

Drought Assessment and Forecasting over Homogenous Climatic Regions of Pakistan



By:

Hamd Ullah

Reg. No. 06-FBAS/PHDST/F16

**Department of Mathematics and Statistics
Faculty of Basic and Applied Sciences
International Islamic University Islamabad
Pakistan
2022**

Accession No. TH-2638/V4

11.3
55.6
HAC

Drought:

Climate change

Statistical forecasting

Environmental forecasting - Pakistan

Drought Assessment and Forecasting over Homogenous Climatic Regions of Pakistan



By:

Hamd Ullah

Reg. No. 06-FBAS/PHDST/F16

Supervised By:

Dr. Muhammad Akbar

Co-Supervised By:

Dr. Firdos Khan

**Department of Mathematics and Statistics
Faculty of Basic and Applied Sciences
International Islamic University Islamabad
Pakistan
2022**

Drought Assessment and Forecasting over Homogenous Climatic Regions of Pakistan

By:

Hamd Ullah

Reg. No. 06-FBAS/PHDST/F16

A DISSERTATION

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

Supervised By:

Dr. Muhammad Akbar

Co-Supervised By:

Dr. Firdos Khan

**Department of Mathematics and Statistics
Faculty of Basic and Applied Sciences
International Islamic University Islamabad
Pakistan
2022**

Author's Declaration

I, **Hamd Ullah** Reg. No. **6-FBAS/PHDST/F16** hereby state that my Ph.D. thesis titled: **Drought Assessment and Forecasting over Homogeneous Climatic Regions of Pakistan** is my own work and has not been submitted previously by me for taking any degree from this university, **International Islamic University, Sector H-10, Islamabad, Pakistan** or anywhere else in the country/world.

At any time if my statement is found to be incorrect even after my Graduation the university has the right to withdraw my Ph.D. degree.


Name of Student: **(Hamd Ullah)**
Reg. No. **6-FBAS/PHDST/F16**
Dated: **26/09/2022**

Plagiarism Undertaking

I solemnly declare that research work presented in the thesis titled: **Drought Assessment and Forecasting over Homogeneous Climatic Regions of Pakistan** is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and University, **International Islamic University, Sector H-10, Islamabad, Pakistan** towards plagiarism. Therefore, I as an Author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of Ph.D. degree, the university reserves the rights to withdraw/revoke my Ph.D. degree and that HEC and the University has the right to publish my name on the HEC/University Website on which names of students are placed who submitted plagiarized thesis.

Student/Author Signature: _____


Name: (Hamd Ullah)

Certificate of Approval

This is to certify that the research work presented in this thesis, entitled: **Drought Assessment and Forecasting over Homogeneous Climatic Regions of Pakistan** was conducted by **Mr. Hamd Ullah**, Reg. No. **6-FBAS/PHDST/F16** under the supervision of **Dr. Muhammad Akbar** no part of this thesis has been submitted anywhere else for any other degree. This thesis is submitted to the **Department of Mathematics & Statistics, FBAS, IIU, Islamabad** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Mathematics, Department of Mathematics & Statistics, Faculty of Basic & Applied Science, International Islamic University, Sector H-10, Islamabad, Pakistan.**

Student Name: Hamd Ullah

Signature: 

Examination Committee:

a) **External Examiner 1:**
Name/Designation/Office Address

Signature:

Dr. Shabbir Ahmad
Assistant Professor,
Department of Statistics,
COMSATS University,
Wah Cant, Pakistan.



b) **External Examiner 2:**
Name/Designation/Office Address)

Signature:

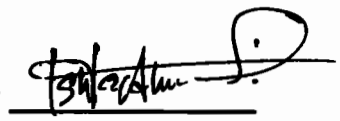
Dr. Iftikhar Hussain Adil
Assistant Professor of Statistics,
Department of Economics,
School of Social Sciences & Humanities (S3H),
National University of Sciences & Technology (NUST);
Sector H-12, Islamabad, Pakistan.



c) **Internal Examiner:**
Name/Designation/Office Address)

Signature:

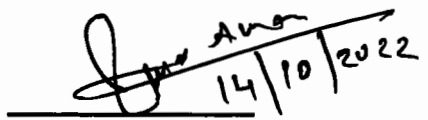
Dr. Ishfaq Ahmad
Associate Professor



Supervisor Name:

Dr. Muhammad Akbar

Signature:


14/10/2022

Co-Supervisor Name:

Dr. Firdos Khan

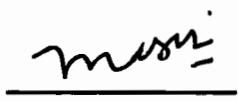
Signature:



Name of HOD:

Prof. Dr. Nasir Ali

Signature:



Name of Dean:

Prof. Dr. Muhammad Irfan Khan

Signature:



Acknowledgement

In the name of ALLAH, the most beneficent, the most merciful.

All the praise is for Allah Almighty who is the most merciful. I am extremely grateful to Allah Almighty for His blessings. My task of completing PhD is solely due to the blessings of Allah who gave me power to accomplish it. I would never have been successful without His support that has never left me helpless in the face of any challenge.

After that, I would like to acknowledge my supervisor **Dr. Muhammad Akbar** for his support and guidance throughout this PhD degree. Moreover, his continuous and vigilant supervision enabled me to understand deeply my field of study and completing this PhD degree. I would also like to acknowledge the knowledgeable and kind co-supervision of **Dr. Firdos Khan**, School of Natural Sciences (SNS), National University of Science and Technology (NUST) Islamabad. He is a great friend and a mentor. I sincerely appreciate invaluable academic and personal support I received from both of my supervisors throughout this study. Without their help, patience and encouragement, this research study would not have been possible. I pray with the depth of my heart that they and their families may receive countless and endless blessings from Almighty Allah. Furthermore, special thanks to **Dr. Ishfaq Ahmad**, all faculty members and supporting staff of Department of Mathematics and Statistics, IIU Islamabad for their sincere comments and suggestions. I am thankful to my PhD classmates and friends like Dr. Muhammad Amjad, Dr. Shafiq-ul-Rehman Khan and others for their moral support.

I would like to specially appreciate my parents, whose affection, support and prayers enabled me to achieve this and many other goals of my life. They encouraged me in every difficult situation. I am also thankful for the support and encouragement of my brothers specially Muhammad Ishaq. Last but not the least, I am very thankful to my beloved wife and cute children Madeeha, Maleeha, Javiria, Mubashira, and Zainab who suffered a lot during this whole period and for their patience, support and encouragement in hard times.

(Hamd Ullah)

Preface

Pakistan is facing serious threat of water scarcity, which requires water resources management. Water scarcity is directly linked up with climate change which includes frequent droughts and floods in different parts of the country. There is high variability w.r.t rainfall, temperature, and climate extremes that affect agriculture, hydropower, drinking water, industry, culture, and consequently the economy of Pakistan. According to German Watch, Pakistan is at 7th position in most vulnerable to climate change and its adverse impacts among the top ten countries in the world. Unfortunately, it has been less developed with low adaptation infrastructure. As a result, monitoring, assessing, and forecasting of droughts using statistical techniques have gained much attention. The drought events become more severe with growing water demand due to expansion of agricultural sector, energy sector, industrial needs, and rapid growth in population of Pakistan.

These and many other problems determine the need to plan these water-related hazardous issues and to know about the drought phenomenon in the country. It is important to know about the areal climatic changes based on regionalization w.r.t drought and to determine the variability according to similar climatic characteristics. Climatic variability over the country requires separate region-wise drought forecasting. Droughts are quantitatively measured by constructing indices. Since a number of drought indices are available in the literature, which may produce different projections and therefore, the analysis may be conducted by using more than one indices. Also, there is need for a deep and comprehensive study based on drought characteristics such as drought duration and severity in the country. Therefore, this study is designed to investigate comprehensive drought risk assessment in Pakistan. In this era of changing climate, it is necessary to be aware of the drought and wet conditions of the country for better water planning and management. It has also significant role in agriculture and socio-economic growth of Pakistan. This study consists of six chapters. A brief description of each chapter is given below:

Chapter 1 is introductory and contains a brief background of drought and water resources. A detailed literature review is included in the chapter. Literature contains several methods and its applications to assess, monitor, and forecast drought risk based on homogenous climatic regions (HCRs) and projections using various drought indices.

On the basis of literature review, research gap is extracted, and objectives of the study are specified. It is found that no statistical study has been conducted to identify homogeneous climatic regions and to assess region-wise drought risk in the case of Pakistan. Moreover, each of the drought index proposed in the literature has some limitations. Hence, objectives of the study are specified as (i). Quantitative measurement of drought phenomenon for the selected metrological stations across Pakistan, (ii). Identification of Homogeneous Climatic Regions (HCRs), (iii). Univariate and Bivariate drought projections at regional level in Pakistan, and (iv). Introduction of a new drought index. Lastly, some details are given about the significance of the study, scheme of the study, selection of metrological stations, and source and type of the data.

Chapter 2 presents the methodology to classify the areal changes of climate by constructing indices, and to construct HCRs w.r.t drought. Climate data of precipitation and temperature of 55 metrological stations across Pakistan are used to quantify drought phenomena by constructing indices for statistical investigation. Drought events for the stations are calculated using series of Reconnaissance Drought Index (RDI) and Standardized Precipitation Index (SPI) at 12-months' time scale based on threshold level of -0.85. The stations have different numbers of events for RDI and SPI and are used to calculate L-moment ratios for each metrological station of the study. Secondly, site characteristics (latitude, longitude, elevation, mean and standard deviation of precipitation) of stations are used to classify the metrological stations through cluster analysis into five subjective homogenous groups. Lastly, discordancy and heterogeneity measures are used for possible heterogeneity of subjective groups. Ultimately, five HCRs w.r.t RDI and SPI are identified over Pakistan. The research work of this chapter has been published in **International Journal of Climatology**; DOI: 10.1002/joc.6214 & **Theoretical and Applied Climatology**; DOI: 10.1007/s00704-020-03109-3.

In Chapter 3, the newly constructed HCRs are considered for future drought projections based on the RDI and SPI indices in Pakistan. According to the areal changes w.r.t drought and water resources, the regions need separate climate forecasting/projections for drought risk assessment. For this purpose, drought events are used to find the most suitable probability distributions from five 3-parameter

extreme value distributions using the L-moment ratio diagram and goodness of fit z-test. The selected probability distributions are estimated using L-moments technique and are used to find three types of drought projections based on statistical method of frequency analysis. Firstly, regional quantiles are calculated to cover a large area for drought risk assessment. These projections are more reliable for planning at a large level due to the maximum number of drought events from multiple sites in a region. Secondly, at-site quantiles are obtained by multiplying the drought mean value of the sites with regional quantile values for planning at the grass-root level using every single site of the study area. These quantiles show high variability and uncertainty among the results of the stations. The at-site quantiles have greater uncertainty compared to regional quantiles due to the lesser number of drought events at only a single site. Thirdly, sometimes there are vast areas with no gauging stations which need to be investigated. Hence, ungauged site projections are obtained to study the areas with no metrological stations using quadratic regression technique. According to 90% error bound and root mean square error, regional drought projections show high similarity at lower return periods but approximately after 10 years of return periods the variability and uncertainty increases. Regions 1 and 2 have no significant chances of drought while regions 3 and 4 have more chances of drought. Region 5 has a moderate drought condition. The results of this chapter are published in **Stochastic Environmental Research and Risk Assessment**; DOI 10.1007/s00477-020-01879-w and **Environmental Processes**; DOI: 10.1007/s40710-020-00478-9.

In **chapter 4**, the drought characteristics i.e., drought duration and drought severity, are used for a deep investigation of drought in Pakistan. Drought is a multi-faceted relationship of several correlated random variables such as drought duration and severity. For this purpose, the RDI and SPI indices are used to calculate the drought series at 12-months moving time scales and extracted the extreme drought events along with duration and severity using threshold level for each station. The extreme drought events are used to construct five bivariate homogenous climatic regions (BHCR) using drought duration and severity for more reliable drought estimates. The two variables are strongly correlated. In case of correlated variables, conventional statistical methods may be unreliable due to different probability distributions for the variables. Such correlated variables can be objectively tackled through copula function even if the variables have different probability distributions. For this purpose, five probability

distributions are checked, and most suitable probability distributions are selected for duration and severity variables within the BHCRs. Similarly, three bivariate single parameter copula functions are checked, and Gumbel-Hougaard (G-H) copula function is selected as best fit copula function through various statistical methods for all the BHCRs. Bivariate joint and conditional regional drought projections are obtained through G-H copula by joining the selected probability distributions using severity-duration frequency (SDF) curves and numerical formulas at 1.25, 2, 5, 10, 25, 50, and 100 years of return periods. For joint drought projections, the primary i.e., T_{OR} & T_{AND} while secondary i.e., T_{KEN} return periods are estimated where T_{KEN} lies in between T_{OR} & T_{AND} . Conditional return periods have high projections at fixed severity (or duration) compared to joint return periods. Regional projections show that regions 1, and 2 have moderate, region 4 has maximum, and region 5 with fewer chances of droughts repeating in the future. Whereas region 3 has mixed results for joint and conditional return periods. In regions 3 and 5, conditional return periods have abrupt changes after 30 months. The research work presented in this chapter have been published in **Arabian Journal of Geosciences**; doi.org/10.1007/s12517-021-08645-4 and **journal of Environmental Modeling and Assessment** (accepted on 17th September 2022).

In chapter 5, a new drought index for measuring drought risk measurement is proposed. Drought is an extreme hydrological event and proper management is necessary to get rid of potential future losses. Drought indices are the statistical tools for understanding and finding drought events and for simplifying the nexus of climatic variables. In this chapter, Standardized Copula-based Drought Index (SCDI) is proposed. The proposed index is based on the idea that the two aridity indices, i.e. UNEP aridity index (UAI) and De-Martone aridity index (DAI) based on precipitation, temperature, and potential evapotranspiration, are combined by employing Copula methodology. Five probability distributions are used to select best-fit distributions for UAI as well as DAI series. Both the drought series are strongly correlated and are suitable to use copula function. Four copula functions are tested to find the best fit Copula function to join the cumulative probabilities of UAI and DAI series. The joint series of cumulative copula function is standardized to get SCDI index for each respective station. Research work of this chapter is has been published in the **International Journal of Environmental Science and Technology**; <https://DOI:10.1007/s13762-022-04411-5>.

Chapter 6 presents conclusion, policy suggestions, and contributions of the study. Major contribution of the study is to classify the whole region of Pakistan into 5 homogeneous climatic regions w.r.t drought conditions using valid statistical techniques to make region-wise forecasting of drought condition while using both univariate as well as bivariate statistical approaches. Based on the conclusions, policy steps are suggested for combating drought risks in each of the region in Pakistan. According to the results, regions 1 and 2 have glaciers and rainfall with rare chances of drought and region 3 has more variability, which may be due to fewer metrological stations. In Region 4, there are low chances of drought occurrence on the basis of univariate analysis, but bivariate analysis provides more reliable results which predicts high and frequent chances of droughts in this region. No severe threats of drought occurrence are predicted in the region 5. Hence major threats of droughts are predicted in regions 3 and 4, which consist of the stations lying in Sind and Baluchistan provinces. Construction of dams in northern areas to store water, overcome deforestation, and incentives for plantation in the regions 3 and 4 are some of the important policy suggestions for policy makers. Lastly, a new multi-scalar Standardized Copula-based Drought Index (SCDI) is introduced by combining two other drought indices using precipitation, temperature, and PET data. It may be considered an important contribution in the literature of drought analysis.

List of Published Articles

1. Ullah H., Akbar M., Khan F. (2020). Construction of homogeneous climatic regions by combining cluster analysis and L-moment approach on the basis of Reconnaissance Drought Index for Pakistan. *International Journal of Climatology*, [USA, John Wiley and Sons Ltd; ISSN: 1097-0088; JCR IF=4.069; Q1; W category of HEC]; 40(1): pp#324-341. <https://doi.org/10.1002/joc.6214>.
2. Ullah H., Akbar M., Khan F. (2020). Droughts' projections in homogeneous climatic regions using Standardized Precipitation Index in Pakistan. *Theoretical and Applied Climatology*, [Austria, Springer-Verlag Wien, ISSN 1434-4483, 0177-798X; JCR IF=3.179; Q2; W category of HEC]; Vol. 140 (2), P#787-803. <https://doi.org/10.1007/s00704-020-03109-3>.
3. Ullah H., Akbar M., Khan F. (2020). Assessment of drought and wet projections in the humid climatic regions for Pakistan. *Stochastic Environmental Research and Risk Assessment*, [USA, Springer New York, ISSN: 14363259, 14363240; JCR IF=3.379; Q1; W category of HEC]; Vol. 34, Issue #12, P#2093-2106. <https://doi.org/10.1007/s00477-020-01879-w>.
4. Ullah, H., Akbar M. (2021). Bivariate homogenous regions and projections based on copula function using RDI and SPI indices for drought risk assessment in Pakistan. *Arabian Journal of Geosciences*. [Springer, Germany; ISSN: 1866-7538 (Electronic), 1866-7511(Print); JCR IF=1.827, Q3; X category of HEC], 14(22), 2338. <https://doi.org/10.1007/s12517-021-08645-4>.
5. Ullah H., Akbar M. (2021). Drought risk analysis for water assessment at gauged and ungauged sites in the low rainfall regions in Pakistan. *Environmental Processes*, [USA, Springer, ISSN: 2198-7491 (Print), 2198-7505 (Online); Scopus indexed, Journal Citation Indicator=0.25, X Category of HEC]; Vol 8, issue 1, P#139-162; <https://doi.org/10.1007/s40710-020-00478-9>
6. Ullah H., Akbar M., Khan F., Amjad M. (2022). A multi-scalar statistical approach to develop Standardized Copula-based Drought Index (SCDI) for drought risk analysis, *International Journal of Environmental Science and Technology* [Iran, Springer, ISSN: 1735-1472; JCR IF=3.51, Q2; W category of HEC]; <https://DOI:10.1007/s13762-022-04411-5>
7. Ullah H., Akbar M. (2022). Bivariate Drought Projections for Water Planning Using Copula Function in Balochistan Pakistan. *Environmental Modeling and Assessment*, (accepted on 17th September 2022). [Springer, Netherlands; ISSN: 1420-2026, 1573-2967; JCR IF=2.016, Q2; X category of HEC].

Table of Contents

List of Table.....	ix
List of Figures.....	xi
List of Abbreviations	xii
1 Introduction	1
1.1 Background	1
1.2 Literature Survey.....	6
1.3 Problem Statement / Research Gap in Pakistan	18
1.4 Objectives of the Study	20
1.5 Study Area and Data	20
2 Construction of Homogeneous Climatic Regions	24
2.1 Background	24
2.2 Material and Methods.....	25
2.2.1 Variables' Construction	25
2.2.3 L-moments Statistical Estimation Approach	33
2.2.4 Index-Flood Procedure.....	34
2.3 Results and Discussions	37
2.3.1 Statistical Measurements of Drought.....	37
2.3.2 Construction of Homogeneous Climatic Regions.....	38
2.3.3 Statistical Validity of the Identified HCRs	45
2.4 Conclusion.....	47
3 Regional Droughts' Projections	49
3.1 Background	49
3.2 Methodology	50
3.2.1 Selection of Best-Fit Probability Distribution	50
3.2.2 Estimation and Inverse Function of Distribution	52
3.2.3 Validation of Quantile Estimates	53
3.2.4 Quantiles at ungauged sites.....	54
3.3 Results and Discussion.....	55
3.3.1 Regional Frequency Analysis	55
3.3.2 At-Sites and Ungauged Sites Projections	64
3.4 Conclusion.....	70
4 Bivariate Drought Projections based on Copula Model(s)	72
4.1 Background	72

4.2	Methodology	73
4.2.1	Statistical Characteristics of Drought	73
4.2.2	Dependence between Drought Characteristics	74
4.2.3	Copula Modelling	74
4.2.4	Tail Dependence of Copula Models	78
4.2.5	Return Periods of Drought Characteristics	79
4.3	Results and Discussion.....	81
4.3.1	Construction of BHCRs	81
4.3.2	Copula modeling for bivariate drought Projections.....	86
4.4	Conclusion.....	104
5	Proposed Standardized Copula-based Drought Index (SCDI)	106
5.1	Background	106
5.2	Methodology	108
5.2.1	Aridity Drought Indices	108
5.2.2	Copula Modeling for SCDI.....	109
5.3	Results.....	112
5.3.1	Construction of SCDI Drought Index	112
5.3.2	Comparison of SCDI with SPI and RDI.....	119
5.4	Conclusion.....	126
6	Conclusions and Recommendations	128
6.1	Summary of the study	128
6.2	Conclusion and Policy Implementation	133
6.3	Future Work	138
	References	139
	Appendix-A.....	164
	Appendix-B.....	166

List of Table

Table 2. 1: Classification of RDI and SPI drought magnitudes for severity levels.	30
Table 2. 2: Critical values for discordancy measures (Hosking and Wallis, 1997)....	35
Table 2. 3: Discordancy and heterogeneity measures for climatic zones of Sheikh et al., (2009).....	40
Table 2. 4: Drought events, discordancy, and heterogeneity measures of the HCR... 	43
Table 2. 5: Regional mean and standard deviation of precipitation and temperature. 	47
Table 3. 1: The GOF Z-test results for the probability distributions of the regions....	58
Table 3. 2: Estimated parameters for the best fitted regional probability distributions of the regions	58
Table 3. 3: Regional projections, RMSE, and error bounds at various return periods.	61
Table 3. 4: At-Site Quantiles using IFP and QR Methods for RDI and SPI at return periods.....	66
Table 3. 5: At-Site quantile estimates for ungauged sites at selected return periods using QR Model for RDI and SPI indices.	69
Table 4. 1: Discordancy Measures for the sites of the five BHCRs w.r.t drought duration and severity using RDI and SPI indices.....	83
Table 4. 2: Heterogeneity Measures of drought duration and severity for RDI and SPI.	85
Table 4. 3: Correlation coefficients between drought duration and severity.....	87
Table 4. 4: Chi-Square goodness-of-fit measures for the five probability distributions.	87
Table 4. 5: Parameters estimates of the best-fitted regional probability distributions of drought duration and severity.	88
Table 4. 6: Parameters estimates of the selected Copula models for the regions.	90
Table 4. 7: Estimates of loglikelihood, AIC, and BIC criteria for the Copula models.	91
Table 4. 8: Dependence measures for the selected estimated Copula models.	92

Table 4. 9: Drought characteristics, Joint {Primary (TOR & TAND) & secondary (TKEN)} and Conditional return periods for BHCRs.	96
Table 5. 1: Results of Shapiro-Wilk goodness of fit test, z-, and p-values for the probability distributions (W-test, z-value & p-value).....	114
Table 5. 2: Results of Cramer Van Mises (S_n) goodness of fit test (parameter estimates/ S_n -test/p-value).....	116
Table 5. 3: Classification of standardized RDI, SCDI, and SPI values, where DI denotes drought index.	119
Table 5. 4: Correlation coefficients of SCDI with SPI and RDI at 3-, 6-, 9-, and 12-months' time scales.....	122
Table 5. 5: Percentages of drought explained in four drought categories by three indices (SCDI/SPI/RDI) using 3-, 6-, 9-, and 12-months' time scales for the selected stations of Baluchistan.	123

List of Figures

Figure 1. 1: Map of Pakistan with international geographical locations.	21
Figure 1. 2: Map of selected meteorological stations of the study area.....	23
Figure 2. 1: Graph of Run Theory using RDI and SPI series for Islamabad Station...37	
Figure 2. 2: Climatic Zones of Pakistan developed by Sheikh et al., (2009).	41
Figure 2. 3: Map of Pakistan showing the developed HCRs.	42
Figure 2. 4: Scatter plot of Mean Annual Precipitation and Elevation.....	45
Figure 2. 5: Region-wise comparison of Precipitation, PET, and Temperature.	47
Figure 3. 1: L-moment ratio diagram for selection of regional probability distributions.....	57
Figure 3. 2: Graphical Projection of RMSE for the regions at RDI and SPI indices. 60	
Figure 3. 3: Regional quantiles with 90% error bounds for the regions.	62
Figure 3. 4: Normal Q-Q plots for the Quadratic Regression equation.	65
Figure 4. 1: Bivariate HCRs of Pakistan based on drought duration and severity.....	85
Figure 4. 2: SDF curves of <i>TOR</i> and <i>TAND</i> return periods for the five regions.....	99
Figure 4. 3: SDF curves of conditional return periods of severity given at 10, 30, 50, and 90 percentile thresholds of durations (months), respectively.	101
Figure 4. 4: SDF curves of conditional return periods of durations (months) given at 10, 30, 50, and 90 percentile severity thresholds, respectively.	103
Figure 5. 1: Comparison of temperature (T) and PET at 3-, 6-, 9-, and 12-months' time scales for Noukundi and Quetta stations.....	117
Figure 5. 2: Comparison of CDFs of selected probability distributions at 3-, 6-, 9- and 12-months' time scales, respectively with the EDF of data for Quetta station.....	118
Figure 5. 3: Comparison of RDI and SPI with proposed SCDI for 3-, 6-, 9- and 12-months' time scales using Nokkundi and Quetta meteorological stations.	120

List of Abbreviations

AIC:	Akaike Information Criterion
AJK:	Azad Jammu and Kashmir
AMDSL:	Annual Maximum Dry Spell Length
AMS:	Annual Maximum Series
ASFA:	At-Site Frequency Analysis
BHCRs:	Bivariate Homogenous Climatic Regions
BIC:	Bayesian Information Criterion
CDF:	Cumulative Probability Functions
DAI:	De-Martone aridity index
EB:	Error Bounds
EDF:	Empirical Distribution Function
ESS:	Error Sum of Square
FCM:	Fuzzy C-Means
GCMs:	General Circulation Models
GEV:	Generalized Extreme Value
G-H:	Gumbel-Hougaard
GLO:	Generalized Logistic
GNO:	Generalized Normal
GOF:	Goodness of Fit
GPA:	Generalized Paretos
HCRs:	Homogeneous Climatic Regions
HKH:	Himalayas, Karakorum, and Hindukush
IFP:	Index-Flood Procedure
KAP:	Kappa
KPK:	Khyber Pakhtoon-Khwa
L-CV:	L-coefficient of variation
L-Kurt:	L-kurtosis

L-MRD:	L-moment ratio diagram
LN3:	Log-Normal Distribution with Three Parameters
L-Skew:	L-skewness
MAP:	Mean Annual Precipitation
MCS:	Monte Carlo simulations
PCA:	Principal Component Analysis
PDS:	Partial Duration Series
PDSI:	Palmer Drought Severity Index
PE3:	Pearson Type-3 distribution
PET:	Potential Evapotranspiration
PMD:	Pakistan Metrological Department
PWM:	Probability-Weighted Moments
QR:	Quadratic Regression
RCM:	Regional Climate Models
RDI:	Reconnaissance Drought Index
RFA:	Regional frequency analysis
RMSE:	Root Means Square Error
SCDI:	Standardized Copula-based Drought Index
SDF:	Severity-Duration Frequency
SDGs:	Sustainable Development Goals
SPEI:	Standardized Precipitation Evapotranspiration Index
SPI:	Standardized Precipitation Index
SPTI:	Standardized Precipitation Temperature Index
UAI:	UNEP Aridity Index
UN:	United Nations
UNEP:	United Nations Environmental Programme
WAK:	Wakeby

Chapter-1

Introduction

1.1 Background

Statistics is a Science that addresses the steps of collecting and improving data, analyzing and modeling objectives, validating and reporting results, and indicating fallacies in case of ignoring statistical reasoning (Weihs and Ickstadt, 2018). Statistical modeling is used for simplifying the complex interactions among factors using mathematical networks (Pearl, 1988; Koller and Friedman, 2009). It is an applied discipline to be used in engineering, environmental sciences, physical sciences, social sciences, business, and so forth (Ramachandran and Tsokos, 2020). Applied statistics is a collection of statistical techniques and applications using collected data for the statistical formulation, explanation, and validity of certain formulas or tests (Sachs, 2012). Hence, appropriate knowledge of statistical methods is necessary for meaningful conclusions about any random phenomenon. Natural and social issues are random events and have a certain degree of uncertainty to occur. Statistical methods are used to quantitatively measure the uncertainty of the results which is one of the main reasons for data collection and analysis (Naghattini, 2017). The climatic events are not constant and show irregular patterns and fluctuations to occur. Therefore, weather forecasts are undeniably uncertain (Wilks, 2011). Climatic events are random in nature and cannot be fully predictable. but statistical inference of data may provide sufficient information through weather forecasting. Environmental statistics are used for planning climatology and risk assessment using extreme climatic events due to their uncertainty of occurrence (Wilks, 2011). In statistics, some common natural issues for research are related to

modeling the effect and risk assessment of climate, environment, and water resource management (Hipel and Fang, 2013).

Drought is a naturally happening phenomenon with damaging properties to ecosystems, water resources, social activities, agriculture, and causes human losses, which is least understood and not easily controllable (Wilhite, 2012; Kis et al., 2017). All these issues are directly or indirectly related to Sustainable Development Goals (SDGs). 17 SDGs of the 2030 agenda set by the United Nations cover the dimensions of social, economic, institutional empowerment, environmental, etc., for the well-being of humans in the world (UN, 2022). These dimensions are assessed and planned based on datasets relevant to each SDG which are collected using statistical methods. Statistical methods work precisely which contributes to the improvement of sustainability due to its services in a better way at each stage in modeling and evaluating SDGs (Istat, 2021).

Drought is a slowly occurring phenomenon of climate and is defined as a persistent precipitation shortage over a region for a definite period (Beran and Rodier 1985). However, temperature, potential evapotranspiration (PET), and streamflow also have a strong role to characterize drought more comprehensively (Vicente-Serrano et al., 2010; Zargar et al., 2011). Bryant (1991) statistically characterized and ranked hazard events e.g., droughts, earthquakes, floods, tropical cyclones, and others based on characteristics including areal extent, degree of severity, economic losses, length of events, long-term impacts, loss of life, suddenness, social effects, and occurrence of associated hazards. Drought is ranked at the top of the hazard events based on most of these characteristics. It is a multivariate climatic issue that has more effects on people and the ecosystem compared to other hazards (Sönmez et al., 2005).

Drought planning and administration have three important steps: (1) monitoring and timely warning, (2) drought risk assessment, and (3) mitigation and early response to drought affected areas (Wilhite and Svoboda, 2000). In the first step, climate data i.e., rainfall, temperature, PET, and stream flow are used through drought indices for drought monitoring. Drought indices are statistical tools for understanding and finding drought events as well as for simplifying the multivariate relationship to manipulate data of single or multiple climate variables (Angelidis et al., 2012). An index gives a more comprehensive understanding of drought phenomenon for decision-making, planning, and management as compared to raw data of the variables (Hayes, 2006). The nature and type of drought indices reveal different conditions about dryness, abnormalities or deferred agricultural and/or hydrological effects like loss of soil moisture and depressed reservoir heights (Zargar et al., 2011). These effects are mainly due to the variables used in the development of drought indices. There are many types of drought indices, and each has its own merits and demerits which make it difficult to select the best one (Mishra and Singh, 2010). A single drought index may not reasonably quantify the severity and amount of drought effects (Heim, 2002). Therefore, more than one drought indices can be used to explain maximum drought variability of an area.

In the second step, drought risk is assessed in the form of regional projections using statistical techniques like frequency analysis. However, for regional projections, homogeneous climatic regions (HCRs) are constructed using drought events and characteristics based on statistical methods. The HCRs are groups of sites with similar drought characteristics and identical probability distribution which provide a basis for reliable impact assessment studies (Hosking and Wallis, 1997; Mirakbari et al., 2010). Hence, identification of HCRs for drought is an important step in devising effective

policies and combating the adverse impacts of climate change at the national level, particularly in the arid areas. Furthermore, statistical techniques are used to assess and forecast drought risk and monitor it at the local or regional level for better planning of agriculture, and water resources to get rid of future losses as much as possible. Regional frequency analysis (RFA) approach based on probability distributions is effectively used for drought risk assessment and forecasting all over the world (Feng et al., 2014; Ganguli and Reddy, 2014; Kaluba et al., 2017; Ullah et al., 2020b, c). Forecasting for one site is calculated using at-site frequency analysis. But mostly in climatic extreme events, there is a shortage of data at a site which minimizes validity of statistical projections (Goyal and Gupta, 2014). Therefore, HCRs are used to solve the problem of small data and to obtain regional drought forecasting using RFA.

Drought is a multi-faceted relationship of several correlated random variables i.e., drought duration and severity, where multivariate statistical analysis gives a better description and explains maximum variability (Mirabbasi et al., 2012). Drought duration and severity are important variables to be used for bivariate projections and drought risk assessment based on severity-duration frequency (SDF) curves (Hailegeorgis et al., 2013; Chebbi et al., 2013). Thus, multivariate statistical analysis may provide more information for drought assessment and forecasting. In the third step, the outcomes in steps one and two are utilized for mitigation and early response to drought affected areas or peoples.

Drought monitoring heavily depends on rainfall, but temperature, which regulates evapotranspiration and affects surface as well as groundwater levels, is also a crucial factor in the presence of drought (Topcu and Seckin, 2016; Qaisrani et al., 2021). Drought severity is increasing because of global warming and temperature is thus

regarded as a key factor in drought evaluation (Zhao and Dai 2015; Hui-Mean et al., 2018). Global warming may also possibly increase evapotranspiration more than precipitation (Trenberth et al., 2014; Vicente-Serrano et al., 2014). Water vapor is increased by 7% which increases the total capacity of precipitation by about 1-2% for warming of 1°C on the globe (Guardian, 2020). Hence, warming of climate might lead to extended droughts, heavy rainfall, and a higher risk of flooding due to melting glaciers that cause human fatalities, economic losses, and social issues. PET is underestimated in arid and semi-arid areas while overestimated in humid and semi-humid areas using the Thornthwaite equation (Jensen et al., 1990; Van der Schrier et al., 2011). Therefore, it is better to define a multifactor index that uses temperature and PET factors along with precipitation to explain more climatic variability in a region.

In this era of climate change, drought is a serious issue that affects water resources, agriculture, and human lives. This issue is worsening in developing countries like Pakistan where drought occurs more frequently. It impacts water availability and quality, which is directly affecting the economy and food supply of the country (Qadri et al., 2018; Qaisrani et al., 2019). However, water management is strongly affected by climate change which includes an increasing number of droughts and floods. There is high climatic variability w.r.t rainfall, temperature changes, and droughts in different parts of Pakistan. Firstly, the northeastern part consists of maximum elevated mountain ranges of Gilgit-Baltistan, Azad Jammu, and Kashmir (AJK), and attached areas of Khyber Pakhtoon-Khwa (KPK) with an extremely wet climate, world's 3rd largest glaciers, heavy rainfalls, and low temperature. Secondly, the southwestern part mostly consists of Balochistan and Sindh provinces which have an extremely arid climate, minimum rainfall, and maximum temperatures and are strongly affected by almost every drought in Pakistan (Anjum et al., 2012; Sajjad et al., 2014). Lastly, another part

consists of Punjab, and attached areas of KPK provinces have minimum elevation and a mixed climate with heavy rainfall in monsoon season, high temperature, and PET.

All the above factors provide a motivation for statistical exploration about droughts characteristics that have hit different areas and to make future projections of drought risks in Pakistan. Such an analysis might be helpful in formulating policy steps for combating worst effects of droughts and for water resources management. Before setting the objectives of the study, a detailed literature review about drought modeling with reference to Pakistan and other countries is given in the following section.

1.2 Literature Survey

Drought is one of the main factors of climate change including floods, wind speed, cyclones, and others. Many studies conducted for various countries are available in the literature. The studies were conducted using statistical techniques to analyze different characteristics of droughts and to make future projections of droughts' duration and severity at regional level in the countries. To identify drought variability, HCRs are constructed using site characteristics and drought events, which are further considered for univariate drought projections through a probabilistic approach of RFA. Drought is a multifaceted relationship of several characteristics e.g., drought duration and drought severity. It can better be assessed when more drought characteristics are considered for bivariate HCRs and projections. In literature, various studies were conducted for different countries using univariate and bivariate approaches to construct HCRs and for drought projections. A brief review of the studies would be helpful for drought modeling and future projections in the case of Pakistan.

Rahmat et al., (2017) used SPI index for the construction of six HCRs based on cluster analysis and modified Andrew's curve in Victoria, Australia. The HCRs were utilized

for arrangements and predictions of drought risk assessment. Alamgir et al., (2020) conducted a study to assess seasonal droughts through SPI index in Bangladesh using severity-area frequency curves. Drought projections were computed using observed precipitation data for the base period of 1961–2005 and compared with precipitation data from nineteen general circulation models (GCMs) with three periods, i.e., 2010–2039, 2040–2069, and 2070–2099. Hence, the study showed the percentage of areas affected by drought with different severity categories. Sadri and Burn, (2011) used data from 36 monitoring sites utilizing average monthly stream flow to construct three HCRs of droughts using drought duration and severity in three provinces of Canada. For HCRs, Fuzzy C-Means (FCM) method and L-moments technique were combined with discordancy and heterogeneity measures. Projections were made at different years of return periods for the regions. Núñez et al., (2011) used expected frequencies of low magnitudes of precipitation totals for 54 stations and observed monthly precipitation for 126 stations with different durations to study drought in arid regions of Chile. Probability distributions were selected and estimated through L-moments approach for return periods of drought using RFA. The 3-parameter Gaucho distribution which is a special form of the 4-parameter Kappa distribution was used to find quantiles at different years of return periods.

Liu et al., (2015) spatially distributed China into eight HCRs using Spatial "K" cluster Analysis by Tree Edge Removal method and Standardized Precipitation Evapotranspiration Index (SPEI) based on rainfall and temperature data of 810 stations ranging from 1961–2013. The study was carried out to support water resource management and agriculture development in China. Several studies were conducted for different areas of China. She et al., (2016) used daily data of rainfall during the rainy season from 1960 to 2014 to calculate the annual maximum dry spell length (AMDSL)

in days for 28 sites in Wei River Basin in China. K-means clustering algorithm with L-moments method was used to construct four HCRs. The regional best-fitted probability distributions were selected for the constructed HCRs to calculate the quantiles of dry spells at various return periods. This areal distribution has a significant role in the water resources management of the country. Feng et al., (2014) compared drought and normal series of 14 sites for a better drought risk assessment in Heihe River Basin, Northwest China. Data of precipitation for the period of 1960-2010 were used to calculate the extreme drought events and normal series using SPI and L-moments method for the selected sites. The stations were classified into three HCRs with different best-fitted regional probability distributions. RFA was applied to estimate the regional quantiles at the selected return periods for both the extreme drought events and normal series. In the regions, there were noticeable areal changes in drought and normal series of extreme precipitation. Zhang *et al.*, (2015) used the methods of FCM and multivariate L-moments homogeneity test based on drought severity, drought duration, and its joint effect for Pearl River basin, China. Daily climate data of 588 stations from 1960 to 2005 were analyzed using SPEI on a 12-months' time scale. The bootstrap sampling technique was used to evaluate the uncertainty using curves of joint probability. Five HCRs were constructed. Bivariate copulas were used to construct joint regional frequency curves, regional probability curves, and find averages w.r.t to drought duration, drought severity, and recurrence time among drought events. She and Xia (2018) conducted a study to assess drought risk using SPEI in the Loess Plateau in China. Exponential and Gamma distributions were best fitted to drought duration and severity variables along with Gumbel-Hougaard (G-H) copula instead of Frank and Clayton copulas. Univariate and bivariate probabilities and return periods were calculated for the joint structure of the variables at selected return periods.

Ghosh and Srinivasan (2016) used gridded precipitation data in the range of 1952-2007 for SPI at a 6-months' time scale in the Southern Peninsula of India and constructed seven HCRs w.r.t drought using k-means cluster analysis. The L-moments were calculated and selected Pearson Type-3 distribution (PE3) for six regions and Wakeby distribution for one region using the goodness-of-fit test. RFA is used to find univariate drought quantiles at various return periods in the regions. Ganguli and Reddy (2014) performed a multivariate analysis of drought risk assessment based on three drought characteristics i.e., drought severity, drought duration, and drought peak in Western India. The drought was modeled using SPI at a 6-months' time scale over the precipitation data ranging from 1896-2005. Trends among the SPI series were found using Mann-Kendall test by splitting the range into three parts i.e., 1896–1931, 1932–1966, and 1967–2005. Western Rajasthan, Saurashtra, Kutch, and Marathwada regions are constructed. For modeling the joint structure of drought characteristics, several Archimedean and elliptical classes of copulas were tested, and Student's t copula was found most suitable using goodness of fit test and tail dependence. The conditional probabilities, as well as joint return periods, were found through the selected copula function. The tri-variate frequency analysis for drought assessment was important over one and two variables frequency analysis and more helpful under changing climate.

Modarres (2009) conducted a study based on AMDSL of 37 rain gauge sites to study the drought risk in Isfahan Province, Iran. Cluster analysis resulted in two HCRs which were finally satisfied as homogenous using the L-moment method with different probability distributions. Sarhadi and Heydarizadeh (2014) used the data of daily precipitation of 67 sites with different lengths between 1951–2006 in Iran to find the AMDSL for drought risk assessment. Ward's cluster method was combined with the L-moments approach and constructed eight HCRs based on heterogeneity measure. The

best-fitted probability distributions were selected using goodness-of-fit test for the regions and found regional projections of AMDSL for drought risk. Amirataee et al., (2018) used monthly rainfall data of 24 stations from 1971 to 2013 across the Urmia Lake basin, Iran. The SPI index with a one-month time scale was applied to extract the variables of drought severity and percentage of drought-covered area. Seven copula functions from various families were used to select the best-fit copula function using selection criteria of AIC, BIC, and RMSE. These criteria selected Frank copula as the best-fitted copula function for the copula-based joint distribution of the severity-area frequency curve. The joint relationship reveals that most of the basin area was affected by severe and extreme drought as well as wet behaviors with significant variabilities. Montaseri et al., (2018) used the Gaussian copula function to compare the traditional and a newly proposed method for the two drought characteristics i.e., duration and severity in Iran. Exponential and Gamma distributions were selected for the two drought characteristics and calculated the joint probabilities and return periods of the sites. Nabaei et al (2019) conducted a study over 102 stations for the drought characteristics of drought duration, severity, and peak extracted from the SPI index in Iran. The best-fitted probability distributions were selected for each drought characteristic using Kolmogorov-Smirnov test and Chi-Square test. The cross-validation Copula Information Criterion was used to find the best-fitted copula from the three Archimedean Copula functions for the three groups of severity-peak, severity-duration, and peak-duration. The selected copula functions were used to calculate the drought projections at various return periods in the country. Bazrafshan, et al., (2020) conducted a study to investigate the drought condition using data from 25 stations for the arid and semi-arid regions in Iran. Drought duration and severity were used to identify three bivariate homogenous climatic regions. Different sets of generalized

logistic, generalized extreme value and Wakeby distributions were chosen as the most suitable marginal probability distributions for the three regions. Moreover, the G-H copula was selected suitable for region 1 while the Gaussian copula was best fitted for regions 2 and 3, respectively to model the joint dependence structure and return periods at selected years for the drought variables.

Santos et al., (2010) used rainfall data of 144 stations ranging from 1910 to 2004 for SPI to analyze spatial and temporal variability of drought in Portugal. The methods of principal component analysis (PCA) and k-means cluster analysis were applied to SPI with multiple time scales and defined three HCRs with different temporal changes of drought. Santos et al., (2011) continued the HCRs to perform frequency analysis of drought risk using L-moments approach. The drought events were extracted using annual maximum and partial duration series at a threshold level of -0.85. The drought projections at several return periods of annual maximum series showed better results compared to partial duration series. Almazroui et al., (2015) classified Saudi Arabia into five HCRs using precipitation and temperature data of 27 stations from 1985–2010 based on PCA technique. The main purpose was to assess economic planning, particularly in the semi-arid and arid areas with high climatic variability. Topcu and Seckin (2016) constructed HCRs through clustering combined with L-moment method using eleven metrological sites in Turkey. SPI at 3-, 6-, 9- and 12-month time scales were used resulting in two HCRs of five and six stations, respectively. The authors noted that elevations play a key role in the construction of HCRs. Regional Frequency Analysis was applied and was calculated drought quantiles at various return periods. Tosunoglu and Can (2016) studied meteorological droughts in Turkey using monthly data from 173 rain gauge stations from 1966 to 2006. PCA method was used to construct seven HCRs w.r.t rainfall in the country. The SPI index was used to determine

the drought characteristics of drought duration and severity that had a high correlation. Various marginal probability distributions were used to find the best fit distributions for the two variables using the Chi-square test. According to the test results, lognormal distribution was a better choice for duration series and generalized Pareto, Gamma, and Weibull distributions for severity series. Different best-fit copula functions were selected out of four different copulas for modeling conditional probabilities and joint return periods for drought durations and severities in the regions.

Dixit and Jayakumar (2022) developed a multivariate drought index (MDI) using copula functions for a more comprehensive drought analysis of meteorological, agricultural, as well as hydrological drought conditions. For MDI index, 4-variate Archimedean copula was performed using precipitation, soil moisture, streamflow, and evapotranspiration data. The MDI is used to find drought severity and duration over the changing climate for river basin in India using 6 Global Climate Models. The results show that precipitation with minimum and maximum temperature are decreasing with lower drought severity and duration according to future scenarios particularly at high emission scenarios. Botai et al., (2020) carried out the study to investigate the joint distribution of drought duration and severity, in the Eastern Cape Province of South Africa. SPI index with 6- and 12-month periods using the monthly data of 22 rainfall stations ranging from 1968-2018 were considered for the analysis. Five bivariate copula functions from Elliptical and Archimedean families were used to find the best-fitted copula function for the dependence measure of selected variables. Gaussian and Joe copula functions were used for the assessment of joint return period of drought duration and drought severity using dual as well as cooperative cases. The Tawn copula functions described the dependence structure with low probability for drought durations whereas high probability for drought severities. The joint return periods for dual cases

showed longer periods over all the univariate return periods of drought risk in the study area. Azam et al., (2018) used SPI index to extract drought duration and severity les using precipitation records of 70 meteorological stations in South Korea. Four HCRs were identified while Pearson type-3 and Kappa distributions were selected as marginal probability distributions for the two drought variables. Similarly, Gaussian and Frank copulas were declared as best fitted to calculate marginal and joint return periods for the regions. According to Achite et al., (2022), meteorological droughts and hydrological droughts are interrelated, and both are studied for planning of water resources in Ouahrane Basin, Northwest Algeria. The climate data of six rainfall stations with one hydrometric station from 1972-2018 were used through SPI and Standardized Runoff Index (SRI) at 1, 2, 3, 4, ..., 12 months timescales. Conditional return periods were found for both types of droughts using copula functions. Results show that mean severity of joint hydro-meteorological drought was 10.19, with 9 months duration, and 0.93 magnitude. Kaluba et al., (2017) used precipitation data from 35 stations and constructed five HCRs to calculate regional quantiles at various return periods of drought using RFA in Zambia. The Generalized Extreme Value distribution was selected using the goodness of fit test to calculate predicted values at different years of return periods for drought. The drought has different conditions in different regions of Zambia.

In the case of Pakistan, some studies containing drought analysis are available in the published literature. Hussain et al., (2011) constructed seven HCRs only for monsoon season (June to September) using climate data of 57 metrological sites for Pakistan. Different clustering methods were combined with Lambert projection method to satisfy the regions. Xie et al., (2013) used precipitation data ranging from 1960-2007 to calculate SPI results for the sites in Pakistan. SPI results were combined with the PCA

method which categorized overall drought patterns in the country. Adnan (2017) used SPI index for drought assessment in Sindh province with a focus on the monsoon season (July–September), using gridded precipitation data from 1951 to 2010. Haroon and Jiahua (2016) combined PCA and SPI results on a 3-months' time scale from January to March, to find drought changes from 1960-2013 all over the country. Khan et al., (2021) considered SPI and Standardized Precipitation Temperature Index (SPTI) indices at 3-month time scales for climate data ranging from 1998-2014 in twelve stations of Punjab province, Pakistan. Drought duration and drought severity variables were extracted which are strongly correlated. Several probability distributions were fitted, where Gamma and Weibull distributions were best fitted based on goodness of fit tests for duration and severity variables of the stations. Similarly, several copula functions were estimated and checked where G-H copula was the most suitable for maximum stations of the study area. This study is limited to only 12 stations in Punjab with no joint or conditional drought projections. Sheikh et al., (2009) distributed Pakistan into six general zones using physiographic and climatic characteristics of the country. Whereas no statistical assessment was done using any kind of data.

Analyses of different studies are based upon different drought indices. Drought indices are statistical measures that use single or multiple climate variables to get drought information. Drought events are extracted from drought indices that are used in constructing HCRs and projections. Hence, drought indices play a vital role in the quantitative measurement of drought to be used in this study. Drought index to be used for risk assessment and future projections play an important role and hence a brief review of drought indices is imperative to present here.

Drought indices has been widely used for assessment and prediction of drought events worldwide which has good effectiveness in managing drought risk assessment and

mitigation schemes (Ndayiragije and Li, 2022), According to Zargar et al., (2011), more than 150 drought indices have been proposed and increasing continuously, and also mentioned the simple, comprehensive, and combined indices categories. All these drought categories are due to the variables and type of indices combined and/or adjusted for the construction of an index. There are many types of droughts, but meteorological, agricultural, and hydrological droughts are common (Wilhite and Glantz, 1985). Another important type of drought is standardized drought indices which are used for monitoring and assessment of drought in a region. These indices are based on proper probability distributions and calculated at various time scales to obtain some nice drought information e.g., drought duration, severity, intensity, peak, beginning, and ending of drought. Standardized drought indices might be used for the evaluation of above mentioned three types of droughts e.g., 1- to 3-months' time scales are suitable for meteorological drought, 3- to 6-months for agricultural droughts while 9-, 12- and so on months are used for hydrological droughts to be used for water resource planning (Mishra and Singh, 2010). According to Datta and Reddy (2022), drought can effectively be monitored using multivariate drought indices like Multivariate Standardized Drought Index (MSDI) instead of univariate drought indices. MSDI was developed by combining precipitation with soil moisture data through copula functions. Further, MSDI is used to find drought duration and severity that highlighted the significance of multivariate analysis for drought risk assessment in Marathwada Region, India. In following we give a brief literature review of standardized drought indices.

McKee et al., (1993) introduced a standardized precipitation index (SPI) which can be computed at multiple time scales using precipitation data. The index is based on probability distribution with specific drought classifications of severity. There is an inverse relation between drought frequency and duration with change in time scale. The

index has some good characteristics of drought intensity, magnitude, beginning, and ending. It is the most used index due to its simplicity and the least data requirement. Tsakiris and Vangelis (2005) suggest that water deficit can be measured using precipitation as the input variable and PET as the output. The logic was used and introduced RDI index to utilize the stated input and output variables with more comprehensive results at several time scales like SPI. There are three steps in its formation: the initial values, normalized, and standardized values, respectively. The index can be more effectively used for hydrological and agricultural purposes and can be compared directly with the UNEP aridity index. The RDI is physically grounded, hence it determines the accumulated deficit between the atmosphere's evaporative requirement and precipitation. Vicente-Serrano et al., (2010) introduced the SPEI index based on the water balance equation by subtracting the accumulated value of PET from precipitation. Temperature data was used to calculate PET which plays the role of temperature variability in drought analysis. The log-logistic probability distribution was selected from several probability distributions using graphical and numerical statistical tools. The index can be calculated at multiple time scales and considered more representative due to PET, particularly for agricultural purposes. Ali et al., (2017) extended the De-Martone aridity index (DAI) to use precipitation and temperature data for a more robust technique of drought assessment. A new SPTI (Ali et al., 2017) has been constructed by adjusting the DAI using temperature for monitoring drought in the area. Further, the SPTI index has been compared with the existing SPI and SPEI indices to check its results by considering 17 sites in Khyber Pakhtunkhwa province, in Pakistan. These include extremely humid and arid sites. The results of SPTI have a strong correlation with SPI and SPEI. It also performed better than SPEI in minimum temperature areas for drought assessment. SPTI can be calculated at multiple time

scales like other standardized indices. Palmer (1965) developed the Palmer Drought Severity Index (PDSI) which is a standardized index with different levels of severities from SPI, RDI, and SPEI. PDSI is based on monthly data of eight types of climate data including evapotranspiration, recharge, runoff, loss, potential evapotranspiration, potential recharge, potential runoff, and potential loss. The index is based on a large set of data, therefore, may not be easy to use particularly in the developing countries where the climate system is weak.

Some important points may be extracted from the above review of literature. Many studies have been conducted to make drought projections based on statistical techniques for several countries including Pakistan. Most of the studies used various statistical techniques for the development of HCRs like PCA method, hierarchical and non-hierarchical clustering, discordancy measure, heterogeneity measures, L-moment approach, etc. Identification of HCRs was considered necessary to determine drought variability in the study areas and for drought projections using a probabilistic technique of RFA. Similarly, the studies used multivariate statistical techniques to explore drought conditions based on drought characteristics. Multivariate L-moments method, discordancy, and heterogeneity measures were used for the construction of bivariate HCRs through different copula functions. The copula models were used for joint projections of drought in the regions. Drought analyses and future projections of the studies provided guidelines for policymakers to make plans for water resources engineering, irrigation, disaster management, and many more. However, a review of literature reveals that no study has been conducted on drought analysis and future projections in the case of Pakistan. There is no comprehensive study that constructed univariate and/or bivariate HCRs for and future projections for drought risk assessment across the whole regions of Pakistan.

Moreover, the existing drought indices have several limitations (Mishra and Singh 2010; Zargar et al., 2011). For example, SPI is based only on precipitation data which is insufficient and explains limited drought variability (Vicente-Serrano et al., 2010; Lee et al., 2017). SPEI and RDI are based on precipitation and PET using Thornthwaite equation. However, PET is underestimated in arid and semiarid areas while overestimated in humid and semi-humid areas using Thornthwaite equation (Jensen et al., 1990; Van der Schrier et al., 2011). SPTI is based on precipitation and temperature data and can better be used in low-temperature regions (Ali et al., 2017). Whereas PDSI may not be easily computable in many regions due to a large amount of data. To overcome or minimize these limitations, a new multi-scaler standardized drought index may be developed using precipitation along with temperature and PET data.

1.3 Problem Statement / Research Gap in Pakistan

Currently, Pakistan has three major risks among others related to climate change and its impacts on population i.e., 1) flooding when melting glaciers, 2) high droughts in the southeast part, and 3) agriculture failure that causes food insecurity. All these risks bring massive disasters including human fatalities, agriculture, and socio-economic losses. Spatial and temporal drought and wet analyses may be used to reduce socio-economic impacts and to improve water resources engineering (Gocic and Trajkovic, 2013). Better water resources engineering has a significant role to fulfill water needs in Pakistan. According to German Watch, Pakistan ranks 7th among the top ten countries in the world in terms of climate change vulnerability and negative impacts (Eckstein, et al, 2016). In the last decades in Pakistan, rainfall-related events like droughts and floods have resulted in decline of agricultural output, livestock production, and human fatalities leading to greater economic losses (Ashraf and Routray, 2015). Pakistan has

severe threats and challenges of water scarcity and water resources management (UNDP, 2016) which also causes a severe crisis in the energy sector.

These and many other problems determine the need to plan these water-related hazardous issues and to know the climatic variability which is crucial for better climate planning, drought risk assessment, and water management in the country. It is important to know about areal climatic changes based on HCRs for drought. Further, the regions need independent climate forecasting to assess the future climatic variability for all areas of the country. Pakistan has severe threats of climate change. Therefore, these steps will help in devising effective policies to combat adverse impacts of climate changes and droughts management at a national level.

However, no study has been conducted to determine HCRs using drought results at annual time scales to consider all the seasons of the year and/or statistical methods for Pakistan. Sheikh et al., (2009) divided Pakistan into six general climatic zones based on physiographic and climatic features. But the study has no quantitative assessment based on climatic data or statistical tests. Furthermore, no study used reliable statistical tools to calculate regional drought projections using drought events and/or drought characteristics all over the country. Therefore, the present study is designed to investigate a complete and comprehensive drought risk assessment in Pakistan by locating HCRs and projections w.r.t drought events and characteristics i.e., drought duration and severity. This study will be helpful for disaster management departments, irrigation planning, policymakers, and water resources engineering to prepare plans for droughts as well as water in the country. Moreover, several drought indices are available in the literature but each of the existing drought indices has some limitations. Statistical techniques may be employed to develop a new drought index in order to

represent drought phenomenon by combining the existing indices. The index is based on precipitation, temperature, and PET data. Hence, this study is planned to work in detail on the following objectives related to drought risk assessment in Pakistan.

1.4 Objectives of the Study

This study is conducted:

1. To quantify drought phenomenon at the selected climatic stations over Pakistan by constructing drought indices.
2. To divide the stations into homogenous climatic regions.
3. To make future projection of drought risk in the homogenous climatic regions as well as at ungauged sites in the country.
4. To make future projections of drought duration and drought severity in the constructed regions using multivariate statistical models.
5. To develop drought severity-duration-frequency (SDF) curves for various return periods.
6. To propose new standardized drought index.

1.5 Study Area and Data

Pakistan lies in South Asia with latitudes from 24°N - 38° N and longitudes from 61°E - 76°E with a total area of 796,096 square kilometers. Map of geographical location along with the neighboring countries of Pakistan is given in Fig. (1.1). Pakistan is partitioned into four provinces of Punjab, KPK, Sindh, and Balochistan, while Northern Areas and AJK are directly administered by the federal government. In the north, the Himalayas, Karakorum, and Hindukush (HKH) are three great mountain ranges that join in a very complex system of mountains, separated by narrow gaps in the rivers.



Figure 1. 1: Map of Pakistan with international geographical locations.

Climate of Pakistan has significant regional variations and is categorized by hot and dry summer in the southern part while cold and wet winter in the extreme northern part of the country. Pakistan is distributed in four different climatic seasons (Sheikh et al., 2009). Firstly, monsoon season (June-September) with heavy rainfall sources in different parts of the country from the Arabian Sea and the Bay of Bengal (Adnan, 2017). Secondly, winter season (December-March) is one of the main sources of rainfall in the country. The Greater Himalayan mountains have mostly snowfall compared to rainfall. Snow and glacier melt keep the Indus Basin Rivers i.e., Indus, Jhelum, and Chenab recurrent during the whole year. Thirdly, the Pre-monsoon period

(April-June) which is very hot and dry, particularly in Balochistan, Sindh, and Southern parts of Punjab. Lastly, post-monsoon season (October-November) is generally very dry and establishes the shift zone between the monsoon and winter rainfall seasons.

Different nature of climate data is used to measure drought in the world including observed climate data, simulated/projected data from different Regional Climate Models (RCM) and GCM models, remote sensing data, and stream flow data of river basins. The nature of data changes the nature of drought measured in a region like metrological drought, agricultural drought, and hydrological drought. Metrological drought research is usually performed based on observed climate data of precipitation and temperature. This study focuses on metrological drought assessment and therefore, the observed climate data of precipitation and temperature are used for the selected metrological stations taken from Pakistan Meteorological Department, Ministry of Climate, Government of Pakistan (PMD, 2018). For statistically more reliable results, only those stations are selected in this study for which data of the climatic variables are available for at least thirty years. On the basis of availability of data, 55 meteorological stations are selected from all over the country given in **Fig. (1.2)**. The information on site characteristics, the geographical description of stations, the length of the data record, and necessary statistics of the sites are given in **Appendix-A**.

Missing values are the main irregularities in the time series of observed precipitation and temperature data where its number varies from site to site given in **Appendix-A**. The station of Ormara has a maximum number of missing observations but it is used to share some of the climatic information in the study results, due to very little number of stations from a vast area in Balochistan province. Regression method is used for

estimating missing observations in the precipitation and temperature data. The whole layout of the study is as follows.

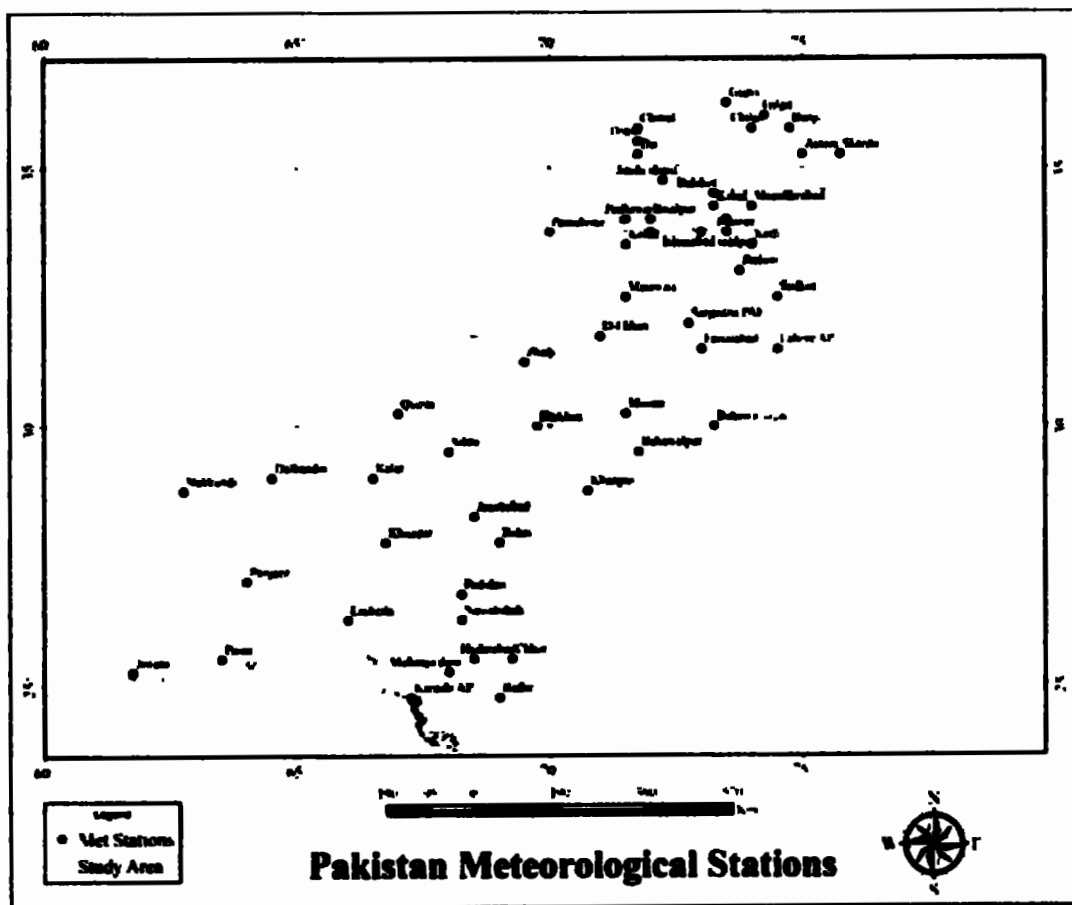


Figure 1. 2: Map of selected meteorological stations of the study area.

Chapter 2

Construction of Homogeneous Climatic Regions

2.1 Background

Assessment of the past climate is a primary step for planning and mitigation of droughts and water resources management (Quesada-Montano et al., 2018). The primary step for drought risk assessment is to classify the study area into HCRs with similar climatic characteristics using valid and relevant statistical techniques. The HCRs w.r.t droughts refer to groups of sites with similar statistical properties that can be used for further planning of drought and water resources management in the regions. This type of regional distribution may be used for regional drought management and improvement measures. Moreover, construction of HCRs is an important practice in hydrology that provide a basis for reliable impact assessment studies (Almazroui et al., 2015; Topçu and Seçkin, 2016; Ghosh and Srinivasan, 2016; Rahmat et al., 2017; Kaluba et al., 2017; Ullah et al., 2019; Ullah et al., 2020).

Several statistical techniques are used for the construction of HCRs with sites of similar drought phenomena for further investigation of drought features and mitigation schemes within the regions. Classification of data is a statistical process for organizing data into similar groups according to some shared characteristics for a more objective and reliable assessment (Timm, 2002; Härdle and Simar, 2019). Cluster analysis is one such standard multivariate statistical technique that can successfully be used for subjective homogenous classes of variables or metrological sites in hydrology. It is based on a criterion to classify the set of data into such classes that minimize the variation within a class while maximizing variation between classes (Rahmat et al., 2017). Some validity tests are used for the ultimate homogeneity of the subjective

homogenous regions. The sites in an HCR always have an identical probability distribution for drought magnitudes except for scaling factor (Mirakbari et al., 2010).

Pakistan consists of regions with high variability due to climatic conditions for rainfall, temperature changes, and drought in different parts of the country (Adnan et al., 2017). It requires to determine the regions of climatic similarity w.r.t. drought conditions which is crucial for better water management in Pakistan. To the best of the author's knowledge, Sheikh et al., (2009) is the only study that has been conducted to mention six climatic zones on the basis of physiographic and climatic characteristics of the country. However, the zones were not statistically or quantitatively assessed for homogeneity using climatic data or tests. Therefore, these zones need homogeneity testing based on observed climate data and statistical measures. Hence, objectives of this chapter include testing homogeneity of these zones and identifying HCRs w.r.t. drought conditions in Pakistan.

2.2 Material and Methods

We require some data on climate variables and suitable statistical techniques to determine HCRs w.r.t. droughts based on RDI and SPI indices in Pakistan. Both the drought indices are used to find drought events for the analysis. The method of clustering algorithm is used for subjective HCRs followed by the statistical tests to validate these subjective HCRs. These are explained in the following sub-sections.

2.2.1 Variables' Construction

The observed climate data of precipitation and temperature are used to construct the necessary variables for the study results. The details about observed data and stations are given in sub-section 1.6 and **Appendix-A**. The variables in this chapter are constructed in the form of drought series using drought indices and PET. The

Thornthwaite method is used for the calculation of PET from the average monthly temperature (T_{mean} , °C) of the sites (Thornthwaite, 1948) as follows:

$$PET = 16 \left(\frac{S}{12} \right) \left(\frac{d}{30} \right) \left(30 * \frac{T_{mean}}{H} \right)^a \quad (2.1)$$

where S is the possible mean monthly sunshine in hours, d is the days of a month, a is calculated using the following equation.

$$a = 6.75 * 10^{-7} H^3 - 7.71 * 10^{-5} H^2 + 1.79 * 10^{-2} H + 0.49 \quad (2.2)$$

where H is the heat index calculated as follows:

$$H = \sum_{l=1}^{12} \left(\frac{T_{l\ mean}}{5} \right)^{1.514} \quad (2.3)$$

There are various drought indices available in the literature, which include Palmer Drought Severity Index (Palmer, 1965), Effective Drought Index (Byun and Wilhite, 1999), Reclamation Drought Index (Weghorst, 1996), etc. However, we have selected RDI and SPI indices which are considered meteorological drought indices and can be calculated for multiple time scales. SPI is a standardized drought index using precipitation data only and was recommended by World Meteorological Organization (WMO) in December 2009. Similarly, RDI is also a standardized drought index using precipitation and PET data which is equally important for agricultural purposes. RDI and SPI indices give more reliable results with 30 or more years of climate data (Karavitis et al., 2011; Mondol et al., 2016).

RDI is a statistical tool to measure dry and wet magnitudes of the station by manipulating climate data. The index is based on probability distributions and is characterized as a meteorological drought index. It is more significant since drought is influenced by both temperature and precipitation, especially when investigating

agricultural droughts. RDI has three stages to be computed. Firstly, the ratio between accumulated totals of precipitation and PET for any time scale is obtained as follows:

$$\alpha_k^{(m)} = \sum_{n=1}^k P_{mn} / \sum_{n=1}^k PET_{mn}, m = 1 \text{ to } N \text{ \& } k = 1, 3, \dots \dots \quad (2.4)$$

where P_{mn} and PET_{mn} are precipitation and potential evapotranspiration (PET), respectively, of the n^{th} month at m^{th} year and N , is for the total number of years.

Secondly, the normalized RDI (RDI_n) is calculated as follows:

$$RDI_{n(k)}^{(m)} = \frac{\alpha_k^{(m)}}{\bar{\alpha}_k} - 1 \quad (2.5)$$

where $\bar{\alpha}_k$ is the mean of the $\alpha_k^{(m)}$ series. The initial construction of RDI_n by Tsakiris and Vangelis, (2005) assumes that $\alpha_k^{(m)}$ values follow the lognormal probability distribution. Thirdly, standardized RDI (RDI_{st}) is calculated as:

$$RDI_{st(k)}^{(m)} = \frac{y_k^{(m)} - \bar{y}_k}{\hat{\sigma}_k} \quad (2.6)$$

where $y_k^{(m)} = \ln(\alpha_k^{(m)})$, $\bar{y}_k$ is the arithmetic mean of $y_k^{(m)}$ and $\hat{\sigma}_k$ is its standard deviation.

However, after detailed analysis at several locations and time scales, it is found that $\alpha_k^{(m)}$ series follow lognormal and gamma distributions, but gamma probability density function shows the best fit for most locations and time scales (Tsakiris et al., (2008).

Therefore, RDI_{st} can be calculated using gamma probability density function as:

$$f(\alpha_k^{(m)}; \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} (\alpha_k^{(m)})^{\alpha-1} e^{-\alpha_k^{(m)}/\beta}, \text{ for } \alpha_k^{(m)} > 0 \quad (2.7)$$

where α is shape, and β is scale parameter estimated by using the maximum likelihood method as follow:

$$\alpha = \frac{1}{4A} \left(1 + \sqrt{1 + \frac{4A}{3}} \right) \quad (2.8)$$

$$\beta = \frac{\bar{\alpha}_k}{\alpha} \quad (2.9)$$

and A is given by:

$$A = \ln(\bar{\alpha}_k) - \frac{1}{n} \sum_{m=1}^n \ln(\alpha_k^{(m)}) \quad (2.10)$$

Where n is the number of observations of $\alpha_k^{(m)}$. Since gamma function is undefined for zeros of $\alpha_k^{(m)}$ series while precipitation may contain zero values. Therefore, zero's probabilities (q) are found as z/n , where z denotes the number of zeros in the $\alpha_k^{(m)}$ series and n denotes length of the time scale.

$$H(\alpha_k^{(m)}) = q + (1 - q)G(\alpha_k^{(m)}) \quad (2.11)$$

where $G(\alpha_k^{(m)})$ is cumulative probability of incomplete gamma function defined as:

$$G(\alpha_k^{(m)}) = \frac{1}{\beta \alpha \Gamma(\alpha)} \int_0^{\alpha_k^{(m)}} (\alpha_k^{(m)})^{\alpha-1} e^{-\alpha_k^{(m)}/\beta} d(\alpha_k^{(m)}) \quad (2.12)$$

The $H(\alpha_k^{(m)})$ is the cumulative probabilities for each $\alpha_k^{(m)}$ value to be converted to standard normal quantiles to get RDI_{st} by Merabti et al., (2017) as follow:

$$RDI_{st} = \Phi^{-1} \left(H(\alpha_k^{(m)}) \right) \quad (2.13)$$

Where $\Phi(\alpha_k^{(m)}; 0, 1)$ represent the standard normal distribution (Tsakiris et al., 2008).

SPI is a simple probabilistic drought index developed by McKee et al., (1993). The index can be found using accumulated precipitation data for various time scales like 1-, 3-, 6-, 9-, 12-months, and so on, which are suitable to understand the possible changes in drought conditions in the area as follow:

$$x = \sum_{n=1}^k P_{mn} \quad m = 1 \text{ to } N \text{ and } k = 1, 3, \dots \dots \quad (2.14)$$

Where P_{mn} is the value of precipitation of the n^{th} month at m^{th} year. Different probability distributions like gamma distribution, exponential distribution, lognormal distribution, and Weibull distribution can be used to calculate SPI index (Guenang and Kamga, 2014). However, gamma distribution is commonly used and considered the most suitable distribution whose distribution function is given below:

$$f(x; \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \text{ for } x > 0 \quad (2.15)$$

where α is shape, and β scale parameters estimated based on maximum likelihood method using the following equations.

$$\alpha = \frac{1}{4A} \left(1 + \sqrt{1 + \frac{4A}{3}} \right) \quad (2.16)$$

$$\beta = \frac{\bar{x}}{\alpha} \quad (2.17)$$

Where $\bar{x}$ denotes mean of drought values and A is calculated by the equation:

$$A = \ln(\bar{x}) - \frac{1}{n} \sum_{i=1}^n \ln(x_i) \quad (2.18)$$

Where n is the number of observations of x. While gamma function is as follows:

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y/\beta} dy \quad (2.20)$$

The cumulative probability ($H(x)$) is defined using the following equation:

$$H(x) = q + (1 - q)G(x) \quad (2.21)$$

Where q is the probability of zeros and $G(x)$ is the incomplete gamma function as:

$$G(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t/\beta} dt \quad (2.22)$$

SPI is found as standard normal quantiles of the $H(x)$ function by Merabti et al., (2017) as follows:

$$SPI = \Phi^{-1}(H(x)) \quad (2.23)$$

TH-26381

Where $\phi(x; 0, 1)$ is used to represent the standard normal distribution.

The cumulative probabilities of $H(x)$ can also be easily transformed to obtain standard values of SPI (also can be used for RDI in the above subsection) using Abramowitz and Stegun (1965) approximation as follow:

$$Z = SPI = - \left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} \right) \quad (2.24)$$

Where $t = \sqrt{\ln \left(\frac{1}{(H(x))^2} \right)}$ and $0 < H(x) \leq 0.5$.

$$Z = SPI = + \left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} \right) \quad (2.25)$$

Where $t = \sqrt{\ln \left(\frac{1}{(1.0 - H(x))^2} \right)}$ and $0.5 < H(x) < 1.0$

where $c_0 = 2.515517$, $c_1 = 0.802853$, $c_2 = 0.010328$, $d_1 = 1.432788$, $d_2 = 0.189269$, and $d_3 = 0.001308$.

SPI and RDI have identical severity levels for drought magnitudes given in Table (2.1).

Table 2. 1: Classification of RDI and SPI drought magnitudes for severity levels.

DI value	Classes
$DI \geq +2.0$	Extremely wet
$+1.5 \leq DI < +2.0$	Severely wet
$+1.0 \leq DI < +1.5$	Moderately wet
$-1.0 \leq DI \leq +1.0$	Near normal
$-1.5 < DI \leq -1.0$	Moderate drought
$-2.0 < DI \leq -1.5$	Severe drought
$DI \leq -2.0$	Extreme drought

Note: The DI stands for Drought Index (RDI and SPI indices)

RDI and SPI are used to quantify and monitor drought conditions of a metrological site.

This study is mainly focusing on hydrological drought risk assessment in Pakistan,

especially in drought affected regions. Therefore, RDI and SPI are limited to only annual time scale for a water year (October-September) to obtain drought information.

A threshold level is used to generate extreme drought data for statistical analysis of extreme events and future planning (Santos et al., 2011). Subject knowledge and experience are necessary to identify threshold levels (Karim et al., 2017). However, small threshold levels give week approximation which causes bias in the estimated returns while the large level increases variance in estimated parameters due to fewer observations (Roth et al., 2016; Ullah et al., 2020). In literature, different threshold levels are used e.g., -0.5 by Liu et al., (2015), -0.8 by Ganguli and Reddy (2012), -0.85 by Santos et al., (2011), whereas -1 by Goyal and Sharma (2016). Drought begins when a negative magnitude of drought series hits -0.85 (Agnew, 2000). Hence in this study, -0.85 is selected as the threshold level for the identification of drought events.

In hydrology, samples are either selected using annual maximum series (AMS) which considers only a single value, or partial duration series (PDS) which considers all the drought values that exceed the threshold level within a year. AMS is commonly used for HCRs and RFA for droughts, floods, and wind speeds (Hassan and Ping, 2012; Shahzadi et al., 2013; She et al., 2016; Fawad et al., 2018). Therefore, in this study, both drought indices are calculated at annual time scales for the selected stations. Run theory method is used to extract drought events as well as drought characteristics (Yevjevich, 1967). Negative signs of the events are ignored in the analysis of drought (Santos et al., 2011; Goyal and Gupta, 2016).

2.2.2 Cluster Analysis

It is important to introduce multivariate statistical methods in climate change analysis (Huth and Pokorna, 2005). One such multivariate statistical technique is cluster analysis

which is used for subjective homogenous grouping of the observations, gauging stations, and/or climatic variables. According to Hosking and Wallis (1997) initially, site characteristics are used for subjective HCRs through clustering methods and then at-site statistics are utilized for testing homogeneity of the subjective regions. Site characteristics are the geographical information of the gauging sites e.g., latitude, longitude, elevation, mean annual precipitation, standard deviation of the site, etc. At-site statistics are numerical measurements obtained from data of meteorological stations such as L-CV, L-Skewness, and L-Kurtosis. Various clustering methods are available in the literature. However, we apply any one of the following two clustering methods. Ward's clustering method (Ward, 1963) with Euclidean distance for the construction of HCRs produces good results in hydrology (Hosking and Wallis, 1997). The same combination of wards and Euclidean distance is also used by Malekinezhad et al., 2011; Hassan and Ping, 2012; Lyra et al., 2014). The Euclidean distance between any two meteorological stations i and j can be calculated as follows.

$$d = \left(\sum_{i,j=1}^n (x_i - x_j)^2 \right)^{1/2} \quad (2.26)$$

where x is any characteristic of the two sites. The ward's algorithm combined with Euclidian distance is used to minimize the error sum of square (ESS) between the objects or observations.

The k-means cluster algorithm is used as it is simpler, flexible, and has convergence and invariance properties (Celebi and Kingravi, 2012). This algorithm combines N sites, with q -dimensional characteristics into k groups in which the distance is minimized between the sites in a group to its center value. The algorithm converges to

the best solution for groups if changing sites would not further decrease the sum of squares of error value (Hartigan and Wong, 1979).

After clustering, the sites of subjective HCRs need not be geographically attached and final. Some necessary adjustments to the sites are made for physical improvement and homogeneity of the HCRs, if required. The L-moments statistical method of estimation is used to perform the Index-Flood Procedure (IFP). The IFP procedure is used to validate the subjective HCRs and find projections. The L-moments method and IFP procedure are explained in the following sub-sections.

2.2.3 L-moments Statistical Estimation Approach

In extreme events like floods and droughts, the probability distributions are mostly skewed with small sample sizes containing outliers. In such cases, the conventional methods of estimation e.g., maximum likelihood method and method of moments, do not give reliable estimates. The conventional estimators give biased and non-normally distributed estimates in small and moderate samples (Wallis et al., 1974). The L-moment estimation method is an alternative system to the conventional moments and gives more reliable and unbiased estimates for the above-referred situation. L-moments are linear combinations of probability-weighted moments introduced by Hosking and Wallis (1993). It is more robust when there are outliers in the data and unbiased in the case of small samples. Let $x_{1:n} \leq x_{2:n} \leq \dots \leq x_{n:n}$ be a sample of drought values in ascending order of magnitude then the first four probability-weighted moments (PWM) are as below:

$$\beta_0 = \frac{1}{n} \sum_{i=1}^n X(i) \quad (2.27)$$

$$\beta_1 = \sum_{i=1}^n \left(\frac{(n-i)}{n(n-1)} \right) X(i) \quad (2.28)$$

$$\beta_2 = \sum_{i=1}^n \left(\frac{(n-i)(n-i-1)}{n(n-1)(n-2)} \right) X(i) \quad (2.29)$$

$$\beta_3 = \sum_{i=1}^n \left(\frac{(n-i)(n-i-1)(n-i-2)}{n(n-1)(n-2)(n-3)} \right) X(i) \quad (2.30)$$

where, $X(i)$ are the ordered values. The four population L-moments are as follows:

$$\lambda_1 = \beta_0 \quad (2.31)$$

$$\lambda_2 = 2\beta_1 - \beta_0 \quad (2.32)$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \quad (2.33)$$

$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \quad (2.34)$$

Hosking (1990) defined L-moment ratios for the population presented as follows.

$$\text{L-coefficient of variation (L-CV): } \tau = \lambda_2 / \lambda_1 \quad (2.35)$$

$$\text{L-skewness (L-Skew): } \tau_3 = \lambda_3 / \lambda_2 \quad (2.36)$$

$$\text{L-kurtosis (L-Kurt): } \tau_4 = \lambda_4 / \lambda_2 \quad (2.37)$$

The sample counterparts of PWMs are represented by b_0 , b_1 , b_2 , and b_3 , L-moments by l_1 , l_2 , l_3 , and l_4 while the L-moment ratios by t , t_3 , and t_4 , respectively.

2.2.4 Index-Flood Procedure

The Index-Flood Procedure (IFP) is a combination of various statistical tests and methods formulated by Hosking and Wallis (1997). This procedure was developed for flood events but later it was equally used for rainfall, wind speed, and droughts (Malekinezhad and Zare-Garizi, 2014; Yin et al., 2015; She et al., 2016; Topcu and Seckin, 2016; Fawad et al., 2018). It has mainly two parts, the first to construct HCR and the second to perform RFA using L-Moment's ratios with the following five steps:

- a) Apply discordancy measures to identify the discordant sites,
- b) Apply heterogeneity measures to the region,

- c) Identify best fit probability distribution for the regions.
- d) Find parameters of the best fit probability distribution, and
- e) Find quantile function for drought projections.

The first two steps are related to HCR and will be applied in this chapter while the remaining steps are used for RFA which would be explained in chapter 3.

Discordancy Measure (D_m): It is a statistical measure used for data screening that checks the appropriateness of the data for any errors or anomalies. The measure is used for the identification of discordant site(s) in a group of sites when the statistical analysis is performed. The discordancy measure is based on sample L-moment ratios such as L-CV, L-skewness, and L-kurtosis using drought magnitudes of the gauging stations.

Let us consider a group containing N sites then the D_m value for the i^{th} gauging site ($i = 1, 2, \dots, N$) is calculated using the following equation:

$$D_m = \frac{1}{3} N (u_i - \bar{u})^T S^{-1} (u_i - \bar{u}) \quad (2.38)$$

Where u_i is a vector of L-moment ratios i.e., $u_i = [t^{(i)} \ t_3^{(i)} \ t_4^{(i)}]^T$, $\bar{u}$ is the mean i.e.,

$\bar{u} = N^{-1} \sum_{i=1}^N u_i$ and S is a matrix of sums of squares and cross products defined as $S = \sum_{i=1}^N (u_i - \bar{u})(u_i - \bar{u})^T$.

When D_m value is large than critical value in Table (2.2), the station is considered discordant.

Table 2. 2: Critical values for discordancy measures (Hosking and Wallis, 1997).

No of sites in a region	5	6	7	8	9	10	11	12	13	14	≥ 15
Critical value	1.333	1.648	1.917	2.140	2.329	2.491	2.632	2.757	2.869	2.971	3.000

Heterogeneity Measures: Heterogeneity measures (H_r , $r = 1, 2, 3$) are statistical measures used to assess the degree of similarity of sites in a group. It is based on

observed and expected estimates of sample L-moment ratios of the sites in a group. For expected estimates, we rely on Monte Carlo simulation by generating N_{simu} similar regions from fitted 4-parameter Kappa probability distribution through regional L-moment ratios. The variation is calculated as standard deviation (S_r , $r=1, 2, 3$) using L-moment ratios as follows:

$$S_1 = \left(\sum_{i=1}^{N_r} N_i (t^i - t^R)^2 / \sum_{i=1}^{N_r} N_i \right)^{1/2} \quad (2.39)$$

$$S_2 = \left(\sum_{i=1}^{N_r} N_i ((t^i - t^R)^2 + (t_3^i - t_3^R)^2) \right)^{1/2} / \sum_{i=1}^{N_r} N_i \quad (2.40)$$

$$S_3 = \left(\sum_{i=1}^{N_r} N_i ((t_3^i - t_3^R)^2 + (t_4^i - t_4^R)^2) \right)^{1/2} / \sum_{i=1}^{N_r} N_i \quad (2.41)$$

Where N_r is total sites in the i^{th} region, t^i is the ratios of the r^{th} site and t^R is the regional average of ratios of all the sites of a region. The simulated regions are assumed to be homogeneous with an equal number of sites having the same length of records as from observed values. The heterogeneity measure of sample L-moment statistics from observed and simulated series are calculated using the relation:

$$H_r = \frac{(S_r - \mu_s)}{\sigma_s}, \text{ for } r=1, 2, 3. \quad (2.42)$$

where μ_s and σ_s denote average and standard deviation from simulated counterparts of observed S_r , respectively. Hosking and Wallis (1997) classified the value of H_r based on its magnitude. If $H_r < 1$ then a region is termed as acceptably homogeneous, if $1 \leq H_r < 2$ then the region is termed as possibly heterogeneous and if $H_r \geq 2$ then it is termed as definitely heterogeneous. H_1 , H_2 , and H_3 are three different heterogeneity measures, where H_1 -statistic is based on L-CV, and is a more significant measure of

heterogeneity, H_2 -statistics is based on the sample L-CV/L-skew and H_3 -statistics is based on the sample L-skew/L-Kurt, respectively.

2.3 Results and Discussions

2.3.1 Statistical Measurements of Drought

Standardized series of RDI and SPI is calculated at a 12-month time scale using precipitation and temperature data of 55 meteorological stations with different lengths of more than thirty years. Drought indices with 12-months (annual) are considered longer time scales to describe hydrological drought and water resources deficiency in a region (Mishra and Singh, 2010; Haroon and Jiahua, 2016). RDI is computed based on precipitation and PET data and therefore, average monthly temperature data is converted to PET using equation (2.1). SPI index is based on only precipitation records of the sites. The method of run theory is used to extract extreme drought events which are less than or equal to -0.85 threshold level. Diagrams of a run theory are calculated for the selected stations of both drought indices. But diagram of only Islamabad station is presented here in Fig. (2.1), due to the larger space required. The stations have a different number of extreme drought events (n_{RDI} , n_{SPI}) given in Table (2.4) for sites. These events are used to calculate sample L-moments and L-moments ratios called at-site statistics for the sites used for identification of HCR and drought projections.

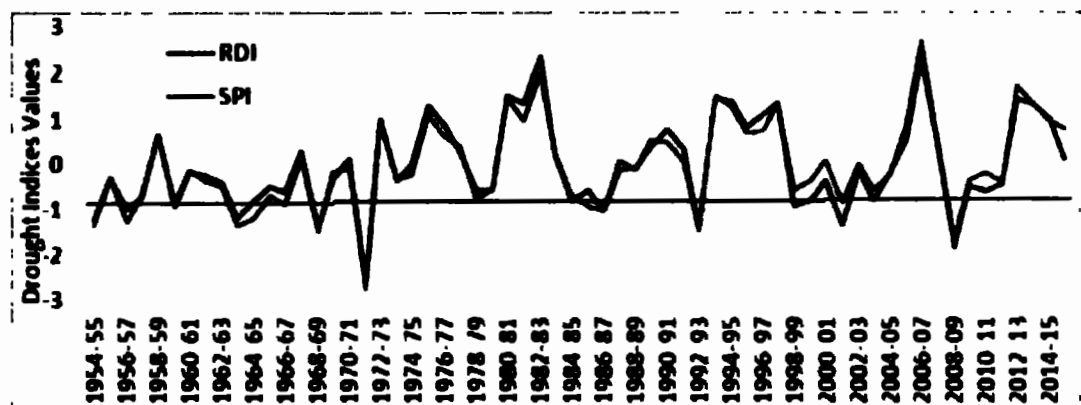


Figure 2. 1: Graph of Run Theory using RDI and SPI series for Islamabad Station.

2.3.2 Construction of Homogeneous Climatic Regions

Detection of homogeneous regions is necessary for more reliable forecasting of droughts (Saf, 2010). HCRs are detected in two steps. As an initial step, the proposed geographical zones of Sheikh et al., (2009) are checked using at-site statistics for possible heterogeneity, given in Fig. (2.2). There are six climatic zones while zones 1 and 5 are internally partitioned into zones "a" and "b". The discordancy and heterogeneity measures are employed using equations (2.38) and (2.42), respectively and the results are presented in Table (2.3). Both parts of zones 1 and 5 are checked individually as well as combined. According to the results, zone 1(a and b), zone 5a, zone 5(a and b), and zone 6 have one discordant station using the discordancy measure based on RDI index. Heterogeneity measures show that zones 1b and 5a are heterogeneous based on both RDI and SPI indices while zones 1(a and b), 3, 5(a and b) and 6 are heterogeneous based on RDI index only. A region is considered homogenous when it satisfies both discordancy and heterogeneity measures. If any of the two statistical measures are not satisfied, the region is concluded as heterogeneous. These statistical measures are considered more robust and reliable in hydrological sciences for the construction of HCRs all over the world (Goyal and Sharma, 2016; Ghosh and Srinivasa, 2016; Fawad et al., 2018). Hence it is finally concluded that HCRs constructed by Sheikh et al., (2009) are not homogeneous using climatic data.

In the next step, a multistage statistical procedure is used to classify the HCRs of Pakistan. In the first stage, k-mean clustering algorithm is performed for the subjective homogenous regions based on RDI and SPI indices using site characteristics i.e., latitude, longitude, elevation, mean annual precipitation, and its standard deviation of the meteorological stations (Hosking and Wallis, 1997). Five subjective homogenous groups are initially located which indicates that elevation has a key role in the

construction of regions. These groups may or may not be statistically homogeneous w.r.t RDI and SPI drought magnitudes. Adjustments are made when some sites are added, changed, or even removed to improve the homogeneity and physical structure of the regions (Hosking and Willias, 1997). Farsadnia et al., (2014) changed several sites from one region to another while deleting some of the sites which created regional heterogeneity. Therefore, in the present study regions are adjusted by changing different meteorological sites from one region to another due to climatic conditions and geographical attachment. Most of the adjustments are made in the first three groups due to maximum variability which contains elevated stations with cold climates and different rainfall patterns.

In the second stage, the at-site statistics i.e., L-CV, L-Skew, and L-Kurt are used for subsequent testing of homogeneity using discordancy and heterogeneity statistical measures. The discordancy measure is used to check for any discordant station(s) in any of the groups. The values of the measures (D_{RDI} and D_{SPI}) are calculated, given in Table (2.4) which indicates that there is no discordant station in any group and explains the critical values given in Table (2.2). Secondly, the heterogeneity measures are used to confirm the ultimate homogeneity of the adjusted regions. Hosking and Willias (1997) suggested that H_1 -statistic is the most powerful measure which has greater power of discrimination and prefers to be used for heterogeneity as compared to the remaining H_2 - and H_3 -statistics. Therefore, the values of H_1 ($H_{1(RDI)}$ and $H_{1(SPI)}$) are used to check the homogeneity of the adjusted regions, given in Table (2.4). All values of H_1 are less than 1 i.e., $H_1 < 1$. Hence it is concluded that the regions are acceptably homogenous based on RDI and SPI indices.

Table 2. 3: Discordancy and heterogeneity measures for climatic zones of Sheikh et al., (2009).

Zones	No of stations	D-sites		H_1 -values		Remarks
		RDI	SPI	RDI	SPI	
Zone-1a	Astor, Bunji, Chilas, Chitral, Dir, Darosh, Gilgit, Gupis, Skardu	0	0	-0.82	-0.28	H
Zone-1b	Balakot, Garhi Dupatta, Islamabad, Jhelum, Kakul, Kotli, Lahore, Murree, Muzaffarabad, Saidu Sharif, Sialkot	0	0	3.19	1.14	H ^T
Zone-1 (a & b)	Stations of Zone 1 (a & b)	1	0	2.24	0.45	H ^T (R)
Zone-2	Cherat, Dera Ismail Khan, Kohat, Parachinar, Peshawar, Risalpur	0	0	0.15	0.47	H
Zone-3	Bahawalnagar, Bahawalpur, Faisalabad, Khanpur, Mianwali, Multan, Raffique, Sargodha	0	0	1.61	-1.74	H ^T (R)
Zone-4	Chhor, Hyderabad, Jacobabad, Nawabshah, Padidan, Rohri	0	0	0.17	-0.26	H
Zone-5a	Barkhan, Kalat, Khuzdar, Lasbela, Quetta, Sibbi, Zhob	1	0	-0.79	1.14	H ^T
Zone-5b	Dalbandin, Nokkundi, Panjgur	0	0	0.36	-1.28	H
Zone-5 (a & b)	Stations of Zone 5 (a & b)	1	0	1.42	0.02	H ^T (R)
Zone-6	Badin, Jiواني, Karachi, Pasni	1	0	2.94	-1.15	H ^T (R)

Note: The D-site and H_1 -value represent discordant sites and the first heterogeneity measure, while the alphabets "H", "H^T" and "H^T(R)" represent the region is Homogenous, Heterogeneous w.r.t RDI and SPI, and Heterogeneous w.r.t RDI, respectively.

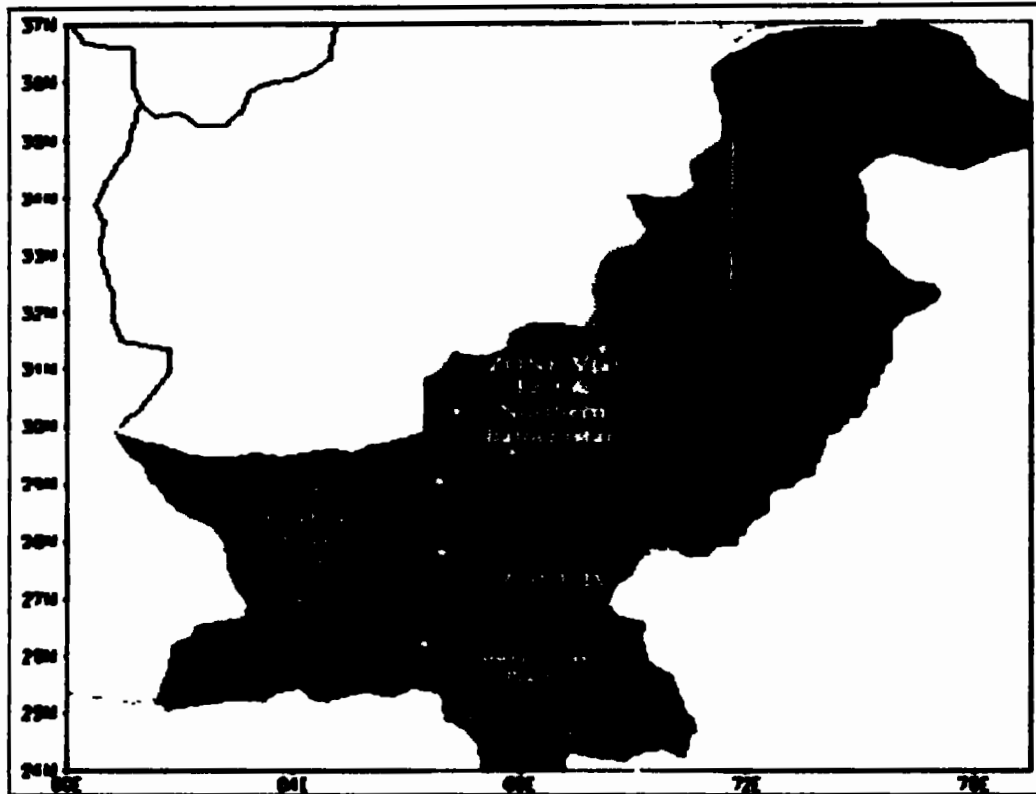


Figure 2. 2: Climatic Zones of Pakistan developed by Sheikh et al., (2009).

The HCR of RDI and SPI indices have changes in Jhelum and Lasbella stations. Both the stations are in region 5 w.r.t RDI index but in regions 2 and 4, respectively for SPI index. The Jhelum station does not change the homogeneity of regions 2 and 5 to be included or excluded but the Lasbella station has significant differences for the indices. The HCR of SPI is more suitable to consider the geographical locations of both the sites while RDI is better as PET has a significant impact on climate in Sindh and Baluchistan provinces. Lasbella station is completely discordant in region 4 w.r.t RDI while it has a discordant value of 3.06 for region 5 w.r.t SPI, hence, the station is discordant.

There are two possibilities to overcome this issue. Firstly, according to Hosking and Wallis (1997), any such stations may be removed from the study. Secondly, Saf, (2010) used robust discordancy measures (D_k^2) for the discordant sites based on L-moments ratios. It is an alternative measure that is approximately equal to the chi-square (χ_{df}^2) distribution with 3 degrees of freedom (df) due to t_1 , t_3 , and t_4 . The D_k values are equated

to a fixed cut-off value of $\sqrt{\chi^2_{(3, 0.975)}}=3.06$ instead of 3.00 significance level, which is square root of chi-square distribution at 0.975 value with 3 degrees of freedom (Neykov et al., 2007). Hence, comparing the D_k value of 3.06 for Lasbella to the cut-off value of chi-square explains the discordancy problem for the SPI index in region 5. Hence, by using robust discordancy measures the RDI regions become suitable for the SPI index. Therefore, we can either use HCRs based on SPI index due to better geographical locations or HCRs through RDI index because of robust discordancy measures. Consequently, the new HCRs for Pakistan are finalized and are given in Fig. (2.3) and Table (2.4).

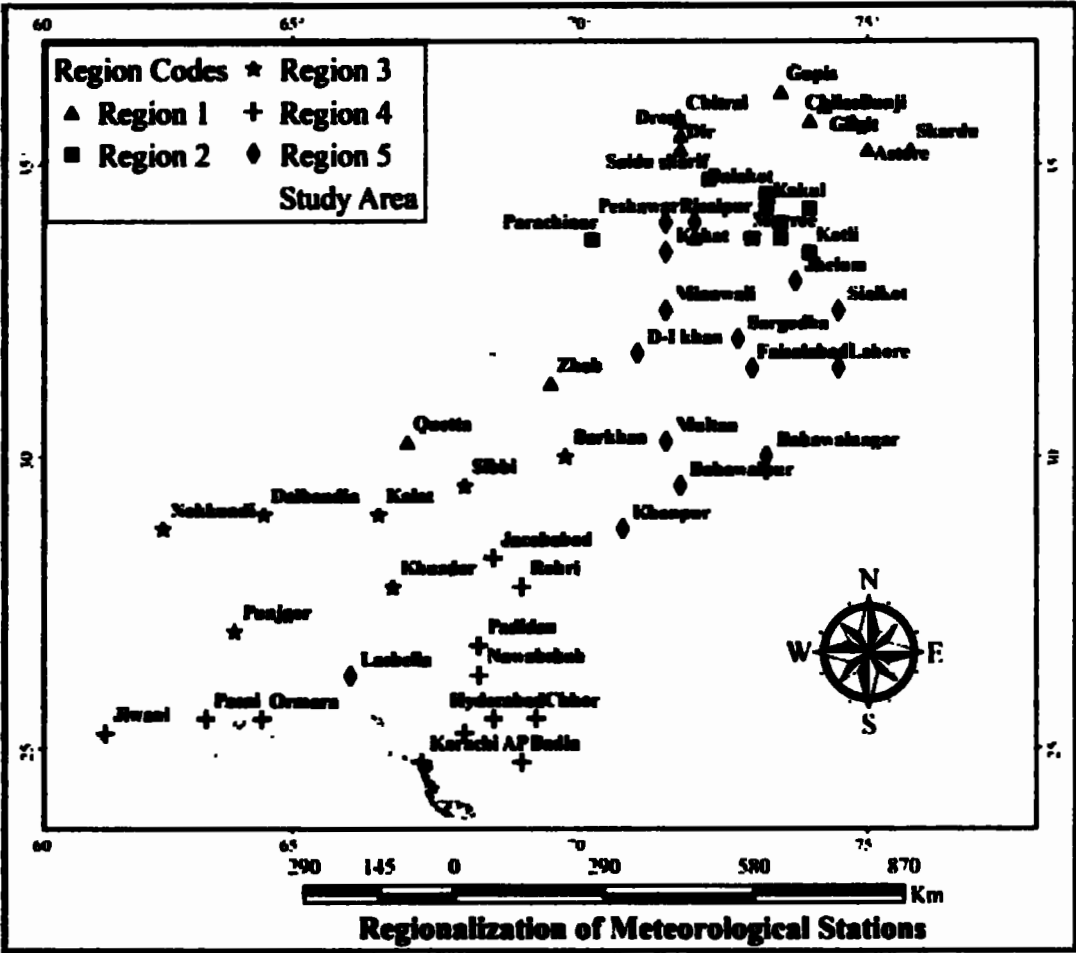


Table 2. 4: Drought events, discordancy, and heterogeneity measures of the HCR.

Region	Stations	D-Values			D-Measure			H-Measure			Regions	Stations	D-Values			D-Measure			D-Measure		
		n_{RDI}	n_{SPI}	D_{RDI}	D_{SPI}	H_1^{RDI}	H_1^{SPI}	n_{RDI}	n_{SPI}	D_{RDI}			D_{SPI}	H_1^{RDI}	H_1^{SPI}	n_{RDI}	n_{SPI}	D_{RDI}	D_{SPI}	H_1^{RDI}	H_1^{SPI}
Region-1												Kotli		12	13	1.29	0.36				
	Astore	12	13	0.55	0.62	-0.02	-0.42					Murree		12	12	0.4	0.66				
	Bunji	14	11	1.17	0.2							Muzaffarabad		9	12	1.31	1.65				
	Cherat	18	16	1.1	0.85							Parachinar		12	12	0.59	1.2				
	Chilas	11	11	0.17	0.2							Saidu Sharif		6	11	0.95	0.31				
	Chitral	9	11	2.32	0.69							Jhelum		Nil	9	Nil	1.14				
	Darosh	20	17	0.2	0.45						Region-3										
	Dir	10	9	0.39	2.19							Barkhan		8	8	0.45	0.45	0.84		-0.74	
	Gilgit	11	12	1.24	2.5							Dalbandin		14	15	1.29	1.9				
	Gupis	7	7	2.27	1.91							Kalat		16	15	0.99	1.2				
Region-2	Quetta	11	14	1.37	1.33							Khuzdar		7	6	1.21	0.24				
	Skardu	14	14	0.38	0.45							Nokkundi		10	10	0.59	0.57				
	Zhob	12	12	0.84	0.63							Panjigur		18	15	1.15	1.4				
												Sibbi		14	12	1.32	1.4				
	Balakot	6	8	1.68	0.57	0.94	-0.07			Region-4											
	Ghari Dupatta	10	10	1.07	2.32							Badin		15	15	0.02	0.16	0.31		-0.05	
	Islamabad	11	11	1.31	1.56							Chhor		14	11	2.37	2.21				
	Kakul	15	11	0.4	0.36							Hyderabad		17	16	0.85	0.21				

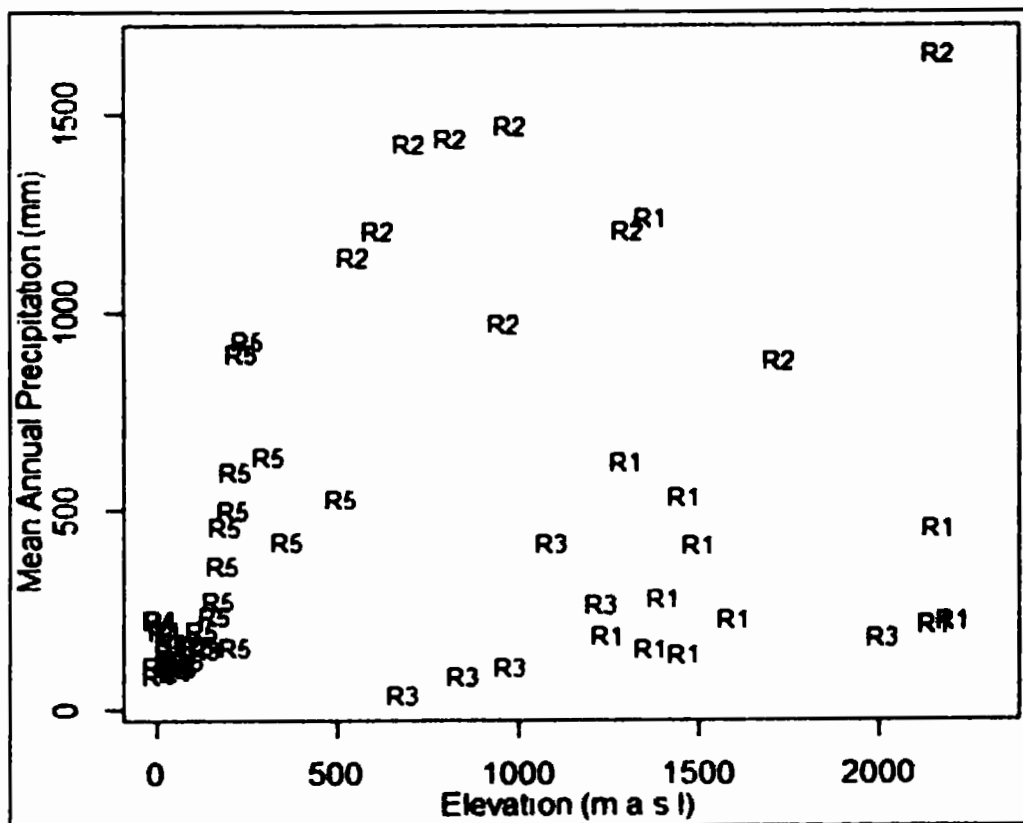
Table 2. 4: Drought events, discordancy, and heterogeneity measures of the HCR.

Region	Stations	D-Values			D-Measure			H-Measure			Regions			Stations			D-Values			D-Measure			D-Measure		
		n_{RDI}	n_{SPI}	n_{SPI}	D_{RDI}	D_{SPI}	D_{SPI}	H_1^{RDI}	H_1^{RDI}	H_1^{SPI}	n_{RDI}	n_{SPI}	n_{SPI}	D_{RDI}	D_{RDI}	D_{SPI}	n_{RDI}	n_{SPI}	n_{SPI}	D_{RDI}	D_{RDI}	D_{SPI}	H_1^{RDI}	H_1^{RDI}	H_1^{SPI}
		20	17	0.43	2.54																				
Region-5	Jacobabad	14	14	0.12	0.46						14	16	1.02	1.48			8	Nil	1.2	Nil					
	Jiwani	18	18	0.48	0.38						15	15	0.14	0.38			15	15	0.14	0.38					
	Karachi	8	8	1.85	1.91						13	16	1.78	1.04			21	21	0.66	1.61					
	Mohin-Jodaro	10	9	1.94	0.47						5	Nil	1.61	Nil			10	11	2.27	1.15					
	Nawabshah	9	10	0.89	1.01						15	15	0.13	0.27			13	13	0.3	0.19					
	Ormara	15	14	0.56	1.33						13	15	1.34	1.84			14	14	0.74	0.76					
	Padidan	20	20	1.49	0.72						18	16	0.56	0.13			18	16	0.56	0.13					
	Passni	15	15	1.00	0.59																				
	Rohri	Nil	5	Nil	1.61																				
	Lasbella																								
	Bahawalnagar	10	9	1.75	1.28	0.79	0.74																		
	Bahawalpur	15	15	0.34	0.58																				
	D.I Khan	16	15	1.75	2.3																				

Note: The n_{RDI} , n_{SPI} , D_{RDI} , D_{SPI} , H_1^{RDI} , and H_1^{SPI} represent the number of drought events, discordancy values of each site, and first heterogeneity value of the region using RDI and SPI indices, respectively.

2.3.3 Statistical Validity of the Identified HCRs

It is important to statistically validate the newly constructed HCRs for the climatic condition using site characteristics and climate data. In Fig. (2.4), the scatter plot of Mean Annual Precipitation (MAP) and elevation of the sites are performed. The plot shows good regional clusters of MAPs with elevation. The MAP increases with the increase in elevation. Topçu and Seçkin (2016) constructed HCRs entirely based on the elevation of the stations. Hence, elevation plays a vital role in the development of HCRs.



elevation are presented in **Table (2.5)**. A brief discussion of the regional results is as follows:

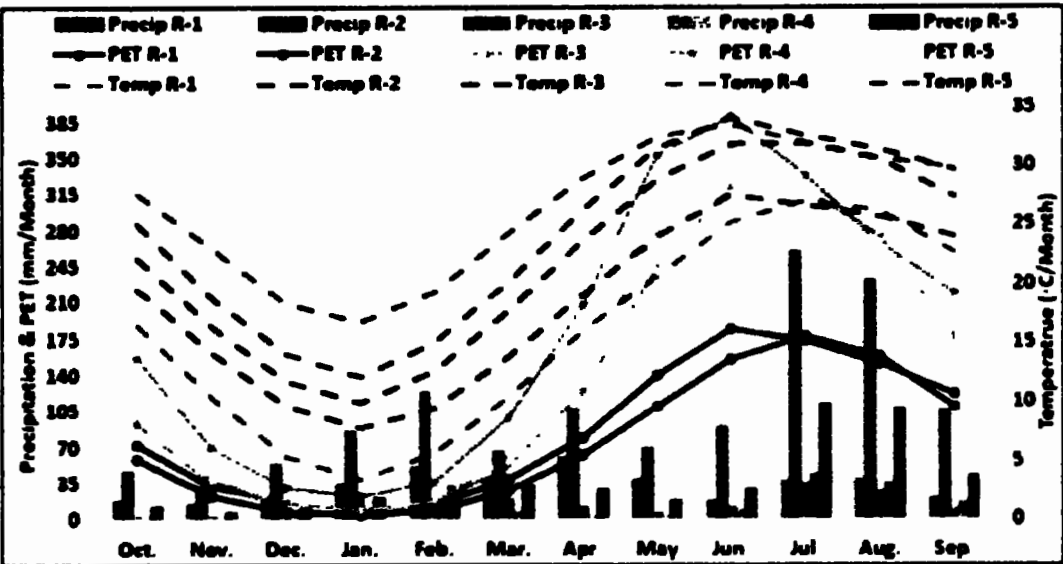
Region 1 has maximum elevated mountain ranges of HKH, joining in a very complex system with 3rd largest glacier in the world. The graph shows maximum regional rainfall, PET, and temperature in the monsoon season from June to September while minimum temperature and PET from December to January. Region 2 lies in the eastern part of the country with high mountains and maximum rainfall, especially during monsoon. It has cold areas and PET with the least value from December to January. Region 3 has a mixed climate with some high mountains in Balochistan. The region is severely drought-prone with some very dry and hyper-arid areas such as Nokkandi, with a minimum annual regional rainfall and maximum temperature and PET. The region has high regional variation in its temperatures. Region 4 contains the hottest and driest part of the country which is highly drought prone. The region has the maximum average minimum temperature and the highest amount of PET compared to other regions. The maximum temperature and PET with least rainfall may be one of the reasons for aridity. The region contains the most arid as well as least elevated areas with plain deserts and coastal lines along the Arabian Sea. Region 5 includes mostly the parts of Punjab and KPK provinces with minimum elevation. The region is highly agricultural land based on both irrigation and rainfall. It has moderate rainfall with heavy rainfall in monsoon season and highest average temperature and PET.

Temperature and PET have significant changes in the last three regions (related to Balochistan, Punjab, and Sindh provinces) and make "S" shape for all the regions. The regions have high variation in their annual average minimum and maximum temperatures. This discussion of regional results shows greater variation among the

variables from humid to hyper-arid stations. The graphical presentation and regional results fully satisfy the construction of homogenous regions.

Table 2. 5: Regional mean and standard deviation of precipitation and temperature.

Region	Average Elevation	Annual Precipitation		Annual Average Temperature	
		Mean	SD	Minimum	Maximum
Region-1	1635.83	390.33	130.64	-2.20	34.95
Region-2	955.38	1205.50	147.10	1.47	35.56
Region-3	998.00	160.08	179.19	1.67	36.83
Region-4	33.33	153.00	170.23	13.63	38.30
Region-5	261.60	447.42	162.62	4.68	41.41



Note: R-1 for 1,2,3,4,5 represent the five HCR for Pakistan, respectively..

Figure 2. 5: Region-wise comparison of Precipitation, PET, and Temperature.

2.4 Conclusion

Homogenous regions of drought are used for planning strategies to cope with the worse condition of water resources in Pakistan. In this chapter, the main objective is to locate HCRs based on RDI and SPI indices using rainfall and temperature data in Pakistan. The existed climatic zones are tested for homogeneity which is not satisfied. Hence, new HCRs are constructed using valid statistical techniques and measures. Cluster

analysis is performed based on site characteristics for subjective HCRs. The L-moments method is applied to calculate discordancy and heterogeneity measures for the ultimate satisfaction of the HCRs. These measures confirmed the five HCRs based on RDI as well as SPI indices in Pakistan. The regions are assessed using monthly and annual results based on observed climatic data which fully confirmed the climatic condition in Pakistan. These results indicate that there is high variation due to the climate in Pakistan from humid to hyper-arid regions. Regions one and two have enough water resources while regions three and four are highly drought-prone. Region five has reasonable rainfall in the monsoon session.

Chapter 3

Regional Droughts' Projections

3.1 Background

Statistical techniques play a key role in future droughts' projections for the assessment of droughts risks. Various probability distributions are employed for droughts' projections using frequency analysis. These forecasts are helpful for the planning of climate-related disasters, management of water supply and storage, and drainage structures (Saf, 2010). Droughts affect large areas with varying intensities and therefore, regional classification w.r.t. varying characteristics of droughts is important for prediction and devising combating strategies for these types of risks. (Quesada-Montano et al., 2018). One such strategy is the Regional Frequency Analysis (RFA) technique which combines different sites of such characteristics for approximate homogeneity and projections over a large area at numerous return periods. Several studies in literature have emphasized the significance of frequency analysis in forecasting various climatic extremes and preparing for better management (Santos et al., 2011; Topçu and Seçkin, 2016; Ghosh, and Srinivasan, 2016; Fawad et al., 2018; Khan et al., 2019). In the previous chapter, we have identified five homogeneous climatic regions of Pakistan using RDI and SPI indices at 12-months' time scales (Ullah et al., 2019; Ullah et al., 2020). Considering the five homogeneous regions, future climate projections based on drought risk assessment are done for each region. Drought mitigation and water resource planning are required to protect against disasters from natural hazards (Quesada-Montano et al., 2018). It is critical to forecast reliable future fluctuations for planning these water-related hazardous situations. Therefore, the newly

constructed five HCRs may be considered to perform comprehensive research work on future drought projections in Pakistan.

This chapter has two main objectives. 1) RFA method is used to find regional drought projections at selected return periods using RDI and SPI. 2) Secondly, to find ungauged sites drought estimates for areas with no metrological sites using Quadratic Regression (QR) technique.

3.2 Methodology

The IFP procedure is described in subsection 2.2.4, which is used to construct HCR and to project droughts at selected return periods. The procedure has five steps, where the first two steps are related to the construction of HCRs and are discussed in the above-mentioned sub-section. While the last three steps of the IFP procedure are used to find the estimates of drought projections and are discussed in the following sub-sections. Additionally, this section contains the methods to find the at-sites projections and estimates of ungauged spots.

3.2.1 Selection of Best-Fit Probability Distribution

The selection of probability distribution is necessary for reliable drought projections, particularly at higher return periods using RFA method. According to Hosking and Wallis (1997) two-parameter probability distributions are not suitable for RFA. Firstly, it may provide biased estimates of quantiles at the tails. Secondly, it does not capture the shape of tail dependence bitterly. Five 3-parameter distributions are more suitable for measuring regional projections. The distributions are Generalized Extreme Value (GEV), Generalized Normal (GNO), Generalized Logistic (GLO), Generalized Pareto (GPA), and Pearson Type-3 (PE3), given in Appendix-B. If none of these distributions are acceptable, then four and five parameters' Kappa (KAP) and Wakeby (WAK)

probability distributions are used. These are mostly required if the HCRs are not properly homogeneous (Hosking and Wallis, 1997; Yin et al., 2015).

There are two ways to select the best-fit probability distribution. Firstly, the graphical approach of the L-moment ratio diagram (L-MRD) is used to compare the closeness of L-skewness and L-kurtosis to regional points using five probability distributions. The method is based on the L-moments procedure. The theoretical curves of the selected distributions are also plotted on the same graph to compare the curves of the distributions with the regional point. A distribution is considered the best fit if its theoretical curve is close to the regional point of L-skewness and L-kurtosis. However, graphical methods provide a rough idea and cannot be used as a single selection method for best-fit distributions (Hosking and Wallis, 1997).

Secondly, the numerical method of the goodness of fit (GOF) Z-test is used which is considered more reasonable and practical. The test is used to find the best fit parent probability distribution for the region based on the Z-test using L-moments. The test assumes that in a homogenous region the data from individual sites within the region have an identical probability distribution, given as follows:

$$Z^{dist} = \frac{(\tau_4^{dist} - \bar{\tau}_4 + \beta_4)}{\sigma_4} \quad (3.1)$$

where 'dist' denote the candidate probability distribution, τ_4^{dist} is used for L-kurtosis using candidate distribution while $\bar{\tau}_4$ is the regional average of L-kurtosis. β_4 Calculates bias between regional average of L-kurtosis and its k^{th} simulated value using equation.

$$\beta_4 = \frac{1}{N_{sim}} \sum_{k=1}^{N_{sim}} (\tau_4^{(k)} - \bar{\tau}) \quad (3.2)$$

Equation (3.2) finds the average difference between simulated and fitted values of L-kurtosis. Whereas σ_4 is given by

$$\sigma_4 = \frac{1}{N_{sim} - 1} \left(\sum_{k=1}^{N_{sim}} (\tau_4^{(k)} - \bar{\tau}_4)^2 - N_{sim} \beta_4^2 \right)^{1/2} \quad (3.3)$$

The hypothesis of candidate probability distributions would be accepted as best fit distribution if $|Z^{dist}| < 1.645$ where 1.645 is a two-sided z-table value at a 10% value of α . If more than one distribution has less than a 1.645 value, then the one with least $|Z^{dist}|$ value is selected as best fit.

3.2.2 Estimation and Inverse Function of Distribution

The L-moments approach discussed in sub-section 2.2.3 is used to estimate the selected best-fit distributions based on extracted drought events. L-moments approach is appropriate due to positive skewness and outliers in drought events. Fitted probability distributions are used to find drought projections. The inverse function is found from regional best fit distribution and is used to calculate dimensionless quantiles which are suitable to all the sites in the region. The inverse function is called regional growth curve ($\hat{q}(F)$) (Stedinger et al., 1993; Hosking and Wallis, 1997). $\hat{q}(F)$ is applied to find regional drought quantiles at selected return periods, known as regional drought projections. The at-site quantiles are found by multiplying mean of drought values of a site with the regional drought quantile values as below:

$$\hat{Q}_i(F) = l_1^{(i)} \hat{q}(F) \quad (3.4)$$

where $\hat{Q}_i(F)$ is the at-site quantile function at non-exceedance probability ($0 < F < 1$) for the site i which is the complement of exceedance probability, $l_1^{(i)}$ is the sample mean of drought events of the i^{th} site called scaling factor, and $\hat{q}(F)$ is the RGC, which represents the quantile value for a specific return period (T) of the normalized regional distribution (Ngongondo et al., 2011). The non-exceedance probability (F) can be defined in the form of return periods (T) as follow:

$$F = 1 - 1/T \quad (3.5)$$

3.2.3 Validation of Quantile Estimates

Estimated quantiles of drought always have uncertainty. To assess the accuracy of regional drought quantiles, Monte Carlo simulations (MCS) method is utilized to produce several similar regions with similar information as observed data (Hosking and Wallis, 1997). For each simulation at the given return periods, the root means square error (RMSE) and a 90 % confidence range of the quantiles are calculated. In each simulation, the quantiles are calculated for many return periods with k^{th} repetition of simulated drought quantile for the i^{th} site at F value presented by $\hat{Q}_i^{(k)}(F)$. Where relative error for the drought quantiles is computed as follows:

$$RE = \frac{(\hat{Q}_i^{(k)}(F) - Q_i(F))}{Q_i(F)} \quad (3.7)$$

While the average relative error (RMSE), for the N repetitions is as below:

$$R_i(F) = \left(\frac{1}{N} \sum_{k=1}^N \left(\frac{(\hat{Q}_i^{(k)}(F) - Q_i(F))}{Q_i(F)} \right)^2 \right)^{1/2} \quad (3.8)$$

Also, the regional average relative error (RMSE) of the estimated quantiles is as below:

$$R^R(F) = \frac{1}{N} \sum_{i=1}^N R_i(F) \quad (3.9)$$

The $(1 - \alpha)100\%$ confidence interval for quantile values is defined as:

$$\frac{\hat{q}(F)}{U_{\frac{\alpha}{2}}(F)} \leq q(F) \leq \frac{\hat{q}(F)}{L_{\frac{\alpha}{2}}(F)} \quad (3.10)$$

Where $L_{\frac{\alpha}{2}}(F)$ and $U_{\frac{\alpha}{2}}(F)$ denote lower and upper intervals using significance level (α).

3.2.4 Quantiles at ungauged sites

The meteorological stations have high variability in terms of distance and homogeneity in Pakistan, mostly in Balochistan province. It is required to evaluate drought conditions at these ungauged locations. The RFA method has the ability of aggregating summary results from several sites in the region. Using equation (3.4), the mean value (l_1) is necessary for a drought estimate at gauged or an ungauged site in a region. The regression method is used to find l_1 for flood estimates at ungauged sites using climatic and physiographic variables i.e., annual average rainfall magnitude, drainage area of the catchment (Kumar et al., 2003; Zaman et al., 2012; Hailegeorgis and Alfredsen, 2017; Khan et al., 2019). Khan et al., (2019) applied QR technique to find l_1 for annual maximum peak flow quantiles at ungauged sites, between the site's means of flood magnitudes and average rainfall in the monsoon season in Pakistan. However, there is no data on the catchment area for rainfall and temperature stations and have non-identical distances from each other in Pakistan shown in Fig. (2.3). On the other hand, there is no rainfall record at ungauged sites to be used as an explanatory variable.

The site coordinates i.e., latitude and longitude, uniquely presents a point on the globe whereas elevation has a vital role in the construction of HCRs (Shahzadi et al., 2013; Topçu and Seçkin, 2016; Ullah et al., 2020). Therefore, the QR model is applied using latitude, longitude, and elevation as explanatory variables and index-drought (l_1) as a response variable. The QR model is used to explain maximum variability and measure non-linear changes in index-drought values, given as follows:

$$\hat{l}_1^{(0)} = \hat{\alpha} + \sum_{i=1}^k \hat{\beta}_i(\text{site characteristics}) + \sum_{i=1}^k \hat{\beta}_{ii}(\text{site characteristics})^2 \quad (3.11)$$

where $\hat{\alpha}$ is the intercept, $\hat{\beta}_i$ and $\hat{\beta}_{ii}$ are regression coefficients while k represents the number of explanatory variables.

3.3 Results and Discussion

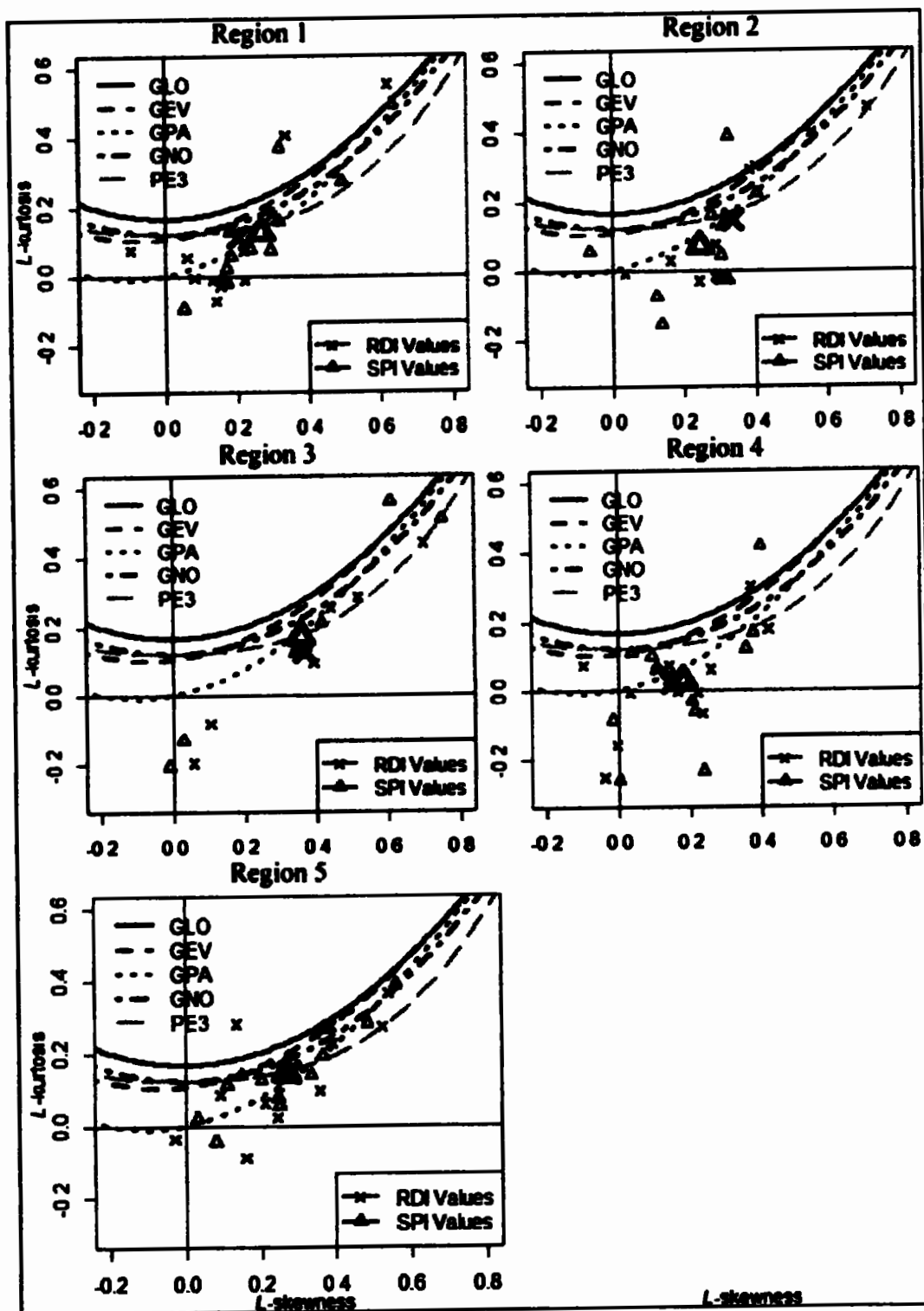
3.3.1 Regional Frequency Analysis

In this chapter, five HCR developed in chapter 2 are considered for drought projections through RDI and SPI extreme events for evaluation, which might be helpful in droughts' risk assessment, and water resources arrangements. The conventional frequency analysis performed in hydrology assumes the necessary properties of statistical data that extreme events belong to a stationary distribution and are independent of one another. In the case of non-stationarity due to climate change and time-dependent processes, the return period might not characterize a comprehensive measure of the probability of failure and its application could lead to false results (Cooley, 2013; Volpi et al., 2015). Checking stationarity and independence between drought events is beyond the scope of this study because mostly drought events take a year or several years in their occurrence. Therefore, we assume that drought events are independent and suitable to apply conventional frequency analysis, where the return period gives better results for RFA.

The selection of the best-fit probability distribution(s) has an important role in the reliable statistical estimates for regional projections, particularly at higher return periods (Saf, 2010). Firstly, L-MRD method for RDI and SPI events is used to match the location and closeness of regional averages of L-skewness versus L-kurtosis and theoretical affiliation of the five distributions, plotted in Fig. (3.1). The lines of distributions in all graphs show that the lines of GPA and PE3 distributions pass close to the regional average points of both RDI and SPI. Secondly, the GOF z-test is

calculated for the five distributions where various distributions are suitable compared to critical values using SPI and RDI in the regions. But the distribution with least z-test value is considered best fit for a region, as presented in Table (3.1). Hence according to z-test values, GPA and PE3 are considered the best-fit regional distributions using SPI and RDI. The L-MRD and GOF test give identical results for the regions and drought indices. The selected regional probability distributions of GPA and PE3 are estimated using extreme drought events of RDI and SPI series through the L-moments technique. The estimated parameters of the best fit regional distributions are given in Table (3.2).

The inverse functions i.e., $(\hat{q}(F))$ are found using regional distributions for drought quantiles at selected return periods, which are given in Table (3.3). Drought quantiles are dimensionless values that represent average drought risk at non-exceedance probability (F) in the future using best-fit distributions for the regions. Drought quantiles are calculated cumulatively, indicating that drought severity grows as the number of return periods increases. Return period is the average time between two drought episodes, where smaller return periods have smaller quantiles while longer return periods have larger quantiles (Cunnane, 1988). The T-year return period implies the risk for occurring an extreme event increases with a ratio of $1/T$ per year (Stedinger et al., 1993; Volpi et al., 2015). Hence, drought risk cumulatively increases with a ratio of $1/T$ each year considering the existing climate data of the regions.



Note: The small symbols of the scatterplot show sites L -skewness and L -kurtosis values while bold symbols denote the regional points, respectively.

Figure 3. 1: L -moment ratio diagram for selection of regional probability distributions.

Table 3. 1: The GOF Z-test results for the probability distributions of the regions.

Region	Index	Probability distributions				
		GLO	GEV	GNO	PE3	GPA
Region-1	RDI	2.97	1.95	1.7	1.2*	-0.43**
	SPI	2.7	1.95	1.54*	0.81*	0.02**
Region-2	RDI	1.98	1.61*	1.2*	0.49**	0.51*
	SPI	2.83	2.19	1.87	1.3*	0.59**
Region-3	RDI	3.29	2.84	2.4	1.64*	1.55**
	SPI	1.69	1.14	0.88	0.40*	0.25**
Region-4	RDI	5.72	4.4	4.22	3.76	1.46**
	SPI	3.92	2.78	2.54	2.03	0.15**
Region-5	RDI	2.43	1.56*	1.12*	0.31**	-0.66*
	SPI	3.27	2.35	1.86	0.98*	-0.01**

Note The symbol () show acceptable while (**) the selected best-fit distributions.*

Table 3. 2: Estimated parameters for the best fitted regional probability distributions of the regions

Region	Index	Dist.	Estimated Parameters		
			Location	Scale	Shape
Region-1	RDI	GPA	0.5925	0.5439	0.3346
	SPI	GPA	0.6001	0.4670	0.1679
Region-2	RDI	PE3	1.0000	0.4419	2.0242
	SPI	GPA	0.6053	0.4684	0.1866
Region-3	RDI	GPA	0.5410	0.4694	0.0227
	SPI	GPA	0.5890	0.4920	0.1990
Region-4	RDI	GPA	0.5567	0.6032	0.3605
	SPI	GPA	0.5640	0.5910	0.3570
Region-5	RDI	PE3	1.0000	0.3370	1.5453
	SPI	GPA	0.6070	0.4612	0.1736

The drought quantiles are interpreted in the shape of RDI, and SPI indices categories of severity levels, given in Table (2.1). The regional drought quantiles are the accumulated values of average future drought risk which show the chances of drought in the regions according to the previously analyzed condition. For example, the first value of the RDI index for region 1 in Table (3.3) means that at a non-exceedance probability of 0.500 (or 2 years of return period), the average drought risk will be less than or equal to the absolute magnitude of 0.929 in any given year if the GPA distribution of drought events does not change. In a long sequence of years, the expected proportion of years in which a drought event smaller than or equal to 0.929 occurs is 0.500 using GPA as the regional best-fit probability distribution. Similarly, the remaining drought quantile magnitudes are interpreted according to the return periods. In the first three regions, the performance of both the drought indices becomes reversed as a drought index with the maximum quantiles becomes less with the increase in return periods and vice versa. Similarly, the last two regions have a similar pattern in drought quantile values with an increase in return periods.

The estimated values always have some degree of uncertainty. The MCS process is used to assess the accuracy of quantiles in the form of RMSE and 90% confidence interval (Error Bound) at different return periods. For this purpose, 1000 MCS are generated through the regional best-fit probability distributions. The RMSE values increase and the 90% error bounds become wider and wider with the increase in return periods, given in Table (3.3). The results of 90% error bounds along with quantile values are graphed in Fig. (3.2) for both the drought indices. In the graphical presentation of the quantiles and error bounds, the dashed lines of the error bounds become wider and wider w.r.t bold lines of quantile estimates. It means that variability and uncertainty increase with the increase in return periods. According to the error

bounds, the uncertainty and variability start in the results of quantiles approximately after 10 years of return periods and increase with the increase in the duration of return periods for all the regions using both the drought indices.

To further assess the accuracy of regional projections the RMSE values are graphed in Fig. (3.3). The RMSE values have significant variability and uncertainty with increasing return periods among the regions. The quantiles are considered better with minimum values of RMSE. Firstly, region 3 has maximum variability and uncertainty in quantile values due to vast areas with few gauging stations and variability in rainfall and elevation records. If there is a large area with few gauging stations then the regional quantiles will have high variability and uncertainty, particularly at higher return periods (Hosking and Wallis, 1997). Secondly, regions 1 and 2 have high variability in the quantiles w.r.t RDI and SPI indices due to high latitudinal areas (Dai et al., 1997). Thirdly, region 5 has more variability in the SPI index. While regions 1 and 5 w.r.t RDI and both indices for region 4 have no significant changes with minimum values of RMSE. Consequently, RDI looks better for regions 1 & 5 and SPI for regions 2 & 3 while both have similar results for region 4.

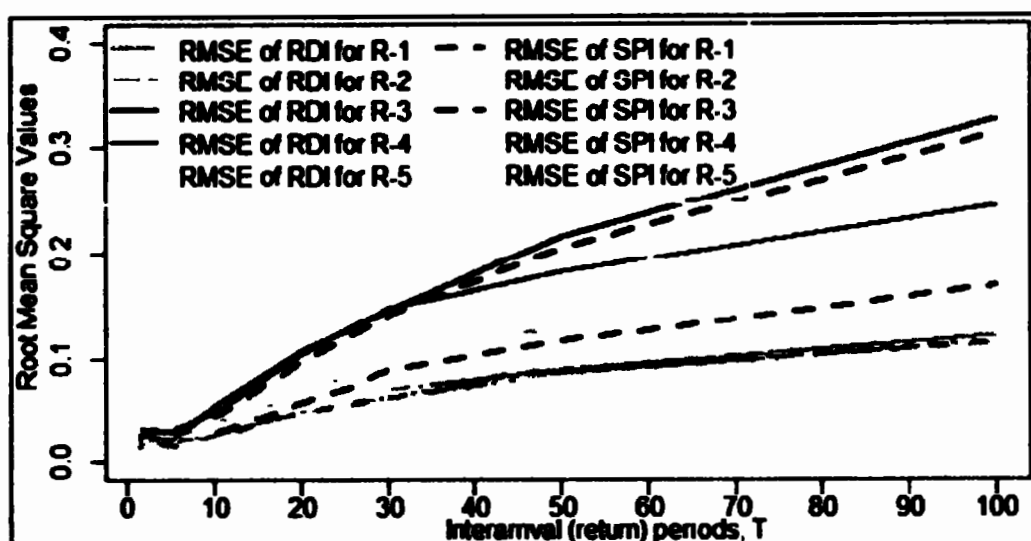


Figure 3. 2: Graphical Projection of RMSE for the regions at RDI and SPI indices.

Table 3. 3: Regional projections, RMSE, and error bounds at various return periods.

Regions	T	$\hat{q}_R(F)$	RMSE	95% EB (RDI)		$\hat{q}_S(F)$	RMSE	95% EB (SPI)	
				$L_{0.025}$	$U_{0.975}$			$L_{0.025}$	$U_{0.975}$
Region-1									
	2	0.929	0.014	0.904	0.951	0.906	0.016	0.878	0.926
	5	1.269	0.016	1.245	1.299	1.259	0.018	1.230	1.289
	10	1.466	0.025	1.431	1.510	1.492	0.029	1.451	1.545
	20	1.621	0.049	1.549	1.712	1.700	0.059	1.625	1.814
	25	1.664	0.058	1.577	1.771	1.761	0.071	1.673	1.902
	50	1.779	0.090	1.639	1.938	1.939	0.115	1.795	2.167
	100	1.870	0.122	1.681	2.082	2.098	0.164	1.885	2.412
Region-2									
	2	0.863	0.032	0.807	0.902	0.911	0.018	0.881	0.935
	5	1.267	0.020	1.237	1.304	1.265	0.020	1.237	1.301
	10	1.575	0.056	1.509	1.687	1.491	0.034	1.448	1.557
	20	1.883	0.107	1.759	2.103	1.686	0.066	1.602	1.825
	25	1.982	0.125	1.839	2.241	1.743	0.079	1.642	1.908
	50	2.290	0.183	2.083	2.685	1.903	0.124	1.742	2.159
	100	2.599	0.245	2.328	3.138	2.042	0.175	1.806	2.380
Region-3									
	2	0.864	0.027	0.814	0.899	0.872	0.027	0.825	0.908
	5	1.283	0.029	1.232	1.324	1.291	0.028	1.244	1.342
	10	1.594	0.048	1.528	1.686	1.591	0.046	1.531	1.683
	20	1.901	0.103	1.784	2.129	1.877	0.099	1.750	2.086
	25	1.998	0.127	1.854	2.286	1.966	0.122	1.811	2.220
	50	2.298	0.216	2.041	2.794	2.235	0.205	1.974	2.670
	100	2.594	0.326	2.204	3.368	2.491	0.306	2.102	3.146
Region-4									
	2	0.927	0.014	0.902	0.948	0.926	0.014	0.901	0.946
	5	1.293	0.015	1.271	1.320	1.292	0.015	1.268	1.317
	10	1.500	0.024	1.471	1.547	1.500	0.026	1.464	1.547
	20	1.662	0.048	1.597	1.751	1.663	0.049	1.593	1.760
	25	1.705	0.056	1.628	1.811	1.707	0.058	1.623	1.820
	50	1.821	0.085	1.700	1.974	1.825	0.086	1.703	1.992
	100	1.912	0.114	1.734	2.117	1.917	0.115	1.751	2.135
Region-5									
	2	0.917	0.016	0.889	0.937	0.915	0.017	0.885	0.939
	5	1.230	0.010	1.214	1.248	1.228	0.012	1.209	1.250
	10	1.449	0.028	1.414	1.504	1.449	0.032	1.403	1.512

Table 3. 3: Regional projections, RMSE, and error bounds at various return periods.

Regions	T	$\hat{q}_R(F)$	RMSE	95% EB (RDI)		$\hat{q}_S(F)$	RMSE	95% EB (SPI)	
				$L_{0.975}$	$U_{0.975}$			$L_{0.975}$	$U_{0.975}$
	20	1.659	0.051	1.600	1.763	1.662	0.058	1.586	1.777
	25	1.726	0.059	1.656	1.847	1.730	0.067	1.642	1.864
	50	1.930	0.084	1.835	2.108	1.937	0.096	1.811	2.132
	100	2.131	0.111	2.007	2.363	2.142	0.126	1.974	2.407

Note: The abbreviations T represent return periods. $\hat{q}_R(F)$ and $\hat{q}_S(F)$ represent the estimated quantiles for RDI and SPI indices at non-exceedance probability F while EB (RDI) and EB (SPI) are used for Error Bounds of RDI and SPI, respectively.

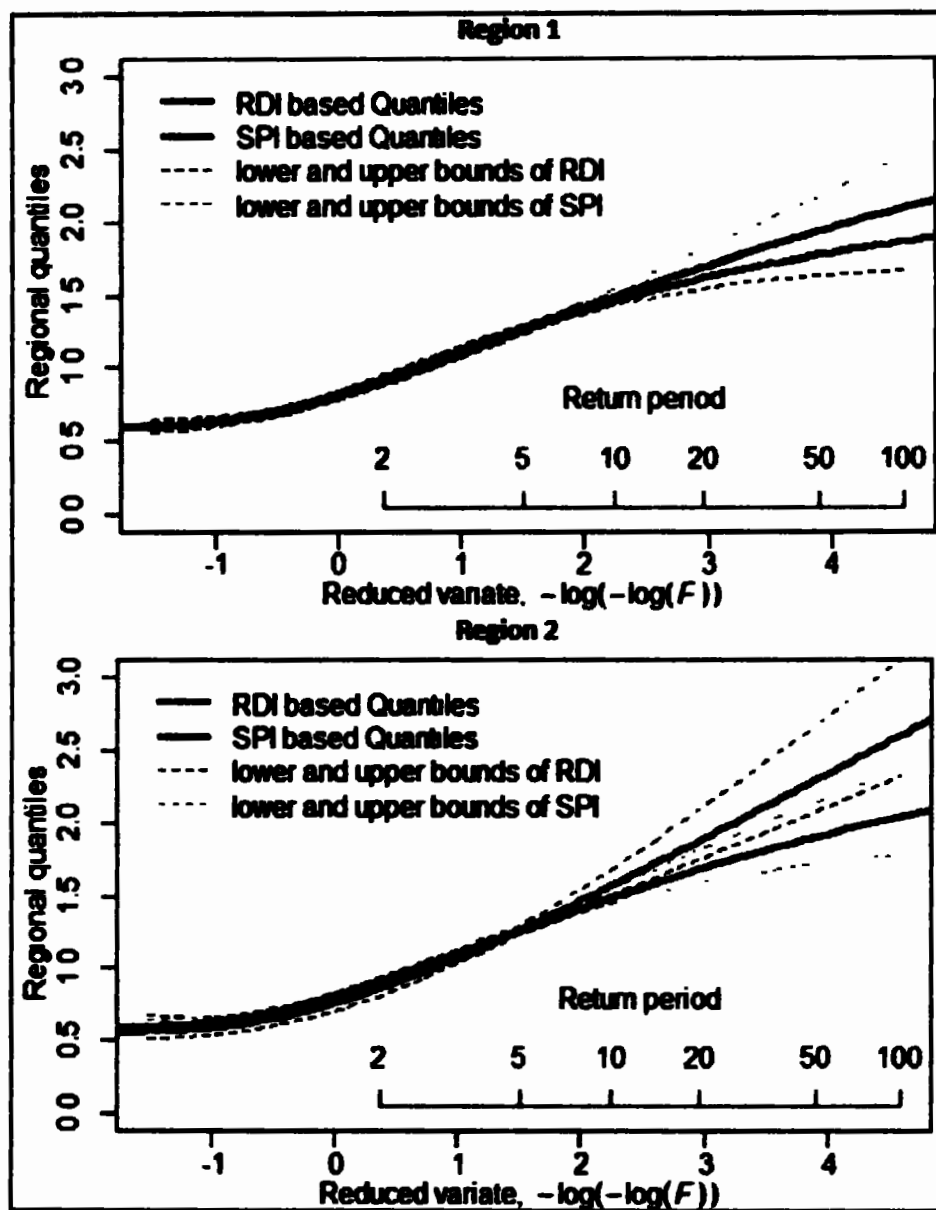


Figure 3. 3: Regional quantiles with 90% error bounds for the regions.

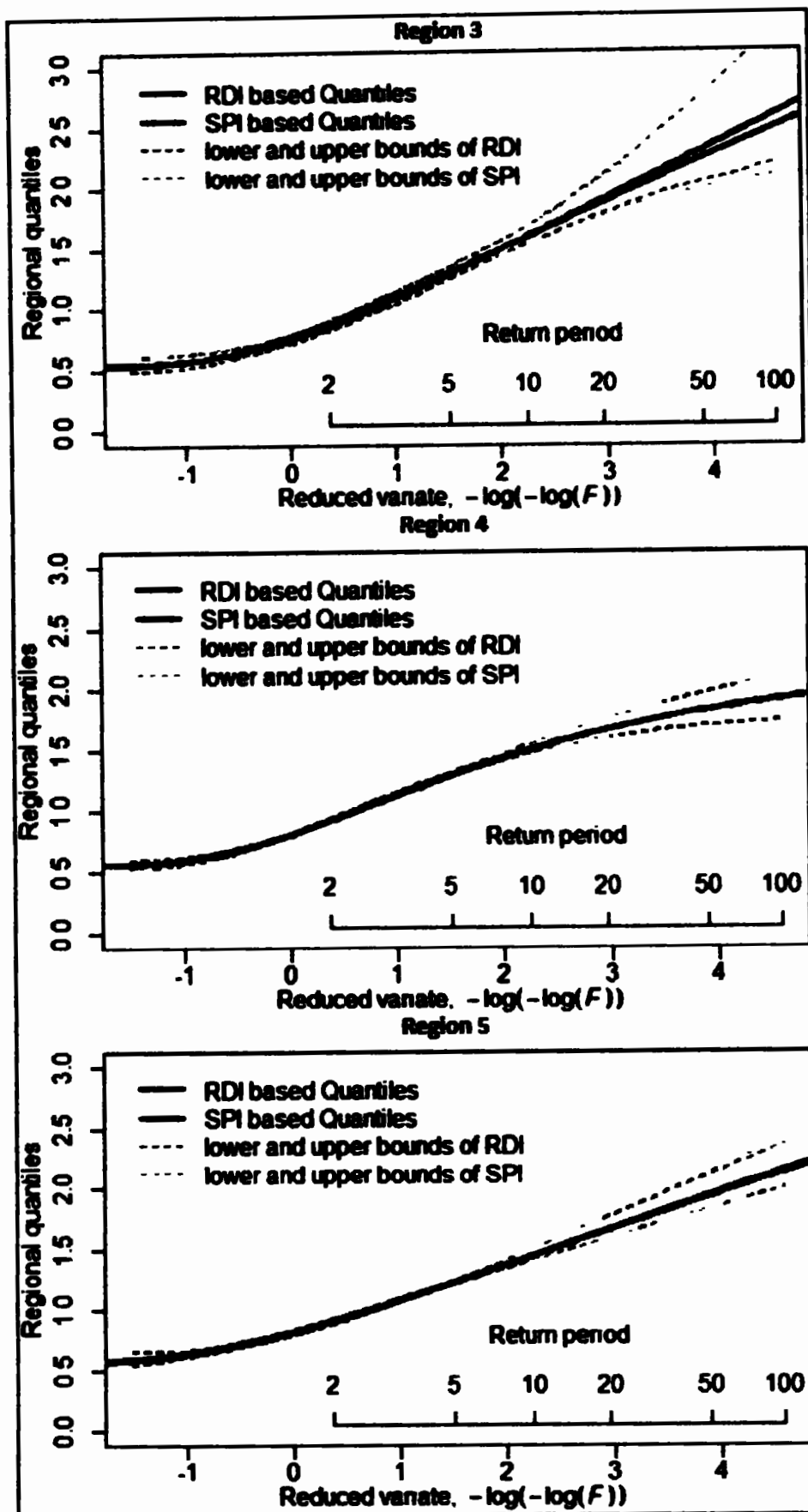


Figure 3. 3: Regional quantiles with 90% error bounds for the regions.

3.3.2 At-Sites and Ungauged Sites Projections

In this era of climate change, many studies in hydrology (drought, flood, wind speed, rainfall) have used the method of return period for short- and long-term projections in the form of quantiles using RFA as well as At-Site Frequency Analysis (ASFA) through IFP (Brito et al., 2018; Das, 2018; Fawad et al., 2018; Fawad et al., 2019; Khan et al., 2019; Ullah and Akbar, 2020). In RFA the results cover a large area while in ASFA the results are obtained at the grass-root level for each site of the study area. At-site quantiles for drought are computed at various years of return periods using equations (3.4) for RDI and SPI indices based on regional best-fit probability distributions. There is significant variation among the site's quantiles of the regions at a similar return period, given in Table (3.4). The scaling factor (l_1) has a vital role which creating a difference from site to site for both drought indices. If a station has many drought values with fluctuations, then l_1 is large and consequently, the at-site quantiles will be large.

Finally, QR model is used to assess future drought estimates at ungauged sites using equation 3.11. Drought scaling factor l_1 is calculated for RDI and SPI. The l_1 and at-site quantiles from IFP at various return periods are used one by one as dependent variables while the site characteristics (latitude, longitude, and elevation) of each site as explanatory variables to fit QR model. The estimates using the fitted QR model are obtained for all the selected sites of the study to check and compare the results with at-site quantiles using the IFP method, presented in Table (3.4). QR estimates give a good approximation with at-site quantiles of the IFP procedure at all selected return periods for the sites. The differences between quantiles are high in sites of region 3 due to rainfall variability and a vast area with fewer stations. Fitted QR models are as follows:

$$\hat{\eta}_1^{(RDI)} = 8.319 - 0.1471Lo_i - 0.1149La_i - 0.00001371E_i + 0.00104Lo_i^2 + 0.00191La_i^2 \quad (3.12)$$

$$\hat{\eta}_1^{(SPI)} = 6.061 - 0.1251Lo_i - 0.01304La_i + 0.00002112E_i + 0.0009219Lo_i^2 - 0.00002764La_i^2 \quad (3.13)$$

In the equations, the site characteristics share differently in their directions in linear and quadratic forms. The use of elevation is important particularly in mountainous sites as the climate has significant change with the elevation change. The normality graphs of RDI and SPI are presented in Fig. (3.4), which shows a good approximation of the relationship among the theoretical quantiles and standardized residuals except for a few stations. The graph of SPI is statistically more suitable as compared to the RDI index. The estimated QR model is used for drought projection at some ungauged sites in the country. The site characteristics are used in QR model and found quantiles at selected return periods for RDI and SPI indices, given in Table (3.5) for some of the selected stations in the country. These estimates give good matching with the areal changes according to estimates of the IFP method. The results confirm that the equation can be used for reliable estimates at any ungauged site within the country or worldwide for drought projections as these site characteristics are easily available.

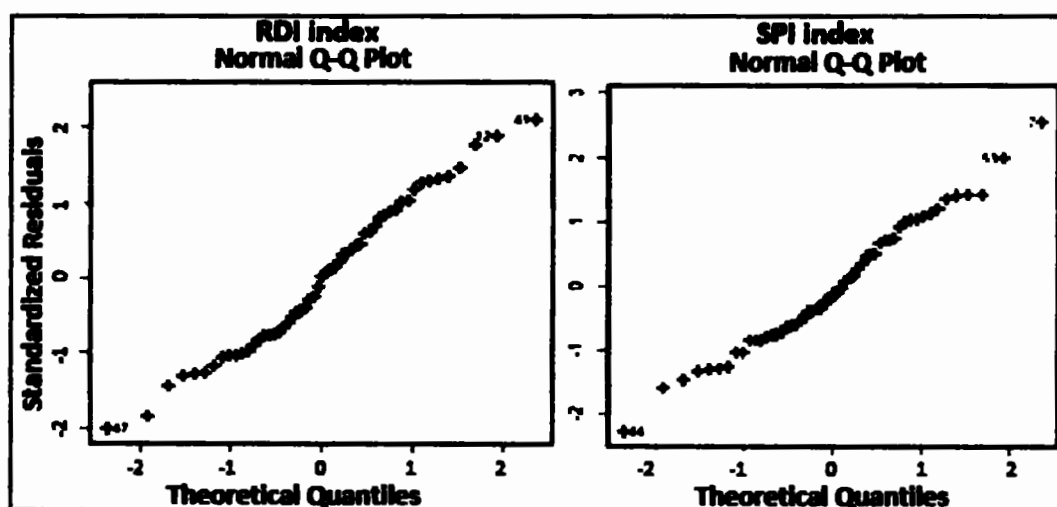


Figure 3. 4: Normal Q-Q plots for the Quadratic Regression equation.

Table 3. 4: At-Site Quantiles using IFP and QR Methods for RDI and SPI at return periods.

Sites	Index	At-Site Quantiles (IFP Method)					At-Site Quantiles (QR Method)				
		2	5	10	25	50	2	5	10	25	50
Astore	RDI	1.44	1.96	2.27	2.57	2.75	1.28	1.80	2.14	2.52	2.78
	SPI	1.36	1.88	2.23	2.64	2.90	1.28	1.77	2.11	2.50	2.76
Badin	RDI	1.34	1.88	2.18	2.47	2.64	1.37	1.87	2.13	2.34	2.44
	SPI	1.31	1.82	2.11	2.40	2.56	1.38	1.90	2.18	2.45	2.59
Bahawalnagar	RDI	1.29	1.73	2.04	2.43	2.72	1.25	1.71	2.04	2.47	2.78
	SPI	1.31	1.81	2.14	2.51	2.76	1.30	1.80	2.13	2.51	2.75
Bahawalpur	RDI	1.29	1.73	2.04	2.43	2.72	1.26	1.73	2.06	2.48	2.79
	SPI	1.29	1.78	2.11	2.48	2.72	1.29	1.80	2.13	2.51	2.76
Balakot	RDI	1.35	1.98	2.46	3.10	3.58	1.29	1.79	2.12	2.52	2.80
	SPI	1.22	1.68	1.98	2.33	2.55	1.26	1.74	2.05	2.41	2.64
Barkhan	RDI	1.37	2.03	2.53	3.17	3.64	1.23	1.75	2.12	2.58	2.92
	SPI	1.39	2.06	2.54	3.13	3.56	1.28	1.82	2.19	2.63	2.94
Bunji	RDI	1.39	1.91	2.20	2.50	2.67	1.32	1.82	2.14	2.49	2.72
	SPI	1.52	2.12	2.51	2.96	3.26	1.27	1.74	2.03	2.36	2.57
Cherat	RDI	1.21	1.65	1.91	2.17	2.31	1.26	1.78	2.13	2.56	2.87
	SPI	1.17	1.62	1.92	2.27	2.50	1.25	1.76	2.09	2.49	2.76
Chilas	RDI	1.49	2.03	2.34	2.66	2.84	1.34	1.84	2.12	2.38	2.53
	SPI	1.48	2.05	2.44	2.87	3.17	1.36	1.88	2.17	2.47	2.64
Chitral	RDI	1.25	1.71	1.97	2.24	2.39	1.32	1.83	2.14	2.49	2.72
	SPI	1.14	1.58	1.87	2.21	2.44	1.26	1.73	2.02	2.34	2.55
Chhor	RDI	1.46	2.04	2.37	2.70	2.88	1.32	1.85	2.18	2.54	2.78
	SPI	1.38	1.91	2.21	2.52	2.69	1.24	1.72	2.03	2.37	2.59
Dalbandin	RDI	1.42	2.10	2.61	3.27	3.76	1.25	1.82	2.23	2.75	3.14
	SPI	1.36	2.02	2.49	3.07	3.49	1.28	1.86	2.26	2.74	3.08
Darosh	RDI	1.21	1.66	1.91	2.17	2.32	1.31	1.84	2.17	2.55	2.80
	SPI	1.17	1.63	1.93	2.28	2.51	1.73	2.04	2.39	2.61	2.77
D-I Khan	RDI	1.30	1.75	2.06	2.45	2.74	1.26	1.74	2.09	2.54	2.88
	SPI	1.36	1.88	2.22	2.61	2.87	1.26	1.76	2.08	2.46	2.72
Dir	RDI	1.31	1.79	2.07	2.35	2.51	1.30	1.83	2.16	2.55	2.81
	SPI	1.34	1.87	2.21	2.61	2.88	1.24	1.73	2.04	2.40	2.63
Faisalabad	RDI	1.40	1.87	2.20	2.63	2.94	1.25	1.72	2.06	2.50	2.82
	SPI	1.32	1.82	2.15	2.53	2.77	1.28	1.77	2.09	2.47	2.72
Ghari	RDI	1.29	1.89	2.35	2.96	3.42	1.28	1.77	2.10	2.51	2.81
Dupatta	SPI	1.41	1.95	2.30	2.70	2.96	1.26	1.74	2.06	2.42	2.66
Gilgit	RDI	1.40	1.91	2.20	2.50	2.67	1.33	1.84	2.15	2.49	2.71
	SPI	1.30	1.81	2.15	2.53	2.79	1.26	1.73	2.02	2.35	2.55

Table 3. 4: At-Site Quantiles using IFP and QR Methods for RDI and SPI at return periods.

Sites	Index	At-Site Quantiles (IFP Method)					At-Site Quantiles (QR Method)				
		2	5	10	25	50	2	5	10	25	50
Gupis	RDI	1.24	1.70	1.96	2.23	2.38	1.32	1.86	2.18	2.52	2.73
	SPI	1.21	1.69	2.00	2.36	2.60	1.26	1.74	2.05	2.40	2.63
Hyderabad	RDI	1.37	1.91	2.21	2.51	2.69	1.34	1.85	2.13	2.41	2.56
	SPI	1.41	1.96	2.28	2.59	2.76	1.36	1.88	2.18	2.49	2.67
Islamabad	RDI	1.17	1.71	2.13	2.68	3.10	1.28	1.76	2.10	2.51	2.81
	SPI	1.27	1.76	2.08	2.43	2.67	1.26	1.74	2.05	2.40	2.63
Jaccobabad	RDI	1.17	1.63	1.89	2.15	2.30	1.27	1.78	2.12	2.53	2.83
	SPI	1.24	1.72	2.00	2.27	2.42	1.30	1.82	2.16	2.54	2.79
Jhelum	RDI	1.39	1.87	2.20	2.62	2.93	1.27	1.74	2.07	2.49	2.81
	SPI	1.32	1.82	2.15	2.53	2.78	1.27	1.75	2.06	2.41	2.64
Jiwani	RDI	1.28	1.78	2.07	2.35	2.51	1.37	1.95	2.31	2.70	2.96
	SPI	1.28	1.78	2.06	2.34	2.50	1.36	1.95	2.31	2.71	2.96
Kakul	RDI	1.17	1.73	2.14	2.70	3.12	1.27	1.78	2.12	2.53	2.82
	SPI	1.42	1.96	2.32	2.72	2.98	1.26	1.75	2.08	2.46	2.71
Kalat	RDI	1.14	1.69	2.10	2.64	3.03	1.22	1.80	2.20	2.71	3.06
	SPI	1.18	1.74	2.15	2.65	3.01	1.29	1.88	2.30	2.81	3.18
Karachi	RDI	1.32	1.85	2.14	2.44	2.60	1.37	1.89	2.17	2.41	2.52
	SPI	1.32	1.83	2.12	2.41	2.57	1.37	1.90	2.20	2.48	2.64
Khanpur	RDI	1.23	1.66	1.95	2.32	2.60	1.26	1.74	2.08	2.48	2.78
	SPI	1.21	1.67	1.97	2.32	2.55	1.30	1.81	2.14	2.52	2.76
Khuzdar	RDI	1.24	1.84	2.29	2.87	3.30	1.26	1.81	2.18	2.62	2.93
	SPI	1.35	2.00	2.46	3.04	3.45	1.31	1.88	2.26	2.71	3.02
Kohat	RDI	1.09	1.46	1.72	2.05	2.29	1.27	1.77	2.11	2.55	2.86
	SPI	1.09	1.51	1.78	2.09	2.30	1.25	1.73	2.05	2.42	2.66
Kotli	RDI	1.12	1.64	2.03	2.56	2.96	1.27	1.75	2.08	2.50	2.81
	SPI	1.18	1.63	1.92	2.26	2.47	1.27	1.75	2.07	2.43	2.67
Lahore	RDI	1.16	1.55	1.83	2.18	2.44	1.25	1.71	2.04	2.47	2.80
	SPI	1.16	1.60	1.89	2.23	2.45	1.29	1.78	2.10	2.48	2.73
Lasbella	RDI	1.23	1.66	1.95	2.32	2.60	1.32	1.85	2.18	2.55	2.79
	SPI	1.44	1.99	2.35	2.77	3.04	1.34	1.88	2.22	2.59	2.83
Mianwali	RDI	1.38	1.85	2.17	2.59	2.90	1.26	1.75	2.09	2.54	2.86
	SPI	1.33	1.83	2.17	2.55	2.80	1.25	1.74	2.06	2.43	2.68
Moin Jodoro	RDI	1.20	1.67	1.94	2.20	2.35	1.35	1.86	2.14	2.41	2.55
	SPI	1.23	1.71	1.98	2.25	2.40	1.36	1.89	2.19	2.49	2.66
Multan	RDI	1.19	1.60	1.89	2.25	2.51	1.25	1.73	2.07	2.51	2.83
	SPI	1.20	1.66	1.96	2.31	2.54	1.28	1.79	2.12	2.50	2.75

Table 3. 4: At-Site Quantiles using IFP and QR Methods for RDI and SPI at return periods.

Sites	Index	At-Site Quantiles (IFP Method)					At-Site Quantiles (QR Method)				
		2	5	10	25	50	2	5	10	25	50
Murree	RDI	1.18	1.73	2.14	2.70	3.12	1.24	1.77	2.13	2.56	2.87
	SPI	1.24	1.72	2.03	2.38	2.61	1.27	1.79	2.15	2.59	2.89
Muzaffarabad	RDI	1.38	2.03	2.52	3.18	3.67	1.29	1.77	2.10	2.50	2.78
	SPI	1.29	1.78	2.10	2.47	2.71	1.26	1.74	2.05	2.39	2.62
NawabShah	RDI	1.41	1.97	2.28	2.59	2.77	1.32	1.83	2.13	2.46	2.66
	SPI	1.47	2.05	2.37	2.70	2.88	1.34	1.87	2.18	2.52	2.72
Nokkundi	RDI	1.37	2.04	2.53	3.18	3.65	1.26	1.85	2.28	2.83	3.24
	SPI	1.38	2.04	2.52	3.11	3.54	1.29	1.88	2.29	2.79	3.15
Ormara	RDI	1.59	2.22	2.57	2.92	3.12	1.35	1.90	2.23	2.57	2.78
	SPI	1.33	1.85	2.15	2.44	2.60	1.35	1.90	2.24	2.59	2.81
Panjgur	RDI	1.15	1.70	2.12	2.65	3.05	1.28	1.86	2.25	2.71	3.02
	SPI	1.27	1.88	2.32	2.86	3.25	1.32	1.91	2.30	2.76	3.08
Parachinar	RDI	1.27	1.87	2.32	2.92	3.38	1.25	1.80	2.17	2.63	2.95
	SPI	1.02	1.42	1.67	1.96	2.15	1.24	1.77	2.12	2.55	2.85
Passni	RDI	1.27	1.77	2.05	2.33	2.49	1.35	1.91	2.25	2.62	2.85
	SPI	1.29	1.80	2.08	2.37	2.53	1.35	1.92	2.26	2.63	2.87
Peshawar	RDI	1.22	1.63	1.92	2.29	2.56	1.29	1.78	2.12	2.54	2.84
	SPI	1.16	1.60	1.90	2.23	2.45	1.24	1.72	2.02	2.37	2.59
Quetta	RDI	1.07	1.46	1.69	1.92	2.05	1.22	1.78	2.19	2.70	3.08
	SPI	1.26	1.75	2.08	2.45	2.70	1.27	1.84	2.24	2.74	3.09
Risalpur	RDI	1.23	1.65	1.94	2.31	2.59	1.29	1.78	2.11	2.52	2.82
	SPI	1.23	1.70	2.00	2.36	2.59	1.25	1.72	2.02	2.36	2.58
Rohri	RDI	1.31	1.83	2.12	2.41	2.57	1.28	1.78	2.11	2.50	2.77
	SPI	1.32	1.84	2.13	2.42	2.58	1.31	1.83	2.16	2.53	2.77
Saidu Sharif	RDI	1.37	2.01	2.49	3.14	3.63	1.30	1.80	2.14	2.53	2.81
	SPI	1.08	1.49	1.75	2.06	2.25	1.25	1.73	2.04	2.38	2.61
Sargodha	RDI	1.14	1.53	1.81	2.15	2.41	1.26	1.73	2.07	2.51	2.83
	SPI	1.13	1.56	1.85	2.17	2.39	1.27	1.76	2.08	2.45	2.70
Sialkot	RDI	1.19	1.60	1.88	2.24	2.50	1.26	1.72	2.05	2.48	2.80
	SPI	1.26	1.74	2.05	2.42	2.65	1.28	1.76	2.08	2.44	2.68
Sibbi	RDI	1.30	1.93	2.40	3.01	3.46	1.26	1.77	2.13	2.59	2.94
	SPI	1.43	2.12	2.62	3.23	3.68	1.28	1.80	2.15	2.56	2.83
Skardu	RDI	1.29	1.77	2.04	2.32	2.47	1.28	1.80	2.13	2.52	2.77
	SPI	1.23	1.72	2.03	2.40	2.64	1.28	1.78	2.12	2.51	2.77
Zhob	RDI	1.34	1.84	2.12	2.41	2.57	1.22	1.76	2.14	2.62	2.98
	SPI	1.30	1.80	2.13	2.52	2.77	1.26	1.80	2.18	2.64	2.96

Table 3. 5: At-Site quantile estimates for ungauged sites at selected return periods using QR Model for RDI and SPI indices.

Stations	Site Characteristics			At-Site Quantiles (RDI)					At-Site Quantiles (SPI)				
	Long	Lat	Elev	2	5	10	25	50	2	5	10	25	50
Bannu	70	33	376	1.251	1.725	2.051	2.459	2.743	1.208	1.77	2.092	2.359	2.599
Dadu	67.75	26.75	41	1.294	1.794	2.097	2.431	2.636	1.298	1.816	2.173	2.454	2.656
D-G Khan	70.75	30	124	1.244	1.71	2.03	2.434	2.718	1.252	1.736	2.114	2.423	2.657
Gawadar	62.25	25.25	7	1.367	1.923	2.248	2.587	2.782	1.349	1.941	2.288	2.612	2.838
Hunza	74.75	36.25	2220	1.309	1.78	2.033	2.268	2.376	1.226	1.847	2.093	2.249	2.432
Jhang	72.5	31.5	157	1.253	1.713	2.036	2.448	2.743	1.257	1.725	2.083	2.413	2.646
Kalam	72.5	35.5	2001	1.275	1.765	2.053	2.354	2.525	1.205	1.83	2.11	2.309	2.528
Larkana	68.25	27.5	51	1.275	1.762	2.07	2.425	2.653	1.279	1.792	2.162	2.442	2.656
Malam Jaba	73	34.75	2465	1.228	1.652	1.879	2.093	2.193	1.151	1.8	2.142	2.175	2.372
Okara	73.5	30.75	590	1.263	1.801	2.202	2.731	3.119	1.343	1.716	2.12	2.715	3.018
Rawalakot	73.75	35.75	1671	1.312	1.837	2.164	2.53	2.753	1.267	1.83	2.084	2.431	2.665
Sahiwal	73.25	31	250	1.26	1.746	2.096	2.55	2.879	1.298	1.717	2.094	2.53	2.785

3.4 Conclusion

RFA method is used for independent drought projections in the HCRs as well as individual metrological stations in Pakistan. The drought events are explored from RDI and SPI series and are used for drought projections in the form of quantiles in each region. The L-MRD and GOF Z-test are used to find the best-fit regional probability distributions that selected GPA probability distribution for the regions while PE3 distribution is selected for regions 2 and Region 3 using RDI only. The different distributions may be due to distinct geographical and climatological conditions.

The probability distributions are estimated through L-moments technique and are used to find three types of drought projections in Pakistan. Firstly, regional drought quantiles are calculated at selected return periods. The MCS process is used to measure the accuracy of drought quantiles by calculating and graphing RMSE and 90% error bounds for the regions. The estimated drought projections show high similarity at lower return periods but approximately after 10 years of return periods the variability and uncertainty increase in projections. Therefore, at higher return periods the estimated results may be used with caution.

Secondly, the at-site quantiles are obtained as a function of average by multiplying the average drought value of each site by regional quantile values at the corresponding return period. The average drought of the sites has a vital role which creating a difference from site to site. The stations with the maximum number and/or higher categories i.e., severe, or extreme drought events, have larger average drought and consequently gave larger at-site estimates. These quantiles show high variability among the sites of the regions and uncertainty at higher return periods. Lastly, the QR relationship is used to estimate drought conditions at ungauged sites. The regions can

be discussed according to the results as follow: Regions 1 and 2 have approximately low projected values of drought. Region two has high values of RMSE and 90% error bounds due to high variability in the occurrence of rainfall amounts. Region 3 has a minimum number of stations with large distances, therefore, the regional, as well as at-site quantile estimates, have high estimates with maximum variability. The region has the highest RMSE and 90% error bound values amongst the regions which show the high uncertainty of estimates. The ungauged site drought estimates are particularly important in this region because of vast areas with no gauging sites. It gives approximately good results compared to the IFP procedure. Region 4 has the minimum projected drought quantiles, RMSE, and 90% error bounds among the regions. Whereas the region has the least rainfall records and maximum drought compared to other parts of the country. It may be due to the continuously occurring less observed rainfall and high temperature where the data series have a statistically greater similarity. Statistically, the more homogenous data sets give the smaller estimates while the more variable data sets give the higher estimates. Regions 3 and 4 are affected by almost all droughts in the country. Region 5 has moderate drought estimates, RMSE and 90% error bound values with no greater threats of droughts according to the results. Both RDI and SPI indices have approximately identical results with little changes.

Chapter 4

Bivariate Drought Projections based on Copula Model(s)

4.1 Background

Drought is a multivariate phenomenon of several correlated variables such as drought duration and drought severity called drought characteristics which gives a more detailed description of the hazard. Therefore, drought can more effectively be investigated through multivariate modeling using drought characteristics. However, drought has a stochastic nature and assumes that drought variables are statistically independent and identically distributed (Cancelliere and Salas, 2004). But these assumptions do not satisfy due to the high correlation between drought variates, which may follow distinct univariate probability distributions (Tosunoglu and Can, 2016).

Different probability distributions may be fitted to drought characteristics. In a similar situation, multivariate drought modeling is quite difficult, and traditional methods of drought frequency analysis cannot precisely explain the relationship between drought characteristics (Song and Singh, 2010; Azam, et al., 2018). Therefore, traditional multivariate distributions like multivariate gamma, multivariate normal, etc., are not possible to use for multivariate drought analysis. Copula models can effectively be used when high correlation exists and different probability distributions are used for the drought characteristics to jointly simulate the drought variables (Cancelliere and Salas, 2004; Salvadori and De Michele, 2010). It can separately estimate the joint dependence structure using probability distributions of the drought characteristics.

There is no such detailed study based on drought characteristics in Pakistan. Therefore, this study is conducted to analyze the region-wise drought conditions in more detail

using copula models w.r.t drought duration and severity extracted from RDI and SPI indices over the HCR in Pakistan. The major objectives of this chapter (study) include: **Firstly**, RDI and SPI indices are used at 12-months' time scales to extract drought characteristics i.e., drought duration and severity for the selected meteorological stations all over Pakistan. **Secondly**, BHCRs are constructed for the drought characteristics. **Thirdly**, the copula model is used to combine the drought characteristics for joint return periods at selected years of return periods. **Lastly**, the joint and conditional severity-duration frequency (SDF) curves are constructed using drought duration and severity for the five homogenous regions.

4.2 Methodology

4.2.1 Statistical Characteristics of Drought

RDI and SPI are statistical methods which are used to find 12-months moving drought series using monthly precipitation and PET data for selected 55 meteorological stations. Runs Theory method may use to find drought characteristics of duration and severity for the sites, given in Fig. (2.1). Drought Durations are the number of consecutive months (M) of moving drought series whose RDI and SPI values are at less than or equal to -0.85. Drought Durations from RDI (D_{RDI}) and SPI (D_{SPI}) are obtained as:

$$D_{RDI} = \sum_{i=1}^M I(RDI_i \leq -0.85), \quad \text{for } M = 1, 2, 3, \dots, T. \quad (4.1)$$

$$D_{SPI} = \sum_{i=1}^M I(SPI_i \leq -0.85), \quad \text{for } M = 1, 2, 3, \dots, T. \quad (4.2)$$

where T represent for the full length of monthly data of a metrological station and $I(RDI_i \leq -0.85)$ is an indicator function denoting value 1 if $(RDI_i \leq -0.85)$, otherwise zero and similarly also for SPI index. Furthermore, the drought severity of

RDI (S_{RDI}) and SPI (S_{SPI}) indices are the sums of drought values within each of the above drought durations defined as follows:

$$S_{RDI} = - \sum_{l=1}^{D_{RDI}} RDI_l \quad (4.3)$$

$$S_{SPI} = - \sum_{l=1}^{D_{SPI}} SPI_l \quad (4.4)$$

4.2.2 Dependence between Drought Characteristics

Copula models have better applications in correlated variables. Let (d_1, s_1) , (d_2, s_2) , ..., (d_n, s_n) be a random sample of n observations for drought duration and severity variables (D , S), respectively. The correlation between these two variables can be quantitatively measured in this study using statistical methods of Pearson correlation coefficient, Spearman rank correlation coefficient, and non-parametric Kendall's tau correlation coefficient. These three measures are tested under the null hypothesis of no association. Null hypothesis of no association is rejected (accepted) by comparing p -value with a 5% level of significance. Kendall tau correlation coefficient is considered more suitable in drought variables (Nelsen 2006).

4.2.3 Copula Modelling

Copula modeling has two main steps. Firstly, selection of best-fit probability distributions for the variables and secondly, selection and estimation of copula models. Several possible probability distributions may be used for a set of data, but the selection of best-fit probability distribution(s) has a key role in reliable statistical estimates of regional projections (Saf, 2010). In statistics, there are many distributions with various parameters. Gamma, Weibull, and other two-parameter distributions are efficiently used in many studies for drought duration and severity (Shiau and Modarres 2009; Mirabbasi et al., 2012; Halwatura et al., 2015). However, one and two-parameter

distributions do not capture the tail effects of extreme events properly (Hosking and Willias, 1997; Chen and Guo, 2019). Hence, several studies used three-parameter probability distributions for the drought duration and severity (Azam et al., 2018; Mortuza et al., 2019). Therefore, we consider both the ideas of literature and select two-parameter Gamma (GAM), Logistic (LOG), Weibull (WBL), and three-parameters GEV and GPA probability distributions given in **Appendix B**. The best fit distributions for drought duration and severity are selected from these five distributions.

In statistical theory, different goodness of fit tests are used for the selection of best-fit probability distributions. Hence chi-square goodness-of-fit test is used to test the candidate probability distributions. Chi-square test statistic is as follows:

$$\chi^2_{dist} = \frac{1}{N} \sum_{i=1}^N \left(\frac{(O_i - E_i)^2}{E_i} \right) \quad (4.5)$$

where 'dist' is used for the candidate probability distribution, O_i is for the observed frequency and E_i is the expected frequency calculated as $E_i = N * P(x_i)$, where N is the number of observations and $P(x_i)$ is the corresponding probability from the candidate probability distribution. The selection criterion for the distribution is based on comparing the chi-square test statistics value with its p-value at a 5% level of significance. A distribution is considered the best fit if it has a minimum chi-square test value with a maximum p-value.

In the second step, copula models are selected for further results of this chapter. In mathematical statistics, the copula model was first developed by Sklar (1959) and later developed by Nelsen (2006), which is a type of multivariate probability distribution. The copula is a technique for modeling two or more dependent variables without involving some complex assumptions about the marginal and joint nature of the

variables (Arnold et al., 2006; Nelsen, 2006). In this study, copula models are used for drought projections through drought duration and severity variables. Copula model is used to diminish the modeling of k-dimensional distribution function $F(\cdot)$. Let D and S be the drought duration and severity while $F_d(d_i)$, and $F_s(s_i)$ be the selected best-fit univariate marginal distribution function, respectively. Then the bivariate Sklar's theorem is expressed as follows:

$$F_{ds}(d, s) = C\{F_d(d_i), F_s(s_i); \theta\}, \quad d, s \in \mathcal{R} \quad (4.6)$$

where θ is the parameter of the selected copula. It is necessary for a copula model that the variables must be uniformly distributed over the interval $[0, 1]$. Therefore, the selected best fit marginal probability distribution functions $F_d(d_i)$, and $F_s(s_i)$ are used to convert the drought duration and severity into cumulative probabilities. Furthermore, if marginal probability distributions are continuous then $C(\cdot)$ is uniquely defined (Nelsen, 2006). While $C(\cdot)$ represents the cumulative distribution function of Elliptical, Archimedean, and/or Extreme Value copulas, to measure the joint dependence between the drought duration and severity. In this study, three bivariate copula models of a single parameter i.e., Clayton from Archimedean, Galambos of Extreme Value family, and G-H are related to both Archimedean and Extreme Value families, are used.

The inversion of Kendall's τ method is used to estimate the copula parameters for the selected functions that use the relationship between Kendall's tau (τ_θ) and copula parameter (θ) (Nelsen, 2006; Genest and Favre, 2007). For Archimedean copulas the relationship between τ_θ and θ is given in **Appendix C**. This relation is based on the generating function of the respective copula as follows:

$$\tau = 1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} dt \quad (4.7)$$

where $\phi(t)$ is the generating function and $\phi'(t)$ is its derivative. While the extreme value copulas are estimated through the Pickands dependence function ($A(t)$) (Pickands, 1981). The mathematical relations of $A(t)$ for the selected extreme value copula models are given in **Appendix C**. Similarly, the same method for extreme values copula is performed through simulation using the bootstrap method as follows:

$$\tau = -1 + 4 \int_0^1 \frac{t(1-t)}{A(t)} dA'(t) \quad (4.8)$$

as $A'(t)$ is the right-hand derivative of $A(t)$ on $(0,1)$. The numerical measurements are always considered more robust as goodness-of-fit criteria for the selection of statistical models (Hosking and Willias, 1997). Several such criteria like Log-likelihood function, Akaike Information Criterion (AIC) (Akaike, 1974), and Bayesian Information Criterion (BIC) (Schwarz, 1978) are used to select the copula models for the regions. Let $f(x; \theta)$ be any function with a random sample of $x_1, x_2, \dots, x_n$ and parameter θ , then the likelihood function can be expressed as follow:

$$\mathcal{L}(\theta; x) = \prod_{i=1}^n f(x_i; \theta) \quad (4.9)$$

While the log-likelihood function is given by:

$$\ell(\theta; x) = \ln(\mathcal{L}(\theta; x)) \quad (4.10)$$

The equation for AIC is expressed as follows:

$$AIC = 2 * p - 2 * \ln(\hat{\mathcal{L}}) \quad (4.11)$$

Where p is the number of parameters of the statistical model and $\hat{\mathcal{L}}$ is the maximum estimated value of the likelihood function. Lastly, the BIC for a random sample of n observations is given as:

$$BIC = p * \ln(n) - 2 * \ln(\hat{\mathcal{L}}) \quad (4.12)$$

In the case of comparing more than one copula model, a model is considered to be the best fit if it has maximum absolute values for log-likelihood function, AIC, and BIC criteria. Similarly, the Kendall tau and Spearman rho correlation coefficients are calculated for the estimated copula models to know which copula model explains maximum dependence for the data. A model is considered the best fit if it has maximum value for the Kendall tau and Spearman's rho correlation coefficients.

4.2.4 Tail Dependence of Copula Models

Tail dependence is the asymptotic dependence of events in the fitted copula model. It is expressed as the probability of extreme events that jointly occur in the lower left and upper right or both corners of the scatter plots of ranked data. In case, if the data of two variables are correlated with extreme events, then copula models with strong tail dependence have particular significance (Huard et al., 2006). A copula that fails to model tail dependence is expected to give misleading projections of extreme events and return periods (Naz et al., 2019). In the case of drought analysis, if a fitted copula model does not capture the tail dependence within drought characteristics, it might give greater uncertainty in the estimates of drought hazard (Tosunoglu and Kisi, 2016). The lower and upper tail dependence is defined as follows:

$$\lambda_U = \lim_{u \rightarrow 1^-} \left[\frac{1 - 2u + C(u, u)}{1 - u} \right] \quad (4.13)$$

$$\lambda_L = \lim_{u \rightarrow 0^+} \left[\frac{C(u, u)}{u} \right] \quad (4.14)$$

Where u is a constant value of a standard uniform variable. Tail behavior is determined entirely by the form of copula, not by the marginal distribution. In drought investigation, upper-tail dependence has greater significance (Azam, et al., 2018).

4.2.5 Return Periods of Drought Characteristics

The joint probability distribution of drought characteristics is an efficient method for predicting and managing droughts (Shiau, 2006). Bivariate cumulative distributions are calculated through copula to measure the joint occurrence of drought duration and severity. The non-exceedance probability of jointly occurring drought variables is as:

$$P(D \leq d_i, S \leq s_i) = F_{ds}(d_i, s_i) = C(F_d(d_i), F_s(s_i)) \quad (4.15)$$

Where $C(\cdot)$ denotes the selected best-fit copula model, $F_d(d_i)$, and $F_s(s_i)$ are the selected best fit marginal probability distribution functions for drought duration and severity. Another type is the joint occurrence exceedance probability that is calculated while exceeding the specific threshold levels as follows:

$$P(D \geq d_i, S \geq s_i) = 1 - F_d(d_i) - F_s(s_i) + C(F_d(d_i), F_s(s_i)) \quad (4.16)$$

The copula model is also used to find conditional probabilities for specific drought duration (d') and severity (s') as follows:

$$P(S \leq s_i | D \geq d') = \frac{F_s(s_i) - C(F_d(d'), F_s(s_i))}{1 - F_d(d')} \quad (4.17)$$

$$P(D \leq d_i | S \geq s') = \frac{F_d(d_i) - C(F_d(d_i), F_s(s'))}{1 - F_s(s')} \quad (4.18)$$

One objective of this study is to find the risk of extreme drought events at various return periods in the future. Frequency analysis approach is used by hydrologists and water resources engineers based on the return period using hydrologic extreme events (Shiau and Shen, 2001). Drought return periods are particularly important due to suitable water usage planning in drought conditions (Serinaldi et al., 2009; Song and Singh, 2010). The univariate drought return periods are calculated as given by (Ganguli & Reddy, 2012; Mortuza et al., 2019; Bazrafshan, et al., 2020):

$$T_D = \frac{E(IAT)}{1 - F_d(d_i)} \quad (4.19)$$

$$T_S = \frac{E(IAT)}{1-F_S(s_l)} \quad (4.20)$$

here T_D and T_S denote the return periods for drought duration and severity, respectively. $E(IAT)$ is the average inter-arrival time of drought events, calculated by the ratio of a total number of years to the total number of droughts (Mortuza et al., 2019).

Primary and secondary return periods are the two types of joint return periods of drought events. The primary joint return periods contain the T_{OR} and the T_{AND} return periods. The T_{OR} is under the condition $D \geq d_l$ or $S \geq s_l$ i.e., either drought severity or duration exceed the specific values. While T_{AND} is calculated when $D \geq d_l$ and $S \geq s_l$ i.e., both drought severity and duration exceed the specific values (Shiau, 2003).

$$T_{OR} = \frac{E(IAT)}{P(D \geq d_l \text{ or } S \geq s_l)} = \frac{E(IAT)}{1 - C(F_d(d_l), F_s(s_l))} \quad (4.21)$$

$$T_{AND} = \frac{E(IAT)}{P(D \geq d_l \text{ and } S \geq s_l)} = \frac{E(IAT)}{1 - F_d(d_l) - F_s(s_l) + C(F_d(d_l), F_s(s_l))} \quad (4.22)$$

The secondary or Kendall's tau return periods are defined by Salvadori et al., (2011). Kendall measure $K_c(q)$ is related to the joint distribution of copula model to describe the risk level at which joint probability for the random variables is at least q -value, at a given probability of $q \in (0,1)$. This type of return period is well-defined, and each group of variables corresponds to a distinct risk area within a given return period as follow:

$$T_{KEN} = \frac{E(IAT)}{P(C(F_d(d_l), F_s(s_l)) \geq q)} = \frac{E(IAT)}{1 - K_c(q)} \quad (4.23)$$

where $K_c(q)$ represents Kendall's distribution function for the selected theoretical copula model at a q^{th} probability value and defined as:

$$K_c(q) = P(C(F_d(d_l), F_s(s_l)) \leq q) \quad (4.24)$$

The $K_c(q)$ can be found in bivariate extreme values as follows (Ghoudi et al., 1998):

$$K_c(q) = q - (1 - \tau_c)q * \ln q \quad (4.25)$$

where τ_c denotes Kendall's tau of selected copula $C(\cdot)$. However, for two-dimensional G-H copula model with parameter θ , the $K_c(q)$ is calculated as follows:

$$K_c(q) = q - \frac{q \cdot \log(q)}{\theta} \quad (4.26)$$

According to (Shiau, 2006), the conditional return periods of droughts are calculated from copula models using the following relations:

$$T_{S|D \geq d'} = \frac{E(IAT)}{1-F_d(d')} * \frac{1}{1-F_d(d')-F_s(s_l)+C(F_d(d'),F_s(s_l))} \quad (4.27)$$

$$T_{D|S \geq s'} = \frac{E(IAT)}{1-F_s(s')} * \frac{1}{1-F_d(d_l)-F_s(s')+C(F_d(d_l),F_s(s'))} \quad (4.28)$$

where $T_{S|D \geq d'}$ represents conditional return periods of drought severity S at a given duration $D \geq d'$ and vice versa. Conditional return periods have particular importance in drought applications. The failure of water resources risk requires studying drought events at a given threshold level of drought duration d' (or severity s').

4.3 Results and Discussion

4.3.1 Construction of BHCRs

Drought is effectively measured using drought duration and severity which are extracted using the method of run theory (Fig. (2.1)). Both the variables are used to find L-moment ratios for calculating the homogeneity measures. In the first step of this chapter, BHCRs are delineated for the group of sites w.r.t both drought duration and severity simultaneously (Yoo et al., 2012; Hao et al., 2017). Firstly, the k-mean clustering algorithm is used to partition the study area into subjective homogenous groups of sites. The site characteristics i.e., latitude, longitude, elevation, and the means of drought duration, severity, and annual precipitation, are used to perform this algorithm. After checking for different k values, $k=5$ is found suitable to construct five homogenous groups based on drought duration and severity. Subjective homogenous

groups of sites are adjusted by changing some of the sites into each other based on geographical positioning, only if it is possible to satisfy the homogeneity.

Secondly, a discordancy measure is used for each station to check the climate data for errors and any discordant site(s) in the adjusted homogenous groups. The discordancy measure shows that in region 4, Nawabshah station does not adjust due to greater than 3 and dropped from further analysis of the study. After the removal of Nawabshah station, the measure is checked again and given in Table (4.1), which satisfied the critical values given in Table (2.2) for drought duration and severity.

Thirdly, for the ultimate checking of homogeneity, the heterogeneity measure is used with three possible results given in Table (4.2). Hosking and Willias (1997) suggest that H_1 is the most powerful measure with greater power of discrimination and is preferred to be used for heterogeneity as compared to others. The results show that all the regions are acceptably homogenous except region-5 for SPI which is possibly homogenous whereas acceptably homogenous w.r.t RDI and no further adjustment is possible in the nearby sites. Hence, the region is considered homogenous for further statistical analysis of the study. These three steps confirm the ultimate homogeneity of five regions based on drought duration and severity w.r.t RDI and SPI with the same sites. Map of the geographical locations of the BHCRs of stations is shown in Fig. (4.1).

Table 4. 1: Discordancy Measures for the sites of the five BHCRs w.r.t drought duration and severity using RDI and SPI indices.

Regions	Sites	RDI: Results				SPI: Results				Regions	Sites	RDI: Results				SPI: Results																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																
		N		D _{Dur}		D _{Sev}		N				D _{Dur}		D _{Sev}		N		D _{Dur}		D _{Sev}																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
		N	D _{Dur}	D _{Sev}	N	D _{Dur}	D _{Sev}	N	D _{Dur}			D _{Sev}	N	D _{Dur}	D _{Sev}	N	D _{Dur}	D _{Sev}	N	D _{Dur}	D _{Sev}																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
Region-1																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																

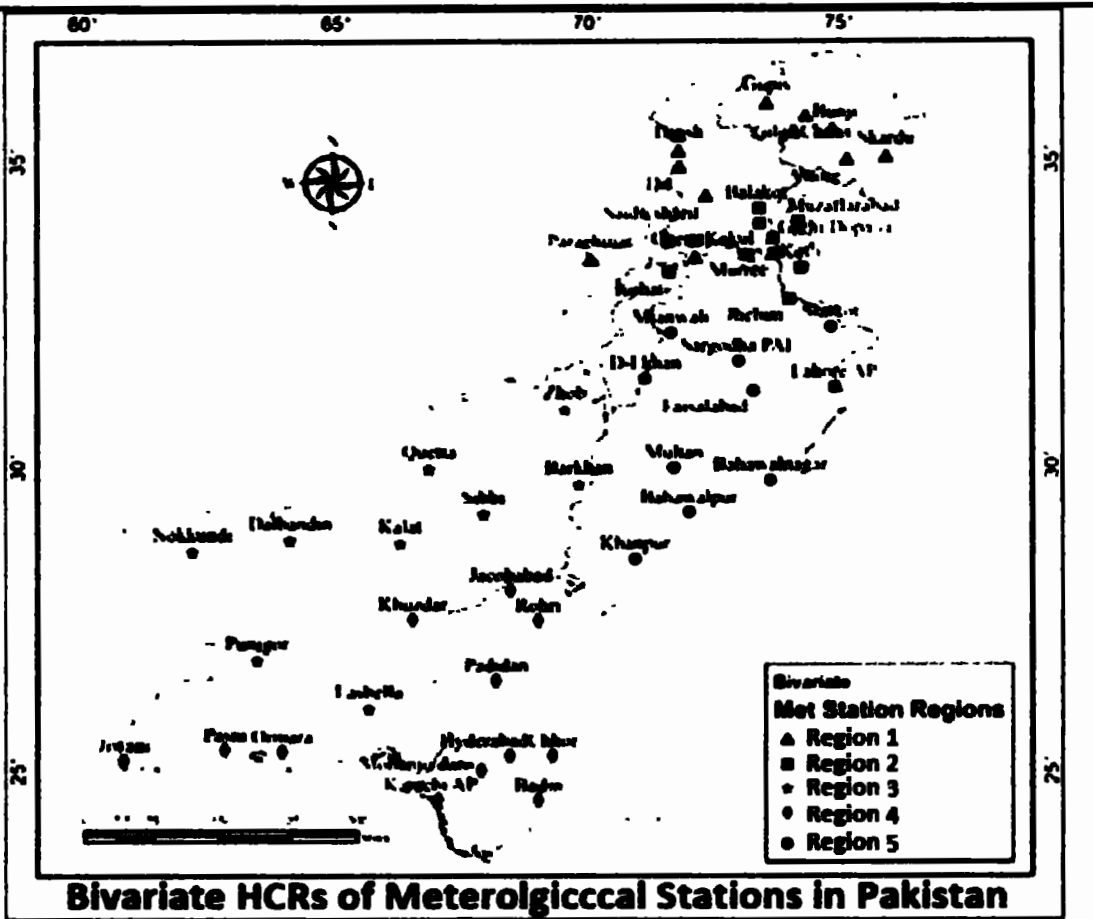
Table 4. 1: Discordancy Measures for the sites of the five BHCRs w.r.t drought duration and severity using RDI and SPI indices.

Regions	Sites	RDI: Results			SPI: Results			Regions	Sites	RDI: Results			SPI: Results					
		N		D _{Dur}	D _{Sev}		N			D _{Dur}	D _{Sev}		N		D _{Dur}	D _{Sev}		
	Islamabad	30	0.26	0.84	28	1.11	1.64		Zhob	24	0.94	1.21	25	1.95	1.35			
Region-5																		
Region-4	Lasbella	12	2.37	2.35	13	0.54	1.03											
	Badin	19	0.85	0.89	19	0.37	0.85		Bahawalnagar	22	0.48	0.51	20	0.86	1.17			
	Chhor	23	0.84	0.75	21	0.72	0.41		Bahawalpur	22	0.2	0.5	22	0.38	0.12			
	Hyderabad	25	1.05	0.75	23	1.02	1.28		D.I Khan	20	1.05	0.79	20	1.59	1.33			
	Jacobabad	29	0.41	0.49	27	0.99	0.44		Faisalabad	24	0.15	0.2	29	0.38	0.86			
	Jiwani	21	1.3	0.78	22	1.57	0.81		Khanpur	20	1.54	1.94	18	1.8	1.96			
	Karachi	19	1.98	1.91	21	0.98	0.87		Lahore	18	1.83	1.57	28	0.86	0.79			
	Mohin Jodaro	22	2.34	2.57	18	2.28	2.29		Mianwali	19	1.77	1.75	17	1.8	1.19			
	Ormara	15	0.19	0.46	15	0.06	0.33		Multan	30	1.56	1.03	29	1.25	0.89			
	Padidan	23	0.92	0.6	20	0.5	0.41		Sargodha	24	0.71	0.42	25	0.35	0.16			
Passni	20	0.33	0.24	19	0.35	0.78		Sialkot	26	0.72	1.29	25	0.73	1.53				
Rohri	22	1.54	1.26	27	1.34	1.18												
Khuzdar	16	0.25	1.3	20	1.82	2.37												

Note: The abbreviations of D_{Dur} and D_{Sev} represent the values of discordancy measures of the BHCRs.

Table 4. 2: Heterogeneity Measures of drought duration and severity for RDI and SPI.

Regions	Type	RDI: Heterogeneity Measure			SPI: Heterogeneity Measure		
		H1	H2	H3	H1	H2	H3
Region-1							
	Severity	-1.64	-1.48	-1.46	-0.1	-0.11	-1.29
	Duration	-0.95	-0.93	-1.07	0.83	1.02	-0.05
Region-2							
	Severity	-0.84	-0.6	-0.08	-1.41	-0.92	-0.41
	Duration	0.37	0.06	-0.16	-1.27	-0.84	-0.42
Region-3							
	Severity	0.87	1.95	2.43	1.3	1.27	1.29
	Duration	0.33	1.58	1.97	1.28	1.83	2.01
Region-4							
	Severity	-2.07	-1.92	-1.15	-2.18	-1.4	-1.44
	Duration	-1.62	-0.61	0.62	-1.68	0.12	-1.26
Region-5							
	Severity	-0.43	-1.29	-0.35	-1.31	-0.03	1.17
	Duration	-0.4	0.5	1.78	-0.82	0.2	1.52



Bivariate HCRs of Meteoroligiccal Stations in Pakistan

Figure 4. 1: Bivariate HCRs of Pakistan based on drought duration and severity.

4.3.2 Copula modeling for bivariate drought Projections

In the second step, quantitative strength between drought characteristics is measured using Kendall's tau, Spearman's rank, and Pearson's correlation coefficients for the BHCRs using the data from drought indices, given in Table (4.3). The measures give a high positive correlation which means that drought characteristics are strongly correlated. Pearson correlation coefficient is better to measure linear dependence and might not give good results when there are outliers in the data (Naz et al., 2019). In such cases, Kendall's tau will be a good choice to describe more variations (Klein, et al., 2011). In such correlated variables, the bivariate copula model is more suitable for further statistical analysis (Tosunoglu and Can, 2016).

In the third step, the selection of best-fitted probability distributions is to be selected. It is necessary for copula modeling and projections of hydrological events even if it does not satisfy some of the statistical assumptions (Saf, 2010). Chi-square goodness-of-fit test is used as a numerical measurement for the selection of best-fit probability distributions to determine whether the candidate probability distribution is acceptable or not. The five probability distributions are estimated, and chi-square test with p-values is given in Table (4.4). The results are compared according to the mentioned criteria in section 4.4 for the selection of best-fit probability distributions. GPA distribution is selected for both drought duration and severity of regions 1, 2, and 5 and for drought severity in region 4 using RDI and SPI indices Weibull distribution is selected for region 3 using RDI index, and in region 4 for duration using RDI and SPI indices while Gamma distribution is selected for region 3 using SPI index only. The selected probability distributions are estimated using the maximum likelihood estimation method, given in Table (4.5). These selected distributions are used in copula modeling in the next step.

Table 4. 3: Correlation coefficients between drought duration and severity.

Regions	RDI: Correlations			SPI: Correlations		
	Kendall	Rho	Pearson	Kendall	Rho	Pearson
Region 1	0.907	0.978	0.979	0.910	0.982	0.975
Region 2	0.901	0.975	0.852	0.898	0.976	0.957
Region 3	0.902	0.980	0.973	0.884	0.964	0.946
Region 4	0.890	0.977	0.969	0.883	0.969	0.961
Region 5	0.914	0.984	0.965	0.902	0.979	0.973

Table 4. 4: Chi-Square goodness-of-fit measures for the five probability distributions.

Regions	Dist.	RDI: Chi-Square values		SPI: Chi-Square values	
		Sev. (p-val)	Dur. (p-val)	Sev. (p-val)	Dur. (p-val)
Region-1					
	GAM	14.16 (0.003)	9.91 (0.019)	16.81 (0.001)	8.18 (0.085)
	WBL	7.77 (0.051)	10.35 (0.035)	11.15 (0.011)	5.44 (0.245)
	LOG	80.49 (0.000)	44.87 (0.000)	52.97 (0.000)	39.02 (0.000)
	GEV	79.14 (0.000)	102.26 (0.000)	75.71 (0.000)	132.40 (0.000)
	GPA	2.10 (0.551)	1.45 (0.694)	6.45 (0.092)	2.62 (0.623)
Region-2					
	GAM	22.75 (0.000)	8.93 (0.003)	11.08 (0.004)	12.99 (0.002)
	WBL	12.99 (0.005)	5.12 (0.024)	11.72 (0.008)	9.18 (0.010)
	LOG	48.61 (0.000)	29.79 (0.000)	61.44 (0.000)	36.28 (0.000)
	GEV	65.36 (0.000)	78.30 (0.000)	89.31 (0.000)	97.22 (0.000)
	GPA	2.47 (0.481)	1.96 (0.375)	9.05 (0.029)	5.42 (0.144)
Region-3					
	GAM	7.31 (0.026)	2.59 (0.274)	1.11 (0.574)	23.53 (0.000)
	WBL	5.54 (0.063)	2.21 (0.331)	1.69 (0.429)	23.55 (0.000)
	LOG	13.99 (0.000)	12.45 (0.000)	24.63 (0.000)	40.12 (0.000)
	GEV	53.09 (0.000)	67.95 (0.000)	88.50 (0.000)	175.37 (0.000)
	GPA	4.49 (0.106)	3.37 (0.185)	6.62 (0.037)	25.70 (0.000)
Region-4					
	GAM	5.69 (0.058)	4.92 (0.027)	3.82 (0.148)	8.17 (0.004)
	WBL	2.73 (0.256)	4.70 (0.030)	1.47 (0.479)	7.12 (0.028)
	LOG	36.23 (0.000)	4.82 (0.000)	36.37 (0.000)	7.18 (0.000)

Table 4. 4: Chi-Square goodness-of-fit measures for the five probability distributions.

Regions	Dist.	RDI: Chi-Square values		SPI: Chi-Square values	
		Sev. (p-val)	Dur. (p-val)	Sev. (p-val)	Dur. (p-val)
Region-5	GEV	35.15 (0.000)	55.27 (0.000)	45.46 (0.000)	55.19 (0.000)
	GPA	0.20 (0.905)	5.20 (0.074)	0.52 (0.769)	5.25 (0.073)
	GAM	1.41 (0.493)	14.91 (0.002)	1.11 (0.574)	23.53 (0.000)
	WBL	1.22 (0.543)	14.24 (0.003)	1.69 (0.429)	23.55 (0.000)
	LOG	22.35 (0.000)	17.16 (0.000)	24.63 (0.000)	40.12 (0.000)
	GEV	64.40 (0.000)	109.71 (0.000)	88.50 (0.000)	175.37 (0.000)
	GPA	4.93 (0.085)	14.94 (0.002)	6.62 (0.037)	25.70 (0.000)

Table 4. 5: Parameters estimates of the best-fitted regional probability distributions of drought duration and severity.

Regions	Type	Dist.	RDI: Est. Parameters			Dist.	SPI: Est. Parameters		
			$\Theta 1$	$\Theta 2$	$\Theta 3$		$\Theta 1$	$\Theta 2$	$\Theta 3$
Region-1									
	Sev.	GPA	-0.108	5.242	-0.371	GPA	-0.084	5.584	-0.296
	Dur.	GPA	0.254	4.172	-0.259	GPA	0.264	4.494	-0.160
Region-2									
	Sev.	GPA	0.134	4.344	-0.350	GPA	0.147	4.328	-0.343
	Dur.	GPA	0.398	3.477	-0.255	GPA	0.402	3.479	-0.209
Region-3									
	Sev.	GPA	-0.078	6.371	-0.337	GPA	-0.008	6.812	-0.307
	Dur.	WBL	----	6.437	1.001	WBL	----	6.848	1.026
Region-4									
	Sev.	GPA	-0.091	6.949	-0.269	GPA	-0.177	7.186	-0.261
	Dur.	GPA	0.327	5.192	-0.194	GPA	0.316	5.051	-0.211
Region-5									
	Sev.	WBL	----	8.218	0.867	GAM	----	13.310	0.666
	Dur.	WBL	----	6.516	1.022	GAM	----	7.180	0.900

Note: $\Theta 1$, $\Theta 2$, and $\Theta 3$ stand for the estimated location, scale, and shape parameters of the selected best-fitted probability distributions, respectively. The dashed spaces in Table show that Weibull and Gamma probability distributions don't have location parameters

In the fourth step, Clayton, Gumbel, and Galambos copula models are initially selected to show the joint relationship between drought duration and drought severity. For fitting a copula model, it is required to transform the observed data of these two variables into uniformly distributed over the interval (0,1). Therefore, the previously selected best fit probability distributions for the regions are used to calculate the cumulative probabilities for each point of drought duration as well as drought severity. Copula parameters are estimated using inversion of Kendall's τ method, given in **Table (4.6)**. Log-likelihood function, AIC, and BIC are calculated as numerical criteria for the selection of best-fit copula functions in each region. AIC and BIC criteria suggest that Clayton copula is a suitable model followed by the G-H copula for modeling the relationship between drought duration and drought severity except region 2 for RDI index only. The results are given in **Table (4.7)**.

Climate and water conditions become more severe for humans as well as ecological systems when the duration and severity of drought extend beyond certain limits (Zhang et al., 2013). It is necessary to investigate and estimate such extreme drought events. The upper-tail dependence of copula models is used to measure such a relationship between drought characteristics. Therefore, it is better to use the copula models with upper-tail dependence for drought analysis (Ganguli and Reddy, 2014; She and Xia, 2018; Azam et al., 2018). The upper-tail dependence is calculated for all copula models given in **Table (4.8)**. Clayton copula has a lack of measuring upper-tail dependence while G-H copula has larger values followed by Galambos copula. Additionally, the Kendall tau and Spearman's rho correlation coefficients for the selected copula models are calculated and the results are given in **Table (4.8)**. A copula model is considered the best fit if it explains the maximum strength of the relationship between the variables. According to the results, the G-H copula model has maximum values for all the regions

using SPI and RDI indices. Lee et al., (2013) suggested that the G-H copula model is appreciated that giving more drought risk information in bivariate frequency analysis. Because of these restrictions and the above discussion, the Clayton copula is dropped due to a lack of upper tail dependence and minimum correlation coefficients whereas the G-H copula model is selected as best-fit for all the regions using RDI and SPI indices. G-H copula model is defined below:

$$C(d, s) = \exp \left[-\{(-\ln(d))^\theta + (-\ln(s))^\theta\}^{1/\theta} \right] \quad (4.40)$$

Where d and s are the CDF values for drought duration and severity, respectively and θ is the value of copula parameter estimated for the regions.

In the last step, the selected copula model is used to calculate the future drought risk at various return periods. According to Table (4.3), drought duration and drought severity are highly correlated variables. Therefore, the univariate frequency analysis may overestimate or underestimate the risk of drought events (Salvadori, 2004). In extreme drought analysis, drought durations and severities have strong impacts, particularly with longer duration and high severity, due to larger effects on water resources forecasting and high risk to the ecological and agricultural system (Azam, et al., 2018).

Table 4. 6: Parameters estimates of the selected Copula models for the regions.

Region	RDI: Copula Parameters			SPI: Copula Parameters		
	Clayton	Gumbel	Galambos	Clayton	Gumbel	Galambos
Region 1	18.164	11.016	10.306	17.062	10.183	9.468
Region 2	15.025	8.747	8.951	15.693	9.009	8.300
Region 3	17.181	7.516	6.799	13.648	6.108	5.391
Region 4	10.802	6.855	6.140	9.627	6.361	5.652
Region 5	17.673	8.099	7.372	15.377	8.637	7.922

Table 4. 7: Estimates of loglikelihood, AIC, and BIC criteria for the Copula models.

Regions	Copula	RDI: Selection Criteria			SPI: Selection Criteria		
		Logik	AIC	BIC	LogLik	AIC	BIC
Region-1							
	Clayton	517.22	-1032.44	-1028.83	535.83	-1069.65	-1066.00
	Gumbel	514.22	-1026.43	-1022.83	520.50	-1039.00	-1035.35
	Galambos	513.55	-1025.11	-1021.51	519.74	-1037.47	-1033.82
Region-2							
	Clayton	540.79	-1079.58	-1075.84	555.58	-1109.16	-1105.42
	Gumbel	542.75	-1083.49	-1079.75	530.13	-1058.26	-1054.52
	Galambos	542.01	-1082.03	-1078.29	529.32	-1056.63	-1052.90
Region-3							
	Clayton	319.46	-636.92	-633.76	275.70	-549.41	-546.28
	Gumbel	255.21	-508.42	-505.26	223.66	-445.32	-442.19
	Galambos	253.35	-504.71	-501.55	221.68	-441.35	-438.22
Region-4							
	Clayton	354.92	-707.84	-704.36	348.55	-695.10	-691.57
	Gumbel	348.49	-694.97	-691.49	345.28	-688.55	-685.02
	Galambos	348.02	-694.05	-690.57	344.44	-686.88	-683.35
Region-5							
	Clayton	431.94	-861.87	-858.46	405.45	-808.89	-805.44
	Gumbel	357.76	-713.52	-710.10	379.52	-757.03	-753.58
	Galambos	356.39	-710.78	-707.37	378.64	-755.29	-751.84

Note: A copula with greater absolute values of Loglikelihood, AIC, and BIC are considered the most suitable according to these criteria.

Table 4. 8: Dependence measures for the selected estimated Copula models.

Regions	Copula	RDI: Measures			SPI: Measures		
		Tau	Rho	TD (L, U)	Tau	Rho	TD (L, U)
Region-1							
	Clayton	0.901	0.985	(0.963, 0)	0.895	0.983	(0.960, 0)
	Gumbel	0.945	0.996	(0, 0.961)	0.941	0.995	(0, 0.959)
	Galambos	0.909	0.988	(0, 0.935)	0.902	0.986	(0, 0.929)
Region-2							
	Clayton	0.883	0.979	(0.955,0)	0.887	0.98	(0.957, 0)
	Gumbel	0.933	0.994	(0, 0.953)	0.936	0.994	(0, 0.955)
	Galambos	0.896	0.985	(0, 0.925)	0.889	0.982	(0, 0.920)
Region-3							
	Clayton	0.896	0.983	(0.960, 0)	0.872	0.975	(0.950, 0)
	Gumbel	0.942	0.995	(0, 0.959)	0.927	0.992	(0, 0.948)
	Galambos	0.867	0.975	(0, 0.903)	0.836	0.962	(0, 0.879)
Region-4							
	Clayton	0.844	0.963	(0.938, 0)	0.828	0.956	(0.931, 0)
	Gumbel	0.907	0.988	(0, 0.934)	0.896	0.984	(0, 0.925)
	Galambos	0.854	0.969	(0, 0.893)	0.843	0.965	(0, 0.885)
Region-5							
	Clayton	0.898	0.984	(0.962, 0)	0.885	0.98	(0.956, 0)
	Gumbel	0.943	0.995	(0, 0.960)	0.935	0.994	(0, 0.954)
	Galambos	0.876	0.978	(0, 0.910)	0.884	0.981	(0, 0.916)

Note: The TD (L, U) represents the lower and upper tail dependence of the selected copula.

Bivariate return periods are used to explain maximum drought risk in the regions by considering drought duration and drought severity values. Primary and secondary return periods are the two types of bivariate joint return periods. Primary return periods of drought are calculated for T_{OR} and T_{AND} return periods where T_{OR} return periods are more practical and called standard return periods (Azam, et al., 2018). While secondary or Kendall return periods (T_{KEN}) give more reliable results for drought risk assessment compared to univariate and primary return periods, which would be helpful to point out high danger areas in the regions (Salvadori and De Michele 2010). It always occurs between T_{OR} and T_{AND} joint return periods (i.e., $T_{OR} \leq T_{KEN} \leq T_{AND}$) and neither over-estimate nor under-estimate drought risk compared to primary return periods.

Similarly, conditional return periods are important for water resources engineers in constructing hydraulic design criteria and risk assessments (Shiau, 2006; Song and Singh, 2010a, b). Conditional return periods can be found for two scenarios. Firstly, conditional return periods are calculated for severity given durations using the conditional distribution function of selected best-fit G-H copula:

$$C(s|d; \theta) = \frac{1}{d} \exp \left(-\{[-\log(d)]^\theta + [-\log(s)]^\theta\}^{1/\theta} \right) \left(1 + \left\{ \frac{-\log(s)}{-\log(d)} \right\}^\theta \right)^{-1+\frac{1}{\theta}} \quad (4.41)$$

Furthermore, the conditional G-H copula for calculating conditional return periods of drought durations given threshold levels of drought severities are as follows:

$$C(d|s; \theta) = \frac{1}{s} \exp \left(-\{[-\log(s)]^\theta + [-\log(d)]^\theta\}^{1/\theta} \right) \left(1 + \left\{ \frac{-\log(d)}{-\log(s)} \right\}^\theta \right)^{-1+\frac{1}{\theta}} \quad (4.42)$$

Bivariate frequency analysis is used and estimated primary (T_{OR} & T_{AND}) and secondary (T_{KEN}) return periods for drought duration and drought severity at univariate return periods of 1.25-, 2-, 5-, 10-, 25-, 50-, and 100-years using RDI and SPI indices. The calculated results of return periods are given in Table (4.9). Firstly, best fit distributions are used for the estimated quantiles of drought duration and severity of the above return periods. The values show that region 2 has least, regions 3 and 5 have largest, while the remaining regions 1 and 4 have in between quantiles of drought duration and severity. Secondly, the best fit regional copula models are used to find joint as well as conditional return periods based on the estimated quantiles of drought duration and severity for the regions. Secondary return periods are found using equation (4.32) for non-exceedance probabilities of the return periods as $q = 0.20, 0.50, 0.80, 0.90, 0.96, 0.98, \text{ and } 0.99$.

Return period is the expected inter-arrival time between drought events of a specific or less magnitude (Haan, 1977, Serinaldi et al., 2009). Let us consider a pair of estimated quantiles for drought severity (4.04) and drought duration (3.42) at 2-year return period

using RDI index in region 1. The G-H copula model is used to find joint and conditional return periods. The T_{OR} primary method gave 4.86 years return period that either drought severity exceeds 4.04 or drought duration exceeds 3.42 months. Similarly, the T_{AND} return period is calculated that both drought severity and drought duration exceed the values 4.04 and 3.42 respectively is 5.31 years whereas the secondary return period (T_{KEN}) is 5.22 years for the same values. The T_{OR} , T_{AND} , and T_{KEN} return periods means that a similar drought may expectedly be repeated after 4.86, 5.31, and 5.22 years, respectively. Conditional return periods for drought severity of 4.04 given that the drought duration is 3.42 is 6.05 years while for the drought duration 3.42 months given that the drought severity is 4.04 is 6.74 years. The estimated values for drought severity and drought duration using SPI are 4.21 and 3.56, respectively at 2 years return period. The joint return periods for T_{OR} , T_{AND} , and T_{KEN} using SPI estimated values are 4.72, 5.19, and 5.09 years respectively whereas conditional return periods for severity given duration and duration given severity are 7.19 and 6.77, respectively.

Similarly, in Table (4.9), the values are calculated using quantiles of drought duration and drought severity for the regions at 1.25-, 2-, 5-, 10-, 25-, 50-, and 100-years of return periods. The regions have changes in joint and conditional return periods due to changes in estimated quantiles of drought duration, drought severity, parameters of regional probability distributions, and copula model. A region has more threats of drought if it has smaller return periods because the drought is expected to repeat soon. According to the results, regions 3 and 5 have larger return periods but also have larger estimated quantiles of drought duration and severity. Region 2 has the least return periods while regions 1 and 4 lie in the middle. However, due to larger differences between the estimated drought duration, drought severity, and return periods, it is difficult to

comment on the severity of the regions. Therefore, for better comparison, we graphically display the return periods through contour lines with similar axis ranges for the regions.

In bivariate frequency analysis, there may be multiple pairs of drought duration and severity for a certain return period which may not be simplified using univariate return periods. In SDF curve, the contour line of any return period shows all possible pairs of the combination of drought duration and severity. Therefore, SDF curves using contour lines for 1.25, 2, 5, 10, 25, 50, and 100 years of return periods would best explain these joint return periods. The SDF curves show the distribution of durations with corresponding severities, given in Fig. (4.2) for the regions using RDI and SPI. It is clear from curves of T_{OR} and T_{AND} return periods, that there are significant changes in results of regions, particularly at higher return periods. According to SDF curves, the T_{OR} return periods are explained by considering a pair of drought duration and severity values. For example, in regions 1, 3, and 4, a drought with a duration of 40 months or severity of 80 is expected to return in 25 years while it will take 50 and 100 years in regions 2 and 5 respectively. Similarly, results for T_{AND} have approximately similar return periods with more changes for the same pair of duration and severity. The variability between SPI and RDI increases with the increase in return periods particularly in regions 2 and 3.

Contour lines show that regions 1, 3, and 4 have relatively more frequent droughts, followed by region 2 while region 5 has rare chances of droughts. Region 1 has the world's 3rd largest glacier with less rainfall and more snowfall whereas snowfall is not considered in the results and resultantly, the region shows more drought conditions. The T_{OR} drought return periods have no bounds that either drought duration or severity can exceed and are always less than T_{AND} return periods as the probability of jointly occurring both variables are less compared to only one out of two variables.

Table 4. 9: Drought characteristics, Joint {Primary (T_{OR} & T_{AND}) & secondary (T_{KEN}) and Conditional return periods for BHCRs.

Regions	Return Periods	RDI: Return Periods						SPI: Return Periods									
		Drought Characteristics			Joint			Drought Characteristics			Joint						
		Severity	Duration	OR	AND	KEN	SID	D S	Conditional	Severity	Duration	OR	AND	KEN	SID	D S	Conditional
Region-4	50	51.71	28.64	138.9	168.3	162.8	211.7	204.8	51.49	29.09	140.2	178.0	168.8	211.7	214.8		
	100	70.33	35.88	277.7	336.8	325.7	423.3	409.5	68.94	35.91	280.3	356.3	337.6	423.4	429.5		
	1.25	1.51	1.51	3.45	3.73	3.68	5.43	5.09	1.47	1.47	3.45	3.75	3.7	4.02	3.7		
	2	5.20	4.18	5.35	6.17	6.00	7.85	7.57	5.28	4.09	5.34	6.23	6.04	5.75	5.48		
	5	13.9	10.13	13.11	15.82	15.19	19.60	18.9	14.2	10.00	13.06	16.00	15.31	14.33	13.69		
	10	22.06	15.39	26.06	31.87	30.50	39.19	37.8	22.51	15.29	25.95	32.26	30.75	28.67	27.38		
	25	35.46	23.53	64.92	80.00	76.42	97.98	94.51	36.07	23.59	64.63	81.02	77.06	71.67	68.46		
	50	48.03	30.71	129.7	160.2	152.9	196.0	189.0	48.72	31.03	129.1	162.3	154.2	143.3	136.9		
	100	63.18	38.94	259.3	320.6	306.0	391.9	378.1	63.87	39.64	258.1	324.8	308.6	286.7	273.9		
	Region-5	1.25	1.40	1.42	3.40	3.64	3.60	5.48	5.13	1.28	1.33	3.41	3.63	3.59	4.88	4.98	
2		4.90	4.04	5.31	5.99	5.85	7.83	7.56	5.14	4.25	5.33	5.97	5.83	7.36	7.44		
5		13.54	9.76	13.04	15.28	14.78	19.57	18.89	13.74	10.24	13.11	15.2	14.74	18.39	18.59		
10		22.10	14.71	25.96	30.74	29.66	39.15	37.77	21.32	15.04	26.09	30.56	29.56	36.77	37.18		
25		36.98	22.18	64.72	77.11	74.28	97.86	94.43	33.01	21.77	65.06	76.64	74.02	91.94	92.94		
50		51.71	28.64	129.3	154.4	148.6	195.7	188.9	43.3	27.16	130.0	153.4	148.1	183.9	185.9		
100		70.33	35.88	258.5	309.0	297.4	391.5	377.7	55.05	32.83	259.9	307.0	296.3	367.7	371.8		

Note: "Drought Characteristics" are drought duration and severity values estimated at 1.25, 2, 5, 10, 25, 50, and 100 return periods, joint return periods contain both primary and secondary return periods while in conditional return periods $S|D$ means $S|D \geq d'$ and $D|S$ means $D|S \geq s'$.

Conditional return periods for drought severity given that threshold levels of drought durations using 10, 30, 50, and 90 percentiles are computed for the regions based on RDI and SPI indices. The conditional SDF curves are presented in Fig. (4.3). Similarly, conditional return periods for drought duration given that threshold levels of drought severity using 10, 30, 50, and 90 percentiles are computed and presented in Fig. (4.4). These percentile levels give different values of drought duration and severity using RDI and SPI given within each graph of the regions. Conditional return periods of both types are increasing continuously with higher changes as the percentiles increases step by step using RDI and SPI. The lines of conditional return periods of drought duration given severity at 90th percentiles for all the regions and severity given duration for region 4 have very highly deviated graphs with higher return periods. The drought durations have much higher conditional return periods compared to severity, particularly for regions 3 and 5. The graph pattern in these two regions for conditional return periods of duration is not smoothly increasing but there are abrupt changes after 30 months duration. All these results show that regions 3 and 5 have fewer chances of drought compared to other regions. Both types of conditional return periods show that region 4 has the most frequent drought compared to other regions. The figures show greater similarity between RDI and SPI lines at smaller percentiles, but changes increase with an increase in percentiles. Rainfall is considered the main source of water depict or drought in a region(s). Prolonged drought with maximum variability and recurrently happening strong rainfall would be the key characteristics of site climatic sensitivity, whereas consistent rainfall and small droughts are characteristics of suitable climatic settings (Halwatura et al., 2015).

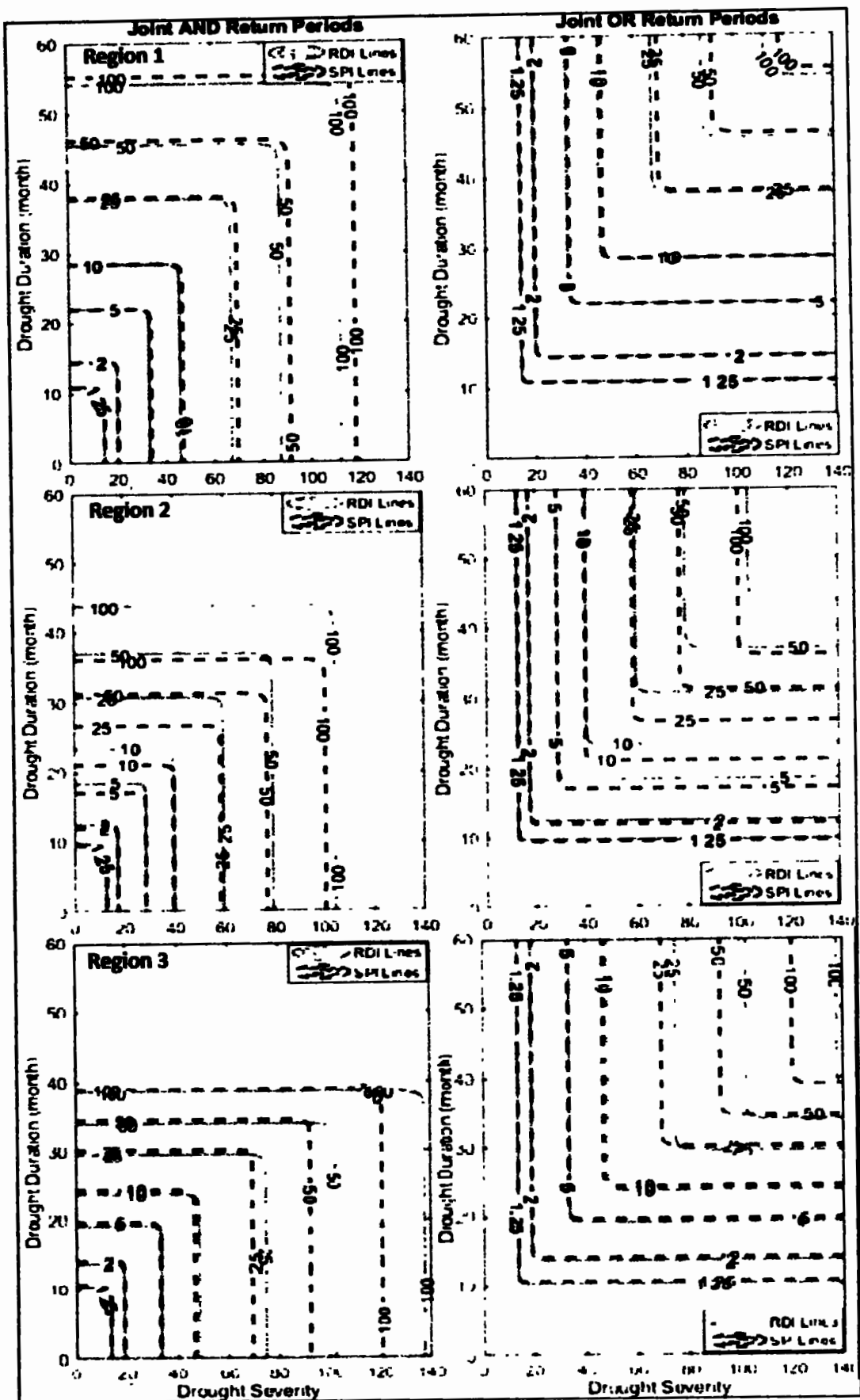


Figure 4. 2: SDF curves of T_{OR} and T_{AND} return periods for the five regions.

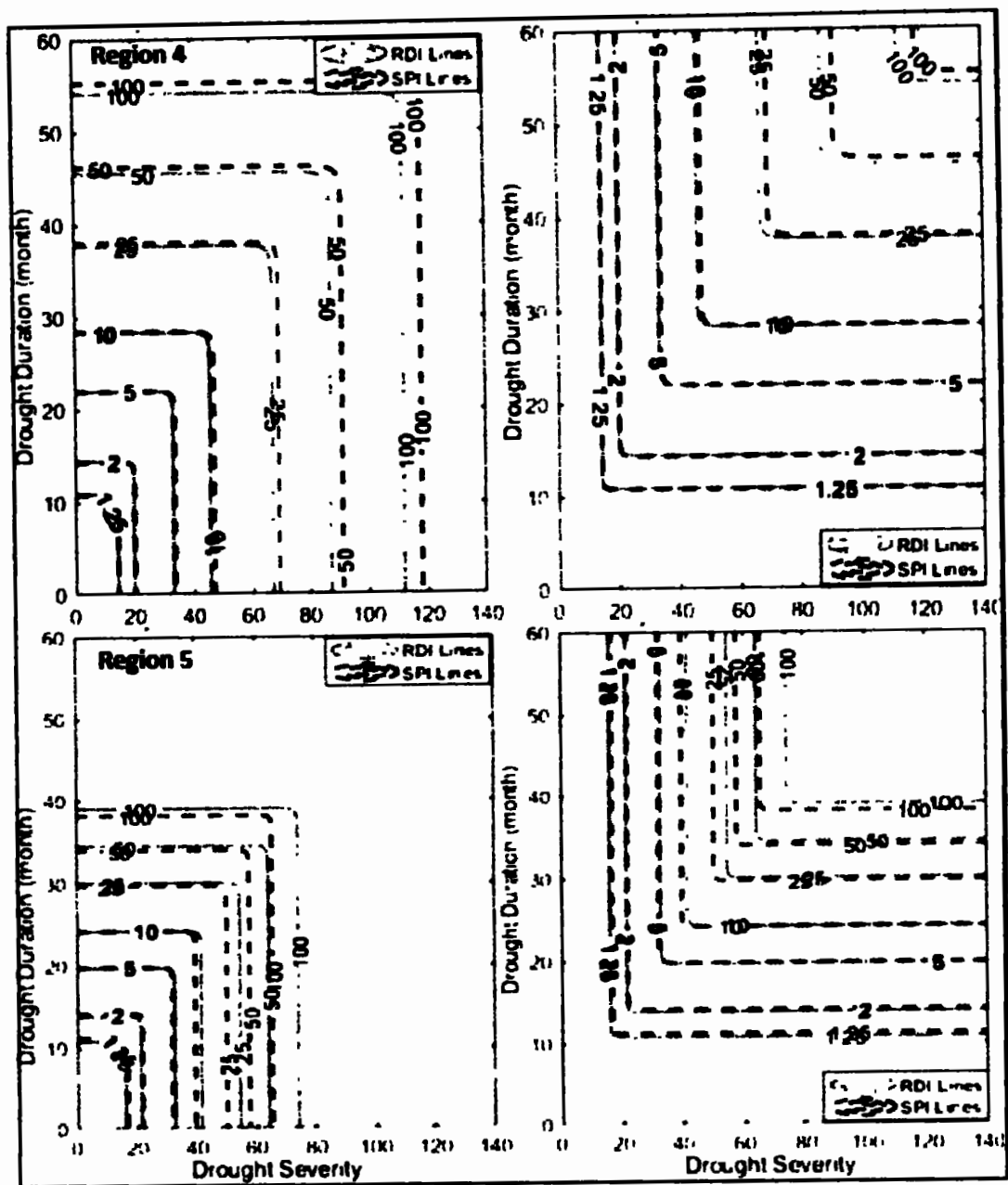


Figure 4. 2: SDF curves of T_{OR} and T_{AND} return periods for the five regions.

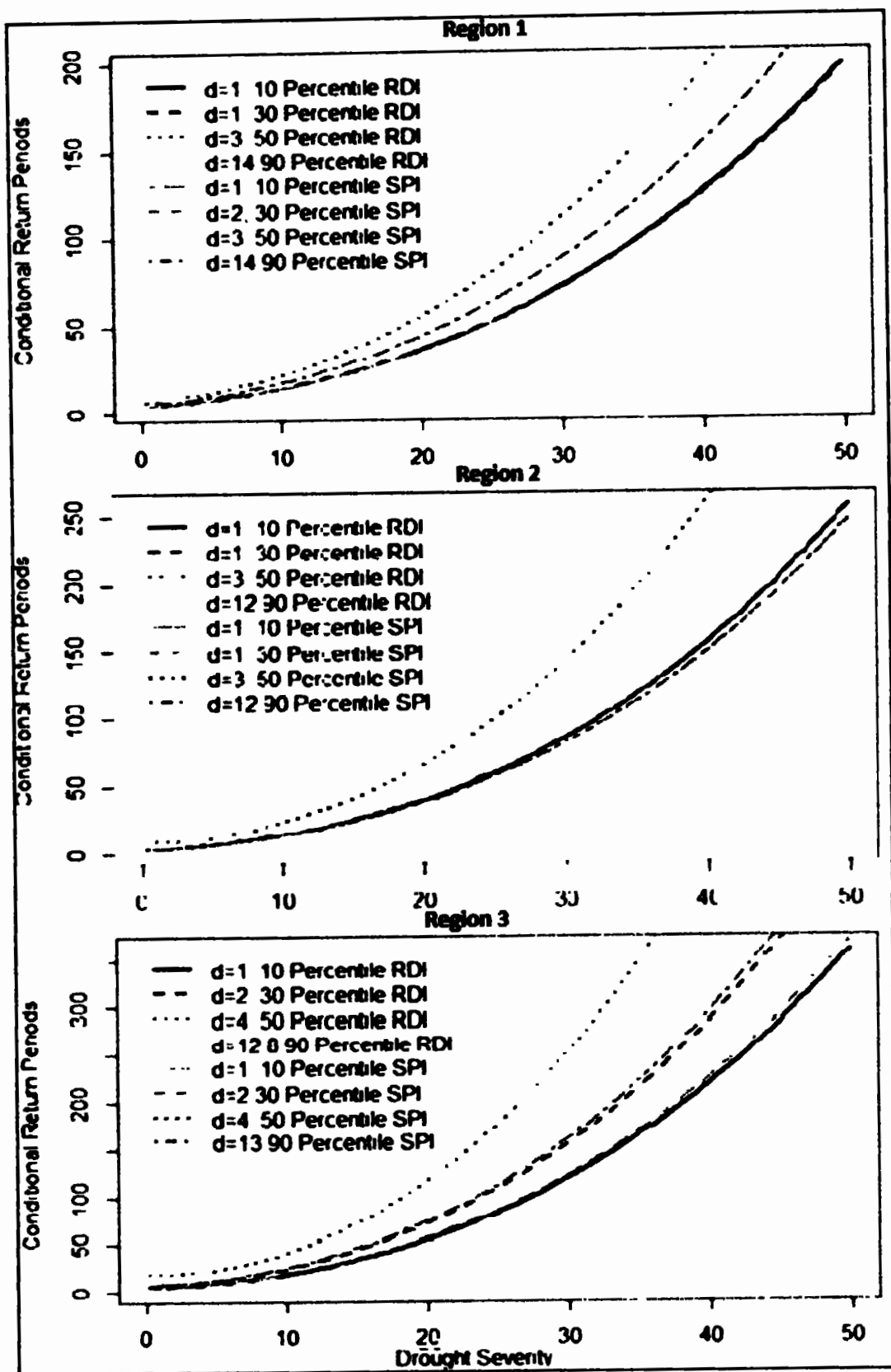


Figure 4. 3: SDF curves of conditional return periods of severity given at 10, 30, 50, and 90 percentile thresholds of durations (months), respectively.

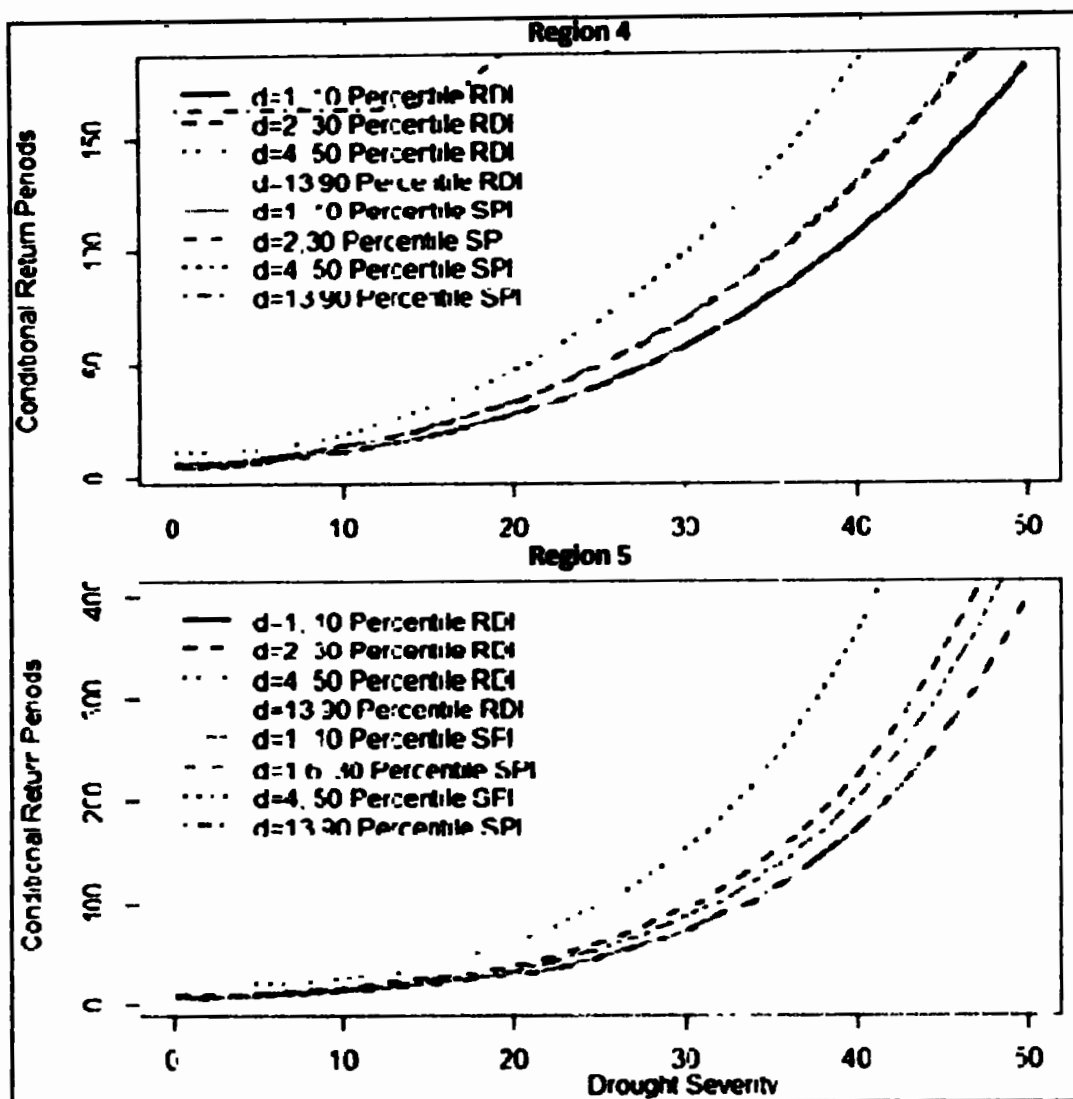


Figure 4. 3: SDF curves of conditional return periods of severity given at 10, 30, 50, and 90 percentile thresholds of durations (months), respectively.

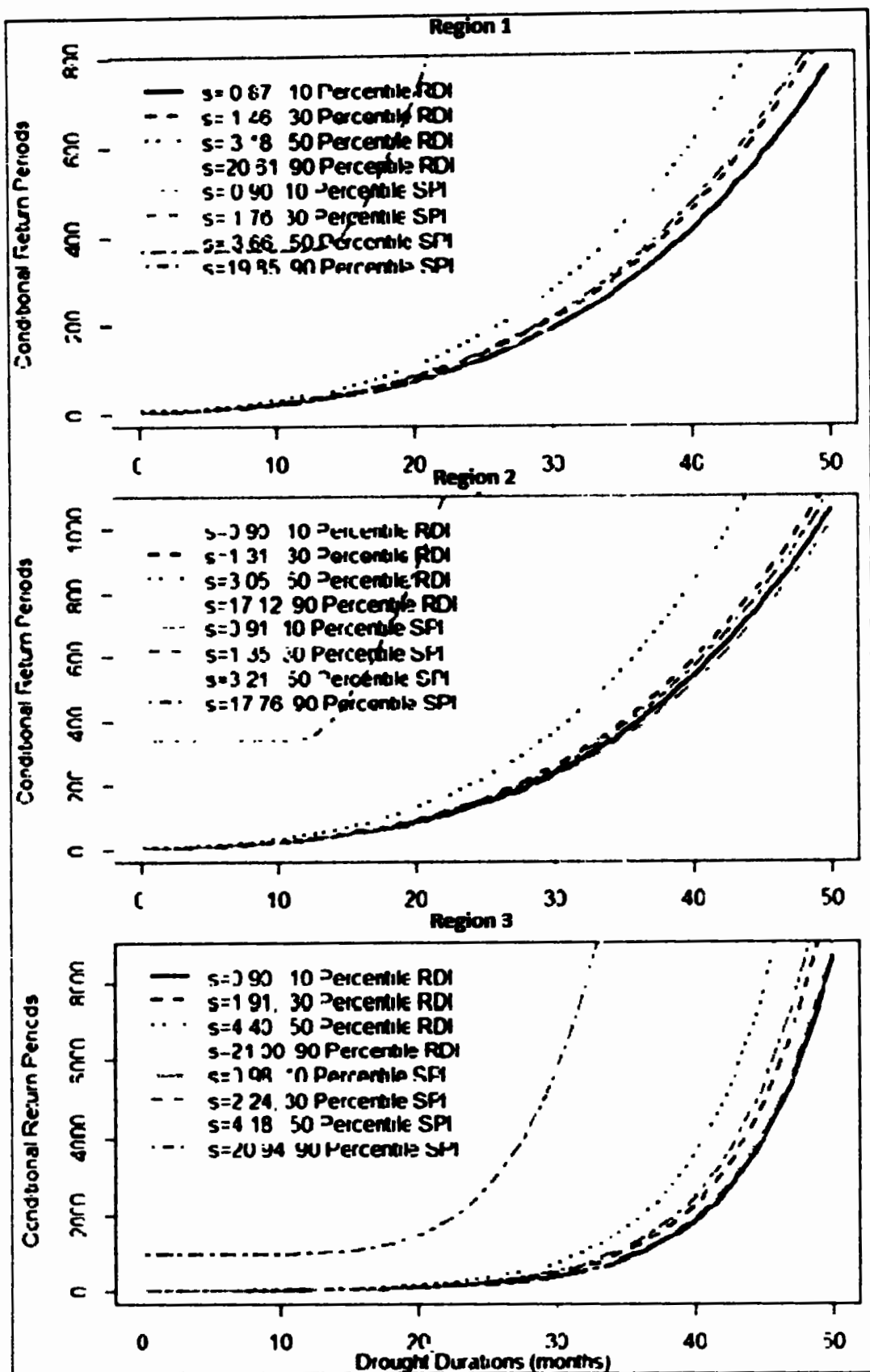


Figure 4. 4: SDF curves of conditional return periods of durations (months) given at 10, 30, 50, and 90 percentile severity thresholds, respectively.

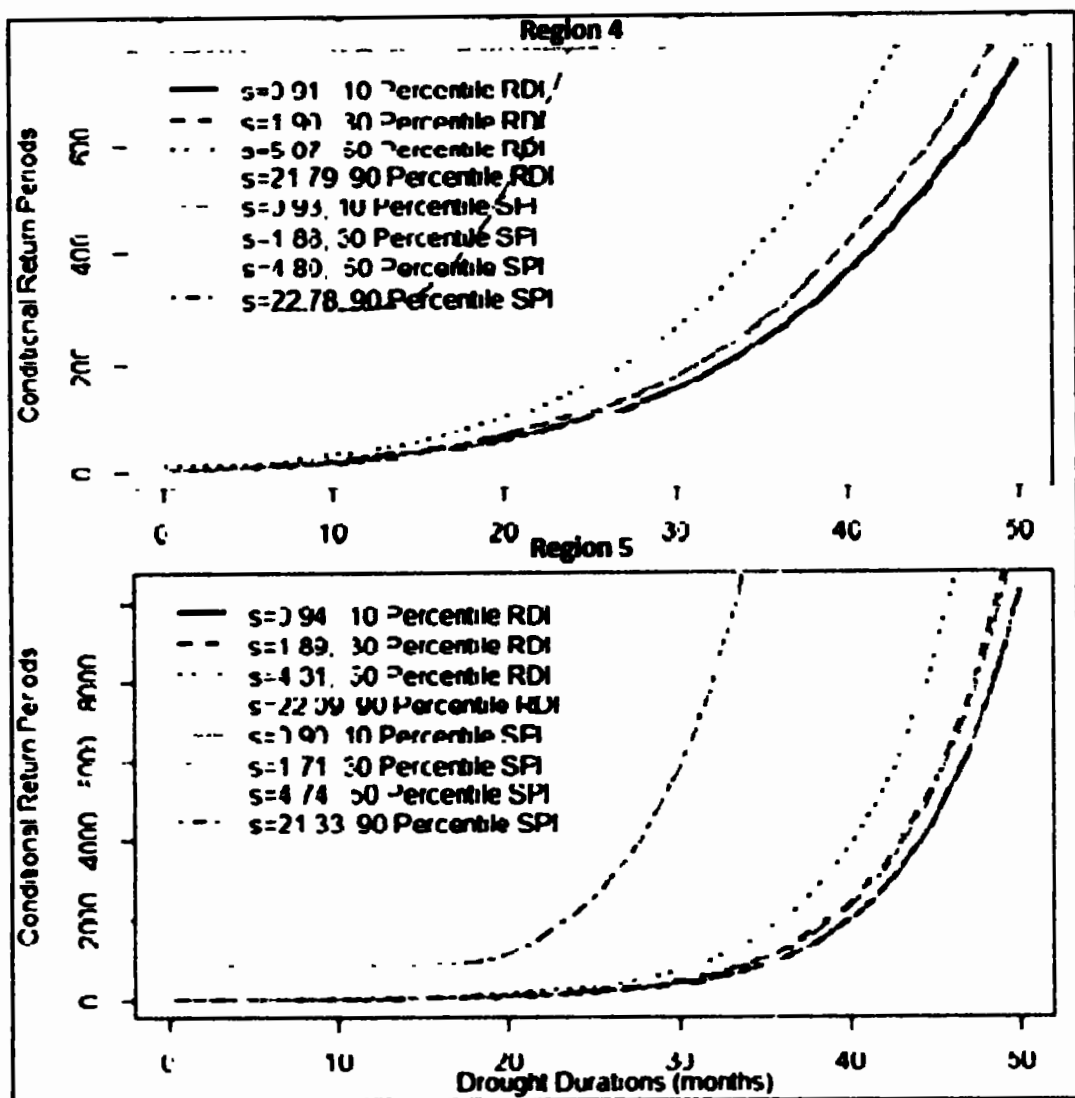


Figure 4. 4: SDF curves of conditional return periods of durations (months) given at 10, 30, 50, and 90 percentile severity thresholds, respectively.

4.4 Conclusion

The drought condition is rapidly improving particularly in the south-west of Pakistan. The values of drought duration and severity are extracted from RDI and SPI series using -0.85 threshold level. These values are used and are calculated L-moments for the construction of five BHCRs using the IFP procedure. The stations in BHCRs have maximum geographical attachment compared to HCRs. Drought assessment based on drought SDF curves is an authentic tool for optimum and reliable planning of water and drought mitigation in the world. Therefore, we used drought SDF curves and calculated

drought projections based on drought duration and severity within the BHCRs at 1.25, 2, 5, 10, 25, 50, and 100 years of return periods.

Statistical methods are used, and best-fit regional probability distributions are selected for drought duration and severity variables. For joint distribution, Clayton, G-H, and Galambos copula models are checked, and G-H model is selected best fit which is important for considering extreme drought risk assessment. The primary and secondary return periods are estimated for the joint nature at selected years and graphed using contour lines. Similarly, conditional return periods have high values compared to joint return periods because mostly drought with such a high severity (or duration) and fixed duration (or severity) takes a long time to occur. The selected return periods show that drought duration and severity increase with the increase in return periods. The RDI and SPI indices have similarities at lower and dissimilarities at higher return periods.

A region with smaller return periods of drought is expected to repeat soon. SDF curves of joint return periods are checked on selected return periods while conditional return periods on selected percentiles of drought duration (or severity). The regions have significant changes, particularly at higher return periods. The drought projections show that regions 1, 3, and 4 have more, region 2 with moderate, and region 5 with fewer chances of droughts repeating. The T_{OR} drought return periods are always less than T_{AND} return periods while T_{KEN} lies in between. Similarly, the conditional return periods have very high values compared to joint return periods, particularly for regions 3 and 5 with abrupt changes after 30 months. Joint and conditional results show that region 3 has mixed projections while region 4 has a frequent chance of drought repeating soon.

Chapter 5

Proposed Standardized Copula-based Drought Index (SCDI)

5.1 Background

In previous chapters, we used SPI and RDI indices for drought risk assessment and forecasting in Pakistan. SPI and RDI are among several other standardized drought indices like SPEI and SPTI with one or more climate data sets i.e., precipitation, temperature, and PET. But these indices have some limitations (Mishra and Singh, 2010; Zargar et al., 2011). For example, SPI is based only on precipitation data which is insufficient and explains limited drought variability (Vicente-Serrano et al., 2010; Lee et al., 2017). SPEI and RDI are based on precipitation and PET using Thornthwaite equation. However, PET is underestimated in arid and semiarid areas while overestimated in humid and semi-humid areas using Thornthwaite equation (Jensen et al., 1990; Van der Schrier et al., 2011). Hence this method is not suggested to be used in low temperature areas (Papadopoulou et al., 2003). There are several other methods of calculating PET such as Blaney-Criddle, Hargreaves and Penman-Monteith, which are based on multiple data sets (Vangelis et al., 2013). Therefore, Thornthwaite equation is the only method that can easily be used in most areas. Ultimately, RDI and SPEI may produce misleading results in such conditions. SPTI is based on precipitation and temperature data which can be used in low temperature areas while SPEI is not suggested for low temperature areas (Ali et al., 2017).

Similarly, drought occurs due to a persistent shortage of precipitation over a region for a definite period (Correia et al., 1994; Wilhite, 2004). Aridity increases and soil

moisture decreases due to high temperature through maximum evaporation from water resources. Rainfall and temperature directly or indirectly influence PET and the natural environment for water resources management (Hounnou and Dedehouanou, 2018). PET has significant role in hydrologic cycle, to be used in drought modeling (Zarch et al., 2015). Global warming and temperature intensify drought severity (Zhao and Dai 2015; Hui-Mean et al., 2018) and possibly increase evapotranspiration more than precipitation (Trenberth et al., 2014; Vicente-Serrano et al., 2014).

To overcome or minimize, the above limitations and problems, we propose a new multi-scaler Standardized Copula-based Drought Index (SCDI) which is based on both temperature and PET along with precipitation and can be used in arid as well as humid regions with multiple time scales if required. The copula models are multivariate statistical functions that may solve the problem of dependence structure within correlated variables and jointly simulate the drought variables with different probability distributions more objectively (Salvadori and De Michele, 2010; Khan et al., 2020; Das et al., 2020; Ullah and Akbar 2021). Copula models have also been used in the construction of drought indices (Kao and Govindaraju, 2010; Kavianpour et al., 2018; Won et al., 2020). Recently, we have used multidimensional copula modeling for constructing micronutrients index at household level in Pakistan (Amjad et al., 2022).

Niemeyer (2008) gives some drought categories including comprehensive and combined drought indices. Comprehensive drought indices use more than one type of climate data while combined drought indices are constructed by joining existing drought indices to develop a new drought index. The objective of this chapter is to work on the development of a new Standardized Copula-based Drought Index (SCDI). **Firstly**, copula model is used to combine the information of existing UAI and DAI

indices which are constructed using data of three climatic variables (precipitation, temperature, and PET) to explain maximum variation. **Secondly**, to validate the results of SCDI, it is compared with existing SPI and RDI using nine metrological stations from Baluchistan province, given in **Appendix-A**. The data of precipitation and temperature are used taken from PMD whereas PET data has been estimated using Thornthwaite equation. Baluchistan is categorized by the mixed climate where the including stations contain climates from hyper-arid to humid. The newly constructed SCDI is equally important for meteorological, agricultural, and hydrological purposes. Therefore, it may be useful for academia as well as for practitioners to get maximum information about drought risk assessment due to maximum climate data.

5.2 Methodology

5.2.1 Aridity Drought Indices

Aridity is the reverse of humidity, which is the degree of lack of effective and life-promoting moisture (American Meteorological Society, 2006). The aridity indices can be calculated for any specific location/station. For the construction of the proposed drought index, we use two aridity indices. i.e., UNEP Aridity Index (UAI) and De-Martone Aridity index (DAI).

The United Nations Environmental Programme (UNEP) aridity index is proposed by UNEP (1993). It is calculated using the ratio between precipitation (P_t) and potential evapotranspiration (PET_t) and can be calculated for various time scales, as follow:

$$(UAI)_k = \sum_{i=1}^k P_i / \sum_{i=1}^k PET_i, \text{ for } k = 1, 2, 3, \dots, 12. \quad (5.1)$$

The DAI index was proposed by De-Martonne (1925), which is used to measure the dryness of an area. It is calculated as follows:

$$(DAI)_k = \left(12/k\right) \sum_{i=1}^k P_i / (\bar{T}_k + 10), \text{ for } k = 1, 2, 3, \dots, 12. \quad (5.2)$$

Where P_i denotes the total precipitation (mm) and $\bar{T}_k$ is the average temperature ($^{\circ}\text{C}$) at the k^{th} time scale. The index can be calculated for any month/season or year, for which monthly values can be considered (Hrnjak et al., 2013). The index values decrease i.e., approach zero when the temperature increases.

5.2.2 Copula Modeling for SCDI

The procedure to select the most suitable probability distribution and copula model is given in detail in chapter 4 (see sub-section 4.4). Several possible probability distributions can be fitted to a set of data. The best-fitted marginal probability distributions have a key role in the estimation of reliable projections of drought risk in a region (Saf, 2010). The one and two-parameter distributions do not capture the tail effect of extreme events properly (Hosking and Wallis, 1997; Chen and Guo, 2019). Hence several studies used three-parameter probability distributions for the drought characteristics (e.g., Azam et al., 2018; Mortuza et al., 2019). However, exponential, gamma and other two-parameter distributions are efficiently used for drought duration and severity (e.g., Shiau and Modarres 2009; Mirabbasi et al., 2012; Halwatura et al., 2015). Therefore, in this study, two-parameter Gamma, Logistic, and Weibull distributions along with three-parameter Generalized Extreme value (GEV) and log-normal distribution with three parameters (LN3) are used to select the most suitable probability distributions for DAI and UAI variables, given in **Appendix-B**. The graphical method of fitting the cumulative probability functions (CDF) of the five probability distributions along with the empirical distribution function (EDF) is used. A distribution is considered better if it is closer to the EDF of the data. As a numerical measure, the Shapiro-Wilk goodness of fit test (Shapiro and Wilk, 1965) is used under

statistical hypothesis testing, whether the candidate probability distribution belongs to a normally distributed population?

$$W = \left(\sum_{i=1}^n b_i y_{(i)} \right)^2 / \sum_{i=1}^n (y_i - \bar{y})^2 \quad (5.3)$$

Where $y_{(i)}$ is the i th value of order statistics of y variable and b_i is the vector of tabulated coefficients. The candidate probability distribution will be rejected if the calculated value of Shapiro-Wilk test statistics is small. This value can be transformed to the standard normal distribution (z-test) based on lognormal distribution to calculate the p-value. A distribution is considered the most suitable distribution if it has a minimum value of z-test statistics with a maximum p-value.

In the second step, the copula models are discussed. The copula model was described by Sklar (1959). It is used to calculate the combined effect of UAI and DAI aridity indices. The selected most suitable marginal distributions are used to calculate the cumulative distribution function of copula models $[C(.)]$ for any family like Elliptical, Archimedean, and Extreme Value copulas. In this study, four single parameter Gaussian, Frank, Gumbel-Hougaard (G-H), and Galambos copula models are used. The Gaussian, Frank, and Galambos copula models belong to the bivariate Elliptical, Archimedean, and Extreme Value families, respectively while the G-H copula is the only function that is related to both the Archimedean and Extreme Value families. These copula models are selected to represent the three copula families. The previously discussed inversion of Kendall's τ method will be used to estimate the copula parameters for the selected copula models that use the relationship between Kendall's tau and copula parameter (θ) (Nelsen 2006; Genest and Favre 2007). The Cramer-Von Mises goodness of fit test (Cramer, 1928; Von Mises, 1928) is used to select the best-

fit copula model. The null hypothesis for the test is that the candidate copula model satisfies the normality conditions and is defined as follows:

$$S_n = \frac{1}{12n} + \sum_{i=1}^n \left(p_{(i)} - \frac{2 * i - 1}{2n} \right)^2 \quad (5.4)$$

The term $p_{(i)} = \phi \left(\frac{y_{(i)} - \bar{y}}{s} \right)$ where $y_{(i)}$ is the value of data with mean $\bar{y}$ and standard deviation of s while ϕ denotes the CDF of standard normal distribution. For the p-value, the S_n values are transformed as $Z = S_n \left(1.0 + 0.75/n \right)$. A copula model is considered the best fit if it has a minimum value of S_n and Z test statistics with a maximum p-value.

Finally, the best-fit copula model is used to find the cumulative probabilities and is transformed to get standardized values of the SCDI index to be classified according to the categories. For standardization the technique of Abramowitz and Stegun, (1948) is used as follow:

$$SCDI = - \left(Q + \frac{a_0 + a_1 Q + a_2 Q^2}{1 + b_1 Q + b_2 Q^2 + b_3 Q^3} \right) \quad (5.5)$$

and

$$Q = \sqrt{\ln \left(\frac{1}{(C(.))^2} \right)}, \text{ where } 0 < C(.) \leq 0.5 \quad (5.6)$$

Similarly

$$SCDI = + \left(Q + \frac{a_0 + a_1 Q + a_2 Q^2}{1 + b_1 Q + b_2 Q^2 + b_3 Q^3} \right) \quad (5.7)$$

and

$$Q = \sqrt{\ln \left(\frac{1}{(1 - C(.))^2} \right)}, \text{ where } 0.5 < C(.) \leq 1 \quad (5.8)$$

Where the given values of constants are $a_0 = 2.515517$, $a_1 = 0.802853$, and $a_2 = 0.010328$ while $b_1 = 1.432788$, $b_2 = 0.985269$, and $b_3 = 0.001308$, and $C(.)$ represent the cumulative probabilities of copula model for the station, respectively. This standardization technique is already utilized by McKee et al., (1993); Tsakiris and Vangelis, (2005); Vicente-Serrano et al., (2010); Ali et al., (2017).

5.3 Results

5.3.1 Construction of SCDI Drought Index

In the construction of SCDI, precipitation, temperature, and PET data are used. Precipitation and temperature data are observed records from the climate while PET is measured using temperature data and latitude of the station based on the Thornthwaite equation. However, PET is not dependent on any single variable such as temperature but is rather based on several other climatic conditions of the area. To check the structure of temperature and PET data, two stations are selected from Balochistan province, first is Quetta which is considered humid and cold while the next is Noukundi which is hyper-arid. The comparison is made at 3-, 6-, 9-, and 12-months' time scales. The graph clearly shows that the movement of temperature and PET is not similar but there are significant changes in the overall pattern of the records of both climate data given in Fig. (5.1). This means that both the variables have their importance and role in the climatic assessment of any area. The effect of PET is different in different parts of the regions. It has smaller values over the cold station of Quetta which shares less part in drought while in the arid station of Noukundi it is very high with greater variabilities. Therefore, it is better to use both temperature and PET data to develop such a drought index which is based on multi types of climate data to explain maximum variability to measure drought conditions more accurately in any part of the world.

To completely validate the results of the SCDI, nine stations in Baluchistan province are selected. The selected stations include humid stations as well as arid stations according to the climatic conditions. **In the first stage**, the three variables are transformed using equations (5.1) and (5.2) to find UAI and DAI indices of interest. These indices can be calculated for any time scale. However, in this study, 3-, 6-, 9- and 12-months' time scales are used to check the results of the SCDI. For 3-months' time scale, the values of precipitation, temperature, and PET are added for any three consecutive months and are used to calculate DAI and UA. The same idea of accumulative values is repeated for 6-, 9-, and 12-months' time scales.

In the second stage, best-fit probability distributions are obtained for UAI and DAI variables by plotting CDFs of the five probability distributions along with the EDF of the transformed variables given only for the Quetta station (**Fig. (5.2)**) to save space. CDF of distributions has approximately similar and better fitting with the EDF of data. For a more robust selection of probability distribution, the Shapiro-Wilk goodness of fit test is used, given in **Table (5.1)**. According to the criterion of goodness of fit test, the small values show that the null hypothesis of normality is probably not suitable. To further investigate, the Shapiro-Wilk test results are transformed to calculate the z- and p-values. The smaller z-values with greater p-values are considered the most acceptable probability distribution. The results show that GEV is the most acceptable probability distribution to be used for both variables in the construction of the SCDI index.

Table 5. 1: Results of Shapiro-Wilk goodness of fit test, z-, and p-values for the probability distributions (W-test, z-value & p-value).

Stations	UNEP Aridity Index (UAI)					De-Martone Aridity Index (DAI)				
	Gamma	Logistic	Weibull	GEV	LN3	Gamma	Logistic	Weibull	GEV	LN3
Barkhan	0.931	0.964	0.97	0.975		0.939	0.973	0.979	0.985	
	2.396	1.034	0.66	0.282	**	2.128	0.429	-0.178	-0.849	**
	0.008	0.15	0.255	0.389		0.017	0.334	0.571	0.802	
Dalbandin	0.960	0.948	0.978	0.984	- 0.984	0.955	0.945	0.973	0.981	0.982
	1.806	2.409	0.534	0.099	0.366	2.062	2.510	0.960	0.197	0.722
	0.035	0.008	0.297	0.540	0.357	0.020	0.006	0.168	0.422	0.235
Jiwani	0.980	0.934	0.987	0.977	0.981	0.979	0.933	0.986	0.976	0.980
	0.257	2.820	-0.721	0.514	0.730	0.319	2.858	-0.608	0.619	0.847
	0.399	0.002	0.765	0.304	0.233	0.375	0.002	0.728	0.268	0.199
Kalat	0.895	0.922	0.908	0.976	0.973	0.895	0.921	0.907	0.976	0.972
	3.903	3.272	3.627	0.688	1.600	3.919	3.302	3.647	0.737	1.656
	0.000	0.001	0.000	0.246	0.055	0.000	0.000	0.000	0.231	0.049
Nokkundi	0.979	0.882	0.976	0.969	0.977	0.978	0.875	0.976	0.972	0.978
	0.124	3.835	0.410	0.973	0.947	0.206	3.958	0.457	0.789	0.770
	0.451	0.000	0.341	0.165	0.172	0.418	0.000	0.324	0.215	0.221

Table 5. 1: Results of Shapiro-Wilk goodness of fit test, z-, and p-values for the probability distributions (W-test, z-value & p-value).

Stations	UNEP Aridity Index (UAI)				De-Martone Aridity Index (DAI)			
	Gamma	Logistic	Weibull	LN3	Gamma	Logistic	Weibull	LN3
Ormara	0.927	0.717	0.939	0.976	0.945	0.774	0.948	0.982 - 0.977
	2.821	5.710	2.434	0.997	2.188	5.234	2.088	0.154 0.909
	0.002	0.000	0.007	0.159	0.014	0.000	0.018	0.561 0.182
Quetta	0.951	0.936	0.902	0.964	0.949	0.936	0.898	0.967 0.963
	2.28	2.827	3.762	2.285	2.335	2.822	3.852	1.387 2.347
	0.011	0.002	0.000	0.011	0.010	0.002	0.000	0.083 0.009
Sibbi	0.945	0.959	0.968	0.987 -	0.916	0.972	0.952	0.986 - 0.985
	2.516	1.855	1.370	0.037	3.418	1.042	2.234	0.437 0.229
	0.006	0.032	0.085	0.515	0.000	0.149	0.013	0.669 0.409
Zhob	0.982	0.978	0.983	0.986 -	0.989	0.981	0.986	0.992 0.991
	-0.251	0.279	-0.312	0.197	-1.173	-0.083	-0.797	-1.844 -1.124
	0.599	0.390	0.622	0.578	0.880	0.533	0.787	0.967 0.869

*Note: ** denote that "The sample skew is not positive. Admissible moment estimates do not exist" while bold values denote best fit distributions.*

In the third stage, the Gaussian, Frank, G-H, and Galambos copula models are used to show the joint relationship between UAI and DAI variables. For fitting a copula model, it is required to transform these two variables into uniformly distributed over the interval (0,1). The previously selected GEV distribution is used to calculate the cumulative probabilities at each point of the variables. Copula parameters are estimated using the inversion of Kendall's τ method, given in Table (5.2). The Cramer-Von Mises (S_n) goodness of fit test is used to select the best fit copula model. A copula model with a smaller S_n value and larger p-values is considered the best fit. According to the results in Table (5.2), different copula models are suitable for different stations in the study area. However, for Nokkundi station both Gaussian and Galambos copula models are best-fit and anyone can be used for the results. This means that no specific but multiple copula models may be checked for a suitable selection to calculate this drought index. Then selected copula model is used to combine the uniformly distributed values of CDFs through the GEV probability distributions for the variables UAI and DAI.

Table 5. 2: Results of Cramer Van Mises (S_n) goodness of fit test (parameter estimates/ S_n -test/p-value).

Stations	Gaussian Copula	Gumbel Copula	Frank Copula	Galambos Copula
Barkhan	0.97/0.03/0.06	6.14/0.03/0.01	31.77/0.02/0.609	5.39/0.01/0.23
Dalbandin	0.99/0.01/0.70	7.37/0.02/0.08	36.34/0.01/0.65	6.64/0.01/0.22
Jiwani	0.99/0.01/0.61	29.18/0.01/0.84	107.5/0.01/0.75	28.46/0.01/0.74
Kalat	0.99/0.01/0.95	11.40/0.03/0.07	109.3/0.01/0.90	6.07/0.06/0.02
Nokkundi	0.99/0.01/0.79	17.41/0.01/0.60	69.80/0.01/0.71	16.66/0.01/0.79
Ormara	0.99/0.02/0.07	9.72/0.03/0.01	45.71/0.02/0.17	8.97/0.01/0.73
Quetta	0.99/0.01/0.82	15.36/0.01/0.57	63.31/0.01/0.57	14.64/0.01/0.76
Sibbi	0.99/0.02/0.31	8.13/0.02/0.09	36.31/0.02/0.22	7.42/0.01/0.54
Zhob	0.99/0.02/0.55	13.27/0.02/0.43	56.19/0.02/0.50	12.55/0.01/0.81

Note: The bold values indicate the best-fit copula model for the stations.

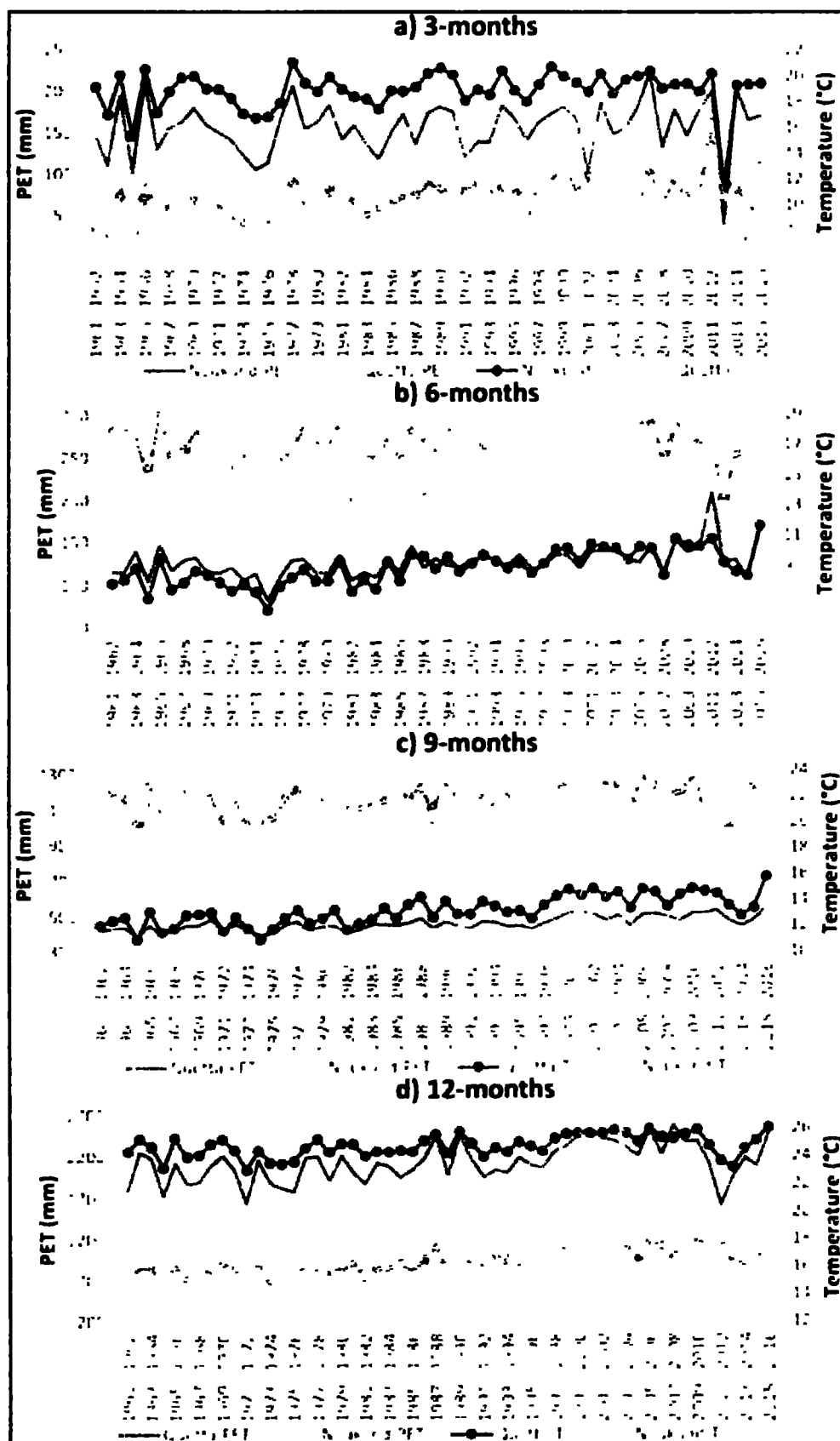


Figure 5. 1: Comparison of temperature (T) and PET at 3-, 6-, 9-, and 12-months' time scales for Noukundi and Quetta stations.

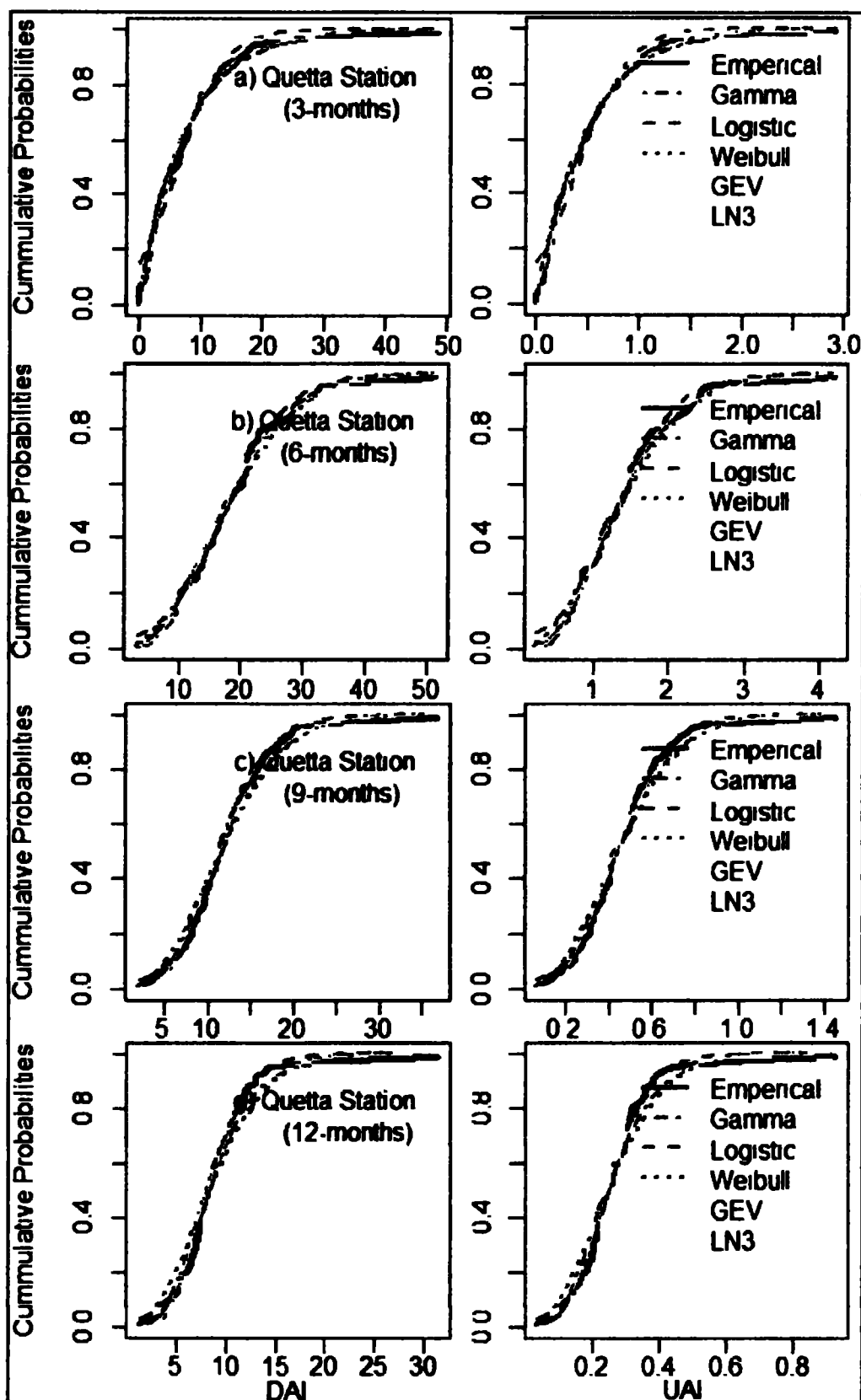


Figure 5. 2: Comparison of CDFs of selected probability distributions at 3-, 6-, 9- and 12-months' time scales, respectively with the EDF of data for Quetta station, only.

In the final stage, a joint series of CDF values for the copula model is obtained and standardized using equations (5.4) and (5.6). These standardized values are classified according to the levels given in Table (5.3), which are the required levels for the SCDI index to be used for drought risk assessment anywhere in the world. The index can be obtained for any time scale as is used in this study for 3-, 6-, 9- and 12-months' time scales.

Table 5. 3: Classification of standardized RDI, SCDI, and SPI values, where DI denotes drought index.

SPI, RDI & SCDI value	Category
$2.0 \leq DI$	Extremely wet
$1.5 \leq DI < 2.0$	Severely wet
$1.0 \leq DI < 1.5$	Moderately wet
$-1.0 < DI < 1.0$	Near normal
$-1.5 < DI \leq -1.0$	Moderately drought
$-2.0 < DI \leq -1.5$	Severely drought
$DI \leq -2.0$	Extremely drought

5.3.2 Comparison of SCDI with SPI and RDI

In this section, the SCDI is compared with the standardized forms of RDI and SPI. Both drought indices are also based on proper probability distribution. The drought series are calculated using RDI, SCDI, and SPI indices at 3-, 6-, 9-, and 12-month time scales for Nokkundi and Quetta stations and graphed given in Fig. (5.4). The three drought indices have a very good approximation of each other and validate the results of the proposed SCDI. The graph also shows changes between SCDI and RDI which means that temperature with PET has a significant effect on drought assessment in any region particularly at arid stations like Nokkundi where the changes in both are more highlighted shown in Fig. (5.2).

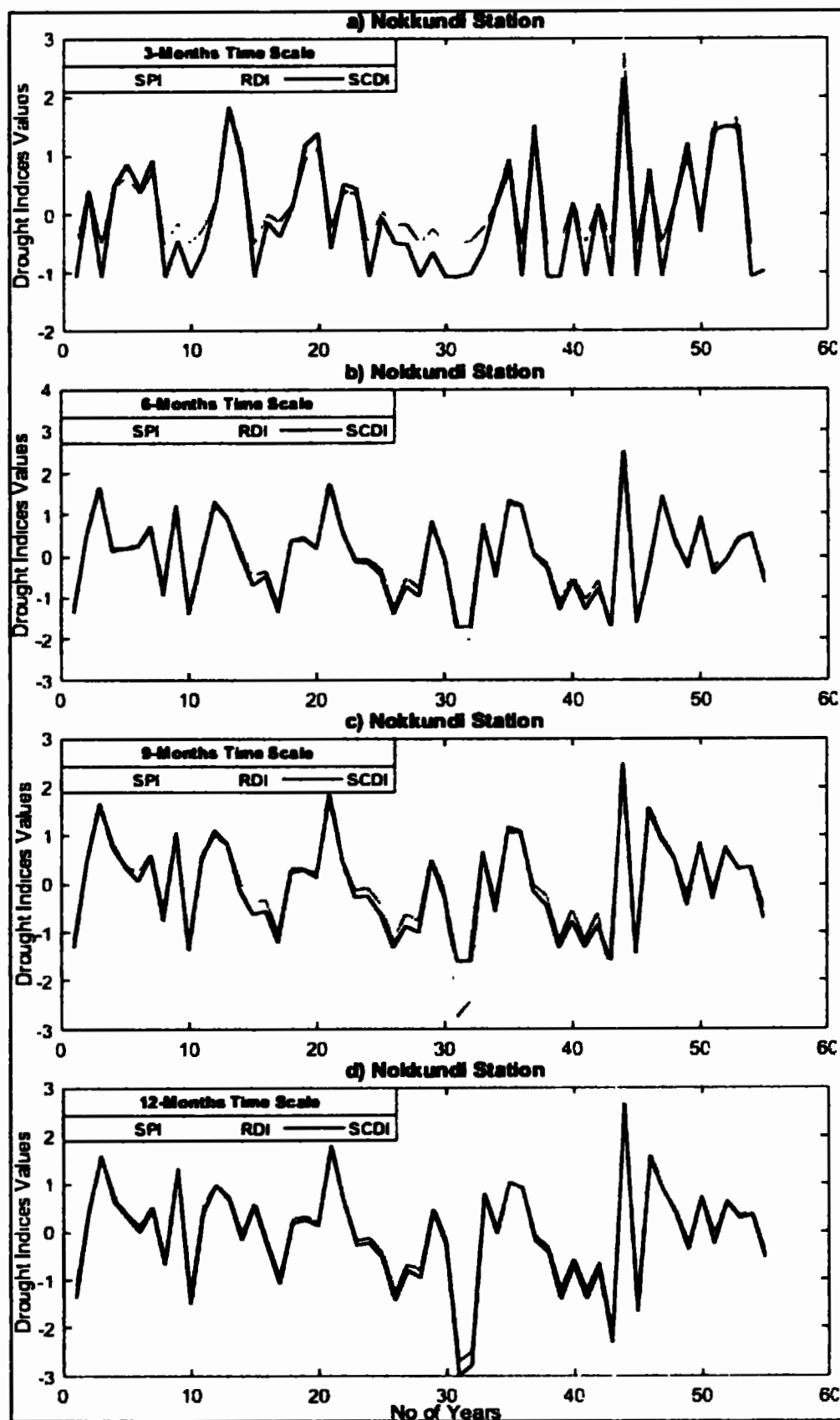


Figure 5.3: Comparison of RDI and SPI indices with the proposed SCDI for 3-, 6-, 9- and 12-months' time scales using Nokkundi and Quetta meteorological stations.

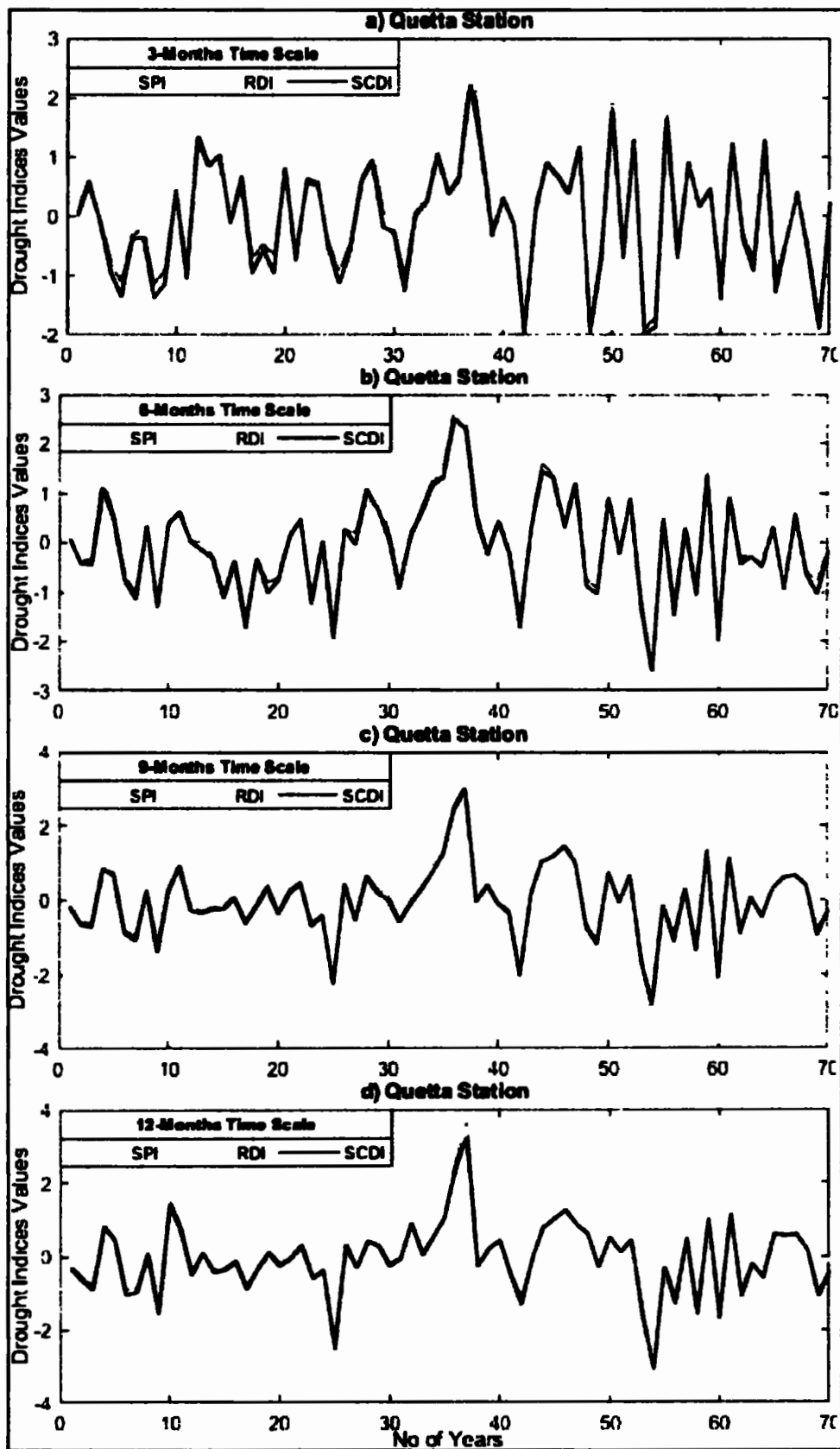


Figure 5. 4: Comparison of RDI and SPI indices with the proposed SCDI for 3-, 6-, 9- and 12-months' time scales using Nokkundi and Quetta meteorological stations.

The strength of the relationship between SCDI with SPI and RDI series is compared for four-time scales. In Table (5.4), values show a very strong correlation coefficient for the stations which indicates that SCDI is highly compatible with SPI and RDI. Furthermore, the percentages of drought explained by each category are calculated for SCDI, SPI, and RDI, given in Table (5.5). In Barkhan station, SCDI has larger percentages for 9- and 12 months' time scales in mild drought category, for 3-, 9-, and 12-months in moderate drought category, for 6-months in severe drought category, and equal values in maximum remaining time scales and categories for the three indices. Similarly, percentages for other stations are given which indicate that in maximum cases SCDI explained maximum drought variability or has equal values compared to SPI and RDI. The wet percentages also have similar values in each category that are not included. Hence, it indicates the importance of SCDI in drought risk assessment.

Table 5. 4: Correlation coefficients of SCDI with SPI and RDI at 3-, 6-, 9-, and 12-months' time scales.

Stations	SPI time scales				RDI time scales			
	3	6	9	12	3	6	9	12
Barkhan	0.958	0.982	0.992	0.983	0.958	0.994	0.994	0.987
Dalbandin	0.880	0.992	0.978	0.974	0.884	0.996	0.982	0.981
Jiwani	0.993	0.992	0.998	0.998	0.992	0.994	0.998	0.998
Kalat	0.913	0.990	0.992	0.968	0.909	0.985	0.993	0.969
Nokkundi	0.966	0.986	0.967	0.999	0.951	0.988	0.971	0.999
Ormara	0.930	0.990	0.967	0.980	0.918	0.988	0.952	0.987
Quetta	0.984	0.980	0.988	0.989	0.990	0.994	0.998	0.997
Sibbi	0.964	0.993	0.994	0.991	0.965	0.994	0.996	0.995
Zhob	0.990	0.992	0.991	0.992	0.988	0.998	0.997	0.996

Table 5. 5: Percentages of drought explained in four drought categories by three indices (SCDI/SPI/RDI) using 3-, 6-, 9-, and 12-months' time scales for the selected stations of Baluchistan.

Stations	MTS	Mild Drought	Mod. Drought	Sev. Drought	Ext. Drought
Barkhan	3-MTS	23.40/48.94/48.94	23.40/0.00/0.00	0.00/0.00/0.00	0.00/0.00/0.00
	6-MTS	34.04/34.04/38.30	12.77/14.89/8.51	4.26/0.00/0.00	2.13/2.13/4.26
	9-MTS	42.55/34.04/34.04	8.51/6.38/8.51	2.13/2.13/2.13	4.26/4.26/4.26
	12-MTS	38.31/31.91/31.91	12.77/10.64/8.51	0.00/0.00/0.00	4.26/4.26/4.26
Dalbandin	3-MTS	16.47/50.59/51.76	3.53/0.00/0.00	16.47/0.00/0.00	0.00/0.00/0.00
	6-MTS	34.12/36.47/37.65	11.76/8.24/7.06	5.88/4.71/5.88	2.35/2.35/2.35
	9-MTS	32.94/29.41/27.06	10.59/8.24/10.59	7.06/3.53/3.53	2.35/3.53/3.53
	12-MTS	37.65/30.59/28.24	9.41/8.24/9.41	4.71/2.35/3.53	3.53/3.53/3.53
Jiwani	3-MTS	24.19/51.61/53.23	29.03/0.00/0.00	0.00/0.00/0.00	0.00/0.00/0.00
	6-MTS	25.81/35.48/35.48	14.52/12.90/11.29	4.84/4.84/6.45	3.23/0.00/0.00
	9-MTS	33.87/33.87/33.87	9.68/9.68/9.68	4.84/8.06/8.06	3.23/0.00/0.00
	12-MTS	29.03/30.65/30.65	12.90/9.68/9.68	6.45/6.45/6.45	1.61/1.61/1.61

Table 5. 5: Percentages of drought explained in four drought categories by three indices (SCDI/SPI/RDI) using 3-, 6-, 9-, and 12-months' time scales for the selected stations of Baluchistan.

Stations	MTS	Mild Drought	Mod. Drought	Sev. Drought	Ext. Drought
Kalat	3-MTS	20.00/50.59/49.41	27.06/0.00/0.00	0.00/0.00/0.00	0.00/0.00/0.00
	6-MTS	27.06/28.24/32.94	11.76/9.41/5.88	3.53/7.06/7.06	3.53/0.00/0.00
	9-MTS	29.41/30.59/29.41	7.06/7.06/7.06	3.53/7.06/7.06	3.53/0.00/0.00
	12-MTS	30.59/29.41/30.59	7.06/5.88/7.06	1.18/7.06/7.06	5.88/0.00/0.00
Nokkundi	3-MTS	21.82/50.91/50.91	32.73/0.00/0.00	0.00/0.00/0.00	0.00/0.00/0.00
	6-MTS	34.55/36.36/30.91	10.91/10.91/10.91	7.27/3.64/5.45	0.00/3.64/1.82
	9-MTS	30.91/34.55/30.91	14.55/10.91/12.73	5.45/0.00/0.00	0.00/5.45/5.45
	12-MTS	30.91/30.91/30.91	9.10/10.91/9.10	3.64/0.00/1.82	5.45/5.45/5.45
Ormara	3-MTS	49.09/45.45/45.45	0.00/0.00/0.00	0.00/0.00/0.00	0.00/0.00/0.00
	6-MTS	25.45/30.91/34.55	9.09/7.27/7.27	10.91/7.27/7.27	0.00/0.00/0.00
	9-MTS	21.82/36.36/40.00	9.09/7.27/7.27	7.27/5.45/5.45	5.45/0.00/0.00
	12-MTS	30.91/34.55/32.73	7.27/5.45/5.45	3.64/7.27/1.82	3.64/1.82/7.27

Table 5. 5: Percentages of drought explained in four drought categories by three indices (SCDI/SPI/RDI) using 3-, 6-, 9-, and 12-months' time scales for the selected stations of Baluchistan.

Stations	MTS	Mild Drought	Mod. Drought	Sev. Drought	Ext. Drought
Quetta	3-MTS	31.43/34.29/37.14	11.43/7.14/7.14	7.14/7.14/7.14	0.00/0.00/0.00
	6-MTS	30.00/27.14/31.43	12.86/11.43/8.57	4.29/5.71/5.71	2.86/1.43/1.43
	9-MTS	37.14/35.71/37.14	7.14/7.14/5.71	1.43/1.43/1.43	5.71/4.29/5.71
	12-MTS	35.71/42.86/35.71	7.14/8.57/10.00	5.71/1.43/1.43	2.86/2.86/2.86
Sibbi	3-MTS	60.00/49.41/49.41	0.00/0.00/0.00	0.00/0.00/0.00	0.00/0.00/0.00
	6-MTS	27.06/25.88/29.41	16.47/14.12/11.76	5.88/5.88/5.88	0.00/0.00/0.00
	9-MTS	32.94/35.29/37.65	11.76/9.41/5.88	9.41/5.88/7.06	0.00/0.00/0.00
	12-MTS	36.47/34.12/32.94	7.06/5.88/7.06	4.71/2.35/2.35	4.71/4.71/4.71
Zhob	3-MTS	32.73/34.55/36.36	21.82/20.00/18.18	0.00/0.00/0.00	0.00/0.00/0.00
	6-MTS	29.09/30.91/32.73	10.91/9.09/7.27	7.27/5.45/7.27	1.82/1.82/1.82
	9-MTS	30.91/34.55/32.73	7.27/9.09/10.91	7.27/3.64/3.64	1.82/1.82/1.82
	12-MTS	32.73/27.27/34.55	10.91/10.91/7.27	5.45/3.64/5.45	3.64/3.64/3.64

Note The bold values indicate the greater percentage of drought explained by SCDI The abbreviation MTS stands for 'months' time scale'

5.4 Conclusion

Drought indices are statistical tools that distinguish drought levels by using one or more climate variables to calculate a single quantitative value. It is a common perception that precipitation is responsible for drought occurrence in a region, but drought can be more comprehensively explained if some additional climate variables like temperature and PET can be used. The existing standardized drought indices have some limitations and need to develop a drought index that is based on maximum climate data to explain maximum drought variability. Therefore, the new multi-scaler proposed SCDI index has temperature and PET along with precipitation and can be used in arid as well as humid regions to explain maximum climate variability.

Nine stations in Baluchistan province are used to validate the results of the newly constructed SCDI which include arid and humid stations. Precipitation, temperature, and PET data sets are used to find the series of UAI and DAI at multiple time scales. Five probability distributions are checked to find the best fit distributions for UAI and DAI series. The qualitative and quantitative results show that GEV is the best fit distribution in most cases and is recommended to be used. In the next step, four copula models are used to select the best-fit copula model, but the result based on Cramer-Von Mises test selects different best fit copula models for the stations. Later, the joint cumulative probabilities of selected copula models are transformed to get standardized values of the SCDI.

The results of SCDI are compared with SPI and RDI which reveal a high correlation among the outputs and shows greater changes in the arid area while similarity in humid areas compared to SPI and RDI. The SCDI has some advantages. Firstly, the whole mathematical process is possible for the estimation and development of the index and

its values. It is to be noted that the index has more flexibility the user can use any suitable probability distributions and copula model to apply this proposed SCDI index. Secondly, the index can be used for metrological, agricultural, and hydrological purposes due to the used climatic variables and multiple time scales from 1-12 months.

Chapter 6

Conclusions and Recommendations

6.1 Summary of the study

Statistical techniques are playing a significant role in the analysis of 17 SDGs of agenda 2030 set by United Nations. Statistical methods work precisely where multiple techniques might be used to check the quality of data and modeling to achieve the main purpose of any SDG. Several SDGs are linked up with the phenomenon of drought. For example, SDGs 1 and 2 are "No poverty" and "Zero hunger" which are related to agriculture, food security, and similar other industries. Agricultural output and drought are inversely linked up with each other. Similarly, SDGs 6 and 7 are "Clean water and Sanitation" and "Affordable and Clean Energy", respectively, which are related to water resources management and the environment of a drought-prone area. Drought strongly affects water resources which are related to clean water and hydropower generation. SDG 13 is "Climate Action" related to temperature increase due to Greenhouse gases and lack of water resources whereas temperature is a major indicator of drought. Consequently, drought affects directly or indirectly several SDGs. Droughts are naturally happening phenomena with damaging properties to ecosystems, social activities, and agriculture. Bryant (1991) statistically characterized and ranked 17 hazard events where drought is ranked at the top of the hazardous events. In recent years' drought has been happening frequently and its effects are severe with variability on hydro-meteorological variables due to climate change. Therefore, monitoring, assessing, and forecasting drought using statistical techniques have gained much attention. Hence, this study has been designed to conduct a statistical analysis of drought

phenomenon threatening a developing country, Pakistan. A summary of the work is as follows.

Chapter 1 explains a brief background of drought and water resources. The statistical significance for drought risk assessment according to the objectives of the study is given in detail. A detailed literature review is included in the chapter. Literature contains several methods and their applications to assess, monitor, and forecast drought risk based on homogenous climatic regions (HCRs) and projections using drought indices. Pakistan has high variability w.r.t rainfall, temperature, climatic extremes, and drought condition which need independent drought projections. However, in the light of literature review, classification of HCRs w.r.t. drought conditions and drought projections using drought events and/or drought characteristics have never been done in the case of Pakistan. Hence, objectives of the study are specified in the light of the above research gap. Locating HCRs, future drought projections based on univariate as well as bivariate probability modeling, and development of a new drought index are set as the important objectives of this study. Lastly, some details are given about the significance of the study in Pakistan, and the source and type of the data. The following chapters have been designed to achieve the specified objectives of this study.

In **chapter 2**, the first two objectives of the study are achieved. As a first objective, drought events for the stations are calculated using the Reconnaissance Drought Index (RDI) and Standardized Precipitation Index (SPI) series. The climate data of monthly total precipitation and average temperature for 55 metrological stations from all over Pakistan are used to quantify numerical results for statistical investigation. The only criterion to select a metrological station for this study is that a station must contain at least thirty years of observed climate data for statistical inference using SPI and RDI

indices. SPI and RDI are standardized drought indices that give useful results such as duration, severity, intensity, peak, start, and end times of a drought event. SPI is based on only precipitation data while RDI is calculated using precipitation and estimated PET using Thornthwaite method. Both the indices are calculated at a 12-months' time scale (October-September) to extract drought events for meteorological stations. The stations have different number of events due to different climate data. A threshold level of -0.85 is used i.e., any drought value less than or equal to -0.85 is considered a drought event.

Second objective of the study is to identify HCRs w.r.t drought for Pakistan. Site characteristics (latitude, longitude, elevation, mean and standard deviation of precipitation) of stations are used to classify the metrological stations through cluster analysis into five subjective homogenous groups. Lastly, discordancy and heterogeneity measures are used for possible heterogeneity of subjective groups. Ultimately, five HCR w.r.t RDI and SPI are classified over whole area of Pakistan.

The HCRs are considered for regional drought projections to achieve third objective of the study in **chapter 3**. Drought events are used to find most suitable probability distributions from five 3-parameter extreme value distributions using L-moment ratio diagram and goodness of fit z-test. The selected probability distributions are estimated through the L-moments method and used for three types of drought projections at selected return periods. Three types of drought projections are obtained using Frequency Analysis at the selected return periods. Firstly, regional quantiles are calculated to cover a large area for drought risk assessment. These projections are more reliable for planning at a large level due to the maximum number of drought events from multiple sites in a region. Secondly, at-site quantiles are obtained by multiplying the drought mean value of the site with regional quantile values for planning at the grass-root level using every

single site of the study area. These quantiles show high variability and uncertainty among the results of the stations. The at-site quantiles have greater uncertainty compared to regional quantiles due to the lesser number of drought events at only a single site. Thirdly, sometimes there are vast areas with no gauging stations which need to be investigated. Hence, ungauged site projections are obtained to study the areas with no metrological stations.

Univariate drought analysis is extended to bivariate drought analysis because drought is a multi-faceted relationship of several correlated random variables such as drought duration and severity. Drought duration and severity are the two important drought characteristics used to denote the length and sum of monthly drought values for a drought event. Climate and water conditions become more severe when drought duration and severity become larger and larger. Hence it is used to further explain any drought event for detailed planning of water resources in a region. The drought events of duration and severity are extracted from RDI and SPI series at 12-months moving time scales using -0.85 threshold level for a deep investigation of drought in Pakistan. In chapter 4 which contain the results of fourth and fifth objectives of this study, these drought characteristics are used to provide a comprehensive understanding of drought conditions and is deeply investigated the drought risk in Pakistan. L-moments results are calculated for the construction of bivariate homogenous climatic regions (BHCRs) using the IFP procedure. Firstly, five BHCRs are constructed using drought duration and severity. The five BHCRs represent the same areas with more attached regional similarities compared to five univariate HCRs. Note that for both HCRs and BHCRs, regions 1 and 2 consist of Gilgit-Baltistan, AJK. and attached areas of KPK, region 3 has Balochistan province, region 4 comprises Sindh province, while region 5 comprises Punjab and attached areas of KPK provinces. However, BHCRs have some differences

in metrological stations from univariate HCRs. The best fit probability distributions are found for drought duration and severity out of five distributions using the chi-square goodness of fit statistical test to be used for drought projections in the BHCRs.

Secondly, three types of joint return periods (T_{OR} , T_{AND} & T_{KEN}) and conditional return periods are calculated based on drought duration and severity for future drought risk assessment. A region with return periods of smaller drought duration and severity is expected to repeat soon and show more chances of drought in a region. The regions have significant changes, particularly at higher return periods. In primary joint return periods T_{AND} is always greater than T_{OR} while the secondary joint return period i.e., T_{KEN} lies in between the primary joint return periods. Similarly, the conditional return periods have high values compared to joint return periods because mostly drought with such a high severity (or duration) given that a fixed duration (or severity) takes a long time to occur. The values for the selected return periods show that drought duration and severity increase with the increase in return periods. RDI and SPI have similarities at lower and dissimilarities at higher return periods because as time increases the impact of temperature and PET also increases. All types of joint and conditional results have their importance according to the situation. We used SDF curves for joint and conditional drought projections based on drought duration and severity within the BHCRs at selected return periods.

In previous chapters, SPI and RDI indices are considered for drought risk assessment and forecasting in Pakistan. There are several standardized indices like SPI, SPEI, SPTI, and RDI with one or more climate data sets i.e., precipitation, temperature, and PET. But these indices have some limitations. As SPI is based only on precipitation which explains limited drought variability. SPEI and RDI are based on precipitation and PET

where PET is underestimated in arid and semiarid areas, overestimated in humid and semi-humid areas while it cannot be estimated where zero temperature using the Thornthwaite equation. SPTI is based on precipitation and temperature data and can bitterly be used in low-temperature regions. However, temperature and PET variables play important role in drought occurrence along with precipitation. Aridity increases and soil moisture decreases due to high temperature through maximum evaporation from water resources.

To overcome or minimize above limitations and problems, in **chapter 5** which comprises the results of sixth objective of this study, we propose a new multi-scaler Standardized Copula-based Drought Index (SCDI). SCDI is based on temperature and PET along with precipitation and can bitterly be used in arid as well as humid regions with multiple time scales if required. The results of two existing aridity indices i.e., UNEP aridity index (UAI) and De-Martone aridity index (DAI) are combined using the copula model. The SCDI explains more drought variation and is equally important for meteorological, agricultural, and hydrological purposes due to the nature of variables used and multiple time scales.

Chapter 6 presents a summary, conclusion and policy implementation, and future work of this Ph.D. dissertation.

6.2 Conclusion and Policy Implementation

This study has divided Pakistan into five Homogeneous Climatic Regions w.r.t drought conditions. Region I has maximum elevated mountain ranges of HKH, and KPK joins in a very complex system with 3rd largest glacier in the world. The region has its maximum rainfall, PET, and temperature in monsoon season (June–September) with some high elevated humid stations of Balochistan. Similarly, it has approximately low

projected values of drought, moderate RMSE, and 90% error bounds which increases with an increase in return periods. The at-site and ungauged site projections have similar results within the region. It has some fluctuations in quantiles w.r.t SPI and RDI. Similarly, joint and conditional return periods show frequent chance of drought in selected years of return periods. However, this region has less rainfall and more snowfall whereas snowfall is not considered in the results and resultantly, the region shows more severe drought conditions. But according to the observed rainfall and snowfall, the region has enough water resources with no frequent chances of drought.

Region 2 has high elevated stations of AJK lies in the eastern part of the country which receives maximum rainfall among the regions, particularly in the monsoon season and minimum temperature and PET. It has high projected values of drought, RMSE, and 90% error bounds due to high variability in the occurrence of rainfall amounts compared to region 1. The at-site and ungauged site projections don't have significant variations within the region. Whereas the joint and conditional regional projections show low to moderate chances of drought to be repeated at selected years of return periods.

Region 3 has a mixed climate with some high mountains and deserts in Balochistan. It is severely drought-prone with some very dry and hyper-arid areas such as Nokkandi, with a minimum annual regional rainfall and maximum temperature and PET. The region has a minimum number of stations with large distances, therefore, regional, as well as at-site quantiles, have high estimates with maximum variability. It has the highest RMSE and 90% error bound values amongst the regions which shows high uncertainty of estimates. The ungauged site drought estimates are particularly important in this region because of vast areas with limited gauging sites. It gives approximately appropriate results compared to the IFP procedure. Similarly, this region has frequent

chances of drought using joint projections while fewer chances using conditional return periods compared to the other regions. Therefore, the region has a mixed condition of drought projections in the future at selected return periods. It may be due to the smaller number of metrological stations including humid and arid stations with larger distances.

Region 4 consists of the driest and least elevated part of the Sindh province which is highly drought prone. It has maximum temperature and PET compared to the other regions with least average regional rainfall. The least rainfall is one of the main reasons for aridity in the region. Unexpectedly, the region has minimum projected drought quantiles, RMSE, and 90% error bounds among the regions. It may be due to consistently occurring less observed rainfall and high temperature with no significant variability. The main reason is that there is consistently a very small amount of annual rainfall in the region with a small value of standard deviation which consequently gives small values of quantile due to data homogeneity. At-site and ungauged site projections have greater similarities within the region. However, the region has the most frequent chances of drought using joint and conditional projections at selected years of return periods which is the actual shape of drought due to least amount of observed rainfall and maximum temperature. Hence it shows the importance of extending the univariate results to bivariate analysis to explain more drought variability within the regions.

Region 5 includes mostly the parts of Punjab and KPK provinces with minimum elevation. It is a highly agricultural region based on both irrigation and rainfall. It has moderate average regional rainfall due to heavy monsoon rains and maximum regional temperature and PET. The region has moderate drought estimates, RMSE, and 90% error bounds. At-site and ungauged site projections have similar projections. Similarly,

the joint and conditional return periods show rare chances of drought compared to other regions in the study area.

Consequently, regions 1 and 2 which consists of Gilgit-Baltistan, AJK, and attached areas of KPK, have maximum rainfall in different months of the year and variability in results. Both the regions have high water resources and cold climates with rare chances of drought in near future. Region 3 which contains stations of Balochistan province, has the maximum variability in all kinds of projections which satisfies the actual condition of the region. However, one of the main reasons is that this region has a large area with fewer meteorological stations. Strangely the quantiles of region 4 which comprise of Sindh province, are small but the bivariate projections show the most frequent chances of drought in the region which is the actual shape of the area. Trend in observed data shows a decrease in rainfall and increase in temperature which increases the chances of drought in regions 3 and 4, respectively. These two regions are strongly affected by almost every drought in the country. Region 5 which comprises of Punjab province and attached areas of KPK province, has no severe threat of drought. The first two regions have a greater deviation in RDI and SPI results while approximately identical results with little changes in the last three regions. It has been noted that rising temperature is accompanied with an increase of PET in the regions.

Following are some of the policy suggestions for the government, non-governmental organizations (NGOs), and public to be adapted for better drought mitigation and water planning in Pakistan. Firstly, Pakistan has significant water resources in the form of world's 3rd largest glaciers and maximum rainfall, particularly in monsoon season. However, due to a lack of water resources management, most of the water is wasted reaching the Arabian Sea. The water from melting glaciers and rainfall in humid as well

as arid regions of the country should be stored in small, medium, or large dams. It will help at local and regional level in the development of agriculture for food security, increase in energy sector, drinking water for humans and animals, problems of aridity, and many more particularly in the arid regions of Balochistan and Sindh provinces. Secondly, the country has a very less covered area with trees and unfortunately, deforestation is common. Removing deforestation and planting trees have a vital role in climate change, increase in rainfall, and mitigation of droughts in the country. The regions need planting of evergreen trees to overcome the problem of aridity particularly, in regions 3 & 4 which has less rainfall, maximum temperature, and wide deserts. Thirdly, the country has frequent droughts and water scarcity but unfortunately, it has been less developed for adaptation infrastructure like cemented canals, irrigation only when needed, etc. The Government as well as the public should use different adaptation measures to cope with the worst consequences of climate change w.r.t droughts and water resources management.

This study has numerous contributions for drought planning and management w.r.t Pakistan as well as existing literature worldwide. Firstly, to identify areal climatic variability the study classifies Pakistan into five univariate HCRs. However, to explain maximum drought variability, drought characteristics are used to construct five bivariate HCRs for Pakistan. Secondly, these univariate and bivariate HCRs are used to find regional drought projections for a joint conclusion of this study. For a more comprehensive drought risk assessment of the study area and a better drought planning, several types of univariate projections like regional, at-site, and ungauged sites along with several types of bivariate drought projections like primary joint, secondary joint, and conditional return periods of duration and severity are calculated. Thirdly, a new approach based on applications of regression approach is introduced for estimating

drought projections at ungauged sites which is important for a more reliable drought risk assessment of the study area. Fourthly, policy guidelines are suggested for drought prone areas of Balochistan and Sind where drought occurrence risks are high in coming years. Lastly, a new multi-scalar Standardized Copula-based Drought Index (SCDI) is developed using precipitation, temperature, and PET data that explain more drought variability due to more climate data.

6.3 Future Work

The following future works are recommended:

- Climate observations of precipitation and temperature play a significant role to investigate drought risk using a maximum number of meteorological stations from the study area. However, in this study, some meteorological stations were dropped due to less than 30 years of climate data. It will be better to adjust these meteorological stations using any other suitable method for drought risk assessment in Pakistan.
- This study has been restricted to RDI and SPI indices using precipitation and temperature data. However, this study area can further be analyzed using different climate variables like streamflow data, daily climate data of precipitation, and temperature.
- Bayesian approach can be used for the development of HCRs and RFA in Pakistan.
- Different climate models like RCM, GCM, and simulated data can be used for future drought risk assessments in the country.
- New drought indices can be constructed using copula functions to combine the data of various climatic variables and/or results of existing drought indices.

References

- Abubakar, H. B., Newete, S. W., & Scholes, M. C. (2020). Drought characterization and trend detection using the reconnaissance drought index for Setsoto Municipality of the Free State Province of South Africa and the impact on maize yield. *Water*, 12(11), 2993.
- Achite, M., Bazrafshan, O., Wałęga, A., Azhdari, Z., Krakauer, N., & Caloiero, T. (2022). Meteorological and Hydrological Drought Risk Assessment Using Multi-Dimensional Copulas in the Wadi Ouahrane Basin in Algeria. *Water*, 14(4), 653.
- Adnan, S., Ullah, K., Gao, S., Khosa, A. H., & Wang, Z. (2017). Shifting of agro-climatic zones, their drought vulnerability, and precipitation and temperature trends in Pakistan. *International Journal of Climatology*, 37, 529-543.
- Adnan, S., Ullah, K., Shuanglin, L., Gao, S., Khan, A. H., & Mahmood, R. (2018). Comparison of various drought indices to monitor drought status in Pakistan. *Climate Dynamics*, 51(5), 1885-1899.
- Agnew, C. T. (2000). Using the SPI to identify drought. *Drought Network News* 12, 6–12.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6), 716-723.
- Alamgir, M., Khan, N., Shahid, S., Yaseen, Z. M., Dewan, A., Hassan, Q., & Rasheed, B. (2020). Evaluating severity–area–frequency (SAF) of seasonal droughts in Bangladesh under climate change scenarios. *Stochastic Environmental Research and Risk Assessment*, 34(2), 447-464.

- Ali, Z., Hussain, I., Faisal, M., Nazir, H. M., Abd-el Moemen, M., Hussain, T., & Shamsuddin, S. (2017). A novel multi-scalar drought index for monitoring drought: the standardized precipitation temperature index. *Water resources management*, 31(15), 4957-4969.
- Almazroui, M., Dambul, R., Islam, M. N., & Jones, P. D. (2015). Principal components-based regionalization of the Saudi Arabian climate. *International Journal of Climatology*, 35(9), 2555-2573.
- American Meteorological Society, (2006). Glossary of Meteorology, <http://amsglossary.allenpress.com/glossary> (accessed 10 November 2020).
- Amirataee, B., Montaseri, M., & Rezaie, H. (2018). Regional analysis and derivation of copula-based drought Severity-Area-Frequency curve in Lake Urmia basin, Iran. *Journal of environmental management*, 206, 134-144.
- Amjad, M., Akbar, M., & Ullah, H. (2022). A copula-based approach for creating an index of micronutrient intakes at household level in Pakistan. *Economics & Human Biology*, 46, 101148. <https://doi.org/10.1016/j.ehb.2022.101148>
- Angelidis, P., Maris, F., Kotsovinos, N., & Hrisanthou, V. (2012). Computation of drought index SPI with alternative distribution functions. *Water resources management*, 26(9), 2453-2473.
- Anjum, S. A., Saleem, M. F., Cheema, M. A., Bilal, M. F., & Khaliq, T. (2012). An assessment to vulnerability, extent, characteristics and severity of drought hazard in Pakistan. *Pakistan Journal of Science*, 64(2).
- Arnold, H., Shevchenko, P. V., & Xiao Lin Luo, X. (2006). Dependence Modelling via the Copula Method. Quantitative Risk Management Group, CSIRO,

Mathematical and Information Sciences, Macquarie University Campus, Australia (Technical Report).

- Ashraf, M., & Routray, J. K. (2015). Spatio-temporal characteristics of precipitation and drought in Balochistan Province, Pakistan. *Natural Hazards*, 77(1), 229-254.
- Azam, M., Maeng, S. J., Kim, H. S., & Murtazaev, A. (2018). Copula-based stochastic simulation for regional drought risk assessment in South Korea. *Water*, 10(4), 359.
- Azimi, S., & Moghaddam, M. A. (2020). Modeling short term rainfall forecast using neural networks, and Gaussian process classification based on the SPI drought index. *Water Resources Management*, 1-37.
- Bazrafshan, O., Zamani, H., Shekari, M., & Singh, V. P. (2020). Regional risk analysis and derivation of copula-based drought for severity-duration curve in arid and semi-arid regions. *Theoretical and Applied Climatology*, 141, 889-905.
- Beran, M., and Rodier, J.A. 1985. Hydrological aspects of drought. *Studies and reports in hydrology 39. UNESCO-WMO, Paris*
- Botai, C. M., Botai, J. O., Adeola, A. M., de Wit, J. P., Ncongwane, K. P., & Zwane, N. N. (2020). Drought Risk Analysis in the Eastern Cape Province of South Africa: The Copula Lens. *Water*, 12(7), 1938.
- Brito, S. S. B., Cunha, A. P. M., Cunningham, C. C., Alvalá, R. C., Marengo, J. A., & Carvalho, M. A. (2018). Frequency, duration and severity of drought in the Semiarid Northeast Brazil region. *International Journal of Climatology*, 38(2), 517-529.

- Bryant, E. A. (1991). *Natural Hazards-Threat, Disaster, Effect, Response*. Cambridge University Press, Cambridge.
- Byun, H. R., & Wilhite, D. A. (1999). Objective quantification of drought severity and duration. *Journal of climate*, 12(9), 2747-2756.
- Cancelliere, A., & Salas, J. D. (2004). Drought length properties for periodic-stochastic hydrologic data. *Water resources research*, 40(2).
- Celebi, M. E., & Kingravi, H. A. (2012). Deterministic initialization of the k-means algorithm using hierarchical clustering. *International Journal of Pattern Recognition and Artificial Intelligence*, 26(07), 1250018.
- Chen, L., & Guo, S. (2019). *Copulas and its application in hydrology and water resources*. Springer Singapore.
- Chen, Y. D., Huang, G., Shao, Q., & Xu, C. Y. (2006). Regional analysis of low flow using L-moments for Dongjiang basin, South China. *Hydrological Sciences Journal*, 51(6), 1051-1064.
- Cooley, D. (2013). Return periods and return levels under climate change. In *Extremes in a changing climate* (pp. 97-114). Springer, Dordrecht.
- Correia, F. N., Santos, M. A., & Rodrigues, R. R. (1991). Reliability in regional drought studies. In *Water resources engineering risk assessment* (pp. 43-62). Springer, Berlin, Heidelberg.
- Cunnane, C. (1988). Methods and merits of regional flood frequency analysis. *Journal of Hydrology*, 100(1-3), 269-290.
- Dai, A., Fung, I. Y., & Del Genio, A. D. (1997). Surface observed global land precipitation variations during 1900–88. *Journal of climate*, 10(11), 2943-2962.

- Das, J., Jha, S., & Goyal, M. K. (2020). Non-stationary and copula-based approach to assess the drought characteristics encompassing climate indices over the Himalayan states in India. *Journal of Hydrology*, 580, 124356.
- Das, S. (2018). Goodness-of-Fit Tests for Generalized Normal Distribution for Use in Hydrological Frequency Analysis. *Pure and Applied Geophysics*, 175(10), 3605-3617.
- Datta, R., & Reddy, M. J. (2022). Bivariate Drought Risk Estimation Using a Multivariate Standardized Drought Index in Marathwada Region, India. In *Water Management: A View from Multidisciplinary Perspectives* (pp. 173-189). Springer, Cham.
- De Martonne, E. (1925). *Traité de Géographie Physique*. Paris: Dunod Editeur.
- Dixit, S., & Jayakumar, K. V. (2022). Spatio-temporal analysis of copula-based probabilistic multivariate drought index using CMIP6 model. *International Journal of Climatology*, 42(8), 4333-4350.
- Eckstein, D., Hufils, M.L. & Wings, M. (2018) Global Climate Risk Index 2019. Berlin. Available at: <http://www.germanwatch.org> [Accessed 29th December 2018].
- Farsadnia, F., Kamrood, M. R., Nia, A. M., Modarres, R., Bray, M. T., Han, D., & Sadatinejad, J. (2014). Identification of homogeneous regions for regionalization of watersheds by two-level self-organizing feature maps. *Journal of Hydrology*, 509, 387-397.

- Fawad, M., Ahmad, I., Nadeem, F. A., Yan, T., & Abbas, A. (2018). Estimation of wind speed using regional frequency analysis based on linear-moments. *International Journal of Climatology*, 38(12), 4431-4444.
- Fawad, M., Yan, T., Chen, L., Huang, K., & Singh, V. P. (2019). Multiparameter probability distributions for at-site frequency analysis of annual maximum wind speed with L-moments for parameter estimation. *Energy*, 181, 724-737.
- Feng, J., Yan, D., Li, C., Gao, Y., & Liu, J. (2014). Regional frequency analysis of extreme precipitation after drought events in the Heihe River Basin, Northwest China. *Journal of Hydrologic Engineering*, 19(6), 1101-1112.
- Ganguli, P., & Reddy, M. J. (2012). Risk assessment of droughts in Gujarat using bivariate copulas. *Water resources management*, 26(11), 3301-3327.
- Ganguli, P., & Reddy, M. J. (2014). Evaluation of trends and multivariate frequency analysis of droughts in three meteorological subdivisions of western India. *International Journal of Climatology*, 34(3), 911-928.
- Genest, C., & Favre, A. C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of hydrologic engineering*, 12(4), 347-368.
- Ghosh, S., & Srinivasan, K. (2016). Analysis of spatio-temporal characteristics and regional frequency of droughts in the southern peninsula of india. *Water resources management*, 30(11), 3879-3898.
- Ghoudi, K., Khoudraji, A., & Rivest, E. L. P. (1998). Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles. *Canadian Journal of Statistics*, 26(1), 187-197.

- Gocic, M., & Trajkovic, S. (2013). Analysis of precipitation and drought data in Serbia over the period 1980–2010. *Journal of Hydrology*, 494, 32–42.
- Goyal, M. K., & Gupta, V. (2014). Identification of homogeneous rainfall regimes in Northeast Region of India using fuzzy cluster analysis. *Water resources management*, 28(13), 4491–4511.
- Goyal, M. K., & Sharma, A. (2016). A fuzzy c-means approach regionalization for analysis of meteorological drought homogeneous regions in western India. *Natural Hazards*, 84(3), 1831–1847.
- Guardian. (2020). Article written by scientists at the Met Office in conjunction with the Guardian <https://www.theguardian.com/environment/2011/dec/15/climate-change-rainfall> (Accessed 4 august, 2020)
- Guenang, G. M., & Kamga, F. M. (2014). Computation of the standardized precipitation index (SPI) and its use to assess drought occurrences in Cameroon over recent decades. *Journal of Applied Meteorology and Climatology*, 53(10), 2310–2324.
- Haan, C. T. (1977) *Statistical methods in hydrology*. The Iowa State University Press, Ames
- Hailegeorgis, T. T., & Alfredsen, K. (2017). Regional flood frequency analysis and prediction in ungauged basins including estimation of major uncertainties for mid-Norway. *Journal of Hydrology: Regional Studies*, 9, 104–126.
- Halwatura, D., Lechner, A. M., & Arnold, S. (2015). Drought severity--duration--frequency curves: a foundation for risk assessment and planning tool for ecosystem establishment in post-mining landscapes. *Hydrology & Earth System Sciences*, 19(2).

- Hao, C., Zhang, J., & Yao, F. (2017). Multivariate drought frequency estimation using copula method in Southwest China. *Theoretical and Applied Climatology*, 127(3-4), 977-991.
- Härdle, W. K., & Simar, L. (2019). *Applied multivariate statistical analysis* (pp. 431-442). Springer International Publishing.
- Haroon, M. A., & Jiahua, Z. (2016). Spatiotemporal analysis of drought variability over Pakistan by Standardized Precipitation Index (SPI). *Pakistan Journal of Meteorology Vol*, 13(25).
- Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)*, 28(1), 100-108.
- Hassan, B. G., & Ping, F. (2012). Formation of Homogenous Regions for Luanhe Basin- by Using L-Moments and Cluster Techniques. *International Journal of Environmental Science and Development*, 3(2), 205.
- Hassan, M. (2016). Water security in Pakistan: Issues and challenges. *United Nations Development Programme Pakistan*, 3(4), 1-34.
- Hayes, M.J. (2006). Drought indices. <<http://www.drought.unl.edu/whatis/indices.htm>> (accessed 15 October 2020).
- Heim, R. R., (2002): A Review of Twentieth-Century Drought Indices Used in the United States. *Bulletin of the American Meteorological Society*, 83, 1149-1165.
- Hipel, K. W., & Fang, L. (Eds.). (2013). *Stochastic and statistical methods in hydrology and environmental engineering: Volume 4: Effective environmental*

management for sustainable development (Vol. 10). Springer Science & Business Media.

Hosking, J. R. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(1), 105-124.

Hosking, J. R. M., & Wallis, J. R. (1993). Some statistics useful in regional frequency analysis. *Water resources research*, 29(2), 271-281.

Hosking, J. R. M., & Wallis, J. R. (1997) *Regional frequency analysis: an approach based on L-moments*. Cambridge University Press. UK.

Hounnou, E. F., & Dedehouanou, H. (2018). Variability of temperature, precipitation and potential evapotranspiration time series analysis in Republic of Benin. *IJAER*, 4, 991-1019.

Hrnjak, I., Lukić, T., Gavrilov, M. B., Marković, S. B., Unkašević, M., & Tošić, I. (2013). Aridity in Vojvodina, Serbia. *Theoretical and Applied Climatology*, 1-10.

Huard, D., Evin, G., & Favre, A. C. (2006). Bayesian copula selection. *Computational Statistics & Data Analysis*, 51(2), 809-822.

Hui-Mean, F., Yusop, Z., & Yusof, F. (2018). Drought analysis and water resource availability using standardised precipitation evapotranspiration index. *Atmospheric Research*, 201, 102-115. <https://doi.org/10.1016/j.atmosres.2017.10.014>

- Hussain, Z. (2011). Application of the regional flood frequency analysis to the upper and lower basins of the Indus River, Pakistan. *Water resources management*, 25(11), 2797-2822.
- Huth, R., & Pokorná, L. (2005). Simultaneous analysis of climatic trends in multiple variables: an example of application of multivariate statistical methods. *International Journal of Climatology*, 25(4), 469-484.
- Istat, (2021). *2021 SDGs Report Statistical Information For 2030 Agenda in Italy*. Istituto nazionale di statistica Via Cesare Balbo, 16 – Roma, Italy.
- Jensen, M. E., Burman, R. D., & Allen, R. G. (1990). Evapotranspiration and irrigation water requirements: American Society of Civil Engineers. *New York*.
- Kaluba, P., Verbist, K. M. J., Cornelis, W. M., & Van Ranst, E. (2017). Spatial mapping of drought in Zambia using regional frequency analysis. *Hydrological sciences journal*, 62(11), 1825-1839.
- Kamruzzaman, M., Cho, J., Jang, M. W., & Hwang, S. (2019). Comparative evaluation of standardized precipitation index (SPI) and effective drought index (EDI) for meteorological drought detection over Bangladesh. *Journal of the Korean Society of Agricultural Engineers*, 61(1), 145-159.
- Kao, S. C., & Govindaraju, R. S. (2010). A copula-based joint deficit index for droughts. *Journal of Hydrology*, 380(1-2), 121-134.
- Karavitis, C. A., Alexandris, S., Tsesmelis, D. E., & Athanasopoulos, G. (2011). Application of the standardized precipitation index (SPI) in Greece. *Water*, 3(3), 787-805.

- Karim, F., Hasan, M., & Marvanek, S. (2017). Evaluating annual maximum and partial duration series for estimating frequency of small magnitude floods. *Water*, 9(7), 481.
- Khan, F., Spöck, G., & Pilz, J. (2020). A novel approach for modelling pattern and spatial dependence structures between climate variables by combining mixture models with copula models. *International Journal of Climatology*, 40(2), 1049-1066.
- Khan, M. A., Faisal, M., Hashmi, M. Z., Nazeer, A., Ali, Z., & Hussain, I. (2021). Modeling drought duration and severity using two-dimensional copula. *Journal of Atmospheric and Solar-Terrestrial Physics*, 214, 105530.
- Khan, M. S. R., Hussain, Z., & Ahmad, I. (2019). A comparison of quadratic regression and artificial neural networks for the estimation of quantiles at ungauged sites in regional frequency analysis. *Applied ecology and environmental research*, 17(3), 6937-6959.
- Kis, A., Pongrácz, R., & Bartholy, J. (2017). Multi-model analysis of regional dry and wet conditions for the Carpathian Region. *International journal of climatology*, 37(13), 4543-4560.
- Klein, B., Schumann, A. H., & Pahlow, M. (2011). Copulas—new risk assessment methodology for dam safety. In *Flood risk assessment and management* (pp. 149-185). Springer, Dordrecht.
- Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. MIT press.
- Kumar, R., Chatterjee, C., Kumar, S., Lohani, A. K., & Singh, R. D. (2003). Development of regional flood frequency relationships using L-moments for

- Middle Ganga Plains Subzone I (f) of India. *Water Resources Management*, 17(4), 243-257.
- Lee, S. H., Yoo, S. H., Choi, J. Y., & Bae, S. (2017). Assessment of the impact of climate change on drought characteristics in the Hwanghae Plain, North Korea using time series SPI and SPEI: 1981–2100. *Water*, 9(8), 579.
- Lee, T., Modarres, R., & Ouara, T. B. (2013). Data-based analysis of bivariate copula tail dependence for drought duration and severity. *Hydrological Processes*, 27(10), 1454-1463.
- Li, L., She, D., Zheng, H., Lin, P., & Yang, Z. L. (2020). Elucidating diverse drought characteristics from two meteorological drought indices (SPI and SPEI) in China. *Journal of Hydrometeorology*, 21(7), 1513-1530.
- Li, Y., & Liu, G. (2020). Risk Analysis of Marine Environmental Elements Based on Kendall Return Period. *Journal of Marine Science and Engineering*, 8(6), 393.
- Liu, X., Wang, S., Zhou, Y., Wang, F., Li, W., & Liu, W. (2015). Regionalization and spatiotemporal variation of drought in China based on standardized precipitation evapotranspiration index (1961–2013). *Advances in meteorology*, 2015.
- Lyra, G. B., Oliveira-Júnior, J. F., & Zeri, M. (2014). Cluster analysis applied to the spatial and temporal variability of monthly rainfall in Alagoas state, Northeast of Brazil. *International Journal of Climatology*, 34(13), 3546-3558.
- M., Seyedabadi, M., & Moazami, S. (2018). Spatial and temporal analysis of drought based on a combined index using copula. *Environmental Earth Sciences*, 77(22), 1-12.

- Malekinezhad, H., & Zare-Garizi, A. (2014). Regional frequency analysis of daily rainfall extremes using L-moments approach. *Atmósfera*, 27(4), 411-427.
- Malekinezhad, H., Nachtnebel, H. P., & Klik, A. (2011). Regionalization approach for extreme flood analysis using L-moments.
- Marini, G., Fontana, N., & Mishra, A. K. (2019). Investigating drought in Apulia region, Italy using SPI and RDI. *Theoretical and Applied Climatology*, 137(1), 383-397.
- McKee, T. B., Doesken, N. J., & Kleist, J. (1993). The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology* 17(22), 179-183.
- Merabti, A., Meddi, M., Martins, D. S., & Pereira, L. S. (2018). Comparing SPI and RDI applied at local scale as influenced by climate. *Water resources management*, 32(3), 1071-1085.
- Mirabbasi, R., Fakheri-Fard, A., & Dinpashoh, Y. (2012). Bivariate drought frequency analysis using the copula method. *Theoretical and Applied Climatology*, 108(1-2), 191-206.
- Mirakbari, M., Ganji, A., & Fallah, S. R. (2010). Regional bivariate frequency analysis of meteorological droughts. *Journal of Hydrologic Engineering*, 15(12), 985-1000.
- Mishra AK. Singh VP. 2010. A review of drought concepts. *J. Hydrol.* 391(1): 202–216.
- Moghim, M. M., Zarei, A. R., & Mahmoudi, M. R. (2020). Seasonal drought forecasting in arid regions, using different time series models and RDI index. *Journal of Water and Climate Change*, 11(3), 633-654.

- Mondol, M. A. H., Das, S. C., & Islam, M. N. (2016). Application of Standardized Precipitation Index to assess meteorological drought in Bangladesh. *Jàmbá: Journal of Disaster Risk Studies*, 8(1).
- Montaseri, M., Amirataee, B., & Rezaie, H. (2018). New approach in bivariate drought duration and severity analysis. *Journal of Hydrology*, 559, 166-181.
- Mortuza, M. R., Moges, E., Demissie, Y., & Li, H. Y. (2019). Historical and future drought in Bangladesh using copula-based bivariate regional frequency analysis. *Theoretical and Applied Climatology*, 135(3), 855-871.
- Nabaei, S., Sharafati, A., Yaseen, Z. M., & Shahid, S. (2019). Copula based assessment of meteorological drought characteristics: regional investigation of Iran. *Agricultural and Forest Meteorology*, 276, 107611.
- Naghetini, M. (Ed.). (2017). *Fundamentals of statistical hydrology*. Cham: Springer International Publishing.
- National Drought Mitigation Center (NDMC). (2014). Drought basics—What is drought? Webpage, NDMC, Lincoln, NE, Available at: <http://drought.unl.edu/DroughtBasics/WhatisDrought.aspx>. (Accessed October 12, 2019).
- Naz, F., Dars, G. H., Ansari, K., Jamro, S., & Krakauer, N. Y. (2020). Drought trends in Balochistan. *Water*, 12(2), 470.
- Ndayiragije, J. M., & Li, F. (2022). Effectiveness of Drought Indices in the Assessment of Different Types of Droughts, Managing and Mitigating Their Effects. *Climate*, 10(9), 125.

- Nelsen, R. B., (2006). *An Introduction to Copulas*, 2nd ed.; Springer Science Business Media: New York, NY, USA.
- Neykov, N. M., Neytchev, P. N., Van Gelder, P. H. A. J. M., & Todorov, V. K. (2007). Robust detection of discordant sites in regional frequency analysis. *Water Resources Research*, 43(6).
- Ngongondo, C. S., Xu, C. Y., Tallaksen, L. M., Alemaw, B., & Chirwa, T. (2011). Regional frequency analysis of rainfall extremes in Southern Malawi using the index rainfall and L-moments approaches. *Stochastic Environmental Research and Risk Assessment*, 25(7), 939-955.
- Niemeyer, S. (2008). New drought indices. *Options Méditerranéennes. Série A: Séminaires Méditerranéens*, 80, 267-274.
- Núñez, J. H., Verbist, K., Wallis, J. R., Schaefer, M. G., Morales, L., & Cornelis, W. M. (2011). Regional frequency analysis for mapping drought events in north-central Chile. *Journal of hydrology*, 405(3-4), 352-366.
- Palmer, W. C. (1965). *Meteorological drought* (Vol. 30). US Department of Commerce, Weather Bureau, Washington, DC. 58 pp.
- Pathak, A. A., & Dodamani, B. M. (2020). Comparison of meteorological drought indices for different climatic regions of an Indian river basin. *Asia-Pacific Journal of Atmospheric Sciences*, 56(4), 563-576.
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan kaufmann.
- PMD (2018). Pakistan Meteorological Department, Ministry of climate, Govt of Pakistan (Accessed 10 January 2018).

- Qadri, S. T., Islam, M. A., Raza, A., Shalaby, M. R., & Sheikh, R. A. (2018). Physico-chemical analysis, classification of ground water, and impact of water quality on the health of people in Khushab City, Pakistan. Qaisrani, Z. N., Baloch, A., Hashim, M., Sami, S. K., Sultan, S. H., & Siddique, M. (2019). Desalination of seawater using lab scale solar plant. *Journal of Applied and Emerging Sciences*, 9(1), pp-63.
- Qaisrani, Z. N., Nuthammachot, N., & Techato, K. (2021). Drought monitoring based on Standardized Precipitation Index and Standardized Precipitation Evapotranspiration Index in the arid zone of Balochistan province, Pakistan. *Arabian Journal of Geosciences*, 14(1), 1-13.
- Quesada-Montano, B., Wetterhall, F., Westerberg, I. K., Hidalgo, H. G., & Halldin, S. (2019). Characterising droughts in Central America with uncertain hydro-meteorological data. *Theoretical and Applied Climatology*, 137(3), 2125-2138.
- Rahmat, S. N., Jayasuriya, N., & Bhuiyan, M. (2017). Identification of homogeneous areas for drought frequency analysis. *International Journal of Integrated Engineering*, 9(2), pp. 18-26.
- Rajsekhar, D., Mishra, A. K., & Singh, V. P. (2013). Regionalization of drought characteristics using an entropy approach. *Journal of Hydrologic Engineering*, 18(7), 870-887.
- Ramachandran, K. M., & Tsokos, C. P. (2020). *Mathematical statistics with applications*. Academic Press.
- Riebsame, W. E., Changnon, S. A., & Karl, T. R. (2019). *Drought and natural resources management in the United States: impacts and implications of the 1987-89 drought*. Routledge.

- Roth, M., Jongbloed, G., & Buishand, T. A. (2016). Threshold selection for regional peaks-over-threshold data. *Journal of Applied Statistics*, 43(7), 1291-1309.
- Sachs, L. (2012). *Applied statistics: a handbook of techniques*. Springer Science & Business Media.
- Sadri, S., & Burn, D. H. (2011). A Fuzzy C-Means approach for regionalization using a bivariate homogeneity and discordancy approach. *Journal of Hydrology*, 401(3-4), 231-239.
- Saf, B. (2010). Assessment of the effects of discordant sites on regional flood frequency analysis. *Journal of hydrology*, 380(3-4), 362-375.
- Sajjad, S. H., Waheed, S. A., Khan, T., Qadri, S. T., & Gilani, N. (2014). Natural Hazards and related contents in curriculum of Geography in Pakistan. *Asian Journal of Natural & Applied Sciences Vol, 3, 2*.
- Salvadori, G. (2004). Bivariate return periods via 2-copulas. *Statistical Methodology*, 1(1-2), 129-144.
- Salvadori, G., & De Michele, C. (2010). Multivariate multiparameter extreme value models and return periods: A copula approach. *Water resources research*, 46(10).
- Santos, J. F., Portela, M. M., & Pulido-Calvo, I. (2011). Regional frequency analysis of droughts in Portugal. *Water Resources Management*, 25(14), 3537.
- Santos, J. F., Pulido-Calvo, I., & Portela, M. M. (2010). Spatial and temporal variability of droughts in Portugal. *Water Resources Research*, 46(3).

- Sarhadi, A. & Heydarizadeh, M. (2014) Regional frequency analysis and spatial pattern characterization of dry spells in Iran. *International Journal of Climatology*, 34(3), 835–848.
- Saud, A., Said, M. A. M., Abdullah, R., & Hatem, A. (2014). Temporal and spatial variability of potential evapotranspiration in semi-Arid Region: Case study the Valleys of Western Region of Iraq. *International Journal of Engineering Science and Technology*, 6(9), 653-660.
- Schneider, S. H. (2011). *Encyclopedia of climate and weather* (Vol. 1). Oxford University Press.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of statistics*, 6(2), 461-464.
- Serinaldi, F., Bonaccorso, B., Cancelliere, A., & Grimaldi, S. (2009). Probabilistic characterization of drought properties through copulas. *Physics and Chemistry of the Earth, Parts a/B/C*, 34(10-12), 596-605.
- Shahzadi, A., Akhter, A. S., & Saf, B. (2013). Regional frequency analysis of annual maximum rainfall in monsoon region of Pakistan using L-moments. *Pakistan Journal of Statistics and Operation Research*, 9(1), 111-136.
- Shaphiro, S., & Wilk, M. B. J. B. (1965). An analysis of variance test for normality. *Biometrika*, 52(3), 591-611.
- She, D., & Xia, J. (2018). Copulas-based drought characteristics analysis and risk assessment across the Loess Plateau of China. *Water Resources Management*, 32(2), 547-564.

- She, D., Xia, J., Zhang, Y., & Shan, L. (2016). Regional frequency analysis of extreme dry spells during rainy season in the Wei River Basin, China. *Advances in Meteorology*, 1-13. <http://dx.doi.org/10.1155/2016/6427568>
- Sheikh, M. M., Manzoor, N., Adnan, M., Ashraf, J., & Khan, A. M. (2009). Climate profile and past climate changes in Pakistan. *Global Change Impact Studies Center (GCISC)-RR-01*.
- Shiau, J. T. (2003). Return period of bivariate distributed extreme hydrological events. *Stochastic environmental research and risk assessment*, 17(1), 42-57.
- Shiau, J. T. (2006). Fitting drought duration and severity with two-dimensional copulas. *Water resources management*, 20(5), 795-815.
- Shiau, J. T., & Modarres, R. (2009). Copula-based drought severity-duration-frequency analysis in Iran. *Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling*, 16(4), 481-489.
- Shiau, J. T., & Shen, H. W. (2001). Recurrence analysis of hydrologic droughts of differing severity. *Journal of water resources planning and management*, 127(1), 30-40.
- Sisto, R., García López, J., Quintanilla, A., de Juanes, Á., Mendoza, D., Lumbreras, J., & Mataix, C. (2020). Quantitative Analysis of the Impact of Public Policies on the Sustainable Development Goals through Budget Allocation and Indicators. *Sustainability*, 12(24), 10583.
- Sivakumar, D., Jiang, Y., & Yahia, E. M. (2011). Maintaining mango (*Mangifera indica* L.) fruit quality during the export chain. *Food Research International*, 44(5), 1254-1263.

- Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8, 229-231.
- Smith, A., Sampson, C., & Bates, P. (2015). Regional flood frequency analysis at the global scale. *Water Resources Research*, 51(1), 539-553.
- Sonali, P., & Nagesh Kumar, D. (2016). Spatio-temporal variability of temperature and potential evapotranspiration over India. *Journal of Water and Climate change*, 7(4), 810-822.
- Song, S., & Singh, V. P. (2010). Frequency analysis of droughts using the Plackett copula and parameter estimation by genetic algorithm. *Stochastic Environmental Research and Risk Assessment*. 24(5), 783-805.
- Sönmez, F. K., Koemuescue, A. U., Erkan, A., & Turgu, E. (2005). An analysis of spatial and temporal dimension of drought vulnerability in Turkey using the standardized precipitation index. *Natural Hazards*, 35(2), 243-264.
- Stedinger, J. R. (1993). Frequency analysis of extreme events. in *Handbook of Hydrology*.
- Svoboda, M., LeComte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., ... & Stephens, S. (2002). The drought monitor. *Bulletin of the American Meteorological Society*, 83(8), 1181-1190.
- Swain, R. B., & Ranganathan, S. (2021). Modeling interlinkages between sustainable development goals using network analysis. *World Development*, 138, 105136.
- Thornthwaite, C. W. (1948). An approach toward a rational classification of climate. *Geographical review*, 38(1), 55-94.
- Timm, N. H. (2002). *Applied multivariate analysis*. Springer.

- Tirivarombo, S., Osupile, D., & Eliasson, P. (2018). Drought monitoring and analysis: standardised precipitation evapotranspiration index (SPEI) and standardised precipitation index (SPI). *Physics and Chemistry of the Earth, Parts A/B/C*, 106, 1-10.
- Topcu, E., & Seckin, N. (2016). Drought analysis of the Seyhan Basin by using standardized precipitation index (SPI) and L-moments. *Journal of Agricultural Sciences*, 22(2), 196-215.
- Tosunoglu, F., & Can, I. (2016). Application of copulas for regional bivariate frequency analysis of meteorological droughts in Turkey. *Natural Hazards*, 82(3), 1457-1477.
- Tosunoglu, F., & Kisi, O. (2016). Joint modelling of annual maximum drought severity and corresponding duration. *Journal of Hydrology*, 543, 406-422.
- Tsakiris, G., & Vangelis, H. (2004). Towards a drought watch system based on spatial SPI. *Water resources management*, 18(1), 1-12.
- Tsakiris, G., & Vangelis, H. J. E. W. (2005). Establishing a drought index incorporating evapotranspiration. *European water*, 9(10), 3-11.
- Tsakiris, G., Nalbantis, I., Pangalou, D., Tigkas, D., & Vangelis, H. (2008, June). Drought meteorological monitoring network design for the reconnaissance drought index (RDI). In *Proceedings of the 1st International Conference "Drought management: scientific and technological innovations"*. Zaragoza, Spain: option Méditerranéennes, series A (No. 80, p. 2008).
- Tsakiris, G., Pangalou, D., Tigkas, D., & Vangelis, H. (2007). Assessing the areal extent of drought. *Water resources management: new approaches and technologies. European water resources association, Chania, Crete-Greece*, 1416.

- Ullah, H., & Akbar, M. (2020). Drought Risk Analysis for Water Assessment at Gauged and Ungauged Sites in the Low Rainfall Regions of Pakistan. *Environmental Processes*, 1-24.
- Ullah, H., & Akbar, M. (2021). Bivariate homogenous regions and projections based on copula function using RDI and SPI indices for drought risk assessment in Pakistan. *Arabian Journal of Geosciences*, 14(22), 1-20.
- Ullah, H., Akbar, M., & Khan, F. (2019). Construction of homogeneous climatic regions by combining cluster analysis and L-moment approach on the basis of Reconnaissance Drought Index for Pakistan. *International Journal of Climatology*, 40(1), 324-341. doi.org/10.1002/joc.6214
- Ullah, H., Akbar, M., & Khan, F. (2020). Droughts' projections in homogeneous climatic regions using Standardized Precipitation Index in Pakistan. *Theoretical and Applied Climatology*, 140(1), 787-803.
- Ullah, H., Akbar, M., & Khan, F. (2020b). Assessment of drought and wet projections in the humid climatic regions for Pakistan. *Stochastic Environmental Research and Risk Assessment*, 34(12), 2093-2106.
- UNEP, (1993) World Atlas of Desertification. The United Nations Environment Programme (UNEP), London.
- United Nations (UN) (2022). THE 17 GOALS | Sustainable Development (un.org) (Accessed 4 March 2022)
- Van der Schrier, G., Jones, P. D., & Briffa, K. R. (2011). The sensitivity of the PDSI to the Thornthwaite and Penman-Monteith parameterizations for potential evapotranspiration. *Journal of Geophysical Research: Atmospheres*, 116(D3).

- Vangelis, H., Tigkas, D., & Tsakiris, G. (2013). The effect of PET method on reconnaissance drought index (RDI) calculation. *Journal of Arid Environments*, 88, 130-140.
- Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of climate*, 23(7), 1696-1718.
- Volpi, E., Fiori, A., Grimaldi, S., Lombardo, F., & Koutsoyiannis, D. (2015). One hundred years of return period: Strengths and limitations. *Water Resources Research*, 51(10), 8570-8585.
- Von Mises, R. (1928). Statistik und wahrheit. *Julius Springer*, 20.
- Wallis, J.R., Matalas, N.C. Slack, J.R. 1974) Just a moment! *Water Resources Research*, 102), 211-221.
- Weghorst, K. M. (1996). The Reclamation Drought Index: guidelines and practical applications. *Bureau of Reclamation. Denver, CO*, 6.
- Wilhite, D. A. (2004). Drought as a natural hazard. in international perspectives on natural disasters; occurrence, mitigation, and consequences, edited by JP Stollman, J. Lidson and LM Dechano.
- Wilhite, D. A. (Ed.). (2012). *Drought assessment, management, and planning: theory and case studies: theory and case studies* (Vol. 2). Springer Science & Business Media.
- Wilhite, D. A., & Svoboda, M. D. (2000). Drought early warning systems in the context of drought preparedness and mitigation. *Early warning systems for drought preparedness and drought management*, 1-21.

- Wilhite, D.A., & Glantz, M.H. (1985). Understanding: the drought phenomenon: the role of definitions. *Water international*, 10(3), pp.111-120.
- Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (Vol. 100). Academic press.
- Won, J., Choi, J., Lee, O., & Kim, S. (2020). Copula-based Joint Drought Index using SPI and EDDI and its application to climate change. *Science of the Total Environment*, 744, 140701.
- Xie, H., Ringler, C., Zhu, T., & Waqas, A. (2013). Droughts in Pakistan: a spatiotemporal variability analysis using the Standardized Precipitation Index. *Water international*, 38(5), 620-631.
- Yevjevich, V. (1967). An objective approach to definitions and investigations of continental hydrologic droughts, *Hydrology papers*, 23, 382-391.
- Yin, Y., Chen, H., Xu, C. Y., Xu, W., Chen, C., & Sun, S. (2016). Spatio-temporal characteristics of the extreme precipitation by L-moment-based index-flood method in the Yangtze River Delta region, China. *Theoretical and Applied Climatology*, 124(3-4), 1005-1022.
- Yoo, J., Kwon, H. H., Kim, T. W., & Ahn, J. H. (2012). Drought frequency analysis using cluster analysis and bivariate probability distribution. *Journal of Hydrology*, 420, 102-111.
- Yue, Y., Shen, S. H., & Wang, Q. (2018). Trend and variability in droughts in Northeast China based on the reconnaissance drought index. *Water*, 10(3), 318.
- Zaman, M. A., Rahman, A., & Haddad, K. (2012). Regional flood frequency analysis in arid regions: A case study for Australia. *Journal of Hydrology*, 475, 74-83.

- Zarch, M. A. A., Sivakumar, B., & Sharma, A. (2015). Droughts in a warming climate: A global assessment of Standardized precipitation index (SPI) and Reconnaissance drought index (RDI). *Journal of Hydrology*, 526, 183-195.
- Zarei, A. R., Moghimi, M. M., & Bahrami, M. (2019). Comparison of reconnaissance drought index (RDI) and effective reconnaissance drought index (eRDI) to evaluate drought severity. *Sustainable Water Resources Management*, 5(3), 1345-1356.
- Zargar, A., Sadiq, R., Naser, B., & Khan, F. I. (2011). A review of drought indices. *Environmental Reviews*, 19, 333-349.
- Zhang, Q., Qi, T., Singh, V. P., Chen, Y. D., & Xiao, M. (2015). Regional frequency analysis of droughts in China: a multivariate perspective. *Water Resources Management*, 29(6), 1767-1787.
- Zhang, Q., Xiao, M., Singh, V. P., & Chen, X. (2013). Copula-based risk evaluation of droughts across the Pearl River basin, China. *Theoretical and applied climatology*, 111(1), 119-131.
- Zhao, T., & Dai, A. (2015). The magnitude and causes of global drought changes in the twenty-first century under a low-moderate emissions scenario. *Journal of climate*, 28(11), 4490-4512.

Appendix-A

Name, record length, site characteristics, and descriptive statistics of Meterological Stations.

S#	Station	Record Length	Missing No's.	Site characteristics				
				Longitude	Latitude	Elevation	MAP	MAT
1	Astore	1954-2013	0	75.00	35.25	2167	456.0	9.7
2	Badin	1931-2016	11	69.00	24.75	11	230.8	26.9
3	Bahawalnagar	1963-2016	8	73.25	30.00	161	238.2	25.6
4	Bahawalpur	1933-2016	19	71.75	29.50	116	162.5	25.7
5	Balakot	1970-2016	1	73.25	34.50	980	1475.9	18.6
6	Barkhan	1969-2016	18	69.75	30.00	1097	418.4	21.7
7	Bunji	1953-2016	3	74.75	35.75	1372	155.9	17.7
8	Cherat	1931-2016	26	72.00	33.75	1301	627.5	17.6
9	Chilas	1953-2016	10	74.00	35.75	1250	188.4	20.2
10	Chitral	1964-2016	8	71.75	35.75	1500	416.8	16.0
11	Chhor	1951-2016	14	69.25	25.50	5	228.4	26.5
12	Dalbandin	1931-2016	37	64.5	29.00	848	82.4	22.5
13	Darosh	1931-2016	0	71.75	35.50	1464	538	17.4
14	D-I Khan	1931-2016	0	71.00	31.75	173	275	24.4
15	Dir	1967-2016	0	71.75	35.25	1369	1241.2	15.5
16	Faisalabad	1931-2016	4	73.00	31.50	183	364.8	24.2
17	Ghari Dupatta	1955-2016	1	73.50	34.00	812	1440.2	19.2
18	Gilgit	1951-2016	0	74.25	36.00	1459	138.2	15.9
19	Gupis	1980-2016	27	73.50	36.25	2155	213.5	12.4
20	Hyderabad	1931-2016	2	68.50	25.50	40	171.8	27.6
21	Islamabad	1951-2016	2	73.00	33.75	543	1145.4	21.6
22	Jaccobabad	1931-2016	5	68.50	28.25	55	119.2	27.6
23	Jhelum	1974-2016	0	73.75	33.00	232	900.6	23.7
24	Jiwani	1954-2016	35	61.75	25.25	56	112.6	25.7
25	Kakul	1952-2016	0	73.25	34.25	1308	1207.8	17.2
26	Kalat	1931-2016	23	66.50	29.00	2015	180	13.8
27	Karachi	1931-2016	7	67.25	24.75	21	203	26.3
28	Khanpur	1952-2016	13	70.75	28.75	87	124.8	25.4
29	Khuzdar	1975-2016	36	66.75	27.75	1231	268.2	22.0
30	Kohat	1951-2016	0	71.50	33.50	510	532.8	23.3
31	Kotli	1952-2016	2	74.00	33.50	613	1209.2	22.0

Name, record length, site characteristics, and descriptive statistics of Meteorological Stations.

S#	Station	Record Length	Missing No's.	Site characteristics				
				Longitude	Latitude	Elevation	MAP	MAT
32	Lahore	1931-2016	2	74.50	31.50	215	601.8	24.5
33	Lasbella	1980-2015	17	66.00	26.25	219	160.9	27.1
34	Mianwali	1959-2016	3	71.50	32.50	210	505.2	24.2
35	Mohin Jodaro	1979-2016	29	68.00	25.25	52	99.1	26.1
36	Multan	1950-2016	3	71.50	30.25	122	202.6	25.4
37	Murree	1959-2016	14	73.50	33.75	2167	1658	12.9
38	Muzaffarabad	1955-2016	0	74.00	34.25	701	1428.9	20.5
39	Nawabshah	1955-2016	13	68.25	26.25	37	155.3	26.8
40	Nokkundi	1961-2016	4	62.75	28.75	682	36.8	24.8
41	Ormara	1961-2016	135	64.50	25.50	7	86.7	23.9
42	Padidan	1933-2016	55	68.25	26.75	46	125.4	26.6
43	Panjgur	1931-2016	4	64.00	27.00	980	108.6	22.3
44	Parachinar	1931-2016	12	70.00	33.75	1725	883.9	14.9
45	Passni	1931-2016	26	63.50	25.50	4	110	25.6
46	Peshawar	1948-2016	0	71.50	34.00	359	425	23.2
47	Quetta	1946-2016	3	67.00	30.25	1600	230.8	16.1
48	Risalpur	1951-2016	12	72.00	34.00	308	637.9	22.2
49	Rohri	1951-2016	6	69.00	27.75	66	105.4	27.0
50	Saidu Sharif	1974-2016	8	72.25	34.75	961	974.4	19.1
51	Sargodha	1957-2016	3	72.75	32.00	187	461.2	24.3
52	Sialkot	1931-2016	3	74.50	32.50	251	934.3	23.3
53	Sibbi	1931-2016	30	68.00	29.50	133	154.2	27.2
54	Skardu	1952-2016	2	75.75	35.25	2209	225.5	11.6
55	Zhob	1961-2016	31	69.50	31.25	1405	278.1	19.3

Note. MAP stands for Mean Annual Precipitation and MAT stands for Mean Annual Temperature of observed Annual precipitation and Annual Temperature records.

Appendix-B

Gamma Distribution

Parameters: Shape (α), Scale (β)

Range of x is: $0 \leq x \leq \infty$.

Probability density function

$$f(x) = \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}$$

CDF

$$F(x) = \frac{\Gamma(\beta, x/\alpha)}{\Gamma(\beta)}$$

Weibull Distribution

Parameters: Shape (α), Scale (β)

Range of x is: $0 \leq x \leq \infty$.

Probability density function

$$f(x) = \frac{\beta x^{\beta-1} e^{-(x/\alpha)^\beta}}{\alpha^\beta}$$

Logistic Distribution

Parameters: Location (ε), Scale (α)

Range of x is: $-\infty \leq x \leq \infty$.

Probability density function

$$f(x) = \frac{e^{-\left(\frac{x-\varepsilon}{\alpha}\right)}}{\alpha \left(1 + e^{-\left(\frac{x-\varepsilon}{\alpha}\right)}\right)^2}$$

CDF

$$F(x) = \frac{1}{1 + e^{-\left(\frac{x-\varepsilon}{\alpha}\right)}}$$

3-parameters lognormal distribution

Parameters: Location (ε), Scale (α), Shape (k)

Range of x is: $-\infty < x \leq \varepsilon + \alpha/k$ if $k > 0$; $-\infty < x < \infty$ if $k = 0$; $\varepsilon + \alpha/k < x \leq \infty$ if $k < 0$.

Probability density function

$$f(x) = \frac{e^{ky-y^2/2}}{\alpha\sqrt{2\pi}} \quad y = \begin{cases} -k^{-1} \log\left(1 - k\left(\frac{x-\varepsilon}{\alpha}\right)\right), & \text{if } k \neq 0 \\ \left(\frac{x-\varepsilon}{\alpha}\right), & \text{if } k = 0 \end{cases}$$

CDF

$F(x) = \Phi(y)$, where Φ is the standard normal CDF.

Generalized Logistic Distribution

Parameters: Location (ε), Scale (α), Shape (k)

Range of x : $-\infty < x \leq \varepsilon + \alpha/k$ if $k > 0$; $-\infty < x < \infty$ if $k = 0$; $\varepsilon + \alpha/k < x \leq \infty$ if $k < 0$.

Probability density function

$$f(x) = \frac{e^{-(1-k)y}}{\alpha(1+e^{-y})^2} \quad y = \begin{cases} -k^{-1} \log \left(1 - k \left(\frac{x-\varepsilon}{\alpha} \right) \right), & \text{if } k \neq 0 \\ \left(\frac{x-\varepsilon}{\alpha} \right), & \text{if } k = 0 \end{cases}$$

CDF

$$F(x) = (1 + e^{-y})^{-1}$$

Generalized Extreme Values Distribution

Parameters: Location (ε), Scale (α), Shape (k)

Range of x : $-\infty < x \leq \varepsilon + \alpha/k$ if $k > 0$; $-\infty < x < \infty$ if $k = 0$; $\varepsilon + \alpha/k < x \leq \infty$ if $k < 0$.

$$f(x) = \frac{e^{-(1-k)y-e^{-y}}}{\alpha} \quad y = \begin{cases} -k^{-1} \log \left(1 - k \left(\frac{x-\varepsilon}{\alpha} \right) \right), & \text{if } k \neq 0 \\ \left(\frac{x-\varepsilon}{\alpha} \right), & \text{if } k = 0 \end{cases}$$

CDF

$$F(x) = e^{-e^{-y}}$$

Generalized Pareto Distribution

Parameters: Location (ε), Scale (α), Shape (k)

Range of x is: $\varepsilon \leq x \leq \varepsilon + \alpha/k$ if $k > 0$; $\varepsilon \leq x < \infty$ if $k \leq 0$.

$$f(x) = \frac{e^{-(1-k)y}}{\alpha} \quad y = \begin{cases} -k^{-1} \log \left(1 - k \left(\frac{x-\varepsilon}{\alpha} \right) \right), & \text{if } k \neq 0 \\ \left(\frac{x-\varepsilon}{\alpha} \right), & \text{if } k = 0 \end{cases}$$

CDF

$$F(x) = 1 - e^{-y}$$

Generalized Normal Distribution

Parameters: Location (ε), Scale (α), Shape (k)

Range of x : $-\infty \leq x \leq \varepsilon + \alpha/k$ if $k > 0$; $-\infty < x \leq \infty$ if $k = 0$; $\varepsilon + \alpha/k \leq x \leq \infty$ if $k < 0$.

$$f(x) = \frac{\phi(y)}{\alpha - k(x - \varepsilon)} \quad y = \begin{cases} -k^{-1} \log \left(1 - k \left(\frac{x - \varepsilon}{\alpha} \right) \right), & \text{if } k \neq 0 \\ \left(\frac{x - \varepsilon}{\alpha} \right), & \text{if } k = 0 \end{cases}$$

ϕ is the standard normal pdf.

CDF

$F(x) = \Phi(y)$, where Φ is the standard normal CDF.

Pearson Type-3 Distribution

Parameters: Location (ε), Scale (α), Shape (k)

Let $a = \frac{4}{k^2}$, $\beta = \frac{1}{2}\alpha|k|$, and $\mu = \varepsilon - 2\alpha/k$

If $k > 0$ than rang of x is $\mu \leq x < \infty$ and

$$f(x) = \frac{(x - \mu)^{a-1} e^{-(x - \mu)/\beta}}{\beta^a \Gamma(a)} \quad F(x) = \frac{G\left(a, \frac{(x - \mu)}{\beta}\right)}{\Gamma(a)}$$

If $k = 0$ than distribution is normal, the range of x is $-\infty < x < \infty$ and

$$f(x) = \phi\left(\frac{x - \varepsilon}{\alpha}\right) \quad F(x) = \Phi\left(\frac{x - \varepsilon}{\alpha}\right)$$

If $k < 0$ than rang of x is $-\infty < x \leq \mu$ and

$$f(x) = \frac{(\mu - x)^{a-1} e^{-(\mu - x)/\beta}}{\beta^a \Gamma(a)} \quad F(x) = 1 - \frac{G\left(a, \frac{(\mu - x)}{\beta}\right)}{\Gamma(a)}$$

Appendix C

Selected Copula functions with generator function, Pickands dependence function, parametric space, and upper-tails dependence measure.

Copula	$C(u_1, u_2)$	$\phi(t)$	$A(t)$ $t \in (0, 1)$	τ_ϑ	UTD
Clayton	$(u_1^{-\vartheta} + u_2^{-\vartheta} - 1)^{-1/\vartheta}$ $\vartheta \in (-1, \infty), \text{ but } \vartheta \neq 0$	$\frac{1}{\vartheta}(t^{-\vartheta} - 1)$	N. A	$\frac{\vartheta}{\vartheta + 2}$	0
Frank	$-\frac{1}{\vartheta} \log \left[1 + \frac{(e^{-\vartheta u_1} - 1)(e^{-\vartheta u_2} - 1)}{e^{-\vartheta} - 1} \right]$ $\vartheta \in (-\infty, \infty), \text{ but } \vartheta \neq 0$	$-\ln \left[\frac{e^{-\vartheta t} - 1}{e^{-\vartheta} - 1} \right]$	N. A	1 $+\frac{4}{\vartheta} \{ * D_1(\vartheta) - 1 \}$	0

Copula	$C(u_1, u_2)$	$\phi(t)$	$A(t)$ $t \in (0,1)$	τ_θ	UTD
Gumbel-Hougaard	$\exp \left[-\{(-\ln u_1)^\theta + (-\ln u_2)^\theta\}^{1/\theta} \right]$ $\theta \in (1, \infty)$	$(-\ln(t))^\theta$	$\{t^\theta + (1 - t)^\theta\}^{1/\theta}$	$\frac{\theta - 1}{\theta}$	$2^{-1/\theta}$
Galambos	$u_1 u_2 e^{-[(-\ln u_1)^{-\theta} + (-\ln u_2)^{-\theta}]^{-1/\theta}}$ $\theta \in (0, \infty)$	N. A	$1 - \{t^{-\theta} + (1 - t)^{-\theta}\}^{-1/\theta}$	-----	$2^{-1/\theta}$

*The Debye function $D_r(\theta)$ is used for positive integer r , $D_r(\theta) = \frac{r}{\theta^r} \int_0^\theta \frac{t^r}{e^t - 1} dt$

Note: UTD denotes the upper-tail dependence measures while N. A is used that the function is not for not applicable for the copula function.

