REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
06-01-2020	Final Report	1-Dec-2014 - 31-Aug-2015

4. TITLE AND SUBTITLE

Final Report: Frameworks for Analysis of Regional, concurrent, Conditional and Non-Stationary Extremes in Geosciences

5a. CONTRACT NUMBER
W911NF-14-1-0684

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES

University of California - Irvine
141 Innovation Drive, Suite 250

Irvine, CA 92697 -7600

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
65951-EV-II.10

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Amir Aghakouchak
UU	UU	UU	UU		19b. TELEPHONE NUMBER 949-824-9350

RPPR Final Report
as of 13-Jan-2020

Agency Code:

Proposal Number: 65951EVII **Agreement Number: W911NF-14-1-0684**
INVESTIGATOR(S):

Name: Amir Aghakouchak
Email: amir.a@uci.edu
Phone Number: 9498249350
Principal: Y

Organization: **University of California - Irvine**
Address: 141 Innovation Drive, Suite 250, Irvine, CA 926977600
Country: USA
DUNS Number: 046705849 EIN: 952226406
Report Date: 30-Nov-2015 Date Received: 06-Jan-2020
Final Report for Period Beginning 01-Dec-2014 and Ending 31-Aug-2015
Title: Frameworks for Analysis of Regional, concurrent, Conditional and Non-Stationary Extremes in Geosciences
Begin Performance Period: 01-Dec-2014 **End Performance Period:** 31-Aug-2015
Report Term: 0-Other
Submitted By: Amir Aghakouchak Email: amir.a@uci.edu
Phone: (949) 824-9350
Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees: 2 **STEM Participants:** 2

Major Goals: The objectives of this STIR project was to investigate the merit of the following model concepts:

(1) A model for regional non-stationary analysis of extremes with constant and time-varying exceedance probability concepts. This will allow analysis of extremes in geosciences across different spatial scales under non-stationary assumption. The model is named Process-informed Nonstationary Extreme Value Analysis (ProNEVA) and can integrate time or a physically-based covariate to describe change in statistics of extremes. The source code of the toolbox along with a Graphical User Interface (GUI) is already freely available to the public.

(2) An empirical Bayesian-based extreme value model for assessing concurrent and conditional extremes. This will allow deriving and assessing the full distribution functions of concurrent (joint) extremes in a changing environment.

(3) A comprehensive and generalized framework for uncertainty assessment of extremes using the concept of Differential Evolution Markov Chain (DE-MC). This model will allow deriving quantitative uncertainty estimates for extremes in a non-stationary world.

Accomplishments: In a publication supported by this project, we show that given nonstationarity, current precipitation Intensity-Duration-Frequency (IDF) curves can substantially underestimate precipitation extremes and thus, they may not be suitable for infrastructure design in a changing climate. We show that a stationary climate assumption may lead to underestimation of extreme precipitation by as much as 60%, which increases the flood risk and failure risk in infrastructure systems. We present a generalized framework for estimating nonstationary IDF curves and their uncertainties using Bayesian inference. The methodology can potentially be integrated in future design concepts.

In this project, we developed a new estimation strategy for estimating the parameters of a new conditional extreme value model. The technique makes use of empirical Bayes estimation for the conditional likelihood that otherwise does not have a simple closed?form expression. The approach is tested on simulations from different types of extreme dependence (and independence) structures, as well as for two real data cases consisting of precipitation analysis conditional on extreme temperature in Boulder, Colorado, and Los Angeles, California, USA. The strategy generally has good coverage when informative priors are used for one of the parameters, except for the independence case where the coverage is low until the sample size reaches about 50. Results for the precipitation and temperature data are found to be consistent with the semi?non?parametric strategy. The presented model can be potentially applied in a wide variety of science fields, especially in earth, environment and climate sciences.

RPPR Final Report
as of 13-Jan-2020

This project developed a generalized approach for uncertainty assessment of extremes using the concept of Differential Evolution Markov Chain (DE-MC).

This project led to development of Process-informed Nonstationary Extreme Value Analysis (ProNEVA).

Training Opportunities: Nothing to Report

Results Dissemination: Cheng L., AghaKouchak A., 2014, Precipitation Intensity-Duration-Frequency Curves for Infrastructure Design in a Changing Climate, Scientific Reports, 4, 7093, doi: 10.1038/srep07093.

Cheng L., Gilleland E., Heaton M.J., AghaKouchak A., 2014, Empirical Bayes estimation for the conditional extreme value model, Stat, 3, 391-406, doi: 10.1002/sta4.71.

Ragno E., AghaKouchak A., Cheng L., Sadegh, M., 2019, A Generalized Framework for Process-informed Nonstationary Extreme Value Analysis, Advances in Water Resources, 130, 270-282, doi: 10.1016/j.advwatres.2019.06.007.

Honors and Awards: The paper Cheng L., AghaKouchak A., 2014, supported by this project, received the 2017 IAHS/STAHY Best Paper Award.

Full Citation: Cheng, L., Aghakouchak, A. Nonstatlonary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate (2014) Scientific Reports, 4, art. no. 7093.

Protocol Activity Status:

Technology Transfer: Nothing to Report

PARTICIPANTS:

Participant Type: PD/PI
Participant: Amir AghaKouchak
Person Months Worked: 1.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Graduate Student (research assistant)
Participant: Linyin Cheng
Person Months Worked: 12.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

ARTICLES:

RPPR Final Report
as of 13-Jan-2020

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Climate Dynamics
Publication Identifier Type: DOI Publication Identifier: 10.1007/s00382-015-2625-y
Volume: 44 Issue: First Page #: 2947
Date Submitted: 1/6/20 12:00AM Date Published: 5/1/15 9:00PM
Publication Location:
Article Title: Non-stationary return levels of CMIP5 multi-model temperature extremes
Authors: Linyin Cheng, Thomas J. Phillips, Amir AghaKouchak
Keywords: Nonstationary, Extremes, Return Period,
Abstract: The objective of this study is to evaluate to what extent the CMIP5 climate model simulations of the climate of the twentieth century can represent observed warm monthly temperature extremes under a changing environment. The biases and spatial patterns of 2-, 10-, 25-, 50- and 100-year return levels of the annual maxima of monthly mean temperature (hereafter, annual temperature maxima) from CMIP5 simulations are compared with those of Climatic Research Unit (CRU) observational data considered under a non-stationary assumption. The results show that CMIP5 climate models collectively underestimate the mean annual maxima over arid and semi-arid regions that are most subject to severe heat waves and droughts. Furthermore, the results indicate that most climate models tend to underestimate the historical annual temperature maxima over the United States and Greenland, while generally disagreeing in their simulations over cold regions.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Stat
Publication Identifier Type: DOI Publication Identifier: 10.1002/sta4.71
Volume: 3 Issue: 1 First Page #: 391
Date Submitted: 1/6/20 12:00AM Date Published: 3/1/14 8:00AM
Publication Location:
Article Title: Empirical Bayes estimation for the conditional extreme value model
Authors: Linyin Cheng, Eric Gilleland, Matthew J. Heaton, Amir AghaKouchak
Keywords: Empirical Bayes, Extreme Value Analysis
Abstract: A new estimation strategy for estimating the parameters of the Heffernan and Tawn conditional extreme value model is proposed. The technique makes use of empirical Bayes estimation for the conditional likelihood that otherwise does not have a simple closed?form expression. The approach is tested on simulations from different types of extreme dependence (and independence) structures, as well as for two real data cases consisting of precipitation analysis conditional on extreme temperature in Boulder, Colorado, and Los Angeles, California, USA. The strategy generally has good coverage when informative priors are used for one of the parameters, except for the independence case where the coverage is low until the sample size reaches about 50. Results for the precipitation and temperature data are found to be consistent with the semi?non?parametric strategy.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 13-Jan-2020

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Scientific Reports
Publication Identifier Type: DOI Publication Identifier: 10.1038/srep07093
Volume: 4 Issue: 1 First Page #:
Date Submitted: 1/6/20 12:00AM Date Published: 11/1/14 2:00PM
Publication Location:
Article Title: Nonstationary Precipitation Intensity-Duration-Frequency Curves for Infrastructure Design in a Changing Climate
Authors: Linyin Cheng, Amir AghaKouchak
Keywords: Extreme precipitation, nonstationarity
Abstract: Extreme climatic events are growing more severe and frequent, calling into question how prepared our infrastructure is to deal with these changes. Current infrastructure design is primarily based on precipitation Intensity-Duration-Frequency (IDF) curves with the so-called stationary assumption, meaning extremes will not vary significantly over time. However, climate change is expected to alter climatic extremes, a concept termed nonstationarity. Here we show that given nonstationarity, current IDF curves can substantially underestimate precipitation extremes and thus, they may not be suitable for infrastructure design in a changing climate. We show that a stationary climate assumption may lead to underestimation of extreme precipitation by as much as 60%, which increases the flood risk and failure risk in infrastructure systems. We then present a generalized framework for estimating nonstationary IDF curves and their uncertainties using Bayesian inference. T

Acknowledged Federal Support: **Y**

Final Report

Army Research Laboratory Award No. W911NF-14-1-0684

Broad Agency Announcement for Basic and Applied Scientific Research Short-Term Innovative Research (STIR):

Frameworks for Analysis of Regional, Concurrent, Conditional and Non-Stationary Extremes in Geosciences

---

PI: **Amir AghaKouchak**
University of California, Irvine
E/4130 Engineering Gateway
Irvine, CA 92697
Tel: (949)-824-9350
Email: amir.a@uci.edu

Submitted to:

United States Department of Defense
Army Research Laboratory

Contents

Title: Frameworks for Analysis of Regional, Concurrent, Conditional and Non-Stationary Extremes in Geosciences
PI: Amir AghaKouchak, University of California-Irvine, Irvine, CA, Email: amir.a@uci.edu

Executive Summary

The objective of this project was to assess the merit and potential applications of innovative new statistical concepts for analysis of extremes in geosciences. Specifically, the proposed study investigated merits of the following model concepts:

- **Objective A**: A model for regional non-stationary analysis of extremes with constant and time-varying exceedance probability concepts. This will allow analysis of extremes in geosciences across different spatial scales under non-stationary assumption. The model is named Process-informed Nonstationary Extreme Value Analysis (ProNEVA) and can integrate time or a physically-based covariate to describe change in statistics of extremes. The source code of the toolbox along with a Graphical User Interface (GUI) is already freely available to the public.

- **Objective B**: A comprehensive and generalized framework for uncertainty assessment of extremes using the concept of Differential Evolution Markov Chain (DE-MC). This model will allow deriving quantitative uncertainty estimates for extremes in a non-stationary world.

- **Objective C**: An empirical Bayesian-based extreme value model for assessing concurrent and conditional extremes. This will allow deriving and assessing the full distribution functions of concurrent (joint) extremes in a changing environment.

In this report, Section 1 summarizes our findings and results for Objectives A and B, whereas Section 2 describes our findings for Objective C.
This project led to **3 published peer-reviewed publications and 2 manuscripts in preparation** fully/partially supported by ARL. The report summarizes the developed methodologies, results and conclusions of the project presented in the following publications:

Cheng L., AghaKouchak A., 2014, Precipitation Intensity-Duration-Frequency Curves for Infrastructure Design in a Changing Climate, Scientific Reports, 4, 7093, doi: 10.1038/srep07093.

Cheng L., Gilleland E., Heaton M.J., AghaKouchak A., 2014, Empirical Bayes estimation for the conditional extreme value model, Stat, 3, 391-406, doi: 10.1002/sta4.71.

Cheng L., AghaKouchak A., Phillips T., 2015, Non-stationary Return Levels of CMIP5 Multi-Model Temperature Extremes, Climate Dynamics, 44(11), 2947-2963, doi: 10.1007/s00382-015-2625-y

This project focused more on the theoretical aspects. We are now using the method developed in Cheng and AghaKouchak, 2014 and Cheng et al., 2015 for analyzing precipitation extremes under different future emission scenarios. Furthermore, we are applying the method developed in Cheng et al., 2014 for assessing concurrent and conditional extremes in joint precipitation-temperature records.

1 Non-Stationary Analysis of Extremes

Abstract: The evolving climate conditions and anthropogenic factors, such as CO_2 emissions, urbanization and population growth, can cause changes in weather and climate extremes. Most current risk assessment models rely on the assumption of stationarity (i.e., no temporal change in statistics of extremes). A number of nonstationary models have been developed with a focus on temporal changes in extremes, ignoring the underlying physical driver. Here, we present Process-based Nonstationary Extreme Value Analysis (ProNEVA) as a generalized tool for incorporating different types of physical drivers (i.e., underlying processes), stationary and nonstationary concepts, and extreme value analysis methods (i.e., annual maxima, peak-over-threshold). ProNEVA builds upon a newly-developed hybrid evolution Markov Chain Monte Carlo (MCMC) approach for numerical parameters estimation and uncertainty assessment. This offers more robust uncertainty estimates of return periods of climatic extremes under both stationary and nonstationary assumptions. ProNEVA is designed as a generalized tool allowing using different types of data and nonstationarity concepts (e.g., physical-based or purely statistical) into account. In this report, we show a wide range of applications describing changes in: annual maxima river discharge in response to urbanization, annual maxima sea levels over time, annual maxima temperatures in response to CO_2 emissions in the atmosphere, and precipitation with a peak-over-threshold approach. ProNEVA is freely available to the public and includes a user-friendly Graphical User Interface (GUI) to enhance its implementation.

1.1 Introduction

Natural hazards pose significant threats to public safety, infrastructure integrity, natural resources, and economic development around the globe. In recent years, the frequency and impacts of extremes have increased substantially in many parts of the world (e.g., Melillo et al., 2014; Coumou and Rahmstorf, 2012; Alexander et al., 2006b; Mazdiyasni et al., 2017; Mallakpour and Villarini, 2017; Hallegatte et al., 2013; Wahl et al., 2015; Vahedifard et al., 2016; Jongman et al., 2014; Aghakouchak et al., 2014). For this reason, there is a great deal of interest in understanding how extreme events will change in the future. Historical observations are the main source of information on extremes (Klemeš, 1974; Koutsoyiannis and Montanari, 2007) and stochastic models are used to infer frequency and variability of extremes based on historical records (e.g., Katz et al., 2002).

Stochastic models used to study extremes can be broadly categorized into two groups: stationary and nonstationary (e.g., Salas and Pielke Sr, 2002; Coles and Pericchi, 2003; Griffis and Stedinger, 2007; Obeysekera and Salas, 2013; Serinaldi and Kilsby, 2015; Madsen et al., 2013; Koutsoyiannis and Montanari, 2015). In a stationary model, the obser-

vations are assumed to be drawn from a probability distribution function with constant parameters (i.e., statistics of extremes do not change over time). In a nonstationary model, however, the parameters of the underlying probability distribution function change over time or in response to a given covariate (i.e., model accounts for changes in statistics of extremes) (Sadegh et al., 2015).

Water resources practices (e.g., flood and precipitation frequency analysis) have traditionally adopted stationary models primarily for the sake of simplicity (Milly et al., 2008), though changes in the water cycle and Earth system processes are inherent (Montanari et al., 2013). Over the past decades, increasing surface temperatures (e.g., Barnett et al., 1999; Villarini et al., 2010; Melillo et al., 2014; Diffenbaugh et al., 2015; Fischer and Knutti, 2015; Mazdiyasni and AghaKouchak, 2015), more intense rainfall events (e.g., Zhang and Li, 2007; Villarini et al., 2010; Min et al., 2011; Marvel and Bonfils, 2013; Westra et al., 2013; Cheng et al., 2014b; Fischer and Knutti, 2016; Mallakpour and Villarini, 2017), changes in river discharge (e.g., Villarini et al., 2009a,b; Hurkmans et al., 2009; Stahl et al., 2010), and sea level rise (e.g., Holgate, 2007; Haigh et al., 2010; Wahl et al., 2011) have been observed and to a great extent attributed to anthropogenic activities (e.g., human-caused climate change, urbanization). The observed hydrologic trends, which can be in response to a physical process (e.g., changes in emissions, temperatures, climatic cycles) or only perceived (statistical) (Matalas, 1997), have challenged the stationary assumption (Milly et al., 2008).

Several studies have promoted the idea of moving away from stationary models to ensure capturing the changing properties of extremes (Milly et al., 2008). However, some have criticized this viewpoint particularly because the assumption of nonstationarity implies adding a deterministic component in the stochastic process, which must be justified by a well-understood process (Koutsoyiannis, 2011; Matalas, 2012; Lins and Cohn, 2011; Koutsoyiannis and Montanari, 2015). Montanari and Koutsoyiannis (2014) noted that more efforts should focus on including relevant physical processes in stochastic models, and suggested stochastic-process-based models as a way to bridge the gap between physically-based models without statistics and statistical models without physics.

Following the recommendation by Montanari and Koutsoyiannis (2014), we propose a generalized framework named *Process-based Nonstationary Extreme Value Analysis* (ProNEVA) in which the nonstationarity component is defined by a temporal or process-based dependence of the observed extremes on an explanatory variable (i.e., a physical driver). Here, process-based dependence corresponds to a process or driver that can alter the statistics of extremes. For example, ProNEVA can be used for analyzing changes in extreme temperatures as a function of CO_2 emissions as the covariate. It is widely recognized that higher amount of CO_2 in the atmosphere results in a warmer climate (e.g., Zwiers et al., 2011; Fischer and Knutti, 2015; Barnett et al., 1999). For this reason, CO_2 emissions can be considered a process-based covariate for explaining temperature extremes. Other exam-

ples include temperature or large scale climatic circulations as covariates for rainfall, and CO_2 concentration or temperature as covariates for sea level rise.

1.2 Background and Method

1.2.1 Process-Based Nonstationarity Extreme Value Analysis

Extreme Value Theory (EVT) provides the bases for estimating the magnitude and frequency of hazardous events (including natural and non-natural extreme events) (Coles, 2001). Most applications utilize either the Generalized Extreme Value distribution (GEV) or the Generalized Pareto distribution (GP) for describing the behavior of extremes. The former is applied to the annual maxima of a variable (e.g., a time series consisting of the most extreme daily rainfall from each year of the record), while the latter is used to describe extremes above a predefined threshold (e.g., all independent river flow values above the flood stage). Both GEV and GP allow incorporating nonstationarity through varying parameters. Several studies have investigated methodologies for testing the assumptions of stationarity and nonstationarity in hydrology, climatology, and earth system sciences (e.g., Katz et al., 2002; Sankarasubramanian and Lall, 2003; Cooley et al., 2007; Mailhot et al., 2007; Huard et al., 2009; Villarini et al., 2009a, 2010; Vogel et al., 2011; Zhu et al., 2012; Willems et al., 2012; Katz, 2013; Obeysekera and Salas, 2013; Salas and Obeysekera, 2014; Rosner et al., 2014; Yilmaz and Perera, 2014; Mirhosseini et al., 2014; Cheng and AghaKouchak, 2014a; Volpi et al., 2015; Read and Vogel, 2015; Sadegh et al., 2015; Krishnaswamy et al., 2015; Mirhosseini et al., 2015; Mondal and Mujumdar, 2015; Lima et al., 2016; Sarhadi and Soulis, 2017; Salas et al., 2018; Yan et al., 2018; Bracken et al., 2018; Ragno et al., 2018).

A number of packages and software tools are currently available including the R-package *ismev* (Gilleland et al., 2013; Gilleland and Katz, 2016) where nonstationarity is modeled as a linear regression function of generic covariates (Gilleland et al., 2013). *extRemes* offers EVA capability and evaluates the underlying uncertainties with respect to parameters (Gilleland and Katz, 2016). *extRemes* also allows tail-dependence analysis and a declustering technique for peak over threshold analysis. The package *climextRemes* (available also in Python) builds upon *extRemes* and includes an estimate of the risk ratio for event attribution analyses. R packages *vgam* and *gamlss* are available for modeling nonstationarity through generalized additive models (see for example Villarini et al. (2009a)). The package *GEVcdn* estimates the parameters of a nonstationary GEV distribution using a conditional density method (Cannon, 2010).

Cheng et al. (2014b) developed a Bayesian-based framework, *Nonstationary Extreme Value Analysis* (NEVA) toolbox that estimates the parameters of GEV and GP distributions and their associated uncertainty for time-dependent extremes (available in Matlab). In the nonstationary case, the parameters are modeled as a linear function of time. NEVA

also includes return level curves based on the concept of expected waiting time (Wigley, 2009; Olsen et al., 1998; Salas and Obeysekera, 2014) and effective return level (Katz et al., 2002). The package *nonstationary Flood Frequency Analysis* estimates the parameters of the Log-Pearson Type III distribution as a linear function of time, based on Bayesian inference approach (Luke et al., 2017). The *tsEVA* toolbox implements the Transformed-Stationary (TS) methodology described in Mentaschi et al. (2016), which comprises of, first, a transformation of a nonstationary time series into a stationary one, so that the stationary EVA theory can be applied, and then a reverse-transformation of the results to include the nonstationary components in the GEV and the GP distributions.

However, the existing tools for implementing EVA under the nonstationary assumption have a number of limitations, mainly lack of a generalized framework for incorporating physical-based covariates. Moreover, most existing tools are incapable of handling parameter estimation in response to a physical covariate (e.g., when the parameters are estimated as a non-linear function of a covariate). To address these limitations, we present ProNEVA, which builds upon NEVA package (Cheng et al., 2014b) but expands to process-based nonstationary extreme value analysis. In addition to stationary EVA, ProNEVA allows nonstationary analyses using user-defined co-variates, which could be time or a physical-based variable. Figure 1 depicts the core structure of ProNEVA.

ProNEVA offers parameter estimation, uncertainty quantification, and a comprehensive assessment of the goodness of fit. The key features of ProNEVA are described as follows: (a) the model includes the most common distribution functions used for extreme value analysis including the GEV, GP, and LP3 distributions; (b) for nonstationary analysis, the users can select both the covariate and the choice of function for describing change in parameters; (c) the covariate can be any user-defined physical covariate; (d) the model also includes a default time-covariate (i.e., describing change over time without a physical covariate); (e) the function describing change in parameters with respect to the covariate can be linear, exponential, or quadratic; (f) the users can select the GP distribution threshold (peak-over-threshold) as a constant value or as a linear quantile regression function of the choice covariate; (g) ProNEVA estimates the distribution parameters based on a Bayesian inference approach; (h) the model allows using a wide range of priors for parameters including the uniform, normal, and gamma distributions; (i) ProNEVA samples from the posterior distribution function of the parameters using a newly-developed hybrid evolution Markov Chain Monte Carlo (MCMC) approach (Sadegh et al., 2017), which provides a more robust numerical parameter estimation and uncertainty quantification; (j) different model diagnostics and model selection indices (e.g., RMSE, AIC, BIC) are implemented to provide supporting information; (k) ProNEVA includes additional exploratory data analysis tools such as the Mann-Kendall test for monotonic trends and the White test for homoscedasticity in time series; (l) in addition to the source code, a Graphical User Interface (GUI) for ProNEVA is also available for easier implementation

(see Supplementary Material); finally, (m) ProNEVA is intended for a broad audience and hence, it is structured such that users can easily customize and modify it based on their needs.

In the reminder of the report, a detailed description of ProNEVA is provided. Four different example application are presented with different variables (e.g., precipitation, sea level, temperature, river discharge) and different covariates (time, CO_2 emissions in the atmosphere, urbanization). ProNEVA can be used for analyzing annual maxima (also known as block maxima) using the GEV and LP3 distributions, and peaks over threshold (POT) or partial duration series using the GP distribution. In the following, we provide a brief overview of the extreme value models and their parameters.

1.2.2 Generalized Extreme Value (GEV)

The GEV function is used to model maxima time series. The National Oceanic and Atmospheric Administration (NOAA), for example, derives precipitation Intensity-Duration-Frequency (IDF) curves based on the GEV distribution. GEV is also widely used in other fields including finance, seismology, reliability assessment (bridge performance assessment (e.g., Ming et al., 2009)). The GEV cumulative distribution function is (Coles, 2001):

$$\Psi_{GEV}(X) = exp\Big\{ -\Big(1 + \xi \cdot \Big(\frac{X-\mu}{\sigma}\Big)\Big)^{-\frac{1}{\xi}}\Big\} \tag{1}$$

for $\xi \cdot \left(\frac{X-\mu}{\sigma}\right) > 0$. μ, σ, and ξ are the parameters of the distribution: μ is the location parameter and represents the center of the distribution; $\sigma > 0$ is the scale parameter and describes the distribution of the data around μ; ξ is the shape parameter and defines the tail behavior of the distribution.

The stationary GEV model can be extended for dependent series by letting the parameters of the distribution be a function of a general covariate X_c, i.e., $\mu(X_c)$, $\sigma(X_c)$, $\xi(X_c)$, (Coles, 2001). Hence, the nonstationary form of eq. 1 is described as:

$$\Psi_{GEV}(X|X_c) = exp\Big\{ -\Big(1 + \xi(X_c)) \cdot \Big(\frac{X-\mu(X_c)}{\sigma(X_c)}\Big)\Big)^{-\frac{1}{\xi(X_c)}}\Big\} \tag{2}$$

For each of the three parameters, the users can select a function to describe change in parameters with respect to time or a covariate (Table S1). The function for each parameter does not constrain the functional relationship used for the other parameters. To ensure the positivity of the scale parameter, $\sigma(X_c)$ is modeled in the log-scale, (Coles, 2001; Katz, 2013). Consequently, the exponential function is not available for $\sigma(X_c)$. Moreover, the shape parameter $\xi(X_c)$ is known to be a difficult parameter to precisely estimate even in the stationary case, (Coles, 2001), especially for short time series, (Papalexiou and Kout-

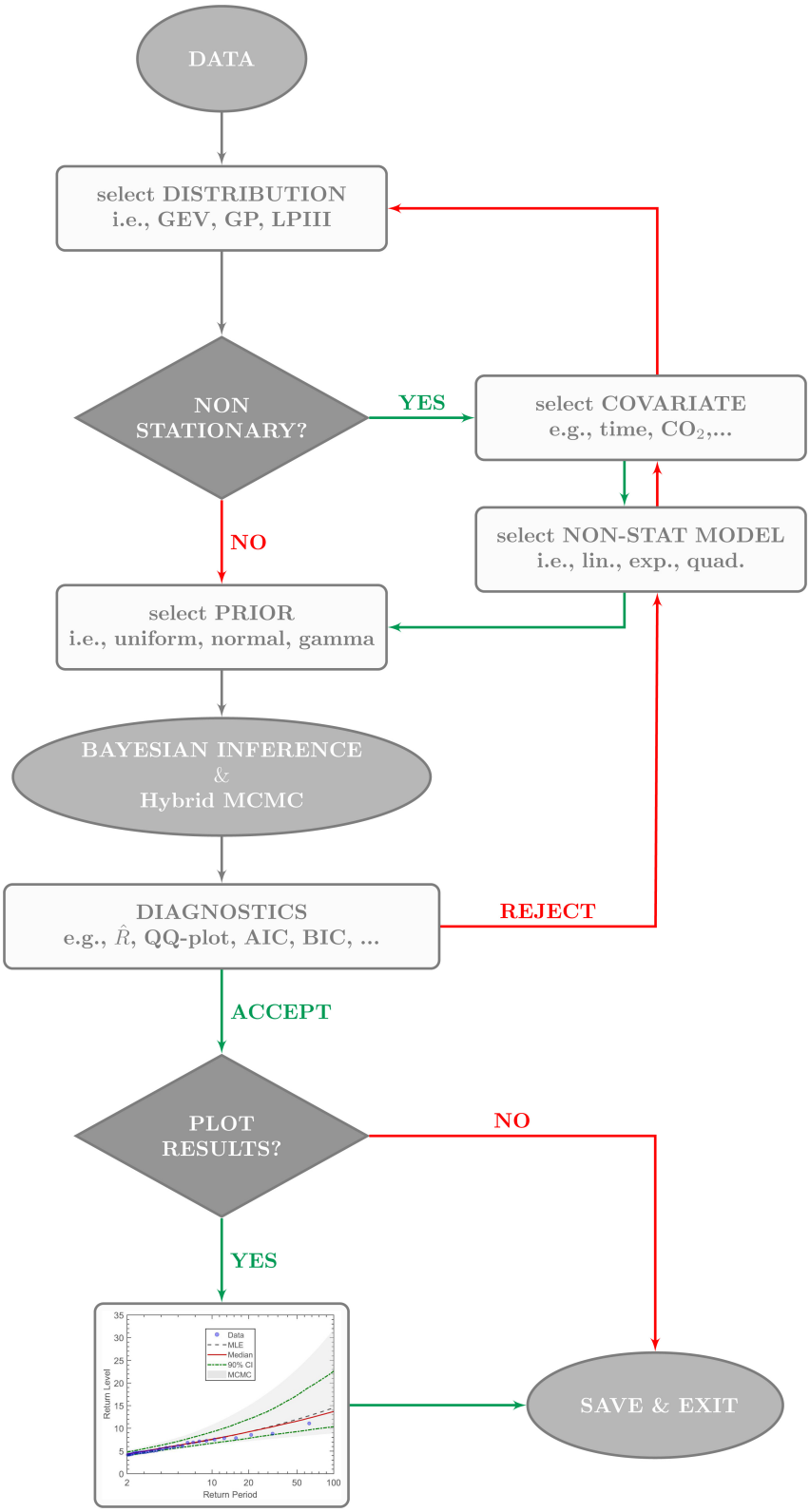

Figure 1: Flowchart representing the core structure of the Matlab Toolbox ProNEVA

soyiannis, 2013). For this reason, only the linear function is included for $\xi(X_c)$.

1.2.3 Generalized Pareto (GP)

The GP distribution is used for time series sampled based on the POT method. The GP distribution has been applied to precipitation (e.g., De Michele and Salvadori, 2003), earthquake data (e.g., Pisarenko and Sornette, 2003), wind speed (Holmes and Moriarty, 1999), and economic data (e.g., Gençay and Selçuk, 2004), among others. The GP cumulative distribution function is as follows (Coles, 2001):

$$\Psi_{GP}(X) = 1 - \left(1 + \xi \cdot \left(\frac{X-u}{\sigma}\right)\right)^{-\frac{1}{\xi}} \tag{3}$$

for a large enough threshold, u, such that $X > u$, $\sigma > 0$ and $\left(1 + \xi \cdot (\frac{X-u}{\sigma})\right) > 0$. In particular, if $\mathbf{X}$ is a block maxima series following a GEV distribution, then the threshold excesses $\{X > u\}$ have a GP distribution. The parameter ξ of the GP distribution is equal to the parameter ξ of the corresponding GEV distribution (Coles, 2001).
In the nonstationary model of the GP distribution, both the threshold value and the parameters of the distribution can be modeled as a function of the user-covariate, (Coles, 2001).

$$\Psi_{GP}(X|X_c) = 1 - \left(1 + \xi(X_c) \cdot \left(\frac{X-u(X_c)}{\sigma(X_c)}\right)\right)^{-\frac{1}{\xi(X_c)}} \tag{4}$$

Analogous to the GEV case, ProNEVA allows incorporating different functional forms for describing change in parameters over time or with respect to a covariate (Table S2). The same considerations for the GEV parameter functional forms are applied to GP distribution too. In addition, the users can specify the type of threshold u. Two quantile-based options are available: constant or linear. In the case of a linear threshold, a linear regression quantile model is adopted. The α-regression quantile function is (Koenker and Bassett, 1978; Kyselý et al., 2010)

$$\mathbf{Y} = \mathbf{X} \cdot \mathbf{U}(\alpha) + \mathbf{r}^{+} - \mathbf{r}^{-} \tag{5}$$

where $0 < \alpha < 1$ is the quantile, $\mathbf{Y}$ is the column vector of n-observations, $\mathbf{X} = [\,\mathbf{X_c} \quad \mathbf{I_n}\,]$ with $\mathbf{X_c}$ being the column vector of covariance and $\mathbf{I_n}$ the n-identity vector, $\mathbf{U} = [\,u_1 \quad u_0\,]'$ is the vector of the regression coefficients, and $\mathbf{r}^{+}$ and $\mathbf{r}^{-}$ are respectively the positive and negative parts of the residuals. Then, $\mathbf{U}(\mathbf{ff})$ is calculated as the optimal solution to eq. 6 (Koenker and Bassett, 1978; Kyselý et al., 2010).

$$\alpha \cdot \mathbf{I_n}' \cdot \mathbf{r}^{+} + (1-\alpha) \cdot \mathbf{I_n}' \cdot \mathbf{r}^{-} := min \tag{6}$$

1.2.4 Log-Pearson Type III (LP3)

The LP3 distribution has been widely used in hydrology for flood frequency analysis particularly after the release of the USGS Bulletin 17B (U.S. Water Resources Council, 1982). However, it has been applied to other studies, such as design magnitude of earthquakes (Gupta and Deshpande, 1994) and evaluation of apple bud burst time and frost risk (Farajzadeh et al., 2010).

The LP3 distribution characterizes the random variable $Q = ln(X)$, given that X follows a Pearson type III (P3) distribution (Griffis et al., 2007). Hereafter, the natural logarithm is used, however any base can be implemented, such as base-10 as in Bulletin 17B (Griffis et al., 2007). The P3 probability density function is

$$\psi_{P3}(X) = \frac{1}{|\beta| \cdot \Gamma(\alpha)} \cdot \Big(\frac{X-\tau}{\beta}\Big)^{\alpha-1} \cdot \exp\Big(-\frac{X-\tau}{\beta}\Big) \tag{7}$$

defined for $\alpha > 0$, $(X-\tau)/\beta > 0$, and $\Gamma(\alpha)$ being a complete gamma function (Griffis et al., 2007). The parameters α, β, and τ are functions of the first three moments, μ_X, σ_X, γ_X, (Griffis et al., 2007):

$$\alpha = 4/\gamma_X^2 \tag{8}$$

$$\beta = (\sigma_X \cdot \gamma_X)/2 \tag{9}$$

$$\tau = \mu_X - 2 \cdot (\sigma_X/\gamma_X) \tag{10}$$

In the case of nonstationary analysis, the first three moments are modeled as a function of the user-defined covariate $\mathbf{X_c}$ (Table S3). The GEV and GP considerations mentioned above hold for the functions to describe change in parameters.

$$\psi_{P3}(X|X_c) = \frac{1}{|\beta(X_c)| \cdot \Gamma(\alpha(X_c))} \cdot \Big(\frac{X-\tau(X_c)}{\beta(X_c)}\Big)^{\alpha(X_c)-1} \cdot \exp\Big(-\frac{X-\tau(X_c)}{\beta(X_c)}\Big) \tag{11}$$

1.2.5 Parameter Estimation: Bayesian Analysis and Markov Chain Monte Carlo Sampling

ProNEVA estimates the parameters of the selected (non)stationary EVA distribution using a Bayesian approach, which provides a robust characterization of the underlying uncertainty derived from both input errors and model selection. Bayesian analysis has been widely implemented for parameter inference and uncertainty quantification (e.g. Thiemann et al., 2001; Gupta et al., 2008; Cheng et al., 2014b; Kwon and Lall, 2016; Sarhadi et al., 2016; Sadegh et al., 2017; Luke et al., 2017) (Sadegh et al., 2018).

Let θ be the parameter of a given distribution and let $\mathbf{Y} = \{y_1, \ldots, y_n\}$ be the set of n observations. Following Bayes theorem, the probability of θ given $\mathbf{Y}$ (posterior) is pro-

portional to the product of the probability of θ (prior) and the probability of $\mathbf{Y}$ given θ (likelihood function). Assuming independence between the observations $\mathbf{Y}$:

$$p(\theta|\mathbf{Y}) \propto \prod_{i=1}^{n} p(\theta) \cdot p(y_i|\theta) \tag{12}$$

The prior brings priori information, which do not depend on the observed data, into the parameter estimation process. The choice of the prior distribution, then, is somehow subjective, and it is based on prior beliefs about the system of interest (Sadegh et al., 2018). The available prior options in ProNEVA include the uniform, normal, and gamma distributions, providing a variety of possibilities. ProNEVA assumes independence of parameters and hence, each parameter requires its own prior.

The likelihood function coincides with the probability density function of the distribution family (i.e. GEV, GP, or LP3) as representative of the data.

In the case of a nonstationary analysis, the vector of parameters ` includes a higher number of elements than in the stationary case, depending on the functional form selected for each of the distribution's parameters.

The posterior distribution is then delineated using a hybrid-evolution MCMC approach proposed by Sadegh et al. (2017). The MCMC simulation searches for the region of interest with multiple chains running in parallel, which share information on the fly. Moreover, the hybrid-evolution MCMC benefits from an intelligent starting point selection (Duan et al., 1993) and employs Adaptive Metropolis (AM) (Roberts and Sahu, 1997; Haario et al., 1999, 2001; Roberts and Rosenthal, 2009), differential evolution (DE) (Storn and Price, 1997; Ter Braak and Vrugt, 2008; Vrugt et al., 2009), and snooker update (Gilks et al., 1994; Ter Braak and Vrugt, 2008) (Sadegh and Vrugt, 2014) algorithms to search the feasible space. The Metropolis ratio is selected to accept/reject the proposed sample, and the $\hat{R}$ informs on the convergence of the chains, which should remain below the critical threshold of 1.2 (Gelman and Shirley, 2011; Cheng et al., 2014b). For a more detailed description of the algorithm, the reader is referred to Sadegh et al. (2017).

1.2.6 Model Diagnostics and Selection

The purpose of fitting a statistical model, whether it is stationary or nonstationary, is to characterize the population from which the data was drawn for further analysis/inference (Coles, 2001). Hence, it is necessary to check the performance of the fitted model to the data (Coles, 2001). We implemented different matrices in the ProNEVA for goodness of fit (GOF) assessment and model selection including: quantile and probability plots for a graphical assessment, two-sample Kolmogorov-Smirnov (KS) test, Aikaike Information Content (AIC), Baysian Information Critiria (BIC), Maximum Likelihhod (ML), Root Mean Square Error (RMSE), and Nash-Sutcliff Efficiency (NSE) coefficient. The hybrid-

evolution MCMC approach (Sadegh et al., 2017) within the Bayesian framework provides an ensemble of solutions for the (non)stationary statistical model fitted to the data. ProNEVA uses the best set of parameters, $\hat{}$, which maximizes the posterior distribution. Marginal posteriors will then provide uncertainty estimates of the estimated parameters.

1.2.7 Standard Transformation

When applied to nonstationary applications, the homogeneity in the distributional assumption requires an adjustment to the traditional GOF techniques (Coles, 2001). Consequently, ProNEVA standardizes the observations based on the underlying distribution family such that the GOF tests can be performed. Table S4 provides information on the transformation methods in ProNEVA. However, it is worth noting that the choice of the reference distribution is arbitrary (Coles, 2001). Here, we selected those transformations that are widely accepted in the literature (Coles, 2001; Koutrouvelis and Canavos, 1999). In the case of a LP3 distribution, the transformation can only be applied when the parameter α is constant (Koutrouvelis and Canavos, 1999). Based on Equation 8, this implies that the transformation can be performed only in the case of constant skewness γ_X.

1.2.8 Probability and Quantile Plots

The probability plot and quantile plot are graphical techniques for evaluating the goodness-of-fit of models. Given an ordered set of n random observations $z_{(1)} < \cdots < z_{(n)}$, the empirical estimate of the probability of $z_{(i)}$ is $\bar{F}_i = i/(n+1)$, where $(n+1)$ guarantees $\bar{F}_i \neq 1$ (Coles, 2001). Assuming $\hat{F}$ as the estimated unknown distribution function of the population, the probability plot consists of the points (Coles, 2001)

$$\left\{ \left(\hat{F}(z_{(i)}); \quad \frac{i}{n+1} \right) : \quad i = 1, \cdots, n \right\} \tag{13}$$

Analogously, the quantile plot contains the points (Coles, 2001)

$$\left\{ \hat{F}^{-1}\left(\frac{i}{n+1} \right); \quad z_{(i)} \right) : \quad i = 1, \cdots, n \right\} \tag{14}$$

In both the probability and quantile plots, $\hat{F}$ is a reasonable fit if the points are close along the unit diagonal (Coles, 2001). Moreover, both plots provide the same information, however on a different scale. Indeed, it is important to investigate both scales, because what seems an acceptable fit in one scale could be a poor fit in the other (Coles, 2001).

1.2.9 Kolmogorov-Smirnov Test

The two-sample Kolmogorov-Smirnov (KS) test is a non-parametric hypothesis testing technique which compares two samples, $Z^{(1)}$ and $Z^{(2)}$, to assess whether they belong to the same population (Massey, 1951). Being $F_{Z^{(1)}}(z)$ and $F_{Z^{(2)}}(z)$ the (unknown) statistical distributions of $Z^{(1)}$ and $Z^{(2)}$ respectively, the null-hypothesis H_0 is $F_{Z^{(1)}}(z) = F_{Z^{(2)}}(z)$, against alternatives. The KS test statistics D^* is:

$$D^* = \max_x(|F_{Z^{(1)}}(z)) - F_{Z^{(2)}}(z)|) \tag{15}$$

H_0 is rejected when the p_{value} of the test is equal to or exceeds the selected α-level of significance, e.g., 5%. We implemented the KS test in ProNEVA as one of the methods to test the goodness-of-fit of the model. Specifically, ProNEVA generates 1000 random samples from the fitted statistical distribution or, in the case of a nonstationary analysis, from the reference distribution. Then, the KS test is performed between the random samples and the input (original or transformed) data. Finally, the rejection rate (RR), eq. 16, is provided as a GOF index.

$$RR = \frac{\sum(H_0 \quad rejected)}{1000} \tag{16}$$

1.2.10 Model Selection based on Model Complexity

A Model showing desirable level of performance efficiency with the minimum number of parameters, i.e., a parsimonious model (Serago and Vogel, 2018), is usually preferred over a model with similar performance but more parameters - e.g, a nonstationary with more parameters relative to a simpler stationary model (Serinaldi and Kilsby, 2015; Luke et al., 2017). Consequently, ProNEVA evaluates different GOF matrices (i.e., AIC, BIC), which account for the number of parameters within the numerical model.
The Akaike Information Criterion (AIC) (Akaike, 1974, 1998; Aho et al., 2014) is formulated as follows

$$AIC = 2 \cdot (D - \hat{L}) \tag{17}$$

where D is the number of parameters of the statistical model and $\hat{L}$ is the log-likelihood function evaluated at $\hat{}$. The model associated with a lower AIC is considered a better fit.

The Bayesian Information Content (BIC) (Schwarz, 1978) is defined as

$$BIC = D \cdot ln(N) - 2 \cdot \hat{L} \tag{18}$$

where N is the length of records. Similar to AIC, the model with lower BIC results a better fit.

1.2.11 Model Selection based on Minimum Residual

Root Mean Square Error (RMSE) and Nash-Sutcliff Efficiency (NSE) coefficient are two matrices widely used in hydrology and climatology as GOF measurements (Sadegh et al., 2018). The focus of both is to minimize the residuals. The vector of residual **RES** is defined as

$$\mathbf{RES} = \left(\left(\hat{F}^{-1}\left(\frac{1}{n+1}\right) - z_{(1)}\right), \cdots, \left(\hat{F}^{-1}\left(\frac{i}{n+1}\right) - z_{(i)}\right), \cdots, \left(\hat{F}^{-1}\left(\frac{n}{n+1}\right) - z_{(n)}\right)\right); \tag{19}$$

following the same notation used for defining the quantile plot. Hence,

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} RES_i^2}{n-1}} \tag{20}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n} RES_i^2}{\sum_{i=1}^{n} (z_{(i)} - mean(z))^2} \tag{21}$$

A perfect fit is associated with $RMSE = 0$ and $NSE = 1$, given $RMSE \in [0, \text{inf})$ and $NSE \in [-\text{inf}, 1)$.

1.2.12 Predictive Distribution

The primary objective of a statistical inference is to predict unobserved events (Renard et al., 2013a). EVA, for example, provides the basis for estimating loads for infrastructure design and risk assessment of natural hazards (e.g., floods, extreme rainfall events). Considering a Bayesian viewpoint, the predictive distribution can be written as (Renard et al., 2013a):

$$f(\mathbf{z}|\mathbf{X}) = \int f(\mathbf{z}, \boldsymbol{\theta}|\mathbf{X}) \cdot d\boldsymbol{\theta} = \int f(\mathbf{z}|\boldsymbol{\theta}) \cdot f(\boldsymbol{\theta}|\mathbf{X}) \cdot d\boldsymbol{\theta} \tag{22}$$

where $\mathbf{X}$ is the observed data, $\mathbf{z}$ is a grid at which $f(\mathbf{z}|\mathbf{X})$ will be evaluated, $\boldsymbol{\theta}$ is the vector of parameters, $f(\mathbf{z}|\boldsymbol{\theta})$ is the pdf of the selected distribution (i.e., GEV, GP, LP3), and $f(\boldsymbol{\theta}|\mathbf{X})$ is the posterior distribution function. The predictive distribution function relies on the fitted distribution function over the parameter space, and uses the posterior distribution for uncertainty estimation (Renard et al., 2013a). In practice, eq. 22 often cannot be derived analytically. Therefore, Renard et al. (2013a) suggest to numerically evaluate it using the MCMC-derived ensemble of solutions sampled from the posterior distribution. The probability density of the k_{th}-element of the vector $\mathbf{z}$ is:

$$\hat{f}(z_k|\mathbf{X}) = \frac{1}{N_{sim}} \cdot \sum_{i=1}^{N_{sim}} f(z_k|\theta_i) \tag{23}$$

In the nonstationary case, the predictive pdf is a function of the covariate, since the distribution parameters depend on the covariates. For this reason, ProNEVA provides the predictive pdf for a number of predefined values of the covariates.

1.3 Return Level Curves under Nonstationarity

Given a time series of annual maxima, the Return Level (RL) is defined as the quantile Q_i for which the probability of an annual maximum exceeding the selected quantile is q_i (Cooley, 2013). Under the stationary assumption, the characteristics of the statistical model are constant over time, meaning that the probability q does not change on a yearly basis. The concept of Return Period (RP) is defined as the inverse of the probability of exceedance, $T_i = \frac{1}{q_i}$ in years. For example, assuming T_i=100 year for an event indicates that the event has 0.01 probability of occurrence in each year (Cooley, 2013). Under the stationary assumption, there is a one-to-one relationship between RL and RP (Cooley, 2013). Therefore, the RL curves are defined by the following points:

$$\Big(\Big(T_i; Q_i\Big), \quad T_i > 1yr, \quad i = 1, \cdots\Big) \tag{24}$$

RL curves are traditionally used for defining extreme design loads for infrastructure design and risk assessment of natural hazards. However, in a nonstationary context both RP and RL terms become ambiguous (Cooley, 2013) and numerous studies have attempted to address the issue. For nonstationary analysis, ProNEVA integrates two different proposed concept: the expected waiting time (Salas and Obeysekera, 2014), for default time-covariate only, and the effective RL curves Katz et al. (2002).

1.3.1 Effective Return Level

Katz et al. (2002) proposed the concept of effective design value (or effective return level), which is defined as q-quantile, Q varying as a function of the covariate (i.e, time or physical). Therefore, for a constant value of $RP = 1/q$, where q is the yearly exceedance probability, the effective RL curves is defined by the points

$$\Big(\Big(x_c, Q_q(x_c)\Big), \quad q \in [0, 1]\Big) \tag{25}$$

where x_c is the covariate, and $Q_q(x_c)$ is the q-quantile.

1.3.2 Expected Waiting Time

Wigley (2009) first introduced the concept of waiting time, i.e., the expected waiting time until an event of magnitude Q_i is exceeded, in which the probability of exceedance in

each year,q_i, changes over time. Olsen et al. (1998) and, later, Salas and Obeysekera (2014) provided a comprehensive mathematical description of the suggested concept.

The event Q_{q_0} is defined as the event with the exceedance probability at time $t = 0$ equal to q_0. Under nonstationary conditions, at time $t = 1$ the probability of exceedance of Q_{q_0} will be q_1, at time $t = 2$, it will be q_2, and so on. Given the selected statistical model F_Q with characteristics θ_t, $q_t = 1 - F_Q(Q_{q_0}, \theta_t)$. Hence, the probability of the event to exceed Q_{q_0} at time m is given by (Salas and Obeysekera, 2014):

$$f(m) = q_m \cdot \prod_{t=1}^{m-1}(1 - q_t) \tag{26}$$

where $f(1) = q_1$ and $f(m) = 1$. The cumulative distribution function (cdf) of a geometrical distribution (eq. 26) is:

$$F_X(x) = \sum_{i=1}^{x} f(i) = \sum_{i=1}^{x} q_i \cdot \prod_{t=1}^{i-1}(1 - q_t) = 1 - \prod_{t=1}^{x}(1 - q_t) \tag{27}$$

where x is the time at which the event occurs, $x = 1, \cdots, x_{max}$, $F_X(1) = q_1$, and $F_X(x_{max}) = 1$. Therefore, the expected waiting time (or RP) in which for the first time the occurring event exceeds Q_{q0} can be derived as

$$T = E(X) = \sum_{x=1}^{x_{max}} x \cdot f(x) = \sum_{x=1}^{x_{max}} x \cdot p_x \prod_{t=1}^{x-1}(1 - p_t) \tag{28}$$

Cooley (2013) simplifies eq. (28) as:

$$T = E(X) = 1 + \sum_{x=1}^{x_{max}} \prod_{t=1}^{x-1}(1 - p_t) \tag{29}$$

which gives the return period under nonstationary conditions, and it is consistent with the definition of RP in the stationary case (Salas and Obeysekera, 2014).

1.4 Explanatory Analysis: Mann-Kendall and White Tests

With the intention of providing additional explanatory data analysis, ProNEVA includes two different tests including the Mann-Kendall (MK) monotonic trend test and the White Test (WT) for evaluating homoscedasticity in the record. These tests can be used to decide whether to incorporate a trend function in one or more of the model parameters or not (i.e., deciding whether to use a stationary or nonstationary model). However, these tests are optional and are not an integral part of ProNEVA.

1.4.1 Mann-Kendall

MK trend test is a widely used test for detecting temporal monotonic changes in the data (Mann, 1945; Kendall, 1955) and thus, it has been applied for detecting nonstationarity in time series (Villarini et al., 2009a; Cheng et al., 2014b). MK evaluates the monotonic trend of the time series $\mathbf{X} = \{x_1, \cdots, x_i, \cdots, x_n\}$ based on the Kendall's S-statistics, which is the difference between the numbers of concordant and discordant pairs (Villarini et al., 2009a),

$$S = \sum_{i<j} sign(x_i - x_j) \tag{30}$$

When $\mathbf{X}$ is independently and randomly distributed, $S = 0$ (Villarini et al., 2009a). For large samples, S tends to normality, and so it is possible to test the Null-Hypothesis (H_0) of no monotonic trend ($S = 0$) against the alternative, at the α-level of significance. This test is most useful for temporal nonstationary analysis. When using a process-based covariate, however, ProNEVA replaces the MK test with the zero slope test for a linear regression model between the time series of the data $\mathbf{X}$ and the covariate $\mathbf{X_c}$ (Shumway and Stoffer, 2011). Given $\hat{\beta}_1$ the estimated slope parameter and β_1 the true value:

$$t_{n-2} = \frac{\hat{\beta}_1 - \beta_1}{s_{\beta_1}} = \frac{\hat{\beta_1}}{s_{\beta_1}} \tag{31}$$

where s_{β_1} is the standard error of β_1, n is the number of observations, and t_{n-2} is a t-distribution with n-2 degrees of freedom. H_0: $\beta_1 = 0$ is performed at 5% level of significance.

1.4.2 White Test

Given the regression model:

$$\mathbf{y} = \beta_0 + \beta_1 \cdot \mathbf{x_c} + u \tag{32}$$

the homoscedasitcity assumption requires that $var\{u^2|\mathbf{x_c}\} = \sigma^2$. In the case of heteroscedasticity then, the residuals of a linear regression model will vary with the dependent variable $\mathbf{x_c}$, (Wooldridge, 2002). White (1980) proposed a test, the WT for heteroscedasticity, based on the estimation of u as (Wooldridge, 2002):

$$\mathbf{u^2} = \delta_0 + \delta_1 \cdot \hat{\mathbf{y}} + \delta_2 \cdot \hat{\mathbf{y}}^\mathbf{2} + error \tag{33}$$

where $\hat{\mathbf{y}}$ is the fitted value $\hat{\mathbf{y}} = \beta_0 + \beta_1 \cdot \mathbf{x}$. H_0 is then $\delta_1 = 0$, $\delta_2 = 0$. Being $R^2_{u^2}$ the $R - squared$ for the regression in equation 33, the Lagrange Multiplier (LM) statistic is calculated as:

$$LM = n \cdot R^2_{u^2} \tag{34}$$

which follow a χ^2_2 distribution. The test is performed at a 5% level of significance.

1.5 Results

As previously discussed, the changes in extremes observed over the past years can stem from changes to different physical processes. In order to account for the observed changes, we need statistical tools that are able to incorporate the cause of variability, which can be represented as time-covariate or a physical-based covariate. In the following, we show example applications of ProNEVA under both stationay and nonstationary assumptions including modeling changes induced by different type of covariates (both temporal and process-based changes).

In the first application, we analyze discharge data from Ferson Creek (St. Charles, IL), which has experienced intense urban development over the years. Urbanization has a direct effect on the amount of water discharged at the catchment outlet, since it increases impervious surfaces. For this reason, we use a process-based nonstationary LP3 model for fitting discharge data, in which the covariate is represented by percent of urbanized catchment area. The second application involves temperature maxima data averaged over the Contiguous United States. Many studies have shown that the amount of CO_2 in the atmosphere causes temperatures to increase. For this reason, we fit temperature data to a nonstationary GEV model, in which the covariate is represented by CO_2 emissions in the atmosphere to include the underlying physical relationship. In the third application, we investigate sea level in the city of Trieste (Italy), which has increased over the years. In this case, we adopted a temporal nonstationary GEV model. The last application involves precipitation data for New Orleans, LA in which we fit a stationary GP model, given that there is no evidence of change in statistics of extremes.

1.5.1 Application 1: Modeling discharge with urbanization as the physical driver

Since 1980, Ferson Creek (St. Charles, IL) basin has experienced land use land cover changes due to urbanization. The percent of urban areas within the catchment has increased from 20% in 1980 to almost 65% in 2010. River discharge highly depends on the land use and land cover of the basin as it determines the ratio of infiltration to direct runoff (Figure 2). Here, urbanization can be considered as a known physical process that has altered the runoff in the basin. To incorporate the known physical process, we investigate annual maxima discharge of the Ferson Creek (station USGS 05551200) using a processed-based nonstationary LP3 model, in which the covariate, X_c, is the percent of urbanized area. LP3 is widely used for modeling discharge data (Bulletin 17B, U.S. Water Resources Council (1982)). We select a nonstationary model in which the parameter μ (mean) is an exponential function of the covariate X_c. We adopt normal priors for the

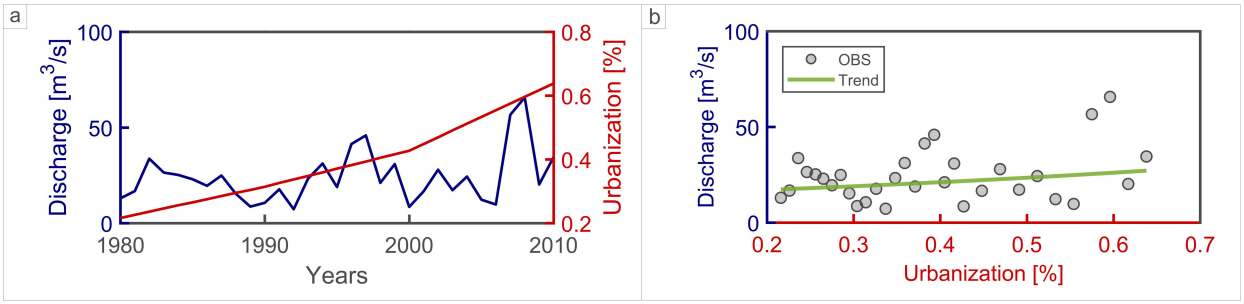

Figure 2: ProNEVA results for Application 1: Modeling discharge in Ferson Creek with urbanization as the physical driver of change. a) Discharge data and percent of urbanization in the basin; b) Discharge data as a function of urbanization.

LP3 parameters. Figure 3.b shows the results of the process-based nonstationary analysis for an arbitrary value of urbanized area, here 37%. For the sake of comparison, Figure 3.a displays the results when a stationary model is implemented. It is worth noting that the nonstationary model (Figure 3.b) fits extreme discharge values (high values of return period) better than the stationary model (Figure 3.a). While based on the AIC and BIC diagnostic tests, the stationary model and the nonstationary model perform rather similarly, the RMSE of the nonstationary model (25.06 m^3/s) is considerably lower than that of the stationary model (77.58 m^3/s).

Urbanization alters the runoff in the basin by reducing the amount of water that infiltrates and increasing the amount of direct runoff. Figure 3.c shows the ability of the statistical model to incorporate this physical process. As anticipated, the expected (ensemble median) nonstationary return level curve associated with a 62% of urbanized area returns higher values of discharge than the one associated with a 37% of urbanized area. For example, under the nonstationary assumption, the magnitude of a 50-year event is 62.47 m^3/s for 37% of urbanized area, similar to the stationary case. However, the magnitude of the 50-year event increases to 78.11 m^3/s (25% more) for 62% of urbanized area. On the contrary, the stationary analysis estimates a 50-year event as an event with magnitude 63.74 m^3/s, independent of the level of urbanization of the catchment. The result demonstrates that a combination between statistical concepts and physical processes is required for correctly estimating the expected magnitude of an event. Figure 3.d displays the effective return level curves (Katz et al. (2002)) which summarize the impact of urbanization on discharge by describing return levels as functions of the selected covariate (x-axis), here urbanization.

1.5.2 Application 2: Modeling temperature with CO_2 as the physical covariate

Over the past decades, many studies have reported higher surface temperature (e.g.: Zhang et al., 2006; Stott et al., 2010; Melillo et al., 2014; Zwiers et al., 2011), mainly due

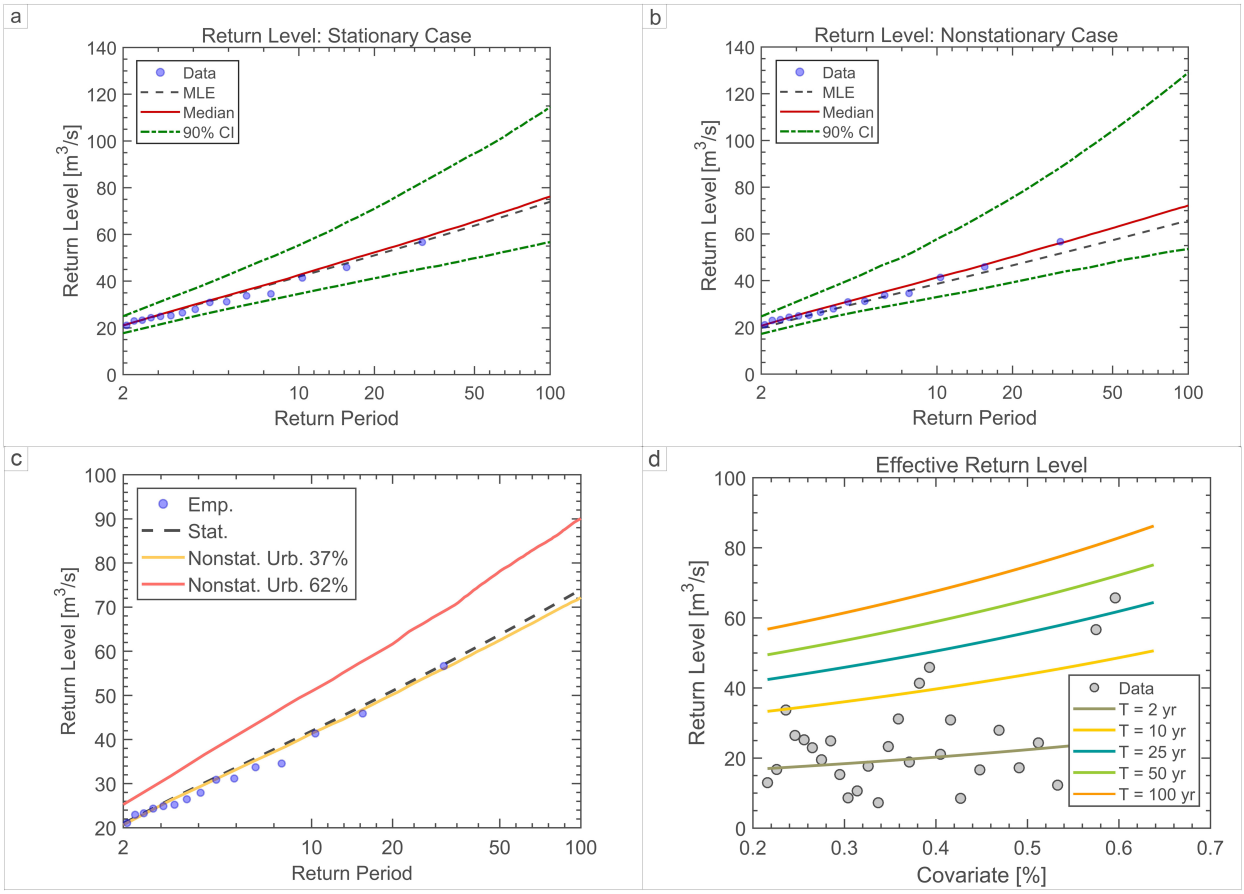

Figure 3: ProNEVA results for Application 1: Modeling discharge in Ferson Creek with urbanization as the physical driver of change. a) Return Level curves based on a stationary model; b) Return Level base on a nonstationary model considering an urbanization area equal to 37% of the catchment area; c) Expected return level curves, i.e. ensemble medians, under stationary and nonstationary assumption; d) Effective return period, i.e. return period as a function of the percent of urbanized area.

to anthropogenic activities as a consequence of increase in greenhouse gasses concentration in the atmosphere. Therefore, we investigate annual maxima surface temperature for the Contiguous United States available from NOAA (NCDC archive - `https://www.ncdc.noaa.gov/cag/national/time-series`) using a process-based nonstationary GEV model in which the user-covariate is represented by CO_2 emissions over the US (Figure 4.a). Territorial fossil fuel CO_2 emissions data are available on Global Carbon Atlas `http://www.globalcarbonatlas.org/en/CO2-emissions` (Boden et al., 2017; BP, 2017; UNFCCC, 2017). To incorporate the observed relationship between temperature and CO_2 in the statistical model (Figure 4.b), we select a model in which the location and the scale parameters of the GEV distribution are linear functions of the covariate, while the shape parameter is constant. We assume normal priors. Figure 5.b shows the results of the non-

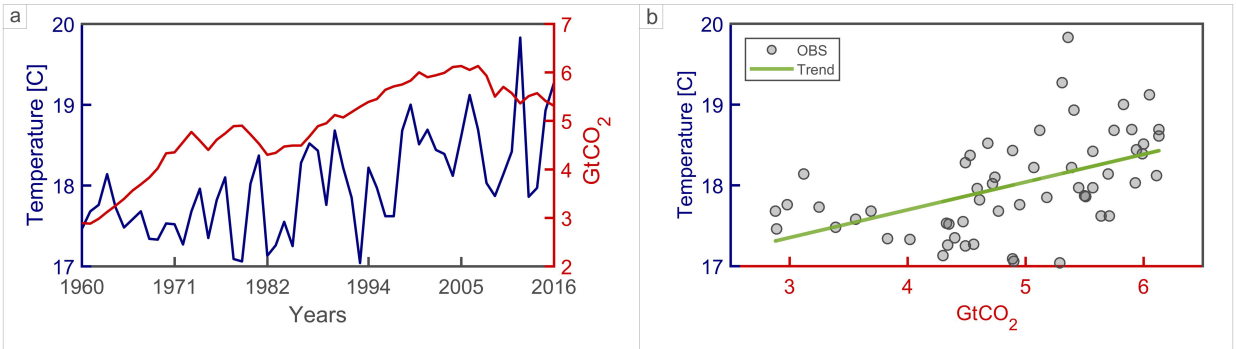

Figure 4: ProNEVA results for Application 2: Modeling temperature maxima with CO_2 emissions as the physical covariate. a) Temperature and CO_2 time series; b) Annual temperature maxima as a function of CO_2 emissions in the atmosphere

stationary model for a value of CO_2 equal to 4.9 $GtCO_2$. For comparison, we also plot the results when a stationary model is selected in Figure 5.a. One can see that the nonstationary model better captures the observed extreme events, particularly events associated with higher values of CO_2. Moreover, the diagnostics tests confirm that the nonstationary model is a better fit. For the nonstationary model, the AIC and the BIC are 93.91 and 104.13, respectively. When the stationary model is considered, both the AIC and BIC increase to 104.98 and 111.11, respectively. Lower values of AIC and BIC indicate a a superior model performance. The advantage of the AIC and BIC for model selection is their ability to account for the number of model parameters: models with higher number of parameters are penalized. Figure S1 shows the effective return level as a function of CO_2 emissions. The results show how temperature extremes change in response to the increasing CO_2 emissions (here, the physical co-variate). For example, looking at the expected magnitude of a 50-year event, the temperature increases of about 4%, from 18.79 °C to 19.5 °C, when the CO_2 emissions increase from 4.49 $GtCO_2$ to 5.51 $GtCO_2$. The results are consistent with the expectation that higher CO_2 leads to a warmer climate, indicating that the statistical nonstationary model is able to model the observed physical relationship between temperature and CO_2.

1.5.3 Application 3: Modeling sea level rise with time as the covariate

The coastal city of Trieste (Italy) has been experiencing an increase in sea level height over the years (Figure S2). Given the observed trend, we investigate annual maxima sea level data from the Permanent Service for Mean Sea Level (PSMSL - station ID 154) by adopting a temporal nonstationary GEV model. The purpose of this example is to show that ProNEVA can also be used for traditional temoral nonstationary analysis. The location and scale parameters of the GEV distribution are modeled as linear functions of the

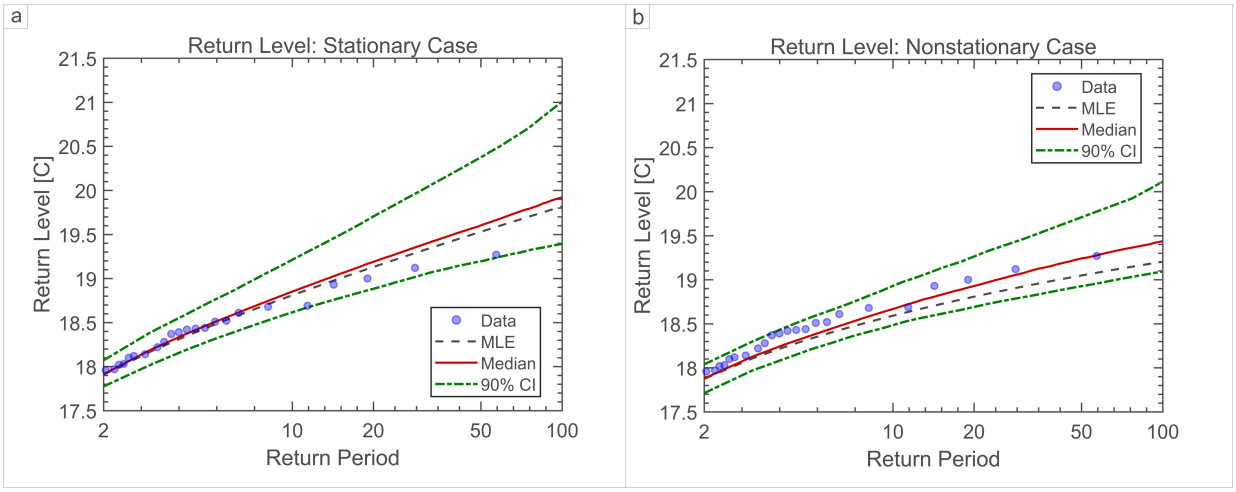

Figure 5: ProNEVA results for Application 2: Modeling temperature maxima with CO_2 emissions as the physical covariate. a) Return Level curves based on a stationary model; b) Return Level base on a nonstationary model considering CO_2 emissions equal to 4.9 $GtCO_2$.

time-covarite. However, as mentioned in methodology section, different alternative are available for users. The shape parameter is kept constant and we use normal priors for parameter estimation. Figure 6.b shows the return level curves for a fixed value of the time-covariate equal to 45 years from the first observation (i.e., 45 years into the future from the beginning of the data). The nonstationary analysis in Figure 6.b provides a better performance that the stationary model in Figure 6.a. Both the AIC and the BIC values confirm that the nonstationary model is the best choice to represent sea level observations in a changing climate. The AIC for the nonstationary model is 976.69, while it is 992.74 for the stationary model. Similarly, the BIC for the nonstationary model is 989.08, while it is 1000 for the stationary model. Lower values for AIC and BIC indicates a better model. The value of the temporal covariate should be regarded as the time at which we estimate expected values of, as in this specific case, sea level. The expected (ensemble median) nonstationary return level curves in Figure 6.c refer to three different time at which we evaluate sea level: 45, 85, and 133 years from the first observation. Here, 133 years from the first observation is beyond the period of observations (88 years) meaning that we project into the future the observed trend and we infer from there. The observed increasing trend in the sea level records results in increasing values of sea level for higher value of the temporal covariate (Figure 6.c). For example, a 50 year event is equal to 7296.3 mm for time equal to 45 years from the first observation, 7349.3 mm for 85 years, and 7410.4 mm for 133 years. We register about 2% increase in sea level when the time of the first observation changes from 45 to 133 years, confirming the ability of the nonstationary model to reproduce the increasing trend in observations. On the contrary, the stationary analysis

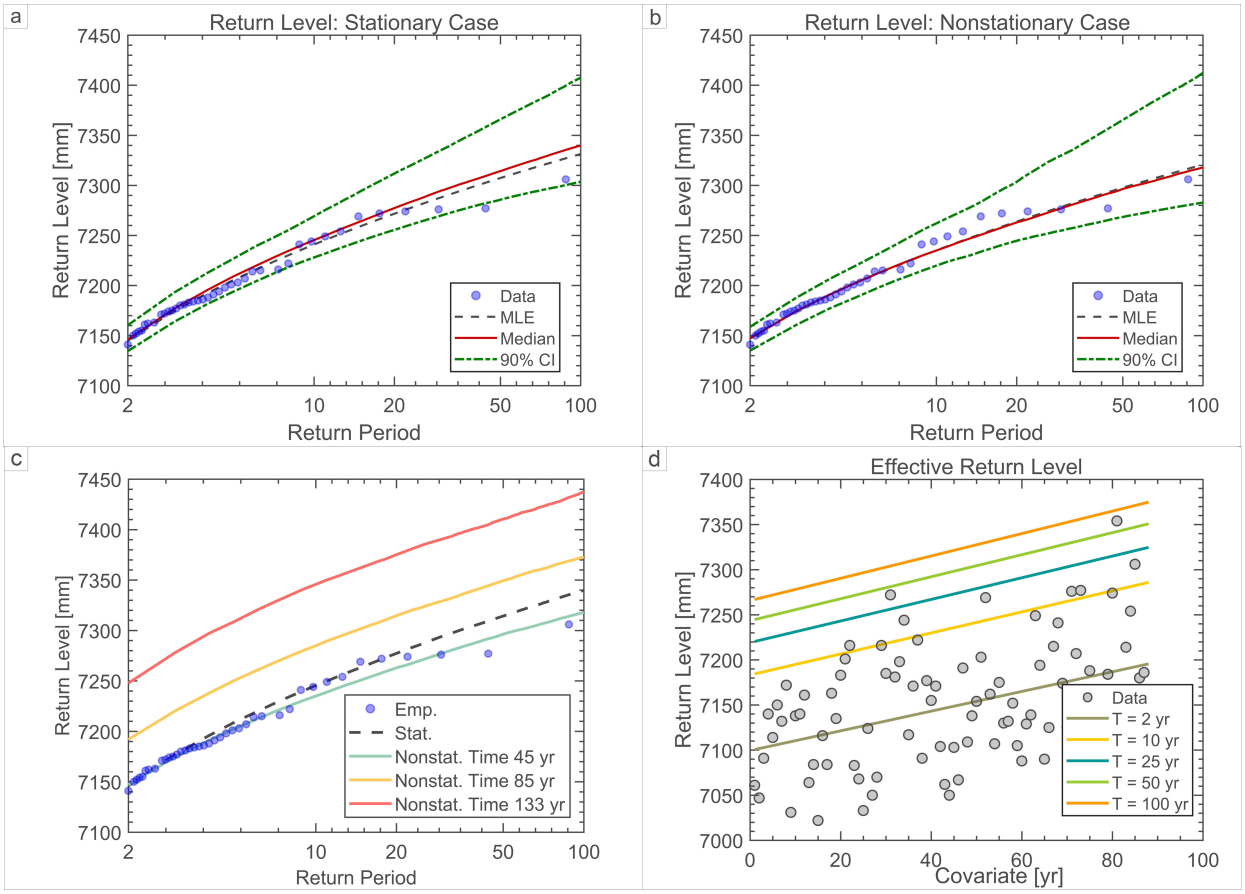

Figure 6: ProNEVA results for Application 3: Modeling sea level rise with time as the covariate. a) Return Level curves based on a stationary model; b) Return Level base on a nonstationary model considering equal to 45 years from the first observation; c) Expected return level curves, i.e. ensemble medians, under stationary and nonstationary assumption; d) Effective return period, i.e. return period as a function of the covariate, here time.

return a 50-year sea level equal to 7314.3 mm regardless of the first observation. Figure 6.d shows the effective return level (Katz et al., 2002) curves, which capture the variability over time (here, the covariate) in the observed data. In the case of a nonstationary model with a temporal covariate, it is possible to evaluate the expected waiting time (Wigley, 2009; Olsen et al., 1998; Salas and Obeysekera, 2014), which incorporates the observed changes in the sea level over time in the estimation of return periods. Figure S3 shows that the current return periods (lower x-axis) will change considering the observed non-stationarity (upper x-asis). For example, the 100-year sea level estimated at t_0 (beginning of the simulation) turns into a 40-year event when the observed trend over time in sea level values is taken into account.

1.5.4 Application 4: Modeling precipitation under a stationary assumption

As last example application, we focus on an example based on the Generalized Pareto (GP) concept for peak-over-threshold extreme value analysis. We investigate a time series of precipitation from New Orleans, LA that does not exhibit changes in statistics of extremes. We obtain daily precipitation from the National Climatic Data Center (NCDC) archive (`https://www.ncdc.noaa.gov/cdo-web/`) for the city of New Orleans, station GHCND:USW00012930. Given that we are interested in heavy precipitation events, we use a GP distribution to focus on values above a high threshold (i.e., avoid including non-extreme values). We extract precipitation excesses considering a constant threshold of the 98th-percentile of daily precipitation values (Figure S4). For this application we select a stationary GP model, given that we do not have physical evidence to justify a more complex model. However, for the sake of comparison, we perform a nonstationary analysis considering the scale parameter as a linear function of time. Figure 7.a represents the return level curves based on a stationary model, while Figure 7.b depicts return level curves for a value of the covariate (here time) equal to half of the period of observation. From a comparison between the two models, the stationary model performs better. The stationary model returns values of the AIC and BIC equal to 713.3 and 721.14, respectively. For the nonstationary model the values of the AIC and BIC are slightly higher (715.02 and 726.79, respectively). The results of this example application suggests that when no evidence of changes due to a physical process can be identified, ProNEVA favors the simplest form of model that represents the historical observations.

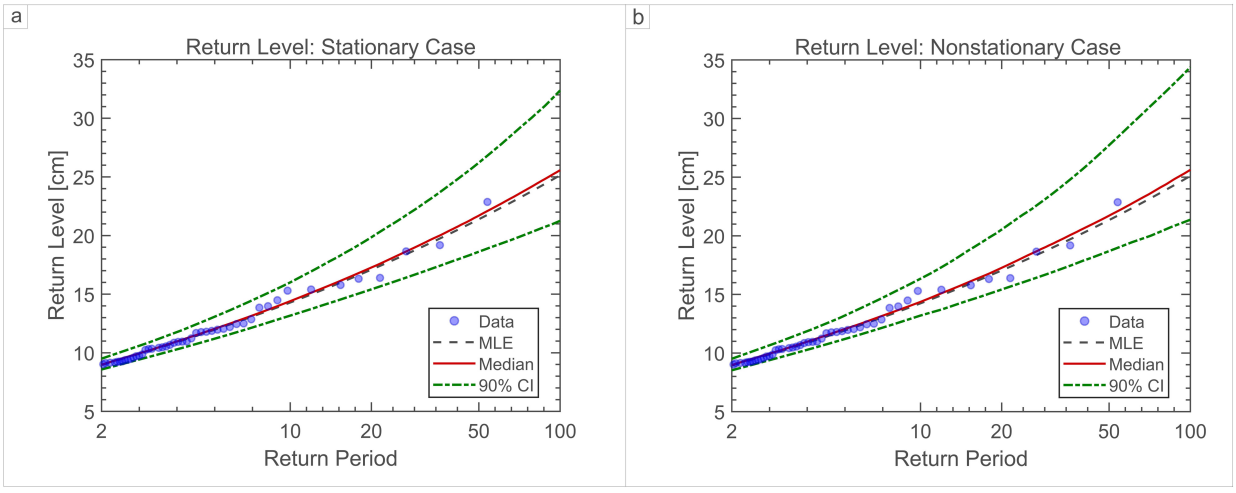

Figure 7: ProNEVA results for Application 4: Modeling precipitation under a stationary assumption. a) Return Level curves under the stationary assumption; b) Return Level curves under the temporal nonstationary assumption for a value of the covariate within the period of observation.

1.6 Conclusion

The ability to reliably estimate the expected magnitude and frequency of extreme events is fundamental for improving design concepts and risk assessment methods. This is particularly important for extreme events that have significant impacts on societies, infrastructure and human lives, such as extreme precipitation events causing flooding and landslides.

The observed increase in extreme events and their impacts reported from around the world has motivated moving away from the so-called stationary approach to ensure capturing the changing properties of extremes (Milly et al., 2008). However, there are opposing opinions and perspective on the need and also form of suitable nonstationary models for extreme value analysis. Most of the existing tools for implementing extreme value analysis under the nonstationary assumption have a number of limitations including lack of a generalized framework for: incorporating physically based covariates; and estimating parameters which depend on a generic physical covariate. To address the above limitations, we propose a generalized framework entitled *Process-based Nonstationary Extreme Value Analysis* (ProNEVA) in which the nonstationarity component is defined by a temporal or process-based dependence of the observed extremes on a physical driver (e.g., change in runoff in response to urbanization, or change in extreme temperatures in response to CO_2 emissions). ProNEVA offers temporal and process-based stationary and nonstationary extreme value analysis, parameter estimation, uncertainty quantification, and a comprehensive assessment of the goodness of fit.

Here we applied ProNEVA to four different types of applications describing change in: extreme river discharge in response to urbanization, extreme sea levels over time, extreme temperatures in response to CO_2 emissions in the atmosphere. We have also demonstrated a peak-over-threshold approach using precipitation data. The results indicate that ProNEVA offers reliable estimates when considering a physical-process or time as a covriate.

The source code of ProNEVA is freely available to the scientific community. A graphical user inter face (GUI) version of the model, Figure 8, is also available to facilitate its applications (see Supporting Information). We hope that ProNEVA motivates more physically-based nonstationary analysis of extreme events.

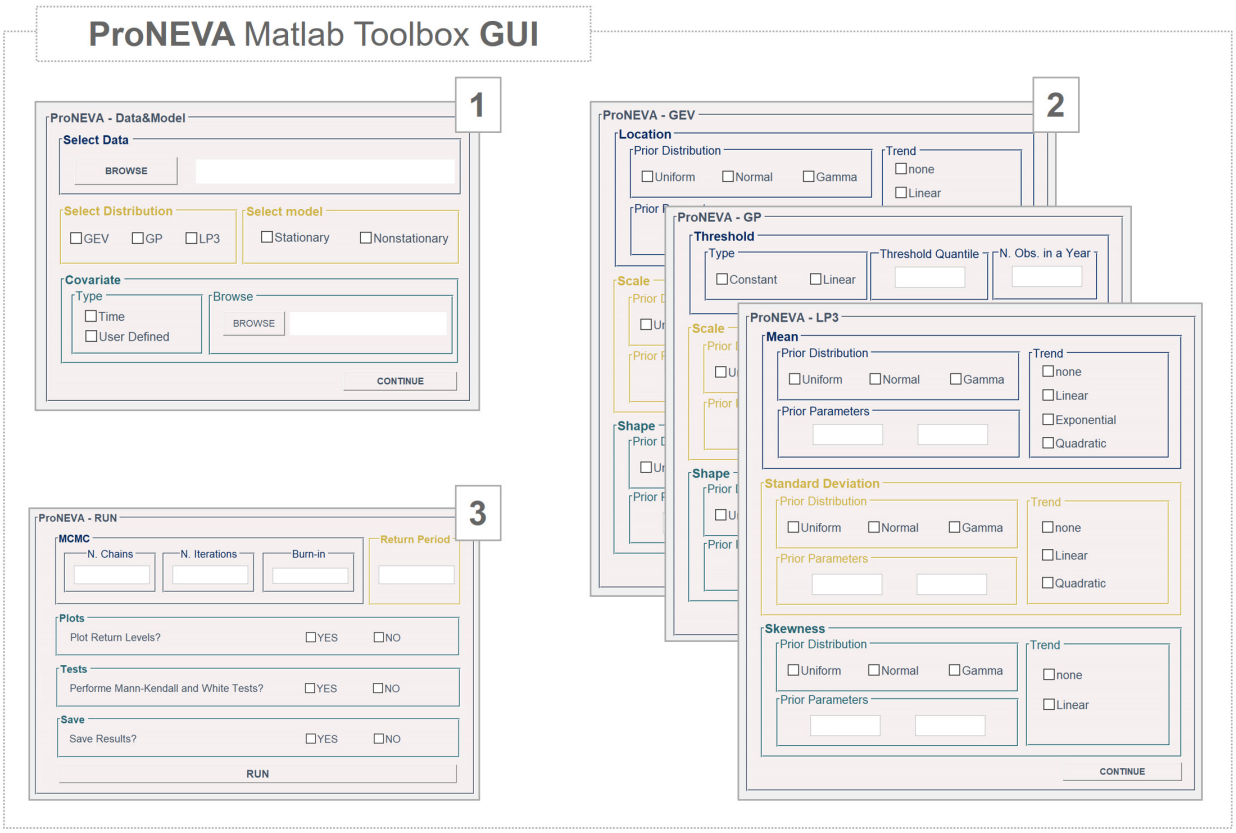

Figure 8: ProNEVA Graphical User Interface (GUI). 1) Interface for uploading data and selecting the choice of distribution (GEV/GP/LP3) and model (stationary/nonstationary) type; 2) Interface specific to the choice of distribution for selecting priors and nonstationarity model; 3) Interface for selecting MCMC information and additional operations (e.g., additional exploratory analyses, saving results saving, plotting options.

2 An Empirical Bayesian-Based Extreme Value Model

Abstract: A new estimation strategy for estimating the parameters of the Heffernan and Tawn conditional extreme value model is proposed. The technique makes use of empirical Bayes estimation for the conditional likelihood that otherwise does not have a simple closed-form expression. The approach is tested on simulations from different types of extreme dependence (and independence) structures, as well as for two real data cases consisting of precipitation analysis conditional on extreme temperature in Boulder, Colorado and Los Angeles, California, U.S.A. The strategy generally has good coverage when informative priors are used for one of the parameters, except for the independence case where the coverage is low until the sample size reaches about 50. Results for the precipitation and temperature data are found to be consistent with the semi non-parametric strategy. The presented model can be potentially applied in a wide variety of science fields, especially in earth, environment, and climate sciences.

2.1 Introduction

Heffernan and Tawn (2004) introduced an important new methodology for modeling multivariate extreme values through a conditional distribution framework that has certain advantages over the usual multivariate extreme value analysis techniques. For example, it is based on threshold exceedances, rather than block maxima, thereby allowing more of the available data to be used for estimation. Moreover, it enables one to model cases whether they are asymptotically dependent or independent using one unified framework.

A drawback to the approach concerns the semi non-parametric estimation procedure for the model as no simple, closed-form distribution exists, in general, without assuming a specific dependence structure for the conditional distribution. The initially proposed approach involves maximum likelihood estimation for the marginal distributions fit to data above a high threshold, Gaussian estimation for the conditional means and standard deviations, and pseudo-likelihood estimation for combining the conditional distributions into a multivariate family. A relatively complicated bootstrap procedure is employed to account for all of the levels of uncertainty in the model. While some advances have been made to this procedure, there still lacks a more fluid estimation and inference strategy for the conditional extreme value model.

To circumvent these issues, we propose to couch the entire estimation procedure into a Bayesian framework. To handle the lack of a closed-form likelihood for the conditional model, empirical Bayes techniques are used, which Owen (1988) developed as a robust alternative to classical likelihood approaches or the bootstrap (cf. Owen, 1990; Mengersen et al., 2013). It was demonstrated that, for some categories of statistical models, when

the likelihood function is numerically unavailable or not entirely known, empirical likelihood methods can be used to bypass simulations from the model while converging in the number of observations. Empirical likelihood has been shown in a wide range of situations to have properties analogous to a real likelihood (cf. Li, 1995; Jing, 1995; Chen, 1994; Qin and Lawless, 1994; Hall and La Scala, 1990). Although further investigation of this methodology is needed, it appears to be a valuable approach in distribution-free contexts.

It is found, here, through simulation results that the approach performs fairly well for most types of dependence structures. In some cases, strong prior information is required, but is also readily obtained through some relatively simple techniques. Results are also compared with the estimation strategy described in Heffernan and Tawn (2004) using the Laplace transformation and a constrained likelihood per Keef et al. (2013a), and our proposed strategy is found to yield consistent estimated parameters.

2.2 Model and Estimation Methodology

Before describing the estimation method proposed here, it is useful to give a brief background on the HT model and previous estimation approaches (section 2.3). Further, we give an intermediate method that assumes a known extremal dependence structure and only accounts for the uncertainty in the dependence parameters of the model in section 2.4, which we use for comparison with our main approach, which is described subsequently in section 2.5.

2.3 The conditional extreme value model

The model introduced by Heffernan and Tawn (2004, henceforth, HT) is perhaps most easily summarized by following Heffernan and Resnick (2007). First, let $\mathbf{X}' = X_1', \ldots, X_n'$ and $\mathbf{Y}' = Y_1', \ldots, Y_n'$ represent two series of independent and identically distributed (henceforth, i.i.d.) random variables that are possibly dependent on each other. Let $\mathbf{X} = X_1 = h_X(X_1'), \ldots, X_n = h_X(X_n')$ and $\mathbf{Y} = Y_1 = h_Y(Y_1'), \ldots, Y_n = h_Y(Y_n')$ be suitably transformed random variables so that they have standardized marginal distributions. Heffernan and Tawn (2004) transformed the variables to the Gumbel scale, whereas Keef et al. (2013a) apply a Laplace transformation. Then, assume that there exist normalizing functions $a(Y)$ and $b(Y)$ such that for $y > 0$

$$\Pr\left\{Y - u > y, \frac{X - a(Y)}{b(Y)} \leqslant z | Y > u\right\} \longrightarrow \exp(-y)G(z) \text{ as } u \longrightarrow \infty \tag{35}$$

where G is a non-degenerate distribution. Through examining a wide class of copula dependence models, Heffernan and Tawn (2004) found that the forms for $a(Y)$ and $b(Y)$

fell into the simple class

$$a(Y) = \alpha Y \text{ and } b(Y) = Y^{\beta}. \tag{36}$$

For positively associated variables X and Y, $\alpha \in [0,1]$ and $\beta \in (-\infty, 1)$. They also found a slightly more complicated form for negatively associated X and Y. Keef et al. (2013a) used a Laplace transformation to ensure Equation (36) was valid for either case, which then gives $\alpha \in [-1,1]$. The parameters α and β are interdependent and control the dependence between the variables X and Y. With the Laplace transformation, $\alpha < 0$ implies negative dependence, and $\alpha > 0$ implies positive dependence. Weakly dependent X and Y are possible as $\alpha \to 0$, but it is also possible that strong dependence exists even when $\alpha = 0$ (cf. Heffernan and Tawn, 2004). The parameter β measures the variability of the dependence with highly negative values indicating lower variability.

To briefly describe the originally proposed estimation procedure, let

$$Z = (X - a(Y))/b(Y). \tag{37}$$

Then, from Equations (35), (36) and (37),

$$X_{|Y>u} = \alpha Y + Y^{\beta} Z_{|Y>u} \tag{38}$$

where the subscripts emphasize the dependence on Y's being extreme. The key role in estimating the joint distribution function of X and Y, conditional on $Y > u$ for u large, is to know the parameters α and β and the distribution function G. Equation (38) motivates a means for estimation of the parameters α and β, which is an active area of research (see e.g., Keef et al., 2009a,b; Lamb et al., 2010; Jonathan et al., 2013), and estimation of G is performed through resampling from the empirical distribution function of the "residual" vectors Z in the HT model after achieving reasonable estimates for α and β.

The current estimation method from the HT model is semi-parametric and involves the following steps:

1. Estimate the marginal distribution function's for each variable separately.

2. Transform each variable in order that they each follow a standard marginal distribution function.

3. Estimate the parameters of the parametric model conditional on large values of the conditioning variable.

4. Information about G (e.g., functions such as the mean and variance, etc.) can be simulated using the empirical distribution function of the estimated standardized residuals. Back transformation can be used to put these estimates onto the original scale.

Heffernan and Tawn (2004) suggested using a hybrid, semi-parametric, model for step 1 of the following form that accounts for both the extreme and non-extreme values (Coles and Tawn, 1991). To simplify notation, let $X = X_i$ and $x = x_i$ be single instances of the random variable and its realization, respectively.

$$\widehat{F}_{X'}(x') = \begin{cases} 1 - (1 - \widetilde{F}_{u_{X'}}(x'))(1 + \xi_{X'}(x' - u_{X'})/\sigma_{X'})_+^{-1/\xi_{X'}} & \text{for } x' > u_{X'} \\ \widetilde{F}_{X'}(x') & \text{for } x' \leqslant u_{X'} \end{cases} \tag{39}$$

where $\widetilde{F}_X$ is the empirical distribution function of the X_i values; $\xi_{X'}$ is the (marginal) shape parameter and $\sigma_{X'} > 0$ is the (marginal) scale parameter (the subscript emphasizing they are the parameters associated with the distribution function for $\mathbf{X}'$), of the generalized Pareto (GP) distribution function as a model for the upper tail of the univariate extremal excesses over a high threshold (i.e., $u_{X'}$ in Equation 39). Similarly for $\mathbf{Y}'$.

The Laplace transformation can be used in step 2, which is given by

$$X_i = \begin{cases} \log\left\{2\widehat{F}_{X'}(x_i')\right\} & \text{for } x_i' < \widehat{F}_{X'}^{-1}(\frac{1}{2}) \\ -\log\left\{2(1 - \widehat{F}_{X'}(x_i'))\right\} & \text{for } x_i' \geqslant \widehat{F}_{X'}^{-1}(\frac{1}{2}) \end{cases} \tag{40}$$

where $\widehat{F}_{X'}$ is estimated according to Equation (39) using maximum likelihood estimation for the GP portion and empirical estimation for $\widetilde{F}_{X'}$ (similarly for Y_i').

Heffernan and Tawn (2004) used non-linear least squares estimation in Equation (38) to estimate α and β for each X_i under the working assumption that Z follows a normal distribution function. Obviously, the assumption of a normal distribution function for Z is inappropriate as it implies that $X_{|Y>u}$ is also normally distributed, which generally is not the case. To counteract the inherent estimation bias from this approach, Keef et al. (2013a) imposed a joint constraint on the dependence parameters (α, β) in order to limit the upper quantiles of $X_{|Y>u}$ to be less than or equal to $x_{F_{|Y>u}}$, the value that would be observed under asymptotic dependence. From the estimates in step 3, Heffernan and Tawn (2004) obtain new estimates $\widehat{z}_i = (x_{i|y>u} - \widehat{a}_i(y_i))/\widehat{b}_i(y_i)$ from which simulations from $\widehat{G}$ are obtained.

Keef et al. (2013b) propose an alternative means for estimating α that does not involve Z. First, different q-th quantiles of $X|Y$ are estimated empirically from Equation (38) for different Y values that fall within two different intervals as $Y_i \in (u - \delta_u, u + s_u)$ and $Y_i \in (\nu - \delta_\nu, \nu + s_\nu)$, where the intervals are around the conditional threshold u, and $\nu > u$ within the range of the observation of Y_i. They then take the median of the estimated α values as $\hat{\alpha}$. See Keef et al. (2013b) for a more thorough description of the approach. This fast estimation procedure is used in our main method described in section 2.5 in order to garner prior information on α, and subsequently β.

To incorporate the uncertainty inference at each stage of the estimation procedure, a bootstrap procedure is proposed by Heffernan and Tawn (2004). Although bootstrapping is a reasonable method for inference, it can be computationally expensive for some larger datasets. The strategy proposed here obviates the need to do any further simulations, as the uncertainty information can be directly obtained from the simulated posterior distribution, as well as the prior distribution and likelihood values.

Finally, it is important to note that the conditional HT approach differs from that of incorporating covariates into the parameters of a univariate extreme value distribution function in that a distribution for values of one variate is conditional on only the extreme values of another variable. Therefore, the dependence is on the processes themselves rather than indirectly through distributional parameters (cf. Gilleland et al., 2013; Jonathan et al., 2012).

2.4 Bayesian estimation under a known dependence structure

Following notation, and especially Equation (35) from above, the aim is to find

$$[\mathbf{Z}, Y \mid Y > u] = [\mathbf{Z}][Y|Y > u] \tag{41}$$

where we use the bracket $[\cdot]$ notation to denote a general distribution function.

Assuming $[Y'|Y' > u_{Y'}]$, which can be derived through back-transformation of Y, is GP leaves only the first term in Equation (41) to be determined. Following Heffernan and Tawn (2004), $Z \sim G_Z$ for some distribution function G_Z, which in the known dependence structure case has a (calculable) density $g_Z(z_1, \ldots, z_k)dz_1 \cdots dz_k$. In the examples that follow we will assume $Z \sim N(\cdot, \cdot)$ but other choices are possible.

By Equation (38), $X = \alpha Y + Y^\beta Z$ for large values of Y. Applying rules of transformation, the density of X is,

$$g_X(x) = \left[\frac{1}{Y^\beta}\right] g_Z\left(\frac{x_1 - \alpha Y}{Y^\beta}\right) dx \tag{42}$$

From Equation (40), recall that $\mathbf{X}$ represents a transformed variable and $\mathbf{X}'$ represents the variable on the original scale. we have that

$$\begin{aligned} X &= F^{-1}\left(F_{X'}(x')\right) \\ &= h_X(x') \end{aligned} \tag{43}$$

where $F_{X'}$ is the distribution function for X' with density $f_{X'}$ and F is the common (e.g., Laplace or Gumbel) distribution function with density f and h_X is the associated transformation function. Furthermore, the conditioning variable Y is also a transformed variable

$Y = F^{-1}(F_{Y'}(y')) = h_Y(y')$. Substituting these into Equation (42) and applying differentiation rules, we arrive at the likelihood for observations X' conditional on $Y' > u_{Y'}$,

$$g_{X'}(x'|Y' = y' > u_{Y'}) = g_Z\left(\frac{h_X(x') - \alpha h_Y(y')}{h_Y^{\beta}(y')}\right) \left[\frac{1}{[F^{-1}(F_{Y'}(y'))]^{\beta}} \frac{f_{X'}(x')}{f\left(F^{-1}\left(F_{X'}(x')\right)\right)}\right] dx', \tag{44}$$

which can be evaluated in closed form.

The main drawback of this approach is that G_Z is known and has a calculable density g_Z. As stated above, we will assume G_Z is the Gaussian distribution but, admittedly, this is an incorrect assumption when considering extremes. However, because Heffernan and Tawn (2004) use the Gaussian distribution to obtain initial estimate of α and β, we feel this is a close alternative for comparison. The proposed empirical Bayes estimation procedure (section 2.5) will circumvent the need to assume closed form for G_Z.

2.5 Empirical Bayes estimation

It is desired to obtain estimates $\hat{\alpha}$, $\hat{\beta}$ and $\hat{G}$. However, because G does not have a simple, closed-form expression and we only need to simulate from G, only the estimates $\hat{\alpha}$ and $\hat{\beta}$ are required. That is, $p(\alpha, \beta) = [\alpha, \beta | X, Y > u_Y]$ is sought. However, it is important to glean uncertainty information from every component of the model, including the marginal distributions of the untransformed variables X' and Y'. To do so, we consider X and Z to be two distinct, if highly dependent, random variables, despite that in the HT model one is completely determined by the other. That is, we modify Equation (38) to be (dropping the $|Y > u_Y$ notation)

$$X = \alpha Y + Y^{\beta} Z + \varepsilon, \tag{45}$$

where ε is i.i.d. noise. In fact, because $\hat{G}$ is desired to be found, $\mathbf{Z}$ could be considered as additional unknown parameters to be sought. However, such a scheme would require estimation of many parameters, which would be highly inefficient and unnecessary. Instead, we look on ε as a nuisance term, giving it zero or unity as its prior depending on whether resulting values for X are within its range or not.

Employing the above modification and from Equation (35), and for simplicity of notation, allowing $\vec{\theta}$ to represent all of the marginal distribution parameters for X' and Y',

and $\vec{\eta}$ to represent any hyper parameters, we have that

$$\begin{aligned}\left[\alpha, \beta, \vec{\theta} \mid \mathbf{Z}, \mathbf{X}, \mathbf{Y} > u_Y\right] &\propto \left[\mathbf{Z}, \mathbf{X}, \mathbf{Y} \mid \alpha, \beta, \vec{\theta}, \mathbf{Y} > u_Y\right] \left[\alpha, \beta, \vec{\theta} \mid \vec{\eta}\right] [\vec{\eta}] \\ &= \left[\mathbf{X}, \mathbf{Y} \mid \mathbf{Z}, \alpha, \beta, \vec{\theta}, \mathbf{Y} > u_Y\right] [\mathbf{Z} \mid \alpha, \beta, \mathbf{Y} > u_Y] \\ &\quad \left[\alpha, \beta, \vec{\theta} \mid \vec{\eta}\right] [\vec{\eta}] \\ &= \left[\mathbf{X} \mid \mathbf{Y}, \mathbf{Z}, \alpha, \beta, \vec{\theta}\right] \left[\mathbf{Y} \mid \mathbf{Z}, \alpha, \beta, \vec{\theta}, \mathbf{Y} > u_Y\right] \\ &\quad [\mathbf{Z} \mid \alpha, \beta, \mathbf{Y} > u_Y] \left[\alpha, \beta, \vec{\theta} \mid \vec{\eta}\right] [\vec{\eta}]\,,\end{aligned} \tag{46}$$

where the last line is arrived at by assuming conditional independence and by Equation (35).

The first two terms $[\mathbf{X} \mid \ldots]$ and $[\mathbf{Y} \mid \ldots]$ are simply the marginal distributions defined by Equation 39 and the GP distribution function (recalling that a back transformation involving the conditioning arguments is first required). For simplicity, here, α, β and $\vec{\theta}$ are taken to be independent and $\vec{\eta}$ is taken to be unity as long as the values fall within their support.

In theory, the prior knowledge on parameters does not depend on the observations Y, and should therefore be specified without using observations, but rather by using any external source of knowledge (Renard et al., 2013b). Thus, we suppose that prior information of parameters should not produce much effect on the simulation results. However, we found that in this specific problem, prior information does matter. Therefore, relatively informative priors are preferred here.

To obtain prior information on α, we employ the fast estimation method introduced by Keef et al. (2013b) and briefly described in section 2.3. An initial estimate for β is achieved through a linear regression on the log-transformed Equation (37); namely,

$$\hat{\beta}\log(Y) + \log(Z) = \log(X - \hat{\alpha}(Y)) \text{ given } Y > u \tag{47}$$

with Z some random residual, the informative initial of $\hat{\beta}$ is obtained and, subsequently, we have some knowledge about the prior distributions for α and β.

All that remains is the primary term of interest, $G = [\mathbf{Z} \mid \alpha, \beta, \mathbf{Y} > u_Y]$. Because G has no simple closed-form expression in general, we employ empirical Bayesian estimation for this term. Empirical likelihood provides likelihood ratio statistics for parameters by profiling a nonparametric likelihood; the approach is analoguous to that used for parametric models (Qin and Lawless, 1994). Owen (1990) showed that for d-variate i.i.d. random vectors $\mathbf{Y}$ (each variate as $y_1, ..., y_n$), with an unknown distribution density f, mean μ_θ and variance σ_θ^2, the approach applies to quite general parameters $\theta(f)$, where θ is the parameter associated with f. Rather than defining the likelihood from the density f as usual, the empirical likelihood method starts by defining parameters of interest θ as

functionals of f, for instance as moments of f, and then profiles a nonparametric likelihood (Mengersen et al., 2013). More precisely, given a set of constraints of the form

$$E_f[h(\mathbf{Y}, \theta)] = 0 \tag{48}$$

where the dimension of h sets the number of constraints unequivocally defining θ, the empirical likelihood is defined as

$$L(\theta|\mathbf{y}) = \max_p \prod_{i=1}^{n} p_i \tag{49}$$

for p in the set

$$p \in [0,1]^n, \sum p_i = 1, \sum_i p_i h(y_i, \theta) = 0 \tag{50}$$

where $p_1 ... p_n$ are nonnegative real numbers summing to unity. The validation of the empirical likelihood approximation is also provided by Owen (1988, 1990). He has proved under mild conditions, if θ satisfies Equation (48), then $-2\log(\frac{L(\theta|\mathbf{y})}{n^{-n}}) \rightarrow \chi^2_d$ in distribution when $n \rightarrow \infty$ and note that n^{-n} is the maximum of $L(\theta|\mathbf{y})$.

In general, the basic idea in this approach is to maximize the empirical likelihood (see Equation 49) subject to constraints provided by Equation (48) which reflect the characteristics of the quantity of interest. For instance, in the one-dimensional case when $\theta = E_f[\mathbf{Y}]$, the empirical likelihood in θ is the maximum of the product $(p_1 ... p_n)$ under the constraint $p_1 y_1 + ... + p_n y_n = \theta$. Solving Equation (49) is based on the Newton-Lagrange algorithm and more are derived with details in Mengersen et al. (2013), Qin and Lawless (1994) and Owen (1990, 1988). Because of its ability to conduct a nonparametric inference without knowledge of higher order moments of the distribution while implicitly taking them into consideration (according to Chen and Cui, 2003), when applying to the conditional likelihood estimation in this study, the first,

$$E[\mathbf{Y} - \mu_\theta] = 0 \tag{51}$$

and second,

$$E[(\mathbf{Y} - \mu_\theta)^2 - \sigma^2_\theta] = 0 \tag{52}$$

statistical moments of "residual" vectors Z (in Equation 37) and conditional vectors of $X|Y$ in the HT model are used as sufficient constraints to estimate the empirical likelihood of G. And if Z has the mean vector μ and vector of standard deviation σ, the respective conditional mean and standard deviation vectors of $X|Y$, for $Y > u$, are $\alpha Y + \mu Y^\beta$ and σY^β, respectively (see Keef et al., 2013b). Now we have all the components in the Bayesian framework (see Equation (46)) to derive the empirical Bayes estimation.

To estimate the parameters using Bayesian inference, a large number of realizations

is generated from the parameters' posterior distributions, using the Differential Evolution Markov Chain (DE-MC; ter Braak, 2006). The DE-MC utilizes the genetic algorithm Differential Evolution (DE; Storn and Price, 1997) for global optimization over real parameter space with Markov Chain Monte Carlo (MCMC) approach (ter Braak, 2006; Gilks et al., 1996). The advantages of simplicity, speed of calculation and convergence makes DE-MC favorable over the conventional MCMC (ter Braak, 2006). In this model, for example, five Markov Chains are constructed in parallel, and are allowed to learn from each other by generating candidate draws based on two random parent Markov chains such that the equilibrium distribution is the target posterior distribution (see ter Braak, 2006; Gelman and Shirley, 2011). Meanwhile, the uncertainty of each parameter in Equation 46 is estimated. This approach has been used in extreme value analysis (Cheng and AghaKouchak, 2014b; Renard et al., 2013b).

2.6 Simulation Experiment

In this study, the results of the conditional extreme value analysis (henceforth, conditional EVA) simulated by the proposed empirical Bayes estimation approach are compared with the originally proposed estimation strategy using the R (R Core Team, 2013) package `texmex` developed by Southworth and Heffernan (2010), which allows for the constrained estimation of the dependence parameters with the Laplace transformation on the marginal variables following Keef et al. (2013a).

Ledford and Tawn (1996) identified four classes of extremal dependence. The first class is that of asymptotically dependent distributions. The other three classes comprise distributions with asymptotically independent dependence structures exhibiting positive extremal dependence, near extremal independence and negative extremal dependence for a d-dimensional variable $\mathbf{Y}$. These three classes correspond respectively to joint extremes of $\mathbf{Y}$ occurring more often than, approximately as often as or less often than joint extremes if all components of the variable were independent (see Ledford and Tawn, 1996, for more details). In this study, the focus is to clarify the performance of interpreting dependence structure (see Equation 36) by the empirical Bayes estimation approach for different types of dependence derived in detail in Section 8 of Heffernan and Tawn (2004). We choose the extremal dependent types below to be simulated, which are also described in Keef et al. (2013a).

Independence. Here $\alpha = \beta = 0$ and G factorizes into Laplace distribution functions.

Asymptotic dependence. Here $\alpha = \pm 1$ (with $+1$ indicating positive dependence and -1 implying negative dependence) and $\beta = 0$ and G takes a range of forms.

Asymptotic independence. Variable X is asymptotically independent of variable Y if $-1 < \alpha < 1$.

The simulated data for analyzing these three types are randomly generated from bivariate extreme value distributions using the R (R Core Team, 2013) package `POT` (Ribatet, 2006). Mainly two types of models are used as listed in Table 1. The experiment is particularly designed to see the performance in estimating dependence parameters α and β with different exceedence sizes based on the proposed empirical Bayesian approach. Initially, sample sizes of 1000, 3000, 5000, 10000 are generated and the high threshold of 99% quantile is selected for all cases to keep independence between Z and Y, thus exceedence sizes of 10, 30, 50 and 100 are left to be experimented on (see an example in Figure 10 and Laplace-transformed results in Figure 9). All simulation cases are repeated for 800 trials, to see the percentage of times (over 800 trials) that the true parameter(s) fell within the estimated 95% credible interval (henceforth, 95% CI), and to explore how the sample sizes (i.e., 10, 30, 50 and 100) might affect estimation inference. Ideally, the percentage of fall-in times for the true parameter(s) should be approximately 0.95 when considering a 95% CI. Another aspect of this experiment also compares simultions given vague priors (uniform distributions with wide support and random initials) for dependence parameters with relative informative ones to check the consequential effect from initials and priors.

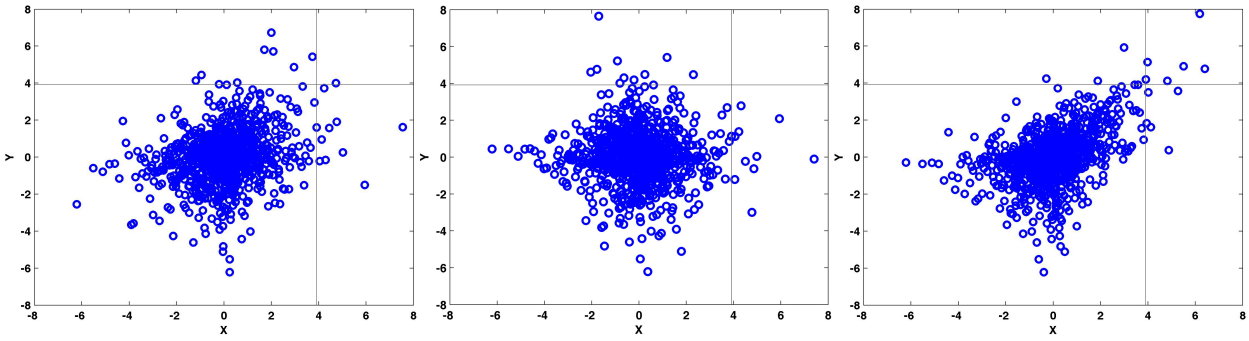

Figure 9: Scatter plot of randomly generated data of sample size 1000. The marginal threshold level, corresponding to the 0.99 quantile is shown in grey lines. Data are shown on Laplace transformed scales. Asymptotic Independence (left), Independence (middle), Asymptotic dependence (right)

The selected three different forms of dependence structures described earlier (Figures 10 and 9) are tested with the proposed empirical Bayes estimation approach. For the simulation experiment, with exceedence sizes of 10, 30, 50 and 100, the percentage of times that the true parameter(s) fell within the estimated 95% CI are shown for parameters α and β individually, as well as for when both parameters fell within the bounds simultaneously in Tables 2 and 3. Table 2 summarizes the simulation results associated with *vague priors* for dependence parameters, while Table 3 displays the results having used *informative priors*. In both tables, it is clear that results improve, if only slightly, with increasing sample sizes, except for the asymptotic dependence case, where the estimation

Table 1: Dependence models for bivariate extreme value distributions used in this study.

Dependence Models	Negative Logistic (nlog)	Logistic (log)
Formula	$V(x,y) = \frac{1}{x} + \frac{1}{y} - (x^{\gamma} + y^{\gamma})^{-\frac{1}{\gamma}}$	$V(x,y) = (x^{-1/\gamma} + y^{-1/\gamma})^{\gamma}$
Pickands' Dependence	$A(\omega) = 1 - [(1-\omega)^{-\gamma} + (\omega)^{-\gamma}]^{-\frac{1}{\gamma}}$	$A : [0,1] \to [0,1]; \omega \mapsto [(1-\omega)^{\frac{1}{\gamma}} + (\omega)^{\frac{1}{\gamma}}]^{\gamma}$
Independence	$\gamma \to 0$	$\gamma = 1$
Total dependence	$\gamma \to +\infty$	$\gamma \to 0$

performs relatively poorly for the parameter, α. This inefficiency may be caused by the fact that it is a special case of estimating a single point in a continuous parameter space (at least for one type of exact dependence).

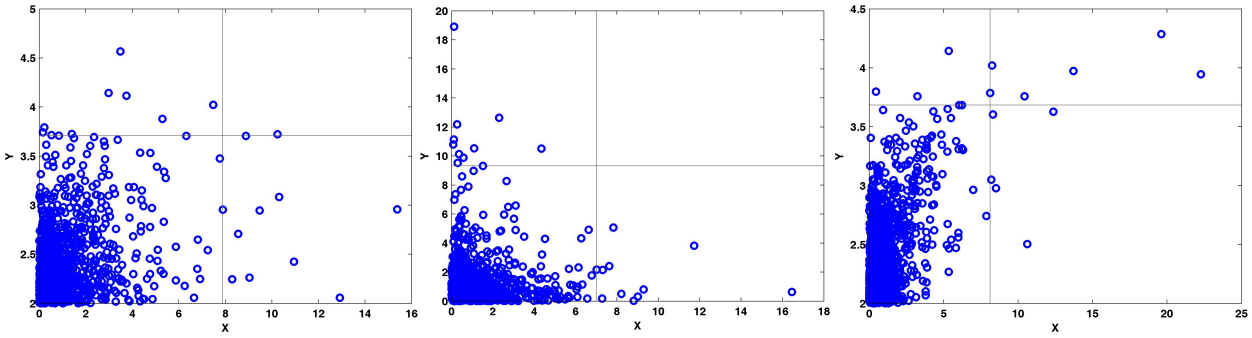

Figure 10: Scatter plot of randomly generated data of sample size 1000. The marginal threshold level, corresponding to the 0.99 quantile is shown in grey lines. Data are shown on original scales. Asymptotic Independence (left), Independence (middle), Asymptotic dependence (right)

By comparing Table 2 and Table 3, with *Informative Priors*, the performance of hit percentage (the true parameter(s) fell within the estimated 95% CI in over 800 trials) for either α or β, or both simultaneously are much better. For example, in the Independence case, with exceedences of 50, Table 3 shows that the individual parameter as well as both simultaneously have the hit percentage over 0.9, while in Table 2 the inference performance is 0.5 on average. That is, over 800 trials, the true parameter(s) fell within the estimated 95% CI with *Informative Priors* over 720 times, while it fell in the interval approximately only 400 times, on average, when using *Vague Priors*. In some cases, such as in the asymptotic independence case with exceedences over 10 (e.g., exceedence size of 50 and 100, then the percentage for both is around 0.914 and 0.915), or in the independence simulation with larger sample size (e.g., same exceedence sizes, the percentage for α is about 0.961 and for

β is approximately 0.958), the percentage even reaches over the ideal situation which is around 0.95. The identifiability issue with β may be because the model at some point is actually $X = \alpha Y + Y^{\beta}(Z - \nu_Z)/\zeta_Z$, where ν_Z and ζ_Z are the mean and standard deviation vectors of Z, respectively, and as noticed, it is not possible to differentiate β from ν_Z and ζ_Z.

Table 2: Results from fitting the conditional extreme value model to simulated data using the empirical Bayes estimation approach with *Vague Priors* proposed here. For each (exceedence) sample size, the percentage (of 800 trials) given is the percentage of times that the *true* parameter(s) fell within the estimated 95% CI's. Results are shown for parameters individually, as well as for when both parameters fell within the bounds simultaneously.

	exceedence size	10	30	50	100
Asymptotic Independence	α	0.796	0.666	0.745	0.701
	β	0.821	0.883	0.938	0.874
	both	0.661	0.588	0.713	0.630
Independence	α	0.800	0.639	0.773	0.695
	β	0.499	0.409	0.396	0.349
	both	0.433	0.301	0.336	0.296
Asymptotic Dependence	α	0.674	0.518	0.585	0.278
	β	0.685	0.674	0.633	0.484
	both	0.439	0.331	0.345	0.135

To compare the empirical Bayes approach to an approach with known dependence structure, Table 4 shows results using informative prior distributions. Comparing Tables 3 and 4, the known dependence structure approach is comparable for the asymptotic independence and independence cases. Clearly, however, for the asymptotic dependence case assuming a known dependence structure is sub-optimal.

We feel that it is reasonable to obtain prior information for β so that this issue is not a major concern. Overall, simulation results show that the proposed approach performs fairly well for most types of dependence structures, but strong and reasonable prior information for β is generally necessary. And the benefit from increasing samples is not so obvious as using *Informative Priors*.

2.7 Temperature and Precipitation test case

In this study, the proposed empirical Bayes estimation approach is further applied on two real data cases, and results are compared with the (slightly modified) HT estimation strategy as implemented by the R (R Core Team, 2013) package `texmex` (Southworth and Heffernan, 2010). Monthly observations of precipitation and temperature from the Climatic

Table 3: Results from fitting the conditional extreme value model to simulated data using the empirical Bayes estimation approach *Informative Priors* proposed here. For each (exceedence) sample size, the percentage (of 800 trials) given is the percentage of times that the *true* parameter(s) fell within the estimated 95% CI's. Results are shown for parameters individually, as well as for when both parameters fell within the bounds simultaneously.

	exceedence size	10	30	50	100
Asymptotic Independence	α	0.941	0.708	0.946	0.923
	β	0.941	0.991	0.960	0.993
	both	0.890	0.704	0.914	0.915
Independence	α	0.965	0.975	0.961	0.940
	β	0.781	0.786	0.934	0.958
	both	0.764	0.768	0.901	0.905
Asymptotic Dependence	α	0.844	0.585	0.741	0.995
	β	0.956	0.929	0.980	0.636
	both	0.823	0.560	0.730	0.634

Research Unit (CRU; New et al., 2000; Mitchell and Jones, 2005) regridded in a common 2 2-degree spatial resolution, are used to provide the test case data. Historical monthly precipitation and temperature records are available from 1901 to 2009. CRU observations have been validated and used in numerous studies of historical climate variability (e.g. Tanarhte et al., 2012; Hao et al., 2013).

To identify extreme conditions, two grid points in the central (Latitude 40.02^o N, Longitude 105.27^o W) and western (Latitude 34.05^o N, Longitude 118.24^o W) United States are selected for conditional extreme dependence structure analysis in Figure 11. The two locations are close to urban areas in Boulder, Colorado and Los Angeles, California where long-term observation stations have been available. In both cases, we consider the precipitation conditional on the temperature's being extreme, i.e., precip $\mid$ temp $> u$, for u large. The marginal threshold level (i.e., u) of temperature at the two locations is corresponding to the 0.97 quantile. For the upper tail of precipitation data, the threshold level for Los Angeles is taken to be the 0.97 quantile, and for Boulder, the 0.99 quantile. The proposed empirical Bayes approach is applied to infer the dependence structure parameters, α describing the dependence strength between precipitation and extremal temperature and β outlining the dependence variability, along with the scale and shape parameters for the upper tail of precipitation and temperature distributions.

Using the CRU precipitation and temperature monthly data (see Figure 11), analyzing precip $\mid$ temp $> u$, for u large, the dependence structure controlling parameters, α and β, and distribution parameters of σ and ξ fitted with the GP distribution function for each variable at two locations (Boulder and Los Angeles) are presented in Table 5. The parame-

Table 4: Results from fitting the conditional extreme value model to simulated data using the Bayesian approach with known dependence structure and *Informative Priors*. For each (exceedence) sample size, the percentage (of 800 trials) given is the percentage of times that the *true* parameter(s) fell within the estimated 95% CI's. Results are shown for parameters individually, as well as for when both parameters fell within the bounds simultaneously.

	exceedence size	10	30	50	100
Asymptotic Independence	α	0.920	0.914	0.923	0.923
	β	0.845	0.871	0.911	0.943
	both	0.801	0.832	0.843	0.877
Independence	α	0.932	0.912	0.926	0.953
	β	0.784	0.793	0.842	0.878
	both	0.729	0.723	0.802	0.882
Asymptotic Dependence	α	0.675	0.698	0.712	0.734
	β	0.687	0.656	0.766	0.865
	both	0.598	0.668	0.700	0.699

ters derived by the empirical Bayes estimation approach and the HT model are simulated with *informative priors* for the α and β. In both locations, the dependence relationship of precipitation given extreme temperature tends to show asymptotic independence, and in Los Angeles, it is near extremal independence, while in Boulder, it is towards negative extremal dependence. Looking at the parameter β, it appears that the variability of the dependence are relatively lower in Los Angeles than that in Boulder. Parameters σ_1 and ξ_1 describe the scale and shape of temperature data which indicates a bounded upper tail distribution, while σ_2 and ξ_2 stand for the precipitation distribution. In Los Angeles, the precipitation distribution shows a heavy tail property (indicated by $\xi_2 > 0$), while in Boulder, it shows a bounded upper tail (see $\xi_2 < 0$). Table 5 also compares results using the empirical Bayes estimation approach and the R (R Core Team, 2013) package `texmex` for the HT method. In general, from the table, we can see that all the parameters, including dependence structure parameters α and β, and GP distribution function parameters σ and ξ, inferred by the two approaches are consistent with each other.

2.8 Summary, conclusions and discussion

The conditional EVA approach introduced by Heffernan and Tawn (2004) is an important new methodology for modeling multivariate extreme values through a conditional distribution framework. Although this approach does not require a priori knowledge of the dependence structure nor that the variables be simultaneously extreme, a difficulty for estimating the parameters is that no simple, closed-form distribution exists in general

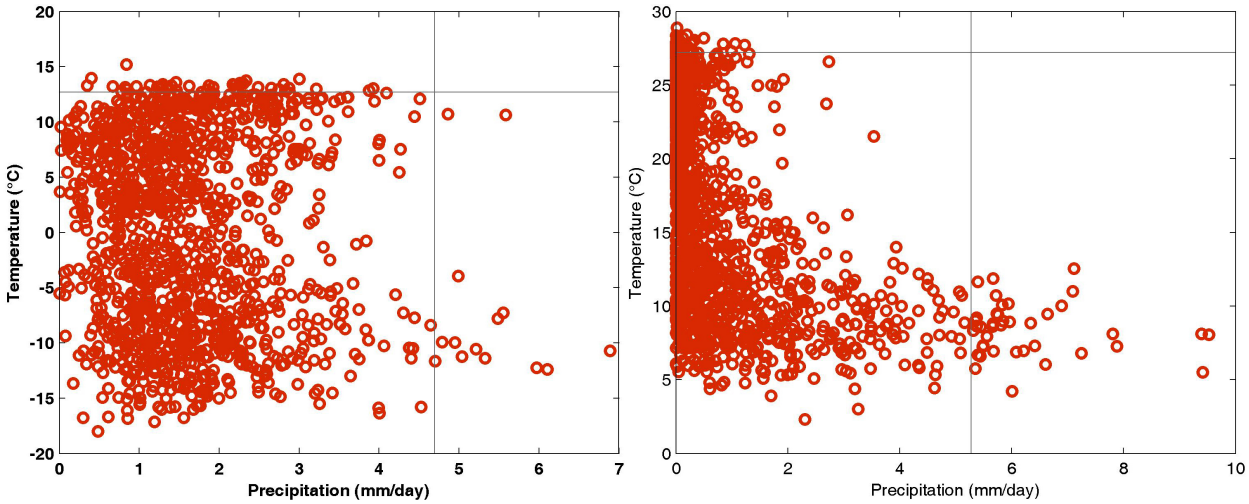

Figure 11: Scatter plot of Precipitation and Temperature data at Boulder, Colorado, U.S.A. (left), and Los Angeles, California, U.S.A. (right). The marginal threshold level for temperature, corresponding to the 0.99 quantile at both locations; quantile of 0.99 at Boulder and quantile of 0.97 at Los Angeles for the precipitation, respectively, are shown in gray lines.

for G. Therefore in the original approach, several estimation methods and constraints are mixed together to evaluate the G distribution function and counteract the inherent bias. This disadvantage motivates the development of the empirical Bayes estimation approach proposed in this study.

Simulations are employed to reproduce known dependence structures with 800 repeated trials for each of three types of dependence and sample sizes, and to test how well the estimation procedure performs. Simulations show generally good coverage of credible intervals. However, the parameter β is relatively hard to infer precisely, and sometimes it is not unique, so strong prior information for β is generally necessary, which is the primary hold-back in this approach and might require further refinement.

The identifiability issue with β may result from ignoring the mean (ν_Z) and standard deviation (ζ_Z) vectors of Z, which are difficult to differentiate from β. To possibly solve this issue, one may include ν_Z and ζ_Z as parameters of interest in the empirical likelihood estimation by imposing other prior knowledge for those parameters, but still, the identifiability problem might not disappear. Another possibility is that one may explore and include higher order moments (other than first and second moments in this study) of the conditional distribution as empirical likelihood constraints, which may also be difficult to identify without any additional assumption because the conditional distribution is generally, numerically unknown.

As for the inefficiency resulting from trying to estimate a single point in a contin-

Table 5: Comparision results from fitting the conditional extreme value model to real data using the empirical Bayes estimation approach and `texmex`.

Los Angeles	exceedence	α	β	σ_1	ξ_1	σ_2	ξ_2
Empirical Bayes	40	-0.065	-0.559	0.733	-0.367	0.820	0.193
	Std. Dev.	0.070	0.214	0.028	0.043	0.044	0.046
`texmex`	40	-0.071	-0.568	0.688	-0.338	0.849	0.140
Boulder	exceedence	α	β	σ_1	ξ_1	σ_2	ξ_2
Empirical Bayes	39	-0.230	-0.270	0.683	-0.143	0.890	-0.251
	Std. Dev.	0.074	0.334	0.033	0.027	0.012	0.060
`texmex`	39	-0.249	-0.284	0.648	-0.149	0.780	-0.143

uous parameter space, a possible extension for the estimation model would be to use a reversible jump Markov chain to include probability terms for those particular cases, namely $\alpha = 0, \beta = 0, \alpha = \pm 1, \beta = 1$, analogously as proposed in the univariate setting by Stephenson and Tawn (2004) for the shape parameter, $\xi = 0$ versus $\xi \neq 0$. Such a scheme might help with the estimation in the asymptotic dependence case and will be explored in a future study.

Additionally, precipitation data conditional on having extreme temperature is also analyzed and compared to the (slightly modified) estimation strategy proposed by Heffernan and Tawn (2004). Resulting parameter estimates are found to be consistent with those from the originally proposed estimation strategy from Heffernan and Tawn (2004).

The presented model can be potentially applied in a wide variety of science fields including finance, earth science, environmental science, and biology. Particularly, this model can be used for assessing spatial climatic extremes (e.g. Gilleland et al., 2013). A myriad of papers show climatic extremes have been changing and are projected to change in the future (e.g. Wehner, 2013; Field et al., 2012; Schubert and Lim, 2013; Easterling et al., 2000; Alexander et al., 2006a; Cheng et al., 2014a). Even concurrent extremes (e.g., joint precipitation and temperature extremes) have been reported to have increased/changed over time (Hao et al., 2013). The proposed methodology allows assessing one extreme variable conditioned on another and hence, we expect it to be a useful tool for conditional extreme value analysis.

References

Aghakouchak, A., D. Feldman, M. J. Stewardson, J.-D. Saphores, S. Grant, and B. Sanders, 2014: Australia?s drought: Lessons for california. *Science*, **343(6178)**, 1430–1431.

Aho, K., D. Derryberry, and T. Peterson, 2014: Model selection for ecologists: the worldviews of AIC and BIC. *Ecology*, **95(3)**, 631–636.

Akaike, H., 1974: A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19(6)**, 716–723.

—, 1998: Information Theory and an Extension of the Maximum Likelihood Principle BT - Selected Papers of Hirotugu Akaike. Springer New York, New York, NY, ISBN 978-1-4612-1694-0, pp. 199–213.

Alexander, L., X. Zhang, T. Peterson, J. Caesar, B. Gleason, A. Klein Tank, M. Haylock, D. Collins, B. Trewin, F. Rahimzadeh, A. Tagipour, P. Ambenje, K. Rupa Kumar, J. Revadekar, and G. Griffiths, 2006a: Global observed changes in daily climate extremes of temperature. *Journal of Geophysical Research*, **111**, D05109.

Alexander, L. V., X. Zhang, T. C. Peterson, J. Caesar, B. Gleason, A. M. Klein Tank, M. Haylock, D. Collins, B. Trewin, F. Rahimzadeh, A. Tagipour, K. Rupa Kumar, J. Revadekar, G. Griffiths, L. Vincent, D. B. Stephenson, J. Burn, E. Aguilar, M. Brunet, M. Taylor, M. New, P. Zhai, M. Rusticucci, and J. L. Vazquez-Aguirre, 2006b: Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research Atmospheres*, **111(5)**, 1–22.

Barnett, T. P., K. Hasselmann, M. Chelliah, T. Delworth, G. Hegerl, P. Jones, E. Rasmusson, E. Roeckner, C. Ropelewski, B. Santer, and S. Tett, 1999: Detection and attribution of recent climate change: A status report. *Bulletin of the American Meteorological Society*, **80(12)**, 2631–2659.

Boden, T., G. Marland, and R. Andres, 2017: Global, Regional, and Nation al Fossil Fuel CO2 Emissions. Carbon Dioxide Information Analysis Center. Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., USA. http://cdiac.ess-dive.lbl.gov/trends/emis/meth_reg.html.

BP, 2017: Statistical Review of World Energy. http://www.bp.com/en/global/corporate/energy-economics.html.

Bracken, C., K. D. Holman, B. Rajagopalan, and H. Moradkhani, 2018: A Bayesian Hierarchical Approach to Multivariate Nonstationary Hydrologic Frequency Analysis. *Water Resources Research*, 243–255.

Cannon, A. J., 2010: A flexible nonlinear modelling framework for nonstationary generalized extreme value analysis in hydroclimatology. *Hydrological Processes*, **24(6)**, 673–685.

Chen, S. X., 1994: Comparing empirical likelihood and bootstrap hypothesis tests. *Journal of Multivariate Analysis*, **51(2)**, 277–293.

Chen, S. X. and H. Cui, 2003: An extended empirical likelihood for generalized linear models. *Statistica Sinica*, **13(1)**, 69–82.

Cheng, L. and A. AghaKouchak, 2014a: Nonstationary precipitation Intensity-Duration-Frequency curves for infrastructure design in a changing climate. *Scientific reports*, **4**, 7093.

—, 2014b: Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate. *Scientific Reports*, **4**, 7093, doi: 10.1038/srep07093.

Cheng, L., A. AghaKouchak, E. Gilleland, and R. Katz, 2014a: Non-stationary extreme value analysis in a changing climate. *Climatic Change*, **127(2)**, 353–369, doi: 10.1007/s10584-014-1254-5.

Cheng, L., A. AghaKouchak, E. Gilleland, and R. W. Katz, 2014b: Non-stationary extreme value analysis in a changing climate. *Climatic Change*, **127(2)**, 353–369.

Coles, S. and L. Pericchi, 2003: Antecipating Catastrophes Through Extreme Value Modeling. *Journal of the Royal Statistical Society - Series C - Applied Statistics*, **52(4)**, 405–416.

Coles, S. G., 2001: *An introduction to Statistical Modeling of Extreme Values*. Springer, ISBN 1852334592.

Coles, S. G. and J. A. Tawn, 1991: Modelling extreme multivariate events. *J. R. Statist. Soc. B*, **53**, 377–392.

Cooley, D., 2013: *Return Periods and Return Levels Under Climate Change*, Springer Netherlands, Dordrecht, chap. 4. ISBN 978-94-007-4479-0, pp. 97–114.

Cooley, D., D. Nychka, and P. Naveau, 2007: Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, **102(479)**, 824–840.

Coumou, D. and S. Rahmstorf, 2012: A decade of weather extremes. *Nature Climate Change*, **2(7)**, 491–496.

De Michele, C. and G. Salvadori, 2003: A Generalized Pareto intensity-duration model of storm rainfall exploiting 2-Copulas. *Journal of Geophysical Research Atmospheres*, **108**, 1–11.

Diffenbaugh, N. S., D. L. Swain, and D. Touma, 2015: Anthropogenic warming has increased drought risk in California. *Proceedings of the National Academy of Sciences*, **112(13)**, 3931–3936.

Duan, Q. Y., V. K. Gupta, and S. Sorooshian, 1993: Shuffled complex evolution approach for effective and efficient global minimization. *Journal of Optimization Theory and Applications*, **76(3)**, 501–521.

Easterling, D., G. Meehl, C. Parmesan, S. Changnon, T. Karl, and L. Mearns, 2000: Climate extremes: Observations, modeling, and impacts. *Science*, **289(5487)**, 2068–2074.

Farajzadeh, M., M. Rahimi, G. A. Kamali, and T. Mavrommatis, 2010: Modelling apple tree bud burst time and frost risk in Iran. *Meteorological Applications*, **17(1)**, 45–52.

Field, C. B., V. Barros, T. F. Stocker, D. Qin, D. Dokken, K. Ebi, M. Mastrandrea, K. Mach, G. Plattner, S. Allen, et al., 2012: *Managing the risks of extreme events and disasters to advance climate change adaptation*. Cambridge University Press Cambridge.

Fischer, E. M. and R. Knutti, 2015: Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes. *Nature Clim. Change*, **5(6)**, 560–564.

—, 2016: Observed heavy precipitation increase confirms theory and early models. *Nature Climate Change*, **6(11)**, 986–991.

Gelman, A. and K. Shirley, 2011: Inference from simulations and monitoring convergence. *Handbook of Markov Chain Monte Carlo*, 163–174.

Gençay, R. and F. Selçuk, 2004: Extreme value theory and Value-at-Risk: Relative performance in emerging markets. *International Journal of Forecasting*, **20(2)**, 287–303.

Gilks, W., G. Roberts, and E. George, 1994: Adaptive Direction Sampling. *Journal of the Royal Statistical Society. Series D (The Statistician)*, **43(1)**, 179–189.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter, 1996: Introducing Markov chain Monte Carlo. *Markov chain Monte Carlo in practice*, 1–19.

Gilleland, E., B. G. Brown, and C. M. Ammann, 2013: Spatial extreme value analysis to project extremes of large-scale indicators for severe weather. *Environmetrics*, **24(6)**, 418–432.

Gilleland, E. and R. W. Katz, 2016: extRemes 2.0: An Extreme Value Analysis Package in R. *Journal of Statistical Software*, **72(8)**.

Griffis, V. W., M. Asce, J. R. Stedinger, and M. Asce, 2007: Log-Pearson Type 3 Distribution and Its Application in Flood Frequency Analysis . I : Distribution Characteristics. *Journal of Hydrologic Engineering*, **12(October)**, 482–491.

Griffis, V. W. and J. R. Stedinger, 2007: Incorporating Climate Change and Variability into Bulletin 17B LP3 Model.

Gupta, H. V., T. Wagener, and Y. Liu, 2008: Reconciling theory with observations: elements of a diagnostic approach to model evaluation. *Hydrological Processes*, **22(November 2008)**, 3802–3813.

Gupta, I. D. and V. C. Deshpande, 1994: Application of Log-Pearson Type-3 Distribution for Evaluation of Design Earthquake Magnitude. *Journal of the Institution of Engineers (India), Civil Engineering Division*, **75**, 129–134.

Haario, H., E. Saksman, and J. Tamminen, 1999: Adaptive proposal distribution for random walk Metropolis algorithm. *Computational Statistics*, **14(3)**, 375–395.

—, 2001: An adaptive Metropolis algorithm. *Bernoulli*, **7(2)**, 223–242.

Haigh, I., R. Nicholls, and N. Wells, 2010: Assessing changes in extreme sea levels: Application to the English Channel, 1900-2006. *Continental Shelf Research*, **30(9)**, 1042–1055.

Hall, P. and B. La Scala, 1990: Methodology and algorithms of empirical likelihood. *International Statistical Review/Revue Internationale de Statistique*, 109–127.

Hallegatte, S., C. Green, R. J. Nicholls, and J. Corfee-Morlot, 2013: Future flood losses in major coastal cities. *Nature Climate Change*, **3(9)**, 802–806.

Hao, Z., A. AghaKouchak, and T. Phillips, 2013: Changes in concurrent monthly precipitation and temperature extremes. *Environmental Research Letters*, **8**, 034014, doi:10.1088/1748-9326/8/3/034014.

Heffernan, J. E. and S. I. Resnick, 2007: Limit laws for random vectors with an extreme component. *The Annals of Applied Probability*, **17(2)**, 537–571.

Heffernan, J. E. and J. A. Tawn, 2004: A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **66(3)**, 497–546.

Holgate, S. J., 2007: On the decadal rates of sea level change during the twentieth century. *Geophysical Research Letters*, **34(1)**, 2001–2004.

Holmes, J. D. and W. W. Moriarty, 1999: Application of the generalized Pareto distribution to extreme value analysis in wind engineering. *Journal of Wind Engineering and Industrial Aerodynamics*, **83(1)**, 1–10.

Huard, D., A. Mailhot, and S. Duchesne, 2009: Bayesian estimation of intensity-duration-frequency curves and of the return period associated to a given rainfall event. *Stochastic Environmental Research and Risk Assessment*, **24(3)**, 337–347.

Hurkmans, R. T. W. L., W. Terink, R. Uijlenhoet, E. J. Moors, P. A. Troch, and P. H. Verburg, 2009: Effects of land use changes on streamflow generation in the Rhine basin. *Water Resources Research*, **45(6)**, 1–15.

Jing, B.-Y., 1995: Two-sample empirical likelihood method. *Statistics & probability letters*, **24(4)**, 315–319.

Jonathan, P., K. Ewans, and J. Flynn, 2012: Joint modelling of vertical profiles of large ocean currents. *Ocean Engineering*, **42**, 195–204.

Jonathan, P., K. Ewans, and D. Randell, 2013: Joint modelling of extreme ocean environments incorporating covariate effects. *Coastal Engineering*, **79**, 22–31.

Jongman, B., S. Hochrainer-Stigler, L. Feyen, J. C. J. H. Aerts, R. Mechler, W. J. W. Botzen, L. M. Bouwer, G. Pflug, R. Rojas, and P. J. Ward, 2014: Increasing stress on disaster-risk finance due to large floods. *Nature Climate Change*, **4(4)**, 264–268.

Katz, R. W., 2013: Statistical Methods for Nonstationary Extremes. In AghaKouchak, A., D. Easterling, K. Hsu, S. Schubert, and S. Sorooshian, eds., *Extremes in a Changing Climate: Detection, Analysis and Uncertainty*, Springer Netherlands, Dordrecht, ISBN 978-94-007-4479-0, pp. 15–37.

Katz, R. W., M. B. Parlange, and P. Naveau, 2002: Statistics of extremes in hydrology. *Advances in Water Resources*, **25(8-12)**, 1287–1304.

Keef, C., I. Papastathopoulos, and J. A. Tawn, 2013a: Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the heffernan and tawn model. *Journal of Multivariate Analysis*, **115**, 396–404.

Keef, C., C. Svensson, and J. A. Tawn, 2009a: Spatial dependence in extreme river flows and precipitation for Great Britain. *Journal of Hydrology*, **378(3–4)**, 240 – 252.

Keef, C., J. Tawn, and C. Svensson, 2009b: Spatial risk assessment for extreme river flows. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **58(5)**, 601–618.

Keef, C., J. A. Tawn, and R. Lamb, 2013b: Estimating the probability of widespread flood events. *Environmetrics*, **24(1)**, 13–21.

Kendall, M. G., 1955: Rank correlation methods.

Klemeš, V., 1974: The Hurst Phenomenon: A puzzle? *Water Resources Research*, **10(4)**, 675–688.

Koenker, R. and G. J. Bassett, 1978: Regression Quantiles. *Econometrica*, **46(1)**, 33–50.

Koutrouvelis, I. A. and G. C. Canavos, 1999: Estimation in the Pearson type 3 distribution. *Water Resources Research*, **35(9)**, 2693–2704.

Koutsoyiannis, D., 2011: Hurst-Kolmogorov Dynamics and Uncertainty. *Journal of the American Water Resources Association*, **47(3)**, 481–495.

Koutsoyiannis, D. and A. Montanari, 2007: Statistical analysis of hydroclimatic time series: Uncertainty and insights. *Water Resources Research*, **43(5)**, 1–9.

—, 2015: Negligent killing of scientific concepts: the stationary case. *Hydrological Sciences Journal*, **60(7-8)**, 1174–1183.

Krishnaswamy, J., S. Vaidyanathan, B. Rajagopalan, M. Bonell, M. Sankaran, R. S. Bhalla, and S. Badiger, 2015: Non-stationary and non-linear influence of ENSO and Indian Ocean Dipole on the variability of Indian monsoon rainfall and extreme rain events. *Climate Dynamics*, **45(1)**, 175–184.

Kwon, H. H. and U. Lall, 2016: A copula-based nonstationary frequency analysis for the 20122015 drought in California. *Water Resources Research*, **52(7)**, 5662–5675.

Kyselý, J., J. Picek, and R. Beranová, 2010: Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold. *Global and Planetary Change*, **72(1-2)**, 55–68.

Lamb, R., C. Keef, J. Tawn, S. Laeger, I. Meadowcroft, S. Surendran, P. Dunning, and C. Batstone, 2010: A new method to assess the risk of local and widespread flooding on rivers and coasts. *Journal of Flood Risk Management*, **3(4)**, 323–336.

Ledford, A. W. and J. A. Tawn, 1996: Statistics for near independence in multivariate extreme values. *Biometrika*, **83(1)**, 169–187.

Li, G., 1995: Nonparametric likelihood ratio estimation of probabilities for truncated data. *Journal of the American Statistical Association*, **90(431)**, 997–1003.

Lima, C. H. R., H.-H. Kwon, and J.-Y. Kim, 2016: A Bayesian beta distribution model for estimating rainfall IDF curves in a changing climate. *Journal of Hydrology*, **540(Supplement C)**, 744–756.

Lins, H. F. and T. A. Cohn, 2011: Stationarity: Wanted dead or alive? *Journal of the American Water Resources Association*, **47(3)**, 475–480.

Luke, A., J. A. Vrugt, A. AghaKouchak, R. Matthew, and B. F. Sanders, 2017: Predicting nonstationary flood frequencies: Evidence supports an updated stationarity thesis in the United States. *Water Resources Research*, **53(7)**, 5469–5494.

Madsen, H., D. Lawrence, M. Lang, M. Martinkova, and T. Kjeldsen, 2013: A review of applied methods in europe for flood-frequency analysis in a changing environment.

Mailhot, A., S. Duchesne, D. Caya, and G. Talbot, 2007: Assessment of future change in intensity-duration-frequency (IDF) curves for Southern Quebec using the Canadian Regional Climate Model (CRCM). *Journal of Hydrology*, **347(1-2)**, 197–210.

Mallakpour, I. and G. Villarini, 2017: Analysis of changes in the magnitude, frequency, and seasonality of heavy precipitation over the contiguous USA. *Theoretical and Applied Climatology*, **130(1-2)**, 345–363.

Mann, H. B., 1945: Nonparametric Tests Against Trend. *Econometrica*, **13(3)**, 245–259.

Marvel, K. and C. Bonfils, 2013: Identifying external influences on global precipitation. *Proceedings of the National Academy of Sciences*, **110(48)**, 19301–19306.

Massey, F. J. J., 1951: Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, **46(253)**, 68– 78.

Matalas, N. C., 1997: Stochastic hydrology in the context of climate change.

—, 2012: Comment on the Announced Death of Stationarity. *Journal of Water Resources Planning and Management*, **138(4)**, 311–312.

Mazdiyasni, O. and A. AghaKouchak, 2015: Substantial increase in concurrent droughts and heatwaves in the United States. *Proceedings of the National Academy of Sciences*, **112(37)**, 11484–11489.

Mazdiyasni, O., A. Aghakouchak, S. J. Davis, S. Madadgar, A. Mehran, E. Ragno, M. Sadegh, A. Sengupta, S. Ghosh, C. T. Dhanya, and M. Niknejad, 2017: Increasing probability of mortality during Indian heat waves. *Science Advances*, 1–6.

Melillo, J. M., T. T. Richmond, and G. W. Yohe, 2014: *Climate Change Impacts in the United States*. ISBN 9780160923883.

Mengersen, K. L., P. Pudlo, and C. P. Robert, 2013: Bayesian computation via empirical likelihood. *Proceedings of the National Academy of Sciences*, **110(4)**, 1321–1326.

Mentaschi, L., M. Vousdoukas, E. Voukouvalas, L. Sartini, L. Feyen, G. Besio, and L. Alfieri, 2016: The transformed-stationary approach: a generic and simplified methodology for non-stationary extreme value analysis. *Hydrology and Earth System Sciences*, **20(9)**, 3527–3547.

Milly, P., B. Julio, F. Malin, M. Robert, W. Zbigniew, P. Dennis, and J. Ronald, 2008: Stationarity is dead: Whither water management? *Science*, **319(5863)**, 573–574.

Min, S.-K., X. Zhang, F. W. Zwiers, and G. C. Hegerl, 2011: Human contribution to more-intense precipitation extremes. *Nature*, **470(7334)**, 378–381.

Ming, L., F. D. M., and K. Sunyong, 2009: Bridge System Performance Assessment from Structural Health Monitoring: A Case Study. *Journal of Structural Engineering*, **135(6)**, 733–742.

Mirhosseini, G., P. Srivastava, and X. Fang, 2014: Developing Rainfall Intensity-Duration-Frequency (IDF) Curves for Alabama under Future Climate Scenarios using Artificial Neural Network (ANN). *Journal of Hydrologic Engineering*, **04014022(11)**, 1–10.

Mirhosseini, G., P. Srivastava, and A. Sharifi, 2015: Developing Probability-Based IDF Curves Using Kernel Density Estimator. *Journal of Hydrologic Engineering*, **20(9)**.

Mitchell, T. and P. Jones, 2005: An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology*, **25(6)**, 693–712.

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

Montanari, A. and D. Koutsoyiannis, 2014: Modeling and mitigating natural hazards: Stationary is immortal! *Water resources research*, **50(12)**, 9748–9756.

Montanari, A., G. Young, H. H. Savenije, D. Hughes, T. Wagener, L. L. Ren, D. Koutsoyiannis, C. Cudennec, E. Toth, S. Grimaldi, G. Blöschl, M. Sivapalan, K. Beven, H. Gupta, M. Hipsey, B. Schaefli, B. Arheimer, E. Boegh, S. J. Schymanski, G. Di Baldassarre, B. Yu, P. Hubert, Y. Huang, A. Schumann, D. A. Post, V. Srinivasan, C. Harman, S. Thompson, M. Rogger, A. Viglione, H. McMillan, G. Characklis, Z. Pang, and V. Belyaev, 2013: "Panta Rhei-Everything Flows": Change in hydrology and society-The IAHS Scientific Decade 2013-2022. *Hydrological Sciences Journal*, **58(6)**, 1256–1275.

New, M., M. Hulme, and P. Jones, 2000: Representing twentieth-century space-time climate variability. Part II: Development of 1901-96 monthly grids of terrestrial surface climate. *Journal of Climate*, **13(13)**, 2217–2238.

Obeysekera, J. and J. D. Salas, 2013: Quantifying the Uncertainty of Design Floods Under Non-Stationary Conditions. *Journal of Hydrologic Engineering*, **19(7)**, 1438–1446.

Olsen, J. R., J. H. Lambert, and Y. Y. Haimes, 1998: Risk of extreme events under nonstationary conditions. *Risk Analysis*, **18(4)**, 497–510.

Owen, A., 1990: Empirical likelihood ratio confidence regions. *The Annals of Statistics*, **18(1)**, 90–120.

Owen, A. B., 1988: Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75(2)**, 237–249.

Papalexiou, S. M. and D. Koutsoyiannis, 2013: Battle of extreme value distributions : A global survey on extreme daily rainfall. *Water Resources Research*, **49(1)**, 187–201.

Pisarenko, V. F. and D. Sornette, 2003: Characterization of the Frequency of Extreme Earthquake Events by the Generalized Pareto Distribution. *pure and applied geophysics*, **160(12)**, 2343–2364.

Qin, J. and J. Lawless, 1994: Empirical likelihood and general estimating equations. *The Annals of Statistics*, 300–325.

R Core Team, 2013: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ragno, E., A. AghaKouchak, C. A. Love, L. Cheng, F. Vahedifard, and C. H. R. Lima, 2018: Quantifying Changes in Future Intensity-Duration-Frequency Curves Using Multi-Model Ensemble Simulations. *Water Resources Research*, 1–38.

Read, L. K. and R. M. Vogel, 2015: Reliability, return periods, and risk under nonstationarity. *Water Resources Research*, **51(1)**, 6381–6398.

Renard, B., X. Sun, and M. Lang, 2013a: Bayesian Methods for Non-stationary Extreme Value Analysis. In AghaKouchak, A., D. Easterling, K. Hsu, S. Schubert, and S. Sorooshian, eds., *Extremes in a Changing Climate: Detection, Analysis and Uncertainty*, Springer Netherlands, Dordrecht, ISBN 978-94-007-4479-0, pp. 39–95.

—, 2013b: Bayesian Methods for Non-stationary Extreme Value Analysis. In *Extremes in a Changing Climate*, Springer, doi: 10.1007/978-94-007-4479-0 3.

Ribatet, M. A., 2006: *A User ' s Guide to the POT Package (Version 1.0)*.

Roberts, G. O. and J. S. Rosenthal, 2009: Examples of Adaptive MCMC. *Journal of Computational and Graphical Statistics*, **18(2)**, 349–367.

Roberts, G. O. and S. K. Sahu, 1997: Updating Schemes, Correlation Structure, Blocking and Parameterization for the Gibbs Sampler. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **59(2)**, 291–317.

Rosner, A., R. M. Vogel, and P. H. Kirshen, 2014: A risk-based approach to flood management decisions in a nonstationary world. *Water Resources Research*, **50(3)**, 1928 – 1942.

Sadegh, M., H. Moftakhari, H. V. Gupta, E. Ragno, O. Mazdiyasni, B. Sanders, R. Matthew, and A. AghaKouchak, 2018: Multi-hazard scenarios for analysis of compound extreme events. *Geophysical Research Letters*.

Sadegh, M., E. Ragno, and A. Aghakouchak, 2017: Multivariate Copula Analysis Toolbox (MvCAT): Describing dependence and underlying uncertainty using a Bayesian framework. *Water Resources Research*, 1–18.

Sadegh, M. and J. A. Vrugt, 2014: Approximate bayesian computation using markov chain monte carlo simulation: Dream (abc). *Water Resources Research*, **50(8)**, 6767–6787.

Sadegh, M., J. A. Vrugt, C. Xu, and E. Volpi, 2015: The stationarity paradigm revisited: Hypothesis testing using diagnostics, summary metrics, and DREAM(ABC). *Water Resources Research*, **51(11)**, 9207–9231.

Salas, J., J. Obeysekera, and R. Vogel, 2018: Techniques for assessing water infrastructure for nonstationary extreme events: a review. *Hydrological Sciences Journal*, **00(00)**, 02626667.2018.1426858.

Salas, J. D. and J. Obeysekera, 2014: Revisiting the Concepts of Return Period and Risk for Nonstationary Hydrologic Extreme Events. *Journal of Hydrologic Engineering*, **(March)**, 554–568.

Salas, J. D. and R. A. Pielke Sr, 2002: Stochastic characteristics and modeling of hydroclimatic processes. *Handbook of Weather, Climate, and Water*, 585–603.

Sankarasubramanian, A. and U. Lall, 2003: Flood quantiles in a changing climate: Seasonal forecasts and causal relations. *Water Resources Research*, **39(5)**.

Sarhadi, A., D. H. Burn, M. Concepción Ausín, and M. P. Wiper, 2016: Time-varying nonstationary multivariate risk analysis using a dynamic Bayesian copula. *Water Resources Research*, **52(3)**, 2327–2349.

Sarhadi, A. and E. D. Soulis, 2017: Time-varying extreme rainfall intensity-duration-frequency curves in a changing climate. *Geophysical Research Letters*, 1–10.

Schubert, S. and Y.-K. Lim, 2013: Climate variability andweather extremes: Model-simulated and historical data. In *Extremes in a Changing Climate*, Springer, doi: 10.1007/978-94-007-4479-0 9.

Schwarz, G., 1978: Estimating the Dimension of a Model. *Ann. Statist.*, **6(2)**, 461–464.

Serago, J. M. and R. M. Vogel, 2018: Parsimonious nonstationary flood frequency analysis. *Advances in Water Resources*, **112(November 2017)**, 1–16.

Serinaldi, F. and C. G. Kilsby, 2015: Stationarity is undead: Uncertainty dominates the distribution of extremes. *Advances in Water Resources*, **77**, 17–36.

Shumway, R. H. and D. S. Stoffer, 2011: *Time Series Analysis and Its Applications With R Examples*. Springer Science+Business Media, Inc, New York, 3rd ed., ISBN 9781441978646.

Southworth, H. and J. Heffernan, 2010: texmex: Threshold exceedences and multivariate extremes. *R package version*, **1**.

Stahl, K., H. Hisdal, J. Hannaford, L. M. Tallaksen, H. A. Van Lanen, E. Sauquet, S. Demuth, M. Fendekova, and J. Jodar, 2010: Streamflow trends in Europe: Evidence from a dataset of near-natural catchments. *Hydrology and Earth System Sciences*, **14(12)**, 2367–2382.

Stephenson, A. and J. Tawn, 2004: Bayesian inference for extremes: accounting for the three extremal types. *Extremes*, **7(4)**, 291–307.

Storn, R. and K. Price, 1997: Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, **11(4)**, 341–359.

Stott, P. A., N. P. Gillett, G. C. Hegerl, D. J. Karoly, D. A. Stone, X. Zhang, and F. Zwiers, 2010: Detection and attribution of climate change: a regional perspective. *Wiley Interdisciplinary Reviews: Climate Change*, **1(2)**, 192–211.

Tanarhte, M., P. Hadjinicolaou, and J. Lelieveld, 2012: Intercomparison of temperature and precipitation data sets based on observations in the Mediterranean and the Middle East. *Journal of Geophysical Research*, **117(D12)**, D12102.

ter Braak, C., 2006: A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces. *Statistics and Computing*, **16(3)**, 239–249.

Ter Braak, C. J. F. and J. A. Vrugt, 2008: Differential Evolution Markov Chain with snooker updater and fewer chains. *Statistics and Computing*, **18(4)**, 435–446.

Thiemann, M., M. Trosset, H. Gupta, and S. Sorooshian, 2001: Bayesian recursive parameter estimation for hydrologic models Water Resources Research Volume 37, Issue 10. *Water Resources Research*, **37(10)**, 2521–2535.

UNFCCC, 2017: National Inventory Submissions. United Nations Framework Convention on Climate Change.

U.S. Water Resources Council, 1982: Guidelines for Determining Flood Flow Frequency. *Bulletin 17B: Reston, Virginia, Hydrology Subcommittee, Office of Water Data Coordination, U.S. Geological Survey*, 182.

Vahedifard, F., J. D. Robinson, and A. AghaKouchak, 2016: Can Protracted Drought Undermine the Structural Integrity of California's Earthen Levees? *Journal of Geotechnical and Geoenvironmental Engineering*, **142(6)**, 02516001.

Villarini, G., F. Serinaldi, J. A. Smith, and W. F. Krajewski, 2009a: On the stationarity of annual flood peaks in the continental United States during the 20th century. *Water Resources Research*, **45(8)**, 1–17.

Villarini, G., J. A. Smith, and F. Napolitano, 2010: Nonstationary modeling of a long record of rainfall and temperature over Rome. *Advances in Water Resources*, **33(10)**, 1256–1267.

Villarini, G., J. A. Smith, F. Serinaldi, J. Bales, P. D. Bates, and W. F. Krajewski, 2009b: Flood frequency analysis for nonstationary annual peak records in an urban drainage basin. *Advances in Water Resources*, **32(8)**, 1255–1266.

Vogel, R. M., C. Yaindl, and M. Walter, 2011: Nonstationarity: Flood Magnification and Recurrence Reduction Factors in the United States. *JAWRA Journal of the American Water Resources Association*, **47(3)**, 464–474.

Volpi, E., A. Fiori, S. Grimaldi, F. Lombardo, and D. Koutsoyiannis, 2015: One hundred years of return period: Strengths and limitations. *Water Resources Research*, **51(10)**, 8570–8585.

Vrugt, J., C. Ter Braak, C. Diks, B. Robinson, J. Hyman, and D. Hingdon, 2009: Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspace Sampling.

Wahl, T., S. Jain, J. Bender, S. D. Meyers, and M. E. Luther, 2015: Increasing risk of compound flooding from storm surge and rainfall for major US cities. *Nature Clim. Change*, **5(12)**, 1093–1097.

Wahl, T., J. Jensen, T. Frank, and I. D. Haigh, 2011: Improved estimates of mean sea level changes in the German Bight over the last 166 years. *Ocean Dynamics*, **61(5)**, 701–715.

Wehner, M., 2013: Methods of projecting future changes in extremes. In *Extremes in a Changing Climate*, Springer, doi: 10.1007/978-94-007-4479-0 8.

Westra, S., L. V. Alexander, and F. W. Zwiers, 2013: Global Increasing Trends in Annual Maximum Daily Precipitation. *Journal of Climate*, **26(11)**, 3904–3918.

White, H., 1980: A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, **48(4)**, 817–838.

Wigley, T. M., 2009: The effect of changing climate on the frequency of absolute extreme events. *Climatic Change*, **97(1)**, 67–76.

Willems, P., K. Arnbjerg-Nielsen, J. Olsson, and V. T. V. Nguyen, 2012: Climate change impact assessment on urban rainfall extremes and urban drainage: Methods and shortcomings. *Atmospheric Research*, **103**, 106–118.

Wooldridge, J. M., 2002: *Introductory Econometrics*. ISBN 0262232197.

Yan, H., N. Sun, M. Wigmosta, R. Skaggs, Z. Hou, and R. Leung, 2018: Next-Generation Intensity-Duration-Frequency Curves for Hydrologic Design in Snow-Dominated Environments. *Water Resources Research*, 1–16.

Yilmaz, A. G. and B. J. C. Perera, 2014: Extreme Rainfall Nonstationarity Investigation and Intensity Frequency Duration Relationship. *Journal of Hydrologic Engineering*, **19(6)**, 1160–1172.

Zhang, Q. and H. Li, 2007: Moea/d: A multiobjective evolutionary algorithm based on decomposition. *Evolutionary Computation, IEEE Transactions on*, **11(6)**, 712–731.

Zhang, X., F. W. Zwiers, and P. A. Stott, 2006: Multimodel Multisignal Climate Change Detection at Regional Scale. *Journal of Climate*, **19(17)**, 4294–4307.

Zhu, J., M. C. Stone, and W. Forsee, 2012: Analysis of potential impacts of climate change on Intensity - Duration - Frequency (IDF) relationships for six regions in the United States. *Journal of Water and Climate Change*, **3(3)**, 185–196.

Zwiers, F., L. Alexander, G. Hegerl, J. Kossin, T. Knutson, P. Naveau, N. Nicholls, C. SchΣr, S. Seneviratne, X. Zhang, et al., 2011: Community paper on climate extremes challenges in estimating and understanding recent changes in the frequency and intensity of extreme climate and weather events world climate research programme open science conference 24-28 october 2011 denver, co, usa.