MODELLING EXTREME HYDROLOGIC EVENTS UNDER NONSTATIONARY CONDITION IN HO CHI MINH CITY, VIETNAM

Submitted in partial fulfillment of the requirement for the award of the degree of

Doctor of Philosophy

by

Le Thi Hoa Binh

Roll No. 716110

Supervisors

Prof. N.V. Umamahesh

Prof. E. V. Rathnam

DEPARTMENT OF CIVIL ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY

WARANGAL-506004, INDIA

FEBRUARY 2020

NATIONAL INSTITUTE OF TECHNOLOGY

WARANGAL

CERTIFICATE

This is to certify that the thesis entitled "**Modelling extreme hydrologic events under nonstationary condition in Ho Chi Minh City, Vietnam**" being submitted by **Ms. Le Thi Hoa Binh** for award of the degree of Doctor of Philosophy to the Faculty of Engineering and Technology of **National Institute of Technology Warangal** is a record of bonafide research work carried out by her under my supervision and it has not been submitted elsewhere for award for any degree.

N.V. Umamahesh

Professor

Thesis supervisor

E. V. Rathnam

Professor

Thesis co-supervisor

Department of Civil Engineering

National Institute of Technology Warangal (T.S.) INDIA

Declaration

This is to certify that the work presented in the thesis entitled "**Modelling extreme hydrologic events under nonstationary condition in Ho Chi Minh City, Vietnam**" is a bonafide work done by me under the supervision of **Prof. N.V. Umamahesh** and **Prof. E. V. Rathnam** and was not submitted elsewhere for the award of any degree.

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have nor misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the Institute and can also evoke penal action from sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Le Thi Hoa Binh

Roll No. 716110

Date:

Abstract

The statistical theory of extremes has been widely used in scientific fields including hydrology, water resources engineering, environmental sciences, finance, public health and so on. As a foundation concept, the statistical methods based on the extreme value theory require the assumption of stationarity in extreme event time series. It means that the occurrence probability of extreme event is not expected to change over time. In recent years, changes in rainfall characteristics and hydrological cycle have been reported for many places of the world due to the change in global climate related to human activities. These changes propose that the assumption of stationarity in hydro-meteorological time series becomes doubtful and may not be suitable in engineering design applications.

As the biggest economic city in the South of Vietnam, Ho Chi Minh City (HCMC) is an example of an emerging coastal megacity dressing the increases of exposure levels to climate risks. In HCMC, heavy rainfall, which is considered as a main cause of floods, witnessed an increase in frequency and magnitude during the last few decades. Although nonstationarity in extreme rainfall has been proved in many places of the world, research into nonstationarity feature in extreme rainfall in HCMC has not been paid attention thoroughly. The covariate Time is usually chosen in previous studies in the literature which aim to model nonstationary extreme rainfall. However, directly using time covariate based linear form in nonstationary modelling may create more bias. Further, it is documented that heavy rainfall in HCMC is influenced by the physical processes. Therefore, in this thesis, Multi-Objective Genetic Algorithm (MOGA) based method is used for modelling nonstationary extreme rainfall over HCMC. From the study results, it is observed that the MOGA based method can be used to develop less bias and good quality nonstationary models which can be used to model spatial variation of extreme rainfall over HCMC.

Recently, human intervention and climate change have been suggested to be the causes of changes in extreme water level which impacts on the likelihood of flooding, especially in coastal areas. In many studies, extreme water level frequency analysis has been developed under nonstationary condition, in which the parameters of a given distribution vary with time or several climatological variables. However, the water level shows unique characteristic as

they are strongly impacted by local influences, therefore the covariates used for nonstationary extreme water level modelling should be chosen with respect to the area of interest. With the above reasons, it is important to consider local variables which have strong physical associations with the process of floods for studying nonstationary extreme water level. Therefore, in this thesis, four local covariates, i.e. rainfall, sea level, urbanization growth and outflows from upstream reservoirs are used to develop nonstationary extreme water level models. The results from this thesis indicate that the nonstationary approach using local covariates is suitable for modelling extreme water level for HCMC. Additionally, based on the best chosen statistical models, sea level and urbanization are found to be the significant influences on nonstationarity in extreme water level at all surveyed stations. Moreover, it could be found that the extreme water level values derived from the stationary models are underestimated relative to the best nonstationary models for all stations.

Flood hazard maps provide essential information for flood risk management and mitigation purposes. Basically, the inputs for flood modelling used to create the flood hazard maps are based on the assumption of data stationarity for flood frequency analysis. However, the changes in the behavior of the climate system can lead to the nonstationarity in extreme events as well as flood series. Therefore, two flood sources, i.e. extreme rainfall and sea level are modelled under nonstationary condition before entering into the flood simulation model as initial inputs. MIKE FLOOD, which is a coupled hydrodynamic model, is used to simulate the flood regime. The coupled hydrodynamic model has been developed for cross-sections based on channel modelling with one-dimensional model and linking these floodplain modelling with two-dimensional model. The high-resolution topographical data derived from Light Detection and Ranging (LiDAR) data and flexible meshes generation are used as the input data for hydrodynamic model to simulate the flood regime for the study area. From the results, the spatial variation of flood hazards indicates that the regions located along both sides of riverbanks are expected to experience a significant increase in the area flooded. Besides, it is also noted that the floodplain extent is larger based on the assumption of nonstationarity.

Univariate frequency analysis can be effective if the infrastructure design is based on a single flood variable. Otherwise, univariate frequency analysis may not provide the complete behavior of flood characteristics. Recently, multivariate frequency analysis has proven to be a

practical approach by researches, especially in flood frequency analysis. Till date, the use of nonstationary approach in flood frequency analysis could be found in many studies which mainly focused on a single random variable, but only a few studies available related to nonstationary multivariate frequency analysis. Therefore, a part of this thesis is to model the multivariate based on nonstationary copula approach for flood variables, i.e rainfall and water level. Copulas are applied to overcome the restriction of classical multivariate flood frequency analysis by choosing the marginal distribution from different types of the probability distribution function. Furthermore, the joint probability of rainfall and water level are constructed using different approaches, which provide more options in choosing appropriate data sample for analysis. From the results, it is noticed that the Generalized Extreme Value (GEV) distribution is suggested as the most appropriate marginal distribution for modeling the flood variables. The joint return periods of rainfall and water level obtained through the optimal copula and marginal distribution show the significant differences between the samples. It means that a reciprocal situation can be found, when a higher value of rainfall corresponds to a lower value of water level and vice versa.

Overal, the findings of this thesis is that the nonstationary approach using local covariates is suitable for modelling extreme hydrologic water level for HCMC. Furthermore, it is observed that the MOGA based method can be used to develop less bias and good quality nonstationary models which can be used to model spatial variation of extreme rainfall over HCMC. In the flood hazard analysis, the result indicated that the regions located along both sides of riverbanks are expected to experience a significant increase in the area flooded. Besides, it is also noted that the floodplain extent is larger based on the assumption of nonstationarity. In addition, it is noticed that the GEV distribution is suggested as the most appropriate marginal distribution for modeling the flood variables in the multivariate flood frequency analysis.

Content

List of Figures

List of Tables

Nomenclature

The following list gives the notations used in chapters of the thesis.

μ	Location parameter of the GEV
$\hat{\mu}$	Estimated value of the GEV's location parameter
σ	Scale parameter of the GEV
$\hat{\sigma}$	Estimated value of the GEV's scale parameter
ξ	Shape parameter of the GEV
$\hat{\xi}$	Estimated value of the GEV's shape parameter
F	Objective function
Θ	Distribution's parameters
$L(\Theta\|X)$	Log-likelihood function
$\chi 2$ (d)	Chi-square distribution with "d" degree of freedom
ε	Residual of nonstationary GEV derived quantiles
$\hat{\varepsilon}$	Sorted nonstationary GEV derived quantiles' residuals
T	Return period
Z	Return level
f (t)	Covariate based on Time
P	Probability
α	Scale parameter of the GPD
$\hat{\alpha}$	Estimated value of the GPD's scale parameter

β	Shape parameter of the GPD
$\hat{\beta}$	Estimated value of the GPD's shape parameter
u	Threshold
F(x,y)	Bivariate distribution function
F(x), F(y)	Marginal distribution function
C	Copula function
θ	Copula parameter
$T_{X,Y}^{OR}$, $T_{X,Y}^{AND}$	OR and AND joint return periods
λ	Mean interarrival time
ϕ	Cumulative density function of the standard normal distribution
Φ	Cumulative density function of the multivariate normal distribution

Abbreviations

1D	One-dimensional
2D	Two-dimensional
3D	Three-dimensional
AAL	Average Annual Losses
ADB	Asian Development Bank
ADF	Augmented Dickey-Fuller
AIC	Akaike Information Criterion
AICc	Corrected Akaike Information Criterion
AM	Annual Maxima
ANFIS	Adaptive Neuro-Fuzzy Inference Systems
ANN	Artificial Neural Network
AO	Arctic Oscillation
CDF	Cumulative Distribution Function
D	Flood Depth
DEM	Digital Elevation Models
DHI	Danish Hydraulic Institute
ELM	Extreme Learning Machine
ENSO	El Nino-Southern Oscillation
ESRL	Earth System Research Laboratory

EVA	Extreme Value analysis
EVT	Extreme Value Theory
GA	Generic Algorithm
GDP	Gross Domestic Product
GEV	Generalised Extreme Value
GIS	Geographic Information System
GMSL	Global Mean Sea Level
GPD	Generalised Pareto Distribution
GUM	Gumbel
HCMC	Ho Chi Minh City
HEC	Hydrologic Engineering Centre
HEC-HMS	Hydrologic Modelling System
HRU	Hydrologic Response Units
IHACRES	Identification of Unit Hydrographs and Component Flows from Rainfall, Evaporation and Streamflow Data
IOD	Indian Ocean Dipole
IPCC	Intergovernmental Panel on Climate Change
KPSS	Kwiatkowski-Phillips-Schmidt-Shin
LiDAR	Light Detecting and Ranging
LLH	Log-Likelihood
LN	Log Normal

LULC	Land Use Land Cover
MDC	Macdinhchi
MEI	Multivariate ENSO Index
M-K	Mann-Kendall
MOGA	Multi-objective Genetic Algorithm
MSS	Multispectral Scanner
NAO	North Atlantic Oscillation
NHMS	National Hydro–Meteorological Service
NOAA	National Oceanic and Atmospheric Administration
NSE	Nash and Sutcliffe
OLI	Operational Land Imager
PDO	Pacific Decadal Oscillation
PNA	Pacific/North American Pattern
POT	Peak over Threshold
P-P	Probability-Probability
Q-Q	Quantile-Quantile
RMSE	Root Mean Square Error
RSR	Root Mean Square Error-observation Standard Deviation Ratio
SCS	Soil Conservation Service
SE	Standard Error
SIMHYD	Simplified Hydrolog

SMARG	Soil Moisture Accounting and Routing Model
SOI	Southern Oscillation Index
SWAT	Soil and Water Assessment Tool
SWM	Stanford Watershed Model
TDM	Thudaumot
TS	Thematic Mapper
TSN	Tansonnhat
USACE	United States Army Corps of Engineers
V	Flood Velocity
WEI	Weibull
WNF	Hybrid ANFIS with Multi-resolution Analysis using Wavelets
WNN	Wavelet Neural Networks

Chapter 1

Introduction

1.1 Urban flood

Floods could be considered as one of the most devastating natural disasters, impacting on the social economy, human life and natural environment. For the period of 1980 to 2009, floods were responsible for approximately 540,000 deaths, and have affected about 2.8 billion people across the world (Doocy et al., 2013). Flood exposure can be acute significantly in coastal cities, especially coastal cities in developing countries undergoing population, economic growth and urban expansion are experiencing more flooding (Hallegatte et al., 2013, Nicholls et al., 2008, Lasage et al., 2014, Adikari et al., 2010). Globally, seventeenth out of the twenty most vulnerable cities in 2005 are from developing countries (Figure 1. 1) which might not possess the flood defense systems as good as the richer ones.

Currently, urban flooding continues to be a problematic concern worldwide with complicated changes in frequency and magnitude of events due to both the natural processes and human activities (Birikundavyi et al., 2002, Ishak et al., 2013, Badrzadeh et al., 2015).

The increases in extreme events (e.g. heavy rainfall and coastal storm surges) and rapid urbanization have been widely recognized for their significant contribution in exacerbating urban flood (Badrzadeh et al., 2015, Biswas and Jayawardena, 2014, Bergstrom, 1976). Unfortunately, based on the recent flooding damages recorded in many places across the world, it seems likely that the increasing unpredictability of natural disasters and floods are now exceeding present-day design considerations (Perrin et al., 2003).

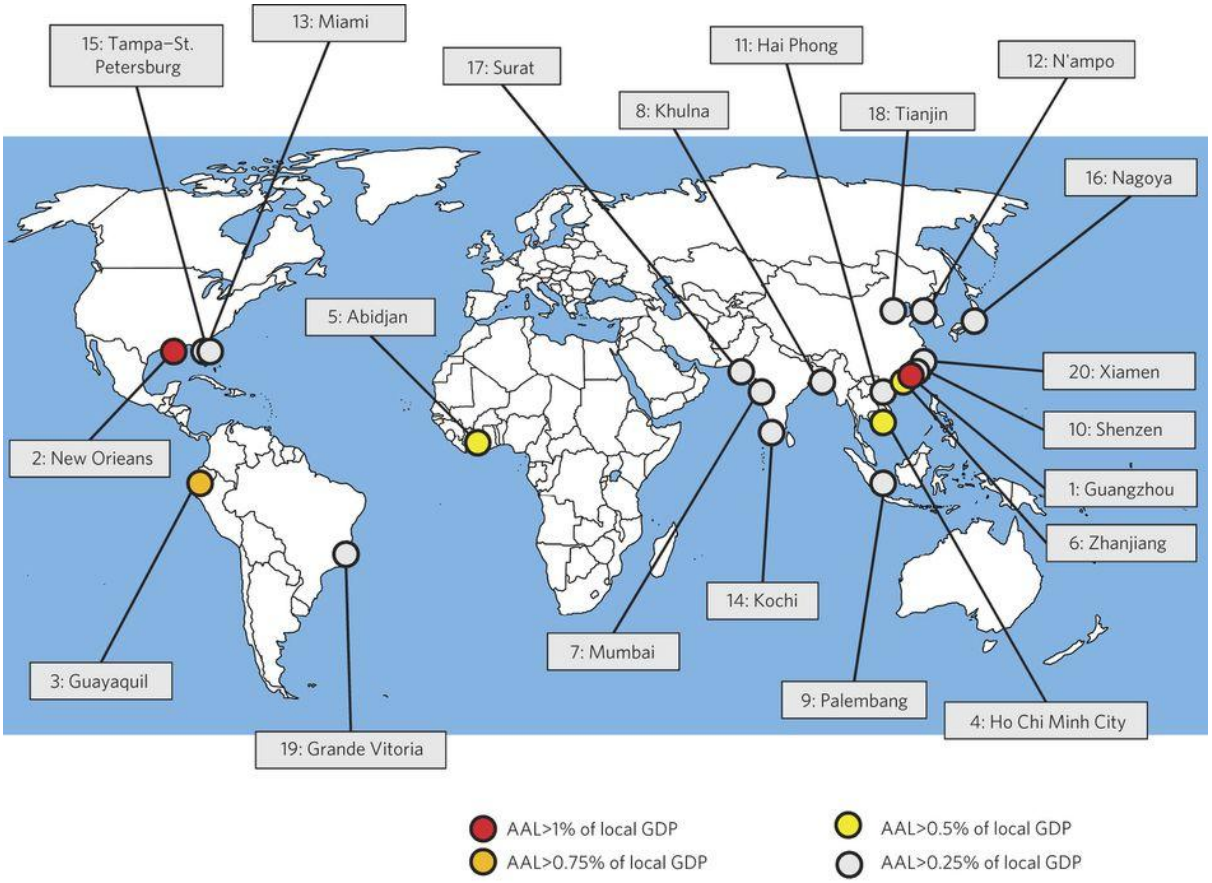


Figure 1. 1: The twenty most vulnerable cities in 2005 according to the ratio of average annual losses (AAL) to the city's gross domestic product (GDP) (Hallegatte et al., 2013)

1.2 Extreme value theory

Extreme value theory (EVT) is unique as a statistical tool since it develops techniques and models to describe the unusual rather than the usual (Coles, 2001). The distinguishing feature of extreme value analysis (EVA) is the ability to quantify the behavior of unusually large (or small) values even when those values are scarce. Particularly, EVA usually requires

levels of a process event that are more extreme than that have been already observed. For example, a dam is normally required to protect against all flood events that it is likely to experience within its lifespan of 100 years. Unfortunately, the observed flood data might only be available for a shorter period, say 30 years. Therefore, it seems impossible to estimate what floods might happen over the next 100 years when providing only 30 years of historical data. Fortunately, EVT provides a framework that enables to extrapolate data from historically observed levels to unobserved levels as such type (Coles, 2001, Mockler et al., 2016). EVT has already become one of the most important statistical disciplines for modelling extreme events over the last few decades. The applications of EVT have been found in various scientific fields including hydrology, environmental sciences, finance, public health and so on.

It is a fact that the frequency analysis of extreme events (e.g. water level and rainfall) plays an essential role in engineering practice, especially in flood structure defense design or flood risk management. As a foundation concept, the statistical methods based on the extreme value theory require the assumption of stationarity in extreme event time series. It means that the occurrence probability of extreme event is not expected to change over time. Nevertheless, it is recently suggested by many scientists that stationarity may no longer be suitable in hydro-meteorological frequency analysis.

1.3 Nonstationarity in extreme events

In the most intuitive sense, stationary means that the statistical properties of a process do not change over time while nonstationary time series shows the trends, seasonal effects and other structures depend on time. In recent years, changes in rainfall characteristics and hydrological cycle have been reported for many places of the world due to the change in global climate related to human activities (Berg et al., 2013, Trenberth, 2011, Groisman et al., 2005, Milly et al., 2008). These changes propose that the assumption of stationarity in hydro-meteorological time series becomes doubtful (Sugahara et al., 2009, Khaliq et al., 2006). Also, it is stated that “Stationarity is dead” (Milly et al., 2008).

In fact, the atmosphere and ocean have warmed over decades, and human intervention has been condemned to be partly responsible for global warming (Min et al., 2011, Petheram et al., 2012). For every 1^{o}C warming, the atmosphere’s water holding capacity increases by

7%, which results in more extreme rainfall (Berg et al., 2013, Trenberth, 2011). Besides, the characteristics of rainfall such as intensity, frequency and duration are also changed due to the influence of global warming (Trenberth et al., 2003). In the other hand, the recent studies have reported that extreme rainfall is influenced by the physical processes such as the El Nino-Southern Oscillation (ENSO), the North Atlantic Oscillation (NAO), the Indian Ocean Dipole (IOD) or the Pacific Decadal Oscillation (PDO) (Villafuerte et al., 2015, Mondal and Mujumdar, 2015, Agilan and Umamahesh, 2015, Kenyon and Hegerl, 2010, Cai and Rensch, 2012). That is a reason why the large-scale climate variables have been commonly used in frequency analysis as the factors causing nonstationarity in extreme rainfall. For example, Villafuerte et al. (2015) found that ENSO has a significant impact on the changes in extreme rainfall in the Philippines. Mondal and Mujumdar (2015) reported that global warming, local temperature changes and ENSO play a significant role in causing nonstationarity in rainfall extremes over India.

Similar to extreme rainfall, low-flow patterns in many places are also proved to have nonstationary features in it. To be more specific, in urban areas, natural land surfaces have been replaced by artificial surfaces to meet the requirement of residential and industrial purposes, thereby results in increasing the magnitude and frequency of floods (Li et al., 2015). Prosdocimi et al. (2015) also indicated that increasing urban levels affect significantly on high flows. Also, recent studies demonstrate the impacts of meteorological patterns and catchment conditions on the low-flows series (Du et al., 2015, Shin and Kim, 2017). In addition to the effects of local processes, the global processes (e.g. ENSO, NAO, PDO and so on) also possibly contribute to the changes in extreme water level as well as flood series (Li et al., 2015, Menéndez and Woodworth, 2010, López and Francés, 2013). To sum up, the stationary condition may no longer suitable, and the concept of nonstationarity should be used in the frequency analysis of extreme water level.

1.4 Flood hazard mapping

In last few decades, the effects of climatic changes and sea level rise have been putting an additional pressure which could increase flood vulnerability by effecting magnitude and frequency of floods (Bates et al., 2005, Nicholls and Cazenave, 2010, Purvis et al., 2008, Karamouz et al., 2017). In reducing damages and losses, flood hazard mapping becomes

priority information since it significantly contributes to the flood warning systems as well as the flood risk management scheme. However, assessing the flood risk at the river basin is not a simple task, because of the complex nature of flood generation caused by a combination of precipitation, river basin characteristics and human activities. Fortunately, the development of numerical flood modelling in recent years, namely the availability of advanced flood modelling and modern survey techniques for collection of high-quality input data for those models, allows to simulate flood behavior and to also study the characteristics of future floods (Alkema, 2007).

Till now, a variety of models have been developed for providing flood information. A methodology that combines the advantageous features of one-dimensional (1D) and two-dimensional (2D) hydraulic models and also the high-resolution of topographic data, are typically applied. Flood hazard maps show the intensity of floods and their associated exceedance probability (Di Baldassarre et al., 2010). One of the common approaches of flood inundation modelling is the use of deterministic based on single simulation (Ali, 2018). In the deterministic approach, three main issues in developing the flood hazard maps using hydrodynamic models such as the topography data resolution, the hydrodynamic model simulation and the design flood hydrograph estimation are commonly considered.

In most of the flood hazard studies, flood depth is widely used to classify the hazard index (Sharif et al., 2016, Komi et al., 2017, Garrote et al., 2016, Alfieri et al., 2014). Nevertheless, flood hazard includes many elements such as the stability of human body, buildings and vehicles in floodwaters (Xia et al., 2011). Therefore, a single parameter cannot completely assess the potential damage of flood flows on people, buildings and vehicles. In previous studies, the combination of the flood depth (D) and velocity (V) has been used as a proxy for the force of the floodwaters to access the instability of human body, vehicles as well as the failure of buildings in floodwaters (Xia et al., 2011, Xia et al., 2014, Kreibich et al., 2009). Therefore, it is suggested that the flood hazard maps can be classified using the combined flood hazard curves derived from the flood depth and velocity thresholds.

1.5 Nonstationary multivariate frequency analysis

It is a fact that univariate frequency analysis can be effective if the infrastructure design is based on a single flood variable. Otherwise, univariate frequency analysis may not

provide the complete behavior of flood characteristics (Chebana and Ouarda, 2011). Recently, multivariate frequency analysis has proven to be a practical approach by researches, especially in flood frequency analysis.

Till date, the use of nonstationary approach in flood frequency analysis could be found in many studies which mainly focused on a single random variable, but only a few studies available related to nonstationary multivariate frequency analysis. However, most of these studies used time as the explanatory variable of the marginal distribution and dependence parameters. As a matter of fact, the change of paradigm from stationarity to nonstationarity can be affected by many causes (e.g. land use and land cover change or climate change). The effects of these factors may not exactly follow the passage of time. More importantly, the changes in the climate or watershed characteristics have affected not only in the nonstationarity of individual hydrological series but also in the dependence structure between the different hydrological series. Therefore, using time as a covariate in nonstationary modelling may suffer some limitations. The physical processes which reflect the physical relationship to hydro-meteorological events should be considered in nonstationary multivariate frequency analysis.

1.6 Motivation for the study

Floods could be considered as one of the most devastating natural disasters, impacting millions of people every year across the world (Jongman et al., 2012, Hallegatte et al., 2013, Lasage et al., 2014, Karamouz et al., 2017). Recently, human intervention and climate change have been suggested to be the causes of changes in extreme events such as rainfall and water level which impact on the likelihood of flooding, especially in coastal areas. Since flood exposure is continuously increasing in coastal zones, there is a growing demand of estimations of the magnitude and frequency of extreme events for the design of coastal defense structure as well as flood risk management purpose. Also, as is stated in the previous section, the nonstationary behavior in hydro-meteorological time series has recently been studied and developed. Therefore, it is necessary to take nonstationarity into account when modelling extreme events.

One of the key factors in preventing and reducing flood damages and a number of lives lost is to provide flood risk assessment information through flood hazard maps. Basically, the

inputs for flood modelling used to create the flood hazard maps base on the assumption of data stationarity for flood frequency analysis. However, the changes in the behavior of the climate system can lead to the nonstationarity in flood series, hence stationary assumption may lead to incorrect flood risk information. Hence, a part of this thesis is dedicated to develop flood hazard maps under nonstationary condition.

Coastal zones are commonly vulnerable to floods caused by the combination of multiple sources. In such areas, floods can result in massive damage when heavy rainfalls occur concurrently with coastal storm surges, thereby resulting in huge socio-economic losses (Karamouz et al., 2014, Hunt, 2005). For these regions, univariate frequency analysis may no longer be effective to describe floods that are characterized by several correlated variables. It is, therefore, necessary to consider the joint probability of flood-caused sources in the evaluation and management of flood risk. Therefore, part of this thesis is dedicated to model the joint probability using the copula-based bivariate frequency analysis, which considers the nonstationary behavior in the flood series.

1.7 Objectives of the study

With this background and appreciating the significance of the studies on extreme hydrologic events, flood hazard mapping and flood frequency analysis, the objectives of the study have been formulated. The objectives of this study are listed as follows: (i) Detecting and analyzing the trend in the extreme hydrologic variables; (ii) Modelling spatial variation of extreme precipitation for the study area under nonstationary condition; (iii) Modelling the flood frequency estimation under nonstationary condition; (iv) Developing inundation maps for study area considering the changes in environment; and (v) Modelling the multivariate based on nonstationary copula approach for flood variables.

1.8 Scope of the Study

It is a fact that, extreme rainfall occurrence is controlled by not only one but also many physical processes. Besides, extreme rainfall is recently proved to have nonstationary feature in time series and continuously result in severe floods. It is therefore suggested that extreme rainfall should be analyzed under nonstationary condition before using as initial information of infrastructure design purpose or making decision purpose. This study suggests an

appropriate method which considers the impact of physical processes to model extreme rainfall under nonstationary condition and investigates the spatial variation of extreme rainfall over a study area.

Recently, human intervention and climate change have been suggested to be the causes of changes in extreme water level which impacts on the likelihood of flooding, especially in coastal areas. In many studies, extreme water level frequency analysis has been developed under nonstationary condition, in which the parameters of a given distribution vary with time or several climatological variables. However, incorporating all of the physical processes may increase the bias of nonstationary modelling. Moreover, water level shows unique characteristic as they are strongly impacted by local influences, therefore the covariates used for nonstationary extreme water level modelling should be chosen with respect to the area of interest. With the above reasons, it is important to consider local variables which have strong physical associations with the process of floods for studying nonstationary extreme water level. Thus, in this study, the most significant physical processes for modelling extreme water level are identified.

As an important role of flood hazard mapping and floodplain extent in making decision, or establishing flood warning systems, it is suggested that the flood sources (e.g. rainfall, water level, upstream outflows and sea level rise) should be analysed under both stationary and nonstationary conditions before using as initial inputs of hydrological and hydrodynamic models since the global climate is continuously changing and unpredictable. The surveyed cross-sections and high-resolution topographical data are used as the input data for the coupled hydrodynamic model to simulate the flood regime. Furthermore, the multi-scale mesh modelling approach, where fine resolution is applied for channel and raised embankment areas and coarser resolution is developed for uniform topographic height are used to develop the hydrodynamic model in this study. The coupled hydrodynamic model in which channel flow is linked to floodplain flow using lateral connection is used to improve accuracy the flood inundation results without the significant increasing computational requirement of the hydraulic model. The flood depth and velocity obtained from the hydrodynamic model are used to develop high-resolution flood hazard maps for the study area.

Single variable flood frequency analysis does not give a comprehensive understanding and assessment of the actual behavior of flood phenomena. This approach can lead to high uncertainty or failure of guidelines in water resources planning, operation and design of hydraulic structure and floodplain zoning. Therefore, it is essential to study the multivariate probability behavior of flood correlated variables, especially under the changing environment. Copulas are widely used for multivariate analysis in various fields. The main advantage of copulas is that the dependence structure is independently modeled with the marginal distribution that allows for multivariate distribution with different margins and full coverage of dependence structure. The essential step in the modelling processing copula is the selection of copula function, which is the best fit for the data sample. This study suggests that the copula function should be selected based on the dependence structure of the variable. Furthermore, the performances of Frank, Clayton, Gaussian and Plackett copulas for an asymptotic independent variable are also assessed.

1.9 Outline of the thesis

Literature review related to modelling nonstationary extreme events, nonstationary approach in multivariate frequency analysis, hydrological and hydraulic modelling and flood hazard mapping are briefly presented in Chapter 2.

Chapter 3 presents the trends analysis in extreme rainfall for eighteen locations in the study area. Besides, the modelling the spatial variation of extreme precipitation under nonstationary condition is also introduced in this chapter.

The most significant physical processes which cause the nonstationarity in extreme water level series are identified and presented in Chapter 4. Besides, the comparison of the estimated extreme water level corresponding to different return periods between the stationary and nonstationary cases is also presented.

In Chapter 5, the analysis of hydro-meteorological events under the nonstationary condition which is used as the input of flood modelling is computed. The developed flood inundation model using the coupled 1D-2D hydrodynamic model with high-resolution topography data is also presented. Furthermore, the high-resolution flood hazard maps, which

are quantified by considering the flood depth and velocity in combination, are also established in this chapter.

Chapter 6 presents the joint probability of correlated flood variables (i.e. water level and rainfall) using nonstationary copula-based bivariate frequency analysis. As such, the nonstationary behavior is modelled for dependence structure and marginal distributions by using local physical processes as covariates. The estimated joint return period of selected variables using the optimal copula and marginal distribution is introduced in this chapter.

Chapter 7 presents the summary of the study, the conclusions arrived and some recommendation for further research activities based on the conclusions from study on modelling nonstationary extreme events.

Chapter 2

Literature review

2.1 Introduction

In this chapter, literature related to modelling nonstationary extreme events, nonstationary multivariate frequency analysis, hydrological and hydraulic modelling and flood hazard mapping are briefly discussed. In detail, the nonstationarity in extreme events time series, i.e. rainfall and water level due to the changes in various physical processes which are diving extreme occurrences is discussed. Then, the studies related to nonstationary multivariate frequency analysis is presented. Hydrological and hydraulic modelling and the flood hazards estimation based on the deterministic approach are presented in the next two sections.

The overview of the entire study is organized as follows. The trend of time series is firstly tested using statistical test. Then the nonstationary univariate and bivariate frequency anlsysis are developed for extreme hydrologic events. In the nonstationary univariate frequency analysis, extreme rainfall and water level are modeled using time and local physical

components as covariates. And then, the return level of these extreme hydrologic events are used as input

for hydrodynamic model to develop the high-resolution flood hazard for the study area. The marginal distributions and dependence structures are also modeled under nonstationary condition in the nonstationary bivariate frequency analysis part. The joint return period of each sample from the combination of these extreme hydrologic events are assessed. The general methodology flowchart is shown in Figure 2.1.

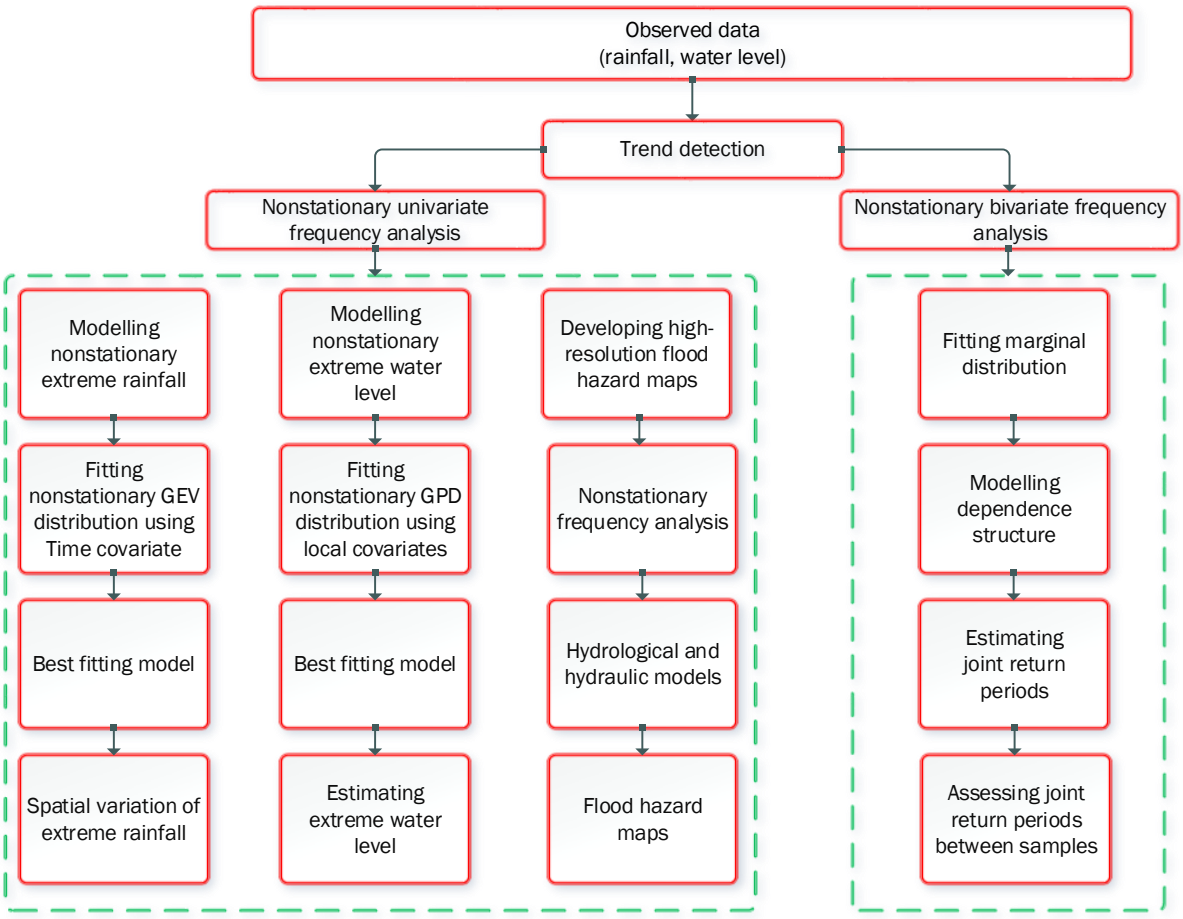


Figure 2.1: Overal methodology in this study

2.2 Modelling nonstationary extreme events

Although the nonstationarity in the hydrological regime has been widely accepted in recent years by hydrologists, it is not easy to provide adequate evidence for supporting it due

to measurement bias, shortage in data and the high variability of hydrological processes in nature (Strupczewski et al., 2001). When having statistical evidence of nonstationarity in time series and having identified a time trend, it can be considered as the resultant of various external and internal long-lasting actions in a basin such as human activities, natural variation of climate and so on. Once the time series are nonstationary, modelling the statistical characteristic of the nonstationarity is needed. There are two main methods for hydro-meteorological frequency analysis, namely non-parametric and parametric methods. While non-parametric methods show their limitations for hydraulic design due to the requirement of sample size, parametric methods have become widely used methods for fitting probability distribution to data since they can produce asymptotically efficient and unbiased estimated (Strupczewski et al., 2001). During the last few years, the nonstationarity in extreme hydro-meteorological events is modelled by many researchers for different applications. In this sub-section, such studies are briefly discussed.

Sugahara et al. (2009) assessed the frequency of nonstationary extreme rainfall in the Sao Paulo City, Brazil, for the period of 1933- 2005. In that study, authors used the Generalized Pareto Distribution (GPD) to fit extreme rainfall series extracted using the peak over threshold (POT). Besides, different percentiles of rainfall are used to define a threshold for extracting POT series. Then, four GDP models, including one stationary and three nonstationary models, are constructed for each extracted time series. Three covariates, i.e. the annual cycle, linear trend and both annual cycle and linear trend, were used in constructing nonstationary models. In that study, the corrected version of Akaike Information Criterion (AICc) was used to identify the best model among the four models. The results from that study showed that the nonstationary model with a linear trend in the scale parameter is found to be the best model. In addition, 0.99 quantiles of daily rainfall amount have increased by 40 mm between 1933 and 2005.

Villafuerte et al. (2015) modeled the nonstationarity in extreme rainfall in the Philippines over the period 1911-2010. In particular, authors used GEV distribution and linked the location parameter of GEV distribution with two covariates (i.e. global mean temperature and ENSO). The best fitting model is selected using the Akaike Information Criterion (AIC) and the likelihood ratio test. The significant changes in extreme rainfall in the study area were found to be related to the near-surface global mean temperature and ENSO.

The results also imply a potential intensification and an increase in the occurrence of extreme rainfall in the future since the global mean temperature continues to rise.

Agilan and Umamahesh (2016b) developed nonstationary rainfall Intensity- Duration-Frequency (IDF) curves for the Hyderabad City, India. In that study, authors modeled the character of rainfall using GEV distribution and allowed the parameters of GEV distribution to vary with five physical covariates (i.e. urbanization, ENSO, IOD, local temperature changes and global warming) and Time covariate. Authors used AIC and the likelihood ratio test to choose the best model. In addition, to develop the nonstationary rainfall IDF curves, authors used 95^{th} percentile of location and scale parameters values in historical observed. The results showed that urbanization and local temperature changes are the best covariates for short duration rainfall, and global processes are the best covariates for long duration rainfall. Moreover, the findings also indicated that the covariate Time never qualified as the best covariate. Similar to above discussed studies, there are many studies which model nonstationarity in extreme rainfall in different parts of the world include, but are not limited to Beguería et al. (2011), Villafuerte and Matsumoto (2015), Panagoulia et al. (2014), Agilan and Umamahesh (2015), Mondal and Mujumdar (2015),Wi et al. (2016), Yilmaz et al. (2016).

Mudersbach and Jensen (2010) analyzed the frequency of nonstationary extreme water level of the German North Sea. First, authors used the non-parametric Mann-Kendall (M-K) test to detect significant trends in annual maximum water level and found that time series have a significant trend on the basis of the 95% significant level. Further, authors modeled the annual maximum water level series using GEV distribution. In details, authors allowed the location and scale parameter of GEV distribution to vary with time. The results showed that the nonstationary GEV approach is suitable for determining coastal design water level.

Masina and Lamberti (2013) investigated the spatial and temporal changes in extreme water level in the North Adriatic based on historical data from the Trieste, Venice, Porto Corsini and Rimini tide gauges. In that study, authors modeled the extreme water level series using GEV distribution and allowed the distribution parameters to vary nonlinearly. In particular, the regional climate indices such as the NAO and Arctic Oscillation (AO) are used in analyzing the variability of the extreme water level along the Northern Adriatic coast. The

findings indicated that the AO is found to be more influential than NAO on extreme sea level in the study area.

Razmi et al. (2017) modeled the nonstationary extreme water level in a coastal part of New York City for a period of 1920-2015. In that study, authors used annual maxima (AM) and POT approaches to extract data time series. The extracted time series were checked for potential trend and nonstationarity using statistical tests including M-K, Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) testes and were found to be significantly increasing. The GEV distribution and GPD distribution were applied as the probability distribution functions on the extracted data under nonstationary approach. Authors allowed the location and scale parameters to vary with time, while the shape parameter was considered to be constant. Ten stationary and nonstationary GEV models were fitted to the selected data. The results showed that the design values of extreme water level under nonstationary condition were significantly different from those obtained under the stationary assumption. Authors also recommended that nonstationary frequency analysis should be used to estimate values of hydrologic variables in different design periods. Similar efforts to model extreme water level under nonstationary condition include, but are not limited to Méndez et al. (2007), Menéndez and Woodworth (2010), Masina and Lamberti (2013), Skjong et al. (2013), Serafin and Ruggiero (2014).

2.3 Nonstationary approach in multivariate frequency analysis

Multivariate parametric distributions which have been extended from univariate distribution have been commonly used to model multivariate extreme events (e.g. flood, rainfall, wind and wave). Traditionally, multivariate frequency analysis is based on stationary approach. However, as mentioned earlier, due to the changing environments and human activities, the statistical characteristics of hydrological series in watersheds may be altered, hence leading to nonstationary feature in times series. Therefore, the use of nonstationary approach in multivariate frequency analysis has now become more and more essential for hydrology design under changing environments.

Bender et al. (2014) used bivariate nonstationary approach to investigate the time-dependent behavior of bivariate hydrological design parameters. In that study, the flood peak and volume time series of the Rhine River providing 191 years of data are used for analysis.

Authors used GEV distribution to model the marginals and Archimedean copulas to model the dependence structure between flood peak and flood volume. The authors found that the influence of trends in the marginal distribution parameters on the corresponding design values is substantially larger than trends in the dependence measure.

Karamouz et al. (2017) suggested an integrated framework to delineate floodplain and assessment of flood damage for Manhattan in New York City by considering the joint effect of inland and coastal flooding. In that study, rainfall and water level data rea tested for frequency analysis considering data stationarity and nonstationarity. Authors used GEV distribution to obtain and compare extreme rainfall and water level magnitudes in different return periods. The bivariate Gaussian, *t*, Clayton and Gumbel copulas are used to model dependence structure between rainfall and water level in that study. A geographic information system (GIS)-based model, using depth-damage functions, land use data, digital elevation model (DEM) and raster maps, is used to estimate flood damage. The results from that study show that floodplain extent and estimation of flood damage are increased when data nonstationarity is considered. Similar efforts using nonstationary approach in multivariate frequency analysis include, but are not limited to Jiang et al. (2015), Sarhadi et al. (2016), Ahn and Palmer (2016).

2.4 Hydrological and hydraulic modelling

2.4.1 Hydrological model

Hydrological modelling involves formulating the mathematical models to describe the hydrological processes such as surface flow, rainfall, infiltration, snowmelt, interception, evapotranspiration as well as their interactions. Hydrological modelling is considered as a valuable tool for researchers in the field of water resources planning and management. Till date, various hydrological models have been developed and applied in small, large or very complex basins across the world to access the effects of climate change on water resources in general and floods in particular.

Rainfall-runoff models can be classified according to the physical processed involved into modelling as well as model input and parameters. It can be classified as lumped and distributed model based on the model parameters as a function of space and time. Model is

deterministic if a set of input values will always produce the same output values, while a model is stochastic if the input values need not produce the same output values. Besides, Models can also be classified as static and dynamic models based on time factors. As such, a static model excludes time while a dynamic model includes time. In a nutshell, the most popular and important classifications are empirical, conceptual and physically-based models (Devia et al., 2015). This sub-section is devoted to introduce briefly the characteristics of these three hydrological models.

Empirical models (Metric models) are primarily based on observations and seek to characterize system response from those data, without considering the changes in the catchment. In details, the mathematical equations in these models are derived from concurrent input and output data and not from the physical processes within catchment. Unit hydrograph, statistical models (i.e. linear and non-linear regressions) and machine learning techniques (e.g. artificial neural network, fuzzy regression and k nearest neighbor) are commonly used in the empirical models. The machine learning techniques are widely used and proved to be appropriate tools for hydrological modelling, especially in stream-flow prediction. However, the models derived from the machine learning techniques seem not to add any scientific knowledge or improved understanding in the field of hydrology.

Badrzadeh et al. (2015) developed models applied for real-time runoff forecasting at Casino station on Richmond River, Australia based on four different approaches, namely traditional artificial neural network (ANN), adaptive neuro-fuzzy inference systems (ANFIS), wavelet neural networks (WNN) and hybrid ANFIS with multi-resolution analysis using wavelets (WNF). The results confirmed the robustness of hybrid wavelet-based models compared to ANN, adaptive neuro-fuzzy inference systems, wavelet neural networks. Yaseen et al. (2016) forecasted the monthly stream-flow discharge rate in the Tigris River, Irag using the extreme learning machine (ELM) method. The results from that study showed a good improvement using ELM model than support vector regression and generalized regression neural network in hydrological forecasting problems. Similar efforts using empirical models in hydrological events prediction include, but are not limited to Birikundavyi et al. (2002), Chang and Chang (2006), Biswas and Jayawardena (2014), Badrzadeh et al. (2015).

Conceptual models (Parameter models) mimic the hydrological processes by conceptualizing the catchment as a number of interconnected storages. Conceptual models also considered physical law but in profoundly simplified form. The mathematic functions are used in this method to describe the movement of water within the catchment. The model parameters are assessed not only from field data but also through calibration. Conceptual models are useful for various purposes and they can be used to infill the lost data or reconstruction of flow sequences.

One of the well-known conceptual models is the Stanford Watershed Model (SWM) elaborated by Crawford and Linsley (1966). The Sacramento model (Bergstrom, 1976) and the GR4J model (Perrin et al., 2003) are other well-known conceptual rainfall-runoff models with different complexities. Vaze et al. (2011) investigated the implications of different rainfall inputs on the calibration and simulation of four conceptual models using data from 240 catchments across southeast Australia. Four rainfall-runoff models including Simplified Hydrolog (SIMHYD) (Porter and McMahon, 1971), Sacramento, Soil Moisture Accounting and Routing Model (SMARG) (Kachroo, 1992), and Identification of Unit Hydrographs and Component Flows from Rainfall, Evaporation and Streamflow Data (IHACRES) (Jakeman et al., 1990) were used. In that study, the results indicated that the better spatial rainfall representation gives better estimates of mean annual runoff for the SMARG, SIMHYD and the Sacramento models. Similar to that study, there are many studies which show that applications of conceptual models on river flow predictions. These include, but are not limited to Petheram et al. (2012), Shin et al. (2015), Mockler et al. (2016), Shin and Kim (2017), Onyutha (2019).

Physically-based models are based on physical law and theoretical principles. These models are characterized by parameters derived by field measurements and have a direct physical significance. The models use a spatial discretization based on grid, hillslopes or some hydrologic response units. Therefore, these models can be highly appropriate when a high level of spatial discretization is needed in modelling. The physically-based models can have many advantages compared to other models because of the use of parameters having a physical interpretation. The limitation of these models is that large data needed, scale-related problems and overparameterization.

MIKE SHE is one of the best-known models, which was developed by a consortium of European institutes such as Danish Hydraulic Institute (DHI), British Institute of hydrology and French consulting agency SOGREAH. MIKE SHE is a fully distributed, physically-based, distributed model capable of both single event and continuous simulations. The model can able to simulate hydrology in plot field and watershed scale The physical based on nature for the model lends inclusion of topography and watershed characteristics (i.e. soil, vegetation and weather parameter sets). Besides, physically-based models such as SWAT and the Hydrologic Engineering Center-Hydrologic Modelling System (HEC-HMS) are also used universally to estimate runoff in both gauged and ungauged watersheds. SWAT is a complex physically-based, continuous model and was designed to forecast the impact of watershed management practices on hydrology, sediment, water quality and agriculture production on the gauge and ungauged basins. The model simulates a watershed by dividing it into sub-basins which are further subdivided into Hydrologic Response Units (HRU). For each HRU in every sub-basin, SWAT simulates the soil water balances, groundwater flow, lateral flow, channel routing, evapotranspiration, crop growth and nutrient uptake, pond and wetland balances, soil pesticide degradation and in-stream transformation nutrients and pesticides. HEC-HMS model is a physically-based distributed model, designed to simulate the rainfall-runoff process of dendritic watershed systems. This model has been widely used to simulate and forecast streamflows in humid, tropical, subtropical and arid watersheds.

Golmohammadi et al. (2014) used MIKE SHE, Soil and Water Assessment Tool (SWAT) and Agricultural Policy Environment extender models to simulate the streamflows of the Canagagigue watershed in the Grand River basin, Canada. The results indicated that the mean daily and monthly flow simulated by MIKE SHE was much better than other models. Three hydrological models, i.e. NAM, SWAT and MIKE SHE, used to model the combined impact of climate change and land use change on hydrology for a catchment in Denmark in a study of Karlsson et al. (2016). The results indicated that substantial changes in discharge extreme due to the changing of land use. Similar efforts using physically-based models in simulating rainfall-runoff include, but are not limited to Larsen et al. (2014), Lin et al. (2015), Kabiri et al. (2015), Cibin et al. (2016), Teng et al. (2018).

2.4.2 Hydraulic model

Hydraulic modelling of surface water is an important element for hydrological and geomorphological applications, especially in floodplain flow simulation. Hydraulic models can be classified by spatial dimension, namely one-dimensional (1D), two-dimensional (2D) and three-dimensional (3D) hydraulic models. The 1D models are the most widespread approach due to numerical stability and computational efficiency. The 1D models can solve the problems of flood flows in open channel with the assumption that the water level is confined within its riverbanks. In contrast, when water is over the riverbank, the 2D models may be the most appropriate solution for flood simulation. The 2D models can solve full shallow water equations which are able to simulate timing and duration of flooding with high accuracy. The water flows can be simulated by 3D models. However, 3D models are time-consuming with relatively low efficiency compared to 1D and 2D models (Liu et al., 2015).

HEC-RAS and MIKE 11 are two of the worldwide-known 1D hydraulic models. HEC-RAS has been developed by the Hydrologic Engineering Centre (HEC) of United States Army Corps of Engineers (USACE), while MIKE 11 was developed by the Danish Hydraulic Institute. HEC-RAS can simulate steady and unsteady flows in the river channels and floodplains, while MIKE 11 is applied for unsteady flow simulation in rivers and floodplains. Other 1D models were developed for flow simulation including FLDWAV and FLUCOMP (Fread and Lewis, 1988, Ervine and MacLeod, 1999). Ahmed (2010) developed a watershed model using MIKE 11 for the Rideau Valley Watershed, Ontario. In that study, a detailed model including 532 km of rivers and lakes, 106 basins, 122 bridges and culverts and 20 water control structures were developed. Authors used observed streamflow data for a period of 5 years for calibration and an additional 5 years of data for validation. The results showed that the developed model could simulate the hydrological processes with a reasonable to high degree of accuracy. Timbadiya et al. (2011) simulated unsteady flows of Tapi River from Ukai dam to Surat city by HEC-RAS model using field surveyed geometric data of the stream. In particular, authors used the flood flows of the year 1998, 2003 and 2006 for model simulation. The performance statistics from that study revealed that the simulated flood flows are in close agreement with the observed flows, therefore the model can be used for flood forecasting in lower Tapi River. Similar efforts using 1D models in flow simulation include,

but are not limited to Yoshida and Dittrich (2002), Zhang and Shen (2007), Mu and Zhang (2007), Paz et al. (2009), Timbadiya et al. (2014b).

In recent years, a variety of 2D hydraulic models have been developed for providing flood information. Among these models, TELEMAC-2D, MIKE 21, TUFLOW, Delft3D-FLOW and SOBEK are the most widespread models used for flood-prone areas modelling and mapping. Teng et al. (2015) developed inundation maps for the Murrumbidgee region and Macquarie-Castlereagh region, Australia. In details, authors used MIKE 21 hydraulic model and airborne laser altimetry digital elevation model to derive the floodplain storages. The results from MIKE 21 model (i.e. inundation extent, volume and water depth) showed a good agreement (above 85% agreement) with those obtained from high satellite imageries. Authors suggested that the model is suitable for practical floodplain inundation simulation as well as scenario modelling under both current and future climate conditions. Similar to that study, there are many studies which show that applications of 2D hydraulic models in flood inundation simulation. These include, but are not limited to Bates et al. (2005), Dutta et al. (2007), Abu-Aly et al. (2014), Karim et al. (2015), Costabile and Macchione (2015), Shen et al. (2015), Ticehurst et al. (2015).

2.5 Flood hazard mapping

In the last few decades, the effects of climatic change and sea level rise have been creating additional pressure which could increase flood vulnerability by affecting the magnitude and frequency of floods (Bates et al., 2005, Nicholls and Cazenave, 2010, Purvis et al., 2008, Karamouz et al., 2017). In terms of reducing damages and losses, flood hazard mapping has become a priority, since the information significantly contributes to flooding warning systems, as well as flood risk management schemes.

There are two main approaches to develop the flood hazard maps, namely deterministic and probability approaches. The most common representation of simulation results is a deterministic flood inundation map based on a single simulation. Probabilistic flood mapping designed to incorporate uncertainty from input data and model parameters, represent spatial and temporal risk and present flood maps in terms of probabilities and percentages.

2.5.1 Probabilistic approach

In the probabilistic approach, the process of floodplain mapping requires certain steps. These steps include: (i) the setting up of flood inundation models; (ii) sensitivity analysis of the model using historical flood data and (iii) ensemble simulation using an uncertainty design event. The probabilistic approach, which is based on ensemble simulation, does not necessarily require the use of physical behavior of the river and floodplain models. Di Baldassarre et al. (2010) compared two different methods (i.e. deterministic and probabilistic) for flood hazard mapping using 2D hydrodynamic model. Their study indicated that flood hazard mapping using probability approach seems to be more reliable. Kalyanapu et al. (2012) used Monte Carlo based 2D flood inundation framework for estimating flood hazard mapping. Their study showed that the probability-weighted flood risk approach provides improved accuracy of flood risk estimation.

However, the main disadvantages of using physically-based 2D hydraulic models in probability frameworks, have been the simulation time required for each simulation. Simulating hundreds of flood events with these computational speeds would take large computer time making 2D model application counterproductive (Timbadiya et al., 2014a). A probability analysis with 2D hydraulic models has been limited to a smaller number of scenarios and smaller spatial domains. Besides, Aronica et al. (2012) suggested that flood inundation probability alone may be insensitive to discharge in relatively steep urban catchments and maybe a limited measure of flood hazard. Moreover, Thompson and Frazier (2014) supposed that a few probabilistic flood hazard maps were limited with respect to the hazard behavior they modelled. These models could also be computationally expensive and parameterization was difficult to compute for forces that were not fully predictable.

2.5.2 Deterministic approach

In a deterministic approach, floodplain maps consist of the construction of a physically-based fully 2D hydraulic model, calibration and validation of the model using historical flood event, using the best-fit statistical model to generate the design flood hydrograph and elaboration of the model results to generate flood hazard maps. Deterministic modelling tools have widely been applied because they are capable of translating changes in input parameters into a change in flood characteristics.

Flood inundation depth and inundation extent can be computed using computational models based on solutions of the full or approximate form of the shallow water equations. 2D hydrodynamic models are identified as the appropriate tools for simulating the flow of water over flat terrain and complex topography. 2D hydrodynamic model results provide further opportunity to develop more meaningful hazard maps by incorporating additional hazard parameters.

The high-resolution flood hazard maps, which was developed using the advanced deterministic and probability approaches, can provide complete information about the physical hazard and reduce uncertainty found in traditional approaches. Masood and Takeuchi (2012) developed flood hazard maps using 1D hydrodynamic model for the city of Dhaka in Bangladesh. Their studies used a simple form of deterministic approach in establishing flood hazard maps. Mazzoleni et al. (2013) suggested a semi-probabilistic approach to develop the hazard map due to embankment-overtopping for the Po River basin. They used 1D and 2D hydrodynamic models to simulate the hydrodynamic regime (i.e. water depth and flow velocity) and the flood hazard maps were obtained using the hazard curves, which combined different flood parameters (i.e. flood extent, water depth and flow velocity). Similar efforts using deterministic approach combined to high-resolution data in flood and inundation mapping include, but are not limited to Moore (2011), Sampson et al. (2012), Shen et al. (2015), Papaioannou et al. (2016).

2.6 Conclusions

In this chapter, an overview of the literature on statistical modelling nonstationary extreme events, hydrological modelling, hydraulic modelling and flood hazard mapping are presented. It seems that the extreme events are increasing in frequency, duration and magnitude in many places across the world, and are likely to cause more intense and frequent floods. Therefore, modelling characteristic of extreme events is essential and has been paid more attention from many researchers over the last few years.

It is the fact that the atmosphere and ocean have warmed over decades, and human intervention has been condemned to be partly responsible for global warming (Min et al., 2011, Petheram et al., 2012). For every $1^{o}C$ warming, the atmosphere's water holding capacity increase by 7%, which results in more extreme rainfall (Berg et al., 2013, Trenberth,

2011). In addition, natural land surfaces have been replaced by artificial surfaces to meet the requirement of residential and industrial purposes, thereby results in increasing the magnitude and frequency of floods (Li et al., 2015). Prosdocimi et al. (2015) also indicated that the increases in urban levels affect significantly on high-flows. Also, recent studies demonstrate the impacts of meteorological patterns and catchment conditions on the low-flows series (Du et al., 2015, Shin and Kim, 2017). Moreover, the global processes (e.g. ENSO, NAO, PDO and so on) also possibly contribute to the changes in extreme events as well as flood series. However, incorporating all of physical processes may increase the bias of nonstationary extremes modelling. Hence, most of this thesis focuses on modelling extreme events concerning covariates of the nonstationary models.

One of the measures to mitigate the flood damage is providing useful information through floodplain areas, the spatial distribution of flood hazard. Therefore, it is of great importance for understanding flood hazard at river scale. Flood modelling inputs that are used to create flood hazard maps are normally based on the assumption of data stationarity for flood frequency analysis. However, changes in the behavior of climate systems can lead to nonstationarity in flood series. Hence, a part of this thesis is dedicated to develop flood hazard maps for HCMC, Vietnam, under nonstationary condition.

Coastal flooding is predicted to be increasing significantly in developing-country cities that are undergoing urban extension and economic growth. In such areas, floods can result in massive damage when heavy rainfalls occur concurrently with coastal storm surges, thereby resulting in huge socio-economic losses. Under the changing environments, multivariate flood frequency analysis, which carefully considered the nonstationary behavior in the flood series, can provide more information for flood mitigation. Hence, a part of this thesis focuses on a nonstationary bivariate approach and its application for HCMC where there has been increasing vulnerability to floods from multiple sources.

Chapter 3

Modelling spatial variation of extreme precipitation over Ho Chi Minh City under nonstationary condition

3.1 Introduction

In recent years, changes in rainfall characteristics and hydrological cycle have been reported for many places of the world, especially increasing in rainfall extremes, due to the change in global climate related to human activities (Berg et al., 2013, Trenberth, 2011, Groisman et al., 2005, Milly et al., 2008). These changes propose that the assumption of stationarity in hydro-meteorological time series may be no longer suitable (Sugahara et al., 2009, Khaliq et al., 2006). As such, the concept of nonstationary extreme value analysis has been well developed and used in modelling the behavior of rainfall extremes in many regions as Taiwan, China, West Central Florida, South Korea and Greece and so on (Chu et al., 2013, Feng et al., 2007, Nadarajah, 2005, Park et al., 2011, Kioutsioukis et al., 2010, Westra et al., 2013).

In other hand, the previous studies have reported that extreme rainfall is influenced by the physical processes such as the ENSO, NAO, IOD or PDO (Villafuerte et al., 2015, Mondal and Mujumdar, 2015, Agilan and Umamahesh, 2015, Kenyon and Hegerl, 2010, Cai and Rensch, 2012). That is a reason why the large-scale climate variables have been commonly used in frequency analysis as the factors causing nonstationarity in extreme rainfall. For example, Villafuerte et al. (2015) found that ENSO has a significant impact on the changes in extreme rainfall in the Philippines. Mondal and Mujumdar (2015) reported that global warming, local temperature changes and ENSO play a significant role in causing nonstationarity in rainfall extremes over India. Hence, the physical processes associated with extreme rainfall have a periodicity component in it (Agilan and Umamahesh, 2016a). Introducing periodicity feature in extreme rainfall frequency analysis is therefore needed, which could significantly impact on the chosen design values.

Most of the studies in modelling extreme rainfall under nonstationary condition, the linear trend was commonly adopted to express the function of parameters of chosen distribution. In particular, Wi et al. (2016) constructed nonstationary GEV distribution and GPD models by introducing linear trend in location and scale parameters. Villafuerte et al. (2015) investigated the changes in extreme rainfall in the Philippines using GEV distribution with linear form of location parameter. Cheng and AghaKouchak (2014) used a nonstationary GEV distribution with linear trend in location parameter to develop the rainfall intensity-duration-frequency curves. However, Agilan and Umamahesh (2016a) recommended that using time covariate based linear form could lead to increase the bias of nonstationary model. Um et al. (2017) suggested that nonlinear function could be a useful option when applied to the nonstationary frequency analysis of extreme rainfall. And the use of flexible nonlinear forms to model nonstationarity in extreme rainfall could be found in many researches (Sugahara et al., 2009, Panagoulia et al., 2014, Agilan and Umamahesh, 2016b, Yilmaz et al., 2016). Thus, among many nonlinear forms that were established and used in the past, choosing an appropriate form for modelling nonstationary extreme rainfall mimicking all involved physical processes is essential.

The main objective of this study is to model the spatial variation of extreme rainfall over HCMC, a flood-prone city in the South of Vietnam, using appropriate nonstationary GEV model. In order to address this objective, the observed data is firstly checked for the

possible trend by using the M-K test. The nonlinear trend in the extreme rainfall time series is developed by the use of MOGA. The best model is chosen by the AICc and the likelihood ratio test. Finally, the best nonstationary model is used to investigate the spatial variation of extreme rainfall corresponding to different return periods.

3.2 Study area

HCMC is located in the South of Vietnam and belongs to a transitional region between the southeastern and Mekong River Delta regions. This city is considered as a typical example of a vulnerable coastal city. Much of HCMC is located in low-lying lands of the Saigon-Dongnai River basin that are prone to frequent flooding (World Bank, 2010, Lasage et al., 2014). Total 154 of the city's 322 communes and wards have a history of regular flooding, affecting 12 percent of the HCMC population (around 971,000 people) (ADB, 2010). Governed by a tropical monsoon climate, the annual average temperatures in HCMC range from 26^{o}C to 27^{o}C, and seasons are separated into wet and dry. Annual average rainfall is around 2,000 mm which mainly occurs in the wet season from May to October. The heavy rainfall in a short period is considered as the main cause of floods in rainy season (Le Vo, 2007, ADB, 2010).

In last few decades, beside huge challenges related to fast-growing population, urbanization and industrialization (Le Vo, 2007, World Bank, 2010), HCMC also has to cope with climate problems such as increases in frequency and magnitude of extreme rainfall events (ADB, 2010). In addition, the impacts of ENSO and PDO on rainfall regimes in Vietnam has been investigated by recent studies (Gobin et al., 2015, Yen et al., 2011, Chan and Zhou, 2005, Nguyen et al., 2014, Chen et al., 2013), which partly influence the extreme rainfall in HCMC. Since extreme rainfall continuously result in severe floods and inundations, it is necessary to detect the trends and develop the spatial variation of extreme rainfall over the entire HCMC, which can be used for the city government in urban planning or infrastructure design purpose.

3.3 Data

In this study, the daily rainfall data are carefully selected from the National Hydro–Meteorological Service (NHMS) of Vietnam. The stations which had numerous days of

missing data in a year are excluded from this study. Only rain gauges with longer rainfall records are selected for this study because the length of data records has a significant influence on the accuracy of parameter estimation of the extreme value distribution (Yilmaz et al., 2016). In particular, the dataset contains 8 stations within HCMC and 10 stations outside HCMC (i.e. Binh Duong, Dong Nai, Ba Ria Vung Tau, Long An and Tay Ninh provinces) that have long-term precipitation observations and spread over the entire study area (Figure 3. 1) is used. The average record length is 40 years. The details of selected stations, including name, location and the length of data, are shown in Table 3. 1.

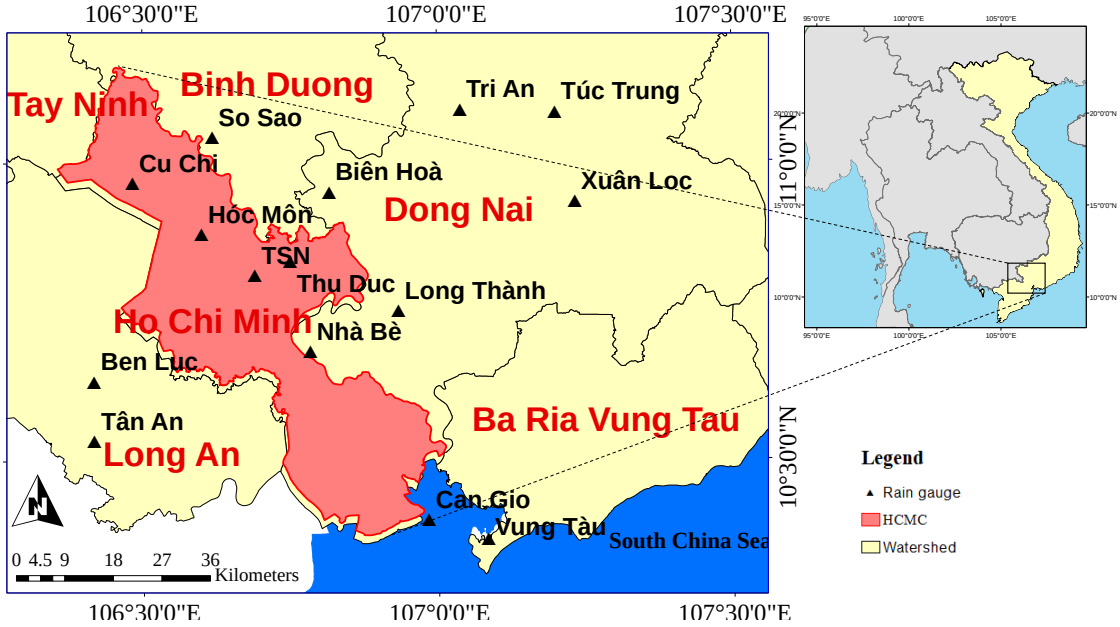
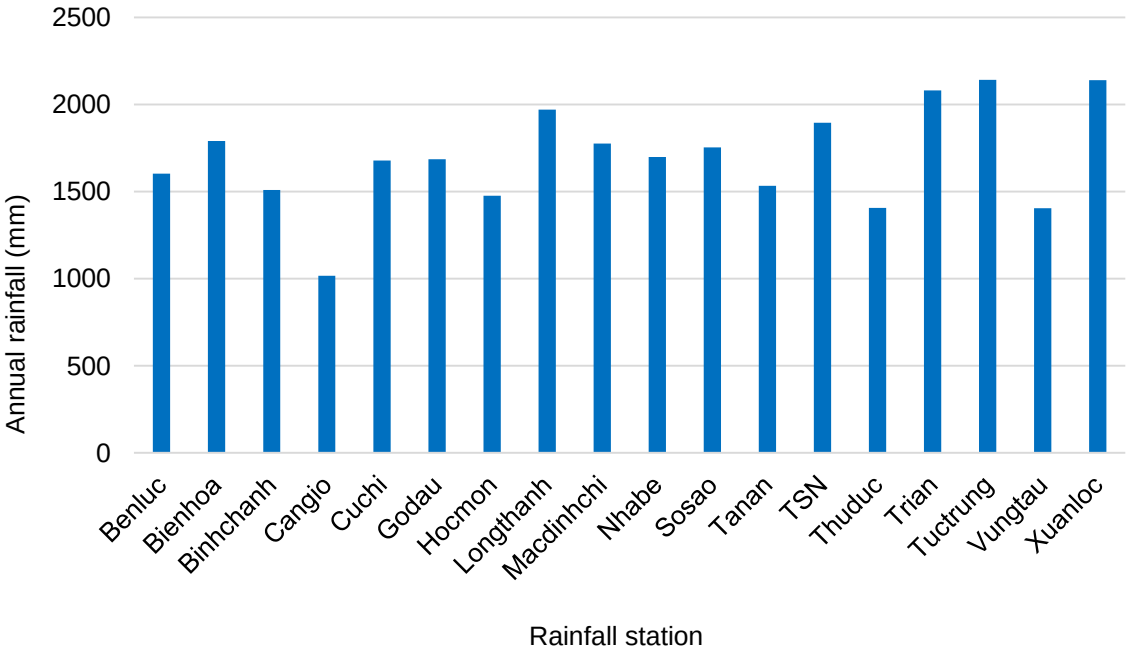


Figure 3. 1: Study area and locations of selected rainfall stations

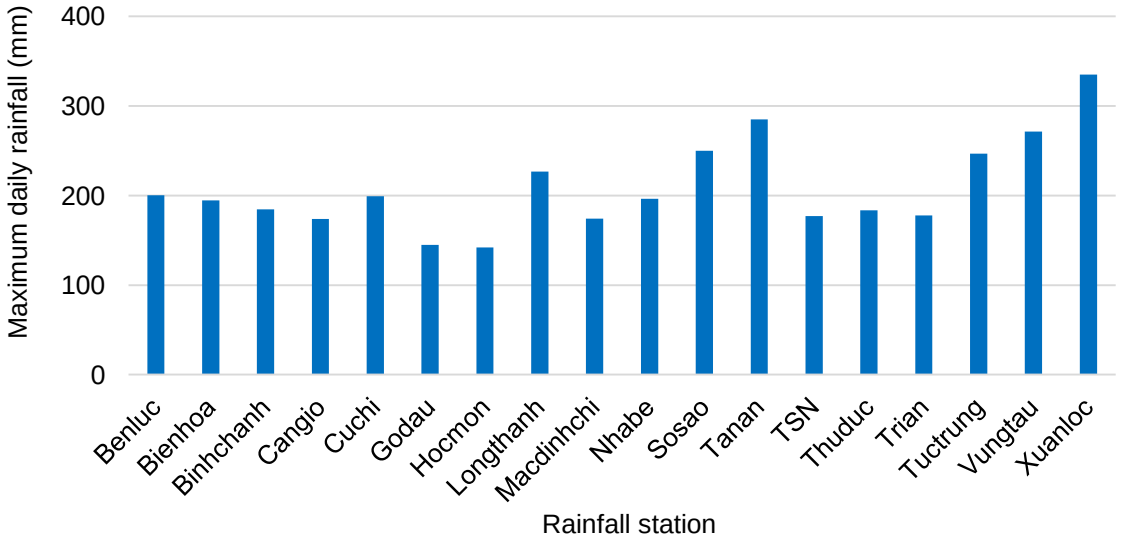
Table 3. 1: Details of selected rain gauges for the study

Region	Station name	Degrees, minutes and seconds		Data period
		Longitude	Latitude	
HCMC	Cuchi	106°29'00"	10°58'00"	1980-2014
	Hocmon	106°36'00"	10°53'00"	1980-2014
	Thuduc	106°45'00"	10°50'00"	1980-2014
	Macdinhchi (MDC)	106°42'01"	10°47'02"	1980-2014
	Tansonnhat (TSN)	106°40'00"	10°49'00"	1956-2014
	Binhchanh	106°44'00"	10°44'00"	1980-2014
	Nhabe	106°47'00"	10°41'00"	1980-2014
	Cangio	106°59'00"	10°24'00"	1980-2014
Binh Duong	Sosao	106°37'07"	11°02'33"	1958-2013
Tay Ninh	Godau	106°12'00"	11°09'36"	1980-2014
Dong Nai	Bienhoa	106°49'30"	10°57'25"	1958-2015
	Longthanh	106°35'00"	11°49'00"	1980-2014
	Tuctrung	107°12'00"	11°05'00"	1978-2015
	Xuanloc	107°14'00"	10°56'00"	1949-2015
	Trian	107°02'22"	11°05'15"	1980-2014
Long An	Benluc	106°25'00"	10°38'00"	1980-2014
	Tanan	106°25'00"	10°32'00"	1980-2014
Ba Ria Vung Tau	Vungtau	107°05'00"	10°22'00"	1949-2015

Figure 3. 2a provides the average annual rainfall of all selected stations. It can be seen that there is a wide variation in the average annual rainfall amount between rain gauges. The stations located in Dong Nai province have the high values of annual rainfall, around 2,000 mm (e.g. Longthanh, Trian, Tuctrung and Xuanloc). Whilst the stations located within HCMC have the lower values of annual rainfall, especially in Cangio station. Figure 3. 2b shows the maximum daily rainfall over the surveyed period of all stations. It is observed that the highest daily rainfall above 300 mm occurred in Xuanloc station, whereas the lowest value of 140 mm occurred in Hocmon station. Figure 3. 3 shows the maximum annual daily rainfall time series with the linear trends of Tansonnhat (TSN) and Hocmon stations during the observed period. Through this figure, time variability of mean and standard deviation of the two series from neighboring stations can be seen. The dates of maximum daily rainfall of both these stations for a period of 1980-2014 are also provided in Figure A. 1.



(a)



(b)

Figure 3. 2: Diagram of a) average annual rainfall and b) maximum daily rainfall of all surveyed stations

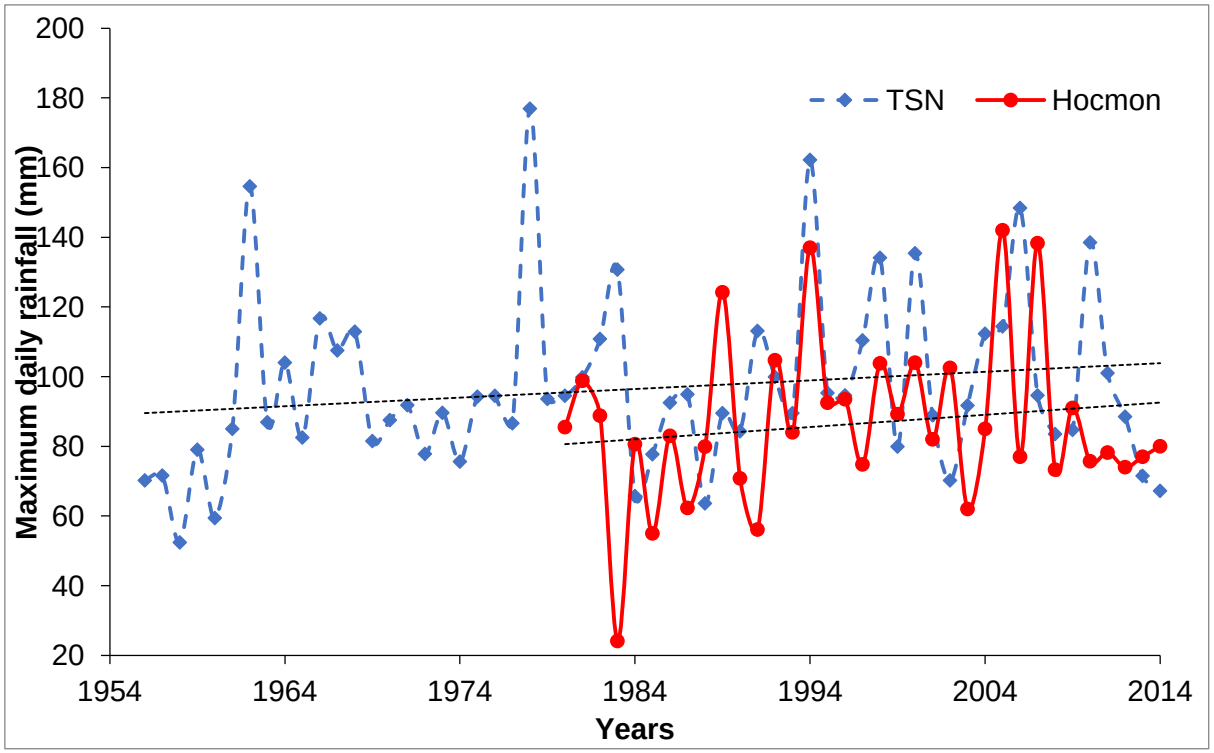


Figure 3. 3: The maximum annual daily rainfall time series with the linear trends (dotted lines) of Tansonnhat (TSN) and Hocmon stations during the observed period

3.4 Methodology

The methodology of this study is organized as follows. The nonstationarity in rainfall time series is firstly detected using statistical test. Then the nonstationary GEV models are developed in which location parameter is expressed as a function of non-linear trend. The best model for each station could be found by the AICc and the likelihood ratio test. Based on the best models, the spatial variation of extreme rainfall over in HCMC and adjacent areas are mapped corresponding to the return periods of 5, 25 and 50 years. The methodology flowchart is shown in Figure 3. 4.

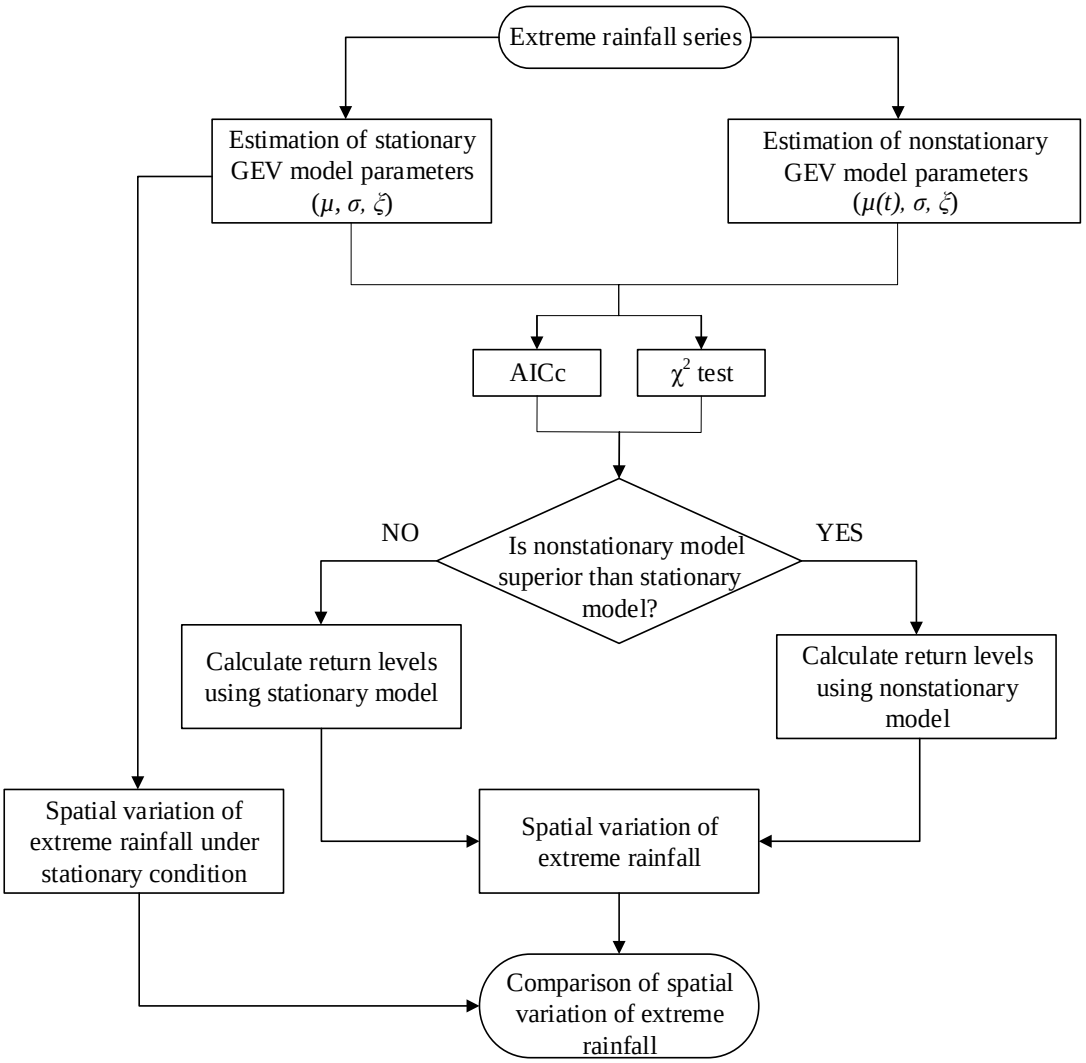


Figure 3. 4: Flowchart to determine the return levels

3.4.1 Trend test

Climatic extremes, particularly heavy rainfall events, have significantly increased in the past few decades due to human activities, urbanization and global climate change (Cheng and AghaKouchak, 2014, Berg et al., 2013, Trenberth, 2011, Min et al., 2011). Hence, the rainfall time series may have a nonstationary component. Normally, trend analysis is used to detect the nonstationarity signal in the rainfall time series. The M-K test (Mann, 1945, Kendall, 1962), a non-parametric statistical test, is widely used to analyze the monotonic trends in series of environmental data, climate data or hydrological data (Katz, 2013, Pohlert, 2016). The resultant M-K test statistic (*Tau*) indicates how strong the trend is and whether it is increasing or decreasing. It has been used by a number of researchers to access significant

trends in the extreme precipitation data (Rakhecha and Soman, 1994, Guhathakurta et al., 2011, Pingale et al., 2014, Zhang et al., 2008, Douglas and Fairbank, 2010, Westra et al., 2013). In addition, Sen's slope estimator is also used to validate the trend analysis in this study.

3.4.2 GEV model development

The asymptotic distribution of extreme rainfall series extracted using the annual maximum method is the GEV distribution and it has been previously used to model the extreme rainfall series under nonstationary condition (Agilan and Umamahesh, 2016a, Cheng et al., 2014, Villafuerte et al., 2015, Yilmaz et al., 2016). Therefore, in this study, the GEV distribution is used to model extreme rainfall series. Suppose $x = x_1, x_2, x_3, \ldots, x_n$ denote the annual maximum rainfall of n independent and identically distributed random variables. The cumulative distribution function of the GEV is given by Eq. (3.1) (Coles, 2001, Katz et al., 2002):

$$F(x;\mu,\sigma,\xi) = exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, \quad 1+\xi\left(\frac{x-\mu}{\sigma}\right) > 0, \sigma > 0 \qquad (3.1)$$

where μ, σ and ξ denote the location, scale and shape parameters. The location parameter (μ) specifies the center of the distribution, the scale parameter (σ) represents the size of deviations around the location parameter, and the shape parameter (ξ) governs the tail behavior of the GEV distribution. The GEV has three types of distribution determined by the sign of the shape parameter, i.e. Fréchet-type ($\xi > 0$), Weibull type ($\xi < 0$), and Gumbel type ($\xi = 0$).

Nonstationary GEV model

In the nonstationary case, the parameters of the models are allowed to vary with covariates (e.g. time or climate variables) (Coles, 2001). In this study, nonstationary models are developed considering time as covariate. In particular, the parameters of the GEV distribution are expressed as a function of time, known as $\mu(t)$, $\sigma(t)$, $\xi(t)$, and $t = 1,2,\ldots n$. For this study, the nonstationarity did not consider in scale (σ) and shape parameter (ξ). Because the precise estimation of ξ is difficult, and it is unrealistic to assume it as a smooth function of time (Coles, 2001). Besides, modelling temporal changes in σ and ξ reliably requires long-term observations which are usually unavailable for practical applications (Cheng et al.,

2014). Hence, under nonstationary condition, the parameters of GEV model are expressed as a function of the covariate [*f(t)*] (Eq. 3.2):

$$\mu(t) = \mu_0 + \mu_1 \times f(t); \qquad \sigma(t) = \sigma; \qquad \xi(t) = \xi \qquad (3.2)$$

where μ_1 denotes as the slope parameter and it represents the trend in the location parameter due to covariate *f (t)*.

3.4.3 Parameters estimation

The method of maximum-likelihood has been widely used for estimating the parameters of nonstationary GEV model. The maximum-likelihood estimates of μ, σ and ξ are taken to be those values which maximize the likelihood function (Katz, 2013). For the nonstationary model, the likelihood function can be represented as a function of parameters (i.e. μ_0, μ_1, σ, ξ). Let x_1, x_2,…, x_n be annual maximum precipitation series of n years. The log-likelihood function can be written as follows:

For $\xi \neq 0$,

$$L(\mu, \sigma, \xi|X) = -n \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{n} \log \left[1 + \xi \left(\frac{x_i - (\mu_0 + \mu_1 \times f(T))}{\sigma}\right)\right] - \sum_{i=1}^{n} \left[1 + \xi xi - (\mu 0 + \mu 1 \times f(T) \sigma - 1 \xi, 1 + \xi xi - (\mu 0 + \mu 1 \times f(T) \sigma > 0\right.$$

(3.3)

For $\xi = 0$,

$$L(\mu, \sigma|X) = -n \log \sigma - \sum_{i=1}^{n} \left(\frac{x_i - (\mu_0 + \mu_1 \times f(T))}{\sigma}\right) - \sum_{i=1}^{n} \exp\left[-\left(\frac{x_i - (\mu_0 + \mu_1 \times f(T))}{\sigma}\right)\right] \quad (3.4)$$

For the purpose of optimization, minimization of negative log-likelihood (Katz, 2013) can be adopted to arrive at the estimates of parameters instead of maximizing log-likelihood. Therefore, minimization the negative log-likelihood function is used for parameters estimation in this study.

As mentioned earlier, extreme rainfall events are affected by global climate change and many physical processes which have a different periodicity. Besides, directly using time covariate based linear form in nonstationary modelling may create more bias. Agilan and Umamahesh (2016a) suggested a non-linear form which is based on time and concerns the

incorporation of both long-term trend and periodicity concurrently. The covariate equation is given as follows (Agilan and Umamahesh, 2016a):

$$f(t) = t^{a_1} + sin(a_2 \times t) + cos(a_3 \times t) \quad (3.5)$$

where a_1 is the variable that controls the magnitude of the long-term trend, and variables a_2 and a_3 control the periodicity. The estimation of a_1, a_2 and a_3 values is implemented by the MOGA. The genetic algorithm (GA), one of the efficient global search methods, is a computerized search and optimization algorithm based on the mechanics of natural genetics and natural selection. Unlike single objective problem, multi-objective problem based on multiple objectives functions is usually considered for estimating robust solution. In this study, MOGA is used to estimate the value of a_1, a_2 and a_3. As such, AICc, a distance function based on correlation coefficient (r) and root mean square error (RMSE) are considered as three objective functions of the GA is used to identify the best solution. They are given equations (3.6), (3.7) and (3.8). The AICc will measure the bias of the nonstationary model while other two functions wil quantify the quality. Minimizing all three objective functions will produce optimal values of variables a_1,a_2 and a_3.

$$F_1 = AICc \quad (3.6)$$

$$F_2 = (1 - r) \times 100 = \left(1 - \frac{\sum_{i=1}^{n}(E_i - \bar{E})(M_i - \bar{M}}{\sqrt{\sum_{i=1}^{n}(E_i - \bar{E})^2}\sqrt{\sum_{i=1}^{n}(M_i - \bar{M})^2}} \right) \times 100 \quad (3.7)$$

$$F_3 = \frac{1}{n}\sum_{i=1}^{n}(E_i - M_i)^2 \quad (3.8)$$

where, E_i and M_i are empirical and model quantile, $\bar{E}$ and $\bar{M}$ are the mean of empirical and model quantiles, respectively.

The flowchart of optimizing the values of a_1,a_2 and a_3 is presented in Figure 3. 5.The package nsga2R in R programming language is used to perform MOGA in this study.

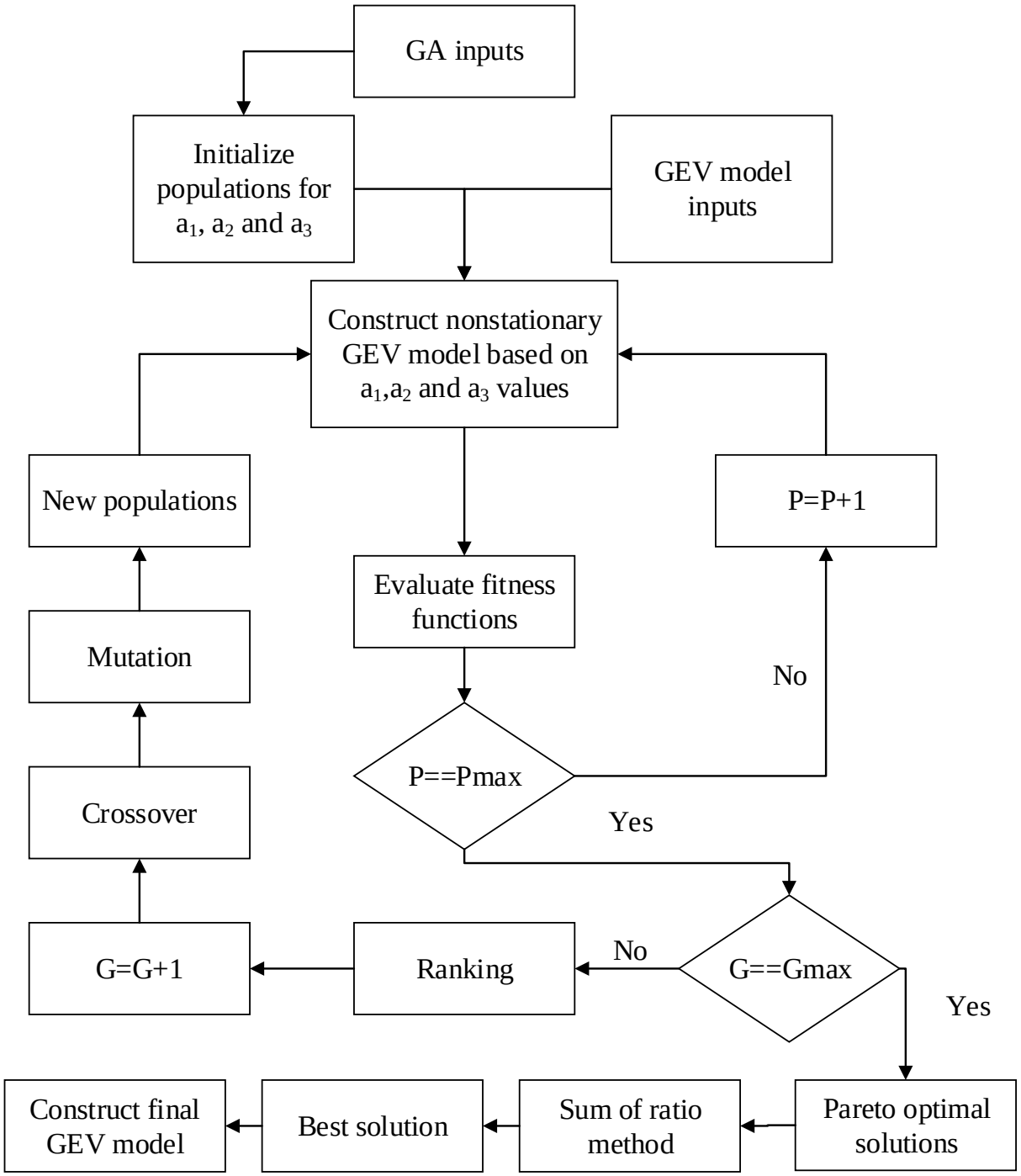


Figure 3. 5: Flowchart for optimizing the values of a_1,a_2 and a_3 (Agilan and Umamahesh, 2016a)

3.4.4 Model selection

Selection of the best model is a complex process and need to consider several different measures. A single measure may fail to determine an appropriate model, which lead to

underestimate or overestimate the probability of extreme rainfall. In this part of thesis, two measures are used to identify the best model for annual maximum extreme rainfall. Besides, the graphical approach (The probability-probability and quantile-quantile plots) is also used to check the quality of fitted model.

The Akaike Information Criterion

The AIC has been used commonly to select the best model among candidate models. Since there is a tendency for a model with more number of parameters to perform better. AIC is normally used as a performance measure when models with different number of parameters are being compared. In the comparison of AIC values between various models, the model with the lowest AIC value is considered to be the most efficient, and hence should be selected. However, Hurvich and Tsai (1995) showed that the AIC may have serious deficiencies, and they recommend a corrected version, namely AICc, which was developed for small samples to mitigate the bias and avoid overfitting the data. Thus, AICc is used for selecting the appropriate model in this study. The AICc is given by Eq. 3.9:

$$AICc = -2L(\Theta|X) + 2k + \frac{2k(k+1)}{n-k-1} \tag{3.9}$$

where n is the sample size, k is the number of parameters in a given model, $-\log L(\Theta|X)$ is the minimized negative log-likelihood function. In addition, Burnham and Anderson (2004) suggested a corrected version of AICc, (denoted Δ_i) which is used in this work for ranking and comparison among the GEV models.

$$\Delta_i = \text{AICc} - \min(\text{AICc}) \tag{3.10}$$

where min(AICc) is the smallest value of AICc among candidate models. The model having $\Delta_i = 0$ is considered as the best model, whereas the model with larger Δ_i is less plausible. The model which has $\Delta_i \leq 2$ is considered reasonable selection for the given time series (Burnham and Anderson, 2004, Agilan and Umamahesh, 2016a).

The likelihood ratio test

The likelihood ratio test allows to determine the significance of the trend parameter in the nonstationary model by comparing negative log-likelihood of stationary and nonstationary

models (Katz, 2013). For example, if the stationary model and nonstationary model are denoted by *(M_0)* and *(M_1)* respectively, the negative log-likelihood of model *(M_0)* and *(M_1)* can be written:

$$l_0(M_0) = -\log L(x_1, x_2, \dots, x_n;\ \mu, \sigma, \xi) \tag{3.11}$$

$$l_1(M_1) = -\log L(x_1, x_2, \dots, x_n; a_1, a_2, a_3, \mu_0,\ \mu_1, \sigma, \xi) \tag{3.12}$$

Under null hypothesis of no trend ($\mu_1 = 0$), the likelihood ratio test statistic, based on twice the difference between $l_0(M_0)$ and $l_1(M_1)$ (Eqs. (3.11) and (3.12)), approximately follows Chi-square distribution with four degree of freedom [denoted by χ^2 (4)] as the different between number of parameters in model *(M_0)* and *(M_1)* (Katz, 2013). The test is given by:

$$2\{l_0(M_0) - l_1(M_1)\} \sim \chi^2\ (4) \tag{3.13}$$

The graphical diagnostics

In order to check the quality of fitting for a chosen model, the graphical approach of the probability–probability plot (P-P) and quantile–quantile plot (Q-Q) can be used. To develop the P-P and Q-Q plots, it is necessary to transform the data into a standardized form (Coles, 2001). Here, the standard Gumbel distribution is used, and the transformed variable is defined by (Katz, 2013, Coles, 2001):

$$\varepsilon_i = \frac{1}{\hat{\xi}} \log\left(1 + \hat{\xi}\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right)\right) \tag{3.14}$$

where $\hat{\mu}$, $\hat{\sigma}$, $\hat{\xi}$ are estimated location, scale and shape parameter. Let $\hat{\varepsilon}$ is order value of ε, the P-P plot points and the Q-Q plot points are given by Eq. (3.15) and Eq. (3.16) respectively (Coles, 2001)

$$\left(\frac{i}{n+1}, exp\left(-exp(-\hat{\varepsilon})\right)\right) \tag{3.15}$$

$$\left(-log\left(1 - \frac{i}{(n+1)}\right), \hat{\varepsilon}\right) \tag{3.16}$$

3.4.5 Return level estimation

Once the best model for the given extreme rainfall series is determined, the *T*-year return level z_p corresponding to the *T*-year return period can be obtained. Here, the location parameter value in the nonstationary model varies over time. Some authors suggested the low-risk approach for calculating location parameter by taking the 95 percentiles of the location parameter values in historical observation (Cheng and AghaKouchak, 2014, Agilan and Umamahesh, 2016a). However, in this study, two ways are used to compute return levels by using the mean value of location parameter for the years 1980-1984, say first five years, and the period of 2010-2014, say last five years. It is the fact that the rainfall process has periodicity as it is controlled by many physical processes (such as the ENSO cycle). As mentioned earlier, the rainfall of the study area is controlled by one of the dominant teleconnections, i.e. ENSO cycle. The El Nino and La Nina events (positive and negative phases of the ENSO cycle) will occur once in 2 to 5 years. In other words, the periodicity of the ENSO cycle is less than 5 years. Besides, a study related to nonstationarity analysis of Chawla and Mujumdar (2018) suggested that the hydrologic cycle in a river catchment can change every 5-year period. Consequently, the average of 5 years is calculated thereby the interannual variations in the rainfall can be eliminated. This concept allows comparing the difference between the first and the last periods of rainfall values at a certain station. Estimation of the *T*-year return level for the first (last) five-year period can be given by Eq. (3.17):

$$z_{pF=}\begin{cases}\hat{\mu}_{5F(5L)} - \frac{\hat{\sigma}}{\hat{\xi}}\left[1 - \left\{-log\left(1-\frac{1}{T}\right)\right\}^{-\hat{\xi}}\right], & for\ \hat{\xi} \neq 0\\ \hat{\mu}_{5F(5L)} - \hat{\sigma} log\left\{-log\left(1-\frac{1}{T}\right)\right\}, & for\ \hat{\xi} = 0\end{cases} \qquad (3.17)$$

where $\hat{\mu}_{5F}$ and $\hat{\mu}_{5L}$ are the mean value of the location parameter of first five years and last five years respectively. By substituting the values of estimated parameters into Eq. (3.17), the estimates of the return levels can be obtained.

3.5 Results and discussions

3.5.1 Trends in extreme rainfall

As mentioned earlier, the M-K test is applied to indicate the increasing or decreasing trend in the rainfall data. In Table 3. 2, the results of the M-K test with all rain gauges are shown. The negative value of *Tau* indicates decreasing trend, in contrast, the positive value of *Tau* indicates increasing trend. The decreasing trend has been found in 8 stations, of which 2 stations are inside HCMC (i.e. Thuduc and MDC) and 6 stations outside HCMC (i.e. Sosao, Godau, Tuctrung, Benluc, Tanan and Vungtau). The increasing trend is observed in the annual maximum precipitation series of remaining stations.

Table 3. 2: Results of M-K test for trend for all gauging stations

Stations	*Tau* value	*p*-value	Stations	*Tau* value	*p*-value
Cuchi	0.08	0.504	Godau	-0.15	0.227
Hocmon	0.02	0.898	Bienhoa	0.07	0.497
Thuduc	-0.15	0.218	Longthanh	0.17	0.156
MDC	-0.22	0.080	Tuctrung	-0.06	0.597
TSN	0.15	0.110	Xuanloc	0.16	0.071
Binhchanh	0.08	0.514	Trian	0.07	0.580
Nhabe	0.24	0.057	Benluc	-0.02	0.865
Cangio	0.38	0.001	Tanan	-0.05	0.660
Sosao	-0.12	0.221	Vungtau	-0.03	0.727

In addition, the *p*-value shown in Table 3. 2 points out that only 1 station in the Southeast side (i.e. Cangio) has a significant trend at 5% significant level. The trends in annual maximum series of Macdinhchi, Nhabe, Xuanloc stations are significant at 10% significance level. None of the test statistics of remaining stations is significant at 10% significant level. The results from Sen's slope estimator (Table 3.3) is almost similar to M-K test.

Table 3. 3: Results of Sen's slope estimator for all gauging stations

Stations	Sen's slope value	*p*-value	Stations	Sen's slope value	*p*-value
Cuchi	0.41	0.50	Godau	-0.46	0.22
Hocmon	0.08	0.89	Bienhoa	0.08	0.89
Thuduc	-0.77	0.22	Longthanh	0.78	0.16
MDC	-0.84	0.08	Tuctrung	-0.56	0.15
TSN	-0.04	0.89	Xuanloc	0.03	0.98
Binhchanh	0.29	0.51	Trian	0.25	0.58
Nhabe	0.84	0.06	Benluc	0.06	0.89
Cangio	2.16	0.001	Tanan	-0.75	0.20
Sosao	0.07	0.79	Vungtau	-0.96	0.06

Some researchers found that the characteristic of rainfall is changing (Westra et al., 2013, Min et al., 2011). It includes the increases in extreme rainfall in most places of the world, although only limited rain gauges indicate a statistically significant nonstationary behavior. Cheng and AghaKouchak (2014) also mentioned that ignoring the nonstationarity may lead to remarkable underestimation of extreme events, which may result in the increase in the risk of infrastructure design and construction. Besides, these authors also suggested that the nonstationary condition can be applied to all datasets regardless of their trend, avoiding a subjective significance measure. Hence, the nonstationary GEV model is constructed for all 18 stations in this study.

3.5.2 Nonstationary GEV model

Before developing the stationary and nonstationary GEV models for extreme rainfall analysis, it is required to determine the value of variables a_1, a_2 and a_3 in Eq. (3.5). Upon estimating the value of variables a_1, a_2 and a_3 by MOGA, the stationary and nonstationary GEV models are constructed for each station. The values of a_1, a_2 and a_3 are shown in Table 3.4 for all stations.

Table 3. 4: Estimated values of a_1, a_2 and a_3

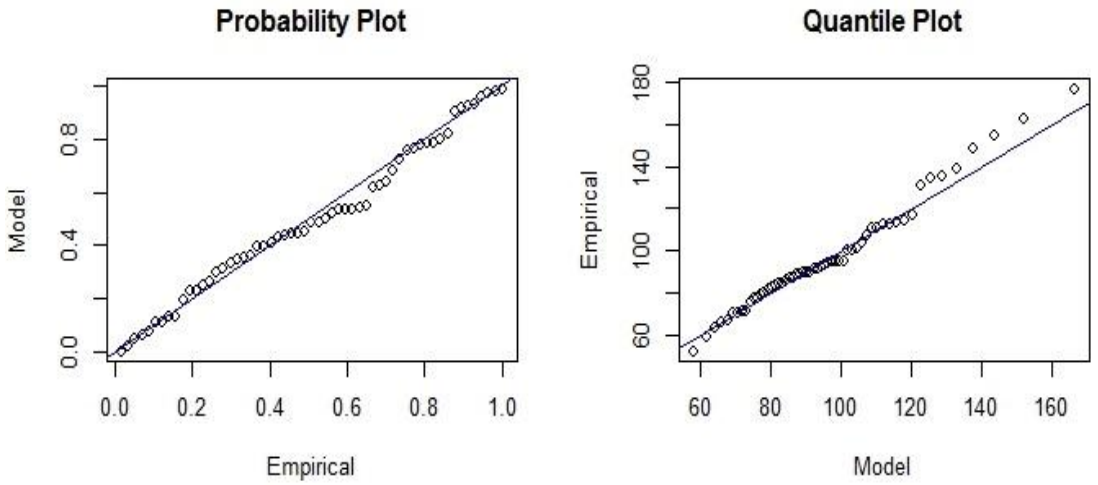
Stations	Variables			Stations	Variables		
	a_1	a_2	a_3		a_1	a_2	a_3
Cuchi	-1.87	5.13	8.31	Godau	-0.92	-5.94	6.04
Hocmon	-4.01	2.99	2.45	Bienhoa	-0.75	7.24	9.83
Thuduc	-0.67	-5.24	6.44	Longthanh	-1.87	6.76	1.40
MDC	-2.35	9.31	2.51	Tuctrung	-10.00	7.61	0.43
TSN	0.15	-0.48	9.18	Xuanloc	-3.77	-4.92	8.30
Binhchanh	-8.32	-8.86	1.85	Trian	-0.87	-9.13	8.00
Nhabe	-0.88	4.76	4.10	Benluc	-10.00	-5.11	1.44
Cangio	-0.29	-6.14	6.16	Tanan	-1.84	1.34	6.02

Sosao	-9.28	7.03	7.83	Vungtau	-1.56	6.50	3.49

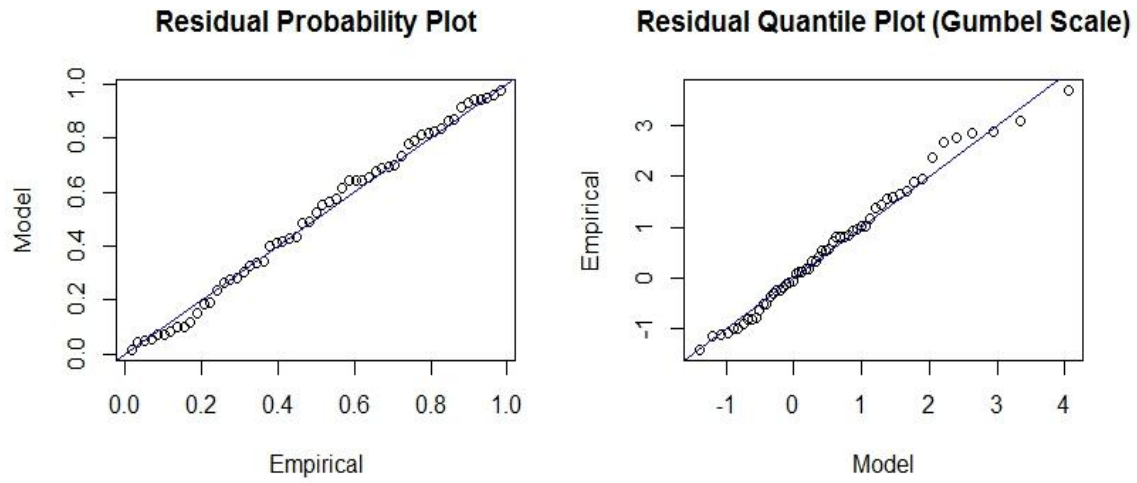
Table 3.5 shows the estimated parameters value of two models along with Δ_i value and likelihood ratio test results. The results show that the nonstationary GEV model is the best model for all stations. Further, as mentioned in the methodology section, the P-P and Q-Q plots are used to check the quality of a fitted model. The diagnostic plots of TSN station are shown in Figure 3.6. It is observed that the nonstationary model shows a better match (Figure 3. 6b) than the stationary model (Figure 3. 6a). The P-P and Q-Q plots for Benluc, Cangio, Nhabe and Xuanloc which are shown in Figure A. 2, A. 3, A. 4 and A. 5 respectively, also indicate a good fit of extreme rainfall data by nonstationary GEV models.

Table 3. 5: Parameter estimates, likelihood ratio test results and Δ_i values

Stations	Parameters value of stationary model				Parameters value of nonstationary model					Likelihood ratio test
	μ	σ	ξ	Δ_i	μ_0	μ_1	σ	ξ	Δ_i	*p*-value
Cuchi	84.32	24.80	0.09	3.08	83.66	19.01	17.53	0.35	0.00	5.97E-03
Hocmon	77.74	22.93	-0.22	3.09	79.55	13.93	18.78	-0.26	0.00	5.95E-03
Thuduc	79.28	29.18	-0.07	4.62	79.92	20.23	20.75	0.11	0.00	2.85E-03
MDC	92.32	23.15	-0.04	5.65	94.06	14.56	16.48	0.08	0.00	1.58E-03
TSN	85.27	19.42	0.01	15.84	84.51	10.23	13.12	0.31	0.00	3.68E-05
Binhchanh	71.07	21.97	0.07	5.07	70.20	14.06	14.36	0.39	0.00	2.33E-03
Nhabe	77.27	23.80	0.06	0.32	74.72	-10.06	15.87	0.44	0.00	1.52E-02
Cangio	41.67	30.25	0.17	11.29	45.14	-22.58	20.17	0.32	0.00	1.48E-04
Sosao	87.66	25.67	0.01	9.47	87.45	-13.70	19.32	0.18	0.00	6.48E-04
Godau	75.03	20.78	-0.13	18.72	73.47	-16.71	10.03	0.39	0.00	4.68E-06
Bienhoa	86.02	20.34	0.10	10.76	86.66	-10.82	15.26	0.25	0.00	3.45E-04
Longthanh	90.00	29.56	0.03	1.79	91.66	-16.51	23.20	0.11	0.00	1.05E-02
Tuctrung	98.63	22.83	0.10	7.92	97.46	-12.01	14.99	0.40	0.00	8.04E-04
Xuanloc	88.79	24.23	0.15	10.54	87.21	11.02	18.01	0.39	0.00	4.62E-04
Trian	91.65	17.45	-0.02	1.82	91.98	-11.24	13.27	0.12	0.00	1.04E-02
Benluc	88.88	24.63	0.12	5.65	87.55	-10.63	17.47	0.39	0.00	1.91E-03
Tanan	78.28	24.12	0.08	4.52	76.30	-17.06	15.79	0.40	0.00	3.12E-03
Vungtau	80.66	24.10	0.16	12.56	81.21	11.23	18.76	0.30	0.00	1.91E-04



(a) Stationary model



(b) Nonstationary model

Figure 3. 6: P-P and Q-Q plots of Tansonnhat (TSN) station

3.5.3 Spatial variation of rainfall extremes

Figure 3. 7 shows the annual maximum daily rainfall return levels corresponding to 5, 25 and 50 years return periods over HCMC. In the nonstationary case, extreme rainfall values

of each station are estimated for the periods of first five years and last five years by using mean value of location parameter for the years of 1980-1984 and 2010-2014 respectively.

Under the nonstationary condition, there is a large variation among rain gauges for the first five years' period compared to current years (the last five years' period). For instance, the estimated return levels corresponding to the 5-year return period vary between 55 to 210 mm per day for the first five years' period, whereas those values for the last five years' period are within range of 100-144 mm per day.

Figure 3. 7 also shows that the differences in estimate rainfall values derived from nonstationary models are more significant in Tanan, Cangio, Thuduc, Godau when comparing first five years and last five years periods. In more detail, the difference in extreme rainfall estimates between these periods reaches 66.5 mm for Tanan station corresponding to 50-year return period, while those values for Cangio, Godau and Thuduc are approximately 53 mm, 37 mm and 20 mm respectively. Remaining stations do not show significant differences in estimated extreme rainfall between concerned periods.

Besides, it can be seen that the values of the return levels based on the best nonstationary models have a significant difference compared with stationary models. In other word, the magnitudes of extreme rainfall under the stationary condition are much lower than those under the nonstationary condition for most of the stations, especially in the South and Northeast of the study area. Considering Tanan station as an example, the 50-year annual maximum daily rainfall values are 189 and 260 mm per day under stationary and nonstationary (for last five years' period) conditions respectively. These results point out that the assumption of stationarity could lead to underestimation of extreme rainfall events, hence choosing the design value for hydraulic structures under stationary or nonstationary condition should be considered thoroughly.

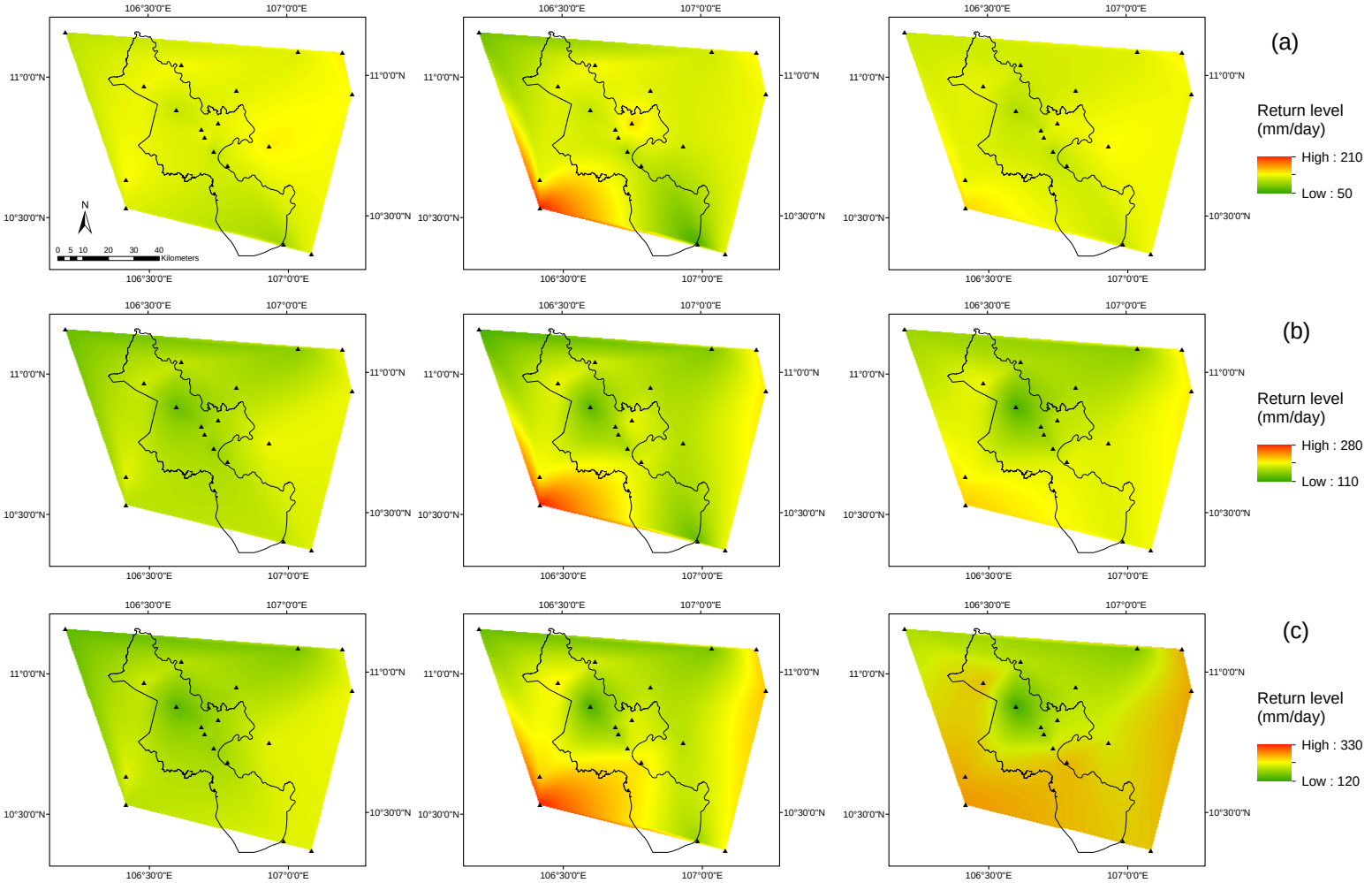


Figure 3. 7: The variation of return levels for the return period of (a) 5-year; (b) 25-year; (c) 50-year

3.6 Summary and conclusions

Under the influence of global climate change and physical processes as ENSO and PDO, extreme rainfall in HCMC has been proven to increase in frequency and magnitude in the last few decades. Therefore, the reality of nonstationary extreme rainfall should be paid more attention in the design of water infrastructure and flood mitigation projects since the extreme value distribution models with constant parameters may no longer be suitable. This present paper is aimed to model the spatial variation of extreme precipitation at HCMC and adjacent areas under nonstationary condition. In particular, the trend in time series is firstly determined by the non-parametric method, known as M-K test, before constructing nonstationary GEV model. In order to introduce nonstationarity into extreme rainfall frequency analysis, a nonlinear trend representing the long-term trend and periodicity of physical processes is suggested for the location parameter, whereas the scale and shape parameters (σ and ξ) are kept constant. The best nonstationary model for each station is found based on AICc value and likelihood ratio test. Based on the results of M-K test, only one station has a significant trend at the 5% significant level, and three stations have significant trend at 10% significant level. The remaining stations do not show a significant trend in the annual maximum precipitation data.

The findings indicate that the nonstationary model can be considered as the best model for modelling extreme rainfall when comparing to stationary model. The chosen nonstationary GEV model also has better goodness of fit performance. Moreover, the extreme rainfall estimates under the stationary condition are much lower than those under the nonstationary condition in a major part of study area.

Regarding two concerned periods, first five years and last five years, it can be seen that the differences in estimate rainfall values derived from nonstationary models are significant, especially in Tanan, Cangio, Thuduc, Godau stations.

In a nutshell, the present study is regarded as an important step towards flood mitigation projects and flood risk management in HCMC. Not only because it is the first of its kind, as authors' knowledge, but also because of the proposed method, which considers the impact of global climate change and physical processes on extreme rainfall of study area. In other word, it is suggested that extreme rainfall should be analysed under both stationary

and nonstationary condition before using as initial inputs of hydrological and hydrodynamic models since the global climate is continuously changing and unpredictable. The findings are also able to provide useful information on nonstationary extreme rainfall of HCMC for decision makers in choosing appropriate design values.

Chapter 4

Modelling nonstationary extreme water level considering local covariates

4.1 Introduction

Coastal cities are vulnerable to flood risk which has dramatically increased over the past decades, impacting millions of people every year across the world (Jongman et al., 2012, Hallegatte et al., 2013, Lasage et al., 2014). Especially, coastal cities in developing countries experiencing population, economic growth and urban expansion are experiencing more flooding (Hallegatte et al., 2013, Nicholls et al., 2008, Lasage et al., 2014, Adikari et al., 2010). Since flood exposure is continuously increasing in coastal areas, there is a growing demand of estimations of the magnitude and frequency of extreme water level for the design of coastal defense structure as well as flood risk management purpose.

As is shown in many papers, understanding the stochastic behavior of extreme water level is an essential need for planning and design flood defense structures (Arns et al., 2013b, Katz, 2013). And the assessment of extreme water level commonly included some form of statistical analysis based on the extreme value theory (Arns et al., 2015, Bulteau et al., 2015, Mudersbach and Jensen, 2010). The extreme value analysis can be based both on AM which normally utilize the GEV distribution, and POT in which the GPD is commonly recommended (Katz, 2013). The AM approach consists of modelling a sequence of maximum values taken from blocks or periods of equal value such as maximum daily rainfall (Sugahara et al., 2009). However, it seems an inappropriate approach if other data on extremes are available or one block happens to contain more extreme events than others (Coles, 2001). Moreover, a small sample is a critical problem in parameter estimation (Sugahara et al., 2009). In contrast, the POT approach is much more efficient as it considers all values exceeding a certain threshold, instead of simply choosing the maximum value. It means that a POT derived sample may not comprise only one or fixed number of events per year. Besides, since the water level objectives are normally expressed in term of certain critical thresholds (e.g. flood alarming rate), the POT method is apparently more suitable and better interpretable in the flood risks context.

Traditionally, these statistical methods based on the extreme value theory require the assumption of stationarity in hydrologic time series. However, under climate variability related to human activities, the extreme events in many places are proved to have nonstationary features in it. That is why the stationary condition may no longer be suitable, and the concept of nonstationary extreme value analysis has been improved and is used more frequently in analysis of extreme water level in low-land areas (Skjong et al., 2013, Serafin and Ruggiero, 2014, Arns et al., 2015, Mudersbach and Jensen, 2010, Méndez et al., 2007, Menéndez and Woodworth, 2010).

However, introducing nonstationarity into hydrologic frequency analysis should be performed with care, especially under environment changes due to the influences of climate change and human activities (Yan et al., 2017). The nonstationarity can be applied after careful testing using statistical or empirical analysis. Detection of trend is a critical issue as it has a significant impact on the nonstationary analysis results (Strupczewski et al., 2001, Šraj

et al., 2016). Once the trend in time series is detected, extreme water level can be modelled under nonstationary condition.

In most of the studies focusing extreme water level analysis under nonstationary approach, the parameters of the chosen distribution functions are commonly dependent on time (Coles, 2001, Obeysekera and Park, 2012, Mudersbach and Jensen, 2010). Some other papers showed that the parameters could vary in both time and several climatological variables (Katz et al., 2002, Méndez et al., 2007, Coles, 2001, Menéndez and Woodworth, 2010). Masina and Lamberti (2013) used regional climate indices such as the NAO and AO in analyzing the variability of the extreme water level along the Northern Adriatic coast. Serafin and Ruggiero (2014) conducted covariates analysis considering some climate indices, such as the Pacific/North American Pattern (PNA); the Southern Oscillation Index (SOI), and the Multivariate ENSO Index (MEI). Nevertheless, the water level shows unique characteristic as they are strongly impacted by local influences (Arns et al., 2013a). In other word, climate and catchment characteristics play an important role in generating water flows. Yan et al. (2017) suggested that the covariates selected for nonstationary modelling should have strong physical associations with the process of events (e.g. flood). Besides, the selected covariates for a particular area may have different effects in other geographical areas, hence covariates should be chosen with respect to the area of interest (Agilan and Umamahesh, 2016b). Therefore, instead of using global climatological variables, local variables which reflect the physical relationship to water level are used in this study area.

As is stated in previous chapter, HCMC is an example of an emerging coastal megacity dressing the increases of exposure levels to climate risks (Storch and Downes, 2011). HCMC appeared amongst the top ten most risk cities in term of exposure population (Lasage et al., 2014, Nicholls et al., 2008, Hallegatte et al., 2013, Storch and Downes, 2011, ADB, 2010, Dasgupta et al., 2011, World Bank, 2010). By 2070, this flood-prone city is expected to be in the top five cities in terms of population exposed to coastal flooding (Hanson et al., 2011, Storch and Downes, 2011). Although the city's government has made a lot of efforts to solve flooding and inundation issues, the situation has not improved significantly.

Located in the downstream of the Saigon-Dongnai River system, HCMC's topographic and geographic conditions make it extremely sensitive to various flood sources. Most of the studies showed that the common causes of flooding in this city are high tide, heavy rain, rapid urbanization growth and high discharge released from upstream reservoirs (Lasage et al., 2014, Storch and Downes, 2011, ADB, 2010, World Bank, 2010). The impact of each component on flooding is clearly different, and the way in which the components combine to generate water extremes also differs for different parts of HCMC.

There are several studies on flooding analysis in HCMC. However, the main limitation of previous studies is the focus on sea level and urbanization, hence normally ignoring the causes of rainfall and discharge released from the upstream. For example, Storch and Downes (2011) quantified current and future flood risks in HCMC by the combinations of urban development and sea level rise scenarios. Lasage et al. (2014) similarly focused on future scenarios concerning the sea level rise and urban growth, then suggested some different adaptive measures with benefit-cost analyses for only District 4 in HCMC. The Asian Development Bank (ADB, 2010) reported the systematical assessment climate-related risks at HCMC and suggested some adaptation plans for this city without analyzing the causes of floods thoroughly. Therefore, without a proper assessment of the causes of floods, quantifying inundation levels and managing these risks will be a problematic challenge for the local government once the flood risks in HCMC are increasing and are expected continuously.

Till now, flooding has become one of the most pressing issues such as in HCMC, it is necessary to investigate the best information on extreme water level, which may effectively help for engineering design and flood risk-based management. However, most existing flood control structures in HCMC have been built based on the assumption of stationary flood frequency analysis which may no longer be suitable for design purpose. In addition, in the authors' knowledge, the application of statistical distribution in analysis of variability of extreme water level events under nonstationary condition and their linkage to the causes of floods in HCMC could not be found in any papers so far.

The main objective of this study is to model the nonstationarity in the extreme water level in HCMC and identify the most significant physical processes which cause the nonstationarity in the series. Four local covariates (i.e. precipitation, mean sea level,

urbanization growth and outflows from upstream reservoirs) are considered due to their close relationship with extreme water level events. In addition, the stationary models for selected gauging stations are also constructed for comparison. The best model which comprises the most significant covariate(s) for each station is chosen based on the AICc and the likelihood ratio test. Finally, based on the best models, estimated extreme water level corresponding to different return periods are computed and compared with stationary case.

4.2 Data

4.2.1 Water level

In this study, the hourly water level data from four sites located along Saigon-Dongnai River (i.e. Phuan, Nhabe, Bienhoa, and Thudaumot) are used in the present analysis. The Phuan and Nhabe stations are located in HCMC, while the Bienhoa and Thudaumot (TDM) stations belong to adjacent areas. These data were provided by the NHMS of Vietnam and are not publicly available. These time series span through 34 years from 01-01-1981 to 31-12-2014 and there is no missing data. Figure 4. 1 shows HCMC and locations of these gauges. To give a visual change in extreme water level over time, the plots for annual maximum hourly water level over the period of 1981 to 2014 at four stations are shown in Figure 4. 2. It can be seen that the annual maximum hourly water level at all stations has positive trend toward the end of surveyed period.

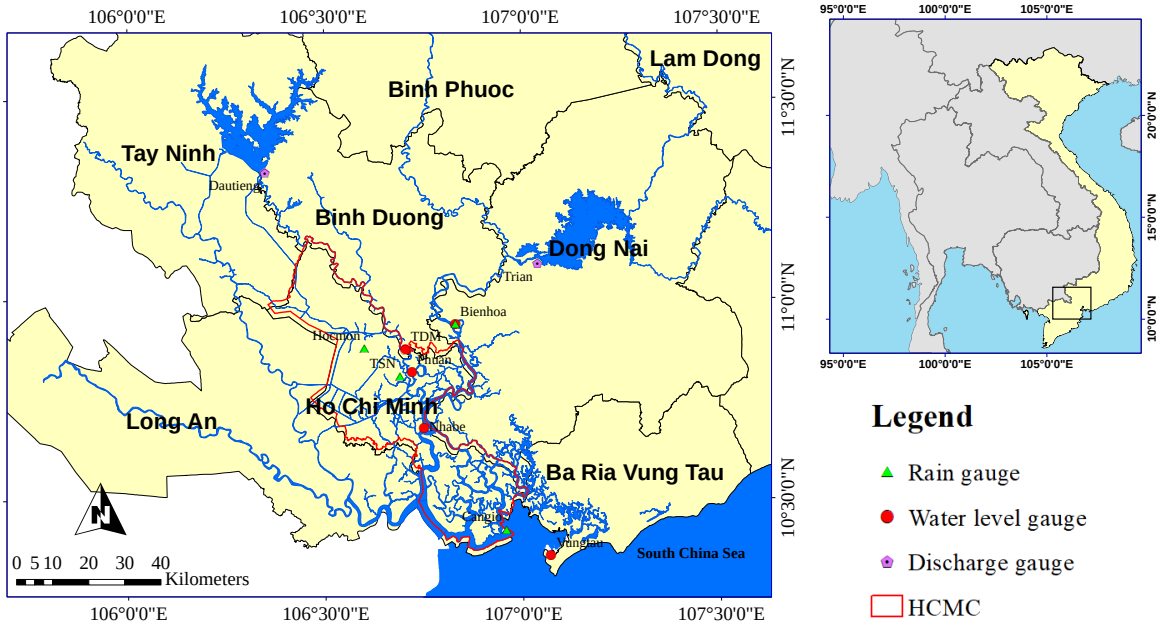


Figure 4. 1: HCMC and locations of gauging stations

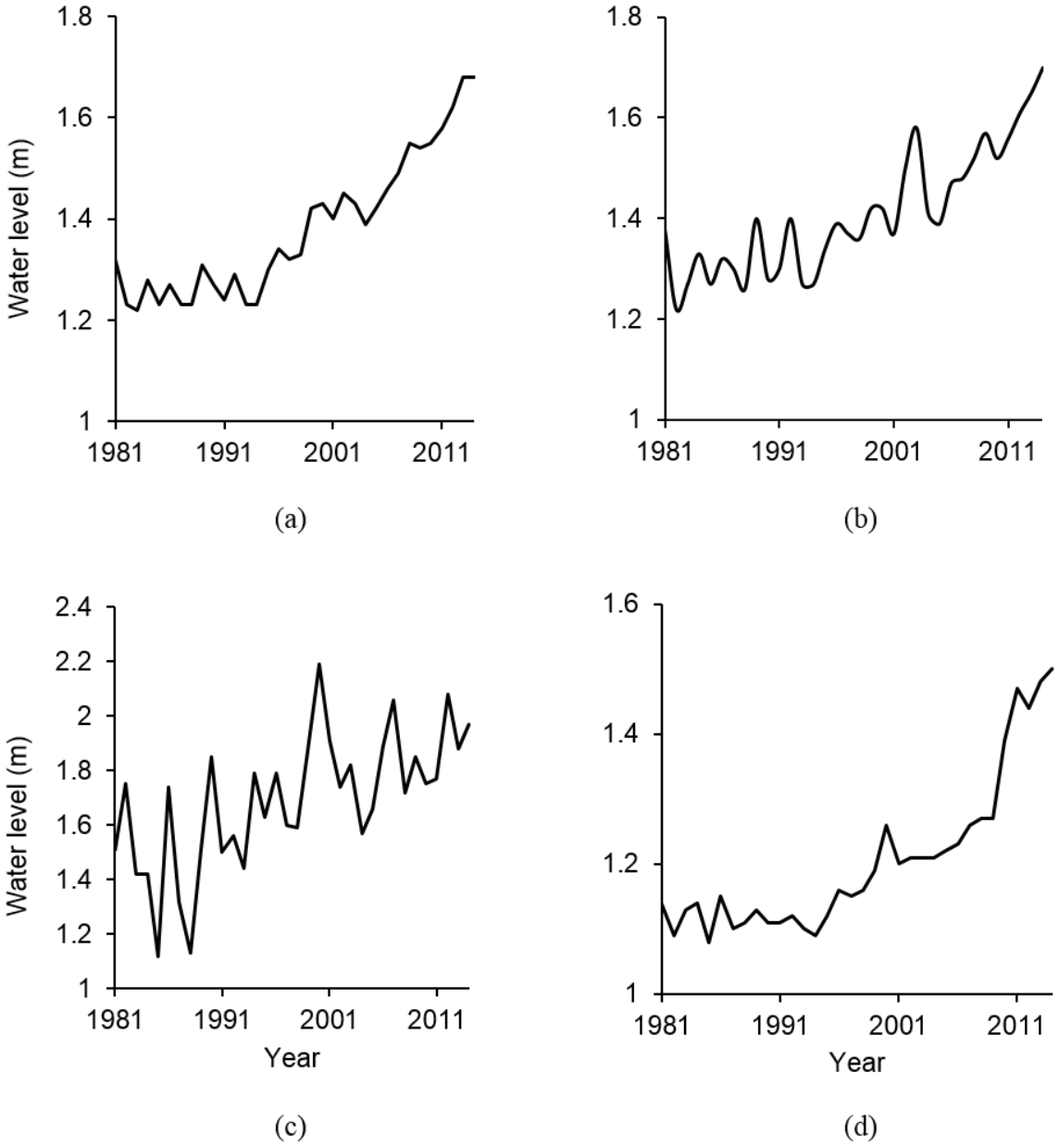


Figure 4. 2: Annual maximum hourly water level at (a) Phuan, (b) Nhabe, (c) Bienhoa, (d) TDM

4.2.2 Data for covariates

Precipitation

HCMC has a tropical monsoon climate with explicit wet and dry season variations in rainfall. Annual average rainfall is around 2,000 mm which mainly occurs in the rainy season

from May to October (approximate 90% of yearly rainfall). Rainfall with more intensity is estimated to become a major threat to the city (ADB, 2010).

Climate variability causes flood nonstationarity via the direct influence on precipitation patterns (Li et al., 2015). Besides, it is widely accepted by previous studies that using meteorological variables as covariates could be more effective and have clearer physical meaning for modelling nonstationary extreme hydrological events, especially for flood risk analysis (Du et al., 2015, Gilroy and McCuen, 2012, Šraj et al., 2016). For this reason, precipitation, a meteorological variable, is chosen as the first covariate in modelling extreme water level because of the close physical relationship between precipitation and water level. The daily precipitation data was provided by NHMS, and was chosen corresponding selected daily water level data. The selected gauge is the same or nearest corresponding to each water level station. Four selected rain gauges, i.e. Bienhoa, Hocmon, TSN, Cangio, with their locations are shown in Figure 4. 1.

Sea level

Major parts of HCMC are located in low-lying lands that are crossed by a complex network of canals and rivers connected to the South China Sea. This makes the city particularly prone to regular flooding and inundation linked to even just high tide (World Bank, 2010). In addition, the situation has become more serious since the mean sea level increased by 0.20 m during the past 50 years (Storch and Downes, 2011) and is expected to rise continuously in the future. Therefore, the choice of sea level as a covariate for extreme water level characteristics assessment is reasonable. The hourly observed sea level data is obtained from NHMS for the period of 1981 to 2014. This data is recorded at Vungtau station (107.05^{o}E and 10.22^{o}N), and it is not published.

Urbanization

In 2008, more than half of world's population lived in urban areas (UN-Habitat, 2011), and it is estimated that the towns and cities will make up 80 percent of the human population (UNFPA, 2007). As a result of urbanization, the citizens are now facing numerous problematic environment challenges such as floods, landslides and other natural disasters (UN-Habitat, 2011, UNFPA, 2007).

In last few decades, HCMC has become the most populous settlement and an important port city for Southeast Asia. The population of HCMC was approximately 7.1 million in 2010, that number is estimated to reach 10 million by 2020 (Labaeye et al., 2012, Storch and Downes, 2011). In order to meet the requirement of residential and industrial purposes, natural land surfaces have been replaced by artificial surfaces that may result in increasing the air temperature in urban areas (Shepherd et al., 2002) as well as magnitude and frequency of floods (Li et al., 2015). Moreover, the storage capacity of the city's water network has also reduced by construction activities which are particularly changing the natural flow of rivers and narrowing floodplain areas. Consequently, HCMC already experiences climate-related environment hazards caused by unsustainable urban development (Gravert and Wiechmann, 2016).

In this study, the built-up land of HCMC is determined from an urban structure type classification for the entire city, which is modeled from high-resolution Landsat remote sensing data captured from 1975 to 2016. The Landsat images used in this study were obtained from the U.S Geological Survey, with one Landsat 2 Multispectral Scanner (MSS) image, six Landsat 5 Thematic Mapper (TM) images and one Landsat 8 Operational Land Imager (OLI) image (Table 4.1). The Semi-Automatic Classification Plugin (Congedo, 2016) for QGIS (QGIS, 2015) is used for the purpose of the classification. In more detail, the Land Cover Signature Classification combined with the Spectral Angle Mapping algorithm are applied for classifying the satellite images. The land use land cover (LULC) maps created from satellite images are shown in Figure 4. 3. The LULC maps of 1988, 1995, 2004 and 2016 are shown in this thesis.

Table 4. 1: The results from satellite image classification

Satellite	Date of capture	Sensor	Resolution (m)	Built-up land (km^2)
Landsat 2	01-04-1975	MSS	60	37.34
Landsat 5	30-01-1988	TM	30	43.82
Landsat 5	02-02-1995	TM	30	60.38
Landsat 5	21-06-1999	TM	30	144.08
Landsat 5	11-12-2004	TM	30	179.17
Landsat 5	03-02-2007	TM	30	217.24
Landsat 5	04-02-2010	TM	30	238.88
Landsat 8	28-02-2016	OLI	30	278.96

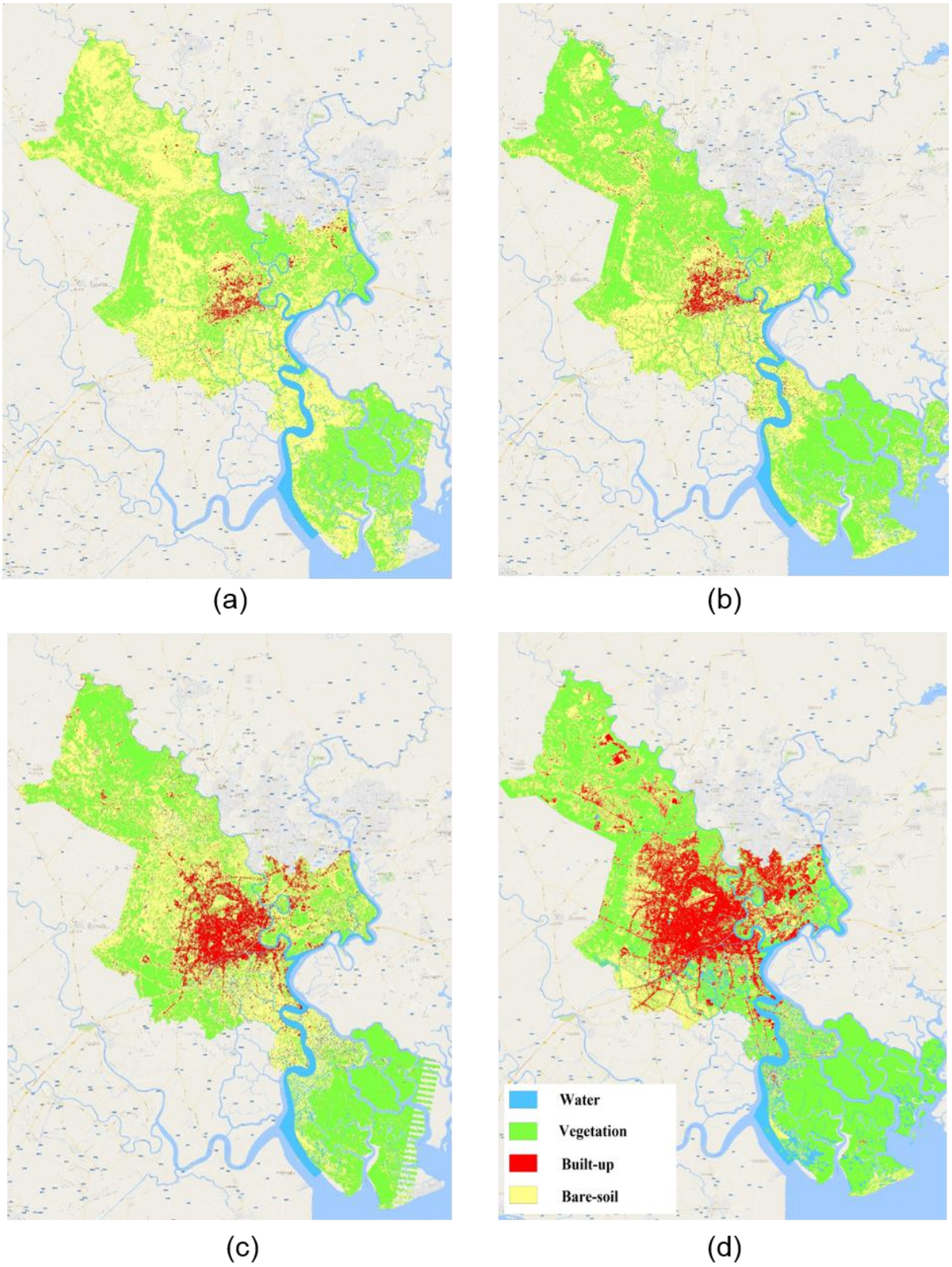


Figure 4. 3: LULC map of (a) 1988, (b) 1995, (c) 2004, (d) 2016

From the list of satellite images used and built-up land area extracted from the LULC maps (Table 4. 1), the growth of the urban area is modeled by a function as follows:

$$B = 7.023E\text{-}44e^{0.05212x} \quad (4.1)$$

where B is built-up land (km^2) and x is the corresponding year. In Eq. (4.1), the root mean square error (RMSE) of the model is 27.96 and the coefficient of determination (R^2) is 0.924. Based on Eq. (4.1), the built-up area of HCMC is calculated for each year between 1975 and 2016.

Upstream flows

The Saigon-Dongnai river, being one of the largest rivers in the south of Vietnam, plays an important role in social and economic development in HCMC. The upper watershed of the Saigon-Dongnai river which drains HCMC, is well-regulated with dams and reservoirs, i.e. the man-made Dautieng and Trian reservoirs. These reservoirs are the main source of energy and water supply for HCMC and surrounding areas (World Bank, 2010, Minh et al., 2007). Outflows from these reservoirs are connected to urban canals and then empty into the South China Sea.

In recent years, many authors have studied the possible effects of human activities on hydrological regimes in a certain river basin using statistical methods. To be more specific, they tried to develop nonstationary models to analyze flood characteristics in river basins considering the presence of reservoirs as a covariate (López and Francés, 2013, Zhang et al., 2015, Machado et al., 2015). Here, the released discharge from Dautieng and Trian reservoirs are considered as covariates to analyze extreme water level of downstream in our study. The outflow from Trian hydropower connects directly to three water level locations, i.e. Bienhoa, Phuan and Nhabe. Meanwhile, the outflow starting from Dautieng reservoir has relative influences on TDM, Phuan and Nhabe water level stations. The daily released discharge data is provided from Trian Hydropower Joint Stock Company and Dautieng-Phuochoa Limited Company for the period of 01-01-1981 to 31-12-2014.

4.3 Methodology

The flowchart of methodology used in this study is presented in Figure 4. 4. Firstly, extreme water level in four station are selected using peak over threshold technique. After that, nonstationarity in hourly water level time series identified using statistical test. Thirdly, based on four physical processes, i.e. rainfall, sea level, urbanization growth, outflows from upstream reservoirs and their combinations, the nonstationary models are developed. The stationary models are also developed for comparison purpose. And then, the best model which comprises the most significant covariate(s) for each selected location could be found based on the AICc and the likelihood ratio test. Finally, based on the best models, the extreme water level is estimated corresponding different return periods for all selected stations.

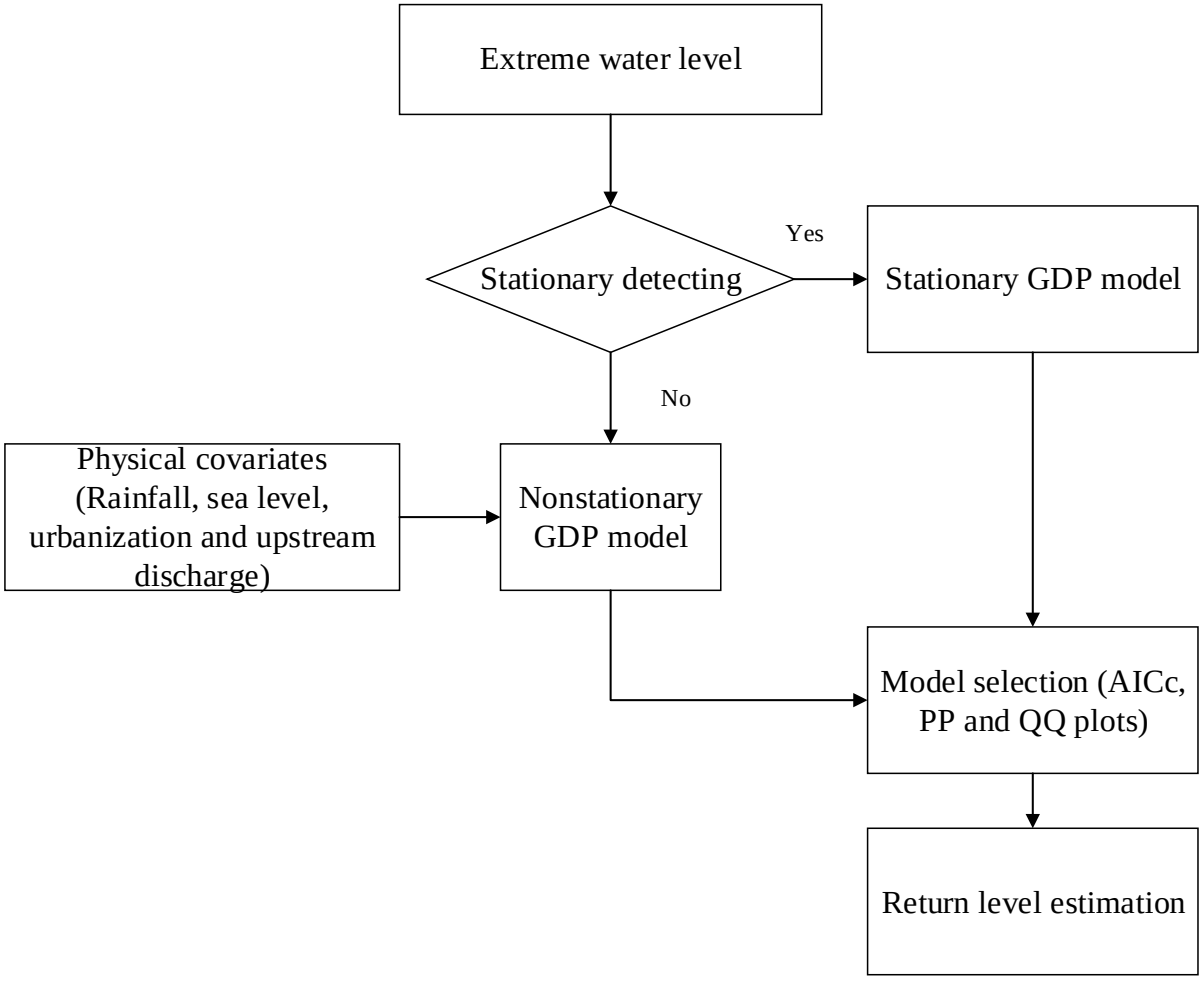


Figure 4.4: Flowchart for methodology

4.3.1 Trend detection

Currently, the concept of nonstationarity in hydrological time series is widely accepted by researchers and the society as well (Strupczewski et al., 2001). Checking nonstationarity should be a prior consideration due to its essential role which has a significant impact on the nonstationary analysis. In this study, the M-K test with a significant level of 0.05 is used to identify the possible trend in the extreme water level.

4.3.2 GPD models for extreme water level

As mentioned earlier, POT approach is apparently more suitable and better interpretable than AM approach in modelling the extreme water level values. Therefore, POT approach is chosen in this study based on its advantage.

The use of POT approach which is linked to the GPD is described as follow. Consider a sequence of n independent and identically distributed random variable $X_1, X_2, \ldots, X_n$ conditioned on $X > u$, where u is a given high threshold. For sufficient high threshold (u), the excess $Y_i = X_i - u$, conditional on $X_i > u$, has an approximate GPD with cumulative distribution function (CDF) defined as (Coles, 2001)

$$F(y;\alpha,\beta) = \begin{cases} 1-\left(1+\beta\frac{y}{\alpha}\right)^{-\frac{1}{\beta}}, & \sigma>0,\ 1+\beta(y/\alpha)>0 \\ 1-exp\left(\frac{-y}{\alpha}\right), & \alpha>0,\ \beta=0 \end{cases} \quad (4.2)$$

where α and β are the scale and shape parameters in that order. The scale parameter (α) is a function of chosen threshold u. Here, the nonstationarity is introduced only in scale parameter. Whereas, the shape parameter is kept constant because precise estimation of β is difficult, and it is unrealistic to assume that it is a smooth function of time (Coles, 2001).

In the nonstationary setting, the parameters are expressed as a function of covariate(s) in the general form

$$log\ \alpha(i) = \alpha_o + \alpha_1 \mathrm{R} + \alpha_2 \mathrm{S} + \alpha_3 \mathrm{U} + \alpha_4 \mathrm{ResT} + \alpha_5 \mathrm{ResD} \quad (4.3)$$

$\beta(i) = \beta$

where *i* denotes the hour; and R, S, U, ResT, ResD denote physical covariates, i.e. Rainfall (R), Sea level (S), Urbanization (U), Upstream discharge released from Trian reservoir (ResT) and Upstream discharge released from Dautieng reservoir (ResD). In the stationary model, the values of covariates R, S, U, ResT, ResD equal zero. The logarithm in Eq. (4.3) is taken to ensure the positive values of the scale parameter. The slope parameter α_1, α_2, α_3, α_4 and α_5 represent the trend due to effects of covariates R, S, U, ResT, ResD in that order. Based on five covariates and their combinations, thirty-one nonstationary models are constructed for Phuan and Nhabe stations, while fifteen nonstationary models are constructed for Bienhoa and TDM stations (Table 4. 2). In addition, an individual covariate or a combination which has significant impacts on the extreme water level in the study area can be found out.

Table 4. 2: Detail of GPD models constructed for extreme water level at the selected locations

Model	Description	Model	Description
Phuan and Nhabe stations			
GPD-0	$Y \sim GP(\alpha, \beta)$	GPD-16	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R + \alpha_2 S + \alpha_3 U), \beta)$
GPD-1	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R), \beta)$	GPD-17	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R + \alpha_2 S + \alpha_3 \text{ ResT}), \beta)$
GPD-2	$Y \sim GP(\exp(\alpha_0 + \alpha_1 S), \beta)$	GPD-18	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R + \alpha_2 S + \alpha_3 \text{ ResD}), \beta)$
GPD-3	$Y \sim GP(\exp(\alpha_0 + \alpha_1 U), \beta)$	GPD-19	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R + \alpha_2 U + \alpha_3 \text{ ResT}), \beta)$
GPD-4	$Y \sim GP(\exp(\alpha_0 + \alpha_1 \text{ ResT}), \beta)$	GPD-20	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R + \alpha_2 U + \alpha_3 \text{ ResD}), \beta)$
GPD-5	$Y \sim GP(\exp(\alpha_0 + \alpha_1 \text{ ResD}), \beta)$	GPD-21	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R + \alpha_2 \text{ ResT} + \alpha_3 \text{ ResD}), \beta)$
GPD-6	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R + \alpha_2 S), \beta)$	GPD-22	$Y \sim GP(\exp(\alpha_0 + \alpha_1 S + \alpha_2 U + \alpha_3 \text{ ResT}), \beta)$
GPD-7	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R + \alpha_2 U), \beta)$	GPD-23	$Y \sim GP(\exp(\alpha_0 + \alpha_1 S + \alpha_2 U + \alpha_3 \text{ ResD}), \beta)$
GPD-8	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R + \alpha_2 \text{ ResT}), \beta)$	GPD-24	$Y \sim GP(\exp(\alpha_0 + \alpha_1 S + \alpha_2 \text{ ResT} + \alpha_3 \text{ ResD}), \beta)$
GPD-9	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R + \alpha_2 \text{ ResD}), \beta)$	GPD-25	$Y \sim GP(\exp(\alpha_0 + \alpha_1 U + \alpha_2 \text{ ResT} + \alpha_3 \text{ ResD}), \beta)$
GPD-10	$Y \sim GP(\exp(\alpha_0 + \alpha_1 S + \alpha_2 U), \beta)$	GPD-26	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R + \alpha_2 S + \varepsilon_3 U + \alpha_4 \text{ ResT}), \beta)$
GPD-11	$Y \sim GP(\exp(\alpha_0 + \alpha_1 S + \alpha_2 \text{ ResT}), \beta)$	GPD-27	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R + \alpha_2 S + \alpha_3 U + \alpha_4 \text{ ResD}), \beta)$
GPD-12	$Y \sim GP(\exp(\alpha_0 + \alpha_1 S + \alpha_2 \text{ ResD}), \beta)$	GPD-28	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R + \alpha_2 S + \alpha_3 \text{ ResT} + \alpha_4 \text{ ResD}), \beta)$
GPD-13	$Y \sim GP(\exp(\alpha_0 + \alpha_1 U + \alpha_2 \text{ ResT}), \beta)$	GPD-29	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R + \alpha_2 U + \alpha_3 \text{ ResT} + \alpha_4 \text{ ResD}), \beta)$
GPD-14	$Y \sim GP(\exp(\alpha_0 + \alpha_1 U + \alpha_2 \text{ ResD}), \beta)$	GPD-30	$Y \sim GP(\exp(\alpha_0 + \alpha_1 S + \alpha_2 U + \alpha_3 \text{ ResT} + \alpha_4 \text{ ResD}), \beta)$
GPD-15	$Y \sim GP(\exp(\alpha_0 + \alpha_1 \text{ ResT} + \alpha_2 \text{ ResD}), \beta)$	GPD-31	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R + \alpha_2 S + \alpha_3 U + \alpha_4 \text{ ResT} + \alpha_5 \text{ ResD}), \beta$

(Table 4. 2 continued)

Model	Description	Model	Description
		Bienhoa station	
GPD-0	$Y \sim GP(\alpha, \beta)$	GPD-8	$Y \sim GP(\exp(\alpha_0 + \alpha_1 S + \alpha_2 U), \beta)$
GPD-1	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R), \beta)$	GPD-9	$Y \sim GP(\exp(\alpha_0 + \alpha_1 S + \alpha_2 \text{ ResT}), \beta)$
GPD-2	$Y \sim GP(\exp(\alpha_0 + \alpha_1 S), \beta)$	GPD-10	$Y \sim GP(\exp(\alpha_0 + \alpha_1 U + \alpha_2 \text{ ResT}), \beta)$
GPD-3	$Y \sim GP(\exp(\alpha_0 + \alpha_1 U), \beta)$	GPD-11	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R + \alpha_2 S + \alpha_3 U), \beta)$
GPD-4	$Y \sim GP(\exp(\alpha_0 + \alpha_1 \text{ ResT}), \beta)$	GPD-12	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R + \alpha_2 S + \alpha_3 \text{ ResT}), \beta)$
GPD-5	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R + \alpha_2 S), \beta)$	GPD-13	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R + \alpha_2 U + \alpha_3 \text{ ResT}), \beta)$
GPD-6	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R + \alpha_2 U), \beta)$	GPD-14	$Y \sim GP(\exp(\alpha_0 + \alpha_1 S + \alpha_2 U + \alpha_3 \text{ ResT}), \beta)$
GPD-7	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R + \alpha_2 \text{ ResT}), \beta)$	GPD-15	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R + \alpha_2 S + \alpha_3 U + \alpha_4 \text{ ResT}), \beta)$
		TDM station	
GPD-0	$Y \sim GP(\alpha, \beta)$	GPD-8	$Y \sim GP(\exp(\alpha_0 + \alpha_1 S + \alpha_2 U), \beta)$
GPD-1	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R), \beta)$	GPD-9	$Y \sim GP(\exp(\alpha_0 + \alpha_1 S + \alpha_2 \text{ ResD}), \beta)$
GPD-2	$Y \sim GP(\exp(\alpha_0 + \alpha_1 S), \beta)$	GPD-10	$Y \sim GP(\exp(\alpha_0 + \alpha_1 U + \alpha_2 \text{ ResD}), \beta)$
GPD-3	$Y \sim GP(\exp(\alpha_0 + \alpha_1 U), \beta)$	GPD-11	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R + \alpha_2 S + \alpha_3 U), \beta)$
GPD-4	$Y \sim GP(\exp(\alpha_0 + \alpha_1 \text{ ResD}), \beta)$	GPD-12	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R + \alpha_2 S + \alpha_3 \text{ ResD}), \beta)$
GPD-5	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R + \alpha_2 S), \beta)$	GPD-13	$Y \sim GP(\exp(\alpha_0 + \alpha_1 R + \alpha_2 U + \alpha_3 \text{ ResD}), \beta)$

GPD-6	$Y \sim GP\left(\exp\left(\alpha_0 + \alpha_1 R + \alpha_2 U\right), \beta\right)$	GPD-14	$Y \sim GP\left(\exp\left(\alpha_0 + \alpha_1 S + \alpha_2 U + \alpha_3 \text{ ResD}\right), \beta\right)$
GPD-7	$Y \sim GP\left(\exp\left(\alpha_0 + \alpha_1 R + \alpha_2 \text{ ResD}\right), \beta\right)$	GPD-15	$Y \sim GP\left(\exp\left(\alpha_0 + \alpha_1 R + \alpha_2 S + \alpha_3 U + \alpha_4 \text{ ResD}\right), \beta\right)$

The choice of an adequate threshold is a critical step for extreme value analysis in the POT method. If the selected threshold is too low, it is likely to violate the asymptotic basis of the model, leading to bias. If the threshold is too high, it will generate few excesses which may lead to high variance in the model estimate. Thus, threshold choice involves balancing bias and variance (Scarrott and MacDonald, 2012). Lang et al. (1999) mentioned about two different methods that can be adopted for threshold selection: the first one is based on physical criteria whereas the second one is based on mathematical and statistical considerations. The physical approach is usually used in river engineering in which the threshold is defined as the flood level for a specific river. Meanwhile, mathematical and statistic approaches are often recommended for coastal waters (Arns et al., 2013b). In this study, the thresholds for different water-level stations are chosen considering physical approach. For physical approach, the threshold values are selected based on flood alarming rate at each station. These rates are regulated by the Vietnam's government (PMV, 2011).

Water level series can exhibit dependencies, which are mostly related to the same meteorological forcing (Arns et al., 2013b). Therefore, in order to remove these temporal dependencies, some form of de-clustering is necessary (Coles, 2001). Firstly, for a given threshold *u*, a cluster is defined whenever there are consecutive exceedances of this threshold. These clusters are separated from each other by setting a minimum interval which is typically selected to be representative of storm duration on the site under study area (can be one or more days). Then the clustered extreme water level is de-clustered by retaining maximum values of each cluster.

4.3.3 Estimation of parameters

The method of maximum likelihood is used for estimating the parameters of GPD model in this study because it can be easily extended to the nonstationary case (Coles, 2001). Suppose that the values $y_1, y_2, \ldots, y_n$ are the *n* excesses of a threshold *u*. The log-likelihood function is given by (Coles, 2001) as Eq. (4.4).

$$L(\alpha,\beta) = -nlog(\alpha) - \left(1+\frac{1}{\beta}\right)\sum_{i=1}^{n} log\left(1+\frac{\beta y_i}{\alpha}\right), \beta \neq 0$$

$$L(\alpha) = -nlog(\alpha) - \frac{1}{\alpha}\sum_{i=1}^{n} y_i, \beta = 0 \quad (4.4)$$

Katz (2013) suggested that minimization of negative log-likelihood, for the purpose of optimization, can be adopted to arrive at the estimates of parameters instead of maximization. Therefore, minimization the negative log-likelihood function is used for estimating parameters α, α_0, α_1, α_2, α_3, α_4, α_5 and β in this study.

4.3.4 Model selection

As discussed in Section 3.4.4, the best model among different candidate models can be selected by the AICc value. Upon identifying the best model among candidate models (Table 4.2), the significance of the best model against the stationary model can be checked by the likelihood ratio test as shown in Section 3.4.4. Here, the AICc and the likelihood ratio test are used to identify the best model for extreme water level.

4.3.5 Estimation of return levels

The *T*-year return level is the level expected to be exceeded once every *T* years. Because scale parameter in nonstationary model varies with time, a low-risk approach (more conservative) is used by taking the 95 percentiles of the scale parameter values in historical observation ($\hat{\alpha}_{95}$= $Q_{95}(\hat{\alpha}_{t1}, \hat{\alpha}_{t2}, \ldots, \hat{\alpha}_{tn})$) to calculate the return level (Cheng et al., 2014). Let n_y be the number of observations per year. The *T*-year return level (Coles, 2001) is given by

$$z_{T=} \begin{cases} u + \frac{\hat{\alpha}_{95}}{\hat{\beta}}\left[\left(Tn_y\zeta_u\right)^{\hat{\beta}} - 1\right], \ for\ \hat{\beta} \neq 0 \\ u + \hat{\alpha}_{95}\log\left(Tn_y\zeta_u\right), \ for\ \hat{\beta} = 0 \end{cases} \quad (4.5)$$

for a given threshold *u*, where ζ_u= Pr $\{Y > u\}$ is the rate of the data exceeding *u*. The rate parameter has a natural estimator simply given by the number of exceeding observations divided by the total number of observations.

4.4 Results and discussions

4.4.1 Test of nonstationarity for extreme water level time series

As mentioned earlier, test for nonstationarity is performed by checking the water level time series against a linear trend. The significance of this trend is investigated by M-K test and the result is shown in Table 4. 3. It can be seen that the M-K test shows a statistically significant increasing trend ($p << 0.05$) of the water level time series for all selected stations. This also indicates the presence of nonstationarity in water level time series. Hence, the nonstationary condition is used to model extreme water level for all stations in this study.

Table 4. 3: The results of M-K test for all considered stations. The *p*-value < 0.05 indicates a significant trend

Stations	*Tau*	*p*-value
Phuan	0.174	2.22E-16
Nhabe	0.104	0.00049
Bienhoa	0.076	0.00049
TDM	0.381	2.22E-16

4.4.2 Threshold selection and de-clustering

In this study, the value of threshold for each water level station is interpreted through physical approach based on flood alarming rate regulated by Vietnam's government. As such, the water level of 1.6 m may be chosen as the threshold for Bienhoa station. However, due to the topography of Bienhoa station and adjacent areas, the water level of 1.5 m can inundate this area. The number of exceedances corresponding the threshold of 1.5 m is 181 compared with 102 for the threshold of 1.6 m. As reported by Bezak et al. (2014), the threshold value should be high enough, so that the model assumption is not violated, but the truncation level should be as low as possible, so that the highest number of exceedances is selected and more reliable parameter estimates can be made. Hence 1.5 m is selected as the threshold value for Bienhoa station.

Based on typical storm duration on the HCMC's coast, a minimal duration of 3 days is chosen to ensure the independence between two peaks. As such, the extreme water level is automatically separated by 3 days. The threshold values and the number of exceedance after de-clustering are shown in Table 4. 4. The relationships between observed extreme water level after de-clustering and local covariates are represented by scatter plots. The scatter plots for Nhabe station are shown in Figure 4. 5. In Figure 4. 5, the correlation coefficients between extreme water level and covariates are 0.91, 0.50, 0.37, -0.31 and 0.12 for urbanization, sea level, rainfall, outflow from Trian reservoir and outflow from Dautieng reservoir respectively. The scatter plots for Phuan, Bienhoa and TDM are shown in Figure A. 6, A. 7 and A. 8 respectively. It can be seen that the relations between extreme water level and local covariates are statistically significant for most cases.

Table 4. 4: Threshold values for water level gauges

Station	Flood alarming rate (m)	Selected threshold (m)	Number of exceedance (after de-clustering)
Phuan	1.30	1.30	147
Nhabe	1.28	1.28	172
Bienhoa	1.60	1.50	181
TDM	1.10	1.10	150

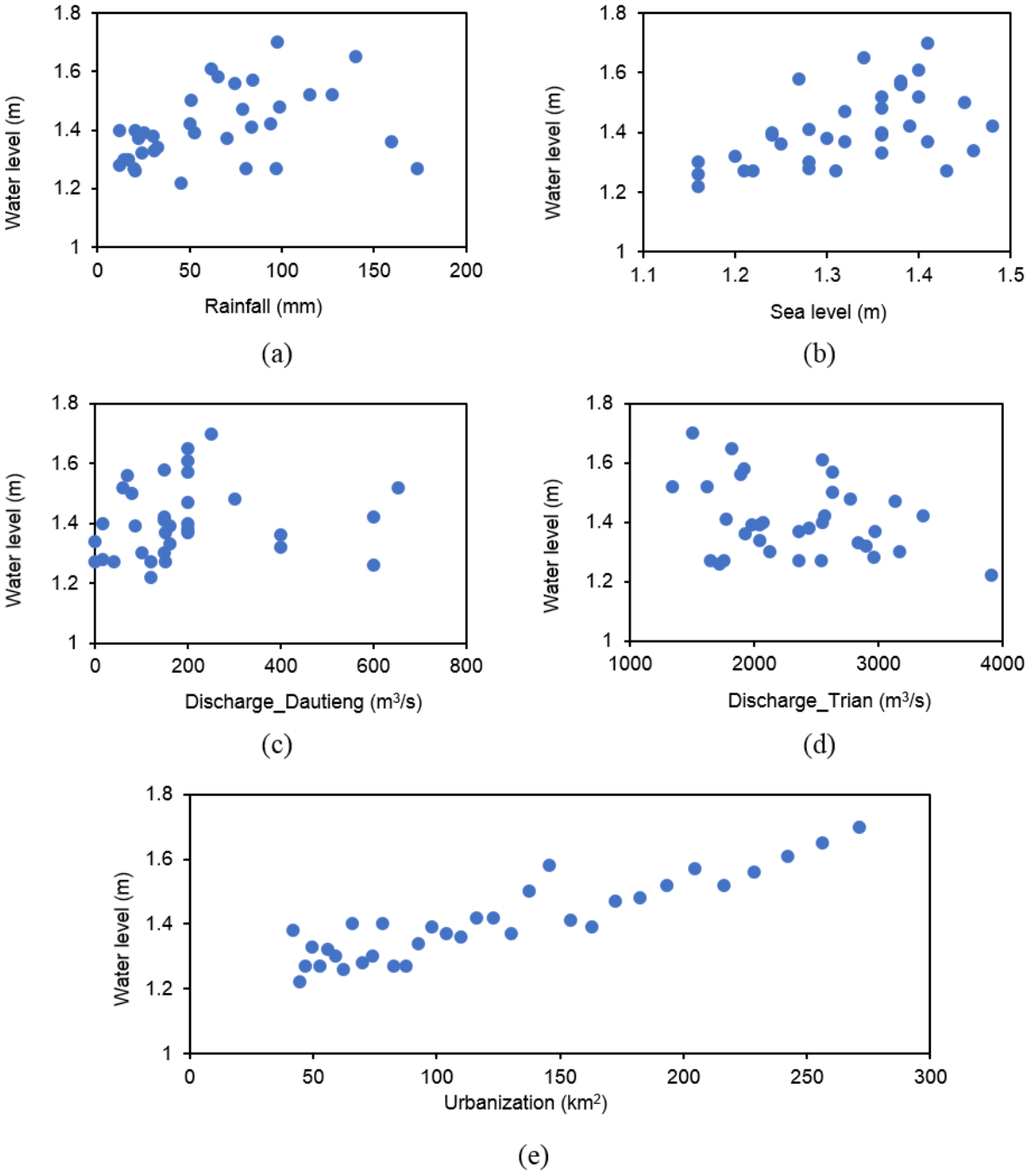


Figure 4. 5: Scatter plots of observed de-clustered extreme water level against the covariate of (a) rainfall, (b) sea level, (c) outflows from Dautieng reservoir, (d) outflows from Trian reservoir and (e) urbanization for Nhabe station

4.4.3 The best models for selected locations

The results for all GPD models fitted to the extreme water level data for the four locations are represented in Table 4. 5. It can be noted that the models with lower AICc should be preferred to models with higher AICc, and the best model is identified as the model which has Δ_i equal to zero. The GPD-26 is found to be the best model for water level extremes for Phuan station. It can be seen that the best model (GPD-26) is based on four covariates such as rainfall, sea level, urbanization and discharge released from Trian reservoir. Moreover, in Phuan station, the stationary model is placed 22nd, and the value of Δ_i between GPD-0 and GPD-26 is 126.1.

For the Nhabe station, the GPD-22 can be considered as the best model for extreme water level. In GPD-22, the linear trend is represented in scale parameter with three covariates of sea level, urbanization and outflow from Trian reservoir. The stationary model is ranked 22 among 32 models.

Regarding Bienhoa station which is closest to Trian reservoir, the model (GPD-14) considering three covariates (i.e. sea level, urbanization and outflow from Trian reservoir) is found to be the best model based on Δ_i value. Including GPD-14, there are seven nonstationary models which are superior to GPD-0. It is also noted that apart from GPD-1, all those superior models are based on outflow released from Trian reservoir.

The GPD-8 is found to be the best model for water level extremes for TDM station. This model included sea level and urbanization as covariates. And the value of Δ_i between GPD-0 and GPD-8 is 95.2. In nine nonstationary models which are superior to GPD-0, sea level and urbanization appear in six models as covariates.

Table 4. 5: The performance of GPD models for extreme water level for all stations

Model	AICc	Δ_i	Model	AICc	Δ_i	Model	AICc	Δ_i
Phuan station								
GPD-26	**-486.5**	**0.0**	GPD-28	-407.8	78.7	GPD-30	-394.9	91.6
GPD-22	-479.8	6.7	GPD-17	-407.0	79.5	GPD-0	-360.4	126.1
GPD-16	-479.5	7.1	GPD-12	-406.8	79.7	GPD-5	-359.7	126.8
GPD-23	-473.2	13.3	GPD-25	-406.7	79.8	GPD-4	-359.7	126.9
GPD-10	-466.1	20.5	GPD-2	-406.2	80.4	GPD-15	-359.1	127.4
GPD-27	-415.1	71.4	GPD-29	-404.7	81.8	GPD-9	-358.1	128.5
GPD-24	-409.2	77.3	GPD-3	-404.3	82.2	GPD-8	-357.7	128.8
GPD-18	-408.7	77.9	GPD-7	-402.3	84.2	GPD-21	-357.0	129.5
GPD-6	-408.4	78.1	GPD-13	-402.3	84.3	GPD-31	-317.8	168.7
GPD-11	-408.0	78.5	GPD-19	-400.3	86.2	GPD-1	-268.9	217.6
GPD-14	-407.9	78.6	GPD-20	-398.1	88.4			
Nhabe station								
GPD-22	**-515.1**	**0.0**	GPD-19	-441.7	73.4	GPD-21	-404.7	110.4
GPD-30	-511.2	3.8	GPD-25	-441.5	73.6	GPD-5	-404.0	111.1
GPD-10	-500.1	14.9	GPD-3	-432.1	82.9	GPD-1	-403.1	111.9
GPD-16	-498.3	16.8	GPD-14	-430.3	84.8	GPD-9	-402.1	113.0
GPD-23	-498.1	17.0	GPD-7	-430.0	85.0	GPD-27	-397.9	117.2
GPD-11	-450.4	64.7	GPD-20	-428.1	86.9	GPD-18	-382.2	132.9
GPD-12	-449.6	65.4	GPD-24	-414.8	100.3	GPD-26	-372.5	142.6
GPD-2	-449.0	66.1	GPD-4	-407.3	107.8	GPD-17	-360.8	154.3
GPD-6	-448.1	67.0	GPD-15	-406.8	108.3	GPD-29	-300.0	215.1
GPD-31	-445.7	69.4	GPD-8	-405.2	109.9	GPD-28	-276.8	238.3
GPD-13	-443.2	71.9	GPD-0	-404.9	110.1			

Model	AICc	Δ_i	Model	AICc	Δ_i	Model	AICc	Δ_i
Bienhoa station								
GPD-14	**-417.6**	**0.0**	GPD-1	-343.6	74.0	GPD-11	-340.6	77.1
GPD-10	-411.6	6.1	GPD-0	-343.4	74.3	GPD-8	-339.8	77.8
GPD-9	-386.0	31.7	GPD-6	-342.4	75.2	GPD-15	-294.2	123.5
GPD-4	-384.9	32.8	GPD-3	-341.9	75.7	GPD-13	-289.4	128.2
GPD-12	-383.8	33.8	GPD-5	-341.8	75.8			
GPD-7	-382.8	34.8	GPD-2	-341.3	76.3			
TDM station								
GPD-8	**-519.1**	**0.0**	GPD-9	-443.3	75.8	GPD-7	-421.3	97.8
GPD-11	-517.6	1.5	GPD-5	-441.4	77.7	GPD-12	-415.5	103.6
GPD-10	-482.5	36.6	GPD-15	-429.7	89.4	GPD-14	-414.1	105.0
GPD-3	-476.0	43.1	GPD-0	-423.9	95.2	GPD-13	-394.0	125.1
GPD-6	-473.9	45.3	GPD-4	-423.5	95.7			
GPD-2	-443.5	75.6	GPD-1	-421.9	97.2			

Note: Models are sorted based on Δ_i values. The bold letters indicate the best GPD models.

To sum up, sea level and urbanization are attributed to be the most significant factors causing nonstationarity in time series since those factors appear in all the best nonstationary models. The outflow from Trian reservoir also has a significant impact on extreme water level in all water level stations located in its downstream. In contrast, the stationary model is not considerable in all cases. The summary of nonstationary analysis is given in Table 4. 6.

Table 4. 6: Summary of nonstationary analysis. The best nonstationary model for each station is shown based on the lowest value of AICc, along with the significant covariate(s)

Station	Best nonstationary model	Significant covariate(s)	Stationary model
Phuan	GPD-26	Rainfall, sea level, urbanization and discharge released from Trian reservoir	24th
Nhabe	GPD-22	Sea level, urbanization and discharge released from Trian reservoir	22nd
Bienhoa	GPD-14	Sea level, urbanization and discharge released from Trian reservoir	8th
TDM	GPD-8	Sea level and urbanization	10th

Maximum-likelihood estimates for scale and shape parameters in the stationary and best nonstationary models are shown in Table 4. 7. This table also demonstrates the likelihood ratio test result (*p*-value). Based on *p*-value, it is clear that the best nonstationary GPD models show superiority over the stationary GPD models for extreme water level of all surveyed locations. Besides, the residual probability and quantile plots for the best nonstationary GPD models for Phuan, Nhabe, Bienhoa and TDM stations are given in Figure 4. 6. It can be seen that the best nonstationary models show a satisfactory fitting the quantiles of extreme water level.

Table 4. 7: Maximum likelihood estimates for parameters of the stationary and best nonstationary models for all stations (SE = standard error of the parameter estimation)

Station	Stationary model		Best nonstationary model							Likelihood ratio test
	α	β	Model	α_0	α_1	α_2	α_3	α_4	β	p-value
	(SE)	(SE)		(SE)	(SE)	(SE)	(SE)	(SE)	(SE)	
Phuan	0.1433	-0.3041	GPD-26	-6.9909	0.0017	3.0355	0.0073	0.0002	-0.8517	3.8E-28
	(0.0153)	(0.0726)		(0.1800)	(2.0E-06)	(0.1609)	(0.0001)	(0.0001)	(0.0811)	
Nhabe	0.1439	-0.2706	GPD-22	-6.1185		2.5076	0.0053	0.0006	-0.7457	4.4E-25
	(0.0140)	(0.0635)		(0.1296)	-	(0.0724)	(0.0001)	(0.0001)	(0.0748)	
Bienhoa	0.1585	-0.1237	GPD-14	-5.2591		0.9482	0.0085	0.0009	-0.3083	2.3E-17
	(0.0154)	(0.0633)		(0.3607)	-	(0.3067)	(2.0E-06)	(2.0E-06)	(0.0427)	
TDM	0.1053	-0.1829	GPD-8	-6.5477		2.1684	0.0095		-0.4940	2.6E-22
	(0.0111)	(0.0685)		(0.1857)	-	(0.1435)	(2.0E-06)	-	(0.0512)	

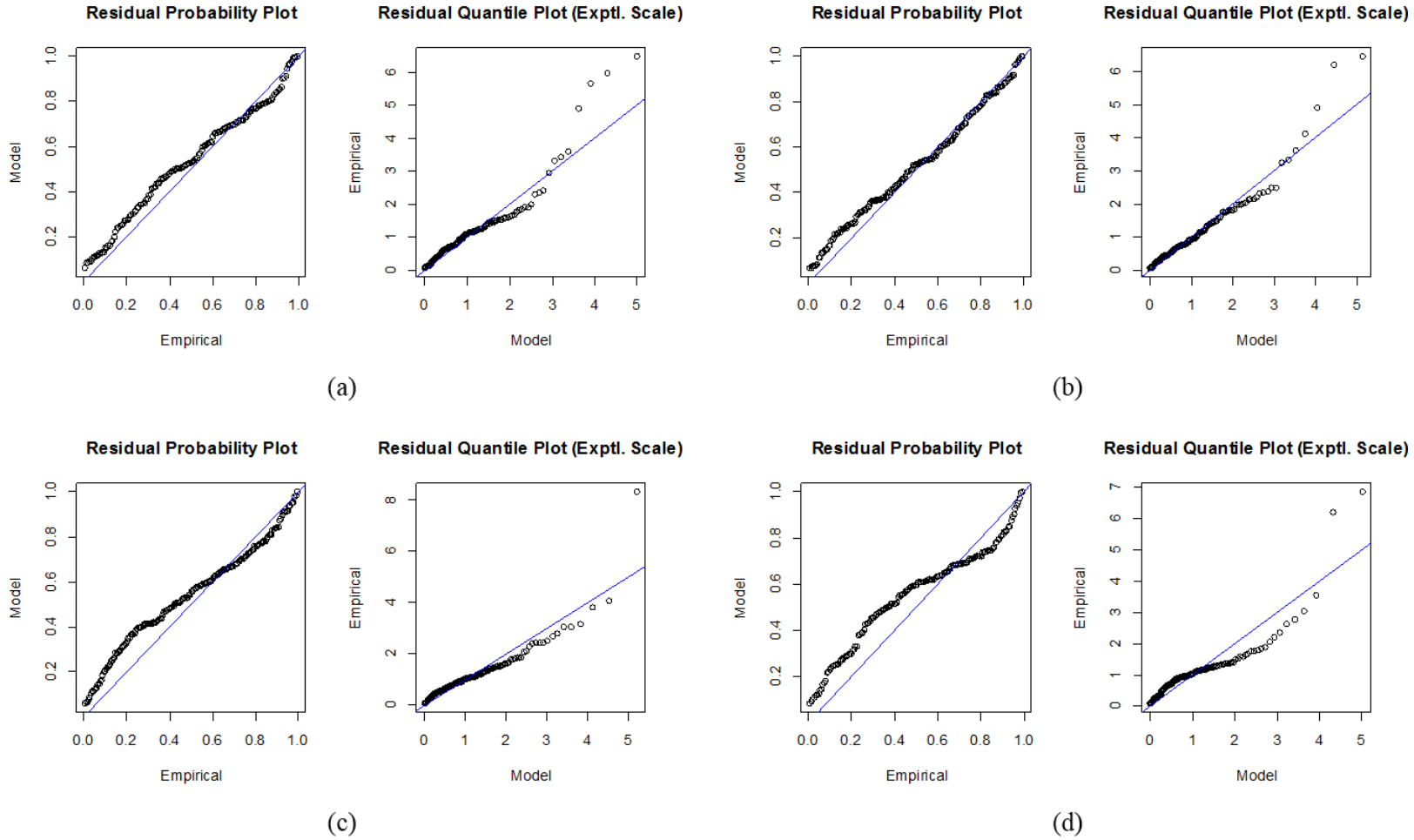


Figure 4. 6: Residual probability and quantile plots for the best nonstationary model for (a) Phuan, (b) Nhabe, (c) Bienhoa and (d) TDM

4.4.4 Extreme water level estimation for different return periods

Figure 4. 7 shows the results from extreme water frequency analysis under stationary and nonstationary conditions for the return period of 100-year. The results for all stations show that the return levels from the best nonstationary model experience significant variability, whereas the stationary model remains unchanged. Taking TDM as an example, a clear increase in extreme water level can be seen over the period of 1981-2014 based on the best nonstationary model (GPD-8) with sea level and urbanization as covariates. Meanwhile the stationary model shows a fixed return level for the same period (Figure 4. 7d). In addition, it can be seen that the stationary model underestimates extreme water level compared to the nonstationary model in recent years (Figure 4. 7d), which could be the result of strong urbanization and rising sea level for this specific example. In a similar way for remaining stations, the 100-year extreme water level is underestimated under stationary condition in recent years, which is shown more apparent for Nhabe and Bienhoa stations (Figure 4. 7b and c).

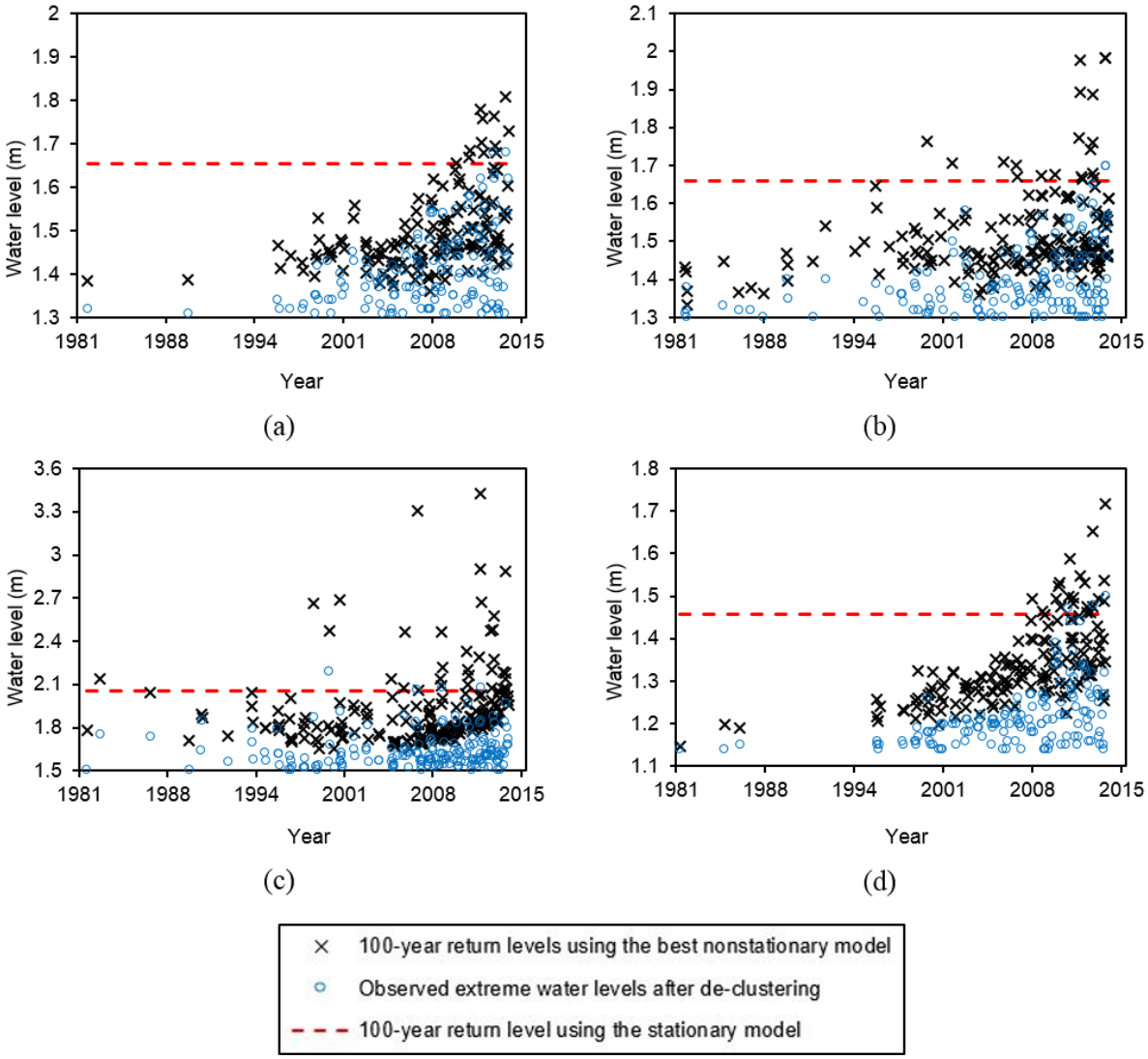


Figure 4. 7: The time series of observed extreme water level after de-clustering (‘o’ markers) in the period 1981-2014, along with 100-year return levels obtained from the stationary model (dashed line) and the best nonstationary model (‘+’ markers) for (a) Phuan, (b) Nhabe, (c) Bienhoa and (d) TDM

In addition, a comparison of the return levels based on the best nonstationary model and common stationary model is shown in Figure 4. 8. It can be seen that the values of the return levels in the best nonstationary models (red curves) shown for all stations have a significant difference compared with stationary models (blue curves). In other word, the extreme water level derived from the stationary models are underestimating for all stations and for all return periods.

In more detail, taking Phuan station as an example, the extreme water level for 1-year period under nonstationary condition (1.69 m) is corresponding to a 100-year return period under stationary condition. In case of Nhabe station, the estimated extreme water level of 100-year return period under nonstationary and stationary conditions are 2.31 m and 1.71 m respectively. With sea level and urbanization as covariates, the estimated extreme water level corresponding to the 100-year return period is 1.9 m for TDM station. While the extreme water level in Bienhoa station is estimated under nonstationary condition to reach a value of 3.23 m (T = 100). To sum up, the difference in extreme water level estimates between stationary and nonstationary models reaches 47% for Bienhoa station, while those values for the remaining stations of Nhabe, Phuan and TDM are 35%, 31% and 25% respectively. The result also emphasizes that choosing the design flood for hydraulic structures using stationary or nonstationary model should be considered thoroughly.

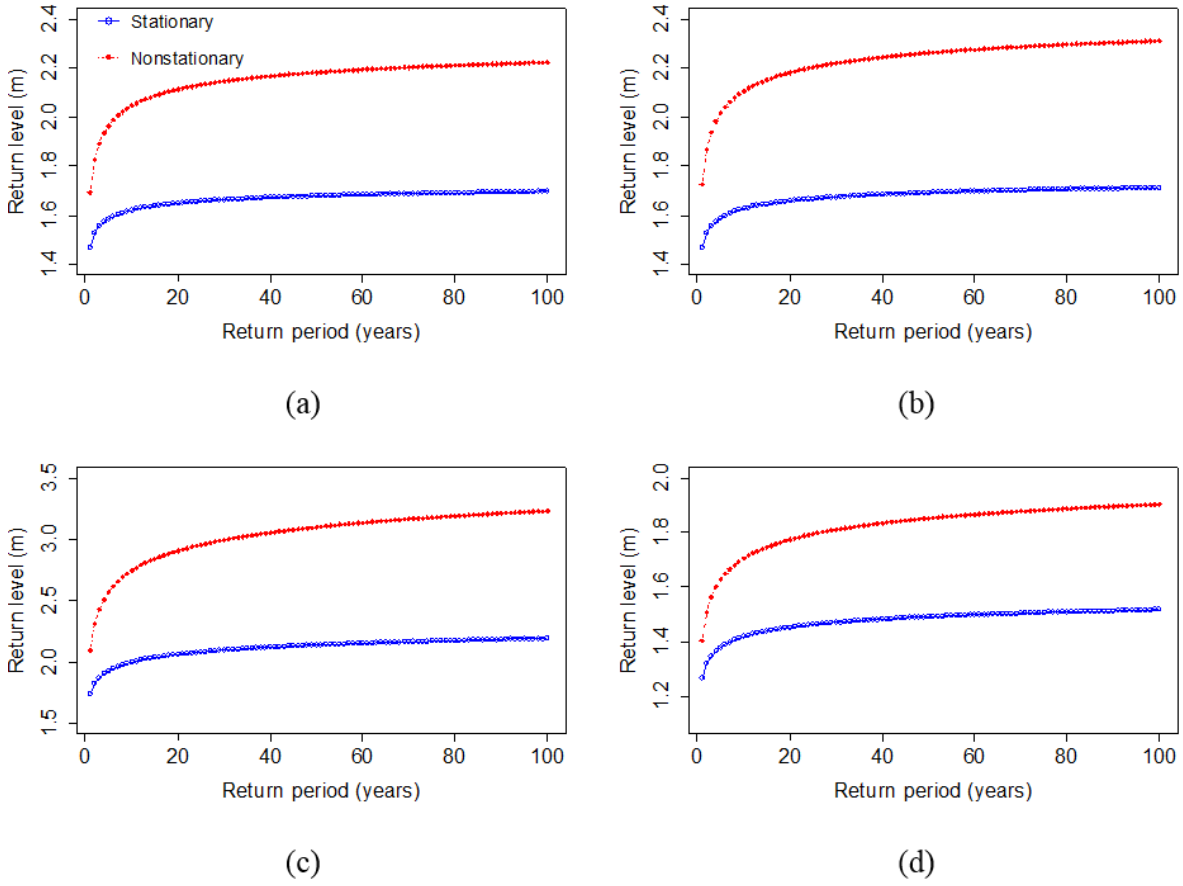


Figure 4. 8: Comparison of return levels between the stationary model and the best nonstationary model for all stations at extreme water level: (a) Phuan; (b) Nhabe; (c) Bienhoa; (d) TDM

4.5 Summary and conclusions

In this study, the modelling of nonstationary extreme water level in HCMC is carried out considering the influences of local physical processes, namely rainfall, sea level, urbanization and outflows from upstream reservoirs due to their relationship to water level. In particular, after detecting the trend in time series by using M-K test, ninety-two nonstationary GPD models (for four stations) are developed based on local physical covariates and their combinations. Besides, the stationary models are also constructed for comparison purpose of showing the dominance of nonstationarity in the extreme water level. The best nonstationary model for each station is found based on AICc value and likelihood ratio test.

The findings from the present study exhibit the significant impacts of local variables on the extreme water level in HCMC. Two factors, sea level and urbanization, are found to be associated with the nonstationarity in extreme water level since those covariates appear in all the best nonstationary models for all considered locations. The impact of outflow from Trian reservoir on extreme water level is more significant than outflow from Dautieng reservoir. In contrast, the stationary model is not superior in all selected locations.

The results also indicate that there are significant differences in the estimated extreme water level between stationary and nonstationary cases. In particular, the extreme water level values derived from the stationary models are underestimated for all stations relative to the best nonstationary models. For instance, the extreme water level for 1-year period under nonstationary condition is corresponding to a 100-year return period under stationary condition. This implies that the existing structures based on stationary assumption of design flood could be likely unsuitable for current level of protection.

Since flooding continues to be considered as a major threat to HCMC and the extreme water level is proved to have nonstationary feature in time series, it poses a huge challenge in term of prevention and mitigation infrastructure design. Our proposed method provides more possible choice for decision makers in selecting appropriate design flood for HCMC as well as for other areas belonging to the Saigon-Dongnai River basin. Although, this study successfully introduces nonstationarity in extreme water level by using local physical processes, it contains some limitations. For instance, the selection of the thresholds is objective, and only GPD is introduced to model the extreme water level. Therefore, it is suggested that more appropriate distributions can be considered for further studies.

Chapter 5

High-resolution flood hazard mapping based on nonstationary frequency analysis

5.1 Introduction

Floods may be considered as among the most devastating natural disasters, impacting millions of people every year across the world (Jongman et al., 2012, Hallegatte et al., 2013, Lasage et al., 2014, Karamouz et al., 2017). In the last few decades, the effects of climatic change and sea level rise have been creating additional pressure which could increase flood vulnerability by effecting the magnitude and frequency of floods (Bates et al., 2005, Nicholls and Cazenave, 2010, Purvis et al., 2008, Karamouz et al., 2017). In terms of reducing damages and losses, flood hazard mapping has become a priority, since the information significantly contributes to flood warning systems, as well as flood risk management schemes. Traditionally, flood control structures have been built based on the assumption of data stationarity for flood frequency analysis (Mudersbach and Jensen, 2010, Salas and

Obeysekera, 2013, Katz, 2013, Šraj et al., 2016, Yilmaz et al., 2016). However, flood series, as recently suggested by many

researchers, have a nonstationary nature due to climate variability (Milly et al., 2008, Gilroy and McCuen, 2012, Ishak et al., 2013). Thus stationary conditions may no longer be appropriate, and the concept of nonstationarity has been improved and used more frequently in analyzing flood events in lowland areas (Binh et al., 2018, López and Francés, 2013, Li et al., 2015, Prosdocimi et al., 2015, Šraj et al., 2016). As such, flood frequency analysis can be performed by taking nonstationarity into account.

Unlike the stationary approach, nonstationary flood frequency analysis, the parameters of the chosen distribution functions are commonly expressed as a function of covariates. Previous studies have commonly used time as a covariate (Šraj et al., 2016, Salas and Obeysekera, 2013). Some other studies have shown that the parameters could vary with several climatological variables, such as ENSO, NAO, AO, PDO, North Pacific Oscillation (NPO), and human-induced environmental factors (López and Francés, 2013, Li et al., 2015, Zhang et al., 2015, Machado et al., 2015, Gilroy and McCuen, 2012). Overall, it can be suggested that the covariates selected for nonstationary modelling should have strong physical associations with the process of flood events and should be able to provide reliable future predictions (Agilan and Umamahesh, 2016b, Yan et al., 2017, Nash and Sutcliffe, 1970).

As mentioned in two previous chapters, HCMC city has been facing climate problems from last few decades, such as increases in frequency and magnitude of extreme rainfall events (ADB, 2010), which have resulted in increasing severe floods and inundation. In addition, sea level rise is likely to have an important influence on the inland reach of tidal flooding, which is expected to be more severe in HCMC in the future (ADB, 2010, World Bank, 2010). However, in most of the studies on flood forecasting in HCMC, rainfall and sea level rise have not been analyzed and assessed thoroughly before being entered into flood simulation models as initial inputs. In other words, they did not use a nonstationary approach to modelling rainfall and sea level that significantly impact on flooding in HCMC. Therefore, without proper assessment of the causes of floods, it is difficult to investigate the best information on flood hazards, which becomes a problematic challenge for local governments as the flood risks in HCMC are increasing continuously.

To date, various models have been developed for providing flood information. Among these, the 1D hydrodynamic model is the most widespread approach, due to its numerical

stability and computational efficiency (Papaioannou et al., 2016, Moore, 2011). However, it may not come up with an accurate result for a complex topography and depends largely on the correct placement of cross-sections (Moore, 2011). In contrast, the 2D hydrodynamic model can accurately model complex topography, geomorphological and sedimentological processes, and has become a standard in flood prediction (Shen et al., 2015, Timbadiya et al., 2014a). Nevertheless, the 2D model is not computationally efficient and may not be suitable for a large area in case of urgent need (Timbadiya et al., 2014a). To combine the advantages of the 1D and 2D hydrodynamic models, an alternative approach has been developed by coupling these models. The coupled 1D-2D hydrodynamic models with their advantages have been widely applied in flood inundation mapping or flood risk estimation (Leandro et al., 2009, Yin et al., 2013, Timbadiya et al., 2014a, Papaioannou et al., 2016).

Furthermore, the flow structure is quite complex in populated areas; therefore, the flow simulation results greatly depend on accurate and high-resolution topographical data (Tsubaki and Fujita, 2010). Among new techniques developed in recent years, the LiDAR technique could improve the accuracy of the DEM for use as model input. The application of high-resolution LiDAR-derived DEM data in flood and inundation simulation can be found in several papers (Moore, 2011, Sampson et al., 2012, Shen et al., 2015, Papaioannou et al., 2016).

The main objective of this study is to address the following issues: (i) modelling the extreme value frequency analysis (i.e. rainfall, sea level and discharge) under nonstationary conditions by considering global and local processes and their possible combinations as covariates; (ii) developing an appropriate flood simulation model based on a coupled 1D-2D hydrodynamic model with high-resolution topography data; (iii) establishing flood hazard maps for different scenarios in a selected area within HCMC, and classifying them based on a combination of flood depth and flood velocity; and (iv) comparison of floodplain extent between the stationary and the nonstationary cases.

5.2 Study area and datasets

As one of the largest river basins in the South of Vietnam, the Saigon-Dongnai River basin provides important sources of water for people in the catchment areas in general and HCMC in particular. The total catchment covers an area of 48,471 km^2, with a mean water

discharge of approx. 47.065×10^9 m^3/year (Merz et al., 2011). The annual average rainfall is 2,000 mm, and the rainy season (April–November) receives 85% of the total annual rainfall. The Saigon-Dongnai River system contains five main rivers: the Dongnai, Saigon, Be, Vamcodong and Vamcotay. The river system drains HCMC before emptying into the South China Sea. The hydrological regime of the Saigon-Dongnai River basin is influenced by a semi-diurnal tide, rainfall and outflows from upstream reservoirs.

5.2.1 Rainfall, Sea level and Discharge

To estimate the values of extreme rainfall, as well as sea level for different flood scenarios, rainfall and sea level data for the period 1980–2014 were collected from 22 rainfall stations which cover the entire catchment area, and one sea level station. In particular, the daily rainfall data was recorded at eight stations located within HCMC and 14 stations outside HCMC (belonging to Binh Duong, Dong Nai, Ba Ria Vung Tau, Long An and Tay Ninh provinces). The hourly observed sea level data was recorded at Vungtau station. These observed data were provided by the NHMS of Vietnam. The estimates of extreme rainfall and sea level for different return periods were used, respectively, as input data for the hydrological model and as downstream boundary conditions for the hydrodynamic model. The locations of the selected gauging stations are shown in Figure 5. 1.

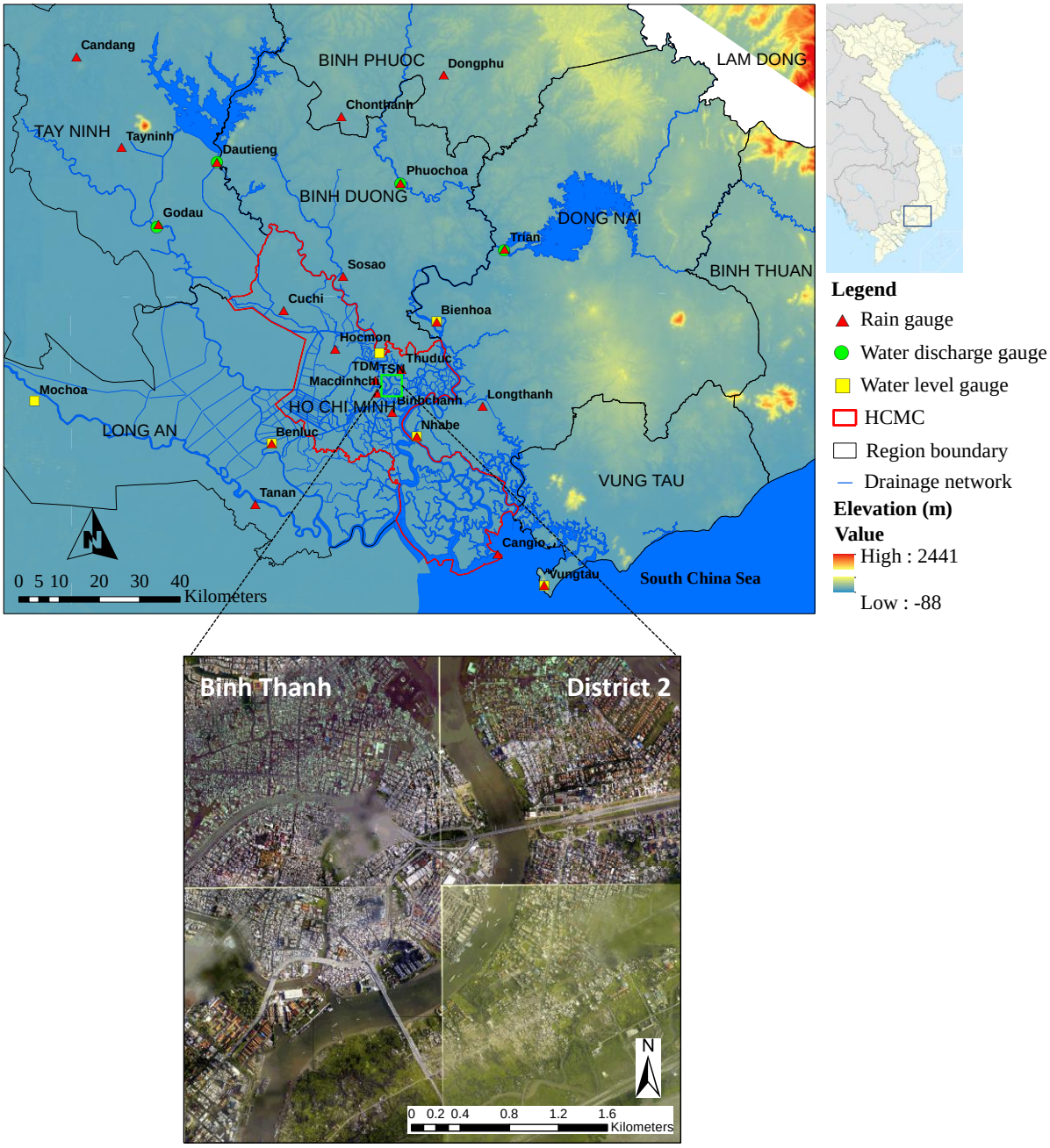


Figure 5. 1: Location of the gauging stations in the Saigon-Dongnai River basin and the wider study area

In addition, water discharge data from upstream areas was also collected and used in the frequency analysis. The daily water discharge data from Go Dau gauging station and three upstream reservoirs, i.e. Trian, Dautieng, Phuochoa, were provided by Trian Hydropower Joint Stock Company and Dautieng-Phuochoa Limited Company for the period 01-01-1980 to 31-12-2014. For the Mochoa station, discharge data is not available; therefore, hourly water-

level data provided by NHMS were used in this study. These data (i.e. water discharge and water level) were used for statistical analysis and to estimate return levels before entering into the hydrodynamic model as upstream boundary conditions. Figure 5. 2 shows the annual maximum daily discharge, annual maximum hourly water level and annual maximum daily rainfall at selected gauges in the Saigon-Dongnai River basin.

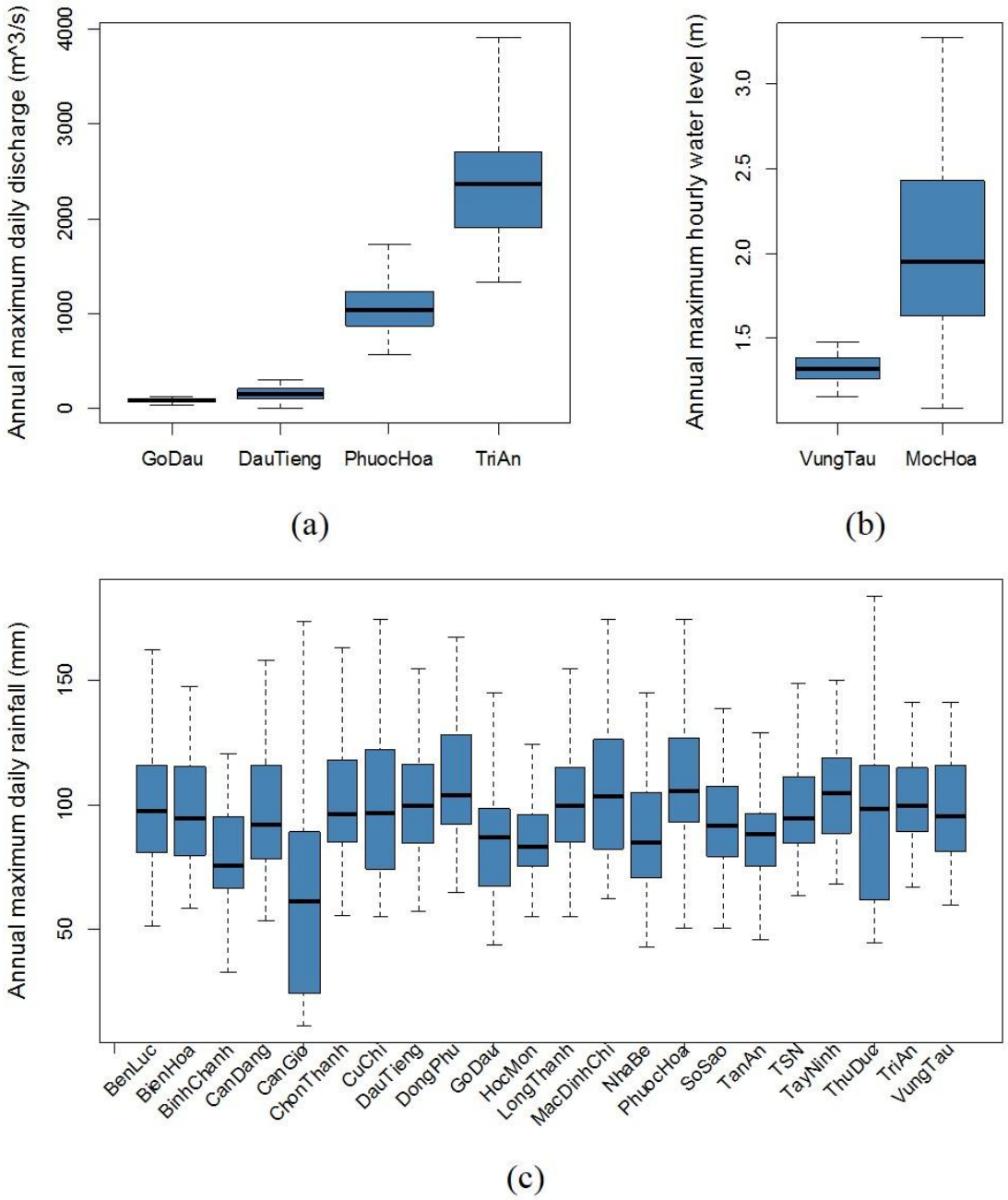


Figure 5. 2: Boxplot of (a) annual maximum daily discharge, (b) annual maximum hourly water level and (c) annual maximum daily rainfall at selected gauges

5.2.2 Data for covariates

In this study, five physical covariates, namely ENSO, PDO, local mean temperature, global warming and global mean sea level rise, were considered, to show the impacts of global and local processes on the nonstationarity of extreme events (i.e. rainfall and sea level).

The ENSO cycle has had a significant impact at local and regional scales through teleconnections influencing the coupled ocean-atmosphere and land systems (Wang et al., 2006). The influences of this pattern on extreme rainfall and flooding have also been indicated in many parts of the world (Gobin et al., 2016, Mondal and Mujumdar, 2015, Villafuerte et al., 2015, Agilan and Umamahesh, 2016b, Li et al., 2015, Ishak et al., 2013). In Vietnam, ENSO has proved to play an important role in climate and to contribute to the inter-annual variation in rainfall in many regions (Yen et al., 2011, Nguyen et al., 2014, Gobin et al., 2016). In this study, monthly sea-surface temperature (SST) anomaly series over the Niño 3.4 region were used as ENSO indicator; this was obtained from the US National Oceanic and Atmospheric Administration (NOAA) Earth System Research Laboratory (ESRL) (available at https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Nino34/).

The PDO, a pattern of Pacific climate variability, has been shown to have significant climatic and environmental impacts across the Pacific Basin (Deng et al., 2013). Together with ENSO, PDO has been investigated by many authors and found to have an influence on the East Asian monsoon, as well as seasonal rainfall patterns (Chan and Zhou, 2005, Chen et al., 2013, Wu, 2013). The PDO index was also extracted from NOAA ESRL (available at https://www.esrl.noaa.gov/psd/data/climateindices/list/), and is used in the nonstationary frequency analysis as a covariate.

Increases in extreme rainfall have been documented in many regions across the world (IPCC, 2012) and human-influenced global warming may be partly responsible (Min et al., 2011). Kunkel et al. (2013) indicated that rising temperatures and subsequent rises in atmospheric moisture content might increase the probable maximum precipitation values. Nevertheless, the physical mechanisms linking local temperatures with rainfall may not be the same as those linking global warming to extreme rainfall changes (Trenberth, 2011, Mondal and Mujumdar, 2015). Therefore, global temperature and local temperatures are chosen as separate covariates for analyzing the extreme rainfall characteristic in this study.

The HadCRUT4 annually observed global average surface air temperature anomaly series (available at http://www.metoffice.gov.uk/hadobs/hadcrut4/) with respect to the 1961–1990 mean was used as an indicator of global warming. The yearly mean temperature data for the period 1980–2014 was provided by the Southern Institute for Water Resources Planning. This data was recorded at four stations located in HCMC: TSN, Bienhoa, Dongphu, Vungtau, and three in adjacent regions (Figure 5. 1). The yearly mean temperature anomaly series based on the 1980–2014 mean was calculated and considered as a covariate.

In recent decades, extreme sea level has been found to have increased in various regions worldwide (Woodworth et al., 2011). Many studies have reported that long-term changes in extreme sea level are generally associated with corresponding increases in mean sea level (Lowe et al., 2010, Letetrel et al., 2010, Weisse et al., 2014, Feng and Tsimplis, 2014). As a coastal city, HCMC is expected to be severely influenced by sea level rise. Hence, using global mean sea level as a covariate in extreme sea level statistical analysis is reasonable. The global mean sea level (GMSL) data used in this study is the up-to-date version of reconstructed GMSL from Church and White (2011) for the period 1980–2014. This data is available at http://www.cmar.csiro.au/sealevel/sl_data_cmar.html.

5.2.3 Soil, land use and DEM

LULC maps, soil maps and 1-m resolution LiDar data were used as input in the hydrological and hydrodynamic models. These data were provided by Ho Chi Minh City Department of Science and Technology and are not publicly available. The main types of soils in the Saigon-Dongnai River basin are alluvial, basalt, grey, black and soft soil. The main types of LULC in the study area are built-up areas, vegetation, bare soil and wetland. These data were used to estimate the parameters in the hydrological and hydrodynamic models.

5.3 Methodology

A flowchart of the proposed methodology for developing flood hazard maps is shown in Figure 5. 3. The flowchart consists of four main sections. Apart from data preparation mentioned in Section 5. 2, the three remaining sections are described as follows:

(i) Annual maximum daily rainfall and hourly sea level time series were used for frequency analysis. The GEV distribution was chosen with the assumption of stationarity and nonstationarity in the time series. The magnitude of extreme rainfall from the best GEV model was used as input data for the hydrological model and, similarly, extreme sea level was used as downstream boundary condition for the hydrodynamic model. Meanwhile, the values of upstream water level and discharge from the stationary GEV model were used as the upstream boundary conditions in the hydrodynamic model

(ii) Estimation of runoff from sub-basins is based on the lumped conceptual rainfall-runoff model. The result from the rainfall–runoff model was used in 1D flow simulation

(iii) Development of the 1D flow model, the 2D flow model and coupling of the 1D-2D models for the river system and identification of the spatial variation of flood hazards corresponding to three flood scenarios derived from a combination of extreme rainfall, sea level and discharge for different return periods (i.e. 25-, 50- and 100-year).

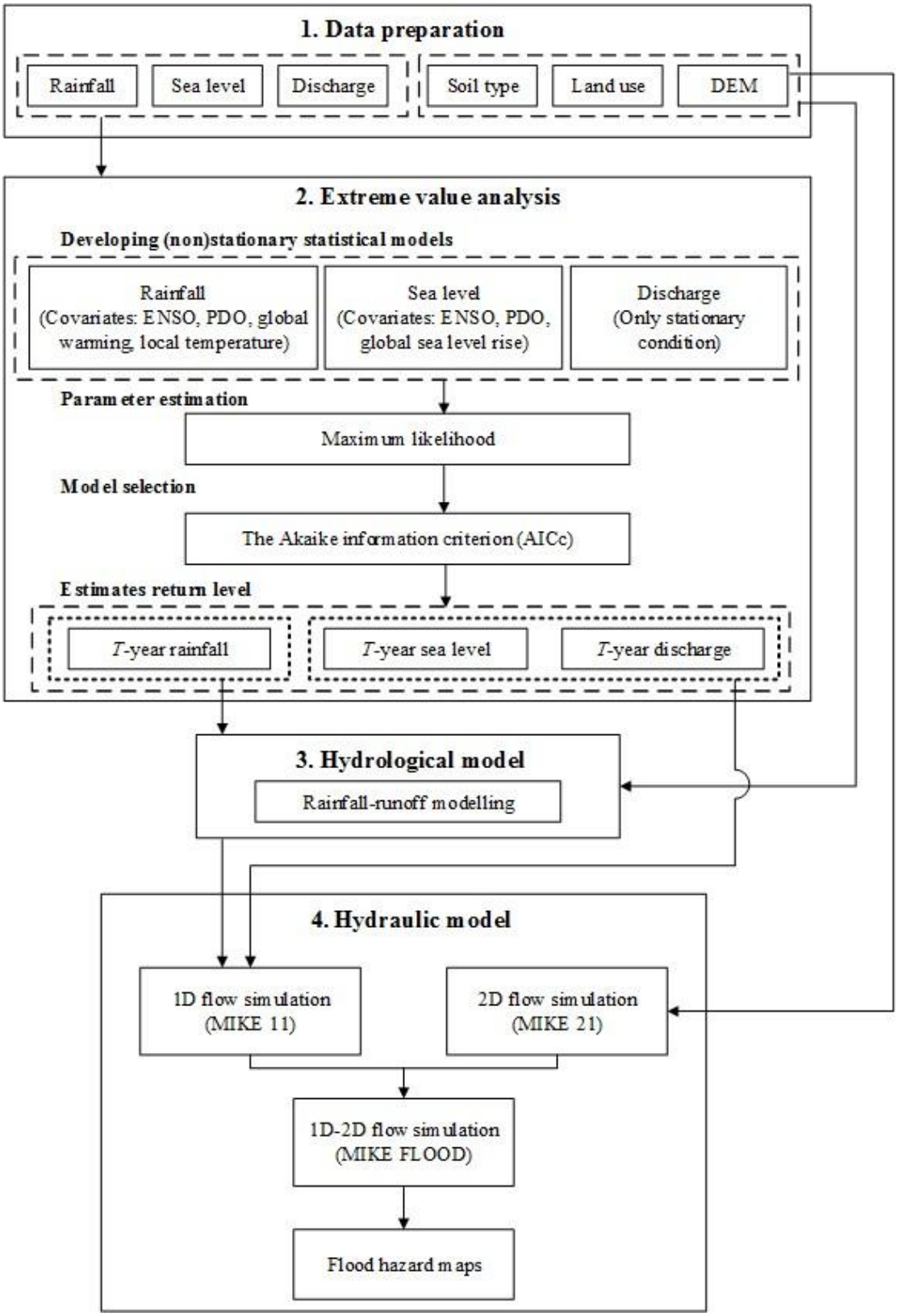


Figure 5. 3: Proposed methodology flowchart for developing flood hazard maps

5.3.1 The GEV model

In this study, the GEV distribution is used for the frequency analysis of extreme events. In the nonstationary setting, the parameters are expressed as a function of covariates in the general form:

Case 1: For extreme rainfall analysis

$$\mu(t) = \mu_0 + \mu_1 C$$
$$\sigma(t) = \exp(\sigma_0 + \sigma_1 C) \quad (5.1)$$

Case 2: For extreme sea level analysis

$$\mu(t) = \mu_0 + \mu_1 C$$
$$\sigma(t) = \exp(\sigma_0 + \sigma_1 C) \quad (5.2)$$

Case 3: For extreme discharge and water-level analysis

$$\mu(t) = \mu_0$$
$$\sigma(t) = \sigma_0 \quad (5.3)$$

where C represents any physical covariate, i.e. the ENSO cycle (E), PDO cycle (P), mean temperature anomaly (GT), local temperature (LT) or global sea level rise (GS). In the stationary model (GEV-0), the values of C equal zero. The exponential in Eqs. (5.1) and (5.2) is taken to ensure positive values of the scale parameter. Based on four covariates and their combinations, 25 nonstationary models were constructed for each rain gauge in the extreme rainfall statistical analysis (Table A. 1). For extreme sea level analysis 13 nonstationary models were constructed based on three covariates and their combinations (for only Vungtau station) (details provided in Table A. 2). Based on these models, individual covariates or combinations that had significant impacts on the extreme rainfall and extreme sea level in the study area were derived. In the case of extreme discharge and extreme water level from upstream, only the stationary condition was used for statistical analysis in this study.

The distribution parameters are estimated by maximum likelihood method. The AICc is used to identify the best fitting distribution for rainfall and sea level. Once the best model for extreme rainfall (or water level or discharge) is determined, the T-year return level z_T

corresponding to the *T*-year return period can be obtained. Unlike in the stationary model, the location and scale parameters of the nonstationary model vary over time. Here, a low-risk approach (more conservative) suggested by Cheng et al. (2014) is used, by taking the 95th percentiles of $\mu(t)$ and $\sigma(t)$ in historical observations to calculate return level in this study, as follows:

$$\hat{\mu}_{95} = Q_{95}(\hat{\mu}_{t1}, \hat{\mu}_{t2}, \ldots, \hat{\mu}_{tn}) \tag{5.4}$$

$$\hat{\sigma}_{95} = Q_{95}(\sigma_{t1}, \sigma_{t2}, \ldots, \sigma_{tn}) \tag{5.5}$$

Estimation of the *T*-year return level can be given by (Coles, 2001) as Eq. (5.6):

$$z_T = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}\left[1 - \left\{-\log\left(1 - \frac{1}{T}\right)\right\}^{-\hat{\xi}}\right] & \text{for } \hat{\xi} \neq 0 \\ \hat{\mu} - \hat{\sigma}\log\left\{-\log\left(1 - \frac{1}{T}\right)\right\} & \text{for } \hat{\xi} = 0 \end{cases} \tag{5.6}$$

By substituting the values of estimated parameters into Eq. (5.6), the estimates of the return levels can be obtained.

5.3.2 Hydrological model

The lumped conceptual rainfall-runoff model is developed by conceptualizing the catchment as a number of interconnected storages, with a set of mathematical equations used to describe the process of the water flow lumped over them (Chiew, 2010). Due to its simplicity, the lumped conceptual rainfall-runoff model has been widely used in previous studies to mimic hydrological processes in catchments (Madsen, 2000, Brirhet and Benaabidate, 2016, Anh et al., 2008). In this study, to estimate the generated runoff in the whole Saigon-Dongnai River basin, it was divided into 216 sub-basins based on land topography. With 22 rain gauges scattered throughout the sub-basins, the Thiessen polygons method was used as an interpolation method to calculate the average depth of rainfall on the area of these sub-basins. The storm runoff was estimated by the Soil Conservation Service (SCS) method (Hjelmfelt Jr, 1991) which is available in the MIKE 11 Unit Hydrograph Model (UHM) developed by the Danish Hydraulic Institute (DHI, 2003). The MIKE 11 UHM is a lumped conceptual rainfall-runoff model that simulates the runoff from a single rainstorm by using the unit hydrograph technique.

The SCS loss method is used to estimated the losses of model. Analysis of storm event rainfall and runoff indicates that there is a threhold which must be exceeded before runoff occurs. It is expresses by Eq. (5.7).

$$\frac{F}{S} = \frac{Q}{P-I} \tag{5.7}$$

where F is actual retention after runoff begins (mm), S is watershed storage (mm), Q is actual direct runoff (mm), P is total rainfall (mm) and I is initial abstraction (mm).

The SCS dimensionless unit hydrograph is used to transform the excess precipitation into a flow hydrograph at the outlet of each basin. The lag time (T_{lag}) is the main parameter for this method. Lag time is the time difference between the centroid of rainfall excess and the centroid of the Direct Runoff Hydrograph (DRH). Lag time can be estimated from the watershed characteristics using Curve Number (CN) by the standard SCS formula and it is given by Eq. (5. 8).

$$T_{lag} = \left(\frac{(L\times 3.28\times 10^3)^{0.8}\times(1000\times CN-9)^{0.7}}{1900\times Y^{0.5}}\right) \tag{5.8}$$

where, T_{lag} is the catchment lag time in hours, L is the hydraulic length measured along the main river in km and Y is the average catchment slope in percent.

The model runs on 24-h rainfall records and potential evaporation. For the flood scenarios, once the extreme rainfall is calculated through the best (non)stationary model for each station, this value is assumed to be uniformly distributed and used as input data for the rainfall-runoff model. The output from the rainfall-runoff model can be used as lateral inflow for the 1D hydrodynamic model.

5.3.3 Hydrodynamic model

1D-2D models are dynamiclly linked in a package called MIKE-FLOOD. The model was developed by the Danish Hydraulic Institute (DHI, 2007). The 1D-2D coupled technique is suitable becauce the study area is reprensented by complex floodplains and river channel. MIKE 11 and MIKE 21 are coupled through of links. There are several type of link can be used in numerous situations. A lateral link is one of the most link which widely used in flood

modelling. Schematisation of lateral links is illustarted in Figure 5.4. It allows a string of MIKE 21 elements to be laterally linked to a given reach in MIKE 11, either a section of a branch or entire branch. Flow through the lateral link is calculated using a structure equation. This type of link is particularly useful for simulating from river channel onto a floodplain (DHI, 2007) . Therefore, in this study, the coupled 1D-2D model MIKE FLOOD was used to simulate the flood inundation for HCMC in the Saigon-Dongnai River basin. To set up the 1D hydrodynamic model which represents the entire river system of the Saigon-Dongnai River basin, the input data included cross-sections, the Manning's *n* roughness coefficients and boundary conditions. There are six boundaries. The hourly sea level data (Vungtau station) was used as a downstream boundary condition, while daily discharge time series from Trian, Dautieng, and Phuochoa reservoirs, daily discharge data from Godau streamflow gauge located in the Vamcotay River, and hourly water level data from Moc Hoa water level gauge located in the Vamcodong River were used as upstream boundary conditions. Besides, five main rivers, i.e. Dongnai, Saigon, Be, Vamcotay and Vamcodong, together with 251 small streams were established, and the details of 300 observed cross-sections were used in this study.

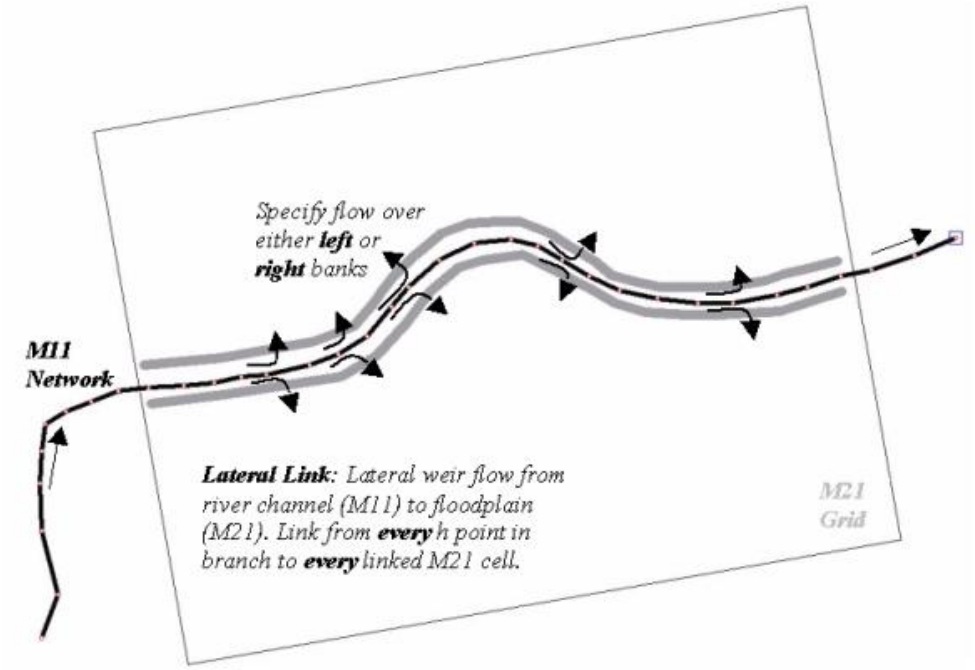


Figure 5.4: Sketch repesenting the application of lateral links; Source: (DHI, 2007, FLOOD, 2011)

The 1D model (MIKE 11) was calibrated for the year 2012 and validated for the year 2011. In the calibration procedure, the Manning's roughness coefficient values were adjusted manually to produce the smallest deviations between the observed and modelled values. The initial Manning's *n* values for channels were chosen from the study of Razmi et al. (2017). Ritter and Muñoz-Carpena (2013) suggested that, in the model performance assessment, one should include at least one absolute value error indicator, one dimensionless index and a graphical technique, which provide a visual comparison between observed data and model calculated values.

In this study, the following statistical criteria were used to assess model performance: the coefficient of determination (R^2), the ratio of root mean square error to standard deviation observations (RSR), the Nash-Sutcliffe efficiency (NSE; Nash and Sutcliffe 1970), and a graphical representation of the relationship between observations and model estimates. The R^2 describes the proportion of the variance in measured data explained by the model, where R^2 = 1 is considered as the perfect match, and values greater than 0.5 are acceptable (Moriasi et al., 2007, Jeong et al., 2010). The RSR incorporates the benefits of error index statistics and includes a normalization factor so that the resulting statistic can apply to various constituents (Moriasi et al., 2007). Values of RSR range from the optimal value of 0 to a large positive value. The NSE, ranging between $-\infty$ and 1.0, is commonly used to access the predictive power of the model (Nguyen et al., 2016). Normally, NSE of 0.65 is considered good for daily results. However, the criteria may be lower for sub-daily outputs and higher for monthly and annual outputs since performance improves as the time interval increases (Jeong et al., 2010). The RSR and NSE are calculated as follows:

$$RSR = \frac{\sqrt{\sum_{i=1}^{N}\left(X_i^{obs} - X_i^{sim}\right)^2}}{\sqrt{\sum_{i=1}^{N}\left(X_i^{obs} - X^{mean}\right)^2}} \tag{5.9}$$

$$NSE = 1 - \frac{\sum_{i=1}^{N}\left(X_i^{obs} - X_i^{sim}\right)^2}{\sum_{i=1}^{N}\left(X_i^{obs} - X^{mean}\right)^2} \tag{5.10}$$

where X_i^{obs} and X_i^{sim} are the *i*th observed and simulated data, X^{mean} is the mean of observed data, and *N* is the total number of observations.

The developed 2D model comprises a flexible mesh with 686,173 elements and 343,498 nodes, which covers study area (16 km^2) mentioned in Section 5.3.5. The Manning's *n* values for floodplain areas are referred to (Timbadiya et al., 2014a). In particular, the Manning's *n* values for the residential areas, agriculture areas and water bodies are 0.2, 0.07 and 0.03, respectively. The surface elevations for the study area were derived from 1-m resolution LiDAR data. Due to the lack of detailed historical data of inundation events, calibration and validation for the 2D model was not carried out.

5.3.4 Flood scenario simulations

Three sets of model simulations were carried out for different flood scenarios derived from a combination of extreme rainfall, sea level and discharge. For the upstream boundary conditions in each scenario, the estimated daily discharge from the frequency analysis (Trian, Dautieng, Phuochoa and Godau) was assumed to be at a constant value. Similarly, the only upstream boundary condition in the Vamcodong River was assigned as estimated water level (Mochoa), which is considered as constant. For the downstream boundary, hourly sea level time series from 24-12-1999 are used, when the highest water level station was recorded at Vungtau. The observed hourly sea level was then amplified until the tidal peak matched the value of extreme sea level gained from the best GEV model. The hourly sea level time series after amplifying were used as downstream boundary conditions for the hydrodynamic model. It is to be noted that the impacts of waves were not considered.

5.3.5 Case study

Located in the downstream of the Saigon-Dongnai River systems, the topographic and geographic conditions of HCMC make it extremely sensitive to various flood sources. HCMC is subject to both regular and extreme flooding. In 2050, 61% of urban land use and 67% of industrial land use are expected to be flooded in an extreme event if the proposed flood control measures are not implemented (World Bank, 2010). However, generating detailed flood hazard maps based on high-resolution topographical data for the whole HCMC, which covers a very large area (approx. 2,095 km^2), requires high computational resources. Our study is limited to the area located close to the Saigon River branch, a distributary of the Saigon-Dongnai River system, covering an area of 16 km^2. This area belongs to Binh Thanh district and District 2 (Figure 5. 1), and is an example of low-lying lands that are prone to

frequent flooding caused by high tide, heavy rainfall and high-water discharge released from the upstream reaches of the river basin. Binh Thanh district represents an urban district with a very high population density, while District 2 represents a semi-urban district, and is considered to become the most strategic metropolitan area in HCMC in the near future. District 2 and Binh Thanh district are estimated to be severely impacted by extreme floods in 2050, with approximately 94% and 82% of their area flooded, respectively (World Bank, 2010). The main land use in this area is residential buildings (i.e. built-up land), intermingled with a few areas of fallow lands. The width of the Saigon River in this area varies between 250 and 350 m, while its average depth is about 20 m.

5.3.6 Flood hazard classification

Flood hazard maps provide essential information for flood risk management and mitigation purposes. In most flood hazard studies, floodwater depth is widely used to classify a hazard index (Sharif et al., 2016, Komi et al., 2017, Garrote et al., 2016, Alfieri et al., 2014). Nevertheless, flood hazard includes many elements, such as the stability of human bodies, buildings and vehicles in floodwaters (Xia et al., 2011). Therefore, a single parameter cannot completely assess the potential damage of flood flows on people, buildings and vehicles. In previous studies, a combination of flood depth and velocity has been used as a proxy for the force of floodwaters to access the instability of the human body and vehicles, as well as the failure of buildings in floodwaters (Xia et al., 2011, Xia et al., 2014, Kreibich et al., 2009). Furthermore, Smith et al. (2014) and AEMI (2014) have indicated that the level of vulnerability of a community is dependent on the strength of the floodwaters, which can be simply described by the depth and speed of floodwaters. They also suggested that flood hazard maps may be classified using combined flood hazard curves derived from flood depth and velocity thresholds. This method provides a basis for categorizing flood hazard based on the intensity of D×V, and it was used in this study. As such, the classification of flood hazard was based on a six-grade scale, ranging from H1 to H6. The class H1 represents a very low hazard, which is generally safe for people, vehicles and buildings, whereas H6 is unsafe for people, vehicles and buildings. The flood hazard classification limits (AEMI, 2014) are presented in Table 5. 1. For further details about flood hazard classification, the reader is referred to Smith et al. (2014) and AEMI (2014).

Table 5. 1: Limits for the flood hazard classification

Hazard class	Flood depth, D (m)	Velocity, V (m/s)	Classification limit ($D \times V$) (m^2/s)	Description
H1	0.3	2.0	≤ 0.3	Generally safe for vehicles, people and buildings
H2	0.5	2.0	≤ 0.6	Unsafe for small vehicles
H3	1.2	2.0	≤ 0.6	Unsafe for vehicles, children and the elderly
H4	2.0	2.0	≤ 1.0	Unsafe for vehicles and people
H5	4.0	4.0	≤ 4.0	Unsafe for vehicles and people; all buildings types vulnerable to structural damage
H6	-	-	> 4.0	Unsafe for vehicles and people; all buildings types vulnerable to failure

5.4 Results and discussions

5.4.1 Flood frequency analysis

The best GEV models were fitted to annual maximum daily rainfall, annual maximum daily water level and annual maximum daily discharge for all gauging stations and the results are presented in Table 5. 2, which also shows the parameter values of appropriate GEV models from the results of frequency analysis. The models with lower AICc should be preferred to those with higher AICc, and the best model is identified as the model with Δ_i equal to zero.

Table 5. 2: Results of the best GEV models and parameter estimation. The best model for each station is based on the lowest value of AICc

Station	Model	μ_0	μ_1	μ_2	μ_3	σ_0	σ_1	σ_2	ξ	Δ_i*
Rainfall										
Ben luc	GEV-2	90.98	-9.66	-	-	22.47	-	-	0.17	1.50
Bienhoa	GEV-0	86.43	-	-	-	21.56	-	-	0.10	-
Binhchanh	GEV-1	68.67	-12.27	-	-	17.51	-	-	0.31	3.25
Candang	GEV-0	85.45	-	-	-	21.86	-	-	0.05	-
Cangio	GEV-7	44.00	-10.81	45.15	-	20.83	-	-	0.35	14.12
Chonthanh	GEV-0	90.56	-	-	-	23.30	-	-	0.05	-
Cuchi	GEV-1	84.78	-19.00	-	-	21.86	-	-	0.14	4.48
Dautieng	GEV-0	90.26	-	-	-	22.14	-	-	0.13	-
Dongphu	GEV-0	97.41	-	-	-	25.31	-	-	0.26	-
Godau	GEV-4	75.88	-12.30	-	-	20.41	-	-	-0.18	0.39
Hocmon	GEV-0	77.74	-	-	-	22.93	-	-	-0.22	-
Longthanh	GEV-5	92.31	16.16	-	-	23.52	-	-	0.22	2.71
Macdinhchi	GEV-3	102.38	-35.12	-	-	21.43	-	-	0.02	0.02
Nhabe	GEV-20	77.85	3.65	-5.30	-	3.32	1.28	-0.79	-0.27	4.20
Phuochoa	GEV-0	95.87	-	-	-	29.05	-	-	0.10	-
Sosao	GEV-0	87.15	-	-	-	21.31	-	-	-0.10	-
Tanan	GEV-23	84.12	-15.28	-21.93	-	2.88	0.72	-1.24	0.32	12.61
Tayninh	GEV-2	94.78	-6.91	-	-	17.86	-	-	0.28	1.31
Thuduc	GEV-10	130.55	-174.18	59.32	-	21.34	-	-	0.26	3.33
Trian	GEV-0	91.65	-	-	-	17.45	-	-	-0.02	-
TSN	GEV-12	88.20	15.37	-6.68	-18.58	13.17	-	-	0.34	2.83
Vungtau	GEV-0	86.17	-	-	-	22.65	-	-	0.31	-
Water level										
Vungtau	GEV-3	1.24	0.002	-	-	0.06	-	-	-0.03	10.07
Mochoa	GEV-0	1.82	-	-	-	0.47	-	-	-0.16	-

Station	Model	μ_0	μ_1	μ_2	μ_3	σ_0	σ_1	σ_2	ξ	Δ_i*
Discharge										
Godau	GEV-0	69.83	-	-	-	22.79	-	-	-0.10	-
Dautieng	GEV-0	113.72	-	-	-	97.73	-	-	0.18	-
Phuochoa	GEV-0	949.74	-	-	-	263.06	-	-	-0.11	-
Trian	GEV-0	2091.25	-	-	-	500.60	-	-	-0.09	-

*Denotes Δ_i between the best model and GEV-0.

The results presented in Table 5. 2 show that the nonstationary GEV models are superior to the stationary GEV models at 12 of the 22 stations studied, while there was no evidence of nonstationarity at the remaining rainfall stations. Taking TSN station as an example, the GEV-12 model was found to be the best model for extreme rainfall based on Δ_i. In the GEV-12 model, the linear trend is represented by the location parameter with three covariates of ENSO, PDO and local temperature, with a value of Δ_i between GEV-0 and GEV-12 is 2.83.

For Vungtau station (sea level), the GEV-3 model that considered global sea level rise as a covariate was found to be the best model. The stationary model is ranked eight among the 15 models, and the value of Δ_i for GEV-3 is 10.07.

The discharge time series at Godau, Dautieng, Phuochoa, Trian stations and the water level time series at Mochoa were assumed to be under stationary conditions, and the maximum-likelihood estimates for location, scale and shape parameters in the stationary models for these stations are also given in Table 5. 2.

Based on the best models, the estimated values of rainfall, sea/water level and discharge corresponding to return periods of 2, 5, 10, 25, 50, and 100 years are provided in Table 5. 3. It can be seen from Table 5.3 that there is a huge difference in the magnitudes of extreme rainfall between gauging stations in the surveyed basin. For example, the extreme rainfall for the 100-year return period at Hocmon station is even lower than the 5-year return period value at Tanan station. Three flood scenarios based on the values of rainfall, sea/water level and discharge corresponding to 25-, 50- and 100-year return periods, presented in Table 5. 3, were used in flood simulation.

Table 5. 3: Variation in return levels for different return periods

Station	Return period (years)					
	2	5	10	25	50	100
Rainfall (mm)						
Benluc	110.54	140.51	163.83	197.94	227.11	259.78
Bienhoa	94.48	121.31	140.83	167.67	189.29	212.31
Binhchanh	86.00	112.72	136.43	175.56	213.10	259.51
Candang	93.54	119.54	137.61	161.46	179.91	198.91
Cangio	107.24	140.17	170.37	221.81	272.64	337.11
Chonthanh	99.18	126.86	146.08	171.42	191.01	211.16
Cuchi	110.53	138.79	160.13	190.52	215.81	243.51
Dautieng	98.57	126.93	148.15	178.10	202.83	229.73
Dongphu	107.14	143.86	174.87	223.81	268.78	322.37
Godau	92.14	111.79	122.80	134.78	142.46	149.20
Hocmon	85.82	107.12	118.61	130.70	138.19	144.59
Long thanh	116.97	149.80	176.52	217.24	253.46	295.40
Macdinhchi (MDC)	110.27	135.04	151.78	173.30	189.55	205.92
Nhabe	105.61	164.81	195.11	225.54	243.57	258.37
Phuochoa	106.72	142.98	169.40	205.84	235.26	266.64
Sosao	94.82	116.87	130.17	145.64	156.22	166.03
Tanan	110.23	176.09	234.99	332.99	427.69	545.52
Tayninh	109.59	135.98	158.69	195.12	229.14	270.24
Thuduc	109.54	140.58	166.86	208.42	246.70	292.40
Trian	98.02	117.37	129.91	145.44	156.74	167.78
TSN	106.87	127.42	146.02	177.33	207.91	246.32
Vungtau	94.96	129.51	160.11	210.59	258.97	318.75
Water level (m)						
Vungtau	1.40	1.47	1.51	1.56	1.60	1.64
Mochoa	1.99	2.45	2.72	3.01	3.20	3.37
Discharge (m^3/s)						
Godau	78.03	101.55	115.71	132.14	143.35	153.72
Dautieng	150.74	281.86	384.51	535.57	665.41	811.51
Phuochoa	1044.26	1313.85	1474.94	1660.57	1786.44	1902.23
Trian	2271.68	2792.79	3109.19	3479.07	3733.43	3970.20

5.4.2 Model calibration and validation

The hourly water level data for the rainy period (from 01-09-2012 to24-09-2012) at four gauging stations (Benluc, Bienhoa, Nhabe, TDM) were used to calibrate the hydrodynamic model. The results of the simulations are presented in Table 5. 4. It may be seen that the RSR for hourly water level varies from 0.32 to 0.66, while R^2 values vary from

0.88 to 0.97. The hourly NSE values range between 0.61 and 0.89, with an average value of 0.71.

Table 5. 4: Calibration and validation results – 1D model

Station	Calibration (hourly water level)			Validation (hourly water discharge)		
	R^2	RSR	NSE	R^2	RSR	NSE
Benluc	0.97	0.55	0.7	0.91	0.4	0.84
Bienhoa	0.88	0.66	0.61	0.77	0.7	0.51
Nhabe	0.96	0.32	0.89	0.81	0.65	0.58
TDM	0.97	0.59	0.65	0.93	0.42	0.82

Validation of the model was performed using hourly discharge data for the period 19-09-2011 to 04-10-2011. The model performance at Benluc and TDM stations was good, with NSE values of 0.84 and 0.82, respectively (Table 5. 4), while the NSE values for Bienhoa and Nhabe were much lower (0.51 and 0.58, respectively).

Figure 5. 5 shows the comparison of observed and simulated hourly water level under calibration (Figure 5. 5a) as well as hourly discharge under validation (Figure 5. 5b) There is a good level of agreement between the observed and simulated water level and discharge at all the gauging stations. Therefore, the model can be used appropriately for the subsequent simulation.

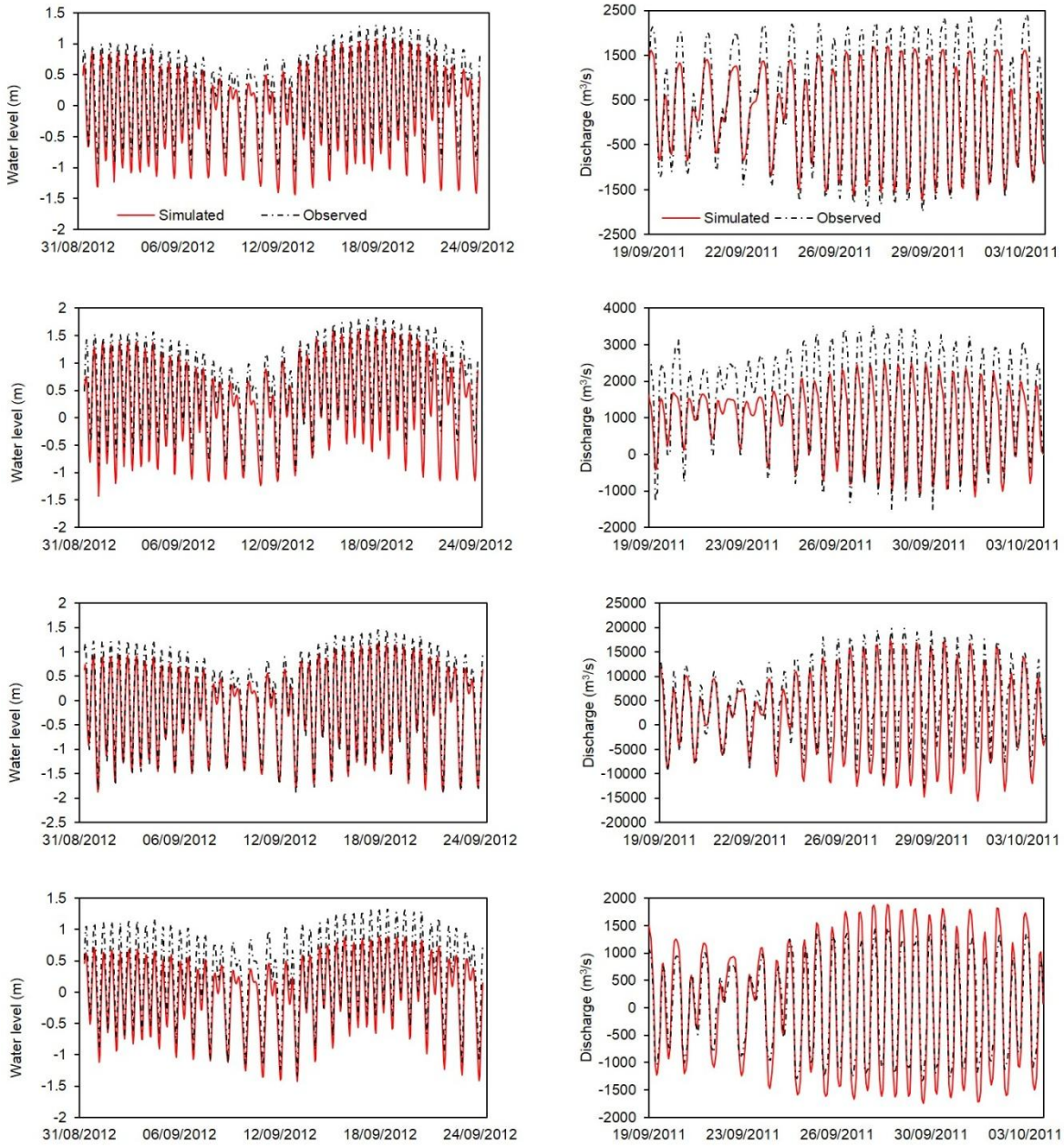


Figure 5. 5: Results of (a) calibration and (b) validation of the flood simulation model at, from top to bottom: Benluc, Bienhoa, Nhabe and TDM stations

5.4.3 Flood hazard maps

Based on the developed model, the flood hazard maps derived from the combination of extreme rainfall, water level and upstream outflows are presented in Figure 5. 6. The maps

represent three different scenarios, 1, 2 and 3 for 25-, 50- and 100-year return period, respectively. The enlarged views are shown in Figure 5. 7a-c, focusing on the large flooded areas (denoted by A1, A2 and A3). It can be seen that areas where the river is surrounded by houses and buildings, or where the riverbed is narrow, the flood inundation extent is wider. Taking a closer look at areas A1, A2 and A3, the floodwater flows freely over the riverbank along roads and alleyways.

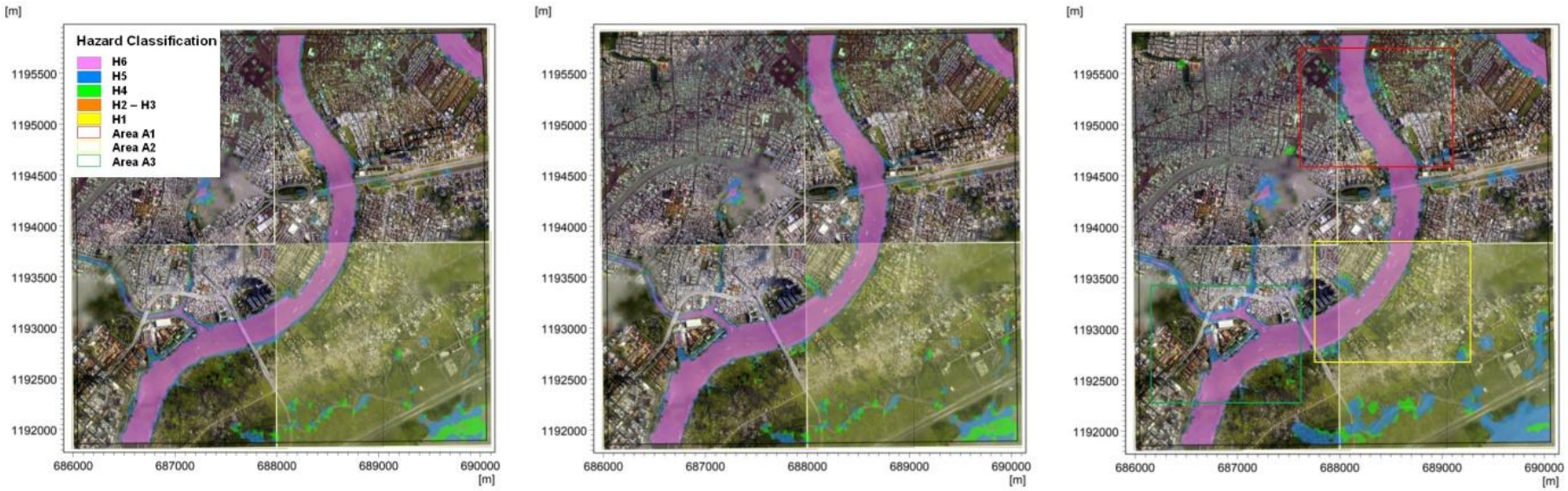


Figure 5. 6: Comparison of flood hazard maps between flood scenarios, with return periods of (a) 25 years, (b) 50 years and (c) 100 years. Flood hazard is classified by considering the combination of flood depth and velocity, ranging from H1 to H6 (see Table 5. 1)

The flood hazard is classified based on six levels (refers to Table 5. 1 in Section 5.3.6), and the area of each flood hazard zone in relation to the entire study area is presented in Figure 5. 7d. However, as the area of H6 flood hazard zone mostly covers the main river surface, it is not shown in Figure 5. 7d. It is clear that the total flooded area increases corresponding to flood events. In particular, the H1 flood hazard zone, which is generally safe for all people, vehicles and infrastructure, covers areas of 0.15, 0.19 and 0.27 km^2 under scenarios 1, 2 and 3, respectively. The cumulative area of H2, H3 and H4 flood hazard zones, where the water depth varies from 0.5 to 2.0 m under Scenario 1, is 0.73 km^2, while for scenarios 2 and 3, the corresponding areas are 0.77 and 0.82 km^2, respectively. Finally, the H5 flood hazard zone, which is unsafe for people, vehicles and buildings, covers 0.49, 0.56 and 0.60 km^2 under scenarios 1, 2 and 3, respectively.

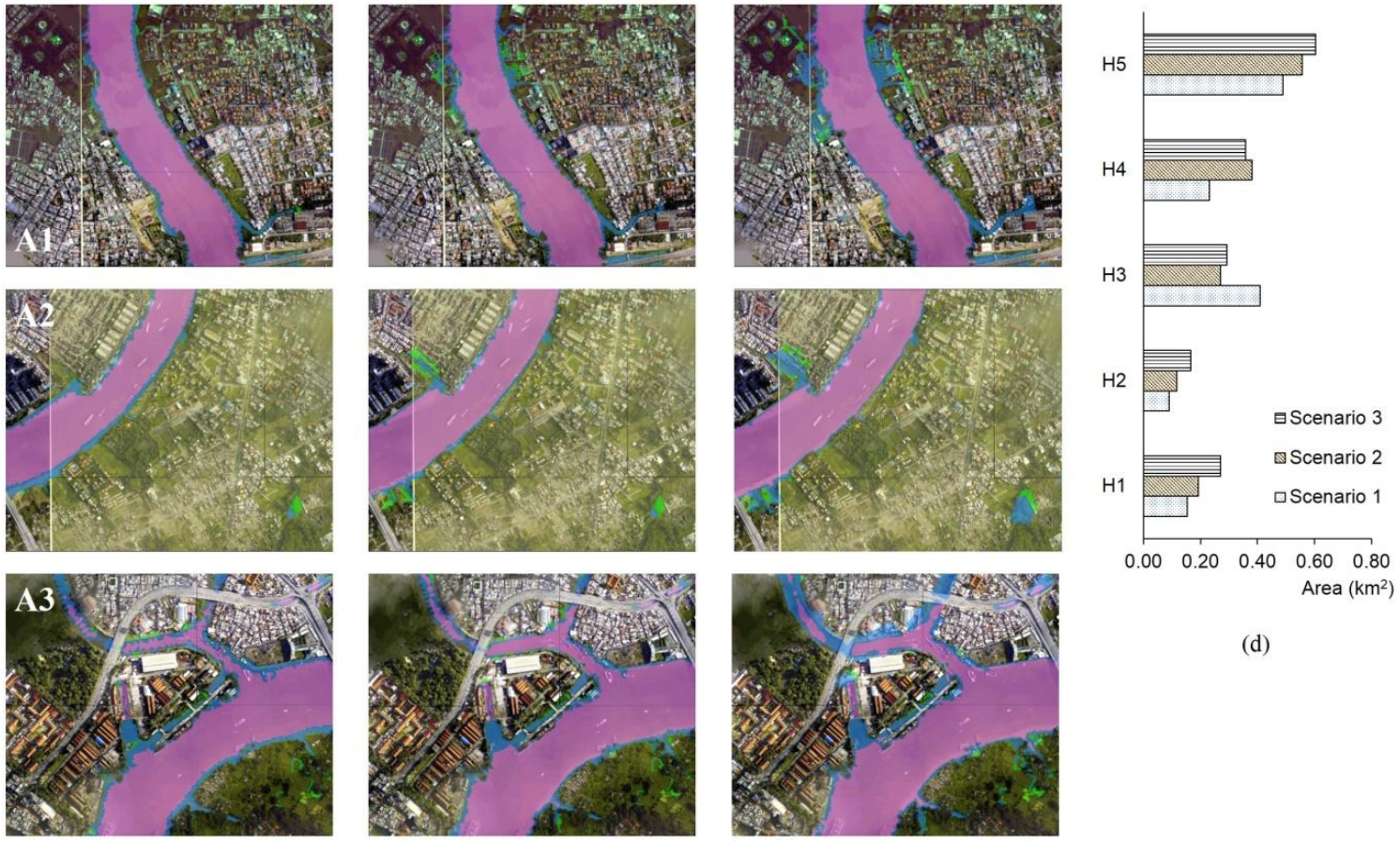


Figure 5. 7: Enlarged view of areas A1, A2 and A3 (Figure 5.5) for different flood scenarios, with return periods of: (a) 25 years, (b) 50 years, and (c) 100 years. (d) Area of each flood hazard zone corresponding to each flood scenario

Furthermore, a floodplain area based on the combination of 50-year extreme rainfall, extreme sea level and extreme discharge under stationary conditions (denoted Sta-50) was developed for comparison purposes and is shown in Figure 5. 8. It can be seen that the flood extent under nonstationary conditions is larger than under stationary condition. The flooded area under Scenario 2 is 0.29 km^2 larger than that under Sta-50. The enlarged views (Figure 5. 8a and b) show two areas in which the difference in flooded area between the two scenarios (Scenario 2 and Sta-50) is particularly remarkable.

The results demonstrate how flood potential can influence human settlements in the study area. Compared to the flood simulation methods used in previous studies (Dang and Kumar, 2017, Lasage et al., 2014), the current method has applicability in flood simulation when considering the nonstationary behavior in time series of extreme events. In comparison with Storch and Downes (2011), ADB (2010) and World Bank (2010), our results show that the flooded areas are discrete and concentrated mainly on both sides of the Saigon River rather than covering entire study area. More importantly, the high-resolution images from our study provide a clear and detailed view of the flooded area, which can contribute to efficient flood risk management, as well as the provision of mitigation strategies by the local government.

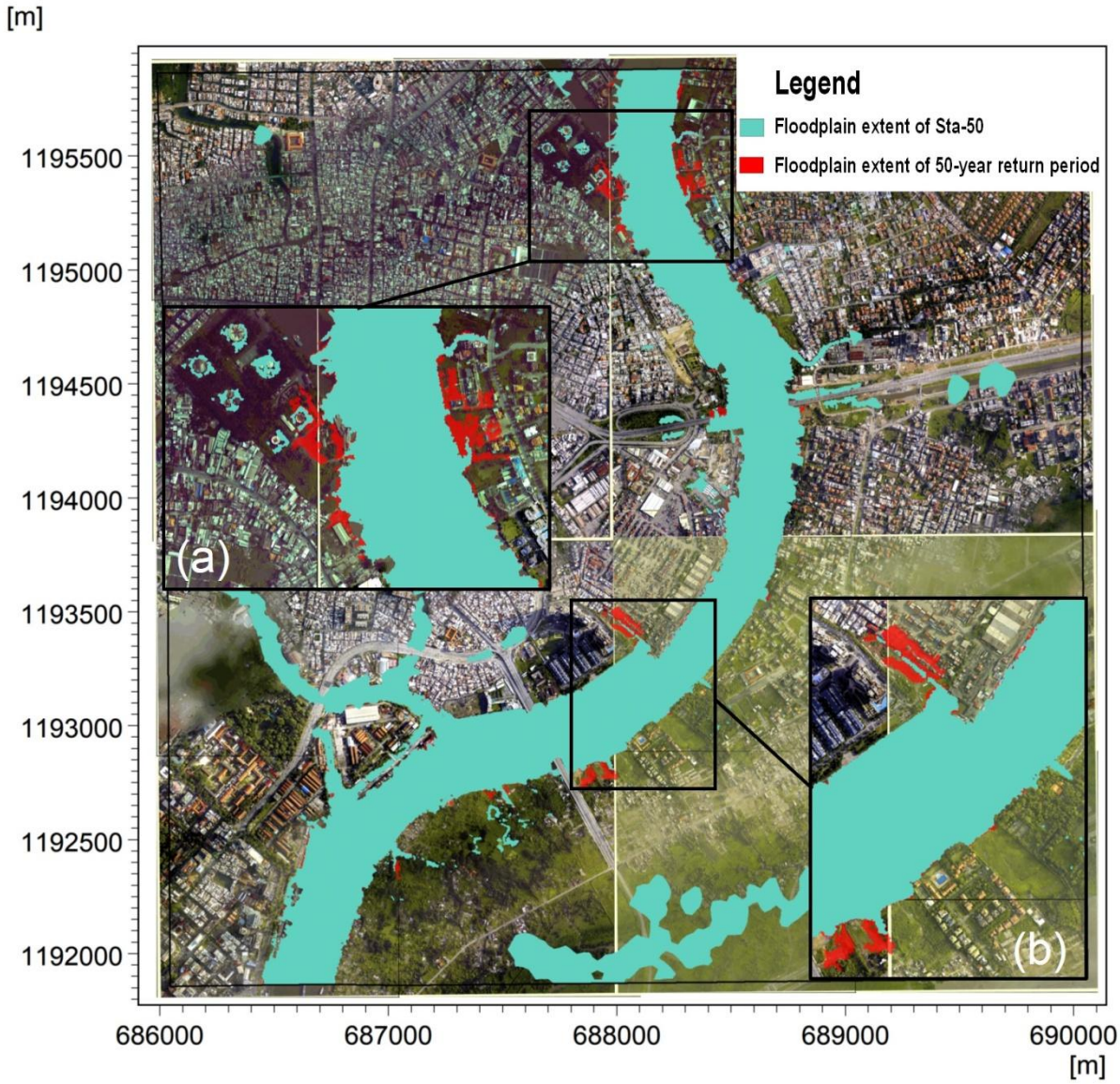


Figure 5. 8: Comparison of floodplain extents of 50-year return period and Sta-50, with (a) and (b) showing zoom-in images of the flooded areas. Sta-50: 50-year extreme rainfall, extreme sea level and extreme discharge under stationary conditions

5.5 Summary and conclusions

One of the key factors in preventing and reducing flood damage and the number of lives lost is the provision of flood risk assessment information through flood hazard maps. In HCMC, extreme rainfall and extreme sea level are considered as the main factors impacting significantly on floodplain extent. These factors are currently proved to have nonstationarity in their time series. It is believed that this study is the first study to develop a flood simulation

model considering nonstationarity in two time series of flood components, i.e. extreme rainfall and extreme sea level; this is different from other research on flooding in HCMC. Moreover, the effects of upstream outflows on flood inundation in HCMC was also considered, which contributed to the model development process. For this purpose, climate indices (ENSO and PDO), global and local temperature, and global mean sea level were used for frequency analysis to investigate the nonstationarity in the extreme rainfall and sea level. The covariates in the best statistical model were attributed as the most significant physical processes causing nonstationarity in the time series. The results of the frequency analysis show that ENSO and PDO are present in the best nonstationary models, hence they can be considered as the main causes of nonstationary behavior in extreme rainfall in the study area. Another finding indicates that global sea level rise has a significant effect on nonstationarity in extreme sea level at Vungtau station.

Flood scenarios were developed based on the results of frequency analysis corresponding to 25-, 50- and 100-year return periods. The calibration and validation results of the MIKE FLOOD model show that the model performs satisfactorily in simulating water flow for the study area. Another notable highlight in this thesis is the use of high-resolution data (i.e. LiDAR) in developing the flood hazard maps of HCMC, which has not been done in previous studies. In this way, the floodplain extent can be defined, and the passage of floodwater can be detected clearly, even at the scale of street networks. The results from the spatial variation of flood hazards indicate that locations along both riverbanks are expected to experience a significant increase in flooded area, with the intensity of D×V reaching 4.0 m^2/s. The percentages of flooded area classified as zones H1–H5 are 8.54%, 9.43% and 10.51% for scenarios 1, 2 and 3, respectively. It is also to be noted that the floodplain extent is larger when based on the assumption of nonstationarity.

Considering the important role of flood hazard mapping and estimating floodplain extent in decision making, or establishing flood warning systems, it is suggested that the flood sources (e.g. rainfall, water level, upstream outflow and sea level rise) are analyzed under both stationary and nonstationary conditions before being used as initial inputs of hydrological and hydrodynamic models, since the global climate is continuously changing and unpredictable. The present work has successfully introduced nonstationarity into flood frequency analysis, thereby influencing the appearance of flood hazard maps. Nevertheless,

several limitations have to be acknowledged. First, the use of 35 years of data might raise concerns about the uncertainty associated with parameter estimation, hence impacts on the extreme value estimates. Second, the missing calibration and validation of the 2D model might lead to an imperfect flood simulation model. However, it could be a suitable option for flood simulation in HCMC since data is not available for a longer duration in this large river basin. The study provides a new approach for flood simulation in HCMC that can be referred to by managers and decision makers, especially for purposes relating to the construction of buildings and infrastructure in flood hazard areas. Moreover, the proposed flood simulation model may be used not only for HCMC but also for other practical cases within the Saigon-Dongnai River basin, where insufficient attention has been paid to detailed flood maps.

Chapter 6

Copula-based bivariate flood frequency analysis under nonstationary condition

6.1 Introduction

Coastal zones are commonly vulnerable to floods caused by the combination of multiple sources. In such areas, heavy precipitation can coincide with coastal storm surges then flooding could be twice as severe, thereby resulting in massive damage to the coastal inhabitants and infrastructures (Karamouz et al., 2014, Hunt, 2005). For these regions, univariate frequency analysis may no longer be effective to describe floods that are characterized by several correlated variables. It is, therefore, necessary to consider the joint probability of flood-caused sources in the evaluation and management of flood risk.

Multivariate parametric distributions have been widely used to model multivariate extreme events (i.e. flood). Nevertheless, this approach contains some limitations since the correlated variables have the same type of marginal distribution, and the dependence structure

could not be described differently from the marginal distribution (Zhang and Singh, 2007). Therefore, the copula theory has been paid more attention and increasingly employed for simulating joint distribution due to its flexibility. The main advantage of using copula functions is that multivariate random events could be described by different family of distributions, and the dependence structure is freely modeled without concerning the marginal distributions. The application of copulas in the fields of hydrology and water engineering could be found in many studies (Renard and Lang, 2007, Zhang and Singh, 2007, Sraj et al., 2015, Masina et al., 2015).

Frequency analysis of hydro-meteorological events has been commonly based on the assumption of stationarity in time series. However, under climate variability related to human activities, changes in rainfall characteristics and hydrological cycle have been reported for many places of the world, which proposes that the assumption of stationarity in hydro-meteorological time series may not valid anymore (Sugahara et al., 2009, Khaliq et al., 2006, Milly et al., 2008, Berg et al., 2013, Šraj et al., 2016, Mudersbach and Jensen, 2010). Till now, the use of nonstationary approach in hydro-meteorological frequency analysis could be found in many studies which mainly focused on a single random variable, but only a few studies available related to nonstationary multivariate frequency analysis. For example, the time-dependent behavior of bivariate hydrological design parameters is investigated by Bender et al. (2014) using nonstationary copula approach. Sarhadi et al. (2016) applied the nonstationary copula analysis to estimate the time-varying joint return period of drought characteristics under nonstationary condition. Jiang et al. (2015) used nonstationary copula to estimate the joint return period of low-flow at two neighboring hydrological stations on the Hanjiang River. Similarly, Ahn and Palmer (2016) performed the nonstationary bivariate frequency analysis for low-flow characteristics in the Connecticut river basin.

However, most of these studies used time as the explanatory variable of the marginal distribution and dependence parameters. As a matter of fact, the change of paradigm from stationarity to nonstationarity can be affected by many causes (e.g. LULC change or climate change). The effects of these factors may not exactly follow the passage of time. More importantly, the changes in the climate or watershed characteristics have affected not only in the nonstationarity of individual hydrological series but also in the dependence structure

between the different hydrological series. In addition, the dependence relationship between different variables could be varied over time, if there is a change in the meteorological forcing processes. Indeed, the study of De Michele et al. (2007) indicated that the dependence of wave height and duration as a function of storm magnitude. Similarly, Corbella and Stretch (2013) showed that the dependence of wave height and period as a function of peak wave power.

Therefore, using time as a covariate in nonstationary modelling may suffer some limitations (Ahn and Palmer, 2016). Agilan and Umamahesh (2016b) indicated that using time as a covariate in nonstationary models is increasing the bias of nonstationarity analysis. López and Francés (2013) also suggested that only time is considered as a covariate may not be entirely correct because the trends can change in the short and long term caused by climate variability and the intensification of human activities. Besides, Šraj et al. (2016) showed the problem with time-varying distribution parameters is that it is difficult to explain why time-varying distribution parameters would change continually in the future in the same way that they did in the past. Jiang et al. (2015) also suggested that physical covariates may be more effective in nonstationary modelling than using time as a covariate. Furthermore, the physical meaning of the estimated parameters will be ambiguous if time is used as covariates in the nonstationary setting.

Recently studies indicated that the hydro-meteorological patterns are strongly affected by local influences within the catchment, hence the covariates selected for nonstationary modelling should have strong physical associations with the process of events (Arns et al., 2013a, Yan et al., 2017). Furthermore, Agilan and Umamahesh (2016b) suggested that covariates should be chosen with respect to the area of interest because the selected covariates for a particular area may have different effects in other geographical areas. Therefore, local variables which reflect the physical relationship to hydro-meteorological events should be considered in nonstationary multivariate frequency analysis.

In hydro-meteorology frequency analysis, the AM and POT approaches have been usually used to extract data sample in extreme values analysis. The AM approach in which considers the maximum value of each year is regarded to be better and safer for the sampling of hydro-meteorological events (Porter and McMahon, 1971). However, flood events can

occur more than once within a year, and a short sample is a critical problem in parameter estimation (Sugahara et al., 2009). Unlike the AM approach, the POT approach considers wider range of events by selecting all excesses over a certain threshold. For low-land areas, especially coastal zones, heavy rainfall can occur several times within any year and coincide with coastal storm surges, which results in flooding. Hence, the extremes may not be missed when POT approach is adopted. In this study, the comparison between different sampling selection approaches is carried out, which may provide more information for analysis.

The main objective of this study is model the joint probability using nonstationary copula-based bivariate frequency analysis for HCMC. To obtain the objective, the nonstationary behavior is modeled for dependence structure and marginal distributions by using local physical processes as covariates. The joint return period (i.e. OR and AND) of selected variables is estimated by the optimal copula and marginal distribution. The results of the joint return periods are presented for three samples to provide important information in choosing appropriate data time series for flood defense design.

6.2 Data

6.2.1 Water level and rainfall

Hourly observed water level data covering a period of 1981- 2014 at Nhabe station along the Saigon-Dongnai River basin are used in this study. Daily rainfall data for the period of 1981 to 2014 collected from Cangio meteorological station. The locations of Nhabe and Cangio stations can be seen in Figure 5.1 in Chapter 5.

In this study, to assess the sensitivity of results to different sampling selections, three samples are extracted, comprising AM and POT approaches as shown below:

- Sample 1: the annual maximum daily rainfall and the annual maximum daily water level for each year are selected
- Sample 2: the annual maximum daily rainfall is firstly obtained, and then the highest water level is chosen within the day of this event
- Sample 3: the POT series of 97.5^{th} percentile daily rainfall is used, and then the highest water level is selected within the day of this event.

For Sample 3, the threshold value of 97.5^{th} percentile of daily events is used to ensure that the threshold should be high enough for the values to be extreme, but not so high to avoid high variance in the model estimates. The duration of 3 days is used to ensure independent events based on typical storm duration on the HCMC's coast. It means that the selected event is automatically separated by 3 days.

6.2.1 Covariates

In this study, to reduce the uncertainty in modelling extreme events, only local variables (i.e. local temperature, urbanization and local sea level) which have a strong relationship to the events (rainfall and water level) are considered as covariates. The temperature data during the period 1981-2014 recorded at TSN was provided by the Southern Institute for Water Resources Planning. Hourly sea level was recorded at Vungtau station from 1981 to 2014 and was provided by the NHMS of Vietnam. The data for urbanization is selected as mentioned in Section 4.2.2 in Chapter 4.

6.3 Methodology

In this study, the time series of rainfall and water level are firstly tested for stationarity by M-K non-parametric test. The dependence between rainfall and water level are evaluated using the Kendall, Spearman and Pearson correlation coefficients. Four probability distributions comprising GEV, Gumbel (GUM), Log-normal (LN) and Weibull (WEI) are used to model rainfall and water level processes within the contexts of nonstationarity. Next, nonstationary copulas are developed to simulate the relationship between rainfall and water level. Finally, the joint return periods of rainfall and water level of the three samples are estimated by the optimal copulas and marginal distributions.

6.3.1 Detecting the nonstationary component in time series

As mentioned earlier, testing nonstationarity in time series should be firstly considered, which crucially influences the statistical analysis result. In this study, the M-K trend test with a significant level of 0.05 is used to identify the trends in rainfall and water level time series.

6.3.2 The dependence analysis for selected variables

The Kendall, Spearman and Pearson correlation coefficients are employed to assess the dependence between the pairs of variables, while K plot (Genest and Boies, 2003) is used for graphical presentation purpose. The correlation coefficient varies between -1 to +1, where -1 and +1 represent a monotonic association between the two variables, and 0 corresponds to the independence. Kendall's τ and Spearman's ρ are two widely used non-parametric tests of detecting the dependence between random variables, and they are not based on any assumptions about the distribution of the data. Meanwhile, Pearson correlation test is based on the assumption of normal distribution of the variables. Besides, the pair of extremal measures (χ and $\bar{\chi}$) (Coles et al., 1999) is also used to detect asymptotical dependence between rainfall and water level. For more details about two measures of extreme dependence (χ and $\bar{\chi}$), the reader is referred to Coles et al. (1999).

6.3.3 Modelling the marginal distributions for both stationary and nonstationary

The probability distribution functions commonly used in hydro-meteorology, namely GEV, GUM, LN and WEI, are chosen for modelling the marginal distribution for both rainfall and water level. The three parameters, namely location, scale and shape, of the marginal distribution are demoted as μ, σ and ξ. The nonstationarity is introduced only in location parameter, while the scale and shape parameters are kept constant. As such, the location parameter can be expressed as a linear function of covariates as follows:

For rainfall frequency analysis

$$\mu(t) = \mu_0 + \mu_1 U + \mu_2 T$$
$$\sigma(t) = \sigma \quad (6.1)$$
$$\xi(t) = \xi$$

For water level frequency analysis

$$\mu(t) = \mu_0 + \mu_1 U + \mu_2 S$$
$$\sigma(t) = \sigma \quad (6.2)$$
$$\xi(t) = \xi$$

where U, T and S denote Urbanization, Temperature, and Sea level respectively. In the stationary model, the coefficients of covariates U, T, and S equal zero. Apart from stationary model, based on two covariates and their combination, three nonstationary models are constructed for each candidate distribution in rainfall/water level frequency analysis. List of marginal distributions and considered covariates used for rainfall and water level frequency analysis is shown in Table A. 3.

The distribution parameters, μ_0, μ_1, μ_2, σ and ξ, are estimated by maximum likelihood method. The AICc is used to identify the best fitting distribution for rainfall and water level. Once the best fitting distribution is selected, the *T*-year return level is calculated based on the low-risk approach mentioned in Section 5.3.1 in Chapter 5.

6.3.4 Modelling the dependence structure for both stationary and nonstationary

Copulas are defined as multivariate distribution functions which link joint probability distributions to their one-dimensional marginal distributions (Badrzadeh et al., 2015). For a bivariate case, a copula function can be expressed as Eq. (6.3) (Chang and Chang, 2006)

$$F(x,y) = C[F(x), F(y)] = C(u,\upsilon) \quad (6.3)$$

where F(x,y) is the joint cumulative distribution function for X (rainfall) and Y (water level) with marginal distribution F(x) and F(y) respectively. C (u,υ) is the copula function of the marginal distribution of rainfall and water level, u = F(x) and υ = F(y) refer to the cumulative frequency distribution respectively.

In order to find the appropriate copulas that best fit to the data sample, several copula functions (i.e. Frank, Clayton, Gaussian and Plackett) are used to model the dependence structure between rainfall and water level in this study. The descriptions of these copula functions along with their parameter are briefly introduced in Table 6. 1. For more detailed description of copulas, the reader is referred to Badrzadeh et al. (2015) and Vaze et al. (2011).

Table 6. 1: Description of the four candidate copulas

Copula	$C_\theta(u,\upsilon)$	Parameter
Frank	$-\frac{1}{\theta} ln[1+\frac{(e^{-\theta u}-1)((e^{-\theta \upsilon}-1)}{e^{-\theta}}]$	$\theta \in (-\infty, \infty)\backslash\{0\}$
Clayton	$(u^{-\Theta}+\upsilon^{-\Theta}-1)^{-1/\Theta}$	$\theta \in (0, \infty)$
Gaussian	$\Phi[\phi^{-1}(u),\phi^{-1}(\upsilon)]$	$\theta \in (-1, 1)$
Plackett	$\frac{1}{2}\frac{1}{\theta-1}\{1+(\theta-1)(u+\upsilon)-[1+$ $\theta-1u+\upsilon 2-4\theta\theta-1u\upsilon 12$	$\theta \in (0, \infty)$

$C_\theta(u,\upsilon)$: copula function, θ: copula parameter, ϕ: cumulative density function of the standard normal distribution and Φ: cumulative density function of the multivariate normal distribution

Similar to the marginal distribution modelling, the copula parameters are modeled using a linear function of covariates of urbanization, local temperature and local sea level as follows:

$$\theta(t) = \theta_0 + \theta_1 \mathrm{U} + \theta_2 \mathrm{T} + \theta_3 \mathrm{S} \quad (6.4)$$

where slope parameters θ_0, θ_1, θ_2, θ_3 represent the trend in the dependent parameter of copula function. The parameters of copulas are estimated by the pseudo maximum likelihood approach. The most appropriate copula is selected using AICc value. Based on three covariates, thirty-two copula functions are developed, the list of these copulas are shown in Table A. 4.

6.3.5 The joint return periods estimation

The joint return period can be defined based on: (i) the probability of rainfall and water level both exceeding certain threshold values (denoted AND); (ii) the probability that rainfall or water level exceeds its threshold values (denoted OR). The joint return period (i.e. AND and OR) is estimated in this study, and these probabilities can be given as (Vaze et al., 2011):

$$T^{AND} = \frac{\lambda}{P(X \geq x \; and \; Y \geq y)} = \frac{\lambda}{1 - F_X(x) - F_Y(y) + C_\theta(\mathrm{u}, \upsilon)}$$

$$T^{OR} = \frac{\lambda}{P(X \geq x \; or Y \geq y)} = \frac{\lambda}{1 - C_\theta(\mathrm{u}, \upsilon)} \quad (6.5)$$

where λ is the mean interarrival time, typically given in years. In case of block maxima, λ equals 1. When applying POT approach, the mean interarrival time can be defined by the observed events.

6.4 Results

6.4.1 Detecting nonstationarity component in time series

The significant trends in rainfall and water level time series are investigated by M-K test for three samples. The test results are shown in Table 6. 2. From the results, the significant increasing trends ($p << 0.05$) have been found in rainfall and water level series for all samples. Therefore, the nonstationary condition is used to model rainfall and water level in this study.

Table 6. 2: M-K test results

Samples	Variables	*Tau*	*p*-value
Sample 1	Rainfall	0.414	0.0006
	Water level	0.694	2.22E-16
Sample 2	Rainfall	0.414	0.0006
	Water level	0.413	0.0006
Sample 3	Rainfall	0.105	0.0205
	Water level	0.249	2.22E-16

6.4.2 Assessing the dependence of selected variables

The dependence between rainfall and sea level are evaluated by the Kendall, Spearman and Pearson correlation coefficients. The variables of rainfall and water level present a moderate correlation between each other when considering AM Samples. In detail, the values of Kendall's tau for Sample 1 and Sample 2 are 0.317 and 0.319 respectively (Table 6. 3). Conversely, when POT series is considered, rainfall and water level show a weaker dependence.

Table 6. 3: Correlation coefficients between rainfall and water level

Correlation coefficients	Pearson's r	Kendall's tau	Spearman's rho
Sample 1	0.372	0.317	0.429
Sample 2	0.305	0.319	0.423
Sample 3	0.103	0.058	0.091

Figure 6. 1 provides the K plots, the Chi and Chi bar plots for the three pairs of observed rainfall and water level corresponding to three samples. Regarding K plots, there is a possible dependence for Sample 1 and Sample 2, with points locate above the line. In contrast, the dependence for Sample 3 is very weak, with points close to the line. Similarly, the Chi and Chi bar plots also show that the dependence between rainfall and water level for Sample 1 and 2 is stronger than for Sample 3. For example, the value of χ (Figure 6. 1c) is nearly equal to 0, it means that the variables rainfall and water level are independent in many cases.

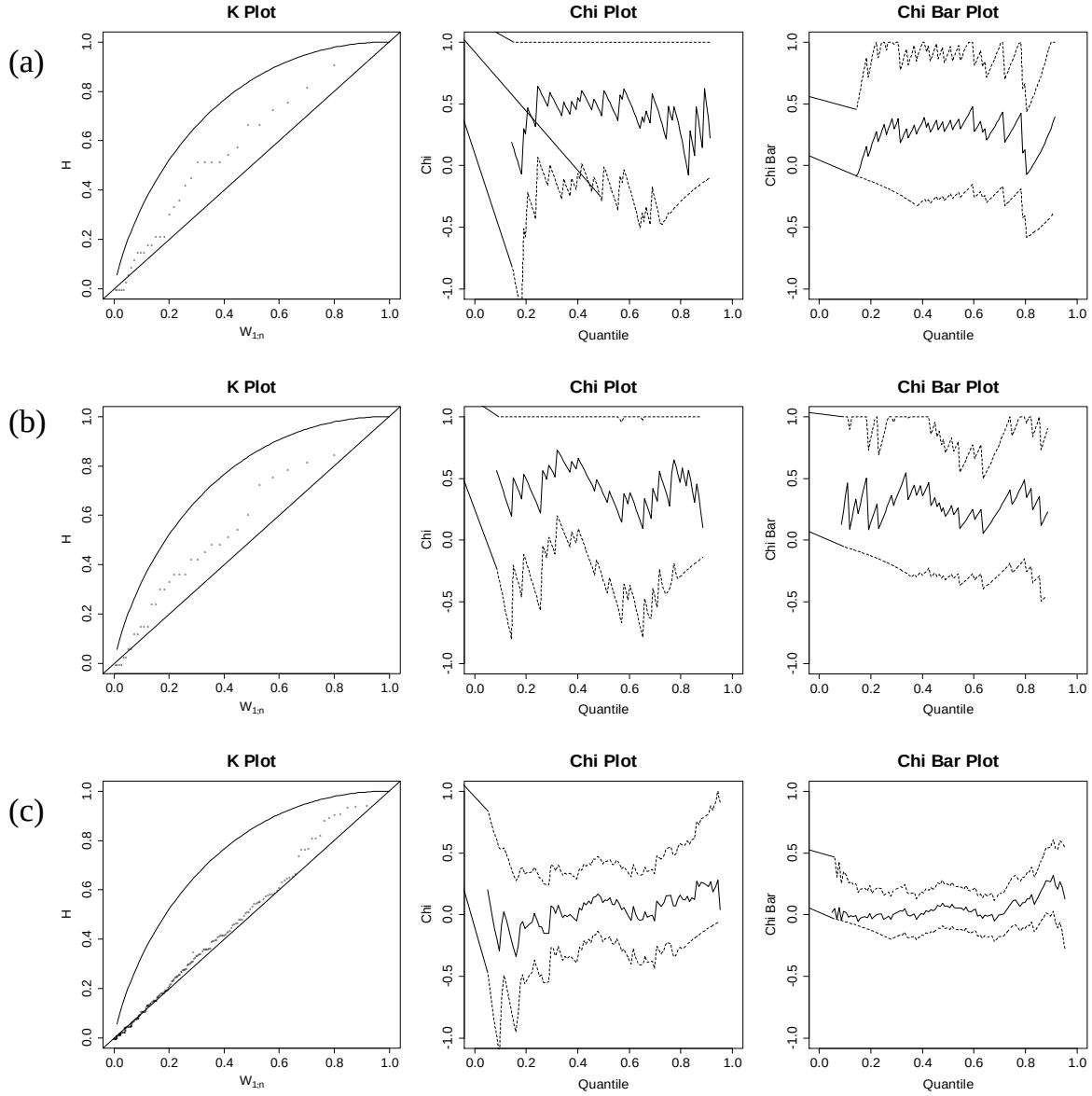


Figure 6. 1: The dependent relationship between rainfall and water level for (a) Sample 1, (b) Sample 2 and (c) Sample 3

6.4.3 Modelling marginal distribution

The values of estimated parameters, along with best fitted marginal distributions for rainfall and water level based on AICc, are presented in Table 6. 4. It can be noted that the model with the lowest value of AICc is identified as the best model for extreme events. The findings show that the nonstationary GEV model based on urbanization covariate (NSGEV-U) is the best model for rainfall for all three samples.

For water level, the nonstationary GUM based on urbanization covariate (NSGUM-U) is considered as the best model for annual maximum daily water level time series (Sample 1). Meanwhile, the NSGEV-US is found to be the most appropriate model for other samples of water level (Sample 2 and 3). In NSGEV-US, the location parameter is expressed as a linear function of urbanization and sea level. For more information about the AICc values for all marginal distributions, reader is referred to Table A. 5.

Figure 6. 2 shows the graphical approach of P-P and Q-Q plots which are used to check the quality of fitting for a chosen model. The P-P and Q-Q plots of stationary GEV (SGEV) and stationary GUM (SGUM) for annual maximum daily rainfall and water level respectively are plotted in Figure 6. 2a and 6. 2c respectively, while those plots of the best nonstationary models (NSGEV-U and NSGUM-U) are shown in Figure 6. 2b and 6. 2d in that order. As can be seen, the best nonstationary models show a better match than the stationary models.

Table 6. 4: The marginal distribution parameters and model selection

Samples	Variables	Best model	Significant covariate(s)	μ_0	μ_1	μ_2	σ	ξ	AICc
Sample 1	Rainfall	NSGEV-U	Urbanization	44.811	21.644	-	19.278	0.351	331.40
	Water level	NSGUM-U	Urbanization	1.385	0.116	-	0.042	-	-102.21
Sample 2	Rainfall	NSGEV-U	Urbanization	44.811	21.644	-	19.278	0.351	331.40
	Water level	NSGEV-US	Urbanization, sea level	0.925	0.074	0.210	0.075	-0.354	-70.73
Sample 3	Rainfall	NSGEV-U	Urbanization	32.181	0.732	-	10.046	0.543	1849.09
	Water level	NSGEV-US	Urbanization, sea level	0.930	0.072	0.201	0.074	-0.299	-520.24

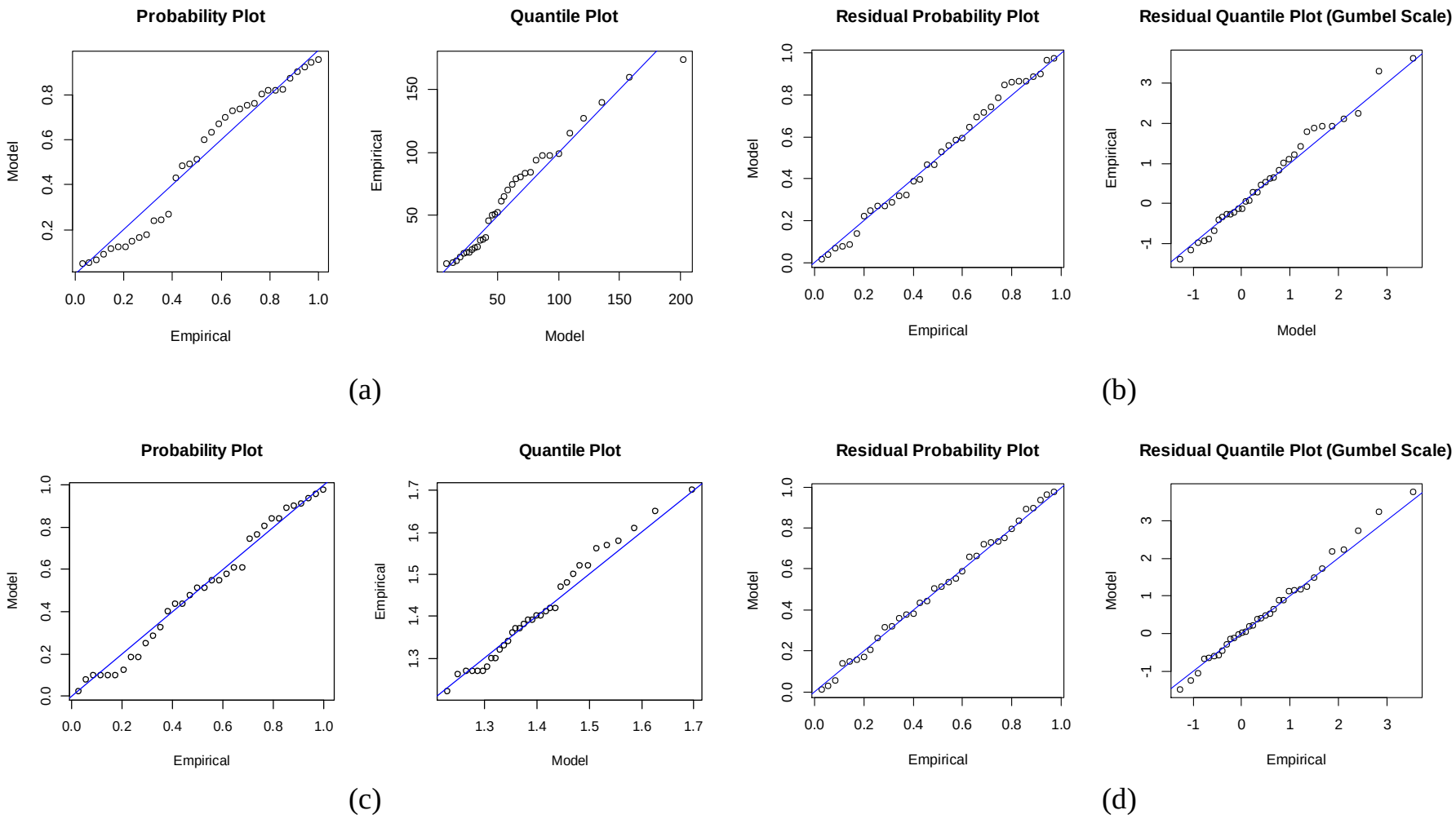


Figure 6. 2: P-P and Q-Q plots of (a) SGEV, (b) NSGEV-U, (c) SGUM and (d) NSGUM-U models for Sample 1

To assess the impact of selection data sample on the marginal distribution modelling, the different return periods are estimated in this study. Table 6. 5 shows the comparison of the return levels based on the best model between the three samples. It is clear that the values of rainfall considering AM series are higher than those values under POT approach. For example, the estimated rainfall level of 2-year return period for Sample 1 corresponds to the 25-year return period for Sample 3. Also for water level, with urbanization as covariate, the nonstationary water level of 2-year return period for Sample 1 is 1.66 m and it is higher than the nonstationary water level of 100-year return level for Sample 3.

Table 6. 5: The return levels based on the best model for rainfall and water level

Return period	Rainfall (mm)			Water level (m)		
	Sample 1	Sample 2	Sample 3	Sample 1	Sample 2	Sample 3
2	122.25	122.25	56.68	1.66	1.43	1.39
10	150.26	150.26	77.71	1.69	1.46	1.42
25	198.03	198.03	120.04	1.73	1.49	1.45
50	245.29	245.29	168.97	1.76	1.50	1.47
100	305.27	305.27	240.05	1.79	1.51	1.49

6.4.4 Modelling the dependence structure

As mentioned earlier, the most appropriate copula is selected based on the AICc value. Table 6. 6 shows the chosen copula followed by its estimated parameters for every sample. As such, nonstationary Clayton copula (NSClayton-US) is most appropriate for Sample 1, with the copula parameter expressed as a linear function of urbanization and sea level. Whilst, stationary Plackett copula (SPlackett) and stationary Gaussian copula (SGaussian) are selected to estimate the joint return period of rainfall and water level for Sample 2 and 3 respectively. The AICc values for copulas are provided in Table A. 6.

Table 6. 6: Parameter estimation and AICc values for selected copulas

Samples	Selected copula	Significant covariate(s)	θ_0	θ_1	θ_2	AICc
Sample 1	NSClayton-US	Urbanization, sea level	-3.308	4.247	-4.137	-7.77
Sample 2	SPlackett	-	1.704	-	-	-6.30
Sample 3	SGaussian	-	0.098	-	-	0.07

6.4.5 Joint return period estimation

The best copula function and marginal distribution are selected to calculate the joint return periods (AND and OR) of rainfall and water level for 5, 10, 50, 100-year periods. Figure 6. 3 provides the joint return periods for Sample 1 which are derived from the best model (Figure 6. 3a) compared to the stationary model (Figure 6. 3b). It is clear that for all return periods, OR-joint return period is higher than AND-joint return period, which occurs for both models. Besides, values of rainfall and water level derived from the stationary model are less than those derived from the best model for almost return periods of AND.

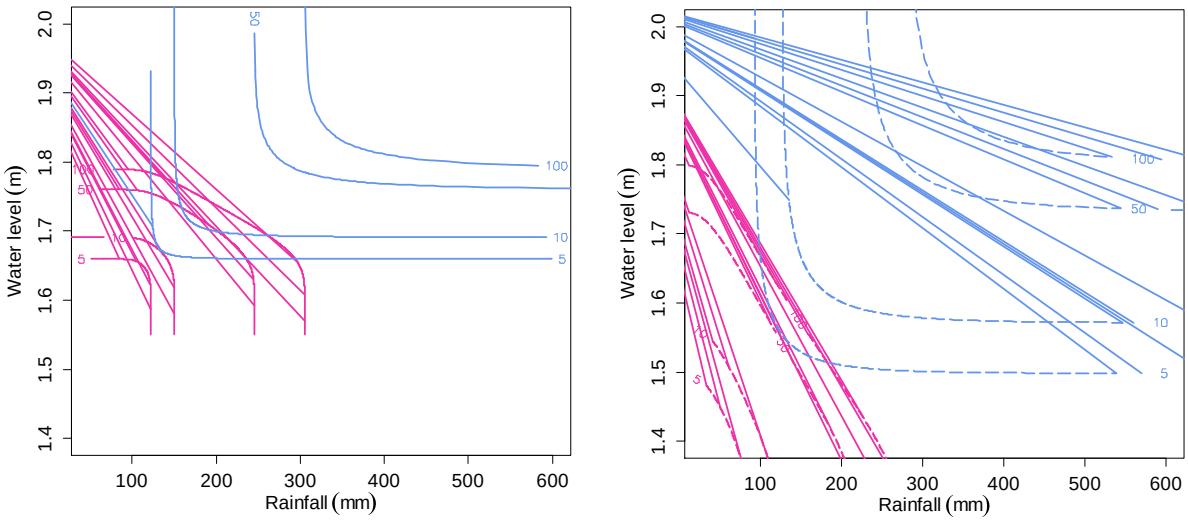


Figure 6. 3: The joint return periods AND (pink color) and OR (blue color) based on: (a) best model (solid lines) and (b) stationary model (dash lines) for Sample 1

Figure 6. 4 shows the comparison of the joint return period of rainfall and water level (OR) between various models for Sample 1. For the more frequent return period (< 10 years), the return levels of rainfall and water level derived from nonstationary marginal distribution in which the parameters are expressed as a function of covariates, are higher than those derived from other models. However, during the less frequent return period, particularly 50-year, values of rainfall and water level are not much different between models. These results indicate that more attention should be focused on nonstationary marginal distribution modelling than dependence structure modelling.

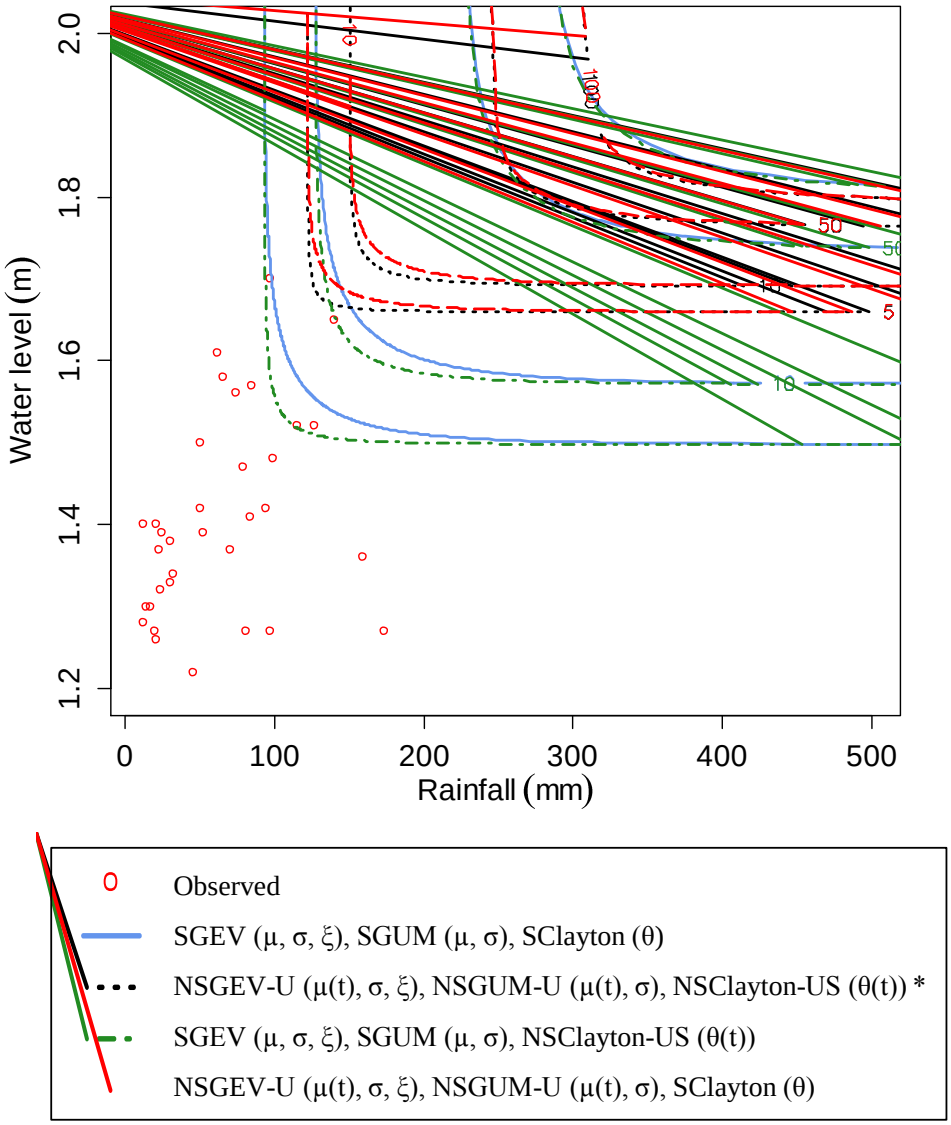


Figure 6. 4: The joint return periods (OR) of rainfall and water level from various models for Sample 1 (* denoted the best model for Sample 1)

The comparison of the joint return periods (OR) of rainfall and water level from the best models corresponding to three samples is also carried out (Figure 6. 5). It is observed that the estimated water level values for Sample 1 are much higher than those for the remaining samples. For the more detailed, the value of rainfall (mm) - water level (m) from the best model are approximately (800 mm -1.80 m) for Sample 1 for the 100-year return period. Meanwhile, those values for Sample 2 and 3 are (800 mm – 1.515 m) and (800 mm – 1.525 m) respectively. In addition, when comparing Sample 1 and Sample 3, a significant difference for the pairs of rainfall and water level occurred for a given return period can be found, with a higher value of rainfall corresponding to a smaller value of water level and vice versa.

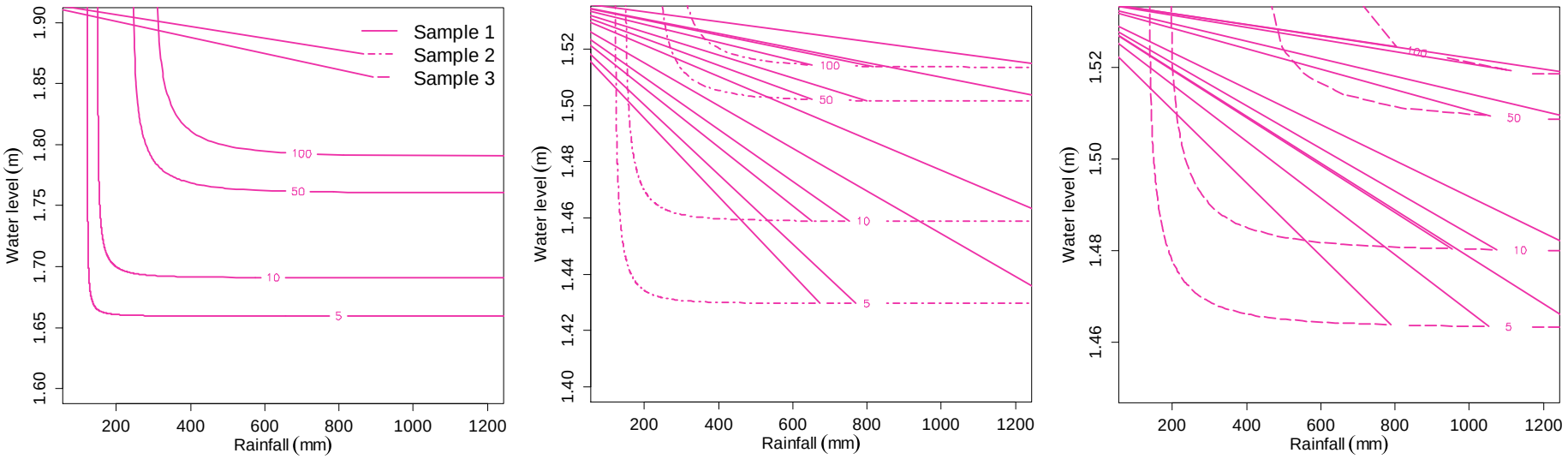


Figure 6. 5: Comparison of the joint return periods of rainfall and water level (OR) from the best models for (a) Sample 1, (b) Sample 2 and (c) Sample 3

6.5 Summary and conclusions

In this present study, a nonstationary bivariate approach is introduced and exemplarily applied to HCMC where there has been increasing vulnerability to floods from multivariate sources. Urbanization, temperature and sea level are used as local covariates for modelling nonstationary the marginal distribution (i.e. GEV, GUM, LN, WEI distributions) and the dependence structure (i.e. Frank, Clayton, Gaussian and Plackett copulas) for extreme rainfall and water level. Since many physical processes control the extreme event occurrence, incorporating all of them may increase the accuracy of modelling, but it may increase the bias of modelling. In this study, local physical processes have proved that they are mostly associated with extreme events in the study area. Besides, among all candidate distributions, GEV seems the most suitable marginal distribution for modelling the flood variables in HCMC by always presenting its lowest value of AICc. It is recommended that GEV can be used as a default distribution for modelling extreme rainfall and water level in HCMC.

The nonstationary copula-based bivariate frequency analysis is conducted based on three samples. The POT approach is commonly considered as a default approach in case of shorter observed data. However, the results of this study indicate that, when considering AM approach (i.e. maximum rainfall and water level), the joint return levels are estimated higher than using other approaches, which may be better and safer for flood design. Therefore, for the river basin with scarce data and affected by several sources of floods, it is necessary to consider both AM and POT approaches, which can provide more choices for stakeholders' decisions.

Our findings also indicate that modelling nonstationarity in the marginal distribution by choosing appropriate covariates, is more important than concentrating on modelling the dependence between rainfall and water level by using nonstationary copulas. These findings support the findings of (Bender et al., 2014) who also concluded that more attention should be paid for modelling the marginal distributions rather than for modelling the dependence.

The proposed method presented in this study provides an overall framework for nonstationary bivariate frequency analysis for HCMC. This method can be applied to other river basins as well as other fields related to water resources management to identify the

appropriate solutions and adaptation strategies under nonstationary behavior of extreme events.

Chapter 7

Summary and conclusions

7.1 Summary

This research shows that the time series of extreme hydrologic events (i.e. rainfall and water level) in HCMC has been proven to increase in frequency and magnitude in last few decades due to the influence of global climate change and physical processes. Besides, this thesis contributes towards modelling extreme events under nonstationary condition by using local physical process as covariates. In addition, the spatial variation of extreme precipitation is also developed from stationary and nonstationary flood frequency analysis. Further, the flood simulation model is developed based on. The high-resolution flood hazard maps, which are quantified by considering the flood depth and velocity in combination, are established using nonstationary frequency analysis and coupled 1D-2D hydrodynamic model with high-resolution topography data. Finally, the last part of this thesis contributes towards investigating the joint probability of flood correlated variables using nonstationary copula-

based bivariate frequency analysis. Furthermore, the performance for an asymptotic independent variable

also assessed. The following paragraphs give the summary and conclusions of this thesis:

- Extreme rainfall in HCMC has been proven to increase in frequency and magnitude in last few decades due to the influence of global climate change and physical processes. Although nonstationarity in extreme rainfall has been proved in many places of the world, research into nonstationarity feature in extreme rainfall in HCMC has not been paid attention thoroughly. Therefore, in Chapter 3, the spatial variation of extreme precipitation over HCMC is modelled under nonstationary condition. However, finding the most significant physical processes which have a close relationship with extreme rainfall is problematic since extreme events are controlled by many physical processes. The study results show that the nonstationary GEV model is found to be superior in capturing extreme precipitation events when compared to the stationary GEV model.. It is also found that the extreme rainfall estimates under the stationary condition are lower than those under the nonstationary condition in most stations.
- The best local covariates for modelling nonstationary extreme water level are studied in Chapter 4. In details, four local covariates, namely rainfall, sea level, urbanization growth and outflows from upstream reservoirs are used to develop ninety-two nonstationary extreme water level models. The stationary models are also developed for comparison purpose. The results indicate that the nonstationary approach using local covariates is suitable for modelling extreme water level in HCMC. Additionally, based on the best chosen statistical models, the significant influences of sea level and urbanization on nonstationarity in extreme water level are found at all surveyed stations. Moreover, it could be found that the extreme water level values derived from the stationary models are underestimated relative to the best nonstationary models for all stations.
- In Chapter 5, flood hazard maps are developed for HCMC under nonstationary conditions using extreme value analysis, a coupled model and high-resolution topographical data derived from LiDAR data. In details, climate indices (ENSO and PDO), global temperature, local temperature and global mean sea

level are used for frequency analysis to investigate the nonstationarity in the extreme rainfall and sea level. The covariates in the best statistical model are attributed as the most significant physical processes causing nonstationarity in the time series. The results of frequency analysis show that ENSO and PDO are present in the best nonstationary models, hence they can be considered as the main causes of nonstationary behavior in extreme rainfall in the study area. Another finding indicates that global sea level rise has a significant effect on nonstationarity in extreme sea level. . The results from the spatial variation of flood hazards indicate that locations along both riverbanks are expected to experience a significant increase in flooded area, especially where the river is surrounded by houses and buildings. It is also to be noted that the floodplain extent is larger when based on the assumption of nonstationarity.

- In the last chapter, urbanization, temperature and sea level are used as local covariates for modelling nonstationarity in the marginal distribution (i.e. GEV, GUM, LN and WEI distributions) and the dependence structure (i.e. Frank, Clayton, Gaussian and Plackett copulas) for extreme rainfall and water level. The nonstationary copula-based bivariate frequency analysis is conducted based on three samples, which may provide more options in choosing appropriate data for analysis. The findings of the study indicate that the GEV distribution is suggested as the most appropriate marginal distribution for modelling the flood variables. The nonstationary Clayton copula, stationary Plackett copula and stationary Gaussian copula are suitable to estimate the joint return period of rainfall and water level for three considered samples. Besides, the results indicate that urbanization is a significant covariate causing nonstationarity. The joint return periods of rainfall and water level obtained through the optimal copula and marginal distribution show the significant differences between the samples. It means that a reciprocal situation can be found, when a higher value of rainfall corresponds to a lower value of water level and vice versa.

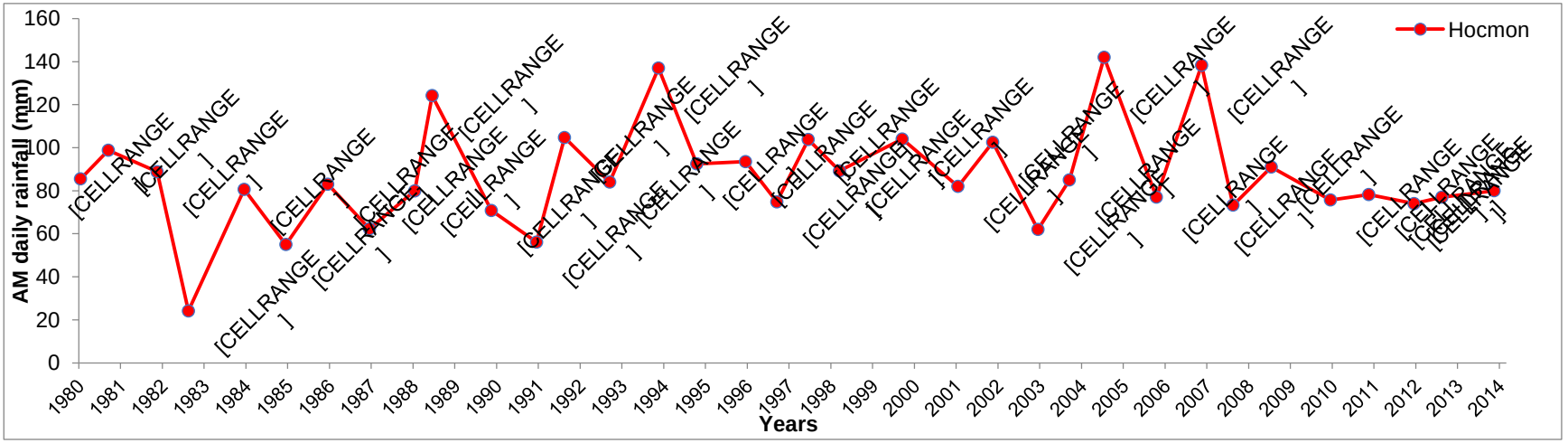
7.2 Scopes for future studies

The work presented in this thesis could be further extended if both relevant data were available and time is not a constraint. Hence, following would be possible future works:

- Assessing covariate and parameter uncertainty extreme events modelling
- Quantifying the uncertainty of the flood hazard maps
- Development of flood hazard management system, including a flood evacuation strategy.

Appendix A

a)


Hocmon
AM daily rainfall (mm)
Years
0
20
40
60
80
100
120
140
160
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014

b)

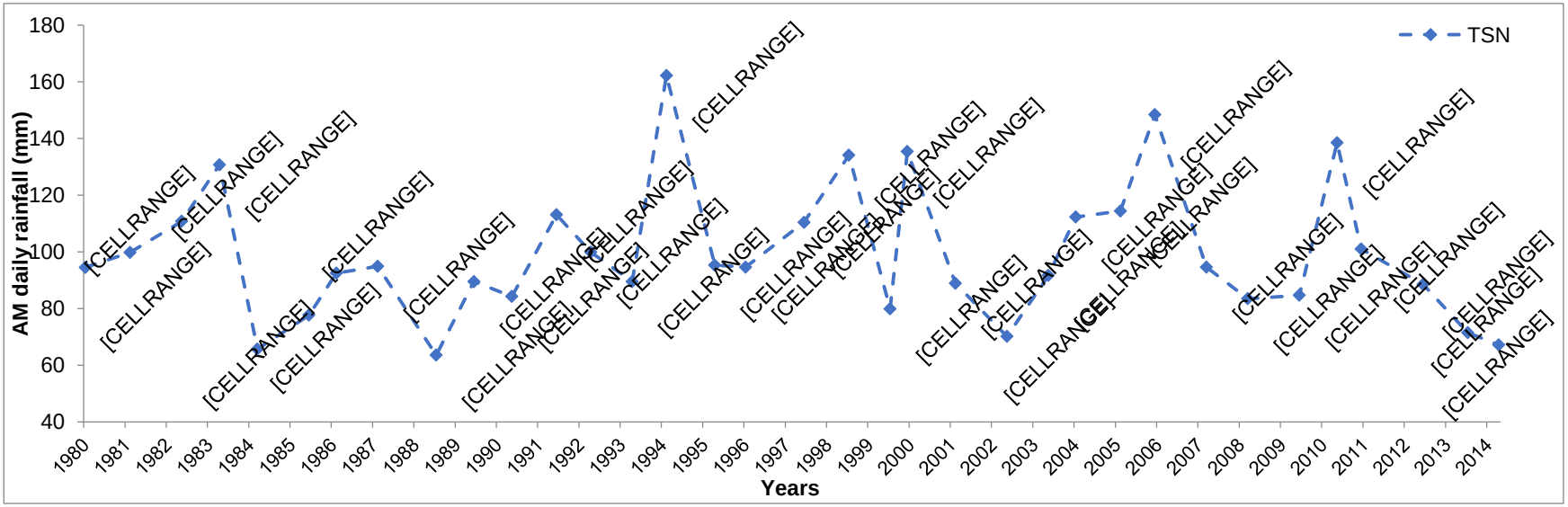
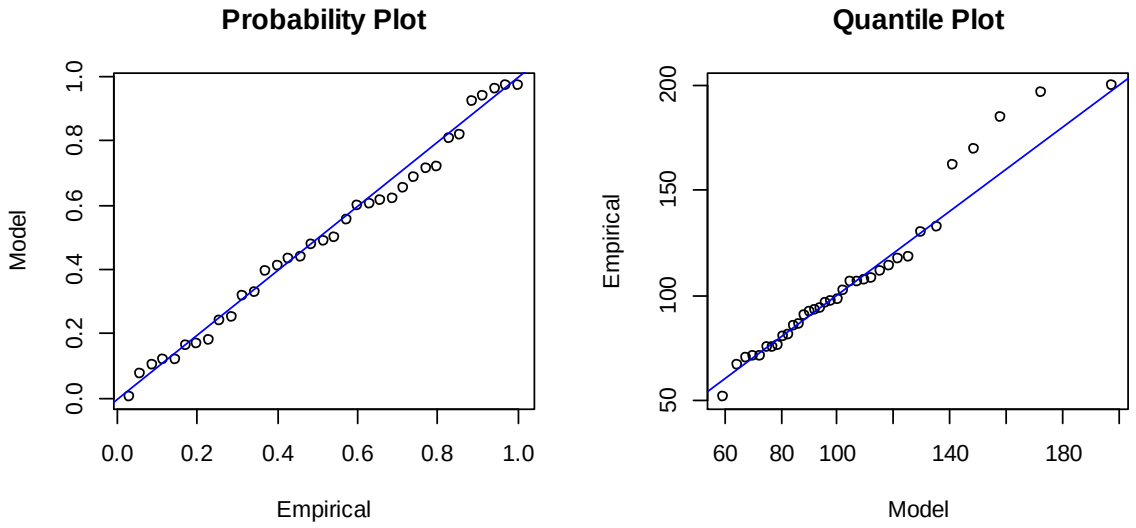
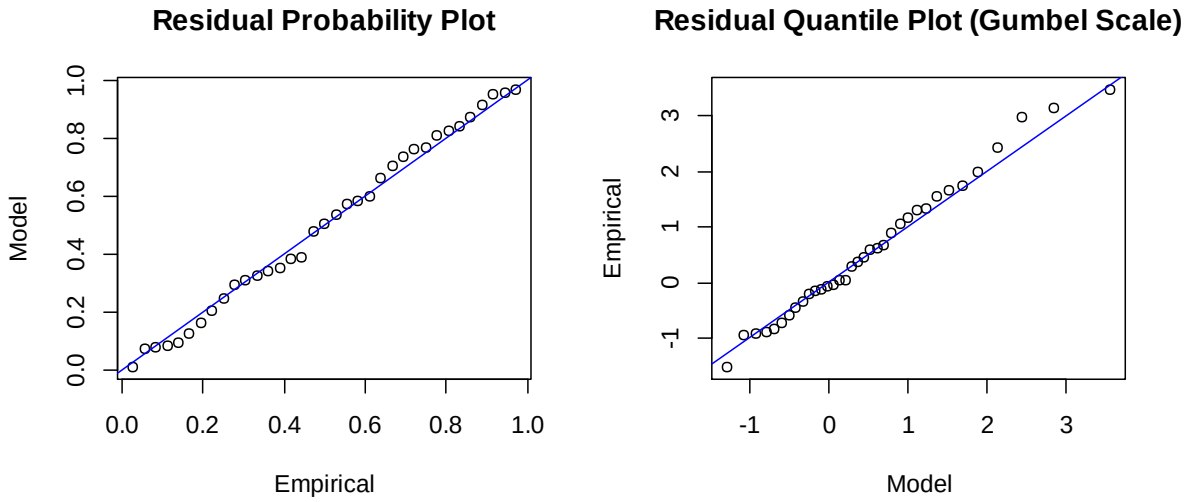


Figure A. 1: Dates of maximum daily rainfall of (a) Hocmon station and (b) TSN station for the period of 1980-2014

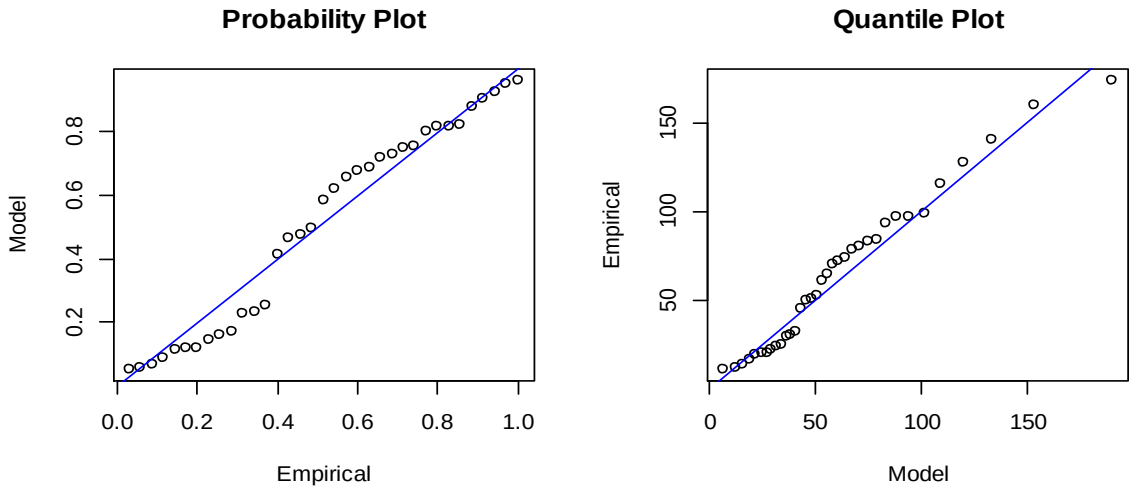


(a) Stationary model

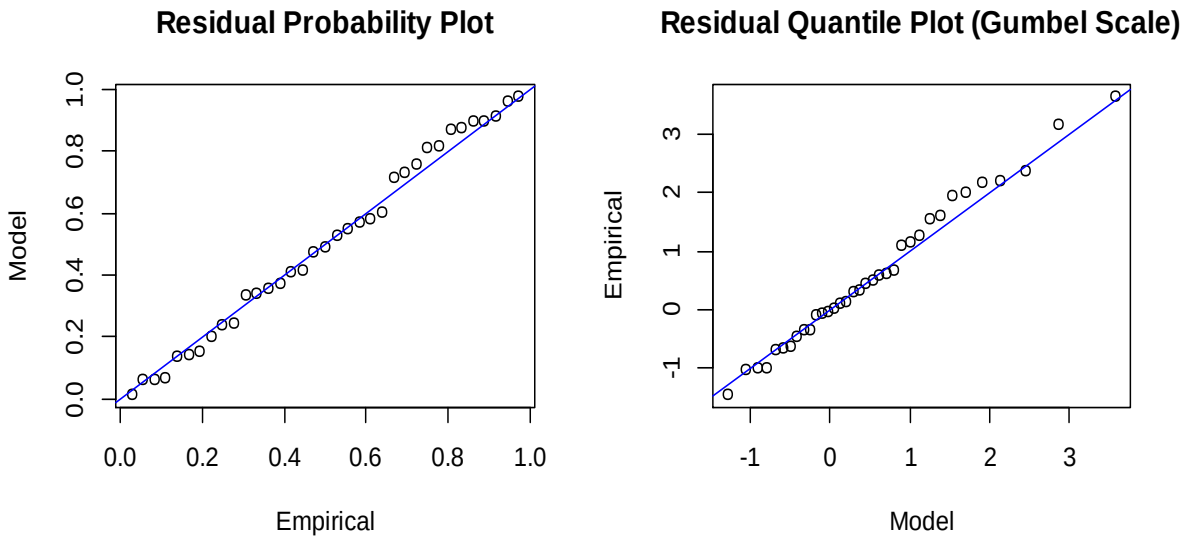


(b) Nonstationary model

Figure A. 2 P-P and Q-Q plots of Benluc station

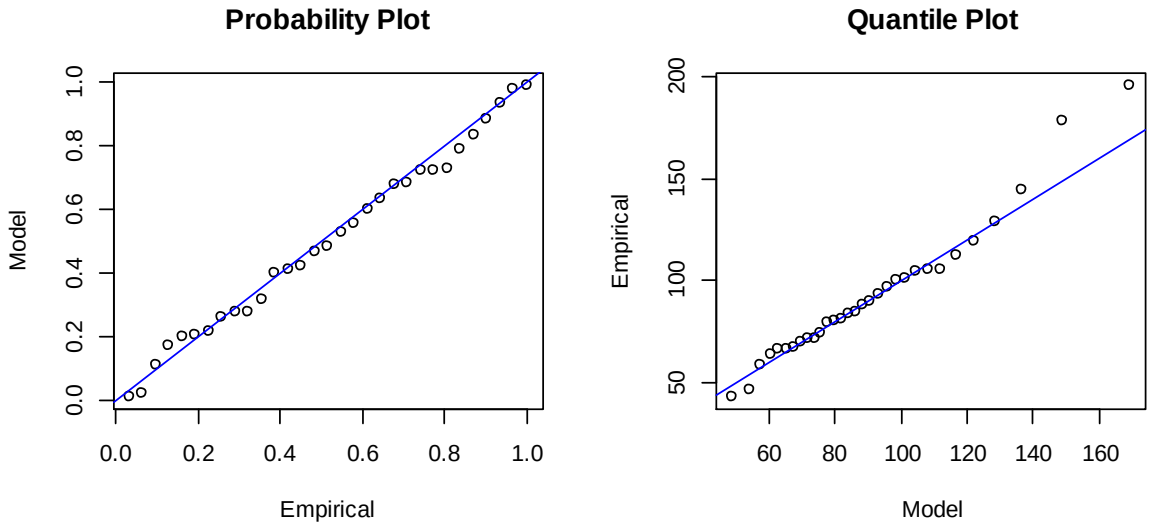


(a) Stationary model

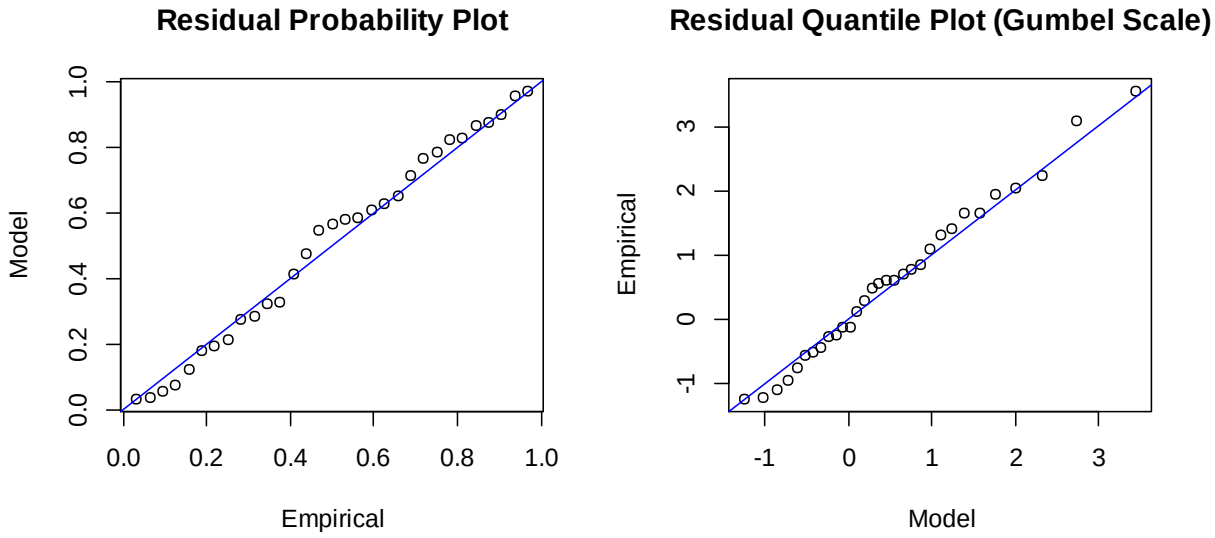


(b) Nonstationary model

Figure A. 3 P-P and Q-Q plots of Cangio station

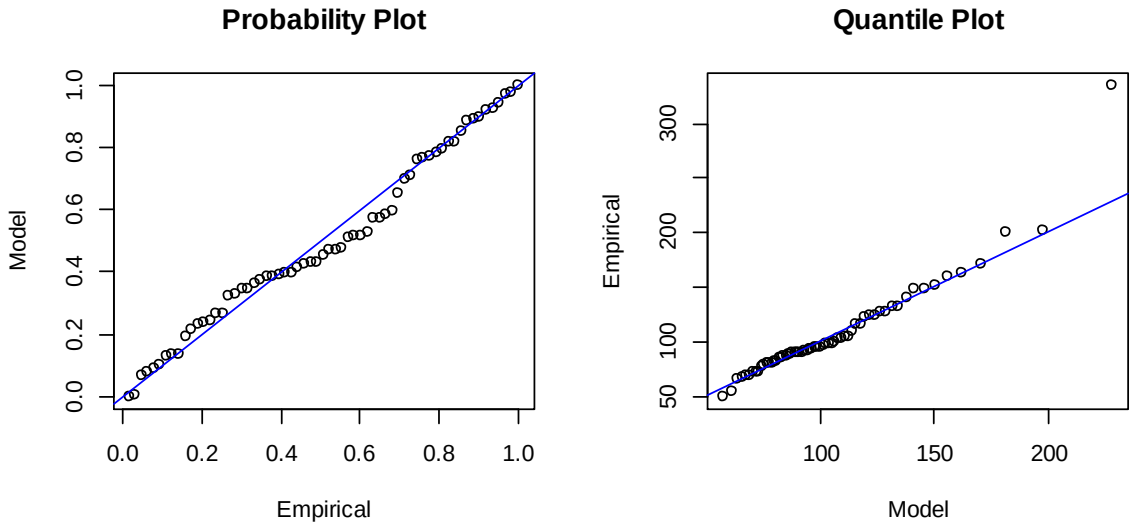


(a) Stationary model

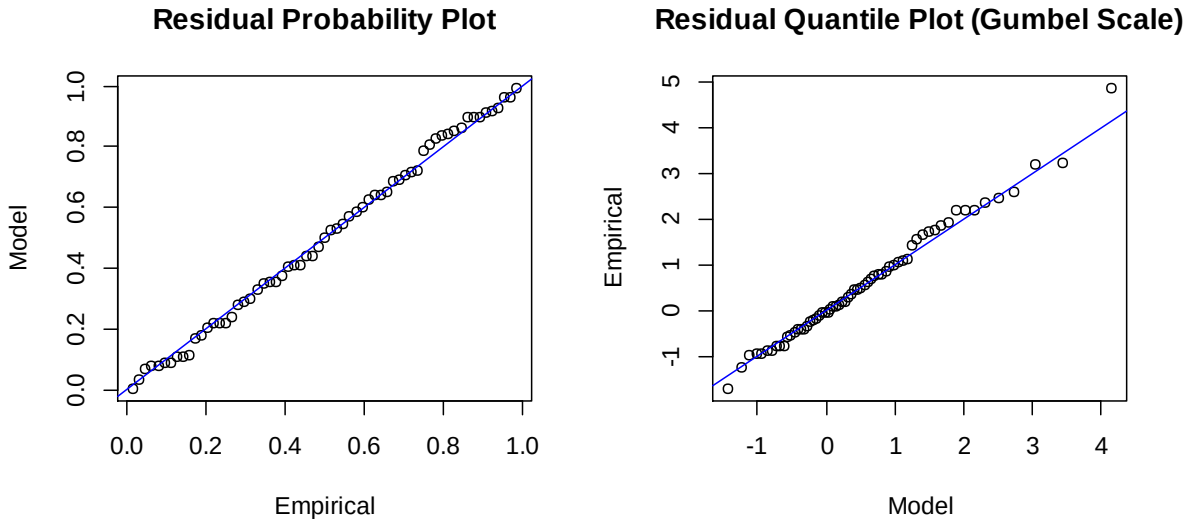


(b) Nonstationary model

Figure A. 4 P-P and Q-Q plots of Nhabe station



(a) Stationary model



(b) Nonstationary model

Figure A. 5 P-P and Q-Q plots of Xuanloc station

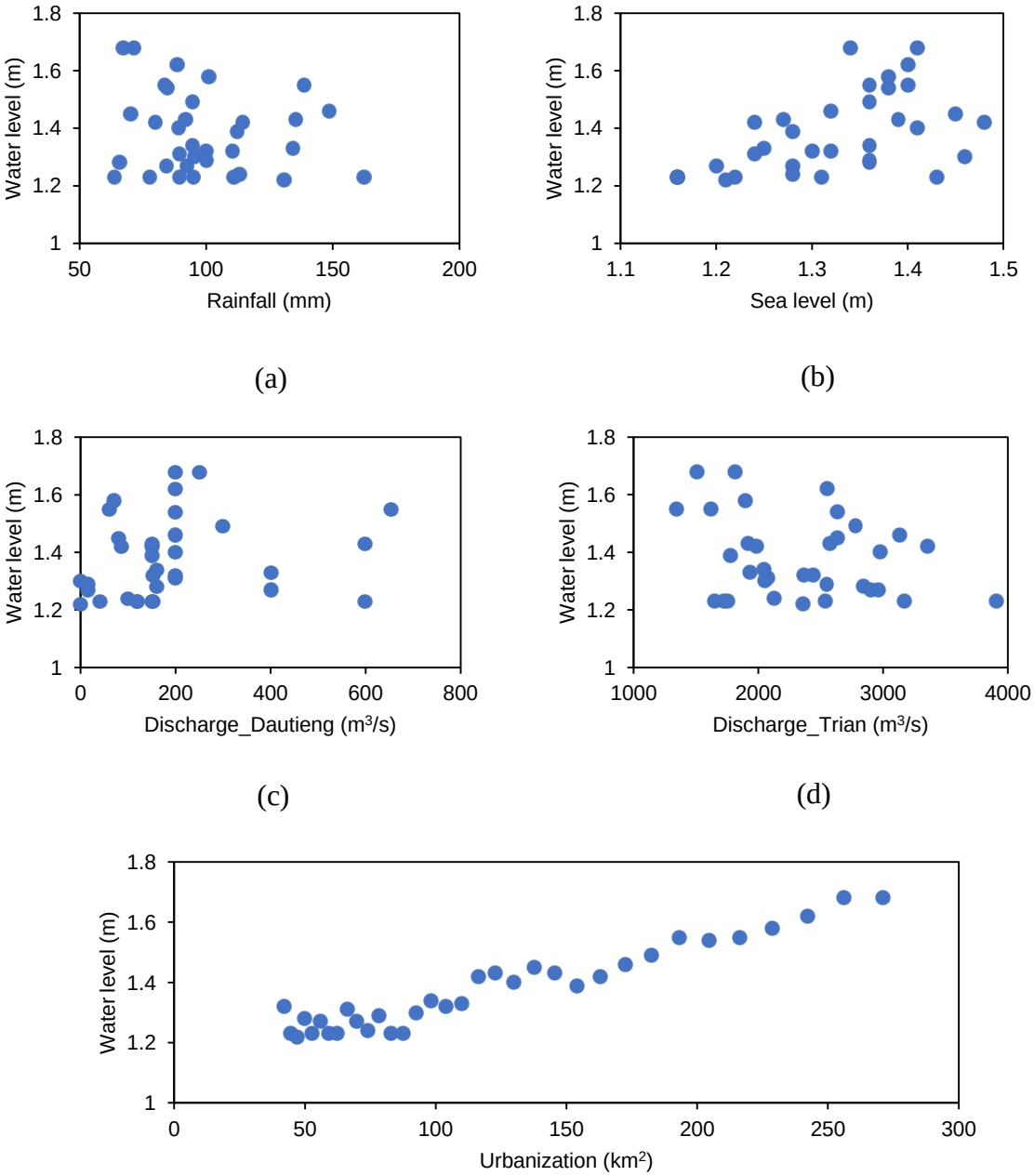


Figure A. 6 Scatter plots of observed de-clustered extreme water level against the covariate of (a) rainfall, (b) sea level, (c) outflows from Dautieng reservoir, (d) outflows from Trian reservoir and (e) urbanization for Phuan station

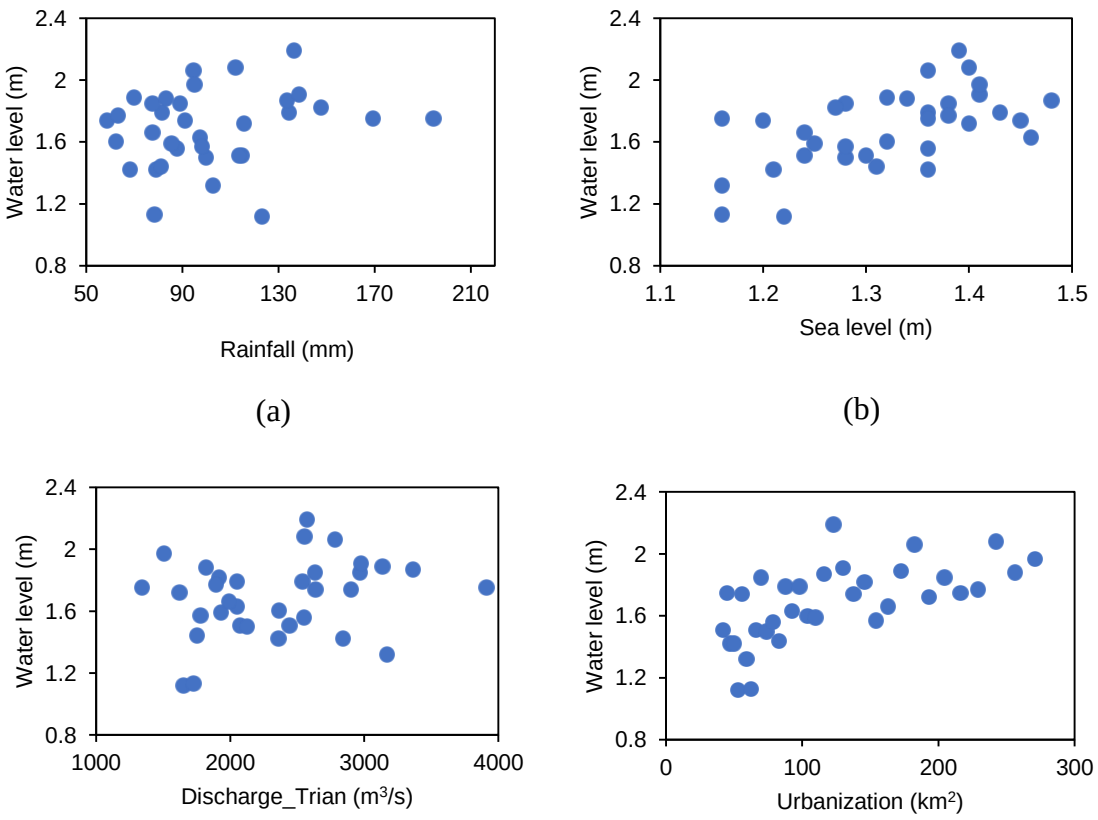


Figure A. 7 Scatter plots of observed de-clustered extreme water level against the covariate of (a) rainfall, (b) sea level, (c) outflows from Trian reservoir and (d) urbanization for Bienhoa station

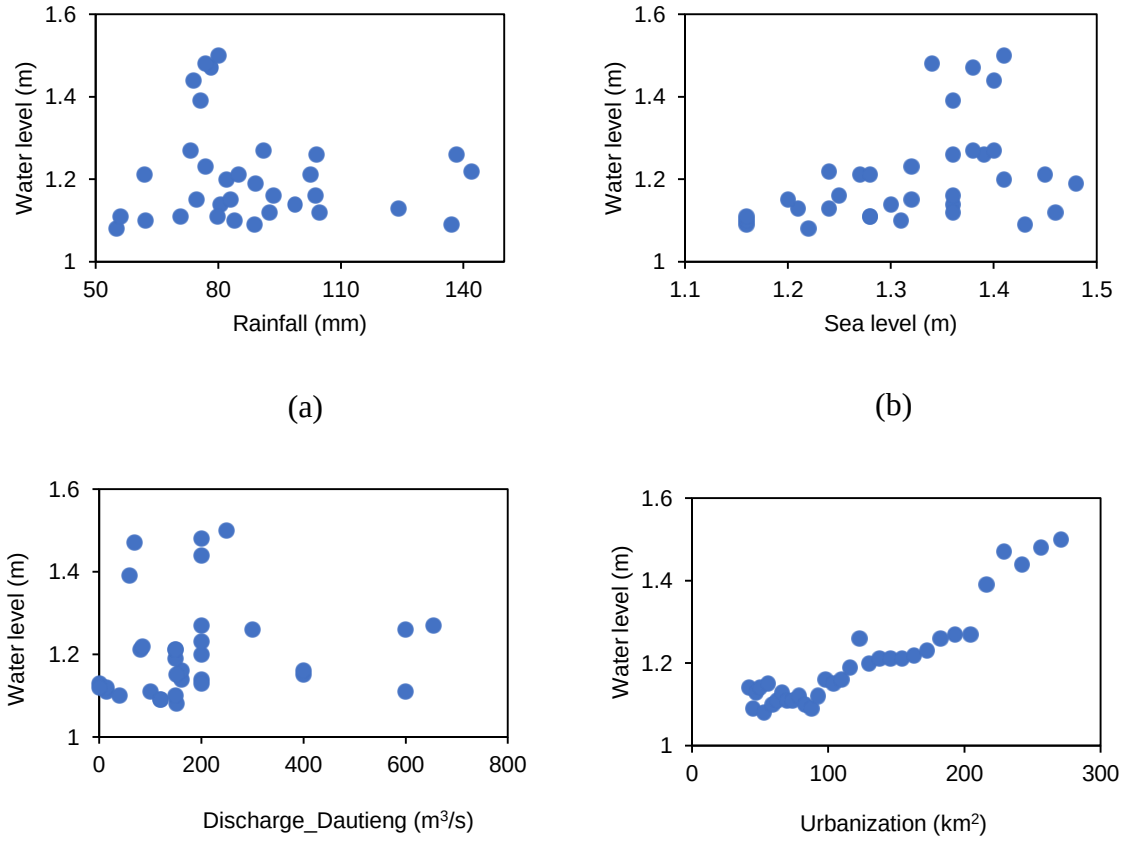


Figure A. 8 Scatter plots of observed de-clustered extreme water level against the covariate of (a) rainfall, (b) sea level, (c) outflows from Dautieng reservoir and (d) urbanization for TDM station

Table A. 1: Details of GEV models constructed for extreme rainfall analysis

Model	Description
GEV-0	$X \sim GEV\left(\mu_0, \sigma_0, \xi\right)$
GEV-1	$X \sim GEV\left(\left(\mu_0 + E\mu_1\right), \sigma_0, \xi\right)$
GEV-2	$X \sim GEV\left(\left(\mu_0 + P\mu_1\right), \sigma_0, \xi\right)$
GEV-3	$X \sim GEV\left(\left(\mu_0 + GT\mu_1\right), \sigma_0, \xi\right)$
GEV-4	$X \sim GEV\left(\left(\mu_0 + LT\mu_1\right), \sigma_0, \xi\right)$
GEV-5	$X \sim GEV\left(\left(\mu_0 + E\mu_1 + P\mu_2\right), \sigma_0, \xi\right)$
GEV-6	$X \sim GEV\left(\left(\mu_0 + E\mu_1 + GT\mu_2\right), \sigma_0, \xi\right)$
GEV-7	$X \sim GEV\left(\left(\mu_0 + E\mu_1 + LT\mu_2\right), \sigma_0, \xi\right)$
GEV-8	$X \sim GEV\left(\left(\mu_0 + P\mu_1 + GT\mu_2\right), \sigma_0, \xi\right)$
GEV-9	$X \sim GEV\left(\left(\mu_0 + P\mu_1 + LT\mu_2\right), \sigma_0, \xi\right)$
GEV-10	$X \sim GEV\left(\left(\mu_0 + GT\mu_1 + LT\mu_2\right), \sigma_0, \xi\right)$
GEV-11	$X \sim GEV\left(\left(\mu_0 + E\mu_1 + P\mu_2 + GT\mu_3\right), \sigma_0, \xi\right)$
GEV-12	$X \sim GEV\left(\left(\mu_0 + E\mu_1 + P\mu_2 + LT\mu_3\right), \sigma_0, \xi\right)$
GEV-13	$X \sim GEV\left(\left(\mu_0 + E\mu_1 + GT\mu_2 + LT\mu_3\right), \sigma_0, \xi\right)$
GEV-14	$X \sim GEV\left(\left(\mu_0 + P\mu_1 + GT\mu_2 + LT\mu_3\right), \sigma_0, \xi\right)$
GEV-15	$X \sim GEV\left(\left(\mu_0 + E\mu_1 + P\mu_2 + GT\mu_3 + LT\mu_4\right), \sigma_0, \xi\right)$
GEV-16	$X \sim GEV\left(\left(\mu_0 + E\mu_1\right), e^{(\sigma_0 + E\sigma_1)}, \xi\right)$
GEV-17	$X \sim GEV\left(\left(\mu_0 + P\mu_1\right), e^{(\sigma_0 + P\sigma_1)}, \xi\right)$
GEV-18	$X \sim GEV\left(\left(\mu_0 + GT\mu_1\right), e^{(\sigma_0 + GT\sigma_1)}, \xi\right)$
GEV-19	$X \sim GEV\left(\left(\mu_0 + LT\mu_1\right), e^{(\sigma_0 + LT\sigma_1)}, \xi\right)$
GEV-20	$X \sim GEV\left(\left(\mu_0 + E\mu_1 + P\mu_2\right), e^{(\sigma_0 + E\sigma_1 + P\sigma_2)}, \xi\right)$
GEV-21	$X \sim GEV\left(\left(\mu_0 + E\mu_1 + GT\mu_2\right), e^{(\sigma_0 + E\sigma_1 + GT\sigma_2)}, \xi\right)$
GEV-22	$X \sim GEV\left(\left(\mu_0 + E\mu_1 + LT\mu_2\right), e^{(\sigma_0 + E\sigma_1 + LT\sigma_2)}, \xi\right)$
GEV-23	$X \sim GEV\left(\left(\mu_0 + P\mu_1 + GT\mu_2\right), e^{(\sigma_0 + P\sigma_1 + GT\sigma_2)}, \xi\right)$
GEV-24	$X \sim GEV\left(\left(\mu_0 + P\mu_1 + LT\mu_2\right), e^{(\sigma_0 + P\sigma_1 + LT\sigma_2)}, \xi\right)$
GEV-25	$X \sim GEV\left(\left(\mu_0 + GT\mu_1 + LT\mu_2\right), e^{(\sigma_0 + GT\sigma_1 + LT\sigma_2)}, \xi\right)$

Table A. 2: Details of GEV models constructed for extreme sea level analysis

Model	Description
GEV-0	$X \sim GEV\left(\mu_0, \sigma_0, \xi\right)$
GEV-1	$X \sim GEV\left(\left(\mu_0 + E\mu_1\right), \sigma_0, \xi\right)$
GEV-2	$X \sim GEV\left(\left(\mu_0 + P\mu_1\right), \sigma_0, \xi\right)$
GEV-3	$X \sim GEV\left(\left(\mu_0 + GS\mu_1\right), \sigma_0, \xi\right)$
GEV-4	$X \sim GEV\left(\left(\mu_0 + E\mu_1 + P\mu_2\right), \sigma_0, \xi\right)$
GEV-5	$X \sim GEV\left(\left(\mu_0 + E\mu_1 + GS\mu_2\right), \sigma_0, \xi\right)$
GEV-6	$X \sim GEV\left(\left(\mu_0 + P\mu_1 + GS\mu_2\right), \sigma_0, \xi\right)$
GEV-7	$X \sim GEV\left(\left(\mu_0 + E\mu_1 + P\mu_2 + GS\mu_3\right), \sigma_0, \xi\right)$
GEV-8	$X \sim GEV\left(\left(\mu_0 + E\mu_1\right), e^{\left(\sigma_0 + E\sigma_1\right)}, \xi\right)$
GEV-9	$X \sim GEV\left(\left(\mu_0 + P\mu_1\right), e^{\left(\sigma_0 + P\sigma_1\right)}, \xi\right)$
GEV-10	$X \sim GEV\left(\left(\mu_0 + GS\mu_1\right), e^{\left(\sigma_0 + GS\sigma_1\right)}, \xi\right)$
GEV-11	$X \sim GEV\left(\left(\mu_0 + E\mu_1 + P\mu_2\right), e^{\left(\sigma_0 + E\sigma_1 + P\sigma_2\right)}, \xi\right)$
GEV-12	$X \sim GEV\left(\left(\mu_0 + E\mu_1 + GS\mu_2\right), e^{\left(\sigma_0 + E\sigma_1 + GS\sigma_2\right)}, \xi\right)$
GEV-13	$X \sim GEV\left(\left(\mu_0 + P\mu_1 + GS\mu_2\right), e^{\left(\sigma_0 + P\sigma_1 + GS\sigma_2\right)}, \xi\right)$

Table A. 3: List of marginal distributions and considered covariate(s) used for rainfall and water level frequency analysis

Rainfall			Water level		
	Considered covariate(s)			Considered covariate(s)	
Model	Urbanization	Temperature	Model	Urbanization	Sea level
SGEV			SGEV		
NSGEV-U	x		NSGEV-U	x	
NSGEV-T		x	NSGEV-S		x
NSGEV-UT	x	x	NSGEV-US	x	x
SGUM			SGUM		
NSGUM-U	x		NSGUM-U	x	
NSGUM-T		x	NSGUM-S		x
NSGUM-UT	x	x	NSGUM-US	x	x
SLN			SLN		
NSLN-U	x		NSLN-U	x	
NSLN-T		x	NSLN-S		x
NSLN-UT	x	x	NSLN-US	x	x
SWEI			SWEI		
NSWEI-U	x		NSWEI-U	x	
NSWEI-T		x	NSWEI-S		x
NSWEI-UT	x	x	NSWEI-US	x	x

Table A. 4: List of copulas and considered covariate(s) used for modelling the dependence between rainfall and water level

	Considered covariate(s)		
Copulas	Urbanization	Temperature	Sea level
SClayton			
NSClayton-U	x		
NSClayton-T		x	
NSClayton-S			x
NSClayton-UT	x	x	
NSClayton-US	x		x
NSClayton-TS		x	x
NSClayton-UTS	x	x	x
SFrank			
NSFrank-U	x		
NSFrank-T		x	
NSFrank-S			x
NSFrank-UT	x	x	
NSFrank-US	x		x
NSFrank-TS		x	x
NSFrank-UTS	x	x	x
SGaussian			
NSGaussian-U	x		
NSGaussian-T		x	
NSGaussian-S			x
NSGaussian-UT	x	x	
NSGaussian-US	x		x

	Considered covariate(s)		
Copulas	Urbanization	Temperature	Sea level
NSGaussian-TS		x	x
NSGaussian-UTS	x	x	x
SPlackett			
NSPlackett-U	x		
NSPlackett-T		x	
NSPlackett-S			x
NSPlackett-UT	x	x	
NSPlackett-US	x		x
NSPlackett-TS		x	x
NSPlackett-UTS	x	x	x

Table A. 5: AICc values for all marginal distributions

Rainfall				Water level			
Models	Sample 1	Sample 2	Sample 3	Models	Sample 1	Sample 2	Sample 3
SGEV	353.024	353.024	1850.206	SGEV	-43.745	8.483	-25.648
SGUM	351.451	351.451	1908.258	SGUM	-46.158	8.246	7.795
SLN	348.185	348.185	1892.357	SLN	-43.808	7.403	-3.542
SWEI	347.365	347.365	1972.834	SWEI	-36.084	7.246	-24.268
NSGEV-U	331.403	331.403	1849.094	NSGEV-U	-99.750	-2.604	-52.654
NSGUM-U	333.418	333.418	1905.505	NSGUM-U	-102.209	0.505	-11.888
NSLN-U	334.475	334.475	1889.877	NSLN-U	-98.824	-2.768	-28.158
NSWEI-U	342.621	342.621	1974.411	NSWEI-U	-88.706	-4.687	-53.807
NSGEV-T	346.805	350.819	1852.260	NSGEV-S	-49.777	-53.867	-373.241
NSGUM-T	344.984	349.675	1910.056	NSGUM-S	-52.270	-52.139	-334.005
NSLN-T	340.904	345.841	1894.370	NSLN-S	-51.905	-50.916	-326.345
NSWEI-T	346.281	348.153	1973.844	NSWEI-S	-45.579	-48.691	-331.725
NSGEV-UT	334.132	333.932	1849.744	NSGEV-US	-97.744	-70.731	-520.239
NSGUM-UT	335.559	335.983	1907.312	NSGUM-US	-100.325	-64.926	-489.369
NSLN-UT	336.954	336.698	1891.268	NSLN-US	-97.508	-60.225	-432.603
NSWEI-UT	344.134	345.176	1974.194	NSWEI-US	-86.196	-68.309	-452.994

Table A. 6: AICc values for all copulas

Copulas	Sample 1	Sample 2	Sample 3	Copulas	Sample 1	Sample 2	Sample 3
SClayton	-1.969	-3.114	1.926	NSClayton-UT	-2.537	0.335	2.884
SFrank	-4.860	-5.182	0.335	NSFrank-UT	-2.268	-0.735	4.065
SGaussian	-2.883	-1.758	0.068	NSGaussian-UT	-0.807	2.434	4.040
SPlackett	-5.214	-6.298	0.406	NSPlackett-UT	-1.907	-1.638	4.165
NSClayton-U	-4.817	-2.032	2.661	NSClayton-US	-7.767	0.371	2.269
NSFrank-U	-4.631	-3.050	2.061	NSFrank-US	-5.247	-1.015	3.378
NSGaussian-U	-3.029	0.048	2.066	NSGaussian-US	-4.743	2.461	3.302
NSPlackett-U	-4.200	-4.036	2.163	NSPlackett-US	-4.247	-2.634	3.477
NSClayton-T	-4.587	-0.983	3.923	NSClayton-TS	-2.195	0.896	5.415
NSFrank-T	-4.460	-3.125	2.199	NSFrank-TS	-3.828	-0.806	3.832
NSGaussian-T	-3.216	0.424	2.066	NSGaussian-TS	-2.601	2.593	3.303
NSPlackett-T	-3.764	-4.049	2.277	NSPlackett-TS	-2.307	-2.588	3.861
NSClayton-S	0.263	-0.883	3.411	NSClayton-UTS	-5.190	2.454	0.452
NSFrank-S	-2.619	-3.082	1.832	NSFrank-UTS	-3.797	1.590	5.445
NSGaussian-S	-0.982	0.465	1.249	NSGaussian-UTS	-4.293	2.691	5.370
NSPlackett-S	-3.041	-4.973	1.840	NSPlackett-UTS	-3.678	-0.242	5.542

References

Past and Future Changes in Extreme Sea Levels and Waves. *Understanding Sea- Level Rise and Variability.*

Abu-Aly, Pasternack, Wyrick, Barker, Massa & Johnson 2014. Effects of LiDAR-derived, spatially distributed vegetation roughness on two-dimensional hydraulics in a gravel-cobble river at flows of 0.2 to 20 times bankfull. *Geomorphology*, 206**,** 468-482.

ADB. 2010. *Ho Chi Minh City Adaptation to Climate Change: Summary Report* [Online]. Asian Development Bank. Available: https://www.adb.org/publications/ho-chi-minh-city-adaptation-climate-change-summary-report [Accessed 19 December 2016].

Adikari, Osti & Noro 2010. Flood- related disaster vulnerability: an impending crisis of megacities in Asia. *Journal of Flood risk management*, 3**,** 185-191.

AEMI 2014. Technical flood risk management guideline: Flood hazard. *Australian Emergency Handbook 7.* Australia Emergency Managerment Institute, Australia.

Agilan & Umamahesh 2015. Detection and attribution of non-stationarity in intensity and frequency of daily and 4-h extreme rainfall of Hyderabad, India. *Journal of Hydrology*, 530**,** 677-697.

Agilan & Umamahesh 2016a. Modelling nonlinear trend for developing non- stationary rainfall intensity–duration–frequency curve. *International Journal of Climatology*.

Agilan & Umamahesh 2016b. What are the best covariates for developing non-stationary rainfall intensity-duration-frequency Relationship? *Advances in Water Resources*.

Ahmed 2010. Numerical modeling of the Rideau valley watershed. *Natural hazards*, 55**,** 63-84.

Ahn & Palmer 2016. Use of a nonstationary copula to predict future bivariate low flow frequency in the Connecticut river basin. *Hydrological Processes*, 30**,** 3518-3532.

Alfieri, Salamon, Bianchi, Neal, Bates & Feyen 2014. Advances in pan- European flood hazard mapping. *Hydrological Processes,* 28**,** 4067-4077.

Ali 2018. *Flood inundation modeling and hazard mapping under uncertainty in the Sungai Johor basin, Malaysia,* CRC Press.

Alkema 2007. *Simulating floods: On the application of a 2D-hydraulic model for flood hazard and risk assessment,* Utrecht University.

Anh, Boxall, Saul & Willems. An evaluation of three lumped conceptual rainfall-runoff models at catchment scale. 2008. Sheffield, UK: The 13th World Water Congress.

Arns, Wahl, Dangendorf & Jensen 2015. The impact of sea level rise on storm surge water levels in the northern part of the German Bight. *Coastal Engineering,* 96**,** 118-131.

Arns, Wahl, Haigh & Jensen. Determining extreme water return levels at un-gauged sites: a case study of the Schleswig-Holsteins coastline and Islands in north-west Germany. Proceedings of the Coastal Dynamics conference 2013, 2013a.

Arns, Wahl, Haigh, Jensen & Pattiaratchi 2013b. Estimating extreme water level probabilities: a comparison of the direct methods and recommendations for best practise. *Coastal Engineering,* 81**,** 51-66.

Aronica, Franza, Bates & Neal 2012. Probabilistic evaluation of flood hazard in urban areas using Monte Carlo simulation. *Hydrological Processes,* 26**,** 3962-3972.

Badrzadeh, Sarukkalige & Jayawardena 2015. Hourly runoff forecasting for flood risk management: Application of various computational intelligence models. *Journal of Hydrology,* 529**,** 1633-1643.

Bates, Dawson, Hall, Horritt, Nicholls, Wicks & Hassan 2005. Simplified two-dimensional numerical modelling of coastal flooding and example applications. *Coastal Engineering,* 52**,** 793-810.

Beguería, Angulo- Martínez, Vicente- Serrano, López- Moreno & El- Kenawy 2011. Assessing trends in extreme precipitation events intensity and magnitude using

non- stationary peaks- over- threshold analysis: a case study in northeast Spain from 1930 to 2006. *International Journal of Climatology,* 31**,** 2102-2114.

Bender, Wahl & Jensen 2014. Multivariate design in the presence of non-stationarity. *Journal of hydrology,* 514**,** 123-130.

Berg, Moseley & Haerter 2013. Strong increase in convective precipitation in response to higher temperatures. *Nature Geoscience,* 6**,** 181-185.

Bergstrom 1976. Development and application of a conceptual runoff model for Scandinavian catchments.

Bezak, Brilly & Šraj 2014. Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis. *Hydrological Sciences Journal,* 59**,** 959-977.

Binh, Umamahesh, Rathnam & Son 2018. Modeling Nonstationary Extreme Water Levels Considering Local Covariates in Ho Chi Minh City, Vietnam. *Journal of Hydrologic Engineering,* 23**,** 04018042.

Birikundavyi, Labib, Trung & Rousselle 2002. Performance of neural networks in daily streamflow forecasting. *Journal of Hydrologic Engineering,* 7**,** 392-398.

Biswas & Jayawardena 2014. Water level prediction by artificial neural network in a flashy transboundary river of Bangladesh. *Global Nest Journal,* 16**,** 432-444.

Brirhet & Benaabidate 2016. Comparison Of Two Hydrological Models (Lumped And Distributed) Over A Pilot Area Of The Issen Watershed In The Souss Basin, Morocco. *European Scientific Journal, ESJ,* 12.

Bulteau, Idier, Lambert & Garcin 2015. How historical information can improve estimation and prediction of extreme coastal water levels: application to the Xynthia event at La Rochelle (France). *Natural Hazards and Earth System Sciences,* 15**,** 1135-1147.

Burnham & Anderson 2004. Multimodel inference understanding AIC and BIC in model selection. *Sociological methods & research,* 33**,** 261-304.

Cai & Rensch 2012. The 2011 southeast Queensland extreme summer rainfall: A confirmation of a negative Pacific Decadal Oscillation phase? *Geophysical Research Letters,* 39.

Chan & Zhou 2005. PDO, ENSO and the early summer monsoon rainfall over south China. *Geophysical Research Letters,* 32.

Chang & Chang 2006. Adaptive neuro-fuzzy inference system for prediction of water level in reservoir. *Advances in water resources,* 29**,** 1-10.

Chawla & Mujumdar 2018. Partitioning uncertainty in streamflow projections under nonstationary model conditions. *Advances in Water Resources,* 112**,** 266-282.

Chebana & Ouarda 2011. Multivariate quantiles in hydrological frequency analysis. *Environmetrics,* 22**,** 63-78.

Chen, Feng & Wu 2013. Roles of ENSO and PDO in the link of the East Asian winter monsoon to the following summer monsoon. *Journal of Climate,* 26**,** 622-635.

Cheng & AghaKouchak 2014. Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate. *Scientific reports,* 4.

Cheng, AghaKouchak, Gilleland & Katz 2014. Non-stationary extreme value analysis in a changing climate. *Climatic change,* 127**,** 353-369.

Chiew 2010. Lumped Conceptual Rainfall- Runoff Models and Simple Water Balance Methods: Overview and Applications in Ungauged and Data Limited Regions. *Geography Compass,* 4**,** 206-225.

Chu, McAleer & Chang 2013. Statistical Modelling of Extreme Rainfall in Taiwan.

Church & White 2011. Sea-level rise from the late 19th to the early 21st century. *Surveys in Geophysics,* 32**,** 585-602.

Cibin, Trybula, Chaubey, Brouder & Volenec 2016. Watershed- scale impacts of bioenergy crops on hydrology and water quality using improved SWAT model. *Gcb Bioenergy,* 8**,** 837-848.

Coles 2001. *An introduction to statistical modeling of extreme values,* Springer.

Coles, Heffernan & Tawn 1999. Dependence measures for extreme value analyses. *Extremes,* 2**,** 339-365.

Congedo 2016. Semi-Automatic Classification Plugin Documentation. *Release,* 4**,** 29.

Corbella & Stretch 2013. Simulating a multivariate sea storm using Archimedean copulas. *Coastal Engineering,* 76**,** 68-78.

Costabile & Macchione 2015. Enhancing river model set-up for 2-D dynamic flood modelling. *Environmental Modelling & Software,* 67**,** 89-107.

Crawford & Linsley 1966. Digital Simulation in Hydrology'Stanford Watershed Model 4.

Dang & Kumar 2017. Application of remote sensing and GIS-based hydrological modelling for flood risk analysis: a case study of District 8, Ho Chi Minh city, Vietnam. *Geomatics, Natural Hazards and Risk***,** 1-20.

Dasgupta, Laplante, Murray & Wheeler 2011. Exposure of developing countries to sea-level rise and storm surges. *Climatic Change,* 106**,** 567-579.

De Michele, Salvadori, Passoni & Vezzoli 2007. A multivariate model of sea storms using copulas. *Coastal Engineering,* 54**,** 734-751.

Deng, Wei, Xie, Ke, Wang, Zeng & Liu 2013. Variations in the Pacific Decadal Oscillation since 1853 in a coral record from the northern South China Sea. *Journal of Geophysical Research: Oceans,* 118**,** 2358-2366.

Devia, Ganasri & Dwarakish 2015. A review on hydrological models. *Aquatic Procedia,* 4**,** 1001-1007.

DHI 2003. MIKE11-A modelling system for Rivers and Channels. *User Guide.* Danish Hydraulic Institute, Denmark.

DHI 2007. MIKE FLOOD. *User Manual.* Danish Hydraulic Institute, Denmark.

Di Baldassarre, Schumann, Bates, Freer & Beven 2010. Flood-plain mapping: a critical discussion of deterministic and probabilistic approaches. *Hydrological Sciences Journal–Journal des Sciences Hydrologiques,* 55**,** 364-376.

Doocy, Daniels, Murray & Kirsch 2013. The human impact of floods: a historical review of events 1980-2009 and systematic literature review. *PLoS currents,* 5.

Douglas & Fairbank 2010. Is precipitation in northern New England becoming more extreme? Statistical analysis of extreme rainfall in Massachusetts, New Hampshire, and Maine and updated estimates of the 100-year storm. *Journal of Hydrologic Engineering,* 16**,** 203-217.

Du, Xiong, Xu, Gippel, Guo & Liu 2015. Return period and risk analysis of nonstationary low-flow series under climate change. *Journal of Hydrology,* 527**,** 234-250.

Dutta, Alam, Umeda, Hayashi & Hironaka 2007. A two- dimensional hydrodynamic model for flood inundation simulation: a case study in the lower Mekong river basin. *Hydrological Processes: An International Journal,* 21**,** 1223-1237.

Ervine & MacLeod 1999. Modelling a river channel with distant floodbanks. *Proceedings of the Institution of Civil Engineers-Water Maritime and Energy,* 136**,** 21-33.

Feng, Nadarajah & Hu 2007. Modeling annual extreme precipitation in China using the generalized extreme value distribution. *Journal of the Meteorological Society of Japan,* 85**,** 599-613.

Feng & Tsimplis 2014. Sea level extremes at the coasts of China. *Journal of Geophysical Research: Oceans,* 119**,** 1593-1608.

FLOOD 2011. 1D-2D Modelling—User Manual. *DHI Water & Environment*.

Fread & Lewis. FLDWAV: A generalized flood routing model. Proceedings of national conference on hydraulic engineering, ASCE, 1988. 668-673.

Garrote, Alvarenga & Díez-Herrero 2016. Quantification of flash flood economic risk using ultra-detailed stage–damage functions and 2-D hydraulic models. *Journal of Hydrology,* 541**,** 611-625.

Genest & Boies 2003. Detecting dependence with Kendall plots. *The American Statistician,* 57**,** 275-284.

Gilroy & McCuen 2012. A nonstationary flood frequency analysis method to adjust for future climate change and urbanization. *Journal of Hydrology,* 414**,** 40-48.

Gobin, Nguyen, Pham & Pham 2015. Heavy rainfall patterns in Vietnam and their relation with ENSO cycles. *International Journal of Climatology*.

Gobin, Nguyen, Pham & Pham 2016. Heavy rainfall patterns in Vietnam and their relation with ENSO cycles. *International Journal of Climatology,* 36**,** 1686-1699.

Golmohammadi, Prasher, Madani & Rudra 2014. Evaluating three hydrological distributed watershed models: MIKE-SHE, APEX, SWAT. *Hydrology,* 1**,** 20-39.

Gravert & Wiechmann 2016. Climate change adaptation governance in the Ho Chi Minh City region. *Sustainable Ho Chi Minh City: Climate Policies for Emerging Mega Cities.* Springer.

Groisman, Knight, Easterling, Karl, Hegerl & Razuvaev 2005. Trends in intense precipitation in the climate record. *Journal of climate,* 18**,** 1326-1350.

Guhathakurta, Sreejith & Menon 2011. Impact of climate change on extreme rainfall events and flood risk in India. *Journal of earth system science,* 120**,** 359.

Hallegatte, Green, Nicholls & Corfee-Morlot 2013. Future flood losses in major coastal cities. *Nature climate change,* 3**,** 802-806.

Hanson, Nicholls, Ranger, Hallegatte, Corfee-Morlot, Herweijer & Chateau 2011. A global ranking of port cities with high exposure to climate extremes. *Climatic change,* 104**,** 89-111.

Hjelmfelt Jr 1991. Investigation of curve number procedure. *Journal of Hydraulic Engineering,* 117**,** 725-737.

Hunt 2005. Inland and coastal flooding: developments in prediction and prevention. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences,* 363**,** 1475-1491.

Hurvich & Tsai 1995. Model selection for extended quasi-likelihood models in small samples. *Biometrics***,** 1077-1084.

IPCC 2012. *Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change*, Cambridge University Press.

Ishak, Rahman, Westra, Sharma & Kuczera 2013. Evaluating the non-stationarity of Australian annual maximum flood. *Journal of Hydrology,* 494**,** 134-145.

Jakeman, Littlewood & Whitehead 1990. Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. *Journal of hydrology,* 117**,** 275-300.

Jeong, Kannan, Arnold, Glick, Gosselink & Srinivasan 2010. Development and integration of sub-hourly rainfall–runoff modeling capability within a watershed model. *Water resources management,* 24**,** 4505-4527.

Jiang, Xiong, Xu & Guo 2015. Bivariate frequency analysis of nonstationary low- flow series based on the time- varying copula. *Hydrological Processes,* 29**,** 1521-1534.

Jongman, Ward & Aerts 2012. Global exposure to river and coastal flooding: Long term trends and changes. *Global Environmental Change,* 22**,** 823-835.

Kabiri, Bai & Chan 2015. Assessment of hydrologic impacts of climate change on the runoff trend in Klang Watershed, Malaysia. *Environmental Earth Sciences,* 73**,** 27-37.

Kachroo 1992. River flow forecasting. Part 5. Applications of a conceptual model. *Journal of Hydrology,* 133**,** 141-178.

Kalyanapu, Judi, McPherson & Burian 2012. M onte C arlo- based flood modelling framework for estimating probability weighted flood risk. *Journal of Flood Risk Management,* 5**,** 37-48.

Karamouz, Ahmadvand & Zahmatkesh 2017. Distributed Hydrologic Modeling of Coastal Flood Inundation and Damage: Nonstationary Approach. *Journal of Irrigation and Drainage Engineering,* 143**,** 04017019.

Karamouz, Zahmatkesh, Goharian & Nazif 2014. Combined impact of inland and coastal floods: mapping knowledge base for development of planning strategies. *Journal of Water Resources Planning and Management,* 141**,** 04014098.

Karim, Dutta, Marvanek, Petheram, Ticehurst, Lerat, Kim & Yang 2015. Assessing the impacts of climate change and dams on floodplain inundation and wetland connectivity in the wet–dry tropics of northern Australia. *Journal of Hydrology,* 522**,** 80-94.

Karlsson, Sonnenborg, Refsgaard, Trolle, Børgesen, Olesen, Jeppesen & Jensen 2016. Combined effects of climate models, hydrological model structures and land use scenarios on hydrological impacts of climate change. *Journal of Hydrology,* 535**,** 301-317.

Katz 2013. Statistical methods for nonstationary extremes. *Extremes in a Changing Climate.* Springer.

Katz, Parlange & Naveau 2002. Statistics of extremes in hydrology. *Advances in water resources,* 25**,** 1287-1304.

Kendall 1962. Rank correlation methods. *In:* 3RD (ed.). New York: Hafner Publishing Company.

Kenyon & Hegerl 2010. Influence of modes of climate variability on global precipitation extremes. *Journal of Climate,* 23**,** 6248-6262.

Khaliq, Ouarda, Ondo, Gachon & Bobée 2006. Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review. *Journal of hydrology,* 329**,** 534-552.

Kioutsioukis, Melas & Zerefos 2010. Statistical assessment of changes in climate extremes over Greece (1955–2002). *International Journal of Climatology,* 30**,** 1723-1737.

Komi, Neal, Trigg & Diekkrüger 2017. Modelling of flood hazard extent in data sparse areas: a case study of the Oti River basin, West Africa. *Journal of Hydrology: Regional Studies,* 10**,** 122-132.

Kreibich, Piroth, Seifert, Maiwald, Kunert, Schwarz, Merz & Thieken 2009. Is flow velocity a significant parameter in flood damage modelling? *Natural Hazards and Earth System Sciences,* 9**,** 1679.

Kunkel, Karl, Easterling, Redmond, Young, Yin & Hennon 2013. Probable maximum precipitation and climate change. *Geophysical Research Letters,* 40**,** 1402-1408.

Labaeye, Brugmann, Van Phuoc, Thanh, Thao, Tuan, Storch & Schinkel 2012. Reality Check: Ho Chi Minh City, Vietnam. *Resilient Cities 2.* Springer.

Lang, Ouarda & Bobée 1999. Towards operational guidelines for over-threshold modeling. *Journal of hydrology,* 225**,** 103-117.

Larsen, Refsgaard, Drews, Butts, Jensen, Christensen & Christensen 2014. Results from a full coupling of the HIRHAM regional climate model and the MIKE SHE hydrological model for a Danish catchment. *Hydrol. Earth Syst. Sci,* 18**,** 4733-4749.

Lasage, Veldkamp, De Moel, Van, Phi, Vellinga & Aerts 2014. Assessment of the effectiveness of flood adaptation strategies for HCMC. *Natural Hazards and Earth System Sciences,* 14**,** 1441-1457.

Le Vo 2007. Urbanization and water management in Ho Chi Minh City, Vietnam-issues, challenges and perspectives. *GeoJournal,* 70**,** 75-89.

Leandro, Chen, Djordjević & Savić 2009. Comparison of 1D/1D and 1D/2D coupled (sewer/surface) hydraulic models for urban flood simulation. *Journal of hydraulic engineering,* 135**,** 495-504.

Letetrel, Marcos, Míguez & Woppelmann 2010. Sea level extremes in Marseille (NW Mediterranean) during 1885–2008. *Continental shelf research,* 30**,** 1267-1274.

Li, Liu & Chen 2015. Evaluation of nonstationarity in annual maximum flood series and the associations with large-scale climate patterns and human activities. *Water Resources Management,* 29**,** 1653-1668.

Lin, Chen, Yao, Chen, Liu, Gao & James 2015. Analyses of landuse change impacts on catchment runoff using different time indicators based on SWAT model. *Ecological Indicators,* 58**,** 55-63.

Liu, Qin, Zhang & Li 2015. A coupled 1D–2D hydrodynamic model for flood simulation in flood detention basin. *Natural Hazards,* 75**,** 1303-1325.

López & Francés 2013. Non-stationary flood frequency analysis in continental Spanish rivers, using climate and reservoir indices as external covariates. *Hydrology and Earth System Sciences,* 17**,** 3189.

Lowe, Woodworth, Knutson, McDonald, McInnes, Woth, Von Storch, Wolf, Swail & Bernier 2010. Past and future changes in extreme sea levels and waves. *In:* CHURCH, J. A., WOODWORTH, P. L., AARUP, T. & WILSON, W. S. (eds.) *Understanding Sea-Level Rise and Variability.* UK: Wiley-Blackwell.

Machado, Botero, López, Francés, Díez-Herrero & Benito 2015. Flood frequency analysis of historical flood data under stationary and non-stationary modelling. *Hydrology and Earth System Sciences,* 19**,** 2561.

Madsen 2000. Automatic calibration of a conceptual rainfall–runoff model using multiple objectives. *Journal of hydrology,* 235**,** 276-288.

Mann 1945. Nonparametric Tests Against Trend. *Econometrica,* 13**,** 245-259.

Masina & Lamberti 2013. A nonstationary analysis for the Northern Adriatic extreme sea levels. *Journal of Geophysical Research: Oceans,* 118**,** 3999-4016.

Masina, Lamberti & Archetti 2015. Coastal flooding: A copula based approach for estimating the joint probability of water levels and waves. *Coastal Engineering,* 97**,** 37-52.

Masood & Takeuchi 2012. Assessment of flood hazard, vulnerability and risk of mid-eastern Dhaka using DEM and 1D hydrodynamic model. *Natural hazards,* 61**,** 757-770.

Mazzoleni, Bacchi, Barontini, Di Baldassarre, Pilotti & Ranzi 2013. Flooding hazard mapping in floodplain areas affected by piping breaches in the Po River, Italy. *Journal of Hydrologic Engineering,* 19**,** 717-731.

Méndez, Menéndez, Luceño & Losada 2007. Analyzing monthly extreme sea levels with a time-dependent GEV model. *Journal of Atmospheric and Oceanic Technology,* 24**,** 894-911.

Menéndez & Woodworth 2010. Changes in extreme high water levels based on a quasi- global tide- gauge data set. *Journal of Geophysical Research: Oceans,* 115.

Merz, Parajka & Blöschl 2011. Time stability of catchment model parameters: Implications for climate impact analyses. *Water Resources Research,* 47.

Milly, Betancourt, Falkenmark, Hirsch, Kundzewicz, Lettenmaier & Stouffer 2008. Stationarity is dead: Whither water management? *Science,* 319**,** 573-574.

Min, Zhang, Zwiers & Hegerl 2011. Human contribution to more-intense precipitation extremes. *Nature,* 470**,** 378-381.

Minh, Minh, Iwata, Kajiwara, Kunisue, Takahashi, Viet, Tuyen & Tanabe 2007. Persistent organic pollutants in sediments from Sai Gon–Dong Nai river Basin, Vietnam: levels and temporal trends. *Archives of Environmental Contamination and Toxicology,* 52**,** 458-465.

Mockler, O'Loughlin & Bruen 2016. Understanding hydrological flow paths in conceptual catchment models using uncertainty and sensitivity analysis. *Computers & geosciences,* 90**,** 66-77.

Mondal & Mujumdar 2015. Modeling non-stationarity in intensity, duration and frequency of extreme rainfall over India. *Journal of Hydrology,* 521**,** 217-231.

Moore. 2011. *Development of a high-resolution 1D/2D coupled flood simulation of Charles City, Iowa.* Master Thesis, University of Iowa.

Moriasi, Arnold, Van Liew, Bingner, Harmel & Veith 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE,* 50**,** 885-900.

Mu & Zhang 2007. Real-time flood forecasting method with 1-D unsteady flow model. *Journal of Hydrodynamics,* 19**,** 150-154.

Mudersbach & Jensen 2010. Nonstationary extreme value analysis of annual maximum water levels for designing coastal structures on the German North Sea coastline. *Journal of Flood Risk Management,* 3**,** 52-62.

Nadarajah 2005. Extremes of daily rainfall in West Central Florida. *Climatic change,* 69**,** 325-342.

Nash & Sutcliffe 1970. River flow forecasting through conceptual models part I - A discussion of principles. *Journal of hydrology,* 10**,** 282-290.

Nguyen, Katzfey & McGregor 2014. Downscaling over Vietnam using the stretched-grid CCAM: verification of the mean and interannual variability of rainfall. *Climate dynamics,* 43**,** 861-879.

Nguyen, Thorstensen, Sorooshian, Hsu, AghaKouchak, Sanders, Koren, Cui & Smith 2016. A high resolution coupled hydrologic–hydraulic model (HiResFlood-UCI) for flash flood modeling. *Journal of Hydrology,* 541**,** 401-420.

Nicholls & Cazenave 2010. Sea-level rise and its impact on coastal zones. *science,* 328**,** 1517-1520.

Nicholls, Hanson, Herweijer, Patmore, Hallegatte, Corfee-Morlot, Château & Muir-Wood 2008. *Ranking port cities with high exposure and vulnerability to climate extremes,* Paris, France.

Obeysekera & Park 2012. Scenario-based projection of extreme sea levels. *Journal of Coastal Research,* 29**,** 1-7.

Onyutha 2019. Hydrological Model Supported by a Step-Wise Calibration against Sub-Flows and Validation of Extreme Flow Events. *Water,* 11**,** 244.

Panagoulia, Economou & Caroni 2014. Stationary and nonstationary generalized extreme value modelling of extreme precipitation over a mountainous area under climate change. *Environmetrics,* 25**,** 29-43.

Papaioannou, Loukas, Vasiliades & Aronica 2016. Flood inundation mapping sensitivity to riverine spatial resolution and modelling approach. *Natural Hazards,* 83**,** 117-132.

Park, Kang, Lee & Kim 2011. Changes in the extreme daily rainfall in South Korea. *International Journal of Climatology,* 31**,** 2290-2299.

Paz, Bravo, Allasia, Collischonn & Tucci 2009. Large-scale hydrodynamic modeling of a complex river network and floodplains. *Journal of hydrologic engineering,* 15**,** 152-165.

Perrin, Michel & Andréassian 2003. Improvement of a parsimonious model for streamflow simulation. *Journal of hydrology,* 279**,** 275-289.

Petheram, Rustomji, Chiew & Vleeshouwer 2012. Rainfall–runoff modelling in northern Australia: A guide to modelling strategies in the tropics. *Journal of Hydrology,* 462**,** 28-41.

Pingale, Khare, Jat & Adamowski 2014. Spatial and temporal trends of mean and extreme rainfall and temperature for the 33 urban centers of the arid and semi-arid state of Rajasthan, India. *Atmospheric Research,* 138**,** 73-90.

PMV. 2011. *Government regulations of tropical depressions, storms, floods alert* [Online]. Vietnam: Prime Minister of Vietnam approvement. Available: http://www.atmt.gov.vn/App_File/laws/6fd9d175-448e-45bb-ac8a-9240221f8d98.PDF [Accessed 28 December 2016].

Pohlert 2016. Non-Parametric Trend Tests and Change-Point Detection.

Porter & McMahon 1971. A model for the simulation of streamflow data from climatic records. *Journal of Hydrology,* 13**,** 297-324.

Prosdocimi, Kjeldsen & Miller 2015. Detection and attribution of urbanization effect on flood extremes using nonstationary flood- frequency models. *Water resources research,* 51**,** 4244-4262.

Purvis, Bates & Hayes 2008. A probabilistic methodology to estimate future coastal flood risk due to sea level rise. *Coastal engineering,* 55**,** 1062-1073.

QGIS 2015. Quantum GIS Geographic Information System. Open Source Geospatial Foundation Project.

Rakhecha & Soman 1994. Trends in the annual extreme rainfall events of 1 to 3 days duration over India. *Theoretical and Applied Climatology,* 48**,** 227-237.

Razmi, Golian & Zahmatkesh 2017. Non-stationary frequency analysis of extreme water level: application of annual maximum series and peak-over threshold approaches. *Water resources management,* 31**,** 2065-2083.

Renard & Lang 2007. Use of a Gaussian copula for multivariate extreme value analysis: some case studies in hydrology. *Advances in Water Resources,* 30**,** 897-912.

Ritter & Muñoz-Carpena 2013. Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments. *Journal of Hydrology,* 480**,** 33-45.

Salas & Obeysekera 2013. Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. *Journal of Hydrologic Engineering,* 19**,** 554-568.

Sampson, Fewtrell, Duncan, Shaad, Horritt & Bates 2012. Use of terrestrial laser scanning data to drive decimetric resolution urban inundation models. *Advances in water resources,* 41**,** 1-17.

Sarhadi, Burn, Concepción Ausín & Wiper 2016. Time varying nonstationary multivariate risk analysis using a dynamic Bayesian copula. *Water Resources Research,* 52**,** 2327-2349.

Scarrott & MacDonald 2012. A review of extreme value threshold es-timation and uncertainty quantification. *REVSTAT–Statistical Journal,* 10**,** 33-60.

Serafin & Ruggiero 2014. Simulating extreme total water levels using a time- dependent, extreme value approach. *Journal of Geophysical Research: Oceans,* 119**,** 6305-6329.

Sharif, Al-Juaidi, Al-Othman, Al-Dousary, Fadda, Jamal-Uddeen & Elhassan 2016. Flood hazards in an urbanizing watershed in Riyadh, Saudi Arabia. *Geomatics, Natural Hazards and Risk,* 7**,** 702-720.

Shen, Wang, Cheng, Rui & Ye 2015. Integration of 2-D hydraulic model and high-resolution lidar-derived DEM for floodplain flow modeling. *Hydrology and Earth System Sciences,* 19**,** 3605-3616.

Shepherd, Pierce & Negri 2002. Rainfall modification by major urban areas: Observations from spaceborne rain radar on the TRMM satellite. *Journal of Applied Meteorology,* 41**,** 689-701.

Shin, Guillaume, Croke & Jakeman 2015. A review of foundational methods for checking the structural identifiability of models: Results for rainfall-runoff. *Journal of Hydrology,* 520**,** 1-16.

Shin & Kim 2017. Assessment of the suitability of rainfall–runoff models by coupling performance statistics and sensitivity analysis. *Hydrology Research,* 48**,** 1192-1213.

Skjong, Naess & Næss 2013. Statistics of Extreme Sea Levels for Locations along the Norwegian Coast. *Journal of Coastal Research,* 29**,** 1029-1048.

Smith, Davey & Cox 2014. Flood hazard. *Water Research Laboratory Technical Report,* 7**,** 17.

Sraj, Bezak & Brilly 2015. Bivariate flood frequency analysis using the copula function: a case study of the Litija station on the Sava River. *Hydrological Processes,* 29**,** 225-238.

Šraj, Viglione, Parajka & Blöschl 2016. The influence of non-stationarity in extreme hydrological events on flood frequency estimation. *Journal of Hydrology and Hydromechanics,* 64**,** 426-437.

Storch & Downes 2011. A scenario-based approach to assess Ho Chi Minh City's urban development strategies against the impact of climate change. *Cities,* 28**,** 517-526.

Strupczewski, Singh & Feluch 2001. Non-stationary approach to at-site flood frequency modelling I. Maximum likelihood estimation. *Journal of Hydrology,* 248**,** 123-142.

Sugahara, Da Rocha & Silveira 2009. Non- stationary frequency analysis of extreme daily rainfall in Sao Paulo, Brazil. *International Journal of Climatology,* 29**,** 1339-1349.

Teng, Huang & Ginis 2018. Hydrological modeling of storm runoff and snowmelt in Taunton River Basin by applications of HEC-HMS and PRMS models. *Natural Hazards,* 91**,** 179-199.

Teng, Vaze, Dutta & Marvanek 2015. Rapid inundation modelling in large floodplains using LiDAR DEM. *Water Resources Management,* 29**,** 2619-2636.

Thompson & Frazier 2014. Deterministic and probabilistic flood modeling for contemporary and future coastal and inland precipitation inundation. *Applied Geography,* 50**,** 1-14.

Ticehurst, Dutta, Karim, Petheram & Guerschman 2015. Improving the accuracy of daily MODIS OWL flood inundation mapping using hydrodynamic modelling. *Natural Hazards,* 78**,** 803-820.

Timbadiya, Patel & Porey 2011. HEC-RAS based hydrodynamic model in prediction of stages of lower Tapi River. *ISH Journal of Hydraulic Engineering,* 17**,** 110-117.

Timbadiya, Patel & Porey 2014a. A 1D–2D coupled hydrodynamic model for river flood prediction in a coastal Urban floodplain. *Journal of Hydrologic Engineering,* 20**,** 05014017.

Timbadiya, Patel & Porey 2014b. One-dimensional hydrodynamic modelling of flooding and stage hydrographs in the lower Tapi River in India. *Current science***,** 708-716.

Trenberth 2011. Changes in precipitation with climate change. *Climate Research,* 47**,** 123-138.

Trenberth, Dai, Rasmussen & Parsons 2003. The changing character of precipitation. *Bulletin of the American Meteorological Society,* 84**,** 1205-1217.

Tsubaki & Fujita 2010. Unstructured grid generation using LiDAR data for urban flood inundation modelling. *Hydrological processes,* 24**,** 1404-1420.

Um, Kim, Markus & Wuebbles 2017. Modeling nonstationary extreme value distributions with nonlinear functions: An application using multiple precipitation projections for US cities. *Journal of Hydrology,* 552**,** 396-406.

UN-Habitat 2011. Cities and climate change: Global report on human settlements 2011. *London: Earthscan*.

UNFPA 2007. State of world population 2007: unleashing the potential of urban growth. *State of world population 2007: unleashing the potential of urban growth.* UNFPA.

Vaze, Post, Chiew, Perraud, Teng & Viney 2011. Conceptual rainfall–runoff model performance with different spatial rainfall inputs. *Journal of Hydrometeorology,* 12**,** 1100-1112.

Villafuerte & Matsumoto 2015. Significant influences of global mean temperature and ENSO on extreme rainfall in Southeast Asia. *Journal of Climate,* 28**,** 1905-1919.

Villafuerte, Matsumoto & Kubota 2015. Changes in extreme rainfall in the Philippines (1911–2010) linked to global mean temperature and ENSO. *International Journal of Climatology,* 35**,** 2033-2044.

Wang, Yang, Saito, Liu & Sun 2006. Interannual and seasonal variation of the Huanghe (Yellow River) water discharge over the past 50 years: connections to impacts from ENSO events and dams. *Global and Planetary Change,* 50**,** 212-225.

Weisse, Bellafiore, Menéndez, Méndez, Nicholls, Umgiesser & Willems 2014. Changing extreme sea levels along European coasts. *Coastal engineering,* 87**,** 4-14.

Westra, Alexander & Zwiers 2013. Global increasing trends in annual maximum daily precipitation. *Journal of Climate,* 26**,** 3904-3918.

Wi, Valdés, Steinschneider & Kim 2016. Non-stationary frequency analysis of extreme precipitation in South Korea using peaks-over-threshold and annual maxima. *Stochastic environmental research and risk assessment,* 30**,** 583-606.

Woodworth, Menéndez & Gehrels 2011. Evidence for century-timescale acceleration in mean sea levels and for recent changes in extreme sea levels. *Surveys in geophysics,* 32**,** 603-618.

World Bank. 2010. *Climate risks and adaptation in Asian coastal megacities: a synthesis report* [Online]. Available: http://documents.worldbank.org/curated/en/866821468339644916/Climate-risks-and-adaptation-in-Asian-coastal-megacities-a-synthesis-report [Accessed 20 December 2016].

Wu 2013. Interannual modulation of the Pacific Decadal Oscillation (PDO) on the low-latitude western North Pacific. *Progress in Oceanography,* 110**,** 49-58.

Xia, Falconer, Wang & Xiao 2014. New criterion for the stability of a human body in floodwaters. *Journal of Hydraulic Research,* 52**,** 93-104.

Xia, Teo, Lin & Falconer 2011. Formula of incipient velocity for flooded vehicles. *Natural Hazards,* 58**,** 1-14.

Yan, Xiong, Guo, Xu, Xia & Du 2017. Comparison of four nonstationary hydrologic design methods for changing environment. *Journal of Hydrology,* 551**,** 132-150.

Yaseen, Jaafar, Deo, Kisi, Adamowski, Quilty & El-Shafie 2016. Stream-flow forecasting using extreme learning machines: A case study in a semi-arid region in Iraq. *Journal of Hydrology,* 542**,** 603-614.

Yen, Chen, Hu, Tzeng, DINH, NGUYEN & Wong 2011. Interannual variation of the fall rainfall in Central Vietnam. *Journal of the Meteorological Society of Japan. Ser. II,* 89**,** 259-270.

Yilmaz, Imteaz & Perera 2016. Investigation of non-stationarity of extreme rainfalls and spatial variability of rainfall intensity–frequency–duration relationships: a case study of Victoria, Australia. *International Journal of Climatology***,** n/a-n/a.

Yin, Yu, Yin, Wang & Xu 2013. Multiple scenario analyses of Huangpu River flooding using a 1D/2D coupled flood inundation model. *Natural hazards,* 66**,** 577-589.

Yoshida & Dittrich 2002. 1D unsteady-state flow simulation of a section of the upper Rhine. *Journal of Hydrology,* 269**,** 79-88.

Zhang, Gu, Singh, Xiao & Chen 2015. Evaluation of flood frequency under non-stationarity resulting from climate indices and reservoir indices in the East River basin, China. *Journal of Hydrology,* 527**,** 565-575.

Zhang & Shen 2007. Study and application of steady flow and unsteady flow mathematical model for channel networks. *Journal of Hydrodynamics, Ser. B,* 19**,** 572-578.

Zhang & Singh 2007. Bivariate rainfall frequency distributions using Archimedean copulas. *Journal of Hydrology,* 332**,** 93-109.

Zhang, Xu, Zhang, Chen, Liu & Lin 2008. Spatial and temporal variability of precipitation maxima during 1960–2005 in the Yangtze River basin and possible association with large-scale circulation. *Journal of Hydrology,* 353**,** 215-227.

Research papers resulting from the thesis

Publications in Peer-Reviewed Journals

- **Le Thi Hoa Binh**, N. V. Umamahesh, E. V. Rathnam & Vu Hai Son 2018. Modeling nonstationary extreme water levels considering local covariates in o Chi Minh City, Vietnam. ***Journal of Hydrologic Engineering***, 23(10), 04018042. DOI:10.1061/(ASCE)HE.1943-5584.0001697

- **Le Thi Hoa Binh**, N. V. Umamahesh & E. V. Rathnam 2019. High-resolution flood hazard mapping based on nonstationary frequency analysis: case study of Ho Chi Minh City, Vietnam. ***Hydrological Sciences Journal,*** Volume 64, Issue 3, https://doi.org/10.1080/02626667.2019.1581363

- **Le Thi Hoa Binh**, V. Agilan, N. V. Umamahesh & E. V. Rathnam 2019. Modelling spatial variation of extreme precipitation over Ho Chi Minh City under nonstationary condition. ***Acta Geophysica.*** https://doi.org/10.1007/s11600-019-00295-1

- **Le Thi Hoa Binh,** E. Venkata Rathnam & N.V. Umamahesh 2019. Copula-based bivariate flood frequency analysis under nonstationary condition: Case study for extreme events in Ho Chi Minh City. ***Water Resources Management (Under review).***

Acknowledgments

I am highly grateful to all those who helped me in completing this thesis possible.

First of all, I would like to specially thank Prof. N. V. Umamahesh and Prof. E. V. Rathnam for encouraging and guiding me as well as for sharing their knowledge and valuable discussions during the doctorate. I am thankful to Prof. K. V. Jayakumar, Prof. P Anand Raj and Prof. M. Sydulu, the members of the Doctoral Scrutiny Committee for their constant support at various stages of the work. I also thank all the Heads of the Department of Civil Engineering during my stay here for their support.

I would like to express my gratitude to the India Council for Cultural Relations, Government of India and Government of Vietnam for the financial assistance through a scholarship to complete this study. I also thank the Thuyloi University, Ho Chi Minh City, Vietnam for the support to take up this research work.

National Hydro-Meteorological Service of Vietnam, Trian Hydropower Joint Stock Company and Dautieng-Phuochoa Limited Company provided me with all the necessary data for this study. I place on record my deep sense of gratitude to the officials of these organizations.

I thank my friends, Agilan, Uday, Sreelatha, Vinay, Kiran, Sruthi and Garima, for their companionship and encouragement, especially when I first came to India.

I am also thankful to all faculty members and another staff member of National Institute of Technology, Warangal for their support during my course in India.

Finally, I would like to express my heartfelt thanks to my family for both financial and emotional support and encouragements.

Le Thi Hoa Binh

Date: