matrix
X_{old}^{imp}	Previously imputed matrix
$x_{1,i}^{B(t)}$	Bootstrap replication data through sampling with replacement
y_i	Observed rainfall data
\bar{y}_i	Average rainfall data over rainfall stations
\hat{y}_i	Predicted rainfall data
$y_{mis}^{(s)}$	Missing values of variable X_s
$y_{obs}^{(s)}$	Observed values of variable X_s
Y_{miss}	Missing values
Y_{obs}	Observed data
y_t	Reconstructed series
\tilde{Y}_T	Reconstructed time series value
Z	Rank of \mathbf{X}'
Z_i	Reconstructed time series value
λ	Eigenvalue
\vec{e}_t	Elements of eigenvector matrix
λ	Eigenvalue that obtained from correlation matrix
θ	Parameter of interest
Σ	Diagonal matrix





Σ_{ii}	Nonnegative real diagonal entries
σ_i	Singular values
σ_i, u_i, v_i	Eigen triple of singular value decomposition
$1(Y_j = v)$	Indicator function

