



HAL
open science

Enhancing Precipitation Hazard Estimation through Intensity-Duration-Area-Frequency (IDAF) Relationships. Application to a Topographically Complex Area

Abubakar Haruna

► **To cite this version:**

Abubakar Haruna. Enhancing Precipitation Hazard Estimation through Intensity-Duration-Area-Frequency (IDAF) Relationships. Application to a Topographically Complex Area. Meteorology. Université Grenoble Alpes [2020-..], 2024. English. NNT : 2024GRALU002 . tel-04632742v2

HAL Id: tel-04632742

<https://hal.science/tel-04632742v2>

Submitted on 22 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : STEP - Sciences de la Terre de l'Environnement et des Planètes

Spécialité : Sciences de la Terre et de l'Environnement

Unité de recherche : Institut des Géosciences de l'Environnement

Améliorer l'estimation des aléas de précipitations grâce aux relations Intensité-Durée-Aire-Fréquence (IDAF). Application à une zone à la topographie complexe

Enhancing Precipitation Hazard Estimation through Intensity-Duration-Area-Frequency (IDAF) Relationships. Application to a Topographically Complex Area

Présentée par :

Abubakar HARUNA

Direction de thèse :

Juliette BLANCHET

CHARGÉE DE RECHERCHE, Université Grenoble Alpes

Directrice de thèse

ANNE-CATHERINE FAVRE

PROFESSEURE DES UNIVERSITÉS, Université Grenoble Alpes

Co-directrice de thèse

Rapporteurs :

Henning RUST

FULL PROFESSOR, Freie Universität Berlin

Gwladys TOULEMONDE

PROFESSEURE DES UNIVERSITÉS, Université de Montpellier

Thèse soutenue publiquement le **31 janvier 2024**, devant le jury composé de :

Clémentine PRIEUR,

PROFESSEURE DES UNIVERSITÉS, Université Grenoble Alpes

Présidente

Juliette BLANCHET,

CHARGÉE DE RECHERCHE HDR, CNRS délégation Alpes

Directrice de thèse

Anne-Catherine FAVRE,

PROFESSEURE DES UNIVERSITÉS, Grenoble INP

Co-directrice de thèse

Henning RUST,

FULL PROFESSOR, Freie Universität Berlin

Rapporteur

Gwladys TOULEMONDE,

PROFESSEURE DES UNIVERSITÉS, Université de Montpellier

Rapporteuse

Peter MOLNAR,

FULL PROFESSOR, Ecole polytechnique fédérale de Zurich

Examinateur

Sophie FUKUTOME,

DOCTEURE EN SCIENCES, MétéoSuisse

Examinatrice



Acknowledgement

I am profoundly grateful to have reached this milestone in my academic journey, and I owe a debt of gratitude to those whose unwavering support and guidance have been instrumental. First and foremost, I extend my deepest appreciation to my two exceptional supervisors, Juliette BLANCHET and Anne-Catherine FAVRE. Your guidance, wisdom, and unwavering support have been the pillars of my success. Your dedication to nurturing my academic growth and pushing me beyond my limits has left an indelible mark on my professional and personal development.

A sincere thank you to the funders of the Extreme Floods in Switzerland (EXCH) project, the Swiss Federal Office of Energy (FOE), and the Swiss Federal Office for the Environment (FOEN), whose generosity made this research possible. To the other members of the project at the Institute of Geosciences and Environment (IGE), Benoit HINGRAY and Guillaume EVIN for the valuable discussions and contributions, and to my co-PhD colleague Kaltrina MALOKU, a big thank you. You have truly enriched the research process and made this endeavor truly fulfilling.

I also extend my appreciation to the members of the HydroMétéorologie, Climat et Interactions avec les Sociétés (HMCIS) team. To the members of the jury, thank you for being part of the journey. Your insightful feedback, constructive criticism, and valuable suggestions have significantly contributed to the refinement of this work.

I want to express my heartfelt thanks to my parents for their unconditional love, encouragement, and sacrifices. Your belief in me has been my source of strength, and I am forever grateful for the values and resilience you instilled in me.

To my incredible wife, *Hayatee*, you have been my rock and my best friend. Your patience, understanding, and unwavering support during the highs and lows of this journey have meant more to me than words can convey. Your belief in my dreams and your sacrifices have been the driving force behind my success. To my wonderful children, Ahmad and Aisha, your laughter, love, and sacrifice made the challenging moments brighter and the joyous ones even more special. Your patience, even during the most challenging times, has been invaluable. You are my inspiration, and I am grateful to have you as a constant source of motivation.

Lastly, to my siblings, your support and care have been a source of comfort and joy. To those not mentioned but equally significant, thank you for being a part of this incredible journey. Your collective influence has shaped my academic and personal growth, and for that, I am forever grateful.

Résumé

À une époque impactée par des phénomènes météorologiques de plus en plus variables qui peuvent modifier profondément les communautés, l'importance de prédictions fiables des précipitations extrêmes à de multiples échelles n'a jamais été aussi prédominante. Malgré sa portée cruciale, une prédiction précise demeure un défi de taille, en particulier dans les régions montagneuses, qui sont particulièrement exposées aux risques associés aux précipitations extrêmes. Par conséquent, des outils plus robustes sont nécessaires pour une prédiction fiable.

Les relations Intensité-Durée-Aire-Fréquence (IDAF) résument les principales caractéristiques statistiques des précipitations extrêmes. Elles sont utilisées pour la quantification des aléas de précipitations et le développement de systèmes d'alerte précoce. Alors que les relations Intensité-Durée-Fréquence (IDF) pour les précipitations ponctuelles ont été largement étudiées, les relations IDAF, prenant en compte la zone d'accumulation, ont reçu beaucoup moins d'attention et, à notre connaissance, seulement pour les extrêmes. Cette thèse vise à modéliser les relations IDAF des précipitations pour toute la gamme des précipitations non nulles dans des zones à la topographie complexe (avec application en Suisse), où des modèles robustes et flexibles sont nécessaires en raison de la forte variabilité spatio-temporelle des précipitations. La grande originalité de cette thèse est que les relations IDAF sont développées pour toute la gamme des intensités de précipitations non nulles, et pas seulement pour les extrêmes. Les distributions marginales robustes des relations modélisées peuvent être utilisées dans les générateurs stochastiques de précipitation.

Premièrement, nous avons identifié un modèle parcimonieux au sein de la famille EGPD (*extended generalized Pareto*) pour modéliser la distribution des intensités non nulles. Deuxièmement, nous avons développé trois modèles de régionalisation pour améliorer la robustesse et la fiabilité des estimations quotidiennes des précipitations. Les résultats montrent que toutes les méthodes régionales offrent une robustesse et une fiabilité accrues en matière de prédiction par rapport au modèle local (sans régionalisation). La méthode spatiale basée sur des modèles additifs généralisés (GAM) a démontré de meilleurs résultats pour les extrêmes, alors que la méthode sur l'approche de la région d'influence a conduit à de meilleurs résultats dans le centre de la distribution.

Troisièmement, nous avons développé des modèles de relations IDF en utilisant toutes les intensités de précipitations non nulles afin d'inclure efficacement les informations disponibles. Trois approches ont été envisagées. La première est basée sur la propriété d'invariance d'échelle des précipitations, la deuxième est basée sur la formulation générale des IDF et la dernière est purement basée sur les données (*data-driven*), employant des équations déterminées empiriquement pour modéliser les relations IDF. Les meilleurs résultats ont été obtenus par la troisième. Des courbes IDF à l'échelle de bassins versants ont été générées à partir de ce modèle pour une utilisation opérationnelle, et les distributions marginales journalières dérivées des modèles sont destinées à être utilisées dans un générateur stochastique de précipitation.

Enfin, nous avons construit des relations IDAF basées sur une approche *data-driven* en utilisant un produit de réanalyse radar. Le modèle nous a permis de caractériser l'aléa de précipitations surfaciques pour un continuum d'échelles spatio-temporelles. Dans l'ensemble, les résultats ont permis de mieux comprendre les variabilités saisonnières et régionales de l'aléa de précipitation en Suisse. Pour de courtes durées, les niveaux les plus élevés sont observés en été, tandis qu'à l'échelle journalière, les niveaux les plus élevés sont observés en automne, notamment au Tessin, région identifiée comme la plus exposée aux précipitations extrêmes à toutes les échelles.

Mots-Clés: Précipitations extrêmes, Courbes IDF, Relations IDAF, Régionalisation, Échelle spatio-temporelle, Alpes suisses.

Abstract

In an era marked by increasingly volatile weather patterns and their profound impact, reliable predictions of extreme precipitation across multiple scales has never been more challenging. Despite its pivotal significance, accurate prediction remains a formidable challenge, especially in mountainous regions that are particularly susceptible to extreme precipitation hazards. As a result, more robust and efficient tools are needed for reliable prediction.

Intensity-Duration-Area-Frequency (IDAF) relationships summarize the main statistical characteristics of extreme precipitation. They are used for areal rainfall hazard quantification, storm characterization, and early warning system development. While Intensity-Duration-Frequency (IDF) relationships for point precipitation have been extensively studied, IDAF relationships, accounting for the area of accumulation, have received far less attention and to our knowledge only for extremes. This thesis aims to model the IDAF relationships for the whole range of non-zero precipitation in topographically complex areas (with application in Switzerland) where robust and flexible models are required due to the strong spatio-temporal variability of precipitation. The key novelty is that IDAF relationships are developed for the whole range of non-zero precipitation intensities, not just extremes. In addition to its usual application, the marginal distributions from the relationships can be utilized in stochastic weather generators.

Four objectives were identified and carefully addressed. First, we identified a parsimonious three-parameter model within the extended generalized Pareto distribution (EGPD) family to model the distribution of non-zero precipitation intensities. Second, we build regionalization models based on three regionalization approaches to improve the robustness and reliability of daily precipitation estimates. The first relies on a fast algorithm that defines distinct homogeneous regions based on upper tail similarity, the second is based on the region-of-influence (ROI) approach, and the third is a spatial approach based on Generalized Additive Model (GAM). All the regional models offered improved robustness and reliability in prediction compared to the local model (without regionalization). The GAM-based method was better in the upper tail, while the ROI method performed better in the bulk of the distribution.

Third, we developed IDF relationships using all non-zero rainfall intensities for 30 min to 72 hr, making efficient use of available information. Three approaches were considered, the first is based on precipitation scale invariance, the second relies on the general IDF formulation, and the last is purely data-driven, employing empirically determined relationships to model the IDF relationships. The best results were shown by the model based on the data-driven approach. It reproduced the known space and time variability of extreme rainfall across Switzerland, catchment-level IDF curves were generated from it for operational use, and the daily marginal distributions derived from the models are intended to be used in a stochastic weather generator currently developed for operational use.

Finally, we constructed IDAF relationships based on a data-driven approach for 1 to 72 hr and 1 to 1,089 km² at each pixel, utilizing a radar-reanalysis product. The model allowed us to characterize areal precipitation hazards for a continuum of spatio-temporal scales. Overall, the results provided insights into the seasonal and regional patterns of precipitation hazards in Switzerland, highlighting the importance of considering multiple spatio-temporal scales when assessing extreme precipitation hazards. For short durations (e.g. 1 hr), the highest levels are almost exclusively observed in summer, while for the daily scale, the highest levels are observed during autumn, particularly in Ticino, a region identified as the most exposed to extreme precipitation across all scales.

Keywords: Extreme Precipitation, IDF curves, IDAF relationships, Regionalization, Spatio-temporal scale, Swiss Alps.

Contents

Acknowledgement	i
Résumé	ii
Abstract	iii
Table of Contents	iv
List of Figures	vi
List of Tables	xiii
I Introduction	1
I.1 Motivation	2
I.2 Extreme floods in Switzerland (EXCH) project	3
I.3 Thesis objectives	4
I.4 Thesis outline	5
II Study area and Data	7
II.1 Study area	8
II.2 Climatology	8
II.2.1 Temperature	8
II.2.2 Precipitation	9
II.3 Data	11
II.3.1 Point precipitation data	11
II.3.2 Gridded data	12
III Statistical methods and hydro-meteorological relationships	15
III.1 Statistical modeling of extreme precipitation	17
III.1.1 Extreme Value Theory	17
III.1.2 Modeling the whole range of non-zero precipitation	20
III.2 Regionalization of at-site estimates of return level with EGPD	27
III.2.1 RFA based on distinct homogenous regions	27
III.2.2 RFA based on region of influence approach (ROI)	28
III.2.3 RFA based on upper tail behavior	28
III.2.4 Spatial method based on Generalized Additive Model (GAM)	29
III.3 Intensity-Duration-Frequency (IDF) relationships	30
III.3.1 Scaling IDF curves	31
III.3.2 General IDF formulation	33
III.3.3 Data-driven IDF curves	34
III.4 Intensity Duration Area Frequency (IDAF) relationships	34
III.4.1 ARF-based IDAF models	36
III.4.2 Data-driven IDAF models	37

III.5	Evaluation framework	37
III.5.1	Split-sampling cross validation	37
III.5.2	Comparison criteria	38
IV	Modeling precipitation intensities with the EGPD in Switzerland	41
IV.1	Parametric families of the EGPD	42
IV.2	Comparison of the EGPD families using daily data	43
IV.2.1	Comparison criteria	43
IV.2.2	Inference	44
IV.2.3	Results	45
IV.3	Left-censored maximum likelihood estimation	48
IV.4	Modeling extreme daily and hourly precipitation with the EGPD	52
IV.4.1	Daily scale	52
IV.4.2	Hourly scale	53
IV.5	Summary	55
V	Regionalization of daily precipitation estimates in Switzerland	56
V.1	Paper 1: Performance-based comparison of regionalization methods to improve the at-site estimates of daily precipitation	57
V.1.1	Introduction	58
V.1.2	Data and study area	60
V.1.3	Candidate Methods	61
V.1.4	Comparison and evaluation criteria	65
V.1.5	Comparison framework	65
V.1.6	Evaluation criteria	66
V.1.7	Results	68
V.1.8	Conclusions and Discussion	72
V.2	Selection of regionalization method for ungauged sites	75
V.2.1	Parameter regionalization	75
V.2.2	Spatial cross-validation framework	75
V.2.3	Result and discussion	76
V.3	Summary	78
VI	Modeling of Intensity-Duration-Frequency (IDF) Relationships in Switzerland	79
VI.1	Paper 2: Modeling Intensity-Duration-Frequency Curves for the Whole Range of Non-Zero Precipitation: A Comparison of Models	80
VI.1.1	Introduction	81
VI.1.2	Data and area under study	83
VI.1.3	Methodology	84
VI.1.4	Results and discussion	96
VI.1.5	Conclusions	106
VI.2	Catchment IDF curves for hydrological applications in Switzerland	108
VI.2.1	Data	108
VI.2.2	IDF curves	108
VI.3	Summary	112
VII	Modeling of Intensity-Duration-Area-Frequency (IDAF) Relationships in Switzerland	113
VII.1	Evaluation of the gridded datasets	114
VII.1.1	Comparison of the empirical values at daily time scales and above	114
VII.1.2	Comparison of return level estimates	118

VII.1.3	Conclusion and discussion	119
VII.2	Paper 3: Modeling Areal Precipitation Hazard: A Data-driven Approach to Model Intensity-Duration-Area-Frequency Relationships using the Full Range of Non-Zero Precipitation in Switzerland	120
VII.2.1	Introduction	121
VII.2.2	Study area and Data	123
VII.2.3	Methodology	125
VII.2.4	Results and Discussion	132
VII.2.5	Conclusions	140
VII.3	Summary	142
VIII	Conclusions and perspectives	143
VIII.1	Conclusion	144
VIII.2	Perspectives	145
VIII.2.1	Modeling in a non-stationarity context	145
VIII.2.2	Comparing EGPD-IDAF models with those from other distributions	147
VIII.2.3	Regional IDAF model	148
VIII.2.4	Accounting for dependence in the IDF and IDAF relationships	149
A	Appendix	151
A.1	Appendix of Chapter III	152
A.1.1	GPD	152
A.2	Appendix of Chapter VI	153
A.2.1	EGPD parameters vs duration	153
A.2.2	Variables and their meaning	158
A.2.3	Catchment IDF curves	159
A.3	Appendix of Chapter VII	164
A.3.1	Comparison criteria	164
	References	165

List of Figures

II.1	Map of Switzerland, the study area. The background color denotes the elevation (m) above sea level. The red-broken lines demarcate the four climatological zones (northwest, northeast, Alps, and south (Ticino)). Some major lakes are highlighted in light blue. The names and locations of some cities are shown in black, while the location of the Jura mountain is shown in blue.	8
II.2	Hypsometric curve of Switzerland based on 1 km by 1 km elevation grid. . . .	9
II.3	Annual mean temperature ($^{\circ}\text{C}$) for the period 1991–2020. Reprinted with permission from MeteoSwiss (2023)	9
II.4	a) Annual precipitation sums (mm) and b) Seasonal precipitation sums (mm) for the period 1991–2020, computed from RhireD (see Section II.3.2.2 for the description of this dataset).	10
II.5	a) Average of 1 hr seasonal precipitation maxima (mm) computed from CombiPrecip and b) Average of 1 day seasonal precipitation maxima (mm) computed from RhiresD. See Section II.3.2.2 for the description of the two datasets.	10
II.6	Map of Switzerland showing the four datasets available for the study. The blue-colored circles represent the 81 rain-gauge stations at 10 min resolution, the red-colored crosses show the locations of the 1176 daily stations, the yellow background shows the extent of RhiresD, and the dark green-colored line with light green background shows the coverage of CombiPrecip (CPC)	12
III.1	Time series of daily precipitation at a station in Zürich from 2001–2010. All the observations above the threshold, taken as the 98% quantile (green-colored line), constitute the peaks over threshold (POT). The annual maxima are highlighted by the red points. In this case, there are 10 annual maxima and 74 POT.	18
III.2	Illustration of a) the density plot and b) the return level plot for the three special cases of the GEV, <i>i.e.</i> Weibull type ($\xi = -0.2$), Gumbel type ($\xi = 0$) and Fréchet type ($\xi = 0.2$). In each case, the location parameter $\mu = 4$ and the scale parameter $\sigma = 1$	19
III.3	Illustration of the a) density plot and b) return level plot for the three special cases of the GPD, <i>i.e.</i> bounded case ($\xi = -0.2$), exponential case ($\xi = 0$) and unbounded case ($\xi = 0.2$). In each case the scale parameter $\sigma_u = 1$	20
III.4	Illustration of the return level uncertainty on the daily data shown in Figure III.1 with a) GEV and b) GPD. In both plots, the points represent the observed values, the black lines represent the fitted return levels, and the gray dashed lines represent the 95% confidence intervals obtained with the delta method (see Coles, 2001 , for the details of the method)	21
III.5	Illustration of some mixture densities. Figure taken from Scarrot and MacDonald (2012)	22

III.6	Illustration of the EGP of Papastathopoulos and Tawn (2013) based on power law defined as $K(u) = u^\kappa$. The density is zero at the origin when $\kappa > 1$ (blue line) and infinite when $\kappa < 1$ (red line). The case where $\kappa = 1$ corresponds to the density of GPD (black line). Adapted from Gamet and Jalbert (2022).	23
III.7	Illustration of the density of MEV (Marani and Ignaccolo, 2015) for various values w , σ and n	24
III.8	Illustration of the EGP of Naveau et al. (2016) based on power law defined as $K(u) = u^\kappa$. The density is shown for various values of κ . In each case, $\sigma = 1$, and $\xi = 0.5$	26
III.9	Sensitivity of theoretical EGP return levels to the values of the shape parameter ξ . The plot is shown for three values of the shape parameter ($\xi = 0, 0.1, 0.2$) with the scale parameter $\sigma = 5$ and $\kappa = 2$	27
III.10	Illustration of Intensity-Duration-Frequency (IDF) curves.	30
III.11	Illustration of Intensity-Duration-Frequency (IDF) curves with the three approaches. Model A is based on the simple scaling assumption. Model B is based on the general IDF formulation and Model C is based on the data-driven approach.	33
III.12	a) Conceptual illustration of IDAF curves in 3-dimension. IDF curves for $A = 81 \text{ km}^2$ (shown in panel b) are obtained by cutting a plane on the IDAF curves in panel a at $A = 81 \text{ km}^2$ (red-colored broken lines). The IAF curves on panel c) are obtained by cutting a plane at $D = 6 \text{ hr}$ on panel a (blue-colored broken lines).	35
III.13	Conceptual illustration of the three comparison criteria, SPAN, FF, and NRMSE_CV. The green line spans from the left to the right tail, indicating that NRMSE_CV is applied to all the quantiles. FF applies only to the maximum value, and SPAN is applied at extrapolation.	38
III.14	Illustration of the FF criteria. Panel a) Shows the histogram of the N values of FF, and the theoretical uniform density (dash-blue line) while the difference between the two densities is shown in panel b). The difference in area ($diff$) = 0.22 in this case.	39
IV.1	Density function corresponding to the four EGP families. In each case $\sigma = 1$, and $\xi = 0.5$	43
IV.2	Seasonal boxplots of the NRMSE across the 500 stations. The boxplots are colored by model. The lower the score, the better the model.	45
IV.3	Seasonal boxplots of relative bias, rBias (%) across the 500 stations. The boxplots are colored by model. The best model is the one with a score = 0.	46
IV.4	Boxplots of the empirical quantiles across the 500 stations. The boxplots are colored by seasons, and they correspond to the non-exceedance probabilities shown along the x -axis.	46
IV.5	Result of model selection according to AIC across the 500 stations. a) Histogram of the number of times a model is selected; b) Map of Switzerland with colored points showing the locations of the stations. Each color represents the model selected at the location.	47
IV.6	Model performance for the case when no censoring is applied (red), and for the case when a censoring of 2 mm (blue) is applied to all the 500 stations (left). The dashed lines colored black and red correspond to NRMSE = 0.05 and 0.1 respectively.	48

IV.7	Effect of the left-censoring threshold C on the performance of each model (x-axis). The performance is measured with the NRSME (y-axis). The dashed lines colored black and red correspond to $\text{NRMSE} = 0.05$ and 0.1 respectively. Each panel corresponds to a season and the boxplots, across the 500 stations, are colored according to the threshold. C_s corresponds to station-specific thresholds, while $C = 0$ mm means no censoring was applied.	49
IV.8	Effect of the left-censoring threshold C on the performance of EGPD-Power for specific quantiles (x-axis). The performance is measured based on rBias (y-axis). The red lines correspond to rBias of 0% . Each panel corresponds to a season and the boxplots, across the 500 stations, are colored according to the threshold. C_s corresponds to station-specific thresholds, while $C = 0$ mm means that we did not apply any left-censoring.	50
IV.9	Quantile-quantile plots for a station located in Airolo in summer. The points are colored according to the censoring threshold used. For this case, $C_s = 9.2$ mm. Panel A shows the qq-plot for the whole range of data. Panel B zooms within the bulk of the distribution (10 to 100 mm), and panel C zooms on the left tail in order to explore the fit below the chosen censoring threshold, <i>i.e.</i> $C_s = 9.2$ mm. The dash-black line represents the diagonal. For a good performance, the points should fall on the diagonal.	50
IV.10	Seasonal maps of the fitted EGPD parameters, κ , σ , ξ (row 1 to 3 respectively) and 100-year return level (bottom row) of daily precipitation amount.	52
IV.11	Seasonal maps of the fitted EGPD parameters, κ , σ , ξ (row 1 to 3 respectively) and 50-year return level (bottom row) of hourly precipitation intensities.	54
V.1	Description of the data used for the study. Left: Map of Switzerland and the neighborhood showing the location of the 1176 daily stations. The color indicates the length of the series, minimum of 20 years and maximum of 156 years. Right: Bar plot showing the number of stations installed in each country for each decade.	61
V.2	Maps of Switzerland showing the optimal number of clusters identified with the PAM algorithm for each season. For each season, the regions identified are color coded. From top left, going clock wise, DJF (2 clusters), MAM (2 clusters), JJA (3 clusters) and SON (3 clusters).	68
V.3	Properties of the ROI identified for each station. Left: Seasonal maps showing the size of the ROI per station. For each season, the size of the circle is proportional to the ROI size, the smallest is 0 [km], and the largest is 100 [km]. The color of the circle indicates whether a local fit is done, or a regional fit. Right: Histogram of the size of the ROI identified. The red points indicates the average number of neighbors identified for each ROI class.	69
V.4	Criteria applied on the bulk of the distribution for each season. Left: Accuracy of the whole distribution as measured by the NRMSE, each boxplot contains 100 values. Right: CRPS score, each boxplot contain 500 points, 1 per station.	71
V.5	Criteria applied on the upper tail for each season. Left: Robustness of the local EGPD and the four candidate models, as measured by the SPAN criteria. The stability is measured with respect to a 100-year return level estimate. Each boxplot contain 50 values. Right: Reliability in prediction of the maxima as measured by the FF criteria, each boxplot contain 100 values.	72
V.6	Map of Switzerland showing the 100year return level for the four seasons as predicted with ROI_EGPD_Full model.	73

V.7	Left panel: Accuracy of the model in the whole distribution (NRMSE). Right panel: Reliability criteria on the far tail (FF). Each criterion is optimized for a value of 1. Each boxplot contains 40 points corresponding to 20 samplings in a spatial cross-validation framework.	76
V.8	Stability of 100, 1,000, and 10,000-year return level estimates, as measured by SPAN in a spatial cross-validation framework, for the four seasons. SPAN is optimized for a value of 1. Each boxplot contains 20 points, with each point corresponding to 1 simulation run. The SPAN is averaged over a 1 by 1 km grid covering the entire Switzerland.	77
VI.1	Map of Switzerland showing the location of the 81 stations. The color of the points indicates the length of the precipitation data in years. The background color shows the elevation above sea level in meters.	84
VI.2	Flowchart of the framework for the study	85
VI.3	Density of the EGPD for $G(v) = v^k$ for $\sigma = 1$, $\xi = 0.5$ and $\kappa = 0.8, 1$ and 2 (adapted from Naveau et al. (2016)). The density of gamma distribution with both shape and scale = 1.4 is shown in blue color. The right panel zooms on the right tail. Indeed the gamma tail decays much faster while having a similar shape in the bulk of the distribution for the case of $\kappa = 2$. The case where $\kappa = 1$ is the exact GP case.	86
VI.4	Illustration of a) Break in scaling of the σ parameter in winter at a station, Robbia in Graubünden. The points colored black are the estimated σ for each duration separately. The models in blue and red are given in Eq. VI.10 and VI.11 respectively. b) Dependence of ξ on duration in summer at a station in Zurich. The black colored points are the estimated ξ for each duration separately. The lines are the fitted linear models whose equations are given in A.2.1. The broken lines represent the 95% confidence interval for the fitted models. The plots of the complete EGPD parameters with duration at these two stations are also given in Appendix A.2.1.2 for the four seasons.	89
VI.5	Boxplots of $1 - \text{NRMSE}_s(d)$ versus duration for the base EGPD model, <i>i.e.</i> , fitted on data of each duration separately. Each boxplot contains 81 points, with each point corresponding to one station.	96
VI.6	Boxplots of the fitted ξ versus duration obtained with the base model. Each boxplot contains 81 points, with each point corresponding to one station.	97
VI.7	Boxplots of the (1-NRMSE) in calibration. Each boxplot contains 1053 points, each point corresponding to one station and duration.	98
VI.8	AIC results for the four seasons. Each panel shows the percentage of stations (y-axis) over which a model is ranked (x-axis) according to the AIC criteria. The ranks are such that '1' is the best and '10' is the worst.	99
VI.9	Boxplots of the regional cross-validation criteria, NRMSE_CV, FF, and SPAN. For the first two criteria, each boxplot contains 2×130 points, corresponding to one regional score for each of the 13 durations and 10 repetitions of the split sampling. For the SPAN, each boxplot contains 130 points. The optimal value for each criterion is equal to 1.	101
VI.10	Simple-scaling (IDF_{ss}) and data-driven ($\text{IDF}_{DD_{global}}$) curves a) in summer at a station in Zurich (North-east). b) in autumn at a station in Locarno (Ticino area in the south). The curves are for the return periods $T = 2, 5, 10, 40,$ and 100 years. The points are the empirical quantiles corresponding to $T = 2, 5,$ and 10 years. The envelopes represent the 95% confidence bounds obtained by block bootstrap.	102

VI.11	Return level estimates (points) and their 95% CI bounds (lines) for return levels $T = 2, 5, 10, 50,$ and 100 years estimated with the simple-scaling (IDF_{ss}), data-driven ($IDF_{DD_{global}}$) and the base model (no linkage of EGPD parameters with duration). The confidence bounds were obtained using block bootstrap. The top row is for a station in Zurich (northeast) in summer for $d = 1$ hour and 24 hours. The second row is for a station in Locarno (Ticino area in the south) in autumn for $d = 1$ hour and 24 hours. In all cases, the broken lines in red, black, and yellow are the empirical levels for $T = 2, 5,$ and 10 years respectively.	103
VI.12	Map of Switzerland showing the seasonal 100-year return level in mm/hr for $d = 1$ hr. Levels predicted with $IDF_{DD_{global}}$	105
VI.13	Map of Switzerland showing the seasonal 100-year return level in mm/hr for $d = 24$ hr. Levels predicted with $IDF_{DD_{global}}$	105
VI.14	Map of the 24 hydrological catchments. The dashed-black line shows the border of Switzerland. The catchments are arranged according to the river they drain and the number in the black box represents the area of each catchment in km^2	109
VI.15	Seasonal IDF curves for 4 out of the 24 large catchments. The lines are the modeled return levels, while the points are the corresponding empirical return levels. Both are colored according to the return period for $T \in \{2, 5, 10, 30, 100\}$ years.	110
VI.16	100-year return level of summer mean areal precipitation for 24 hr duration.	111
VII.1	Seasonal boxplots of the KGE and its sub-components. Each boxplot contains 71 points, 1 point for each pair of rain gauge and the underlying pixel.	115
VII.2	Seasonal boxplots of the KGE and its subcomponents. Each boxplot contains 71 points, 1 point for each pair of rain gauge and the underlying pixel.	117
VII.3	Seasonal boxplots of relative bias (rBias) of a T -year return level estimate for $T \in \{20, 50, 100\}$ years. Each column corresponds to a season, and each row corresponds to a return period. The boxplots contain 71 points, 1 point for each pair of rain gauge and the underlying pixel.	118
VII.4	Map of Switzerland, the study area. The background color denotes the elevation (meters) above mean sea level. The Radar symbols show the location of the five radars in Switzerland, with their names in the white boxes. The name of some cities is shown in black and the name of some mountains and regions are shown in red. The black-colored square shows exemplarily the maximum extent of the rectangular window used for data aggregation, <i>i.e.</i> $1089 km^2$	123
VII.5	a) Conceptual illustration of IDAF curves in 3-dimension. IDF curves for $A = 81 km^2$ (shown in panel b) are obtained by cutting a plane on the IDAF curves in panel a at $A = 81 km^2$ (red-colored broken lines). The IAF curves on panel c) are obtained by cutting a plane at $D = 6$ hr on panel a (blue-colored broken lines).	127
VII.6	Boxplots of linear correlation (a) and probability of detection (POD) (b) for the four seasons. Each boxplot contains 71 points, 1 point for each pair of rain gauge and the underlying CPC pixel.	132
VII.7	Boxplots of relative bias in a 20-yr return level estimate for the four seasons. Each boxplot contains 71 points, 1 point for each pair of rain gauge and the underlying CPC pixel.	133

VII.8	EGPD parameters as a function of duration D and area A at a pixel located in Adelboden (elevation of 1354 m, see Figure VII.4), for the four seasons (columns). The first row is for $\kappa(D,A)$, the second row is for $\sigma(D,A)$, and the last row is for $\xi(D,A)$. In each panel, the lines represent the modeled relationship using the EGPD-IDAF model, and the points show the parameter estimates using the base model. The lines and points are colored by duration.	134
VII.9	Goodness of fit of the EGPD-IDAF model. a) Boxplots of (1-NRMSE) and b) NBias for the four seasons. Each boxplot contains 7,056 by 10 points, 1 point each for a pixel and a spatio-temporal scale (D, A).	135
VII.10	Boxplots of the cross-validation criteria for the four seasons. Each boxplot contains $2 \times 100 \times 40$ points for NRMSE_CV and FF (2 regional scores (<i>i.e.</i> $FF_{reg}^{(12)}$ and $FF_{reg}^{(21)}$) for each pair of (D, A), and 40 resamplings). In the case of SPAN20, each boxplot contains 100×40 points (1 regional score for each pair of (D, A) and 40 resamplings).	136
VII.11	Boxplots of $n95CI_{width}(\%)$ for a 50-year return level estimate, using the base model and the global model. Each boxplot contains $7,056 \times 10$ points (7,056 pixels, 10 durations).	137
VII.12	Application of the fitted EGPD-IDAF model at a pixel location in Adelboden for the summer season. The top row (a) shows some IDF curves for four spatial scales (one per column). The bottom row (b) shows the IAF curves for four temporal scales (one per column). The lines and the points show the modeled and empirical levels respectively, colored by their return periods. The colored envelopes are the 95% confidence intervals of the model estimates obtained by block bootstrap. The 50-year empirical values are not shown due to the short record length of the data	138
VII.13	Same as Figure VII.12 but for autumn at a location in Sion in the Canton of Valais (see Figure VII.4).	139
VII.14	Map of seasonal 20-year return level obtained with the EGPD-IDAF model for the spatio-temporal scales a) ($D = 1$ hr, $A = 1$ km ²) and b) ($D = 24$ hr, $A = 1,089$ km ²). The black-colored square in b) shows exemplarily the maximum extent of the rectangular window used for data aggregation, <i>i.e.</i> 1089 km ²	140
A.1	Relationship between the EGPD κ parameter and duration d for a station in Zurich (KLO). To obtain this, we fitted EGPD to the data of each of the 13 durations separately by maximum likelihood. The black points are the fitted κ . The broken lines in red represent the 95% confidence interval for the fitted model (log-log TR) see Eqn. VI.11.	154
A.2	Same as in Figure A.1, but for the EGPD scale parameter σ	155
A.3	Same as in Figure A.1, but for the EGPD shape parameter ξ . The equations for the fitted models are given in A.2.1. The broken lines represent the 95% confidence interval for the linear-log model.	155
A.4	Same as in Figure A.1, but for a station at Robbia in Graubünder (ROB)	156
A.5	Same as in Figure A.4, but for the EGPD scale parameter σ	156
A.6	Same as in Figure A.4, but for the EGPD shape parameter ξ . The equations for the fitted models are given in A.2.1. The broken lines represent the 95% confidence interval for the linear-log model.	157
A.7	Seasonal IDF curves for 6 out of the 24 large catchments. The lines are the model return levels, while the points are the corresponding empirical return levels. Both are colored according to the return period for $T = 2, 5, 10, 30,$ and 100 years.	160

A.8	Seasonal IDF curves for 6 out of the 24 large catchments. The lines are the model return levels, while the points are the corresponding empirical return levels. Both are colored according to the return period for $T = 2, 5, 10, 30,$ and 100 years. . .	161
A.9	Seasonal IDF curves for 6 out of the 24 large catchments. The lines are the model return levels, while the points are the corresponding empirical return levels. Both are colored according to the return period for $T = 2, 5, 10, 30,$ and 100 years. . .	162
A.10	Seasonal IDF curves for 6 out of the 24 large catchments. The lines are the model return levels, while the points are the corresponding empirical return levels. Both are colored according to the return period for $T = 2, 5, 10, 30,$ and 100 years. . .	163
A.11	a): Boxplots of bias (β) for the four seasons. b): Boxplots of variability ratio (γ) for the four seasons. Each boxplot contains 71 points, 1 point for each pair of gauge and the underlying CPC pixel.	164
A.12	a): Boxplots of the bias in extreme precipitation totals for the four seasons. b): Boxplots of the false alarm ratio (FAR) for the four seasons. Each boxplot contains 71 points, 1 point for each pair of gauges and the underlying CPC pixel. . .	164

List of Tables

II.1	MeteoSwiss radars altitude in m above sea level and year of installation/upgrade to dual-polarized radars	13
V.1	Summary of the regional models that are compared in this study. The first model is the local EGPD model, the next three models are based on regional homogeneity, while the last model is a spatial methods based on GAM. The second column gives the name of the model. The next three columns are the parameters of the EGPD model, and indicates whether the parameter is estimated locally (from the data of the station at hand only) or through regionalization. The last column gives reference to the section where the method is described.	63
V.2	Summary of the comparison results from the four criteria used. For each season and criteria, the model with the highest median is shown. In the case of the CRPS score however, the model with the smallest median CRPS is shown.	72
VI.1	Summary of the IDF models that are compared in this study.	91
VI.2	Summary of the AIC ranking of the models in the four seasons. The model with rank 1 is the best, i.e it has the smallest AIC for most of the stations in that season. A rank of 10 indicates the worst-performing model.	100
VII.1	Contingency table to compare gridded data and rain gauge data	116
A.1	List of variables and their meaning	158



Introduction

Contents

I.1	Motivation	2
I.2	Extreme floods in Switzerland (EXCH) project	3
I.3	Thesis objectives	4
I.4	Thesis outline	5

I.1 Motivation

Extreme precipitation, characterized by its intensity, duration, and spatial extent, has the potential to trigger disasters and associated risks, which pose profound threats to communities, ecosystems, and infrastructure. These include devastating floods, landslides, debris flows, erosion, water quality deterioration, etc. A striking example is the case of Hurricane Ian, a tropical cyclone, which in 2022 caused wind and flooding damage in western Cuba, Florida, and the Carolinas. The disaster resulted in 161 fatalities, insured losses of \$50-65 billion (Chandan et al., 2023), and estimated damage of \$113 billion. The storm hit an area of high economic value, and population growth, and was described as the third-costliest weather disaster in US history (Flinch, 2022).

Mountainous regions are particularly susceptible to extreme precipitation hazards (Frei et al., 2000). One of the primary factors contributing to these hazards is orographic lifting. When moist air masses encounter mountain ranges, they are forced to ascend, cool, and condense, resulting in heavy precipitation. This phenomenon is often intensified when warm, moist air collides with cooler, elevated terrain. The topography of mountainous areas also plays a significant role in their vulnerability to extreme precipitation and the space and time variability (Buytaert et al., 2006). Steep slopes, narrow valleys, and rugged terrain can exacerbate the impact of heavy precipitation or rapid snowmelt. These conditions increase the likelihood of flash floods, landslides, and debris flows, which pose a considerable threat to both the natural environment and human settlements.

The Alpine region (Frei and Schär, 1998), and particularly Switzerland (Stucki et al., 2012), stands out for its vulnerability to extreme precipitation and the resulting flood hazards. In addition, the Alps are influenced by both Atlantic and Mediterranean climatic regimes, which include events of stratiform and convective nature. According to the Swiss Federal Office for the Environment, FOEN¹, nine out of ten Swiss communes have already been affected by natural hazards, and around one-fifth of the Swiss population is currently exposed to the risk of flooding. River flooding, for example, resulting from high precipitation is one of the most devastating and costly natural hazards in Switzerland (Froidevaux et al., 2015). Furthermore, between 1972 and 2022, the cost of the average damage caused by natural hazards in Switzerland, 90% of which are linked to flooding and debris flows, was around CHF 305 million per year¹. In particular, the August 2005 flood in Switzerland led to six deaths and a total of three Billion Swiss francs (3×10^9 CHF) in material damage (FOEN, 2008).

What makes this situation even more critical is the undeniable impact of human activities on our climate. The latest report from the Intergovernmental Panel on Climate Change (IPCC, 2023) makes it clear that human actions, especially the emission of greenhouse gases, have significantly warmed the planet, with the global surface temperature reaching 1.1°C above 1850–1900 in 2011–2020. Switzerland, in particular, has experienced a warming trend almost double the global average (Scherrer et al., 2016b). As a result of this warming trend, heavy precipitation is expected to become more frequent and intense due to the increased moisture-holding capacity of the atmosphere.

Given the potentially catastrophic consequences associated with extreme precipitation-related hazards, there is an imperative demand for more robust and efficient models capable of enhancing our understanding and enabling reliable prediction of extreme precipitation across multiple spatial and temporal scales. This is even more crucial in topographically complex areas where robust and flexible models are required due to the substantial spatio-temporal variability of precipitation. This thesis contributes to this need in topographically complex locations, with particular application in Switzerland, by proposing robust models of Intensity-Duration-Area-Frequency (IDAF) relationships for precipitation. IDAF relationships summarize the main statistical characteristics

¹<https://www.bafu.admin.ch/bafu/en/home/topics/natural-hazards/in-brief.html>. Last access: 06-Nov-2023

of extreme precipitation (return level, return period, duration, and area). They extend the conventional Intensity-Duration-Frequency (IDF) curves by accounting for the spatial extent of precipitation (*i.e.*, the area). These models are crucial for quantifying the risk and magnitude of extreme precipitation events across space and time scales. They provide essential insights into the likelihood and intensity of extreme events over different time and spatial scales. This will enable us not only to assess areal precipitation hazards, but also to better protect against flooding disasters, develop more effective policies, plan for emergency response and flood mitigation, and enhance flood protection measures against the ravages of extreme weather events.

I.2 Extreme floods in Switzerland (EXCH) project

This thesis is part of the *Extreme Floods in Switzerland* (EXCH) project that aims to evaluate the hazards linked to extreme floods in Switzerland. The project is funded by the Federal Office for the Environment (FOEN) and the Swiss Federal Office of Energy (SFOE). It aims to improve the reliability of estimating the risks posed by such events in the future. The outcomes of the project will serve as a foundation for conducting safety assessments of buildings, facilities, and infrastructures crucial for public safety and the Swiss economy.

A particular difficulty in reliable assessments of extremely rare flood events, however, is that very long time series of discharge data are required. Unfortunately, this is hampered by the relatively short historical record of observed discharge data. To overcome this limitation, a particularly appealing way is to employ a chain of hydro-meteorological simulations. This approach employs stochastic weather generators to generate scenarios of long series of precipitation and temperature. The generated synthetic series is then used as input for a hydrological model, which simulates the corresponding discharge scenarios at any location within the studied catchments.

In a previous phase of the project titled "*Extreme flood events on the River Aare and Rhine*" (EXAR) (Andres et al., 2021), two stochastic weather generators, GWEX (Evin et al., 2018) and SCAMP (Raynaud et al., 2020) were developed and employed to generate 30 weather scenarios, each of 10,000-years, for time scales of 1 and 3 days. A lumped hydrological model was then used to simulate the corresponding discharge scenarios (Viviroli et al., 2022). The focus was on large catchments (area > 1,000 km²) located mostly in the northern part of Switzerland (catchment area of the Aare River). Critical structures in the Aare drainage basin include 19 dams and the Mühleberg, Gösgen, and Beznau I and II nuclear power plant (Andres et al., 2021).

In the current phase of the project, the scope has been extended to cover the entirety of Switzerland, with a particular focus on small hydrological catchments. As a consequence, the spatio-temporal scales of interest are now in the range of 10 to 1,000 km², and 1 to 72 h. The project has two specific goals. The first objective is to develop a stochastic weather generator of mean areal precipitation suitable for small catchments. This task is handled by a separate Ph.D thesis. The second objective, which is the primary goal of the current thesis, is to develop models of IDAF relationships of non-zero precipitation. The key deliverable of the thesis to the project is IDF curves for areal precipitation over predefined catchments. In addition, marginal distributions from IDF models developed in the course of the thesis are intended to be used to feed and validate the stochastic weather generator in the first thesis. Further details of the IDAF model are discussed in the following section.

I.3 Thesis objectives

The main goal of the thesis is to develop a model of IDAF relationships of non-zero precipitation to be applicable everywhere in Switzerland. The model will allow assessing the frequency of a given mean areal precipitation intensity, for a given duration (1 hr to 72 hr) and over an area of interest (10 to 1,000 km²). Within the context of this thesis, the interest is to model not only the extremes but the entire range of non-zero precipitation intensities (including low, medium, and extreme intensities).

While IDF relationships for point precipitation have been extensively studied in the literature (Chow, 1962; Burlando and Rosso, 1996; Koutsoyiannis et al., 1998; Blanchet et al., 2016), consideration of the area of accumulation, as provided by the IDAF relationships, has received far less attention and, to the best of our knowledge, only for extremes (De Michele et al., 2001; Overeem et al., 2010; Panthou et al., 2014; Mèlèse et al., 2019). The key novelty of the thesis is that IDAF relationships are developed for the whole range of non-zero precipitation intensities, inclusive of low, medium, and extremes. Modeling IDF and IDAF relationships using all the non-zero data instead of only the extremes, has some advantages. First, by using all the non-zero data, estimation uncertainty is expected to reduce, resulting in more accurate predictions. Secondly, in addition to having IDF and IDAF curves, the robust marginal distributions derived from the IDF models can be used to feed and validate stochastic weather generators, (e.g. the one developed in the EXCH project), for the simulation of synthetic scenarios of mean areal precipitation.

To achieve the goal of the thesis, the specific objectives are formulated in terms of the following research questions:

1. What is the most suitable and parsimonious probability distribution to model the entire range of non-zero precipitation intensities in a topographically complex area?
2. What is the most effective regionalization method to improve the robustness and reliability of daily precipitation estimates in a topographically complex area?
3. What is the best model of IDF relationships for the full range of non-zero precipitation intensities in a topographically complex area?
4. What is an appropriate model of IDAF relationships for the full range of non-zero precipitation intensities in a topographically complex area?

The study area is Switzerland, a topographically complex location with substantial seasonal and regional precipitation variability.

The research questions presented above are designed to help achieve the specific goal of the thesis, which is to develop a model of IDAF relationships of non-zero precipitation. The first question aims to identify the appropriate marginal distribution of non-zero precipitation intensities required in the IDAF relationships. The second question explores various regionalization approaches to enhance the robustness and reliability of the estimates from the marginal distribution identified in the first question. The third question investigates, builds, and compares approaches suitable for modeling IDF relationships of non-zero precipitation (without area) using the distribution identified in the first question. Finally, the fourth question extends the IDF relationships to account for the spatial extent (area) of precipitation by modeling the IDAF relationships of non-zero precipitation (the main goal of the thesis).

I.4 Thesis outline

The thesis is logically organized to address the four research questions outlined in the previous section. The study area is presented in Chapter II, while the theoretical background is presented in Chapter III. The four research questions are addressed in the order of Chapters IV to VII, followed by conclusions and discussions of the perspectives in Chapter VIII.

More specifically:

- Chapter II describes our study area, including the climatology of temperature and precipitation. This chapter also introduces the three precipitation datasets utilized throughout the thesis.
- Chapter III offers a theoretical foundation for the statistical and hydro-meteorological tools employed to address our research questions. It begins with a background on statistical modeling of extreme precipitation, followed by reviews of approaches to modeling the whole distribution of non-zero precipitation (Question 1). The regionalization methods we employ to address Question 2 are then outlined, followed by an introduction of the approaches for modeling IDF (Question 3) and IDAF relationships (Question 4). The chapter concludes by presenting the split-sampling cross-validation framework and the evaluation criteria we apply to compare the various models developed in this thesis.
- Chapter IV is dedicated to answering Question 1, by identifying a parsimonious distribution that can model the entire range of non-zero precipitation in Switzerland. Following the bibliography, a particular focus is given to the Extended Generalized Pareto Distribution (EGPD) of Naveau et al. (2016). Three parametric families of the EGPD are compared based on their ability to model daily precipitation data in Switzerland. The best model is chosen based on its parsimony and then employed to characterize extreme daily and hourly precipitation in Switzerland. Finally, the model is retained for subsequent use in developing IDF and IDAF relationships.
- Chapter V addresses Question 2 by examining regionalization methods to improve the robustness, accuracy, and reliability of daily precipitation estimates with the EGPD. Various regionalization models based on different concepts are developed and compared using the split-sampling cross-validation framework. The results are contained in Paper 1 (Haruna et al., 2022).
- Chapter VI addresses Question 3, starting with a review of existing approaches to constructing IDF relationships. Models of IDF relationships are developed and compared, with the best model selected through split sampling cross-validation. The results are contained in Paper 2 (Haruna et al., 2023b). The chapter concludes with the presentation of catchment-level IDF curves, developed using the best model identified in Paper 2, for hydrological applications in Switzerland.
- Chapter VII tackles the final research question (Question 4) by building models for Intensity-Duration-Frequency-Area (IDAF) relationships using all non-zero precipitation. The chapter begins by evaluating the gridded data, which is required for building IDAF relationships, against rain gauge measurements through point-to-pixel comparison. The IDAF relationships are then developed for various spatio-temporal scales, ranging from 1 hr/1 km² to 72 hr/1000 km². The model is evaluated and applied to assess areal precipitation hazards and build IDAF curves in Switzerland. The findings are presented in Paper 3 (Haruna et al., 2023a).

- Chapter [VIII](#) concludes the thesis by stating the conclusions and discussing the relevant perspectives for the research undertaken.



Study area and Data

Chapter overview

This chapter first introduces the area under study, Switzerland, along with its precipitation and temperature climatology. Next, the precipitation datasets that are used in the course of the thesis are described.

Contents

II.1	Study area	8
II.2	Climatology	8
II.2.1	Temperature	8
II.2.2	Precipitation	9
II.3	Data	11
II.3.1	Point precipitation data	11
II.3.2	Gridded data	12

II.1 Study area

Our study area is Switzerland, a country spanning 41,285 km². Figure II.1 shows the map of the study area. Despite its small size, Switzerland exhibits a complex topography, with elevation ranging from 191 to 4,127 m above sea level. Approximately 30% of the land is situated above 1,500 m (Figure II.2), leading to significant spatial variations in both precipitation intensity and occurrence. Various factors, including the Alps, the Atlantic Ocean, and the Mediterranean Sea influence the country's climate. These factors contribute to seasonal and spatial variability in precipitation, as documented in prior studies (see e.g. Sodemann and Zubler, 2009; Giannakaki and Martius, 2015; Scherrer et al., 2016a).

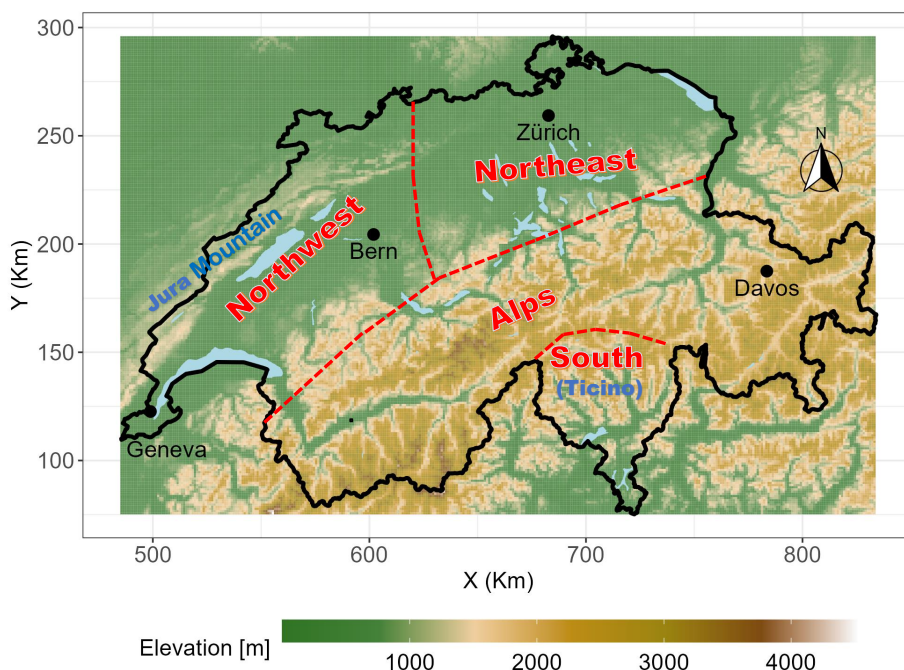


Figure II.1: Map of Switzerland, the study area. The background color denotes the elevation (m) above sea level. The red-broken lines demarcate the four climatological zones (northwest, northeast, Alps, and south (Ticino)). Some major lakes are highlighted in light blue. The names and locations of some cities are shown in black, while the location of the Jura mountain is shown in blue.

II.2 Climatology

II.2.1 Temperature

Figure II.3, reprinted from MeteoSwiss (2023), shows the map of the annual mean temperature in Switzerland for the period 1991–2020. The mean annual temperature ranges from below -9°C to over 14°C . There seems to be a strong correlation between elevation and temperature, with the lowest temperatures observed at the highest elevations in the Alps, and the highest temperatures in the low-lying areas, such as the inner valleys, south of the Alps in Ticino, and north of the Alps in the Swiss plateau. Specific regions with very high temperatures include the area along Lake

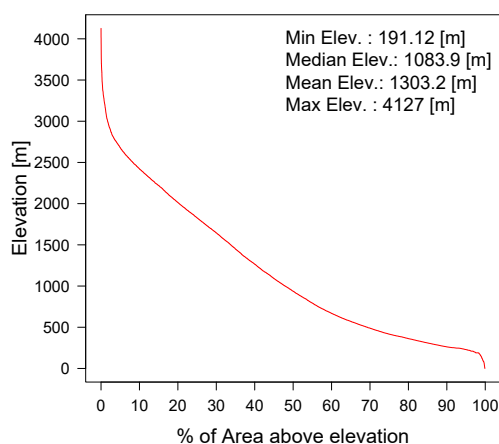


Figure II.2: Hypsometric curve of Switzerland based on 1 km by 1 km elevation grid.

Geneva and the foothills of the Jura mountains in western Switzerland, the inner valleys as well as the Ticino region in the south of the Alps. July and August are the months in which the highest temperatures are recorded, with average maximum temperature exceeding 25°C in the low-lying regions (MeteoSwiss, 2023).

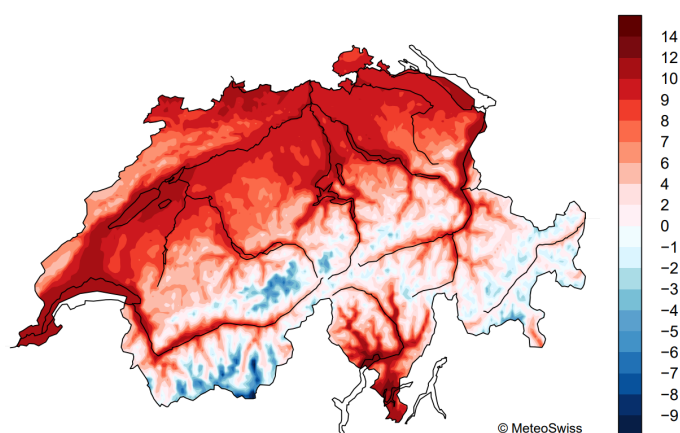


Figure II.3: Annual mean temperature (°C) for the period 1991-2020. Reprinted with permission from MeteoSwiss (2023).

II.2.2 Precipitation

In Switzerland, precipitation patterns exhibit specific regional differences with four main climatic zones (see e.g. Frei and Schär, 1998; Molnar and Burlando, 2008). The four zones are shown in Figure II.1. The first two zones are located in the northern alpine foreland and are divided into the northwest and northeast. The northeast exhibits little variability while the northwest is influenced by the Jura mountain, leading to higher accumulations compared to the northeast. Next is the central part where the Alps is located, and precipitation here is mainly affected by orography and shadowing of the inner valleys. Finally, the south of the Alps, around Lake Maggiore, is heavily influenced by the Mediterranean Sea, leading to high accumulations and intense precipitation.

Annual precipitation sums vary across the regions (Figure II.4a), with the highest annual totals exceeding 2,000 mm in areas like the Alps, the Jura region (northwest), and the Ticino region (south of the Alps). Conversely, shielded from both sides, the inner valleys such as the Rhône and Inn experience minimal annual precipitation, between 400 to 800 mm. In terms of seasonality

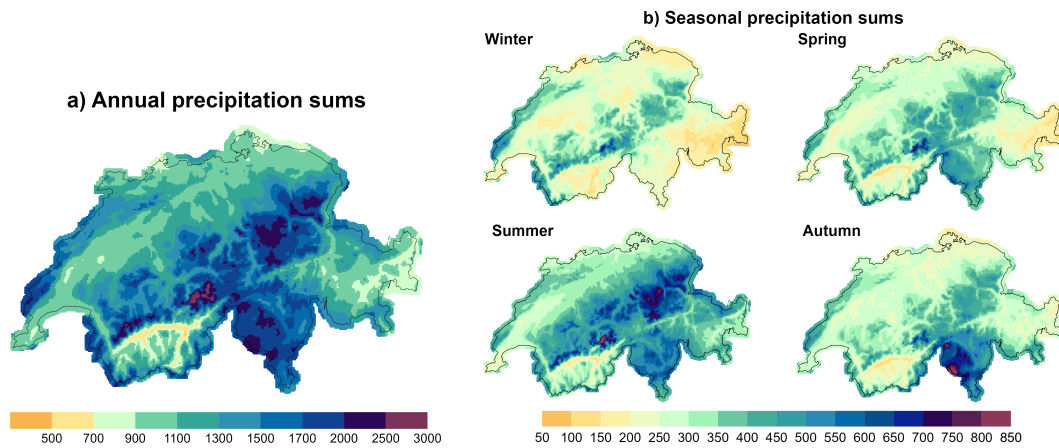


Figure II.4: a) Annual precipitation sums (mm) and b) Seasonal precipitation sums (mm) for the period 1991–2020, computed from *RhiresD* (see Section II.3.2.2 for the description of this dataset).

(Figure II.4b), summer is the primary precipitation season throughout Switzerland. An exception is Ticino where the main season is autumn. Conversely, winter records the lowest precipitation across all regions.

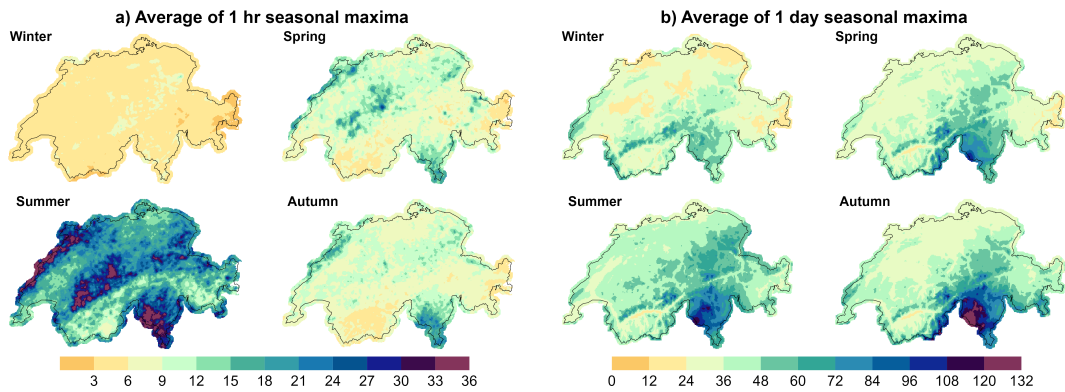


Figure II.5: a) Average of 1 hr seasonal precipitation maxima (mm) computed from *Combiprecip* and b) Average of 1 day seasonal precipitation maxima (mm) computed from *RhiresD*. See Section II.3.2.2 for the description of the two datasets.

Figure II.5 shows the spatial distribution of heavy precipitation, defined here as the average of seasonal maxima. It can be observed that the spatial distribution depends on the accumulation duration. For short-duration accumulations (e.g., 1 hr), the most intense precipitation occurs during summer nationwide, with peak intensities of up to 30 mm/hr in areas like Ticino, Jura, and the northern rim. Since precipitation at this time scale is mainly convective, it is not surprising that the lowest intensities are observed in winter. For longer accumulations (e.g. 1 day), Ticino receives the heaviest precipitation with autumn experiencing maximum 1 day totals surpassing 130 mm. In other regions, heavy precipitation mainly occurs during summer.

II.3 Data

This section introduces the datasets of precipitation measurement that we use in the thesis. Two categories of data are available, point precipitation measurements recorded at rain gauge networks, and gridded data at a spatial resolution of 1 km by 1 km square grids.

II.3.1 Point precipitation data

In general, point data from rain gauge measurement networks provide the longest available measurement for precipitation. They are also considered the most accurate and reliable source of precipitation measurement in comparison to other methods such as weather radar and satellite observations. Despite their reputation, some major sources of uncertainties (e.g. errors induced by wind, wetting, splashing, and evaporation errors) affect the accuracy of rain gauge measurements which leads to systematic bias. Another major limitation of point measurements is their lack of spatial representativeness. These errors contribute inevitably to the overall uncertainty in any statistical analysis with rain gauge data.

In Switzerland, point measurements from rain gauge networks are subjected to substantial and very rigorous quality control checks by the Swiss Federal Office for Meteorology and Climatology (MeteoSwiss) to ensure the reliable quality of the data (MeteoSwiss, 2017; Fukutome et al., 2018). These include checks to ensure that recorded values are within climatologically physical limits, they are consistent with those from nearby gauges, and they satisfy inter-parameter consistency and variability tests.

For our thesis, point precipitation measurements are available at two sampling resolutions. The first measurements at 10 min sampling resolution are recorded using tipping-bucket rain gauges at automatic weather stations across the country. The second measurements are made daily and come from the manual precipitation monitoring network. The two datasets are described in the following sections.

II.3.1.1 10 min precipitation data

The precipitation data at 10 min sampling resolution comes from 81 stations. They are spread across Switzerland and their locations, represented by the blue-colored circles, are shown in Fig. II.6. Out of this total, 71 stations belong to the Swiss Meteorological Network (SwissMetNet) of MeteoSwiss, while 10 belong to the canton of Luzerne, a partner network of MeteoSwiss. The precipitation data is measured with a tipping-bucket gauge of 0.1 mm depth resolution at a sampling resolution of 10 min. Most of the stations at high altitudes are shielded from wind and the tipping gauge is heated in order to account for snow. The sample data has a variable length ranging from a minimum of 20 years to a maximum of 40 years from 1981 to 2020. The stations are located at elevations ranging from a minimum of 203 m, an average of 952.4 m, and a maximum of 3294 m. This dataset is aggregated to 1 hr and used in Chapter IV to model hourly precipitation, while it is aggregated to durations from 30 min to 72 hr in Chapter VI to build Intensity-Duration-Frequency curves.

II.3.1.2 Daily precipitation data

We have a comprehensive dataset from 1176 stations for the daily precipitation measurements. Their locations are shown by the red-colored crosses in Figure II.6. From this total, 500 are located within Switzerland and 676 in the neighboring countries (France, Italy, Austria and Germany). The data has a variable length ranging from a minimum of 20 years to a maximum of 156 years, from 1863 to 2019. Although our study is limited to Switzerland, the use of the data in the neighboring countries is simply to improve our model estimates near the border through a technique called

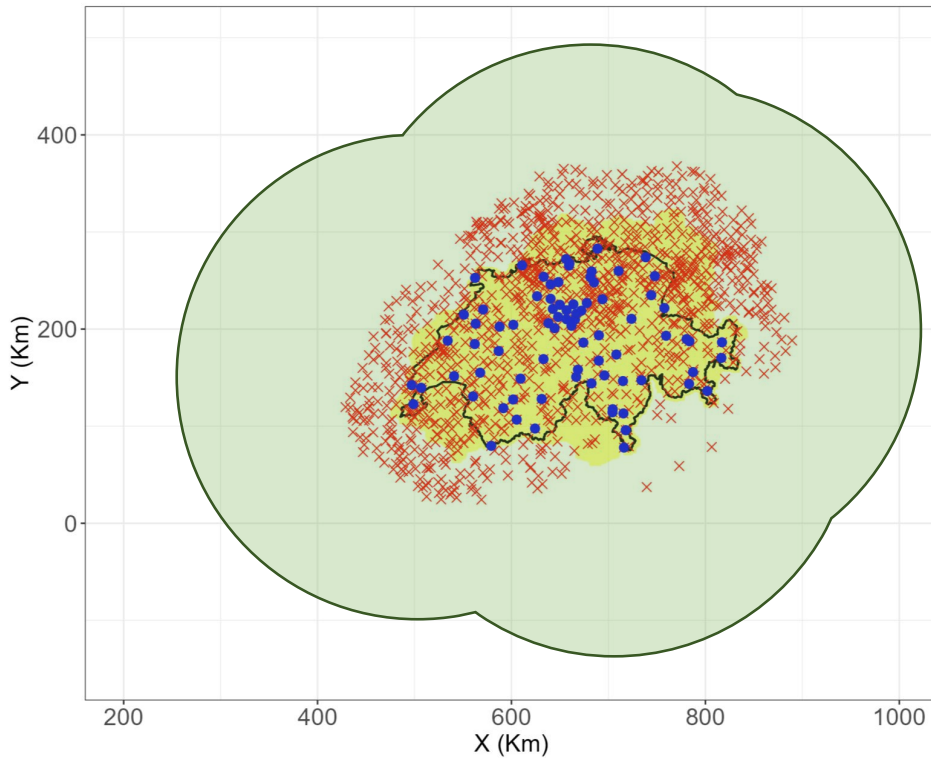


Figure II.6: Map of Switzerland showing the four datasets available for the study. The blue-colored circles represent the 81 rain-gauge stations at 10 min resolution, the red-colored crosses show the locations of the 1176 daily stations, the yellow background shows the extent of RhiresD, and the dark green-colored line with light green background shows the coverage of CombiPrecip (CPC)

regionalization (see Chapter V for more details). In Switzerland, the measuring instruments for daily measurements have evolved with time. It started from completely manual rain gauges (400 cm² ombrometers to 200 cm² Hellmann raingauge) recorded at 06:00 UTC, to fully automatic gages (tipping-buckets and weighing gauges) with a sampling resolution of 10 min (Fukutome et al., 2018; MeteoSwiss, 2021). The daily data from the other countries are also recorded at 06:00 UTC, except Italy where measurements are at 08:00 UTC (Isotta et al., 2014). We use this dataset in Chapter IV to model daily precipitation, and in Chapter V to build regionalized models.

II.3.2 Gridded data

As pointed out earlier, while rain gauges provide direct measurements of precipitation, their major drawback is their limited spatial representativeness. Furthermore, in this thesis, addressing the fourth research question (see Chapter I), which has to do with building IDAF curves, requires information on areal precipitation intensities everywhere in the study area. To obtain this information, we make use of two gridded products, one based on radar-reanalysis called CombiPrecip, and another based on spatial interpolation of daily precipitation data named RhiresD. The two datasets are presented in the following subsections, along with their related uncertainties.

II.3.2.1 Radar-reanalysis product: CombiPrecip

CombiPrecip (CPC) is a radar-reanalysis product resulting from the geostatistical merging of radar and rain gauge in an operational setting (Sideris et al., 2014a). It combines the high accuracy

of rain gauge with the high spatial coverage of radar. The geostatistical merging is through co-kriging with external drift, where the rain gauge data is treated as the primary source, and the radar data as the external drift. Information from rain gauges comes from more than 250 automatic stations at 10 min resolution, and that from the radar comes from five polarimetric C-band Doppler radars that are suitably located to provide the reliable coverage required in the topographically complex area (Table II.1). The radars were initially three C-band Doppler radars located at Albis (installed in 1959), La Dole (1959), and Monte Lema (1993), which were upgraded to polarimetric Doppler radars in 2011. Later, two additional radars were installed, one at Pointe de la Plaine Mortes in 2014 and another at Weissfluhgipfel in 2016, to make a total of five (Germann et al., 2006; Panziera et al., 2018; Germann et al., 2022).

Name	Altitude (m)	Year installed	Year upgraded
Albis	938	1959	2012
La Dole	1682	1959	2011
Monte Lema	1626	1993	2011
Pointe de la Plaine Morte	2937	2014	-
Weissfluhgipfel	2840	2016	-

Table II.1: MeteoSwiss radars altitude in m above sea level and year of installation/upgrade to dual-polarized radars

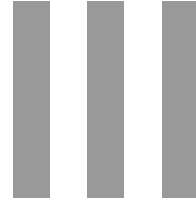
The data from the radar is subjected to substantial quality control before being employed in the CPC algorithm. Treatment of the radar data (Germann et al., 2006) involves clutter elimination through a robust algorithm designed for this purpose, visibility correction resulting from orthographic shielding, correction for vertical profile of reflectivity, and bias correction. This is in addition to an automatic hardware calibration of the radars to check the stability/accuracy of the components and a tailored operational scan strategy (20 elevation sweeps every five min) crucial in mountainous regions such as Switzerland (Germann et al., 2015).

Known limitations and sources of uncertainty in the use of CPC involve the limited length of the data, non-homogeneity of the series due to radar upgrades and evolution of the number of radars, and conditional bias. CPC was used in several applications in Switzerland for extreme value analysis (Panziera et al., 2016), climatological studies (see e.g. Panziera et al., 2018), meteorological forcing of hydrological model (see e.g. Andres et al., 2016), and has been evaluated in several aspects (see e.g. Gabella et al., 2017; Panziera et al., 2018; Gugerli et al., 2020). It is available at hourly temporal resolution and a spatial grid of 1 km by 1 km and extends 100 - 150 km beyond the borders of Switzerland. It has been available since 2005, and 17 years of data, from 1st January 2005 to 31st December 2021 is used to build IDAF curves in Chapter VII.

II.3.2.2 Spatial analysis of daily precipitation: RhiresD

The second gridded data is RhiresD that results from the spatial interpolation of station data at a daily scale (MeteoSwiss, 2021). It contains daily precipitation totals, corresponding to accumulations from 06:00 UTC of the day D to 06:00 UTC of the day $D+1$. It uses information from over 600 manual and automatic gauging stations, within and outside Switzerland, whose density has evolved with time. The interpolation scheme is based on a regionally varying precipitation-topography relationship, estimated by local weighted linear regression. There are two main sources of uncertainty associated with the dataset. The first comes from rain gauge measurement errors such as wind-induced deflections and under-catch that result in underestimation of true precipitation. The second source of error is related to interpolation, which potentially results in the

overestimation of light and the underestimation of intense precipitation. RhiresD has been used in many applications in the literature (see e.g. [Köplin et al., 2014](#); [Panziera et al., 2018](#); [Gugerli et al., 2019](#); [Pfister et al., 2020](#)). A total of 61 years of data, from January 1961 to December 2021, at a spatial resolution of 1 km², is available for this study. We use this data to evaluate the statistics of the CPC data for 24 hr totals in Chapter [VII](#).



Statistical methods and hydro-meteorological relationships

Chapter overview

This chapter introduces the statistical methods and hydro-meteorological relationships applied in the thesis. An overview of the extreme value theory models followed by methods for modeling the entire distribution of non-zero precipitation is presented. Regionalization methods to improve the at-site estimates of return levels are then described. Next, methods for modeling Intensity-Duration-Frequency (IDF) and Intensity-Duration-Area-Frequency (IDAF) relationships of precipitation are presented. Finally, the evaluation framework for the comparison of statistical models developed in the course of the thesis is described.

Contents

III.1	Statistical modeling of extreme precipitation	17
III.1.1	Extreme Value Theory	17
III.1.2	Modeling the whole range of non-zero precipitation	20
III.2	Regionalization of at-site estimates of return level with EGPD	27
III.2.1	RFA based on distinct homogenous regions	27
III.2.2	RFA based on region of influence approach (ROI)	28
III.2.3	RFA based on upper tail behavior	28
III.2.4	Spatial method based on Generalized Additive Model (GAM)	29
III.3	Intensity-Duration-Frequency (IDF) relationships	30
III.3.1	Scaling IDF curves	31
III.3.2	General IDF formulation	33
III.3.3	Data-driven IDF curves	34
III.4	Intensity Duration Area Frequency (IDAF) relationships	34
III.4.1	ARF-based IDAF models	36
III.4.2	Data-driven IDAF models	37

III.5	Evaluation framework	37
III.5.1	Split-sampling cross validation	37
III.5.2	Comparison criteria	38

III.1 Statistical modeling of extreme precipitation

In engineering, the key safety design requirement of a structure, such as a dam, is that it should not fail within its intended design life. This means that the dam should safely evacuate the largest discharge that it will experience within its design life. Such a large discharge is rare or has a very low frequency of occurrence, and in engineering terms, a very high "return period" T . A typical return period for such design based on standards is $T = 10,000$ years. This means that the discharge has on average $1/10,000$ or 0.0001 probability of being exceeded in any given year.

The question is now how to give a reasonable estimate of such an amount that will be used for dimensioning the dam. For reliable estimates, a 10,000-year event requires at least 10,000 years of observed data (Hosking and Wallis, 2005). The main difficulty, however, is that the observed data required to estimate the rare event usually spans only a few decades. Accordingly in the absence of the required length of data, there is a need for a sound mathematical framework to justify the necessary extrapolation to estimate this rare event. Gladly, extreme value theory (EVT) provides this solid framework for the statistical modeling of extreme events (Coles, 2001).

III.1.1 Extreme Value Theory

Under the EVT framework, there are two main classical modeling approaches. The first is an approach based on the block maxima of a time series using the Generalized Extreme Value (GEV) distribution. The second is based on peaks over threshold (POT) using the Generalized Pareto distribution (GPD). Figure III.1 illustrates the block maxima and the POT samples. In this section, we will give a brief account of the theory behind the two approaches. For details of other modeling approaches, the reader is referred to Coles (2001).

III.1.1.1 Approach based on block maxima

Let us consider a sequence of independent and identically distributed (iid) random variables X_1, X_2, \dots, X_n , having the same cumulative distribution F . In our context, X_i can be assumed as the daily precipitation amount recorded at a rain gauge station in Switzerland on day i . Let us also define M_n as the maxima over a block of size n , such that $M_n = \max(X_1, X_2, \dots, X_n)$. If the block size is taken as 1 year, then $n = 365$ and $M_{n=365}$ is the annual maximum (see Figure III.1).

We want to find, or at least, approximate the cumulative distribution function (CDF) of M_n . Knowing the CDF of M_n is enough to compute the return period T corresponding to any non-exceedance probability p from $T = \frac{1}{p}$. Then the question is "What is the CDF of M_n ?"

A first starting point is to obtain the CDF of M_n as :

$$\mathbb{P}(M_n \leq x) = \mathbb{P}(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) = F^n(x).$$

However, this will lead to degeneracy because as $n \rightarrow \infty$, then $F^n(x) = 0$ as $x \rightarrow 0$ or $F^n(x) = 1$ as x gets more extreme.

This degeneracy can be avoided if we consider normalized maxima, that is if we can find some sequence of real numbers $a_n > 0$ and b_n such that $\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right)$ converges in distribution to a non-degenerate distribution function G . This can be written as:

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) = F^n(a_n x + b_n) \xrightarrow{n \rightarrow \infty} G(x).$$

If this condition holds, then we say that F is in the maximum domain of attraction of G , written as $F \in MDA(G)$.

Fisher and Tippett (1928) and Gnedenko (1943) showed that if $a_n > 0$ and b_n exist, then G has to be one of the families of the GEV distribution defined as:

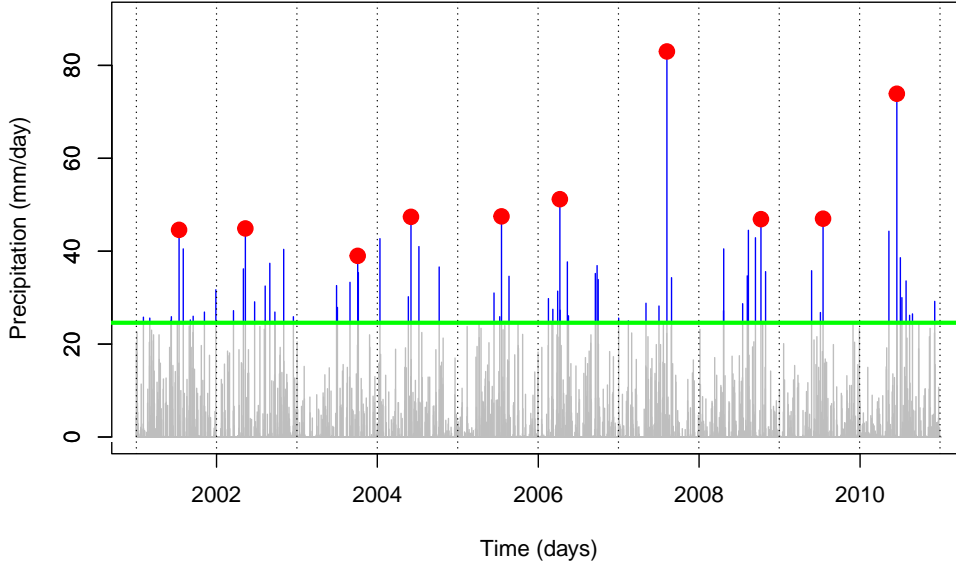


Figure III.1: Time series of daily precipitation at a station in Zürich from 2001–2010. All the observations above the threshold, taken as the 98% quantile (green-colored line), constitute the peaks over threshold (POT). The annual maxima are highlighted by the red points. In this case, there are 10 annual maxima and 74 POT.

$$G_{\mu,\sigma,\xi}(x) = \begin{cases} \exp \left[- \left(1 + \xi \frac{x-\mu}{\sigma} \right)_+^{-1/\xi} \right] & \text{if } \xi \neq 0 \\ \exp \left[- \exp \left(- \frac{x-\mu}{\sigma} \right) \right] & \text{if } \xi = 0 \end{cases}, \quad (\text{III.1})$$

defined on the set $\{x : 1 + \xi(x - \mu)/\sigma > 0\}$, where $\mu \in \mathbb{R}$ is the location, $\sigma > 0$ is the scale and $\xi \in \mathbb{R}$ is the shape parameter.

Three particular cases of the GEV can be identified depending on the value of the shape parameter ξ . Figure III.2 gives an illustration of the density and return level plot for the three cases.

- If $\xi > 0$, then $G_{\mu,\sigma,\xi}$ is Fréchet type, and heavy-tailed.
- If $\xi = 0$, then $G_{\mu,\sigma,\xi}$ is Gumbel type, and light-tailed.
- If $\xi < 0$, then $G_{\mu,\sigma,\xi}$ is Weibull type, and bounded-tailed.

Now that we know that GEV is the asymptotically justified distribution for the block maxima, we can use its CDF to obtain the return level corresponding to any return period T of interest. The T -year return level can be obtained from the quantile function of the GEV distribution as:

$$x_T = \begin{cases} \mu - \frac{\sigma}{\xi} \left\{ 1 - \left[-\log \left(1 - \frac{1}{T} \right) \right]^{-\xi} \right\} & \text{if } \xi \neq 0 \\ \mu - \sigma \log \left[-\log \left(1 - \frac{1}{T} \right) \right] & \text{if } \xi = 0 \end{cases}, \quad (\text{III.2})$$

where x_T is the quantile that is exceeded with a probability $p = \frac{1}{T}$. A plot of x_T as a function of T (called the return level plot) for different values of ξ is shown in Figure III.2. It can be seen that the plot is linear for $\xi = 0$, convex with asymptotic limit for $\xi < 0$ and concave with no finite bound when $\xi > 0$.

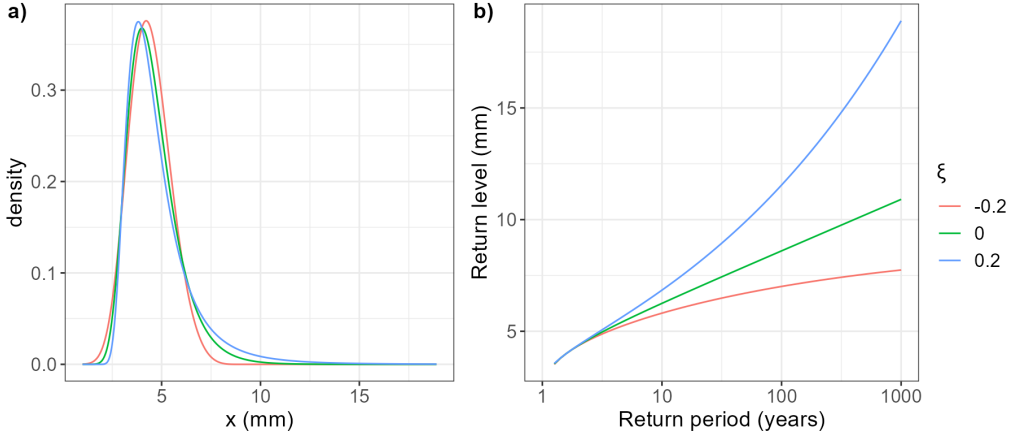


Figure III.2: Illustration of a) the density plot and b) the return level plot for the three special cases of the GEV, i.e. Weibull type ($\xi = -0.2$), Gumbel type ($\xi = 0$) and Fréchet type ($\xi = 0.2$). In each case, the location parameter $\mu = 4$ and the scale parameter $\sigma = 1$.

III.1.1.2 Approach based on peaks over threshold

While the GEV, presented in the previous section, is the classical approach to model extremes, the major drawback lies in the wasteful use of the data, since within each block, only one value is retained, and the other values are discarded. In Figure III.1 for example, only 10 values (highlighted in red), representing the annual maxima in the 10 years time series will be retained, and the others will be discarded. The issue is that some extremes may be contained in the discarded data, especially in cases where the extreme events are sporadic or irregularly spaced. In Figure III.1, the annual maximum of 2003 is not more extreme than other non-annual maxima values (in 2010 for example). It is therefore apparent that a more suitable approach that will use all the extremes, not only the block maxima is needed.

A judicious approach (see Coles, 2001), is to consider all the excesses above a sufficiently high threshold u and use them in the modeling. These excesses are called peaks over threshold (POT), and an example of the POT samples is shown in Figure III.1 in blue. The next task is to find an asymptotically justified model for the POT.

Recall from Section III.1.1.1 that X_1, X_2, \dots, X_n , is a sequence of independent and identically distributed random variables having the same cumulative distribution F . If F belongs to the domain of attraction of G , that is, if the block maxima of X can be approximated by GEV, then Balkema and De Haan (1974) and Pickands III (1975) showed that the limiting distribution of the excesses $(X - u)$, conditioned on $X > u$, is the generalized Pareto (GPD). That is:

$$\mathbb{P}(X - u \leq x | X > u) \xrightarrow{u \rightarrow x_F} H_\xi \left(\frac{x}{\sigma_u} \right) \text{ for } 0 \leq x < x_F - u,$$

with

$$H_\xi \left(\frac{x}{\sigma_u} \right) = \begin{cases} 1 - (1 + \xi \frac{x}{\sigma_u})_+^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - \exp(-\frac{x}{\sigma_u}) & \text{if } \xi = 0 \end{cases}, \quad (\text{III.3})$$

where $a_+ = \max(a, 0)$, $\sigma_u > 0$ is the scale parameter, and $\xi \in \mathbb{R}$ is the shape parameter. The subscript u in σ_u indicates that σ_u depends on the chosen threshold u . GPD therefore has only two parameters, σ_u and ξ . The GEV and the GPD are related, such that $\sigma_u = \sigma + \xi(u - \mu)$ with both distributions having the same shape parameter ξ . μ and σ are respectively the location and scale parameters of the GEV. Three cases of the distribution are possible, the bounded case (short-tailed)

when $\xi < 0$, the exponential case (light-tailed) when $\xi = 0$, and the unbounded case (heavy-tailed) when $\xi > 0$. An illustration of the three cases is shown in Figure III.3.

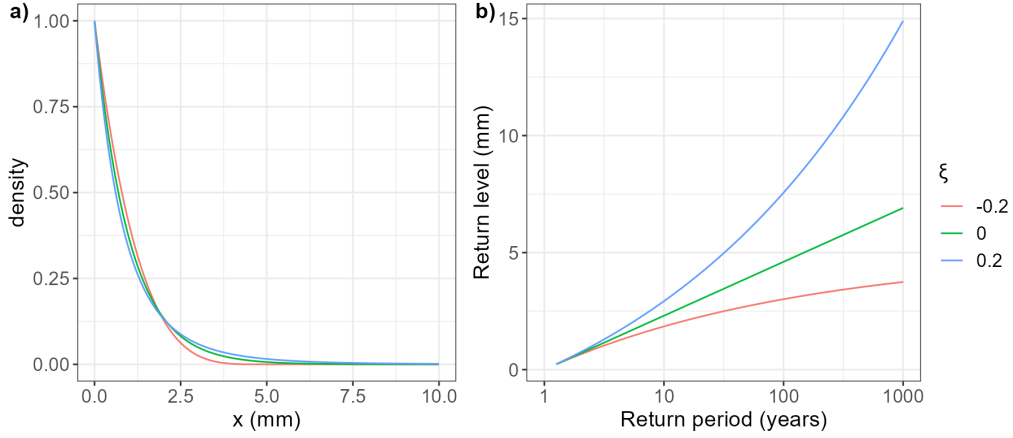


Figure III.3: Illustration of the a) density plot and b) return level plot for the three special cases of the GPD, i.e. bounded case ($\xi = -0.2$), exponential case ($\xi = 0$) and unbounded case ($\xi = 0.2$). In each case the scale parameter $\sigma_u = 1$.

The T -year return level using the GPD is obtained from:

$$y_T = \begin{cases} u + \frac{\sigma_u}{\xi} \left[(T n_y \zeta_u)^\xi - 1 \right] & \text{if } \xi \neq 0 \\ u + \sigma_u \log(T n_y \zeta_u), & \text{if } \xi = 0 \end{cases}, \quad (\text{III.4})$$

where $\zeta_u = \mathbb{P}(X > u)$ is estimated from the sample proportion of the points exceeding u , i.e. $\frac{k}{n}$, with k the number of exceedances of u and n the total number of observations. n_y is the number of observations per year. The return level plot is shown in Figure III.3 b) for different values of ξ .

The use of GPD has significantly improved the estimation uncertainty as reported by Langousis et al. (2016). This is because GPD uses more data compared to GEV. The authors have also highlighted that the GPD performs better in many hydrologic applications, particularly when $\xi > 0$. To demonstrate this, we applied both models to 30 years of daily data shown in Figure III.1. The maximum likelihood estimation was used for fitting the models, and the delta method (refer to Coles, 2001, for the details of the method) was used to estimate the uncertainty. The return level plots, along with the 95% confidence bounds, are presented in Figure III.4. It can be observed that narrower bounds are obtained with the GPD due to the more efficient use of data.

Despite the potential advantage of GPD over GEV, a major drawback in the application of the GPD is the lack of a unified framework for the choice of u , in spite of several advancements (see reviews in Scarrot and MacDonald, 2012; Fukutome et al., 2015; Langousis et al., 2016; Northrop et al., 2017). The correct choice is crucial because of the resulting bias-variance trade-off. A high threshold leads to small bias but increased estimation variance, while a smaller threshold leads to significant bias due to the deviations from a perfect GP model (Coles, 2001).

III.1.2 Modeling the whole range of non-zero precipitation

In addition to the delicate issue of threshold selection in the application of GPD, the issue of inefficient use of hydrologic information nonetheless remains, since the POT still accounts for a very minute fraction of the whole data (less than 5% if a threshold of 95% is used). In Figure III.1, for instance, all the data below the POT (highlighted in grey) will be discarded. Another major drawback is the inability to model the bulk of the distributions. Modeling the whole range of precipitation has applications in water resources management for urban water supplies, hydropower,

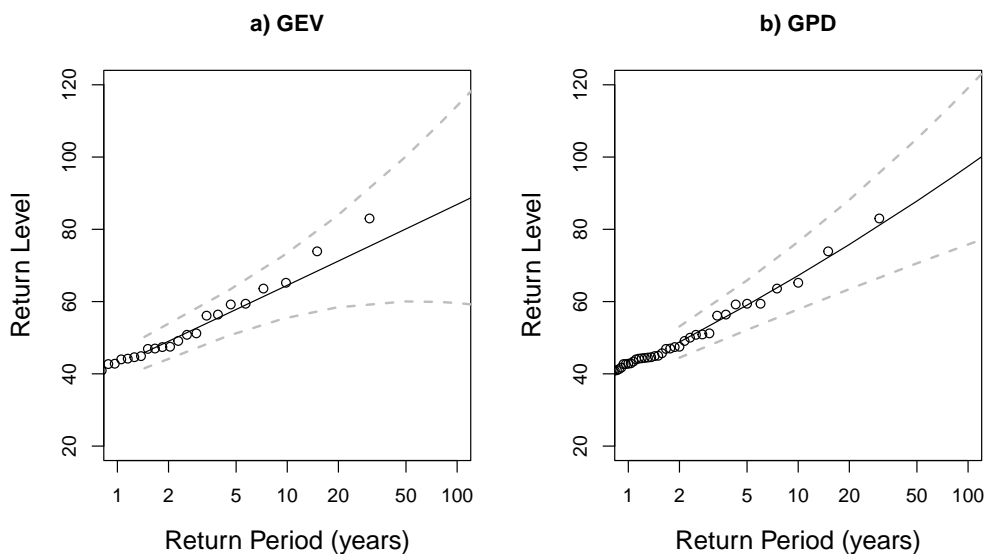


Figure III.4: Illustration of the return level uncertainty on the daily data shown in Figure III.1 with a) GEV and b) GPD. In both plots, the points represent the observed values, the black lines represent the fitted return levels, and the gray dashed lines represent the 95% confidence intervals obtained with the delta method (see Coles, 2001, for the details of the method)

flood forecasts, and irrigation systems. As an example, marginal distribution for the entire range of precipitation is required in stochastic generators (see examples of application in Evin et al., 2018; Viviroli et al., 2022). Simulated precipitation from the generators is used as input to conceptual hydrological models for subsequent flood modeling and risk assessment. Other practical applications are in the evaluation of numerical weather simulations or investigation of the climatology of precipitation events as outlined in Blanchet et al. (2019).

Within the framework of modeling the whole range of precipitation, gamma distribution is often the preferred candidate. However, a major drawback is the potential to underestimate heavy precipitation (Katz et al., 2002). This is because gamma distribution is light-tailed since it is in the Gumbel domain of attraction (Haan and Ferreira, 2006). As a consequence, several approaches were proposed as alternatives to the gamma distribution. In the following sections, we will present a brief review of some approaches for modeling the whole range of non-zero precipitation. While some of the approaches are explicitly for modeling the whole range of non-zero precipitation, others are meant to simply add flexibility to the fit of the GPD in terms of the threshold selection.

III.1.2.1 Models based on GP-Mixtures

Mixture models generally involve modeling the upper tail of the distribution with the asymptotically justified GPD, and another for the bulk of the distribution. Depending on the type of the model for the bulk of the distribution, they can be categorized into parametric, semi-parametric, or non-parametric (Scarrot and MacDonald, 2012; Naveau et al., 2016). In the parametric models, for example, Frigessi et al. (2002) employed the Weibull distribution along with a weighting function, Behrens et al. (2004) used a gamma distribution, and Carreau and Bengio (2009) applied a Gaussian distribution for the bulk of the distribution. In the semiparametric approach, Cabras et al. (2009) approximated the bulk distribution by an equispaced binning of the data, followed by fitting a Poisson log-link GLM to the counts. do Nascimento et al. (2012) on the other hand utilized a weighted mixture of gamma densities for the bulk distribution. For the nonparametric approach,

MacDonald *et al.* (2011) used a kernel density estimator for the bulk of the distribution, while Tancredi *et al.* (2006) employed a flexible mixture of uniform distributions for below-threshold data. An illustration of some mixture densities is shown in Figure III.5. A common drawback of this approach, which is based on mixtures, is its lack of parsimony in terms of the high number of parameters, the resulting inferential challenge, and the lack of robustness in keeping the two tails separated (Scarrot and MacDonald, 2012; Naveau *et al.*, 2016)

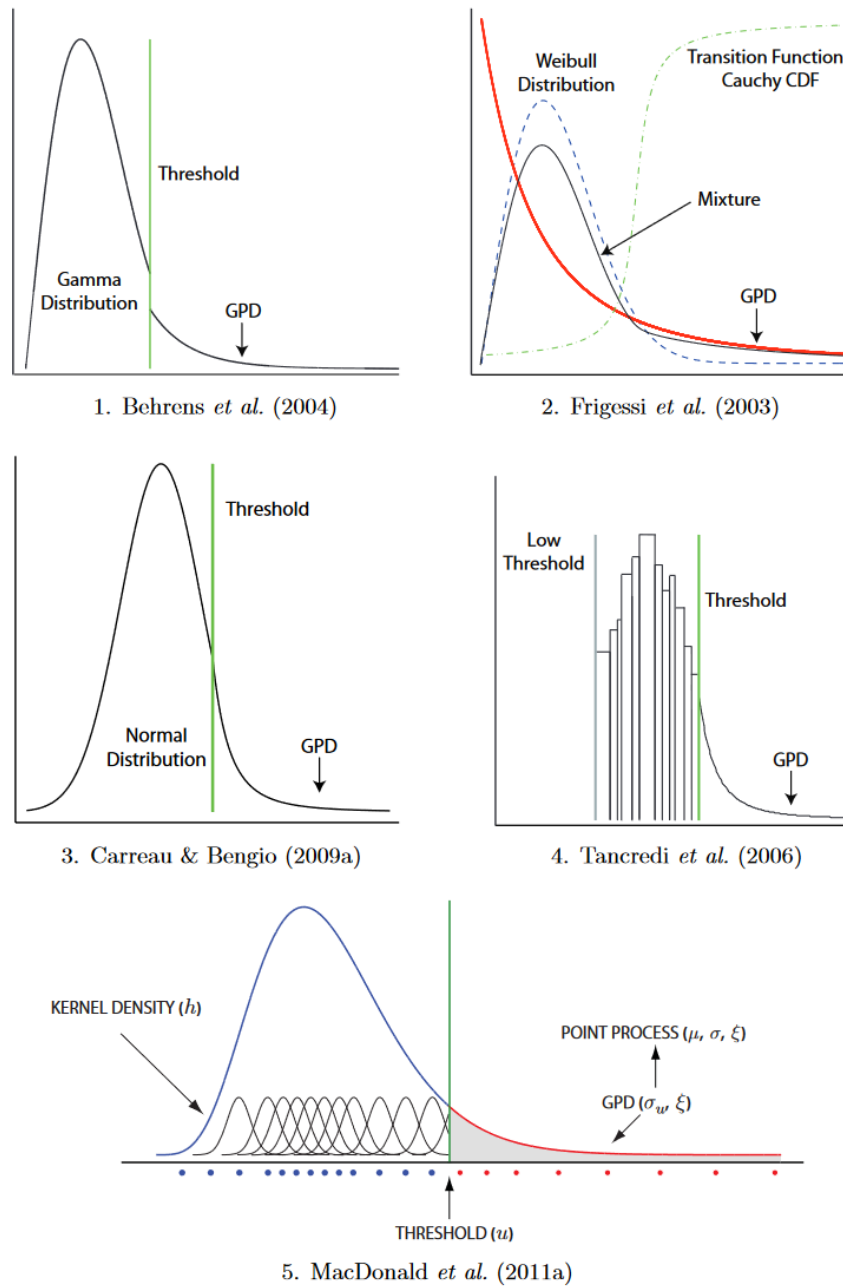


Figure III.5: Illustration of some mixture densities. Figure taken from Scarrot and MacDonald (2012).

III.1.2.2 Models based on probability integral transform

Papastathopoulos and Tawn (2013) proposed an approach to extend the GPD by allowing for the use of a sub-asymptotic threshold while still preserving the GPD distribution in the upper tail. The ultimate goal here is to create models that are more flexible, in terms of the threshold choice, compared to the GPD, thereby yielding an improvement in the estimation of the upper tail. This is a different motivation from the previously mentioned approaches that intended to model the full range of the distribution. The model is constructed through a probability integral transform, where GPD-like random variables are generated by replacing the uniform draws, with something more flexible, that has support on the unit interval $[0, 1]$, and contains the uniform draws as special cases. Three parameter extensions were proposed by the authors, which are based on power law, incomplete beta function, and incomplete gamma. They both have one additional parameter compared to the GPD, and all showed less sensitivity to threshold choice compared to the GPD, in parameter and return level estimates.

Gamet and Jalbert (2022) noticed that the three extended models proposed in Papastathopoulos and Tawn (2013) either have a null or infinite density at the origin (*i.e.* at the location of the threshold). An illustration is given in Figure III.6 for the case of power law defined as $K(u) = u^\kappa$. This is a practical constraint since in many applications where the GPD is used, the density at the threshold is positive and finite. To overcome this, they proposed two more extensions, one based on the truncated normal distribution, and another based on the truncated beta distribution, that satisfy the constraint of being positive and finite at the origin. The authors showed that their extensions provided a better estimate of the upper tail index than the previous extensions. In addition to the use of a sub-asymptotic threshold, it is also possible to use a zero threshold for precipitation.

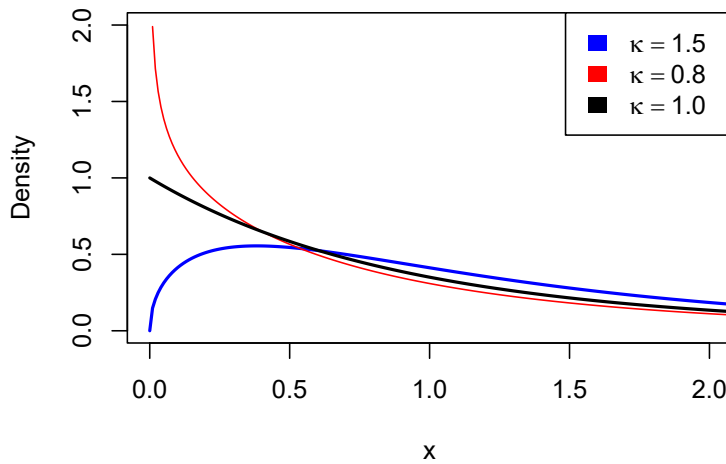


Figure III.6: Illustration of the EGP of Papastathopoulos and Tawn (2013) based on power law defined as $K(u) = u^\kappa$. The density is zero at the origin when $\kappa > 1$ (blue line) and infinite when $\kappa < 1$ (red line). The case where $\kappa = 1$ corresponds to the density of GPD (black line). Adapted from Gamet and Jalbert (2022).

III.1.2.3 Meta-statistical extreme value (MEV) distribution

Beyond the class of GPD extensions, the meta-statistical extreme value (MEV) distribution (Marani and Ignaccolo, 2015) has been widely applied in modeling hydro-meteorological extremes (see Zorzetto et al., 2016; Marra et al., 2018, 2019; Gründemann et al., 2023; Falkensteiner et al., 2023, for examples). This method does not require the asymptotic assumption, like in the EVT, and makes full use of all the data, referred to as "ordinary events" rather than only the extremes (block maxima or peaks over threshold). The ordinary events in each year, defined as all the precipitation events above a small threshold (e.g. 0.3 mm for 3 hr intensities in Gründemann et al., 2023), are modeled with the Weibull distribution. The CDF of the MEV is given as:

$$F(x) = \frac{1}{M} \sum_{j=1}^M \left[1 - e^{\left(-\frac{x}{\sigma_j}\right)^{w_j}} \right]^{n_j}, \quad (\text{III.5})$$

where M is the number of years in the record, n_j is the number of ordinary events in the j^{th} year, $\sigma_j > 0$, and $w_j > 0$ are the Weibull scale and shape parameter for the j^{th} year. The shape parameter w controls the tail decay of the distribution. The tail behavior is exponential when $w = 1$, heavier than exponential when $w < 1$, and fast decaying, while still retaining an infinite upper-end point when $w < 1$ (Gründemann et al., 2023). However, unlike the EVT distribution where the upper tail is completely controlled by the shape parameter ξ , the tail decay of the MEV is also controlled by the inter-annual variations of w_j , σ_j and n_j (Marra et al., 2018). The density of the MEV for the case when the parameters are time-invariant (Schellander et al., 2019), that is $w_j = w$, $\sigma_j = \sigma$ and $n_j = n$ is illustrated in Figure III.7.

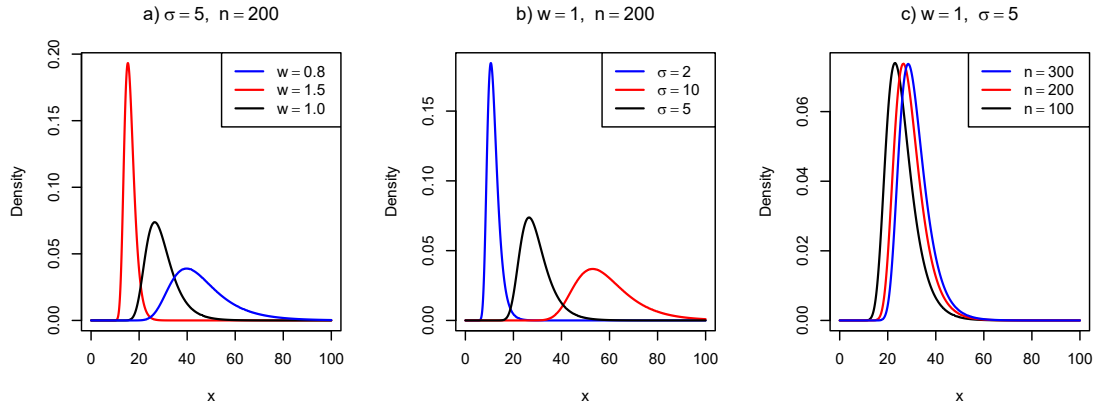


Figure III.7: Illustration of the density of MEV (Marani and Ignaccolo, 2015) for various values w , σ and n .

This approach allows more efficient use of the available data and has been shown to outperform GEV and GPD in estimation uncertainty for longer return periods (Zorzetto et al., 2016), and provided smoother spatial patterns of local extremes (Gründemann et al., 2023). Later extensions of the model were proposed for sub-daily scale (Marra et al., 2018), a simplified version, simplified MEV (SMEV), to account for mixture distributions (Marra et al., 2019), and a temporal MEV (TMEV) to account for sub-yearly evolution of distribution parameters (Falkensteiner et al., 2023). It is worth noting that, while the MEV distribution and its extensions use all the data, the interest is on having more robust estimates of the extreme quantiles rather than having a model for the full range of non-zero precipitation.

III.1.2.4 Extended GPD of Naveau et al. (2016)

Coming back to the class of GPD extension, Naveau et al. (2016) proposed a method that extends the framework of Papastathopoulos and Tawn (2013) to model the full range of non-zero precipitation, not just to improve the upper tail estimation. The delicate choice of the GPD threshold is entirely avoided and both the lower and the upper tails are modeled according to the GPD distribution while allowing a smooth transition between the two tails.

Similar to the approach of Papastathopoulos and Tawn (2013), the extension of the GPD is through probability integral transform. In particular, any continuous random variable can be generated by applying its inverse CDF to uniform draws. For instance, a GP random variable can be simulated from $Y = \sigma H_\xi^{-1}(U)$, where U is a uniform random variable on $[0,1]$. Additional flexibility is achieved by replacing U with $V = K^{-1}(U)$, where K is a continuous CDF that is also defined on $[0,1]$. The random variable $I = \sigma H_\xi^{-1}[K^{-1}(U)]$, is then said to be distributed according to the Extended Generalized Pareto distribution (EGPD) with CDF defined as

$$\mathbb{P}(I \leq i) = K \left[H_\xi \left(\frac{i}{\sigma} \right) \right]. \quad (\text{III.6})$$

The choice of K has to ensure that both the lower and upper tails of I are EVT compliant. Specifically, Naveau et al. (2016) defined some limiting constraints to ensure that EVT compliance of the two tails is achieved. The first constraint given in Equation III.7 ensures that I has an upper tail that is equivalent to the GP tail. If we denote $\bar{K} = 1 - K$ as the tail of K , then

$$\lim_{v \rightarrow 0} \frac{\bar{K}(1-v)}{v} = a, \text{ for some finite } a > 0. \quad (\text{III.7})$$

The second constraint in Equation III.8 ensures that the low precipitation (modeled as $-I$) also has a GPD tail of the Weibull type.

$$\lim_{v \rightarrow 0} \frac{K(v)}{v^s} = c, \text{ for some finite } c > 0 \text{ and } s > 0. \quad (\text{III.8})$$

The authors proposed four parametric families of K that satisfied the above constraints, the simplest of which is a power law, defined as $K(u) = u^\kappa$. Below we show how this model satisfies the constraints in the upper (Equation III.7) and lower (Equation III.8) tails.

Starting with the upper tail, we need to show that the ratio of the upper tails of the EGPD and GPD tails approaches a constant for very large i . That is:

$$\lim_{i \rightarrow \infty} \frac{\bar{K}[H_\xi(\frac{i}{\sigma})]}{\bar{H}_\xi(\frac{i}{\sigma})}.$$

By putting $v = \bar{H}_\xi(\frac{i}{\sigma})$, the limit becomes $\lim_{v \rightarrow 0} \frac{\bar{K}(1-v)}{v}$, and by applying l'Hopitals rule, and taking $K(u) = u^\kappa$, it becomes $\lim_{v \rightarrow 0} \kappa(1-v)^\kappa = \kappa$, for $\kappa > 0$. Hence, the constraint is achieved for a $\kappa > 0$.

Regarding the second constraint in the lower tail, Naveau et al. (2014) showed that under EVT assumption, low precipitation intensities might be adequately described by a power law of the form $\mathbb{P}(I \leq i) \approx \text{constant} \times i^s$, for any small $i \geq 0$ where $s \geq 0$ denotes the shape parameter.

For tail equivalence, the ratio of the tails for low precipitation intensities $\frac{G[H_\xi(i/\sigma)]}{i^s}$ needs to converge to a finite constant. The ratio can be expressed as

$$\frac{K \left[v \frac{H_\xi(v)}{v} \right]}{K(v)} \frac{K(v)}{v^s} \sigma^{-s},$$

with $v = i/\sigma$. Knowing that $\lim_{v \rightarrow 0} \frac{H_\xi(v)}{v} = 1$ from Taylor expansion, the ratio becomes $\sigma^{-s} \frac{K(v)}{v^s}$. Taking $K(u) = u^\kappa$ and applying l'Hôpital's rule, $\lim_{v \rightarrow 0} \frac{K(v)}{v^s}$ becomes $\lim_{v \rightarrow 0} \frac{K(v)}{v^s} = \sigma^{-s} \lim_{v \rightarrow 0} \frac{\kappa v^{\kappa-1}}{s v^{s-1}}$ which simplifies to $\sigma^{-s} \kappa > 0$ with $\kappa = s$. This is why κ controls the shape of the lower tail of this class of EGPD. The lower tail constraint here is also met by the Gamma distribution, and so this model behaves like the Gamma distribution in its lower tail.

The choice of $K(u) = u^\kappa$ therefore satisfies the constraints and ensures EVT compliance in both upper and lower tails. Under this formulation, the CDF of the EGPD is expressed as:

$$\mathbb{P}(I \leq i) = \left[H_\xi \left(\frac{i}{\sigma} \right) \right]^\kappa. \quad (\text{III.9})$$

The model has three parameters. $\kappa > 0$, $\sigma > 0$ is the scale parameter, and $\xi \geq 0$ controls the upper tail. $H_\xi \left(\frac{i}{\sigma} \right)$ is as defined in Equation III.3. With just one additional parameter, κ , the distribution is parsimonious and able to adequately model the full range of non-zero precipitation. The density of the model is shown in Figure III.8. This model has seen wide applications in the modeling of precipitation in several locations (see e.g. Blanchet et al., 2019; de Carvalho et al., 2021), and especially in our study area (see Evin et al., 2018; Le Gall et al., 2022). This is the model we use for the non-zero precipitation intensities, defined as $I = X|X > 0$, all along the thesis, and we will elaborate on this choice in Chapter IV.

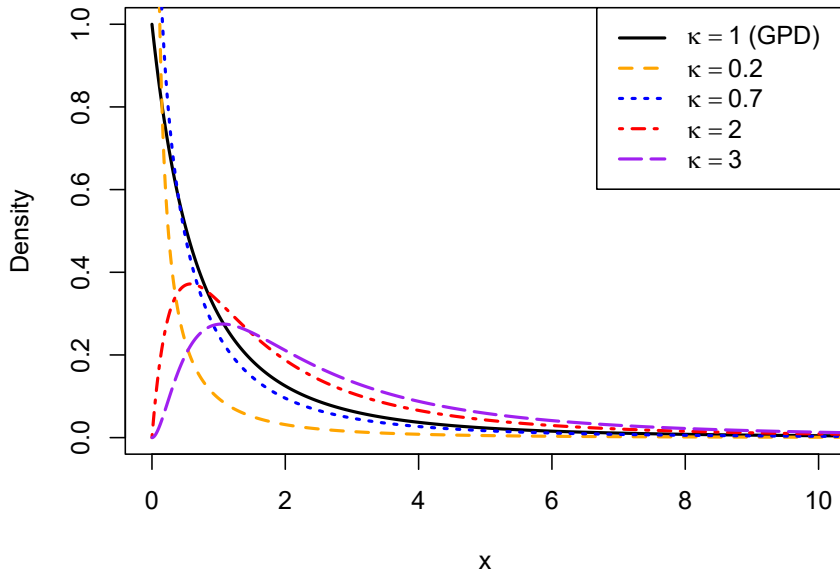


Figure III.8: Illustration of the EGPD of Naveau et al. (2016) based on power law defined as $K(u) = u^\kappa$. The density is shown for various values of κ . In each case, $\sigma = 1$, and $\xi = 0.5$.

III.2 Regionalization of at-site estimates of return level with EGPD

The main difficulty in extreme value analysis is the limited amount of data for model estimation. This is because extremes are scarce by definition and so return level estimates are marred by huge variance. The shape parameter ξ that controls the upper tail behavior in the extreme value distributions is difficult to estimate based on a few decades of data. Figure III.9 illustrates the sensitivity of theoretical return levels to the value of ξ . It can be seen that the 100-year return level for $\xi = 0.2$ is almost double the estimate when $\xi = 0$. Therefore a small error in the estimate of this parameter can lead to large changes in the return level estimates.

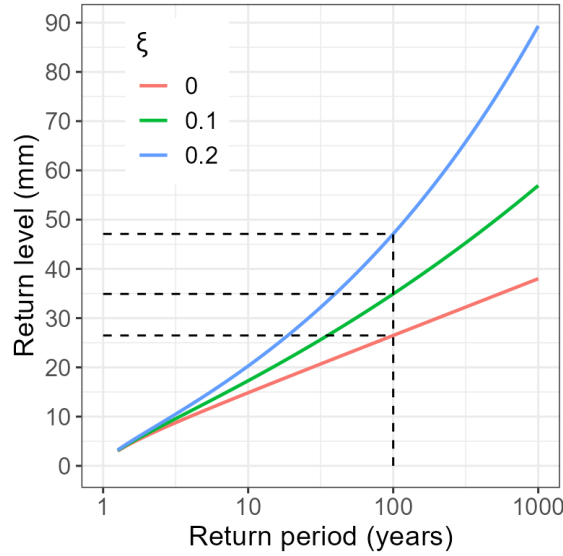


Figure III.9: Sensitivity of theoretical EGPD return levels to the values of the shape parameter ξ . The plot is shown for three values of the shape parameter ($\xi = 0, 0.1, 0.2$) with the scale parameter $\sigma = 5$ and $\kappa = 2$.

Since precipitation is spatial by nature, several studies (e.g. Cunnane, 1988; Burn, 1990; Hosking and Wallis, 2005) proposed the use of observations surrounding the local station to increase the quantity of data available for estimation, thereby reducing estimation uncertainty. In this thesis, we explore some regionalization approaches with the aim of improving the robustness and reliability of the at-site estimates of the EGPD distribution in Switzerland. We focus on two main approaches, the first is based on regional frequency analysis (RFA) that relies on the assumption of homogeneity of regions in terms of their precipitation distribution. The second approach is a spatial method that does not require regional homogeneity, but all the data is pooled in order to estimate the spatial surface of each of the model parameters. The core theoretical concepts of the different methods are presented in this section, while the results are shown in Chapter V.

III.2.1 RFA based on distinct homogenous regions

Regional frequency analysis is defined as the process of polling data from several sites to estimate the frequency distribution (Hosking and Wallis, 2005). The classical RFA method proposed in Hosking and Wallis (2005) involves the definition of homogeneous regions of hydrologically similar sites up to a scaling factor. To identify the regions, cluster analysis, in combination with some carefully chosen covariates such as geographical and atmospheric characteristics, is applied. Cluster analysis is usually performed using clustering algorithms such as K-means, hierarchical clustering, and partitioning around medoids (PAM) (see Kaufman and Rousseeuw, 2005, for de-

tails of these clustering methods). Statistical homogeneity tests are then applied to accept or reject the homogeneous regions. Thereafter, a chosen distribution is fitted to the scaled observations in the identified region, and all stations within this region share the same regional parameters. Station-specific parameters and quantiles can then be inferred by appropriate rescaling. This approach has been applied in several studies with a specific interest in the extremes using GEV distribution and GPD (e.g. [Gaál and Kyselý, 2009](#); [Malekinezhad and Zare-Garizi, 2014](#); [Deidda et al., 2021](#)).

While this classical RFA approach has seen wide applications in the literature, it has some major drawbacks which include the difficulty associated with covariate selection and the application of the homogeneity test. Another major issue is that the method results in contiguous homogeneous regions that are separated by distinct boundaries. This produces abrupt shifts in the return level estimates along the boundaries of the contiguous regions. Such discontinuities cannot be justified, since precipitation, except in cases such as orography, varies seamlessly in space. To circumvent some of these drawbacks, novel methods were proposed, and we present two of these approaches in the next two sections.

III.2.2 RFA based on region of influence approach (ROI)

RFA based on the region of influence method (ROI), as proposed in [Burn \(1990\)](#), is similar in concept to the classical RFA method. It, however, circumvents the drawback of having contiguous regions separated by distinct boundaries. Instead, a region of influence is assigned to each station and all the scaled observations in the identified ROI are used to estimate its regional parameters. This method has been applied by various authors (e.g. [Gaál et al., 2008](#); [Kyselý et al., 2011](#); [Carreau et al., 2013](#); [Evin et al., 2016](#); [Das, 2017, 2019](#))

The application of this method involves several choices. For instance, the choice of scale factor for normalization, the distance metric to define the closeness of each station to every other, the threshold of the radius to define the cutoff for the inclusion of stations into the ROI for a site, and the choice of homogeneity test. These choices have a considerable influence on the method application, so they have to be objectively decided. Different authors in the application of the method have explored some or all of these factors, starting from [Burn \(1990\)](#), to [Gaál et al. \(2008\)](#), and in [Evin et al. \(2016\)](#).

In this thesis, we follow the objectively selected steps and choices similar to [Evin et al. \(2016\)](#) in the application of the method in the Southeastern part of France, containing the Alps (similar to our case). The authors applied the method by considering POT (exceedances of a 70% quantile) of central rainfalls (largest observations in 3-day rainfall events) and on some distributions (Exponential, GPD, and Weibull). We apply the same procedure but on non-zero precipitation and the EGPD model. The application of this regionalization method on daily precipitation is presented and discussed in Chapter [V](#).

III.2.3 RFA based on upper tail behavior

This RFA method was developed by [Le Gall et al. \(2022\)](#) to circumvent some of the issues with the classical RFA method of [Hosking and Wallis \(2005\)](#), which involves covariate selections and the application of homogeneity tests. The method is based on a fast algorithm that defines distinct homogeneous regions relying on their upper tail similarity. It requires neither covariate selection nor the application of any homogeneity test. Only the precipitation data at hand is used, thereby avoiding the need for any covariate. More so, since the regions are delineated based on the similarity of their upper-tail behavior, they are inherently homogeneous, so the application of any homogeneity test is avoided.

To apply the method, the upper tail behavior for each station is summarized by ω , a ratio of probability-weighted moments (PWMs) which is defined as

$$\omega = \frac{3\alpha_2 - \alpha_0}{2\alpha_1 - \alpha_0} - 1, \quad (\text{III.10})$$

where α_j denotes the PWM of order j .

Le Gall et al. (2022) showed that ω summarizes the upper tail behavior of the data at hand. Stations with high values of ω have very intense extremes, and those with low values have less intense extremes. Additionally, they showed that within the EVT framework, ω depends mainly on the shape parameter ξ and the flexibility parameter κ of the EGPD, while being independent of the scale parameter σ . In the case of the GPD and GEV, it only depends on the ξ .

After ω is estimated at each station, the next task is to group the stations into clusters based on their estimated value of ω . From the definition of ω , all the stations within the same cluster are homogeneous in terms of their upper tail similarity. Grouping the stations can be achieved using a clustering algorithm such as K-means, hierarchical clustering, or partitioning around medoids (PAM). In the final step, data from all the stations belonging to the cluster are pooled to estimate the regional parameters according to an algorithm defined by the authors. In the case of the EGPD, one can choose to keep the regional ξ in each cluster, and estimate site-specific κ and σ , or keep both the regional ξ and regional κ and estimate a station-specific scale parameter σ . The former and the latter options are called "semi-regional" and "regional" approaches respectively (Le Gall et al., 2022).

The application of the method for regionalization of extreme precipitation is shown in Chapter V. It will be shown in the chapter that although the method does not use any covariate information, geographically coherent regions, with a few isolated cases, are produced. A drawback of the method is that similar to the classical regional frequency methods, when contiguous regions are formed, they result in abrupt shifts in the regional parameters along the boundaries. Moreover, inherent in cluster analysis, the decision on the optimal number of clusters requires some level of subjectivity. Furthermore, in an application of the method, Le Gall et al. (2022) discovered that the method tends to be "economical" in terms of the number of clusters, *i.e.* it usually results in the formation of only a few homogeneous regions (Le Gall et al., 2022).

III.2.4 Spatial method based on Generalized Additive Model (GAM)

The next method we consider for regionalization is a spatial method that is based on semi-parametric regression using Generalized Additive Models (GAM) (Hastie and Tibshirani, 1986). This is in contrast to the previous methods where regionalization is based on the homogeneity of normalized data or upper tail similarity.

Consider $I(x)$ the daily non-zero precipitation measured at a location in Switzerland, indexed by the covariate x such as longitude, latitude, or altitude. Let us assume that $I(x)$ follows the three-parameter EGPD with parameters $k(x)$, $\sigma(x)$, and $\xi(x)$, that is:

$$I(x) \sim \text{EGPD}[k(x), \sigma(x), \xi(x)], \quad (\text{III.11})$$

where x denotes some covariate and each of the model parameters depends on some form of x .

We assume that each of the three parameters has a smooth-varying relationship with x . The relationship is explained using GAM, represented by smoothing splines written as:


$$\eta_\alpha\{\alpha(x)\} = \sum_{j=1}^{J_\alpha} f_{j,\alpha}(x_{j,\alpha}), \quad (\text{III.12})$$

where $\alpha \in \{\kappa, \sigma, \xi\}$ and $\eta_\alpha(\cdot)$ is a link function. Although the parameters are defined such that $\kappa > 0$, $\sigma > 0$, and $\xi \geq 0$, we use the identity link for all of them. However, we imposed the necessary constraints during optimization to ensure that the parameters remained within the

desired bounds. J_α is the total number of smooth functions for the parameter α , $f_{j,\alpha}(\cdot)$ is the j^{th} smooth function for the parameter α . Each of the smooth functions is approximated in terms of basis function expansions:

$$f(x) = \sum_{k=1}^K \beta_k b_k(x), \quad (\text{III.13})$$

where β_k are the unknown coefficient to be estimated, and $b_k(x)$ are some known basis functions of dimension k . The basis functions can be for example cubic regression splines, thin plate splines, cyclic splines, etc (see [Wood, 2017](#), for other types).

Estimating the EGPD parameters involves simply estimating the coefficients of the relationships (β) in Equation III.13. In this thesis, we estimate β directly from the data using the `evgam`  package ([Youngman, 2020](#)) that we adapted for EGPD. Station-specific EGPD parameters can be obtained from the fitted surfaces, using the information on the covariates at the location.

Examples of applications of this spatial method can be found in [Chavez-Demoulin and Davison \(2005\)](#); [Blanchet and Lehning \(2010\)](#); [Youngman \(2019, 2020\)](#); [Carrer and Gaetan \(2022\)](#). This regionalization method makes the most efficient use of information since all the data in the study area is pooled to estimate the smooth surfaces. The application of this regionalization method on daily precipitation in our study area and a comparison with the other methods is presented in Chapter V.

III.3 Intensity-Duration-Frequency (IDF) relationships

Intensity-Duration-Frequency (IDF) curves are used by engineers and water practitioners for the hydrological designs and management of structures such as sewer lines, culverts, drains, dams, dykes, etc. IDF curves summarize the main characteristics of extreme point precipitation across different timescales, *i.e.* return level, the associated return period, and the duration of the event. An example of IDF curves is shown in Figure III.10. The x -axis represents the duration of a precipitation event, in this case from 1 hr to 72 hr, and the y -axis gives the intensities in mm/hr. Each parallel line represent events with the same probability of exceedance.

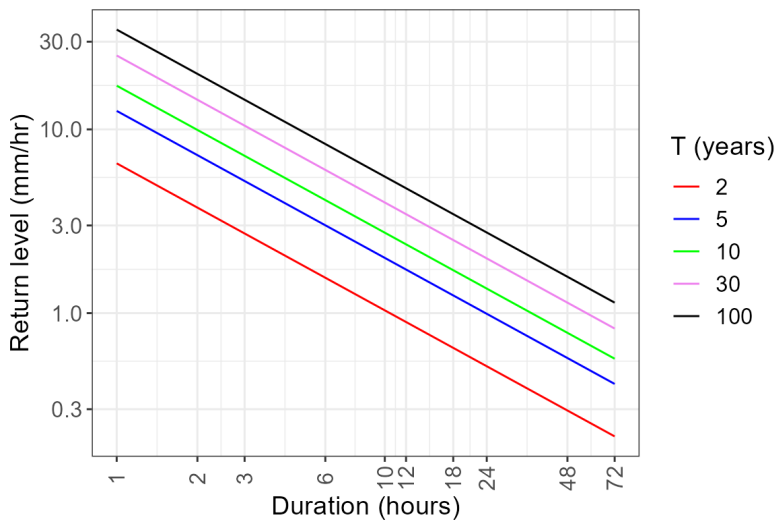


Figure III.10: Illustration of Intensity-Duration-Frequency (IDF) curves.

In this section, we will provide a brief introduction and theory behind the development of IDF curves. Results of the modeling of the IDF curves relationship for our study area are presented in Chapter VI.

Let us define the random variable $I(D)$ as the average non-zero precipitation intensity over the duration D . It is described by the CDF $F_D(i)$ such that $F_D(i) = \mathbb{P}(I(D) < i)$. The exceedance frequency is defined as $p_D(i) = 1 - F_D(i)$. The return period of any non-zero intensity i as a function of p_D is given by $T = \frac{1}{p_D \times \delta_D}$, where δ_D denotes the average number of non-zero precipitation intensities of duration D per year. Consequently, the T -year return level over duration D , $i(T, D)$, is defined as the $(1 - \frac{1}{T \times \delta_D})$ quantile of F_d . By assuming F_d to be the CDF of the EGPD, the IDF curve, $i(T, D)$ is expressed as:

$$i(T, D) = \frac{\sigma(D)}{\xi(D)} \left\{ \left[1 - \left(1 - \frac{1}{T \times \delta_D} \right)^{\frac{1}{\kappa(D)}} \right]^{-\xi(D)} - 1 \right\}, \quad (\text{III.14})$$

where $\kappa(D) > 0$, $\sigma(D) > 0$ and $\xi(D) \geq 0$ are the three EGPD parameters for the duration D , T is the return period in years, δ_D is the average number of non-zero precipitation intensities per year for the duration D and is estimated by the long term average.

The next task is to decide whether each of the three parameters; κ , σ , and ξ depends on D or not, and the exact form of the relationship. In the following subsections, we will present three approaches to determine the type of relationship. The first approach is based on precipitation scale-invariance, the second depends on the general formulation of Koutsoyiannis et al. (1998), and the third is a data-driven approach.

III.3.1 Scaling IDF curves

This approach is based on scale invariance. It has been shown that precipitation exhibits this property within some scales (e.g. Schertzer and Lovejoy, 1987; Gupta and Waymire, 1990, 1993; Over, 1995; Harris et al., 1997; Lima, 1998; Molnar and Burlando, 2008; Veneziano and Lepore, 2012; Paschalis, 2013). This property provides the physical justification for modeling IDF curves, and thus the possibility of inferring return levels of interest across scales. This approach is arguably the most commonly used approach, possibly because of its rich theoretical background, physical basis, and ease of application in regions with scarce availability of sub-daily precipitation series. IDF curves based on this approach can be found in several applications (e.g. Burlando and Rosso, 1996; Menabde et al., 1999; Willems, 2000; Van de Vyver and Demarée, 2010; Panthou et al., 2014; Blanchet et al., 2016; Innocenti et al., 2017; Sane et al., 2018; Mélése et al., 2018).

Scale invariance in the strict sense is defined by Gupta and Waymire (1990) as the property where the probability distribution of $I(D)$ can be inferred from the distribution of $I(D_0)$ at the reference duration D_0 through:

$$I(D) \stackrel{dist}{=} C_\lambda I(D_0), \quad (\text{III.15})$$

where the parameter C_λ determines the type of scaling; simple-scaling or multi-scaling.

A weaker assumption, the so-called "wide sense scaling", implied by Equation III.15, is when the scaling is in the moments according to:

$$\mathbb{E}[I(D)^q] = \left(\frac{D}{D_0} \right)^{-k(q)} \mathbb{E}[I(D_0)^q], \quad (\text{III.16})$$

where q is the order of the moment, $k(q)$ is called the moment scaling function, D_0 is the reference duration.

When the parameter in Equation III.15, $C_\lambda = \left(\frac{D}{D_0} \right)^{-H}$, *i.e.* a scalar that depends only on the ratio of the scales, we have "strict sense simple-scaling" expressed as:

$$I(D) \stackrel{dist}{=} \left(\frac{D}{D_0} \right)^{-H} I(D_0), \quad (\text{III.17})$$

where $0 < H < 1$ is the scaling exponent, and other variables retain their meanings.

Similar to Equation III.16, wide sense simple-scaling, implied by Equation III.17, is when the moment scaling function in Equation III.16 is linear in q , *i.e.* $k(q) = Hq$. This is expressed as:

$$\mathbb{E}[I(D)^q] = \left(\frac{D}{D_0} \right)^{-Hq} \mathbb{E}[I(D_0)^q]. \quad (\text{III.18})$$

Assuming that the random variable $I(D_0)$ follows the EGPD, that is $I(D_0) \sim \text{EGPD}(\kappa_{D_0}, \sigma_{D_0}, \xi_{D_0})$, and combining this with the assumption of simple-scaling in Equation III.17, we have:

$$\mathbb{P}[I(D) \leq x] = \mathbb{P}\left[I(D_0) \leq \left(\frac{D}{D_0} \right)^H x \right].$$

Putting this in terms of the CDF of the EGPD, we have:

$$\mathbb{P}[I(D) \leq x] = \mathbb{P}\left[I(D_0) \leq \left(\frac{D}{D_0} \right)^H x \right] = \left\{ 1 - \left[1 + \xi_{D_0} \frac{\left(\frac{D}{D_0} \right)^H x}{\sigma_{D_0}} \right]_+^{-1/\xi_{D_0}} \right\}^{\kappa_{D_0}} \quad \text{for } \xi_{D_0} \neq 0. \quad (\text{III.19})$$

This can be written as :

$$\mathbb{P}[I(D) \leq x] = \mathbb{P}\left[I(D_0) \leq \left(\frac{D}{D_0} \right)^H x \right] = \left\{ 1 - \left[1 + \xi_{D_0} \frac{x}{\left(\frac{D}{D_0} \right)^{-H} \sigma_{D_0}} \right]_+^{-1/\xi_{D_0}} \right\}^{\kappa_{D_0}} \quad \text{for } \xi_{D_0} \neq 0. \quad (\text{III.20})$$

Therefore, under the simple-scaling assumption, $I(D)$ follows the EGPD with parameters:

$$\begin{cases} \kappa(D) = \kappa_{D_0} \\ \sigma(D) = \left(\frac{D}{D_0} \right)^{-H} \sigma_{D_0} \\ \xi(D) = \xi_{D_0} \end{cases}.$$

By substituting these expressions in Equation III.14, IDF curves under simple scaling assumption can be expressed as

$$i(T, D) = \left(\frac{D}{D_0} \right)^{-H} \frac{\sigma_{D_0}}{\xi_{D_0}} \left\{ \left[1 - \left(1 - \frac{1}{T \times \delta_D} \right)^{\frac{1}{\kappa_{D_0}}} \right]^{-\xi_{D_0}} - 1 \right\}. \quad (\text{III.21})$$

An illustration of this model for $\kappa_0 = 2$, $\sigma_0 = 5$, $\xi_0 = 0.1$, and $H = 0.8$ is shown in left panel (Model A) of Figure III.11. It can be seen that the IDF curves are parallel to each other. We note here that the parallel curves are obtained by assuming that the parameter δ_D is the same across all durations. In practice, however, it decreases with duration, so the IDF relationship is not a separable function of T and D and the curves will not be parallel even with the simple scaling assumption. This is discussed in more detail in Chapter VI using real precipitation data in Switzerland.

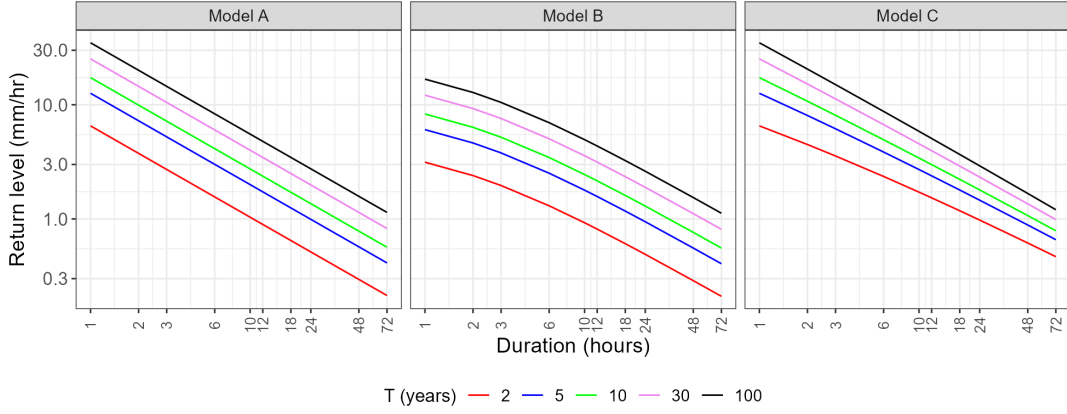


Figure III.11: Illustration of Intensity-Duration-Frequency (IDF) curves with the three approaches. Model A is based on the simple scaling assumption. Model B is based on the general IDF formulation and Model C is based on the data-driven approach.

III.3.2 General IDF formulation

This second approach is based on the general formulation of [Koutsoyiannis et al. \(1998\)](#), a generalization of the various empirical formulations for modeling IDF curves (e.g. [Sherman, 1931](#); [Bernard, 1932](#); [Chow, 1962](#); [Carreteras, 1987](#)). He showed that all of them can be simplified into the following form:

$$i(T, D) = \frac{a(T)}{b(D)}, \quad (\text{III.22})$$

where $b(D) = (D + \theta)^H$. The parameter $\theta > 0$ is the duration offset, and $0 < H < 1$ is the duration exponent. $a(T)$ is the $(1 - \frac{1}{T})$ quantile of the re-scaled intensities $I(T, D)b(D)$.

This formulation is consistent with both probabilistic theories and the physical constraints of scaling across duration. Several applications of this formulation to build IDF curves can be found in the literature (e.g. [Koutsoyiannis et al., 1998](#); [Van de Vyver and Demarée, 2010](#); [Blanchet et al., 2016](#); [Sane et al., 2018](#); [Ulrich et al., 2020](#); [Fauer et al., 2021](#); [Roksvåg et al., 2021](#)). This method can be seen as an extension of the scaling approach in which an additional parameter (θ) is added to allow for the curvature of the return levels curves for short durations ([Menabde et al., 1999](#)).

Under this framework, the EGPD parameters as a function of duration are

$$\begin{cases} \kappa(D) = \kappa_{D_0} \\ \sigma(D) = \left(\frac{D+\theta}{D_0+\theta}\right)^{-H} \sigma_{D_0} \\ \xi(D) = \xi_{D_0} \end{cases} .$$

By substituting these expressions in Equation III.14, IDF curves can be expressed as

$$i(T, D) = \left(\frac{D + \theta}{D_0 + \theta}\right)^{-H} \frac{\sigma_{D_0}}{\xi_{D_0}} \left\{ \left[1 - \left(1 - \frac{1}{T \times \delta_D}\right)^{\frac{1}{\kappa_{D_0}}} \right]^{-\xi_{D_0}} - 1 \right\}. \quad (\text{III.23})$$

An illustration of this model for $\kappa_0 = 2$, $\sigma_0 = 5$, $\xi_0 = 0.1$, $H = 0.8$, and $\theta = 1.5$ is shown in middle panel (Model B) of Figure III.11. It can be seen that θ introduces some curvature for short durations, while the longer durations are not affected. The application and performance of this model are also presented in Chapter VI.

III.3.3 Data-driven IDF curves

The third approach is based on [Overeem et al. \(2008\)](#) and is called the "data-driven" approach. The term "data-driven" is used to show that the functional relationship (linkage) between IDF model parameters and duration is empirically determined from the data itself. The method involves fitting a parametric model, for example, EGPD, to data of each duration. A particular regression model is then fitted for each parameter as a function of duration. As a consequence, the return level of any duration can be inferred from the inverse of the distribution, with parameters obtained from the regression model. This approach does not impose the assumption/existence of scaling. Interestingly, both the two previous approaches can be seen as special cases of the data-driven approach with particular functional relationships imposed on the parameters. We note that, although the other two approaches also consider the link between parameter and duration, the specific forms rely on some underlying theoretical or physical hypothesis. However, with the data-driven method, the functional relationship is empirically determined from the data itself, hence its name.

To illustrate IDF curves with this approach, we assume the following parametric models for the three EGPD parameters:

$$\begin{cases} \kappa(D) = \left(\frac{D}{D_0}\right)^{-b_\kappa} a_\kappa \\ \sigma(D) = \left(\frac{D}{D_0}\right)^{-b_\sigma} a_\sigma \\ \xi(D) = a_\xi + b_\xi \log(D). \end{cases} .$$

IDF curves are shown in the right panel (Model C) of [Figure III.11](#) for $a_\kappa = 2$, $b_\kappa = -0.4$, $a_\sigma = 5$, $b_\sigma = 0.8$, $a_\xi = 0.1$, $b_\xi = -0.02$. It can be seen that, even with δ_D constant, the IDF curves are not parallel, they rather scale differently.

We comment here that unlike in [Overeem et al. \(2008\)](#) where the coefficients in the parametric model of each GEV parameter were fitted separately, we extend his approach by fitting our distribution (EGPD) with parameters as a function of D . Application on real data and further discussions are presented in [Chapter VI](#).

III.4 Intensity Duration Area Frequency (IDAF) relationships

The IDF relationships presented in the previous section do not account for the spatial extent of precipitation, that is area (A). Accounting for the spatial component is essential, especially in hydrological applications, since watersheds and river basins integrate the precipitation falling over their respective areas, leading to the accumulation of runoff and subsequent flood generation.

When this spatial component is accounted for, we have the so-called Intensity-Duration-Area-Frequency (IDAF) relationships. They provide the mathematical link between precipitation intensities (I), durations (D), areas (A), and frequency of occurrence (F). They summarize the main statistical characteristics of extreme precipitation (return level, return period, duration, and area) and are useful tools for engineers and hydrologists in hydrological design (e.g. [Bertini et al., 2020](#)), quantification of areal precipitation hazard (e.g. [Overeem et al., 2010](#); [Panthou et al., 2014](#); [Mélèse et al., 2019](#); [Zhao et al., 2023](#)), storm characterization (e.g. [Ramos et al., 2005](#); [Ceresetti et al., 2012](#); [Blanchet and Mélèse, 2020](#)), and development of early warning systems (e.g. [Panziera et al., 2016](#)).

Let us define $I(D,A)$ as the random variable of non-zero precipitation intensities for a given duration D and area A . We further assume that $I(D,A)$ follows the EGPD, *i.e.*:

$$I(D,A) \sim \text{EGPD}[\kappa(D,A), \sigma(D,A), \xi(D,A)], \quad (\text{III.24})$$

where $\kappa(D,A) > 0$, $\sigma(D,A) > 0$ and $\xi(D,A) \geq 0$ are the three EGPD parameters for the duration D and area A .

Let $F_{D,A}(i)$ be the CDF of $I(D,A)$, such that $F_{D,A}(i) = \mathbb{P}(I_{D,A} < i)$, then the IDAF curve, which is the T -year return level over duration D and area A is defined by the quantile function of $F_{D,A}$, *i.e.*:

$$i(T,D,A) = \frac{\sigma(D,A)}{\xi(D,A)} \left\{ \left[1 - \left(1 - \frac{1}{T \times \delta_{D,A}} \right)^{\frac{1}{\kappa(D,A)}} \right]^{-\xi(D,A)} - 1 \right\}, \quad (\text{III.25})$$

where $\kappa(D,A) > 0$, $\sigma(D,A) > 0$ and $\xi(D,A) \geq 0$ are the three EGPD parameters for the duration D and area A . T is the return period in years, $\delta_{D,A}$ is the average number of non-zero precipitation intensities for duration D and area A per year.

We illustrate a conceptual plot of IDAF curves in Figure III.12. A plot of IDAF curves is 3-dimensional (Figure III.12a), with intensity (I) along the vertical axis, duration (D) along the horizontal axis, and area (A) along the third axis which is perpendicular to the other two axes. For each specific return period (*e.g.* 2-year, 10-year, or 50-year), a curve is shown to visualize how the intensity changes across A and D . However, a much simpler approach is to decouple the 3-dimensional plot into two sub-plots, each in 2-dimension. The first one shows how the intensities of specific return periods change across durations for a fixed area, *i.e.* IDF curves (Figure III.12b), and the second one, a plot of Intensity-Area-Frequency (IAF) curves, shows how the intensities change across areas for a fixed duration (Figure III.12c).

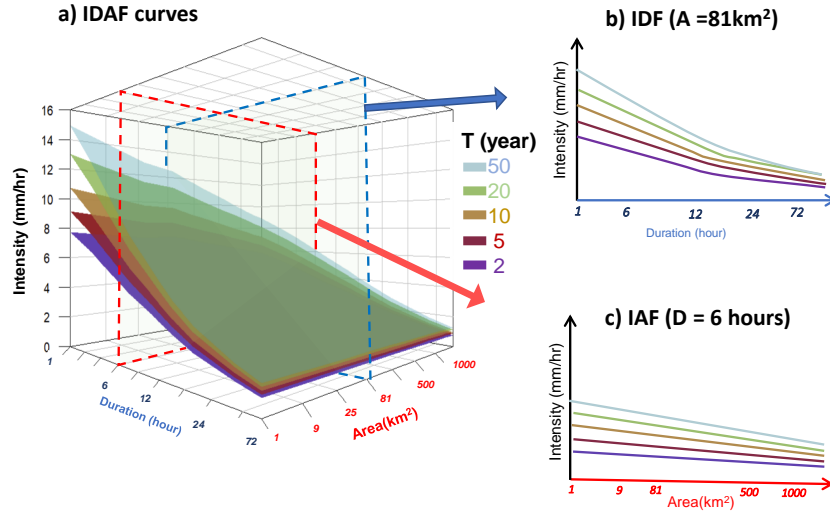


Figure III.12: a) Conceptual illustration of IDAF curves in 3-dimension. IDF curves for $A = 81 \text{ km}^2$ (shown in panel b) are obtained by cutting a plane on the IDAF curves in panel a at $A = 81 \text{ km}^2$ (red-colored broken lines). The IAF curves on panel c) are obtained by cutting a plane at $D = 6 \text{ hr}$ on panel a (blue-colored broken lines).

The literature provides two main methods of modeling IDAF relationships, namely areal reduction factor (ARF) and data-driven IDAF models which are presented in the following subsections.

III.4.1 ARF-based IDAF models

ARF-based IDAF curves are built by coupling IDF curves and ARF. That is

$$i(T, D, A) = i(T, D) \text{ARF}(T, D, A), \quad (\text{III.26})$$

where $i(T, D, A)$ is the IDAF curves, $i(T, D)$ is the IDF curves obtained from Equation III.14 for example, and $0 < \text{ARF}(T, D, A) < 1$ is the areal reduction factor. An example of ARF (Koutsoyiannis and Xanthopoulos, 1999) is given in Equation III.27 for example for the case when ARF is independent of T .

$$\text{ARF}(D, A) = 1 - \frac{0.048 \cdot A^{0.36 - 0.01 \ln(A)}}{D^{0.35}}, \quad (\text{III.27})$$

where A is the area in Km^2 and D is the duration in hr.

Applications of the ARF-based IDAF models can be found in the literature, for example, De Michele et al. (2001) derived an ARF formulation based on the concept of dynamic scaling of precipitation and used it to model IDAF curves in Milan. Later, Ceresetti et al. (2012) used the ARF of De Michele et al. (2001) to model IDAF curves for storm severity assessment in southern France. Panthou et al. (2014) also applied the same ARF formulation to build IDAF models and to characterize areal precipitation in West Africa. Ramos et al. (2005) introduced an empirical ARF formulation to model IDAF curves for storm severity assessment in Marseille. Bertini et al. (2020) employed the ARF formulation in Equation III.27 to build IDAF curves for dam design in Italy. Mèlèse et al. (2019) and Blanchet and Mèlèse (2020) applied an extension of the ARF formulation of De Michele et al. (2001) to build IDAF curves respectively for areal hazards and storm severity assessment in southern France. The extension was to cope with the significant spatio-temporal variability in the mountainous area.

The ARF formulation given in Equation III.27 is one of many examples in the literature, but there are several methods of estimating ARF, ranging from hypothesis-based analytical methods to purely empirical methods (see Svensson and Jones, 2010, for a review). The analytical methods are developed based on some theoretical frameworks that rely on some hypothesis on the relationship between point precipitation and areal precipitation. Examples of such ARF models are those based on the spatial correlation of precipitation (Rodriguez-Iturbe and Mejía, 1974; Sivapalan and Blöschl, 1998), crossing properties (e.g. Bacchi and Ranzi, 1996), or dynamic-scaling and multi-fractal analysis (e.g. De Michele et al., 2001; Veneziano and Langousis, 2005). Empirical methods do not make any distributional assumption on the precipitation process but rely on the observed relationship between point precipitation extremes and areal precipitation extremes. Examples of ARF based on this approach can be found in Svensson and Jones (2010).

ARF generally increases with duration, since events of short duration (e.g. convective events) are more localized and so ARF decreases faster with the area, while those of longer durations extend over a larger area (e.g. frontal events) and so the ARF is less steep. In the same argument, ARF decreases with an increase in area for a given duration. More so, ARF decreases with the return period (frequency of occurrence) since weak precipitation events can extend over large regions, leading to flatter ARF curves, while intense events are more localized, leading to ARF curves that are steeper. A common assumption however is to treat ARF as independent of T (e.g. Koutsoyiannis and Xanthopoulos, 1999; De Michele et al., 2001; Svensson and Jones, 2010; Panthou et al., 2014). This general behavior of ARF depends on the season, catchment characteristics, and the type of precipitation event (Svensson and Jones, 2010) and a strong deviation from this general behavior (e.g. $\text{ARF} > 1$) can be observed in mountainous regions (e.g. Figure 2 of Mèlèse et al., 2019).

III.4.2 Data-driven IDAF models

Beyond the ARF-based IDAF curves modeling approach, [Overeem et al. \(2010\)](#) proposed a purely data-driven approach to model IDAF curves. This involves modeling the parameters of the statistical distribution of the precipitation intensities as functions of duration and area. The type of relationship is empirically determined from the data, with no underlying physical hypothesis such as spatial correlation (as done in [Rodríguez-Iturbe and Mejía, 1974](#)) or scaling (as performed in [De Michele et al., 2001](#)). This can be written as

$$\eta_{\alpha}(\alpha(D,A)) = f_{\alpha}(D,A), \quad (\text{III.28})$$

where $\alpha \in \{\kappa, \sigma, \xi\}$ is a given EGPD parameter, $\eta_{\alpha}(\cdot)$ is a link function, $f_{\alpha}(D,A)$ is a function of D and A (see Equation [VII.5](#) for example).

After fitting the relationships in Equation [III.28](#), the IDAF curves $i(T,D,A)$ can be obtained from Equation [III.25](#) by appropriate substitutions.

In addition, ARF for a given D , A , and T can be estimated from

$$\widehat{\text{ARF}}(T,D,A) = \frac{\hat{i}(T,D,A)}{\hat{i}(T,D,A_0)}, \quad (\text{III.29})$$

where $\hat{i}(T,D,A)$ and $\hat{i}(T,D,A_0)$ denote the T -year return levels of areal and point precipitation respectively, obtained from the fitted IDAF model. A_0 usually corresponds to the raingauge surface area, however, to avoid extrapolation of the relationships in Equation [III.28](#), A_0 can be considered as the smallest A used to fit the relationships. For instance, $A_0 = 1 \text{ km}^2$ in our case where we use CPC data.

As highlighted by [Mélèse et al. \(2019\)](#), this approach has the advantage of being flexible and applicable in cases where the assumptions of the analytical ARF formulations cannot be verified. Results of modeling IDAF relationships in the study area using the data-driven approach are presented and discussed in Chapter [VII](#).

III.5 Evaluation framework

The thesis involves comparing different statistical models and to ensure objective comparison, we need an appropriate evaluation framework and suitable criteria. In this section, we briefly introduce the framework and the criteria that we used to compare the models.

We apply the evaluation framework and the criteria proposed in [Garavaglia et al. \(2011\)](#) and [Renard et al. \(2013\)](#) which involves the use of split-sampling cross-validation to assess the predictive performance of a given model in an independent validation data set. A justification for this approach in comparison to the classical statistical goodness of test fits is that the latter lacks the ability to assess the model's ability to predict unobserved values and that they are also not very efficient for three-parameter distributions (see [Garavaglia et al., 2011](#), and the references therein).

III.5.1 Split-sampling cross validation

In the split-sampling cross-validation, we consider each station and divide the time series into two subsamples of the same length but on different randomly chosen years. We then fit each of the competing models on the first sub-sample and evaluate the predictive performance on the second sub-sample (validation sample). The evaluation is carried out using some well-chosen criteria that are described in the next section. We repeat the same procedure on the second sub-sample (use

it to train the models and validate using the first sub-sample). Since the split sampling is done randomly, we repeat the procedure several times to address sampling bias. We apply the same procedure at the other stations in the study area and select the model that has the best regional performance (average of the scores over all the stations).

III.5.2 Comparison criteria

The criteria are chosen to ensure that the model is robust and reliable in the prediction of extremes as well as the bulk of the distribution. A robust model yields similar estimates of a high quantile when the calibration sample is changed, and a reliable model assigns the correct exceedance probability to high values. Robustness is measured by SPAN (Equation III.30), reliability is measured by FF (Equation III.31), and the accuracy/reliability of the model in the bulk of the distribution is measured by NRMSE_CV (Equation III.32). An illustration of the three criteria is shown in Figure III.13, while their expressions are given below:

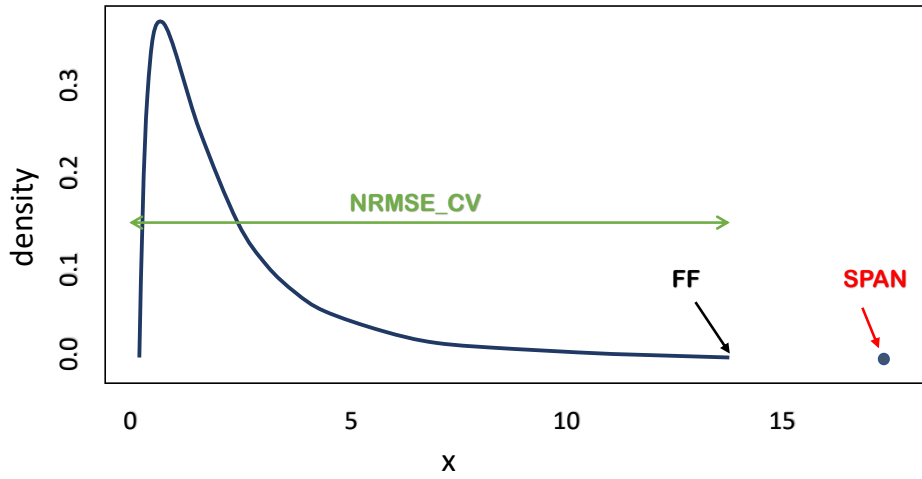


Figure III.13: Conceptual illustration of the three comparison criteria, SPAN, FF, and NRMSE_CV. The green line spans from the left to the right tail, indicating that NRMSE_CV is applied to all the quantiles. FF applies only to the maximum value, and SPAN is applied at extrapolation.

- Robustness: The Robustness criteria SPAN is computed as:

$$\text{SPAN}_{s,T} = \frac{2 \left| \hat{i}_{s,T}^{(1)} - \hat{i}_{s,T}^{(2)} \right|}{\hat{i}_{s,T}^{(1)} + \hat{i}_{s,T}^{(2)}}, \quad (\text{III.30})$$

where $\hat{i}_{s,T}^{(1)}$ and $\hat{i}_{s,T}^{(2)}$ are the T -year return levels estimated from sub-sample 1 and 2 respectively at station s . A SPAN of 0.5 means that the absolute difference between the two return levels is half of their average. A regional value of SPAN in the study area is computed as $\text{SPAN}_{\text{reg},T} = 1 - \frac{1}{N} \sum_{s=1}^N \text{SPAN}_{s,T}$, where N is the total number of stations. A perfectly robust model should have $\text{SPAN}_{\text{reg},T} = 1$. SPAN is usually computed to test the models at extrapolation (e.g. $T = 100$ years, 1,000 years)

- Reliability in predicting the maximum value: The FF criteria is calculated from:

$$FF_s^{(12)} = \left\{ \hat{F}_s^{(1)} \left[\max(i_s^{(2)}) \right] \right\}^{n_s^{(2)}}, \quad (\text{III.31})$$

where $FF_s^{(12)}$ is the cross-validation criteria computed at station s , by predicting the probability of the maximum value in sub-sample 2, of sample size $n_s^{(2)}$ using the model $\hat{F}_s^{(1)}$ fitted on sub-sample 1. $FF_s^{(21)}$ is computed symmetrically.

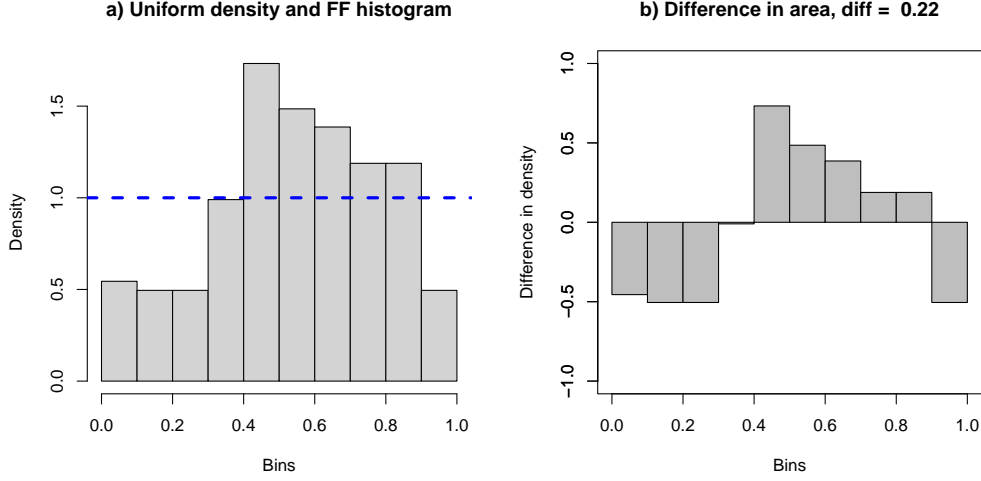


Figure III.14: Illustration of the FF criteria. Panel a) Shows the histogram of the N values of FF, and the theoretical uniform density (dash-blue line) while the difference between the two densities is shown in panel b). The difference in area ($diff$) = 0.22 in this case.

As noted by Renard et al. (2013) and Blanchet et al. (2015), if the fitted model is a good estimate of the true distribution of the data, $FF_s^{(12)}$ should be a realization of a uniform distribution. Hence, the difference in the area, noted $diff$, between the density of a theoretical uniform and that of the N values of $FF_s^{(12)}$ (computed over the N stations), should be close to zero. An illustration of $diff$ is shown in Figure III.14. FF_{reg} at the regional scale, given as $1 - diff$, should therefore take a value of 1 for a reliable model and 0 for a completely unreliable model; the lower the value the less reliable the model is.

- The reliability/accuracy over the entire observations: While the previous scores, FF and SPAN, focus on extremes only, it is important that the model is also reliable in the bulk of the distribution, especially given that we use the EGPD. To measure the reliability of a model in predicting all the observations in cross-validation, we use the normalized root mean square error (NRMSE_CV), which is expressed as:

$$NRMSE_CV_s^{(12)} = \frac{\left[\frac{1}{n_s^{(2)}} \sum_{j=1}^{n_s^{(2)}} \left(i_{s,T_j}^{(2)} - \hat{i}_{s,T_j}^{(1)} \right)^2 \right]^{1/2}}{\hat{i}_s^{(2)}}, \quad (\text{III.32})$$

where $NRMSE_CV_s^{(12)}$ is the score computed at station s , $n_s^{(2)}$ is the sample size of the second sub-sample, $i_{s,T_j}^{(2)}$ is the empirical quantile with return period $T_j = \frac{n_s+1}{j \times \delta}$, δ is the average number of non-zero precipitations per year in sub-sample 2, $\hat{i}_{s,T_j}^{(1)}$ is the corresponding T_j return level estimated from the model fitted on sub-sample 1. The denominator is

the mean of non-zero precipitation in sub-sample 2 at station s computed as $\frac{1}{n_s^{(2)}} \sum_{j=1}^{n_s^{(2)}} i_{s,T_j}^{(2)}$.

Similar to the other criteria, the regional score is computed from $\text{NRMSE_CV}_{\text{reg}}^{(12)} = 1 - \frac{1}{N} \sum_{s=1}^N \text{NRMSE_CV}_s^{(12)}$, and $\text{NRMSE_CV}_{\text{reg}}^{(21)}$ is computed symmetrically. $\text{NRMSE_CV}_{\text{reg}} = 1$ indicates a perfectly accurate model (the model accurately predicts the empirical return levels).

This evaluation framework and criteria are used throughout the study and adapted to each case. They are applied to compare regionalization models (Chapter [V](#)), models of IDF curves (Chapter [VI](#)), and models of IDAF curves (Chapter [VII](#)).

IV

Modeling precipitation intensities with the EGPD in Switzerland

Chapter overview

This chapter answers Question 1, by evaluating the ability of the extended generalized Pareto distribution (EGPD) of [Naveau et al. \(2016\)](#) to model the hourly and daily precipitation intensities in the study area. We also justify the choice of the simplest family, the three-parameter model based on the power law, over more complex candidates. A delicate issue that has to do with the lower censoring threshold in the application of the model is also discussed.

Contents

IV.1	Parametric families of the EGPD	42
IV.2	Comparison of the EGPD families using daily data	43
IV.2.1	Comparison criteria	43
IV.2.2	Inference	44
IV.2.3	Results	45
IV.3	Left-censored maximum likelihood estimation	48
IV.4	Modeling extreme daily and hourly precipitation with the EGPD	52
IV.4.1	Daily scale	52
IV.4.2	Hourly scale	53
IV.5	Summary	55

IV.1 Parametric families of the EGPD

Recall from Chapter III that the CDF of the EGPD is expressed as $\mathbb{P}(I \leq i) = K \left[H_\xi \left(\frac{i}{\sigma} \right) \right]$, and the exact form depends on the choice of the function K . The constraints in the choice of K to ensure that both lower and upper tails are EVT compliant were presented in Equation III.7 and III.8. Using these constraints, Naveau et al. (2016) proposed four parametric families for K written as:

1. Model 1: $K(u) = u^\kappa$ with $\kappa > 0$. The parameter κ controls the shape of the lower tail. The model therefore has a total of three parameters: σ , ξ , and κ . The density function for various values of κ is shown in Figure IV.1. When $\kappa > 1$ (red and purple), the density is zero at the origin, when $\kappa < 1$ (blue and yellow) the density is infinite at the origin, and when $\kappa = 1$ (black) the model reduces to the GPD. It can be observed that the smaller the value of κ , the more the density is concentrated near zero.
2. Model 2: $K(u) = pu^{\kappa_1} + (1-p)u^{\kappa_2}$ with $\kappa_2 \geq \kappa_1 > 0$ and $p \in [0, 1]$. This model increases the flexibility of Model 1 by adopting a mixture of power laws. The lower tail is controlled by κ_1 while κ_2 modifies the shape of the density in its central part. The model therefore has a total of five parameters: σ , ξ , p , κ_1 , and κ_2 . The density function for various values of κ_1 and κ_2 is shown in Figure IV.1. In each case, $p = 0.5$. Similar to the case of Model 1, when $\kappa_1 > 1$ (red and purple), the density is zero at the origin, when $\kappa_1 < 1$ (blue and yellow) the density is infinite at the origin, and when $\kappa_1 = \kappa_2 = 1$ (black) the model reduces to the GPD.
3. Model 3: $K(u) = 1 - Q_{\delta_1} \left[(1-u)^{\delta_1} \right]$ with $\delta_1 > 0$ where Q_{δ_1} is the CDF of a Beta random variable with parameters $1/\delta_1$ and 2, that is:

$$Q_{\delta_1}(u) = \frac{1 + \delta_1}{\delta_1} u^{1/\delta_1} \left(1 - \frac{u}{1 + \delta_1} \right).$$

The model has three parameters: σ , ξ , and δ_1 . The additional parameter δ_1 increases the flexibility in the central part of the distribution. The authors noticed a drawback of this model due to the imposition of a behavior of type χ^2 at the lower tail. This means that the shape of the lower tail is not determined by the data but imposed by the model. This point is illustrated in Figure IV.1, where the density is always zero at the origin. The central part is however modified by the various values of δ_1 . When $\delta_1 \rightarrow \infty$ (black) the model reduces to the GPD.

4. Model 4: $K(u) = \left\{ 1 - Q_{\delta_1} \left[(1-u)^{\delta_1} \right] \right\}^{\delta_2/2}$ with $\delta_1, \delta_2 > 0$. This model extends model 3 by one additional parameter δ_2 to allow flexibility in the lower tail. Hence δ_2 controls the lower part of the distribution, while δ_1 controls the central part. The model thus has a total of four parameters: σ , ξ , δ_1 , and δ_2 . The density of the model is shown in Figure IV.1, and it can be seen that flexibility is now achieved at the lower tail for various values of δ_2 . Model 3 is regained when $\delta_2 = 2$, while GPD is regained when $\delta_2 = 2$ and $\delta_1 \rightarrow \infty$.

In summary, all four models have two parameters in common: scale parameter σ and GPD shape parameter ξ . They however have additional parameters, ranging from one to three, for additional flexibility to model the whole distribution. The authors compared the performance of the four models on hourly data in Lyon, France, and found out that Model 1 with three parameters outperformed the other models in terms of the Akaike information criteria (AIC) (Akaike, 1974). On the other hand, Model 3 with equally three parameters, showed the worst performance due to its lack of flexibility, resulting from the constraint imposed on the lower tail of the distribution.

We comment here that while all the four models presented above are based on parametric forms for K , the choice of K doesn't have to be parametric. For instance, Tencaliec et al. (2020)

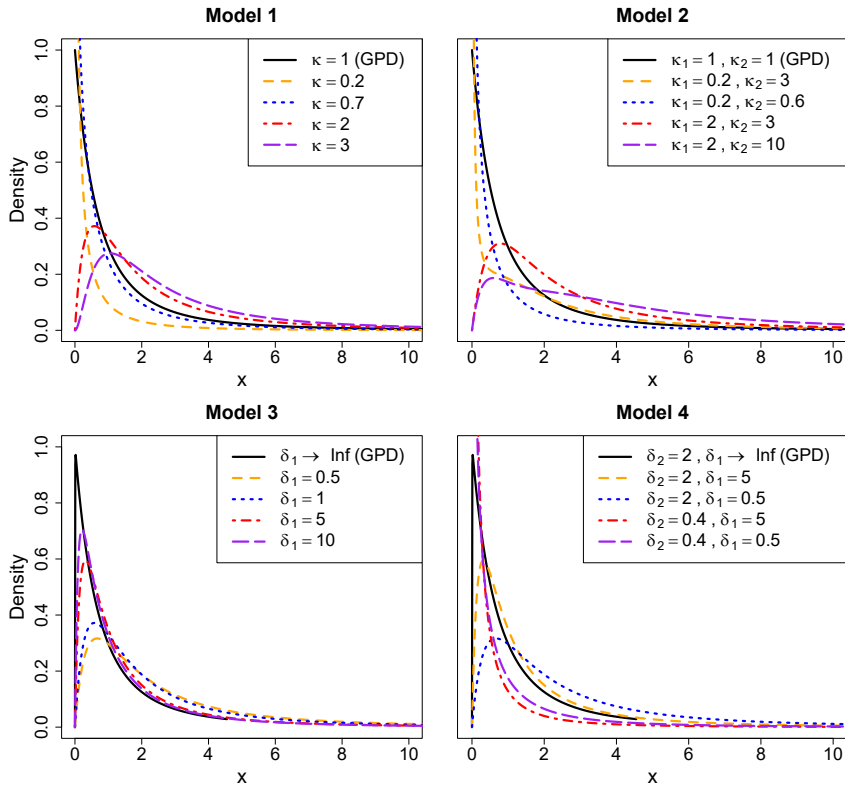


Figure IV.1: Density function corresponding to the four EGPD families. In each case $\sigma = 1$, and $\xi = 0.5$.

introduced a semi-parametric choice of K based on Bernstein polynomial approximations. We however limit our applications to the parametric families because of our interest in modeling parametric IDF and IDAF relationships, where each EGPD parameter can be modeled as a function of duration (for IDF) or duration and area (for IDAF relationships).

IV.2 Comparison of the EGPD families using daily data

In this section, we compare the family of EGPD models presented in the preceding section using daily data in our study area. We will however only consider Models 1, 2, and 4, since Model 3 was shown by the authors to lack adequate flexibility in the lower tail. In the remaining of the chapter, we will refer to the three models as EGPD-Power (Model 1), EGPD-Mixtures (Model 2), and EGPD-Beta (Model 4). We consider a total of 500 daily stations in Switzerland with variable record lengths ranging from 20 years to 156 years. The reader is referred back to Chapter II for a detailed description of the data. Due to the apparent seasonality of precipitation in the study area, the data is divided into four seasons, winter (DJF), spring (MAM), summer (JJA), and autumn (SON). The three models are then fitted on the seasonal precipitation data and their performance is compared.

IV.2.1 Comparison criteria

We compare the performance of the models using three criteria described below:

1. Normalized root mean square error (NRMSE):

$$\text{NRMSE} = \frac{\left[\frac{1}{n} \sum_{j=1}^n (i_{p_j} - \hat{i}_{p_j})^2 \right]^{1/2}}{\bar{i}}, \quad (\text{IV.1})$$

where n is the sample size, i_p is the empirical p -quantile of non-zero precipitation, \hat{i}_p is the corresponding p -quantile estimated from the fitted model, and the denominator \bar{i} is the mean of non-zero precipitation. The normalization is done to enable the comparison of the score across the stations. NRMSE gives an aggregate measure of the goodness-of-fit of the model, with emphasis on the largest quantiles. A perfect model should have a score = 0.

2. Relative bias (rBias):

$$\text{rBias}_p = \frac{i_p - \hat{i}_p}{i_p}, \quad (\text{IV.2})$$

where i_p is the empirical p -quantile of non-zero precipitation and \hat{i}_p is the corresponding estimate from the fitted model. We focus on some specific probabilities ($p = 0.01, 0.1, 0.25, 0.5, 0.75, 0.9, 0.99, 0.999$), to assess the bias across the entire distribution. For an unbiased model, $\text{rBias} = 0$. A positive value indicates that the model is underestimating the empirical quantiles, while a negative value indicates overestimation.

3. Akaike Information Criteria (AIC) ([Akaike, 1974](#)):

$$\text{AIC} = -2 \log(L) + 2n_{param}, \quad (\text{IV.3})$$

where L is the maximized likelihood in Equation IV.4 and n_{param} is the number of parameters to be estimated. AIC rewards the goodness-of-fit, as measured by the likelihood, but also penalizes model complexity in terms of the number of parameters to be estimated. The lower the AIC, the better the model.

IV.2.2 Inference

Two classical inference methods for fitting the models were proposed by the authors. The first one is based on maximum likelihood estimation (MLE), and the second one is based on the probability-weighted moments (PWM) (which are more robust compared to classical moments). The authors found out that the MLE has slightly better performance compared to the PWM approach for simulated data. However, the MLE performed poorly in the case of real data (hourly rainfall in Lyon, France). More specifically, the shape parameter ξ and the corresponding return level estimates were grossly overestimated by the MLE. The authors explained that the poor performance of the MLE was likely due to the effect of data discretization caused by instrumental precision. Since the likelihood applies to each observation, it is affected by the numerous small values close to zero. PWM is however based on summary statistics of the data rather than the individual observations, and so less sensitive to misspecified settings. To improve the parameter inference, the two methods, MLE and PWM, were modified to account for left-censoring and interval-censoring of the data (see [Naveau et al., 2016](#), for details). The inference with the left-censoring approach led to a remarkable performance improvement with the MLE, while the PWM showed only a marginal improvement.

Despite the higher robustness of the PWM compared to MLE, the latter has a key advantage since it can be used in regression settings to account for various cases of non-stationarity, such as temporal, spatial, or spatio-temporal non-stationarity. The non-stationarity is usually achieved by modeling the parameters as a function of some chosen covariates. This is a fundamental requirement in our context because the approach we will use for the development of IDF and IDAF

relationships requires modeling the distribution parameters as a function of covariates, specifically duration, and area. We note however that some attempts have been made to use covariates with PWM, for instance, Ribereau et al. (2008) extended the PWM to model non-stationary GEV. However, a drawback of their approach is that non-stationarity is only possible in the location parameter while the scale and shape parameters have to remain constant. Carreau et al. (2017) allowed the non-stationarity in the GPD scale parameter within homogeneous regions, but the shape parameter had to remain constant. Both approaches therefore involve constraints that limit their general applicability, since non-stationarity, especially in space, can be observed in the scale parameter (e.g. Chavez-Demoulin and Davison, 2005; Davison et al., 2012; Youngman, 2019) and shape parameter (e.g. Chavez-Demoulin and Davison, 2005; Blanchet and Lehning, 2010; Youngman, 2019). With MLE on the other hand, non-stationarity can be flexibly allowed for each parameter. As a consequence, our inference methodology will be based on MLE using the censored likelihood.

As already mentioned, the authors proposed the use of two types of censoring approaches, left-censoring and interval-censoring. They however discussed that interval-censoring is less effective compared to the left-censoring, and so we only consider the latter. The censored likelihood is written as:

$$L_c(\boldsymbol{\psi}) = \prod_{i:x_i < C} F(C) \prod_{i:x_i \geq C} f(x_i), \quad (\text{IV.4})$$

where $C \geq 0$ is a censoring threshold, $\boldsymbol{\psi}$ is the vector of parameters to be estimated, with size ranging from three in the case of EGPD-Power to five for EGPD-Mixtures. F and f denote the CDF and densities of the model respectively. For now, we will use $C = 2$ mm to compare the models, later in Section IV.3 we will elaborate on this point.

IV.2.3 Results

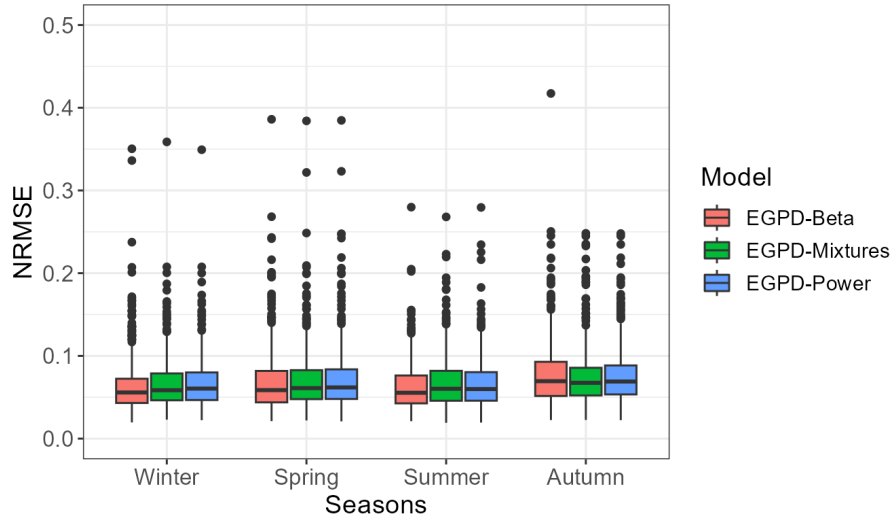


Figure IV.2: Seasonal boxplots of the NRMSE across the 500 stations. The boxplots are colored by model. The lower the score, the better the model.

The boxplot of the NRMSE for the four seasons is shown in Figure IV.2. All four models show generally good model performance, with the median around 0.05, albeit the presence of few outliers in each season. There is no noticeable difference in the performance of the three models, with nearly indistinguishable median scores. EGPD-Beta shows slightly the lowest variability in winter but the most in autumn.

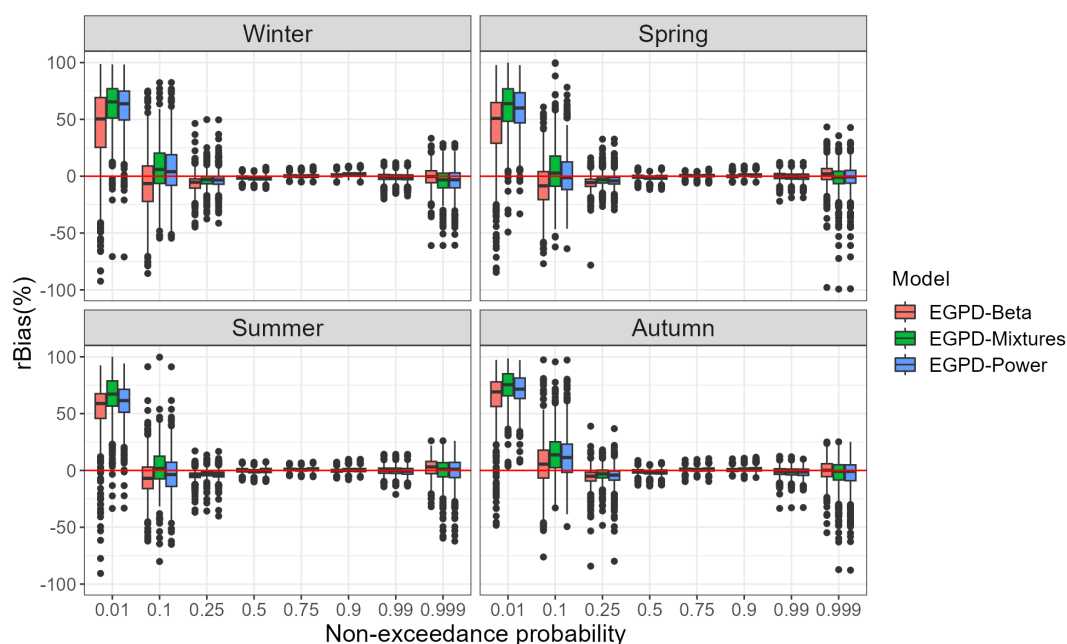


Figure IV.3: Seasonal boxplots of relative bias, $rBias$ (%) across the 500 stations. The boxplots are colored by model. The best model is the one with a score = 0.

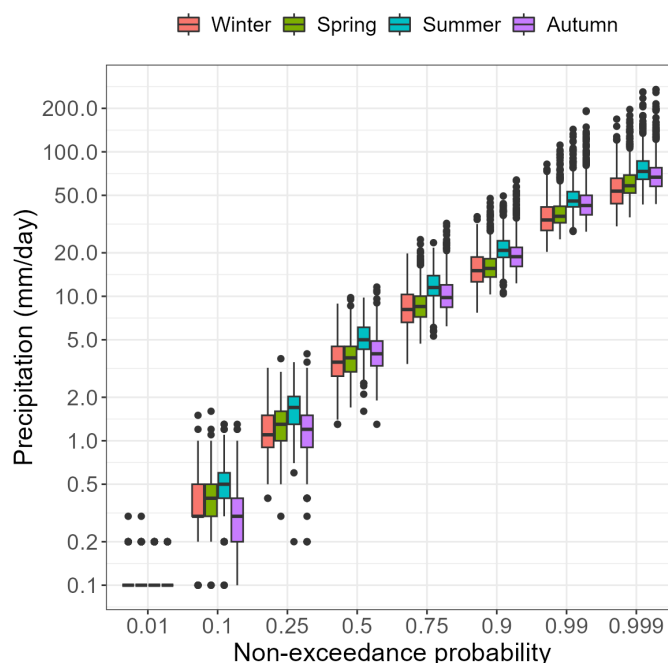


Figure IV.4: Boxplots of the empirical quantiles across the 500 stations. The boxplots are colored by seasons, and they correspond to the non-exceedance probabilities shown along the x-axis.

To compare the models in terms of relative bias, Figure IV.3 shows the boxplots for some specific quantiles that correspond to non-exceedance probabilities of 0.01, 0.1, 0.25, 0.5, 0.75, 0.9, 0.99, and 0.999 of non-zero precipitation. The choice of these quantiles is to assess the potential bias of the models for low, medium, and extreme quantiles. Here also, the models show nearly the same bias for all seasons and quantiles. Notwithstanding, the EGPD-Beta slightly

shows the least bias in comparison to the other models. Furthermore, it can be observed that all three models display the least bias in the bulk and upper tail of the distribution, but show the highest bias in the lower tail of the distribution. To investigate the bias, it is pertinent to understand what these quantiles correspond to in terms of precipitation amount. Figure IV.4 shows the precipitation amounts, across the 500 stations, to which the quantiles correspond. The 0.01-quantile, for example, which is underestimated in each case, corresponds to 0.1 mm, the resolution of the rain gauge. It can therefore be argued that this underestimation is actually caused by the instrumental precision rather than a failure of the models. This is because, while the rain gauge is limited to 0.1 mm, the models can predict any non-zero precipitation, depending on the estimated parameters. The 0.25-quantile has a median ranging between 1 and 1.5 mm, which might be considered small at the daily time scale. This point shows that the observed bias in the lower tail might not translate to the failure of the models in modeling the lower tail.

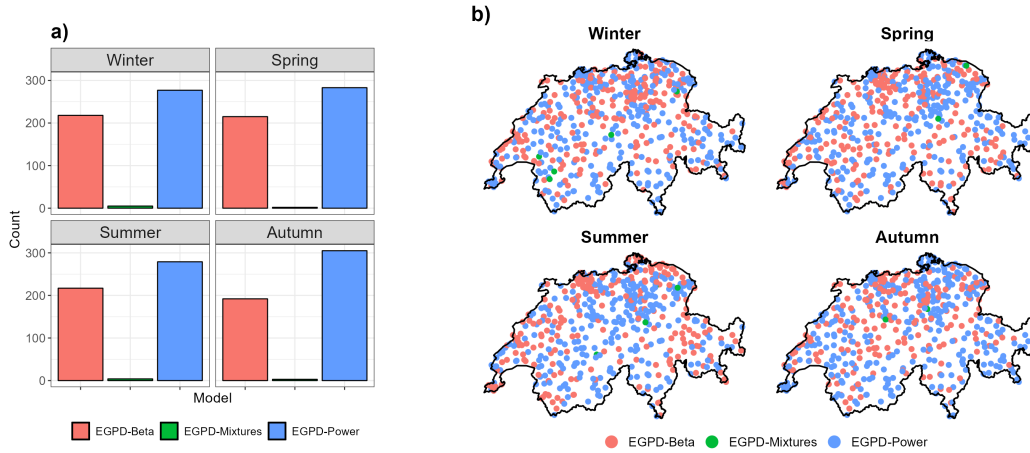


Figure IV.5: Result of model selection according to AIC across the 500 stations. a) Histogram of the number of times a model is selected; b) Map of Switzerland with colored points showing the locations of the stations. Each color represents the model selected at the location.

The two criteria presented above, NRMSE and rBias, reward only the goodness-of-fit of the model, without accounting for the model complexity in terms of the number of parameters. To account for this, we select the model with the lowest AIC score at each of the 500 stations. Figure IV.5 shows the result of the model selection. The histogram on the left (Figure IV.5a) shows that the three-parameter EGPD-Power (colored blue) is the most parsimonious model of all three. It is closely followed by the four-parameter EGPD-Beta (colored pink), while the five-parameter EGPD-Mixtures is only selected in very few stations. The map on the right (Figure IV.5b) shows that there is no specific spatial or seasonal pattern in the model performance to warrant the use of more than one model in our application. Accordingly, we choose the three-parameter EGPD-Power model for the rest of the thesis. As highlighted before, this model has been widely applied in the literature for modeling both daily and sub-daily precipitation (e.g. Naveau et al., 2016; Blanchet et al., 2019; de Carvalho et al., 2021), and especially in our study area (e.g. Evin et al., 2018; Le Gall et al., 2022).

IV.3 Left-censored maximum likelihood estimation

In the previous section, we used the censored likelihood in Equation IV.4 to fit the models using $C = 2$ mm. We will now further investigate and make an attempt to address some potential questions that are likely to arise. For instance, why is censoring necessary in the application of the EGD model? Do all the EGD families presented above require censoring? What is the appropriate censoring threshold to use? How does using the left-censoring threshold affect the performance of the model across the entire distribution, especially the part below the chosen threshold? Below we present our attempt to answer the questions.

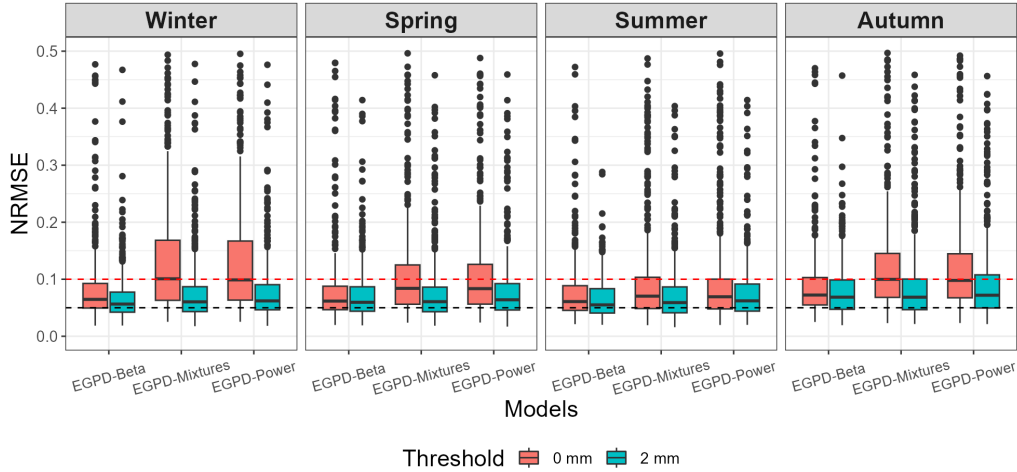


Figure IV.6: Model performance for the case when no censoring is applied (red), and for the case when a censoring of 2 mm (blue) is applied to all the 500 stations (left). The dashed lines colored black and red correspond to $\text{NRMSE} = 0.05$ and 0.1 respectively.

The first question has been answered by the authors of the model, and they attribute it to the effect of discretization of the data due to instrumental precision that strongly affects low values. To counteract this effect, the data is treated as left-censored, and the likelihood in Equation IV.4 is used to estimate the parameters. We need to emphasize that the censoring threshold C employed here should not be confused with the GPD upper threshold u . In the GPD case, the model applies only to the data above u . In our case, however, both the data below and above C are accounted for in the likelihood function, albeit in different ways. The data above C is believed to be observed, and though the density function f is applied to them. For the data below C , it is assumed that their precise magnitude is not known, although they have been observed. All that is known is that they are less than C , and so the distribution function F is applied.

The next question is: do all the EGD families presented above require censoring? To answer this question, Figure IV.6 shows the boxplots of the NRMSE of the three EGD models for the case when no censoring is applied (left) and the case when $C = 2$ mm. We can observe from the figure that the two EGD models based on power law (EGPD-Power and EGD-Mixtures) are more affected compared to EGD-Beta. Even then, when censoring was applied, the performance of the EGD-Beta model improved (see the case of winter for example).

To answer the question of the appropriate censoring threshold to use, we investigate the performance of the models based on three choices of C . In the first two cases, we consider a uniform threshold $C = 1$ mm, and $C = 2$ mm respectively across all stations and seasons. In the third case, we considered a station-specific threshold $C = C_s$. We determine C_s as the threshold that minimizes the NRMSE of the fitted model. Figure IV.7 displays the boxplots of NRMSE for the different values of C . The case when $C = 0$ mm means no censoring was applied. It can be seen

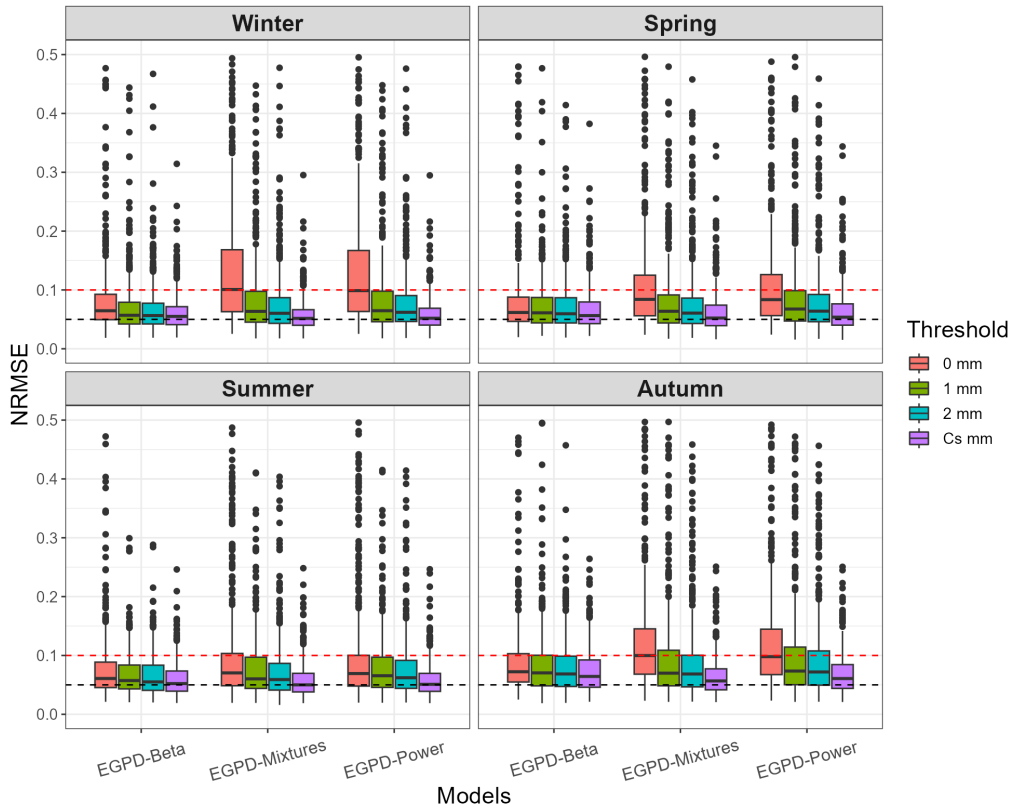


Figure IV.7: Effect of the left-censoring threshold C on the performance of each model (x-axis). The performance is measured with the NRSME (y-axis). The dashed lines colored black and red correspond to $NRMSE = 0.05$ and 0.1 respectively. Each panel corresponds to a season and the boxplots, across the 500 stations, are colored according to the threshold. C_s corresponds to station-specific thresholds, while $C = 0$ mm means no censoring was applied.

that the best result for all seasons and models is when $C = C_s$, that is when the station-specific threshold is applied. This is the case especially when considering the outliers in the boxplots.

Finally, we answer the question of the effect of censored likelihood on the overall distribution, especially the part below the chosen threshold. Figure IV.8 shows the seasonal boxplots of rBias (%) across the 500 stations. The boxplots are colored according to the censoring threshold used. All three models showed similar behavior so we show the results only for the case of EGPD-Power. Looking at all the seasons, all four cases show similar levels of performance for 0.5 to 0.9-quantiles. When we focus on quantiles above the 0.9-quantile, however, the case when $C = 0$ mm shows the worst performance, whereas the case when $C = C_s$ mm shows the best results. Coming down to small quantiles, *i.e.*, 0.25-quantiles and below, the case when $C = C_s$ shows generally the least performance as seen from the large variability. Since $C = C_s$ mm optimizes the NRMSE, it is not surprising that the model performed better for the large values, at the expense of the smaller values. To elaborate on this point, Figure IV.9 shows the quantile-quantile plot for a station in Airolo in summer. We observe that a threshold of 1 mm (green points), and 2 mm (blue points) yield poor performance in the bulk and upper tail (Panel A and B). A threshold of 9.2 mm has to be used to ensure an adequate fit. This however results in a poor fit at the left tail (Panel C) as all the empirical quantiles are underestimated. The implication is that the fitted model with $C = 9.2$ mm can be used for extrapolation, *i.e.* estimation of high return levels, but will result in the underestimation of the small rainfall amounts. Finally, we comment that censoring in general improves the model performance not only in the upper tail but the left tail as well. This point can

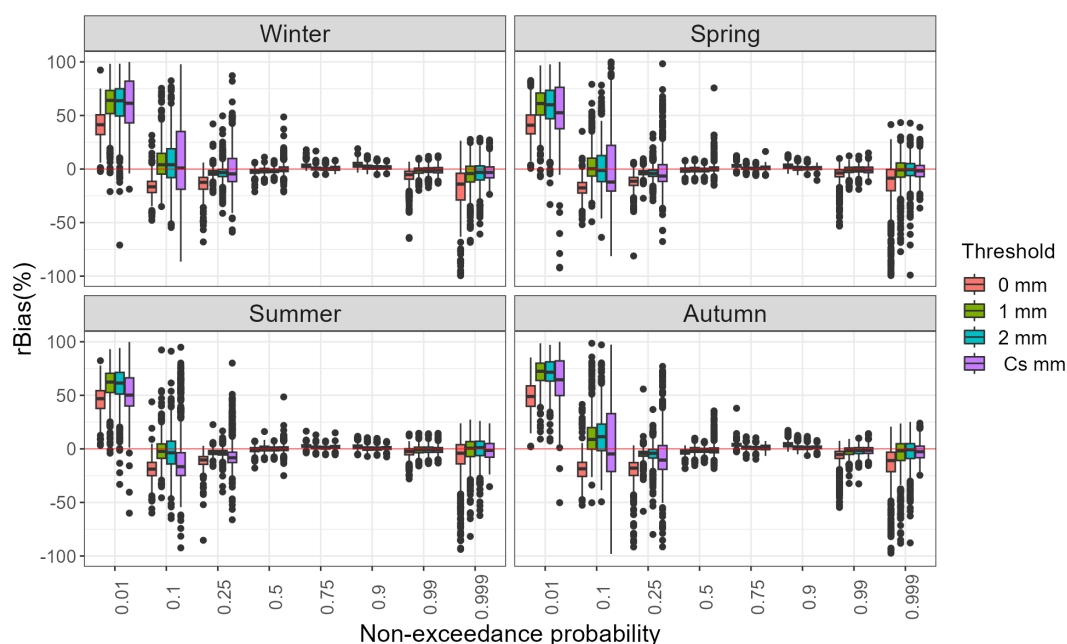


Figure IV.8: Effect of the left-censoring threshold C on the performance of EGPD-Power for specific quantiles (x-axis). The performance is measured based on rBias (y-axis). The red lines correspond to rBias of 0%. Each panel corresponds to a season and the boxplots, across the 500 stations, are colored according to the threshold. C_s corresponds to station-specific thresholds, while $C = 0$ mm means that we did not apply any left-censoring.

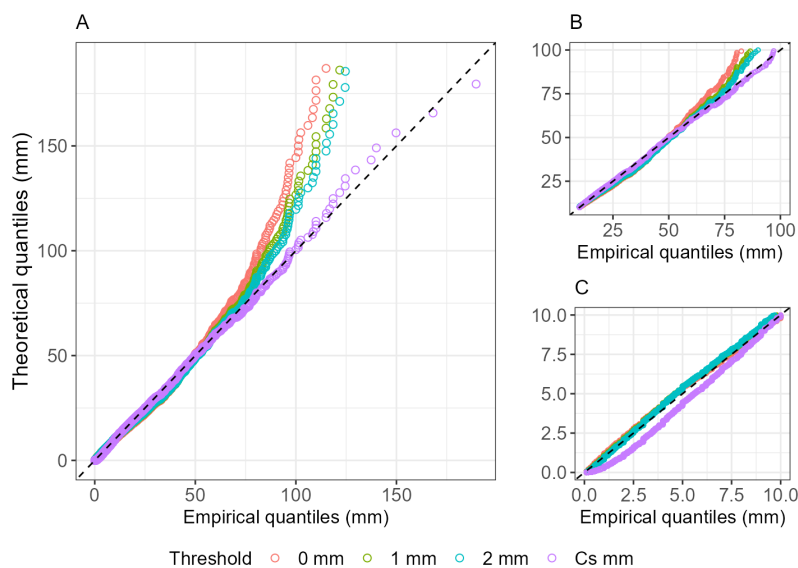


Figure IV.9: Quantile-quantile plots for a station located in Airolo in summer. The points are colored according to the censoring threshold used. For this case, $C_s = 9.2$ mm. Panel A shows the qq-plot for the whole range of data. Panel B zooms within the bulk of the distribution (10 to 100 mm), and panel C zooms on the left tail in order to explore the fit below the chosen censoring threshold, i.e. $C_s = 9.2$ mm. The dash-black line represents the diagonal. For a good performance, the points should fall on the diagonal.

be observed in Figure IV.8, since the case when no censoring is applied ($C = 0$ mm) has a higher

bias (although lower variability) compared to the other cases for 0.1 to 0.5-quantiles.

To summarize, the result in this section revealed that all the models showed improved performance when censoring was applied. It also showed that the models based on power law are most affected by this issue, and require censoring before application. Another point observed is that the best model performance is obtained when station-specific censoring is applied rather than using a generic threshold for all stations and seasons. In this application, we automate the choice of the site-specific threshold choice by finding the value that minimizes the NRMSE. This choice was shown to ultimately improve the model performance, especially in the upper tail, although it resulted in a lower performance in the lower tail of the distribution. The decision to the exact choice depends on the type of applications at hand. Since our target is on the development of models for extreme precipitation, we keep the station-specific threshold that minimizes the NRMSE.

IV.4 Modeling extreme daily and hourly precipitation with the EGPD

The goal of this section is to investigate and understand the seasonal characteristics of extreme precipitation in our study area by modeling non-zero precipitation using the three-parameter EGPD-Power model. Our analysis is conducted at two distinct temporal scales: daily scales, using data from 500 stations, and hourly scales, using data collected from 81 stations. We remind the reader that both datasets have been introduced in Chapter II. Additionally, from this section onward and throughout the rest of this thesis, we will simplify our references to the EGPD-Power model by using the term "EGPD". The fitting of the model is through MLE using the censored likelihood in Equation IV.4 and station-specific threshold. The return level is obtained from

$$i_T = \frac{\sigma}{\xi} \left\{ \left[1 - \left(1 - \frac{1}{T \times \delta} \right)^{\frac{1}{\kappa}} \right]^{-\xi} - 1 \right\}, \quad (\text{IV.5})$$

where $\kappa > 0$, $\sigma > 0$ and $\xi \geq 0$ are the three EGPD parameters, T is the return period in years, δ is the average number of non-zero precipitation intensities per season and is estimated by the long term average.

IV.4.1 Daily scale

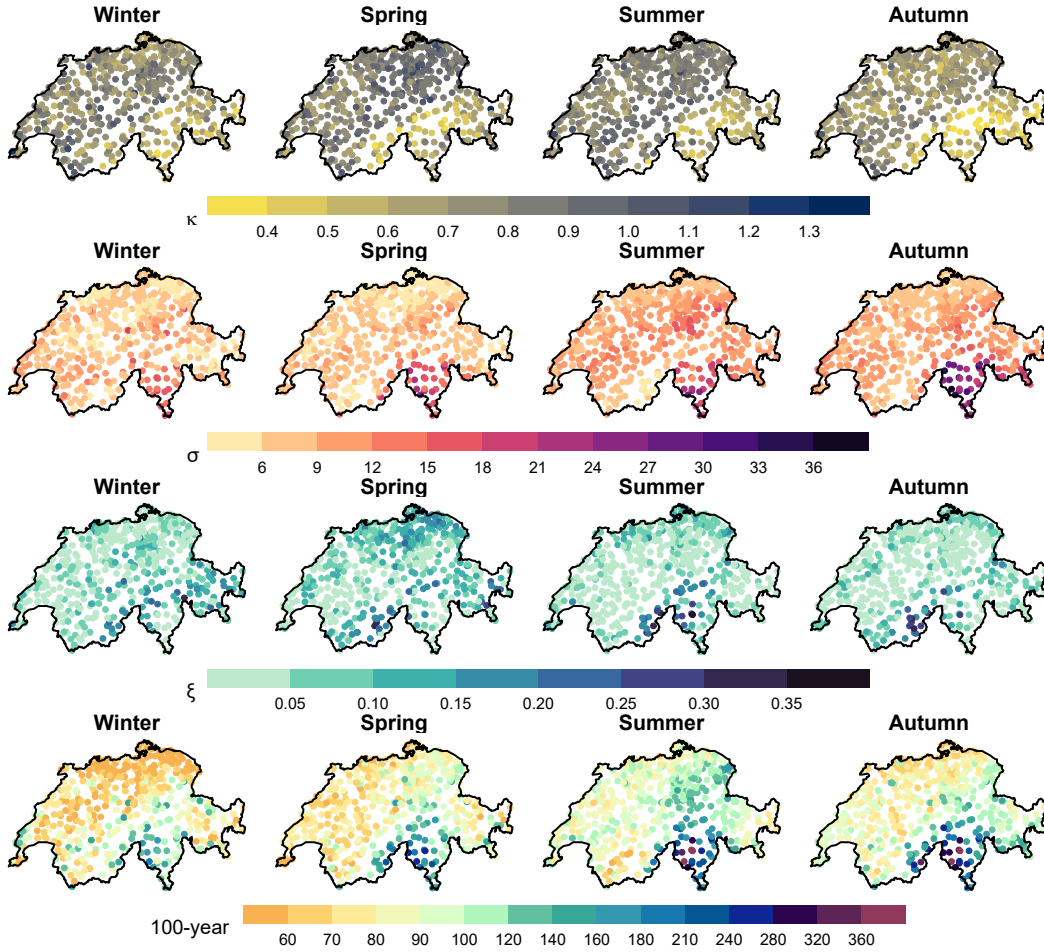


Figure IV.10: Seasonal maps of the fitted EGPD parameters, κ , σ , ξ (row 1 to 3 respectively) and 100-year return level (bottom row) of daily precipitation amount.

Figure IV.10 shows the maps of the three fitted EGPD parameters and 100-year return levels of daily precipitation. Each row in the figure corresponds to a parameter, and each column corresponds to a season. The first row from the top shows the seasonal maps of κ , the parameter that controls the lower tail. As a reminder, when $\kappa < 1$, the lower tail of the density is infinite, and when $\kappa > 1$, the density is zero at the origin, and the larger the value, the lesser the mass is concentrated at the origin. Looking at the maps, the values of the parameter range from 0.2 to 1.25, across all the seasons, with a median of around 0.75 (boxplots not shown here). It can be seen also that there are no large seasonal differences in the magnitude of the parameter. For all seasons, however, the south of Switzerland appears to be the region with the lowest value of κ .

Moving to the second row, the seasonal maps of the scale parameter σ are shown. The parameter represents the spread or variability of non-zero precipitation in the study area. Areas with a large value of σ correspond to locations with large variability of non-zero precipitation and, those with small values correspond to locations with small variability. The maps show that there is both seasonal and regional pattern in the magnitude of the parameter. We observe that all the regions have their lowest values of σ in winter and the largest in summer. An exception is Ticino where the largest values are observed in autumn. Among the regions, the inner valleys possess the lowest values and the Ticino area has the largest values of σ . In the north of Switzerland, the largest values are observed along the northern alpine rim, especially in summer.

The third row shows the seasonal maps of the shape parameter ξ , which controls the upper tail of the EGPD. Around 75% of the stations have $\xi < 0.15$, and a median $\xi < 0.1$ for all seasons. Summer and spring have the largest values of ξ compared to autumn and winter. Unlike the scale parameter, the southern part of the Valais region and the eastern part of Switzerland show the largest values, indicating that more extraordinary precipitation is to be expected in these areas.

Finally, the seasonal maps of the 100-year return level (Equation IV.5) are shown in the bottom row. Similar to the scale parameter, spatial and seasonal patterns are clearly observed. In all regions, the lowest levels are observed in winter. In the north of the Alps, the largest values are observed in summer, while in the south, the largest values are observed in autumn. The similarity of the observed spatial and seasonal pattern of the 100-year return levels with those of the scale parameter indicates that, at least for this return period, σ rather than ξ controls the extreme precipitation. Similar behavior was observed by MeteoSwiss using the GEV (personal communication with Christoph FREI). It is expected however that for much longer return periods, *i.e.* more into the right tail, the spatial pattern of the return levels should be controlled by the the shape parameter ξ .

IV.4.2 Hourly scale

Moving to the hourly scale, Figure IV.11 shows the seasonal maps of the fitted EGPD parameters and 50-year return level. Similar to the daily scale, the case of $\kappa < 1$ is mostly observed everywhere, except in spring where $\kappa > 1$ is observed in the northern plateau. The maps of the scale parameter σ indicate that the largest variability in the hourly non-zero precipitation is observed explicitly in summer and the lowest in winter. Regionally, Ticino in the south presents the largest variability compared to all other regions. The north of the Alps presents similar variability without differences due to the topography. The shape parameter ξ on the other hand shows the largest values in summer, followed by spring, while the lowest values are observed in winter. Regionally, the largest values are observed in the north for spring and summer. On the other hand, in autumn, values that are comparable to those in the north are also observed in Ticino.

The return level maps in the bottom row, show that summer is the season experiencing the largest value of the 50-year return level, and winter is the lowest. In winter, there is no distinct spatial pattern, and the levels in the south are nearly of the same magnitude as those in the north. In spring, the levels in the north and the Ticino start to increase, and they reach their highest values in summer. Similar to the map of σ in summer, there is nearly no noticeable effect of the

topography on the magnitude of the return levels in the north. In autumn, the levels in all regions except Ticino, have nearly the same magnitude as those in spring. Ticino area remains the region subjected to the largest levels irrespective of the season.

Our return level estimates at the hourly scales are similar to those in Fukutome et al. (2015), and climatological discussions for the seasonal and spatial pattern of the return levels can be found therein. For instance, in autumn, the high return levels in Ticino might result from southerly air flows approaching the Alps as the mid-latitude cyclones reach further south as winter approaches. This can lead to intense precipitation events when the warm and humid Mediterranean air is compelled to ascend rapidly over a relatively short distance, ultimately reaching the altitude where free convection can occur. In summer, thunderstorms initiate in the Jura Mountains and along the northern Alpine rim, subsequently migrating towards the plateau region. This movement pattern leads to the occurrence of nearly equivalent levels of extreme precipitation in these areas. In contrast, during the winter season, extreme precipitation at the hourly scale tends to be at its minimum. This phenomenon can be attributed to the fact that extreme precipitation events at this timescale are primarily driven by convection, which is less prevalent during the colder winter months.

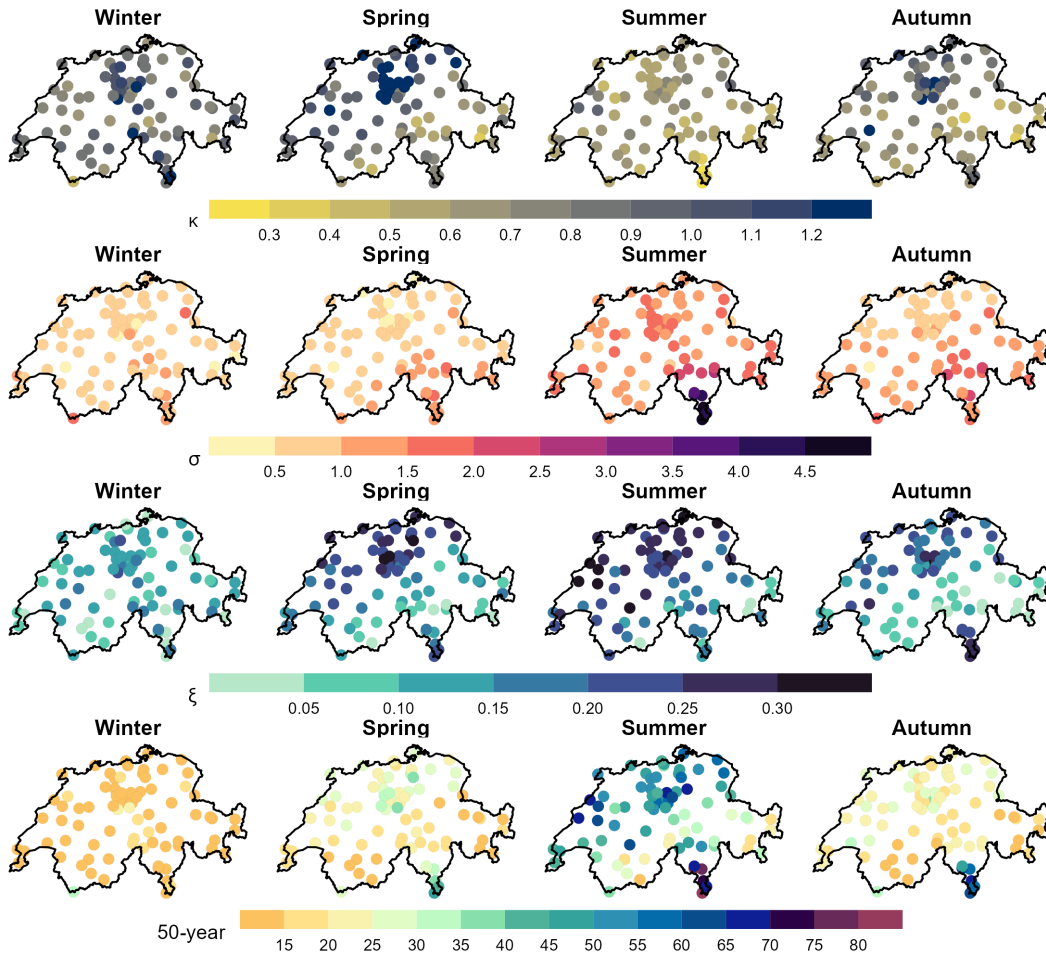


Figure IV.11: Seasonal maps of the fitted EGPD parameters, κ , σ , ξ (row 1 to 3 respectively) and 50-year return level (bottom row) of hourly precipitation intensities.

IV.5 Summary

To sum up, the main take-home message from this chapter is summarized below:

Question 1: What is the most suitable and parsimonious probability distribution to model the entire range of non-zero precipitation intensities in a topographically complex area?

- Three parametric families of the EGPD were compared, and the three-parameter model based on power law appeared to be the most parsimonious. The model is therefore retained for subsequent use in the thesis.
- Left-censored MLE is required for the application of the considered EGPD models, especially for the models based on power law.
- A station-specific threshold choice is required for adequate performance of the model. The use of a uniform threshold might result in poor performance at some stations.
- The use of a high censoring threshold is likely to result in poor model performance at the left tail.
- The three-parameter EGPD model based on power law adequately modeled the hourly and daily extreme precipitation in Switzerland.



Regionalization of daily precipitation estimates in Switzerland

Chapter overview

This chapter addresses Question 2 by presenting the results of the regionalization of daily precipitation. The regionalization is aimed at improving the robustness and reliability of the at-site return level estimates using the EGPD in Switzerland.

Contents

V.1	Paper 1: Performance-based comparison of regionalization methods to improve the at-site estimates of daily precipitation	57
V.1.1	Introduction	58
V.1.2	Data and study area	60
V.1.3	Candidate Methods	61
V.1.4	Comparison and evaluation criteria	65
V.1.5	Comparison framework	65
V.1.6	Evaluation criteria	66
V.1.7	Results	68
V.1.8	Conclusions and Discussion	72
V.2	Selection of regionalization method for ungauged sites	75
V.2.1	Parameter regionalization	75
V.2.2	Spatial cross-validation framework	75
V.2.3	Result and discussion	76
V.3	Summary	78

V.1 Paper 1: Performance-based comparison of regionalization methods to improve the at-site estimates of daily precipitation

Abubakar Haruna¹, Juliette Blanchet² and Anne-Catherine Favre¹

¹University Grenoble Alpes, Grenoble INP, CNRS, IRD, IGE, Grenoble, France,

²University Grenoble Alpes, CNRS, IRD, Grenoble INP, IGE, Grenoble, France.

Article *published* on 01 June 2022 in *Hydrology and Earth System Science (HESS)*

<https://doi.org/10.5194/hess-26-2797-2022>

Preamble to the paper

Recall from Chapter IV that the three-parameter EGPD was shown to be parsimonious and able to adequately model the precipitation intensities in the study area. The main goal of this article is to use the concept of regionalization to improve the robustness and reliability of daily precipitation estimates using the EGPD. To achieve this goal, the article begins by identifying and reviewing three regionalization approaches. The first method is built on the idea of conventional Regional Frequency Analysis (RFA) but is based on a fast algorithm that defines distinct homogeneous regions relying on their upper tail similarity. It includes only the precipitation data at hand without the need for any additional covariate. The second is based on the region-of-influence (ROI) approach in which neighborhoods, containing similar sites, are defined for each station. The third is a spatial method that adopts Generalized Additive Model (GAM) forms for the model parameters. Regionalized EGPD models are then developed and compared using a split-sample cross-validation framework. The article presents several significant contributions. Firstly, regionalization is applied to the EGPD for the first time, a distribution that models both low, medium, and extreme intensities. Secondly, our comparison approach is more comprehensive and based on criteria that focus on the predictive ability of the models. Lastly, regionalization methods, which have not been previously compared, are considered.

Abstract

In this article, we compare the performance of three regionalization approaches in improving the at-site estimates of daily precipitation. The first method is built on the idea of conventional RFA (Regional Frequency Analysis) but is based on a fast algorithm that defines distinct homogeneous regions relying on their upper tail similarity. It uses only the precipitation data at hand without the need for any additional covariate. The second is based on the region-of-influence (ROI) approach in which neighborhoods, containing similar sites, are defined for each station. The third is a spatial method that adopts Generalized Additive Model (GAM) forms for the model parameters. In line with our goal of modeling the whole range of positive precipitation, the chosen marginal distribution model is the Extended Generalized Pareto Distribution (EGPD) on which we apply the three methods. We consider a dense network composed of 1176 daily stations located within Switzerland and in neighboring countries. We compute different criteria to assess the models' performance in the bulk of the distribution and the upper tail. The results show that all the regional methods offered improved robustness over the local EGPD model. While the GAM method is more robust and reliable in the upper tail, the ROI method is better in the bulk of the distribution.

V.1.1 Introduction

Flood events occurring at different time scales pose hazards that are of enormous consequences to life and property. Even though necessary for risk assessments and safe design, reliable prediction remains a challenge and a difficult task. Usually, in the context of risk assessment, river flows are simulated via hydrological models. These models take as inputs, among others, meteorological data such as temperature and precipitation. However, whatever the complexity of the model and how it represents the underlying hydrological behavior of the catchment, the accuracy, robustness, and reliability of the flood predictions rely on the quality of the input data.

Precipitation intensities, the key input signal, is modeled using probabilistic methods. Within this framework, a good probabilistic model should predict precipitation intensities of any return level, whether low, medium, or extreme with reliable accuracy. Gamma distribution, a common choice over models such as log-normal, Weibull, exponential, fails in this aspect, as the tail is too light to model heavy intensities (Katz et al., 2002). Models based on the classical extreme value theory (EVT), such as the Generalized Pareto (GP), can model the upper tail but one has to choose another model for the other intensities below the chosen threshold.

Since GPD has been favored in hydrological applications (see e.g. Langousis et al., 2016), many authors in the framework of modeling the full range observations have considered different approaches to adding flexibility to this model. A common approach is the use of mixture models where GPD is combined with another appropriate model for the bulk of the distribution, (see review in Scarrot and MacDonald, 2012). Mixture models, however, have the drawback of inflating the number of parameters to estimate (Naveau et al., 2016) and thus complexify statistical inference.

As an alternative, Naveau et al. (2016) proposed a model which is an extension of the GP (afterward called EGPD). It has the advantage of avoiding the need for threshold selection (a drawback of GP) while being parsimonious by avoiding the use of mixtures. The model is gamma-like in the lower tail and heavy-tailed (GP) in the upper, with a smooth transition in-between. It is able to model adequately the entire range of positive precipitation and many author within the framework of precipitation modelling have used this model, (e.g. Blanchet et al., 2015; Evin et al., 2018; Tencalec et al., 2020; Le Gall et al., 2022).

Modeling the whole range of precipitation has various practical applications. For instance in

flood risk assessments, where stochastic precipitation generators are used to simulate long series of positive precipitations, extremes included (e.g. in [Evin et al., 2018](#)). The simulated precipitation is then used as input to conceptual hydrological models for the simulation of long series of river flows. Other practical applications are in the evaluation of numerical weather simulations or investigation of the climatology of precipitation events as outlined by [Blanchet et al. \(2019\)](#)

Although EGPD uses all the data to estimate the parameters, the shape parameter which controls the upper tail behavior remains difficult to estimate based on a few decades of data, the usual length of precipitation data. This is because there are usually few extremes exhibiting much variability. As precipitation is spatial by nature, several studies ([Cunnane, 1988](#); [Burn, 1990](#); [Hosking and Wallis, 2005](#)) proposed the use of observations surrounding the local station to increase the quantity of data available for estimation, thereby reducing the uncertainty involved in the estimation.

Different methods exist in the literature to use information surrounding the station at hand (see [Cunnane, 1988](#); [Hosking and Wallis, 2005](#)). Methods based on regional homogeneity (e.g. method of [Hosking and Wallis \(2005\)](#)) pool all observations in hydrologically similar sites to increase the sample size, and by so yielding more accurate estimates of the parameters. Hydrologically similar sites are first defined using cluster analysis and then subjected to some statistical homogeneity tests on the scaled observations. Thereafter, a chosen distribution is fitted to the scaled observations in the identified region, and all stations within this region would share the same regional parameters. Station-specific parameters and quantiles can then be inferred by appropriate scaling. This method has been applied by various authors (e.g. [Gaál and Kysely, 2009](#); [Malekinezhad and Zare-Garizi, 2014](#)) and on various distributions such as the GEV and the GP. Variants of this method exist such as the region of influence (ROI) proposed by [Burn \(1990\)](#), which avoids defining fixed regions, but assigns homogeneous regions (neighborhood of different shapes according to the method) for each site. Scaled observations within the neighborhood of each station are then used to estimate the regional parameters of that station. This method has been applied by various authors (see [Gaál et al., 2008](#); [Kysely et al., 2011](#); [Carreau et al., 2013](#); [Evin et al., 2016](#); [Das, 2017, 2019](#)).

In contrast to the aforementioned methods that generally rely on some covariates such as spatial coordinates to define the homogeneous regions, another variant, recently developed by [Le Gall et al. \(2022\)](#) defines homogeneous regions based on the similarity of their upper tail behavior. This method avoids the use of any covariate but relies completely on the precipitation data at hand. The upper tail behavior for each station is summarized based on a ratio of probability-weighted moments (PWMs) (refer to Eq V.4). Subsequently, a clustering algorithm is used to partition these ratios into distinct homogeneous regions, and then regional parameters can be estimated.

Spatial methods exist in which all the observations from all the stations are pooled and then used to estimate spatial surface for each of the model parameters. The surface for each of the model parameters is defined as a function of some well-chosen covariates such as longitude, latitude, altitude, etc. Estimating the parameters involve simply the estimation of the coefficients of these relationships. From the fitted surfaces, station-specific model parameters can be inferred as a function of the covariates at that specific location. Surfaces that are smooth and flexible can be obtained by fitting generalized additive models (GAM) to the relationships (see [Chavez-Demoulin and Davison, 2005](#); [Blanchet and Lehning, 2010](#); [Youngman, 2019, 2020](#)). Other alternatives to the classical RFA include the Bayesian spatial modeling (see [Madsen et al., 1995](#); [Cooley et al., 2007](#)) and those discussed in [Cunnane \(1988\)](#).

Recent analyses have been done to compare the performance of regional approaches with a particular interest in distributions allowing to model extremes only. [Gaál et al. \(2008\)](#) compared different versions of the ROI method against the classical RFA method of [Hosking and Wallis \(2005\)](#). The ROI versions were distinguished by the choice of the distance metric and the maximum threshold to delineate neighborhoods. For all the methods, GEV distribution was assumed as the underlying distribution. The authors, through a Monte Carlo simulation study, concluded

that the ROI approach was superior to the classical RFA involving distinct clusters. In an interpolation framework, [Carreau et al. \(2013\)](#) compared three methods; spatial interpolation of locally estimated parameters, method of ROI and a rainfall generator called SHYPRE. For the first two methods, GEV distribution was assumed. The author found comparable performance between the ROI and SHYPRE, and a lack of robustness in the method based on interpolation of local parameters. [Deidda et al. \(2021\)](#) also used GEV to compare the classical RFA of [Hosking and Wallis \(2005\)](#) and geostatistical interpolation of locally estimated parameters. They highlighted the limitation of the former in yielding distinct regions and being of less accuracy compared to the latter. Other comparisons include those of [Gaál and Kyselý \(2009\)](#); [Kyselý et al. \(2011\)](#) and [Das \(2019\)](#).

Our approach differs from the aforementioned studies in the following aspects. First, in contrast to the case where the underlying distributions are basically for modeling only extremes (e.g. [Burn, 1990](#); [Gaál et al., 2008](#); [Gaál and Kyselý, 2009](#); [Kyselý et al., 2011](#); [Carreau et al., 2013](#); [Evin et al., 2016](#); [Das, 2019](#); [Deidda et al., 2021](#)), we consider the EGPD that models both low, medium and extreme precipitations. This is in line with our goal of having a robust and reliable model, that can model the whole distribution and not only the extremes. Secondly, our comparison approach is more general and based on [Garavaglia et al. \(2011\)](#) and [Renard et al. \(2013\)](#), focusing on the predictive ability of the models in a cross-validation framework similar to the case of authors such as [Blanchet et al. \(2015\)](#); [Evin et al. \(2016\)](#), rather than simply based on quality of fit (e.g. [Gaál et al., 2008](#); [Kyselý et al., 2011](#); [Deidda et al., 2021](#)). Finally, in our contribution, we compare new methods not previously compared viz-a-viz. The first method defines distinct homogeneous regions based on their similarity in upper tail behavior ([Le Gall et al., 2022](#)). The second method is based on the ROI approach framework of [Evin et al. \(2016\)](#). The last method is a spatial approach that assumes generalized additive models (GAM) forms for the model parameters. For all the methods, we assume the EGPD as the underlying marginal distribution. We apply this comparison to a dense network of over 1100 daily stations located within Switzerland and in the neighboring countries.

The paper is organized as follows: Section [V.1.2](#) presents the data and the study area. Section [V.1.3](#) introduces the competing models while section [V.1.4](#) describes the comparison methodology as well as the criteria used. The results are presented in section [V.1.7](#). Finally, we discuss the conclusion and the relevant perspectives in section [V.1.8](#).

V.1.2 Data and study area

The comparison is made considering daily precipitation observations from 1176 stations shown in Figure [V.1](#). From this total, 500 are located within Switzerland and 676 in the neighboring countries. The data has a variable length ranging from a minimum of 20 years to a maximum of 156 years, from 1863 to 2019. The bar plot in Figure [V.1](#) shows the number of stations installed in each country during each decade of the study period. While the main study area is Switzerland, we use the data in the neighboring countries simply to improve the estimates of the stations located around the border of Switzerland. Consequently, although we use all the stations (both within and outside) for regionalization and model fitting, we apply the performance criteria only on the stations located in Switzerland.

Daily precipitation in Switzerland is characterized by seasonality arising from multiple moisture sources brought by prevailing winds ([Sodemann and Zubler, 2009](#); [Umbricht et al., 2013](#); [Giannakaki and Martius, 2015](#)). It is also characterized by spatial variability both in intensity and occurrence resulting from the complex topography ([Sevruk, 1997](#); [Sevruk et al., 1998](#); [Frei and Schär, 1998](#); [Molnar and Burlando, 2008](#); [Isotta et al., 2014](#)). Winter receives the least precipitation and summer is the main season of precipitation all over Switzerland. An exception is in the case of Ticino in the South, where autumn is the main season. This region is also subject to the heaviest precipitation. In the North of the country, the topography plays an important role; the

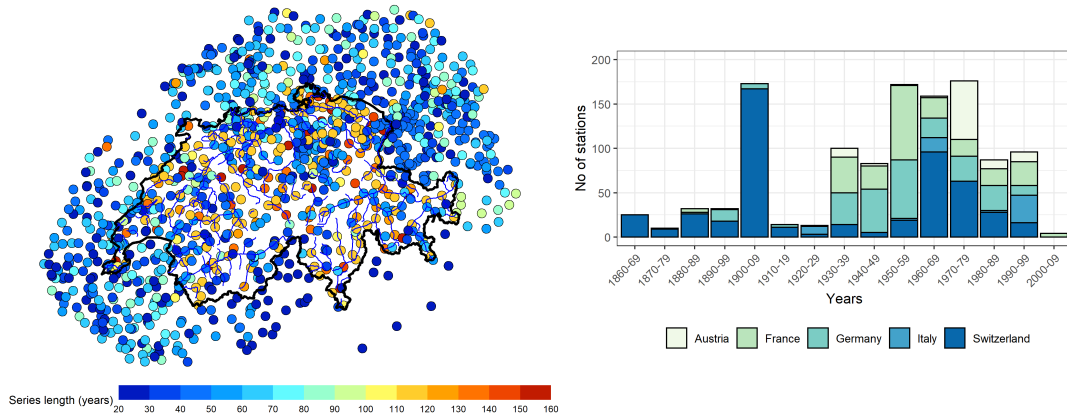


Figure V.1: Description of the data used for the study. Left: Map of Switzerland and the neighborhood showing the location of the 1176 daily stations. The color indicates the length of the series, minimum of 20 years and maximum of 156 years. Right: Bar plot showing the number of stations installed in each country for each decade.

northern rim and the Jura mountain receive heavier precipitation compared to the plateau.

As a result of the marked seasonality, and the importance of taking it into account (Leonard et al., 2008; Garavaglia et al., 2011), we apply a seasonal based analysis approach. We divide the data into the four distinct seasons of three months each: Winter (Dec, Jan, Feb), Spring (Mar, Apr, May), Summer (Jun, Jul, Aug), and Autumn (Sep, Oct, Nov).

V.1.3 Candidate Methods

In this section, we start by presenting the marginal distribution (EGPD). We then give a brief description of three different methods of regionalization that we will use to improve the local estimates of the EGPD. They are: i) Regional frequency analysis based on the upper tail behavior, ii) Region of influence approach (ROI), and iii) Spatial method using generalized additive model (GAM) forms. Finally, we summarize the regional models that are developed based on the three outlined methods of regionalization as applied to the EGPD.

V.1.3.1 Marginal distribution of positive precipitation

We use the marginal distribution of precipitation proposed by Naveau et al. (2016), which is able to model sufficiently the full spectrum of positive (non zero) precipitation. The model is EVT compliant in the upper and lower tail while providing a smooth transition in-between. It provides an alternative to the light-tailed distributions such as Gamma, which can underestimate extremes (Katz et al., 2002). Four parametric families of this model have been proposed by Naveau et al. (2016), and more recently a non-parametric scheme of the transition function by Tencaliec et al. (2020). However, the simplest of the parametric family is parsimonious and can adequately model precipitation intensities without the need for GPD threshold selection (Naveau et al., 2016; Evin et al., 2018; Le Gall et al., 2022). We therefore use this model in our study.

Let X be a random variable representing positive daily precipitation intensity that is distributed according to the EGPD, then the cumulative distribution function (CDF) is given by

$$F(x) = \mathbb{P}(X \leq x) = G \left[H_{\xi} \left(\frac{x}{\sigma} \right) \right], \quad (\text{V.1})$$

where G is any CDF that ensures a smooth transition between the EVT compliant upper and lower tail, and

$$H_{\xi}\left(\frac{x}{\sigma}\right) = \begin{cases} 1 - (1 + \xi \frac{x}{\sigma})_+^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - \exp(-x/\sigma) & \text{if } \xi = 0 \end{cases}, \quad (\text{V.2})$$

with $a_+ = \max(a, 0)$.

For the parsimonious model we use, the function G is simply defined as $G(v) = v^k$. Therefore the model is given as:

$$F(x) = \left[H_{\xi}\left(\frac{x}{\sigma}\right) \right]^k. \quad (\text{V.3})$$

The model has three parameters. $k > 0$ controls the lower tail, $\xi \geq 0$ controls the upper tail, and $\sigma > 0$ is the scale parameter.

Inference of the model parameters can be done through maximum likelihood estimation (MLE), or through the method of probability-weighted moments (PWM).

V.1.3.2 Methods of regionalization

V.1.3.2.a RFA based on upper tail behavior: Classical regional frequency analysis ([Hosking and Wallis, 2005](#)) defines regions that are homogeneous up to a scaling factor. To identify the regions, covariates have to be carefully chosen, which usually include at-site characteristics such as geographical and atmospheric characteristics. However, this information might not be generally available at each station. Homogeneity tests then have to be applied to confirm that the regions are sufficiently similar.

[Le Gall et al. \(2022\)](#) proposed a fast and efficient method to delineate regions based on the homogeneity of their upper tail behavior. The method relies on the precipitation data at hand only without the need for additional covariates. More so, regions identified are inherently homogeneous, thereby avoiding the need for the application of some homogeneity tests. For each station i , a ratio ω given in Eq. [V.4](#) that is based on probability weighted moments (PWM) is obtained.

$$\hat{\omega} = \frac{3\hat{\alpha}_2 - \hat{\alpha}_0}{2\hat{\alpha}_1 - \hat{\alpha}_0} - 1, \quad (\text{V.4})$$

where $\hat{\alpha}_j$ denotes the PWM of order j .

The authors showed that ω summarizes the upper tail behavior of the data at hand, and for the EGPD model, it depends mainly on the ξ parameter (effect of k not very significant, (see [Le Gall et al., 2022](#))). Stations with high values of ω have high intense extremes, and those with low values have less intense extremes. The idea is to classify or form regions with similar values of ω , which is possible using any of the clustering algorithms such as K-means, hierarchical clustering, Partitioning around medoids (PAM), For details of these clustering methods, see [Kaufman and Rousseeuw \(2005\)](#).

V.1.3.2.b RFA based on region of influence approach (ROI): The region-of-influence (ROI) method ([Burn, 1990](#)) is similar in concept to the classical RFA method. It circumvents the drawback of having contiguous regions separated by distinct boundaries that result in “undesirable step changes of the variables and estimated quantiles” ([Gaál et al., 2008](#)). Instead of defining distinct homogeneous regions separated by some boundaries, a region of influence is assigned to each station. All the scaled observations in the identified ROI are used to estimate its regional parameters. To apply this method, several choices have to be made. These involve the choice of the scale

factor, distance metric, radius delimitation, and homogeneity test. The choices influence the application of the method and have to be carefully and objectively decided. Different authors in the application of the methods have explored some or all of these factors, starting from [Burn \(1990\)](#), and in [Gaál et al. \(2008\)](#).

In this work, we follow the objectively selected steps and choices similar to [Evin et al. \(2016\)](#) in the application of the method in the Southeastern part of France. The authors applied the method by considering POT (exceedances of a 70 % quantile) of central precipitations (largest observations in 3-day precipitation events) and on some distributions (Exponential, GPD and Weibull). We apply the same procedure but on positive precipitation and EGPD model.

V.1.3.2.c Spatial method based on Generalized Additive Model (GAM): In contrast to the previous methods where regionalization is based on homogeneity of normalized data or upper tail similarity, this is a regression-based method for fitting the parameters of models, by allowing for spatial non-stationarity of the parameters. Accordingly, we pool all the observations from all the stations to estimate flexible and smooth spatial surfaces for each parameter, relying on the ground that pooling of spatial information can help improve the at-site estimates, and hence the extreme quantiles. In particular, we let the parameters have a generalized additive model (GAM) form, represented by smoothing splines. In effect, we assumed them to have some form of flexible relationship with some covariates x , which can be explained by GAM forms.

V.1.3.3 Regional models

This section summarizes the regional models that are compared in the study. The models are built based on the concepts of the three regionalization methods outlined in section [V.1.3.2](#). [Table V.1](#) presents the four models plus the local EGPD model.

Table V.1: Summary of the regional models that are compared in this study. The first model is the local EGPD model, the next three models are based on regional homogeneity, while the last model is a spatial methods based on GAM. The second column gives the name of the model. The next three columns are the parameters of the EGPD model, and indicates whether the parameter is estimated locally (from the data of the station at hand only) or through regionalization. The last column gives reference to the section where the method is described.

S/N	Model	κ	σ	ξ	Ref. Section
1	Local EGPD	local	local	local	V.1.3.1
2	Omega EGPD	local	local	regional	V.1.3.3.a
3	ROI EGPD Full	neighborhood	neighborhood	neighborhood	V.1.3.3.b
4	ROI EGPD Semi	local	local	neighborhood	V.1.3.3.c
5	GAM EGPD	spatial	spatial	spatial	V.1.3.3.d

V.1.3.3.a Omega EGPD Model: This model is built based on the regionalization method described in section [V.1.3.2.a](#), i.e, RFA based on upper tail behavior. To build this model that relies on regionalization of the shape parameter, the following steps are followed:

1. For each station i , $i = 1, \dots, N$ use the positive data to estimate the ratio $\hat{\omega}_i$.
2. Identification of homogeneous regions: Use an appropriate clustering algorithm alongside an internal validation criteria to decide on the optimal number of homogeneous clusters

based on the estimated $\hat{\omega}_1, \dots, \hat{\omega}_i, \dots, \hat{\omega}_N$. In our case, after doing a simulation study (result not shown), we settled on the PAM algorithm and three criteria, Silhouette (Rousseeuw, 1987), Davies Bouldin (DB) (Davies and Bouldin, 1979), and S_Dbw (Halkidi and Vazirgiannis, 2001).

3. For each homogenous region C ,
 - (a) Fit EGPD locally to find $(k_i, \sigma_i, \text{ and } \xi_i)$.
 - (b) Find the regional shape parameter ξ_r as average of all ξ_i in that region.
 - (c) Fit EGPD locally again to find new estimates $\hat{k}_{i,new}$ and $\hat{\sigma}_{i,new}$, given the estimated ξ_r .

We have also explored other options to estimate ξ_r after obtaining the homogeneous regions:

- The first method involves pooling all the observations in a homogeneous region (cluster) after scaling them by their mean, and then fitting a regional EGPD to estimate the regional parameters $(k^{(R)}, \sigma^{(R)}, \xi^{(R)})$. We then retain $\xi^{(R)}$ and then refit an EGPD locally to estimate k_i and σ_i . Every station in that cluster will have similar $\xi^{(R)}$ but locally estimated k_i and σ_i .
- The second approach is similar to the main method where we take the the average of the locally estimated ξ , but here, we take a weighted average. The idea is that, for each cluster, the locally estimated ξ for the medoid station (the station with the least average dissimilarity to all the other stations in the same cluster) should be assigned the highest weight in the average, all other stations should then have weights as a function of their dissimilarity to this medoid. Thus very similar stations to the medoid should have higher weights, while those that are less similar should have smaller weights. The dissimilarity is measured by the Manhattan distance $|\omega_m - \omega_i|$, where ω is given in Eq. V.4, while the indices m and i denote respectively the medoid and the station i .

After testing these three approaches to estimate ξ_r , by measuring the accuracy of the resulting quantile-quantile plot according to the normalized-root-mean-square-error (NRMSE) (see section V.1.4 for details of this criteria,) results (not shown here) showed that the first method, where we simply take the average of the locally estimated ξ , resulted in the least error. We thus retain this approach in our subsequent analysis.

V.1.3.3.b ROI EGPD Full Model: This model is based on the method of ROI described in V.1.3.2.b.

Let $X_i \sim EGPD(k_i, \sigma_i, \xi_i)$ be the random variable of daily positive precipitation at station i which is distributed according to the EGPD. We assume also that $Y_i = \frac{X_i}{m_i}$ is the daily positive precipitation normalized by a scale factor m_i . If we consider several stations (whose data has been normalized as well) that have similar distribution as Y_i and use the data to estimate the regional parameters $k^{(R)}, \sigma^{(R)}, \text{ and } \xi^{(R)}$, then, Y_i will have parameters $(k^{(R)}, \sigma^{(R)}, \xi^{(R)})$. Accordingly, by back transformation, the unnormalized random variable X_i will have the parameters $(k^{(R)}, m_i \sigma^{(R)}, \xi^{(R)})$. This shows that for a random variable that is distributed according to the EGPD, after regionalization, the parameters k and ξ are those obtained regionally, while the scale parameter σ has to be multiplied by the scale factor for that station.

The general procedure of application is summarized below and the details can be found in Evin et al. (2016).

1. For each station and season, exceedances of a threshold of 95% quantile (POT) are selected and scaled by a factor. The scale factor is the mean of all positive daily precipitation.

2. We start the search from a radius of 2 km starting from the current station. If we find other stations within this radius, we apply the homogeneity test on the scaled POT found in step 1. If the test is positive, we increase the radius by another 2 km and repeat the test. We stop the search when the test fails or when we reach a maximum radius beyond which we doubt the existence of homogeneity. We use 100 km as the upper bound.
3. The distance metric we use is the “crossing distance” (Gottardi et al., 2012) given in Equation V.5. This distance takes into account the effect of elevation and is summed over all the pixels along a straight line between two targeted stations. We use a weight on elevation equal to 20 similar to Evin et al. (2016) to account for the effect of relief. Again, following the same authors, we use the test of Hosking and Wallis (2005) on mean and L-coefficient of variation (L-CV).

$$d = \sqrt{\sum_{\text{pixels}} (\Delta x^2 + \Delta y^2 + 20\Delta z^2)}. \quad (\text{V.5})$$

4. We estimate the regional parameters by a weighted-MLE on the scaled positive observations in the ROI. The target station has the highest weights and the closer the station, the higher the weights.

The full regional model is such that $X_i \sim \text{EGPD}(k^{(R)}, m_i \sigma^{(R)}, \xi^{(R)})$. We call this model - ROI_EGPD_Full afterward.

V.1.3.3.c ROI EGPD Semi Model: This model follows exactly as the ROI_EGPD_Full in the preceding section. The only difference is that here we retain the regional shape parameter $\xi^{(R)}$ obtained from the neighborhood, and then estimate the two other parameters locally, i.e. from only the data at the station i . We refer to this model as ROI_EGPD_Semi.

The semi regional model is such that $X_i \sim \text{EGPD}(k_i, \sigma_i, \xi^{(R)})$.

V.1.3.3.d GAM EGPD Model: For this spatial EGPD model, we have $X(\mathbf{x}) \sim \text{EGPD}(\sigma(\mathbf{x}), k(\mathbf{x}), \xi(\mathbf{x}))$, where \mathbf{x} denotes some covariate, and each of the model parameter depends on some form of \mathbf{x} . The relationship between the model parameter (say α) and the covariate \mathbf{x} is through an identity link:

$$\alpha(\mathbf{x}) = \beta_0 + \sum_{k=1}^K \sum_{d=1}^{D_k} \beta_{kd} b_{kd}(\mathbf{x}), \quad (\text{V.6})$$

where β_{kd} and b_{kd} are respectively the basis coefficients and the basis functions. K is the number of smooths and D_k is the dimension (number of knots) for smooth k .

For the choice of spatial covariates, we use longitude, latitude, and mean daily precipitation because they give better Akaike information criterion (AIC) (Akaike, 1974). To fit EGPD with GAM, we extended the functions already available in the `evgam` R package (Youngman, 2020).

V.1.4 Comparison and evaluation criteria

This section presents the comparison framework and the performance criteria used to compare the regional models.

V.1.5 Comparison framework

The evaluation framework and criteria is as proposed by Garavaglia et al. (2011) and Renard et al. (2013). Garavaglia et al. (2011) and the references there in, argued that the classical statistical goodness of test fits such as Kolmogorov–Smirnov test (Kolmogoroff, 1941; Sminorv, 1944),

Anderson–Darling Test (Anderson and Darling, 1952), Cramer von-Mises criterion (Cramer, 1928; Darling, 1957) lack the ability to assess the models ability to predict unobserved values, and that they are also not very efficient for three-parameter distributions.

Accordingly, we follow a split sampling procedure and a cross-validation framework. For each station i , we used $1/3^{rd}$ of the data by choosing every third observation to reduce temporal dependence. Then, we divide the non-zero observations into two equal sub-samples of the same length but on different years that are randomly chosen. We call the first and second sub-samples $S_i^{(1)}$ and $S_i^{(2)}$. We then fit a model $\hat{F}_i^{(1)}$ and $\hat{F}_i^{(2)}$ on sub-sample $S_i^{(1)}$ and $S_i^{(2)}$ respectively. We then compute the criteria $C_i^{(12)}$ at station i , by comparing $\hat{F}_i^{(1)}$ vs $S_i^{(2)}$ (i.e model fitted on sub-sample 1 vs observations in sub-sample 2). In the same way we compute criteria $C_i^{(21)}$. Given that we have N stations, and so N values of both $C_i^{(12)}$ and $C_i^{(21)}$, the regional score is obtained as the average of these scores. This procedure is repeated 50 times to obtain 50 regional averages of these indices.

V.1.6 Evaluation criteria

For each method, four (4) criteria C are computed. We first judge the methods based on how they accurately fit the entire observations at each site. Next, we compare them in terms of their robustness in extrapolation, *i.e.* how stable a high quantile estimate is, depending on which sub-sample is used in the estimate. We finally judge the performance based on the reliability to predict precipitation maxima.

In the following paragraphs, we describe the four criteria used for the comparison:

V.1.6.1 Accuracy of the whole distribution

The accuracy of the model in predicting the positive observations at a given station is given by the normalized root mean square error (NRMSE) (Blanchet et al., 2019). For each site i , the positive observed values in $S_i^{(2)}$ are associated to their empirical return periods. We then use the fitted model on $S_i^{(1)}$, *i.e.* $\hat{F}_i^{(1)}$, to estimate the modeled quantiles associated to these return levels and finally compute the normalized root mean squared error associated to these quantiles. The normalization is by the average daily precipitation. This score is given as:

$$\text{NRMSE}_i^{(12)} = \frac{\left\{ \frac{1}{n_i^{(2)}} \sum_{k=1}^{n_i^{(2)}} \left(r_{i,T_k}^{(2)} - \hat{r}_{i,T_k}^{(1)} \right)^2 \right\}^{1/2}}{\frac{1}{n_i^{(2)}} \sum_{k=1}^{n_i^{(2)}} r_{i,T_k}^{(2)}}, \quad (\text{V.7})$$

where $\text{NRMSE}_i^{(12)}$ is the score computed at station i , $r_{i,T_k}^{(2)}$ is the k^{th} observation of return period T in $S_i^{(2)}$, $\hat{r}_{i,T_k}^{(1)}$ is the corresponding T return level estimated from $\hat{F}_i^{(1)}$. The denominator $\frac{1}{n_i^{(2)}} \sum_{k=1}^{n_i^{(2)}} r_{i,T_k}^{(2)}$ is the average daily precipitation at site i . Details on the score are given in Blanchet et al. (2019).

Finally, the regional score computed over the N stations, *i.e.* $\text{NRMSE}_{\text{reg}}^{(12)}$ is given as:

$$\text{NRMSE}_{\text{reg}}^{(12)} = 1 - \frac{1}{N} \sum_{i=1}^N \text{NRMSE}_i^{(12)}. \quad (\text{V.8})$$

$\text{NRMSE}_{\text{reg}}^{(21)}$ is computed in similar way. We thus finally have 2×50 values of $\text{NRMSE}_{\text{reg}}$ resulting from the cross-validation on both periods. $\text{NRMSE}_{\text{reg}} = 1$ means a perfect model, and the closer the value is to 1, the more accurate the model is.

V.1.6.2 CRPS

The continuous ranked probability score (CRPS) has been used as a metric to compare the performance of two competing probabilistic forecasts models (Jordan et al., 2018). It gives a combined measure of both spread and reliability of a forecast distribution, given the observation or outcome that is observed.

For a given observation $x_{i,t}$ at station i and time step t that is contained in $S_i^{(1)}$, we have 50 of its quantile estimates coming from the 50 fitted models $F_i^{(2)}$ (models fitted with data not containing $x_{i,t}$). If the method used to estimate all this 50 models is accurate enough, then these 50 quantile estimates should be similar (low spread), and very close to the observed value $x_{i,t}$. The same applies to an observation $x_{i,t}$ contained in $S_i^{(2)}$ when compared to its quantile estimates from the 50 models of $F_i^{(1)}$. Thus the CRPS of $x_{i,t}$ should be low, and when applied to all the observations at station i , the average, $\overline{\text{CRPS}}_i$ should be also low.

The $\overline{\text{CRPS}}_i$ averaged over the observed data from time step $t = 1$ to $t = T_i$ at station i is given as:

$$\overline{\text{CRPS}}_i = \frac{1}{T_i} \sum_{t=1}^{T_i} \int_{\mathbb{R}} \{F_{i,t}(y) - H(y - x_{i,t})\}^2 dy, \quad (\text{V.9})$$

where $H(z)$ denotes the Heaviside function that is 0 if $z \leq 0$ and 1 otherwise. $F_{i,t}(y)$ and $x_{i,t}$ are the CDF of the 50 estimates, and the observed value at time step t of station i respectively. Note that for the cross-validation, if $x_{i,t}$ belongs to $S_i^{(1)}$ (resp. $S_i^{(2)}$), then $F_{i,t}(y)$ will be the CDF of the 50 quantile estimates of the 50 models $F_i^{(2)}$ (resp. $F_i^{(1)}$). The smaller the $\overline{\text{CRPS}}$ score, the better the model.

Given that we have N stations, in the end, we will have N values of $\overline{\text{CRPS}}$ computed for each of the competing models. This is different from the other criteria with 50 or 100 values per model.

V.1.6.3 Stability of high quantile estimate

The robustness of a model is measured by the stability of a high quantile estimated from two sub-samples. A robust model should have a similar estimate of say, a 100-year event, when the sub-sample/calibration data is changed. The SPAN criteria (Garavaglia et al., 2011; Blanchet et al., 2019) gives the measure of the stability of a chosen quantile estimated from two sub-samples.

The score is computed as the absolute difference between the two quantile estimates divided by their average and is given as:

$$\text{SPAN}_{i,T} = \frac{2 \left| \hat{r}_{i,T}^{(1)} - \hat{r}_{i,T}^{(2)} \right|}{\left(\hat{r}_{i,T}^{(1)} + \hat{r}_{i,T}^{(2)} \right)}, \quad (\text{V.10})$$

where $\hat{r}_{i,T}^{(1)}$ and $\hat{r}_{i,T}^{(2)}$ are the T -year return levels estimated from $\hat{F}_i^{(1)}$ and $\hat{F}_i^{(2)}$ respectively at station i .

The score is computed for all the N stations and the regional score $\text{SPAN}_{\text{reg},T}$ is computed as:

$$\text{SPAN}_{\text{reg},T} = 1 - \frac{1}{N} \sum_{i=1}^N \text{SPAN}_{i,T}. \quad (\text{V.11})$$

In the end, we have 50 values of SPAN. A robust model should have a $\text{SPAN}_{\text{reg},T}$ of 1, therefore the closer the value is to 1, the more stable/robust the model is.

V.1.6.4 Reliability in predicting the maximum observed value

The reliability of a model is defined as its ability to associate the correct probability to a given observation. Specifically, the FF criteria measure the reliability of the model in predicting the maximum value in a given sample. The score is defined as:

$$FF_i^{(12)} = \left[\hat{F}_i^{(1)} \left(\max_i^{(2)} \right) \right]^{n_i^{(2)}}, \quad (\text{V.12})$$

where $FF_i^{(12)}$ is the cross validation criteria computed at station i , by predicting the probability of the maximum value in sub-sample 2, $S_i^{(2)}$, of sample size $n_i^{(2)}$ using the model $\hat{F}_i^{(1)}$ fitted on the sub-sample 1, $S_i^{(1)}$.

According to Renard et al. (2013) and Blanchet et al. (2015), if the model is reliable, then $FF_i^{(12)}$ is realization of a uniform distribution. Accordingly, if we compute the score for all the stations, *i.e.* $FF_1^{(12)}, \dots, FF_i^{(12)}, \dots, FF_N^{(12)}$, we should end up with a set $FF^{(12)}$ of N realizations of a uniform distribution. Blanchet et al. (2015) therefore, concluded that, the area between the density of $FF^{(12)}$ and an uniform density should be close to zero.

The regional score $FF_{reg}^{(12)}$ is computed as $1 - \text{AREA}(FF^{(12)})$ and $FF_{reg}^{(21)}$ is computed in similar way. The closer the value is to 1, the more reliable the model is in prediction of the maxima. We have at the end 2×50 values of FF_{reg} .

V.1.7 Results

V.1.7.1 Estimated regions with RFA by upper tail behavior

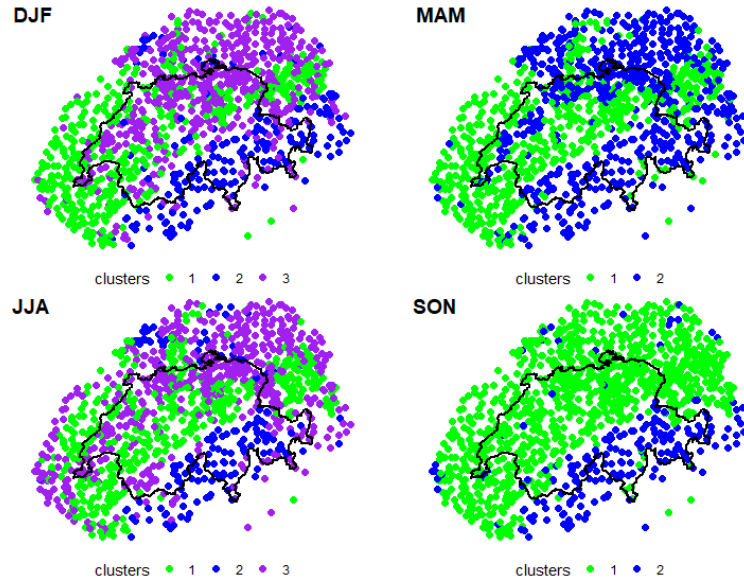


Figure V.2: Maps of Switzerland showing the optimal number of clusters identified with the PAM algorithm for each season. For each season, the regions identified are color coded. From top left, going clock wise, DJF (2 clusters), MAM (2 clusters), JJA (3 clusters) and SON (3 clusters).

Fig V.2 shows the optimal number of clusters identified for each season. We have 3 clusters in the case of winter (DJF), 2 in spring (MAM), 3 clusters in summer (JJA), and 2 in autumn (SON). Notably, although the spatial coordinates are not used, the identified clusters are somehow spatially plausible for each season. Stations in the South are generally in the same cluster. In the North, the stations located in the Northern rim and the Jura generally fall in the same cluster. This is partly

according to our knowledge of the spatial pattern of heavy precipitations in the respective seasons. A few stations for each season, however, appeared in clusters different from their neighbors.

V.1.7.2 Estimated Regions with method of ROI

Figure V.3 presents the neighborhoods found for each station. On the left is the map of Switzerland showing the stations with the size corresponding to the size of the radius identified and the color indicating local fit will be done (for stations without any neighbors) or a regional fit (for stations with at least one neighbor). For those without neighbors, the search terminated sometimes at a very small distance. This means that although they have proximate neighbors according to the crossing distance, homogeneity test failed. For some stations, however, no neighbors are found. These are stations whose closest neighbors are at large crossing distances, and the test failed after application.

On the right of Figure V.3, the histogram of the neighborhood size, as well as the average number of neighbors identified for each class is shown. Only a few stations reach the bound of 100 km. The observed seasonal differences are due to the seasonality and spatial variability of daily precipitation in Switzerland. We recall that the identification of the regions for each station is based on the homogeneity of the scaled extremes. The occurrence of these extremes and their quantity, which affects the test of homogeneity (Evin et al., 2016), depend on the season and hence the observed seasonal differences.

In our subsequent experiments, we keep these neighborhoods and fit accordingly the model versions described in section V.1.3.3.b, and V.1.3.3.c *i.e.* ROI_EGPD_Full and ROI_EGPD_Semi. For all the stations without any neighborhood, we simply fit a local EGPD model.

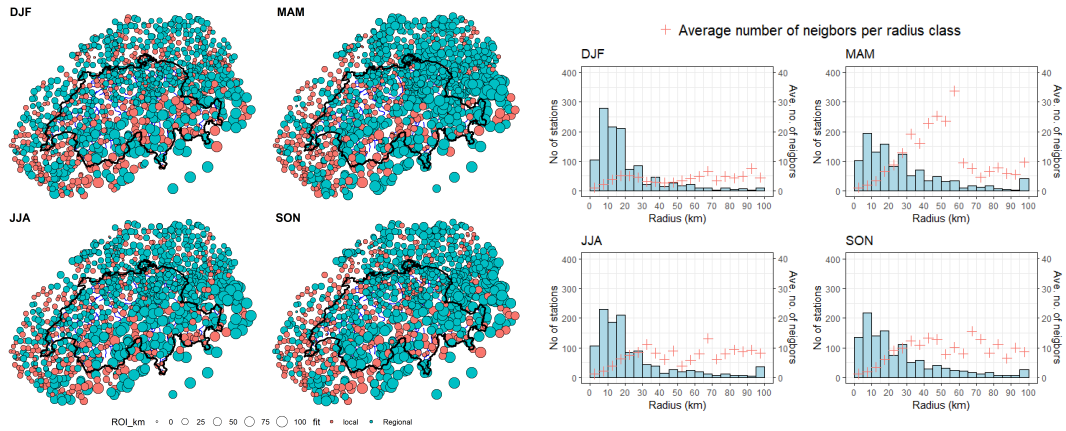


Figure V.3: Properties of the ROI identified for each station. Left: Seasonal maps showing the size of the ROI per station. For each season, the size of the circle is proportional to the ROI size, the smallest is 0 [km], and the largest is 100 [km]. The color of the circle indicates whether a local fit is done, or a regional fit. Right: Histogram of the size of the ROI identified. The red points indicates the average number of neighbors identified for each ROI class.

V.1.7.3 Choice of covariates in GAM

Following the outlined methodology for the spatial model in section V.1.3.2.c, we present in this part the choice of covariate combinations made for the EGPD model.

We use longitude, latitude and the mean daily precipitation to explain the parameters of EGPD, that is k , σ and ξ . Other covariates would be possible but we use these ones because they are

readily available and so estimation at ungauged locations would be possible. After testing different combinations of the covariates, we use the following forms for the relationships:

$$k = s(m) \tag{V.13}$$

$$\sigma = s(lon, lat) + s(m) \tag{V.14}$$

$$\xi = s(lon, lat) \tag{V.15}$$

where $s(lon, lat)$ means a thin plate spline smooth s on the longitude lon and latitude lat , $s(m)$ means a cubic spline on the mean daily precipitation (m).

Although all the three model parameters have to be positive, we still used the identity link function to reduce the complexity in the generation of the gradients of the negative likelihood function of the EGPD model (necessary for model fitting in GAM, see [Wood et al., 2016](#)). We, however, imposed constraints in the likelihood function to ensure that the parameters remain positive.

V.1.7.4 Model comparison

In this section, we present the results of the comparison between the competing models, as judged by the criteria introduced in Section [V.1.4](#). We remind that the sampling was repeated 50 times for all the models, and so for clarity, we show the boxplots of the criteria. Each boxplot contains 50 values of the criteria obtained per run in the case of SPAN, 100 in the case of FF and NRMSE, and 500 in the case of CRPS (500 corresponds to the number of stations in Switzerland, for which the criteria was computed on).

First, the accuracy/reliability of the models in the bulk of the distribution as measured by the NRMSE is shown on the left of Figure [V.4](#). The results are shown per season and the closer the value is to 1, the more accurate the model is. From this result we can clearly see that for all seasons, the two models based on ROI are the most reliable. ROI_EGPD_Full model (where we regionalize according to the method of ROI, all the three parameters of the EGPD model) being the best model compared to ROI_EGPD_Semi model 4 (where only the shape parameter is regionalized, but the other two parameters are locally estimated). The performance of the other regional models is similar and there is no large improvement of these models over the local EGPD model according to this criteria. The CRPS score gives a combined measure of the spread and reliability of the competing models. A model should not only assign the correct probabilities to the observations, but the spread of the probabilities estimated from the different sub samples (100 in our case) should be low as well. We computed this score for all the positive observations at every station. The best model should have the smallest score.

The plot on the right of Figure [V.4](#) presents the seasonal boxplots for the CRPS of the 5 models. Each box plot consists of 500 points, one per station (Recall the scores are computed for only the 500 stations located within Switzerland). From the boxplots and the medians, it is clear that ROI_EGPD_Full has the smallest value of this score. For all seasons, (except summer) the local model has the largest score, showing that the models offer improvement over the local model.

The stability/robustness of the estimate of a 100-year return level in between two periods as measured by the SPAN criteria is shown on the left of Figure [V.5](#). Obviously all the regional models show clear improvement in robustness over the local EGPD model. The spatial model (GAM_EGPD) shows the highest robustness over all the models, except in autumn when it is slightly overtaken by ROI_EGPD_Full model. The results also shows that of all the four regional models, the RFA model based on upper tail (Omega_EGPD) has the smallest robustness. In the case of the two ROI models, ROI_EGPD_Full is more robust compared to the ROI_EGPD_Semi model. The former involves regionalizing all the model three parameters while the latter involves

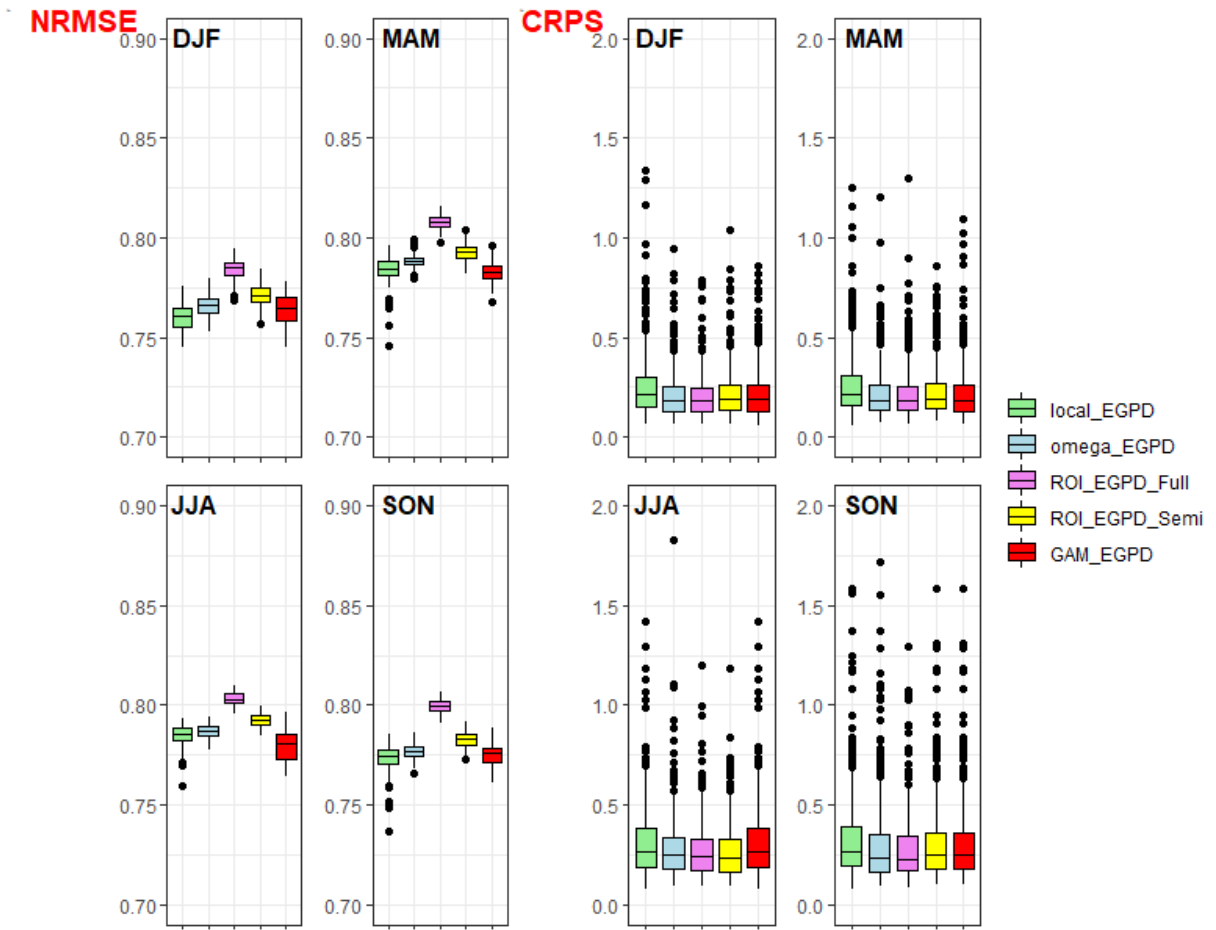


Figure V.4: Criteria applied on the bulk of the distribution for each season. Left: Accuracy of the whole distribution as measured by the NRMSE, each boxplot contains 100 values. Right: CRPS score, each boxplot contain 500 points, 1 per station.

only the shape parameter regionalization. Looking at the two models where only the shape parameter is regionalized, the model based on ROI (ROI_EGPD_Semi) is more robust in all the seasons compared to the model based on upper tail behavior that involves clustering (Omega_EGPD).

Finally, the FF score measures the reliability of the models in the upper tail, more precisely in the prediction of the maximum observed value. This criterion is also optimized at a value of 1. The plot on the right of Figure V.5 shows generally high values of this score for all the models, indicating that they are generally reliable in the prediction of the maximum observed value. For all seasons all the regional models appear to be more reliable compared to the local model. An exception is however in the case of the model (Omega_EGPD) in winter and ROI_EGPD_Semi in summer (looking at the median). In autumn and summer, GAM_EGPD model emerges as the most reliable. Whereas in winter ROI_EGPD_Full is the most reliable, ROI_EGPD_Semi is the most reliable in spring.

To conclude, we present an overall summary of the results in Table V.2 by focusing on the median of the boxplots. We also show the map of seasonal 100-year return level predicted with the ROI_EGPD_Full model in Figure V.6. The maps reveal clear seasonality and spatial pattern. Ticino in the south is subject to the highest levels, especially in the autumn, where up to 400mm can be expected.

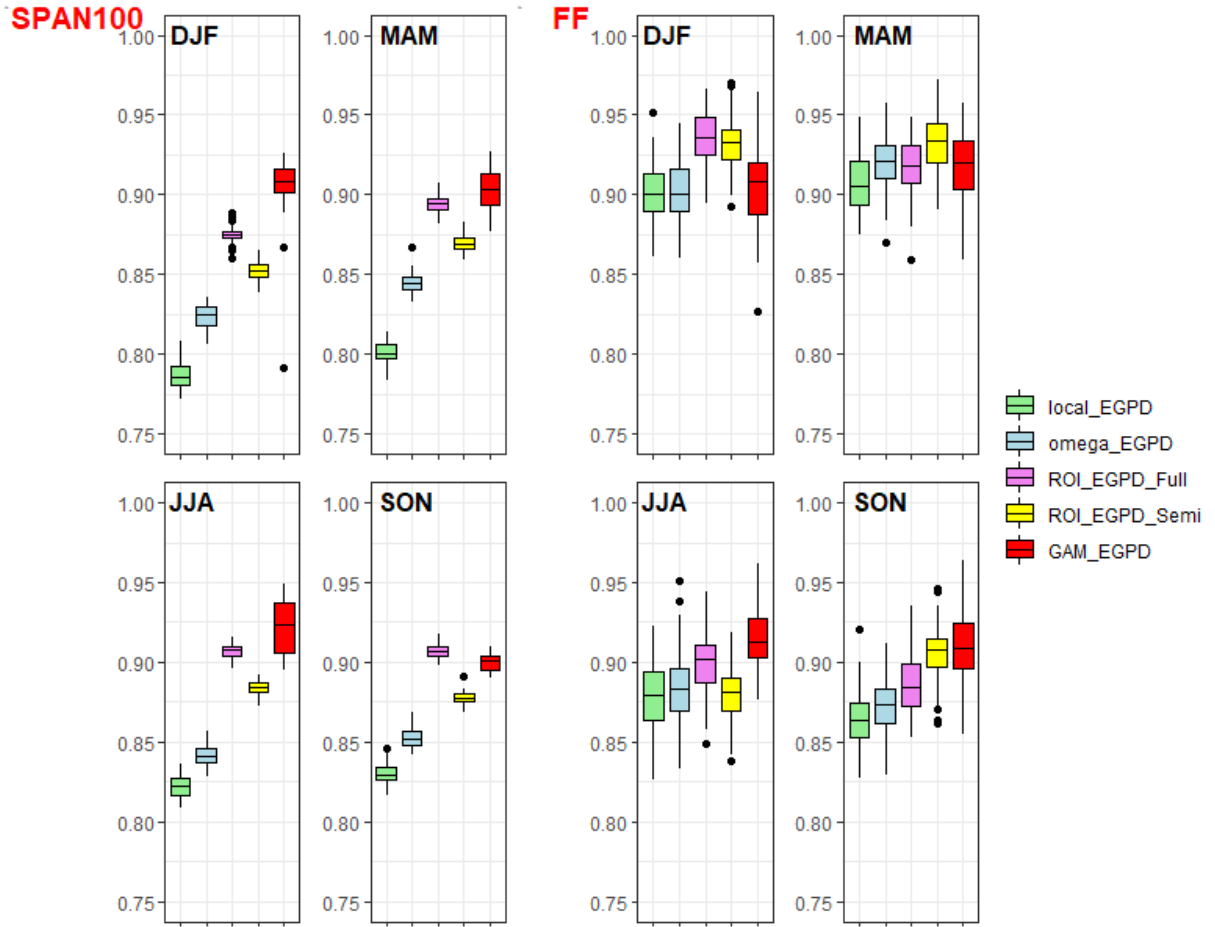


Figure V.5: Criteria applied on the upper tail for each season. Left: Robustness of the local EGPD and the four candidate models, as measured by the SPAN criteria. The stability is measured with respect to a 100-year return level estimate. Each boxplot contain 50 values. Right: Reliability in prediction of the maxima as measured by the FF criteria, each boxplot contain 100 values.

Table V.2: Summary of the comparison results from the four criteria used. For each season and criteria, the model with the highest median is shown. In the case of the CRPS score however, the model with the smallest median CRPS is shown.

Season	NRMSE	CRPS	SPAN 100	FF
Winter	ROI_EGPD_Full	ROI_EGPD_Full	GAM_EGPD	ROI_EGPD_Full
Spring	ROI_EGPD_Full	ROI_EGPD_Full	GAM_EGPD	ROI_EGPD_Semi
Summer	ROI_EGPD_Full	ROI_EGPD_Semi	GAM_EGPD	GAM_EGPD
Autumn	ROI_EGPD_Full	ROI_EGPD_Full	ROI_EGPD_Full	GAM_EGPD

V.1.8 Conclusions and Discussion

The objective of this contribution was to compare three methods to improve the at-site estimates of daily precipitation. By considering a dense network of 1176 stations mainly located in Switzerland, we compared methods based on different philosophies to regionalize the estimation of daily precipitation. The first method defines homogeneous regions based on their upper tail similarity. No covariate is used in the delineation of regions, but only the precipitation data at hand. The

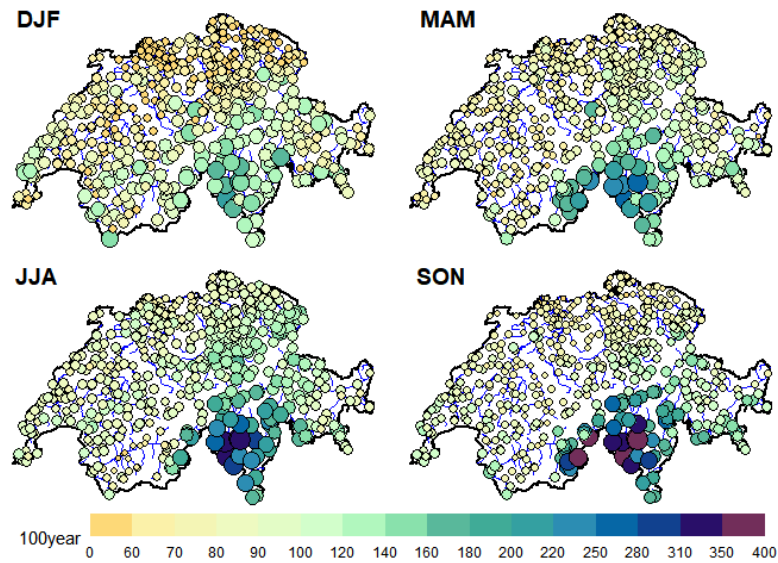


Figure V.6: Map of Switzerland showing the 100year return level for the four seasons as predicted with ROI_EGPD_Full model.

second method avoids defining "hard" clusters but assumes that every station has its homogeneous region that can be identified using homogeneity tests. The third method is spatially based, so all the data is used to estimate smooth and flexible surfaces for the model parameters. Pooling the data to estimate the surfaces thus ensures sharing of information between stations. Using these methods, we built four regional EGD models and compared them. The comparison is based on the accuracy, robustness, and reliability of the models in a cross-validation framework. More precisely, we assessed the performance in both the bulk of the distribution (NRMSE and CRPS) and the upper tail (SPAN and FF).

In contrast to most comparative studies of regionalization approaches that focused on extreme distributions (GEV or POT), we assumed the daily data to follow the EGD. This distribution can adequately model the full spectrum of precipitation intensities. It has the elegant property of being EVT compliant in both the upper and lower tails while providing for a smooth transition in between.

The results showed that regionalization offers improvement in robustness and reliability even in the case of a full-scale model (EGD) that includes the whole data in the estimation of its parameters.

From the four criteria used, the performance depends on the season, but we can still make the following conclusions:

- In terms of the reliability/accuracy over the whole distribution, the ROI model (ROI_EGPD_Full), with all parameters regionalized, emerged as the most accurate.
- Reliability in the prediction of the maxima, as measured by FF, indicated that the GAM model is the most reliable, especially the seasons with the heaviest precipitation (summer and autumn).
- GAM model emerged as the most robust (SPAN) and is followed closely by the ROI model (ROI_EGPD_Full).

In conclusion, two models compete hand in hand; the ROI model (ROI_EGPD_Full) and the GAM model. When we focus on the bulk of the distribution (NRMSE and CRPS), ROI_EGPD_Full is the best model. When we however focus on the far tail (FF and SPAN), the GAM model

is the best. As [Garavaglia et al. \(2011\)](#) pointed out, the two properties of reliability and robustness are complementary. For two models of similar reliability, the model with the best robustness should be preferred. Given this, the GAM model on EGPD, combining both properties in the upper tail can be said to be the preferred method. We note, however, a major drawback of GAM. It requires significant computational time as compared to the ROI, especially in our case where we have a dense network (1176 stations), with long series (up to 156 years for some stations), and we use all the positive precipitation. In practice thus, it would be much easier to use ROI compared to GAM, given that the performance of both is similar, and the former is more reliable when we consider the whole distribution (a feature of interest in our case), not only maxima.

It is worthy to note that in the course of the present study, we focused our evaluation at the station level, where we have observations. A further step will be to assess the models more generally by looking at their performance at ungauged locations in spatial cross-validation as done by [Blanchet and Lehning \(2010\)](#), [Carreau et al. \(2013\)](#) or more generally the framework proposed by [Blanchet et al. \(2019\)](#). In this aspect, the spatial model based on GAM offers a key advantage over the other methods since it inherently results in a regional model that can be applied everywhere. In the case of the other methods, however, the step of choosing the appropriate interpolation technique has to be considered. Also, it is worth mentioning the inherent drawback of conventional RFA approaches involving "hard" clustering, in this case, the Omega_EGPD model. They are known to produce abrupt parameter shifts (in our case, the shape parameter) along the boundaries of the contiguous regions. Again, estimation at ungauged locations between two homogeneous regions with a significant difference in the regional shape parameter will be a difficult decision. The method of ROI however circumvents some of the drawbacks of the conventional RFA approach.

Finally, our approach also assumed the spatial independence of the observations. This assumption will however not be true, especially since we have considered all the positive precipitation. We however expect the benefit of the regional approach to outweigh the consequences of ignoring the spatial dependence ([Hosking and Wallis, 1988](#)), especially since our interest is on the marginal distribution only ([Zheng et al., 2015](#)). An interesting aspect also is to improve the method based on omega (see [V.1.3.2.a](#)) to take into account both the margins and the dependence between sites.

Open Research

The functions to run the EGPD using the GAM model (see Section [V.1.3.3.d](#)) can be found at <https://github.com/abuharuna/egpdGAM>

V.2 Selection of regionalization method for ungauged sites

The comparison in the previous section focused on the performance of the regionalization methods at the station scale. A major point of interest is the issue of ungauged sites, that is locations in the study area where there are no measurements. It will be interesting therefore to check how the models perform at ungauged sites. This involves comparing the models in a spatial cross-validation framework as against the previous case (Section V.1) that involved comparison in a temporal cross-validation framework. We consider two of the models that appeared to compete hand in hand in the previous comparison. They include the GAM_EGPD model based on the spatial approach (GAM) and the ROI_EGPD_Full model based on the region of influence approach (ROI). For ease of reference, we will henceforth simply call them GAM and ROI models respectively.

V.2.1 Parameter regionalization

The GAM model gives automatically a regional model, so the parameters do not need to be interpolated, however, the parameters of the ROI model need to be interpolated. The search for an appropriate technique for interpolation is an immense task of its own since there are different methods, such as linear regression, kriging, regression with smoothing splines, *etc.*, each again with different variants. For instance, within kriging, it is possible to consider ordinary kriging, kriging with external drift, or co-kriging. Fortunately, Blanchet et al. (2019), in a similar context to ours, compared the three methods, along with their variants, and found out that the thin plate spline (TPS) with elevation as external drift is the best method. We therefore follow their footsteps and use the TPS method to interpolate the parameters of the ROI model. In each case, longitude, latitude, and elevation are used as covariates.

V.2.2 Spatial cross-validation framework

In the spatial cross-validation framework, we randomly divide the 1,176 stations into two equal sizes S^1 and S^2 , each containing 50% of the stations (588 stations). These are the same stations used in Section V.1 of this chapter and their locations are shown in Figure V.1, and their description is given in Section V.1.2 of this chapter.

In the first instance, we consider S^1 as the calibration sample and fit the GAM and ROI models to the stations in the sample. The fitting procedure for both models is described in Section V.1.3.3. Next, for the case of the ROI model, we interpolate the fitted EGPD parameters using TPS to estimate the response surface for each parameter. As stated earlier, the GAM model gives the response surface directly for each parameter, so there is no need for further interpolation. We then use the fitted response surfaces from both GAM and ROI to estimate the EGPD parameters at the location of the stations in S^2 to have \hat{F}^1 models. The C^{12} criteria are computed on S^2 stations by comparing their observations with the \hat{F}^1 models. C^{21} criteria are computed symmetrically by using S^2 as the calibration sample and S^1 as the validation sample. The cross-validation procedure is replicated 20 times to address sampling bias.

The criteria we compute in each case are the NRMSE, FF, and SPAN. The reader is referred back to Sections III.5 and V.1.6 for the formulations of these criteria. As a reminder, NRMSE measures the accuracy and reliability of the model over all the observations (the entire distribution), FF assesses the reliability of the model to assign the correct probability to the overall maximum at the validation stations and SPAN evaluates the robustness of a model at extrapolation, *i.e.* the stability of a high return level estimate from the model when the calibration data set is changed. In this application, we compute SPAN for return periods of 100, 1,000, and 10,000 years.

To compute FF and NRMSE, the fitted model is validated against the observations at a given location, so these criteria can only be computed at the validation stations (588 stations). SPAN, on

the other hand, compares return level estimates from two models \hat{F}^1 and \hat{F}^2 at a given location. Hence it can be computed even at ungauged locations by estimating \hat{F}^1 and \hat{F}^2 from the fitted response surface using information from the covariates. Hence, we consider 41,308 fictitious locations situated at a 1 by 1 km grid in Switzerland to compute the SPAN.

V.2.3 Result and discussion

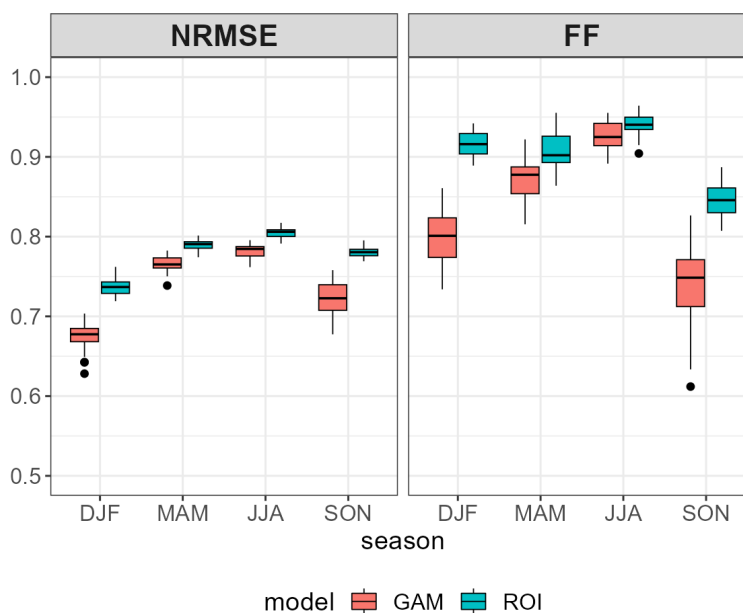


Figure V.7: Left panel: Accuracy of the model in the whole distribution (NRMSE). Right panel: Reliability criteria on the far tail (FF). Each criterion is optimized for a value of 1. Each boxplot contains 40 points corresponding to 20 samplings in a spatial cross-validation framework.

The spatial cross-validation results for the NRMSE and FF criteria are shown in Figure V.7. The best model for both criteria should have a value of one (1). In both cases, the figures reveal that ROI when parameters are interpolated with TPS is the best compared to the GAM model. The robustness criteria shown in Figure V.8 reveal the same result, the ROI model has better performance compared to the GAM model. On an ideal basis, the GAM model should theoretically be the better model since the ROI model relies on post-interpolation of the parameters, however, the results show otherwise. One possible explanation could be that for all seasons, precipitation in Switzerland exhibits significant spatial variability resulting from the interplay of multiple sources and complex topography. Consequently, it was difficult for the GAM model to fit surfaces that were flexible enough to capture/explain this complex variability. Furthermore, we are coupling GAM with a distribution (EGPD) that uses all the non-zero precipitation. This will result in considerable computational demand and difficulty in the optimization of the number of required parameters (coefficients of the basis functions as well as the smoothing parameters). A commonly adopted technique for expediting the computational speed in the case of the large dataset is to set a maximum number of rows, and then to sample the observations without replacement (Youngman, 2020). Nevertheless, since we use all non-zero precipitation, we didn't find this technique practical due to the risk of excluding extreme precipitation events from the analysis.

In summary, based on the results from both temporal and spatial cross-validation, the regional model that combines the ROI approach with TPS interpolations emerges as the most effective method for regionalizing extreme daily precipitation in both gauged and ungauged locations within

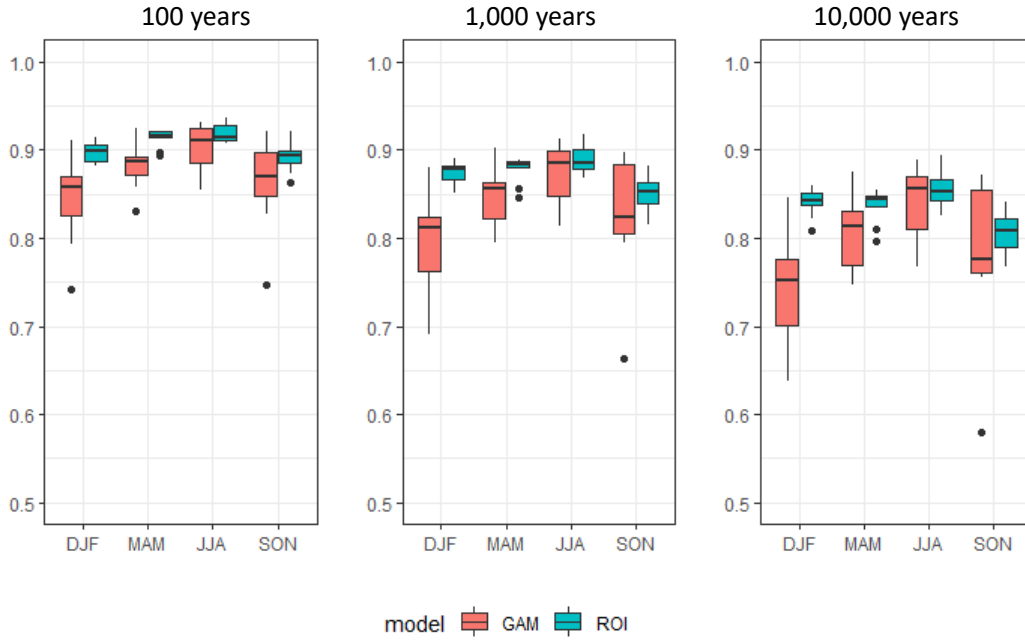


Figure V.8: Stability of 100, 1,000, and 10,000-year return level estimates, as measured by SPAN in a spatial cross-validation framework, for the four seasons. SPAN is optimized for a value of 1. Each boxplot contains 20 points, with each point corresponding to 1 simulation run. The SPAN is averaged over a 1 by 1 km grid covering the entire Switzerland.

our study area. However, a notable limitation of this model is its independent interpolation of the EGP parameters, which overlooks potential dependencies and correlations among the three EGP parameters.

One potential solution to address this limitation involves adopting a hierarchical approach for interpolating the ROI-EGPD parameters, as demonstrated in a prior study by [Deidda et al. \(2021\)](#) for GEV distribution. To apply this approach, four steps, outlined below are followed.

- First, the shape parameter ξ at the station location that was regionalized using the ROI model is interpolated.
- Subsequently, the remaining two parameters, σ , and κ are re-estimated at the station locations, conditioned on the interpolated ξ .
- In the third step, σ is interpolated, while κ is re-estimated conditionally taking into account the interpolated values of both ξ and σ .
- Finally, the re-estimated κ values are interpolated to obtain a fully regional EGP model.

In the future, it will be interesting to employ this hierarchical approach and assess its performance compared to other methodologies. This will contribute valuable insights into the effectiveness of addressing parameter dependencies within the context of extreme precipitation regionalization.

V.3 Summary

To sum up, the main take-home message from this chapter is summarized below:

Question 2: What is the most effective regionalization method to improve the robustness and reliability of daily precipitation estimates in a topographically complex area?

- Three regionalization methods were compared to improve the robustness and reliability of daily precipitation estimates using the EGPD.
- The results show that all the regional models offered improvement compared to the local application of the EGPD.
- Two of the regional models, one based on ROI and the other based on GAM, appeared to compete hand in hand in temporal cross-validation and were further compared in spatial cross-validation.
- The spatial cross-validation results showed that the ROI coupled with thin plate spline interpolation emerged as the best model for both gauged and ungauged locations.
- A possible limitation of the approach is that it neglects the likely correlation between the EGPD parameters and a hierarchical approach was recommended to be explored in the future.

VI

Modeling of Intensity-Duration-Frequency (IDF) Relationships in Switzerland

Chapter overview

This chapter addresses Question 3 by focusing on modeling IDF relationships of precipitation in Switzerland using all the non-zero precipitation intensities. The three-parameter EPGD model is used as the distribution for the non-zero precipitation intensities. Various approaches to building IDF relationships are explored and their performance is compared using some well-chosen criteria. The best approach is retained and used to build IDF curves in Switzerland for hydrological applications.

Contents

VI.1	Paper 2: Modeling Intensity-Duration-Frequency Curves for the Whole Range of Non-Zero Precipitation: A Comparison of Models	80
VI.1.1	Introduction	81
VI.1.2	Data and area under study	83
VI.1.3	Methodology	84
VI.1.4	Results and discussion	96
VI.1.5	Conclusions	106
VI.2	Catchment IDF curves for hydrological applications in Switzerland	108
VI.2.1	Data	108
VI.2.2	IDF curves	108
VI.3	Summary	112

VI.1 Paper 2: Modeling Intensity-Duration-Frequency Curves for the Whole Range of Non-Zero Precipitation: A Comparison of Models

Abubakar Haruna¹, Juliette Blanchet² and Anne-Catherine Favre¹

¹University Grenoble Alpes, Grenoble INP, CNRS, IRD, IGE, Grenoble, France,

²University Grenoble Alpes, CNRS, IRD, Grenoble INP, IGE, Grenoble, France.

Article published on 08 May 2023 in Water Resources Research (WRR)

<https://doi.org/10.1029/2022WR033362>

Preamble to the paper

The goal of this article is to use the extended generalized Pareto distribution (EGPD) to build Intensity-Duration-Frequency (IDF) relationships using non-zero precipitation intensities. To achieve this aim, the article starts by reviewing three outlined approaches for building IDF relationships. The first approach is based on the scale invariance theory, where IDF relationships are built based on the scaling behavior of precipitation. The second method is based on the general IDF formulation of Koutsoyiannis et al. (1998), which generalizes the various traditional IDF formulations. The last approach is called the data-driven approach, where each parameter can vary with duration, and the form of the relationship is empirically determined by the data at hand. Models of IDF relationships are then built based on these three approaches and some extensions to account for scaling breaks and varying shape parameters with duration. This leads to a total of ten IDF models with parameters ranging from four to ten. The models are compared first in calibration and then in a split-sample cross-validation approach. To the best of our knowledge, this article is the first to present an objective comparison of the three outlined IDF curve modeling approaches. Furthermore, while several applications of the EGPD have been presented in the literature, this work presents the first attempt to apply this distribution in IDF curve development. Finally, as far as we know, this study presents the first attempt to model IDF curves for the whole of Switzerland that link parameters with duration, in addition to the traditional IDF curves.

Abstract

Intensity-Duration-Frequency curves are useful in water resources engineering for the planning and design of hydrological structures. As opposed to the common use of only extreme data to build IDF curves, here, we use all the non-zero precipitation intensities, thereby making efficient use of the available information. We use the Extended Generalized Pareto Distribution (EGPD) to model the distribution of the non-zero precipitation intensities. We consider three commonly used approaches for building IDF curves. The first approach is based on the scale-invariance property of precipitation, the second relies on the general IDF formulation of [Koutsoyiannis et al. \(1998\)](#) while the last approach is purely data-driven ([Overeem et al., 2008](#)). Using these three approaches, and some extensions around them, we build a total of 10 models for the IDF curves, and then we compare them in a split-sampling cross-validation framework. We consider a total of 81 stations at 10 min resolution in Switzerland. Due to the marked seasonality of precipitation in the study area, we performed a seasonal-based analysis. The results reveal the model based on the data-driven approach as the best model. It is able to correctly model the observed intensities across duration while being reliable and robust. It is also able to reproduce the space and time variability of extreme precipitation across Switzerland.

VI.1.1 Introduction

Intensity-Duration-Frequency (IDF) curves provide the link between precipitation intensity, duration, and non-exceedance frequency (or rather the return period). It is a very common and useful tool in the area of water resources engineering. IDF curves are practically used to infer high return levels of precipitation intensities for the hydrological designs of structures such as sewer lines, culverts, drains, dams, dykes, etc. They are also used to calibrate/validate stochastic weather generators ([Ritschel et al., 2017](#)).

IDF curves are traditionally modeled by first fitting a statistical model, *e.g.* a Gumbel distribution, to extreme data of each duration separately. Secondly, selected return levels, *e.g.* 2, 5, and 10 years are obtained for each duration from the fitted distribution. And lastly, the inferred return levels are linked to duration by some empirical formulation (*e.g.* [Sherman, 1931](#); [Bernard, 1932](#); [Chow, 1962](#); [Carreteras, 1987](#); [Meylan et al., 2012](#)). Although common and easy to implement, there are several drawbacks to this approach. The approach lacks parsimony because several parameters have to be fitted (a set for each return period). It is not robust in the sense that IDF curves are only available for specific return levels. Thus, whenever a new return level is needed, the process has to be repeated. Another major drawback is that uncertainty in the return levels obtained at the initial steps is not taken into account in the last step. Lastly, there can be intersections between curves of different return levels that cannot be theoretically justified.

To overcome the outlined limitations of the traditional parametric methods, novel approaches were considered to link the different durations together in IDF curves. In general, in spite of the approach, two choices have to be made: the model for the distribution of the precipitation intensities, and the model for the dependence of intensity on duration.

For the choice of the model for the distribution of the precipitation intensities, extreme value distributions are usually considered. For example, in the case of annual maxima series, Generalized Extreme Value (GEV) (*e.g.* [Blanchet et al., 2016](#); [Innocenti et al., 2017](#); [Van de Vyver, 2018](#); [Sane et al., 2018](#); [Mélèse et al., 2018](#); [Ulrich et al., 2020](#); [Jurado et al., 2020](#); [Fauer et al., 2021](#)), or a special case of the GEV, that is Gumbel (*e.g.* [Yu et al., 2004](#); [Agbazo et al., 2016](#); [Chang et al., 2016](#); [Ghanmi et al., 2016](#)). For the peaks over threshold (POT), Generalized Pareto Distribution (GPD) has been used (*e.g.* [Madsen et al., 1995](#); [Ben-Zvi, 2009](#); [Van de Vyver and Demarée, 2010](#)).

Regarding the model for the dependence of intensity on duration, many formulations that are

based on different approaches have been proposed in the literature. Here we identify and focus on three major approaches.

The first approach is based on scale invariance. It has been shown that precipitation exhibits this property within some scales (see [Schertzer and Lovejoy, 1987](#); [Gupta and Waymire, 1990, 1993](#); [Over, 1995](#); [Harris et al., 1997](#); [Lima, 1998](#); [Molnar and Burlando, 2008](#); [Veneziano and Lepore, 2012](#); [Paschalis, 2013](#)). This property provides the physical justification for modeling IDF curves, and thus the possibility of inferring return levels of interest across scales. This approach is arguably the most commonly used approach, possibly because of its rich theoretical background, physical basis, and ease of application in regions with scarce availability of sub-daily precipitation series. IDF curves based on this approach can be found in several applications ([Burlando and Rosso, 1996](#); [Menabde et al., 1999](#); [Willems, 2000](#); [Van de Vyver and Demarée, 2010](#); [Blanchet et al., 2016](#); [Innocenti et al., 2017](#); [Sane et al., 2018](#)).

The second approach is based on the general formulation of [Koutsoyiannis et al. \(1998\)](#), a generalization of the various empirical formulations for modeling IDF curves. This formulation has the key advantage of being a separable function of return levels and duration. It is also consistent with both probabilistic theories and the physical constraints of scaling across duration. Several applications of this formulation to build IDF curves can be found in the literature (e.g. [Koutsoyiannis et al., 1998](#); [Van de Vyver and Demarée, 2010](#); [Blanchet et al., 2016](#); [Sane et al., 2018](#); [Ulrich et al., 2020](#); [Fauer et al., 2021](#); [Roksvåg et al., 2021](#)). This method can be seen as an extension of the scaling approach in which an additional parameter (θ) is added to allow for the curvature of the return levels curves for short durations.

The third approach is based on [Overeem et al. \(2008\)](#) and is called the data-driven approach. Here the functional relationship (linkage) between IDF model parameters and duration is empirically determined from the data itself. The method involves fitting a parametric model, for example, GEV, to data of each duration. A particular regression model is then fitted for each parameter as a function of duration. As a consequence, the return level of any duration can be inferred from the inverse of the distribution, with parameters obtained from the regression model. This approach imposes neither the assumption/existence of scaling nor the separability condition in the case of the general formulation of [Koutsoyiannis et al. \(1998\)](#). Interestingly, both approaches can be seen as special cases of the data-driven approach with particular functional relationships imposed on the parameters. We note that, although the other two approaches also consider the link between parameter and duration, the specific forms rely on some underlying theoretical hypothesis. However, with the data-driven method, the functional relationship is empirically determined from the data itself, hence its name.

There are also nonparametric approaches, which rather than imposing a parametric model on the intensities, use stochastic rainfall models to estimate the IDF curves (for a brief review, see [Langousis and Veneziano, 2007](#); [Veneziano et al., 2007](#); [Tyrakis and Langousis, 2019](#)). Here, we focus on the class that uses parametric models for the intensities.

The common use of GEV and GPD as models for the distribution of the precipitation intensities in IDF curves is justified by the practical use of IDF curves, which is to infer high return levels for hydrological designs. However, there is a major drawback in using these distributions. This is mainly due to poor utilization of already scarce data; only one value per block (in GEV distribution), or only values above a given threshold (as in GPD). This is why [Langousis et al. \(2016\)](#) mentioned that compared to the GEV, GPD is more favored for extreme return level estimation since it makes more efficient use of hydrologic information and hence the resulting decrease in estimation uncertainty. Despite that, the delicate issue of threshold remains with the GPD.

As a remedy, [Naveau et al. \(2016\)](#) proposed the Extended Generalized Pareto Distribution (EGPD) to model all the non-zero precipitation intensities. It has the advantage of using all the information present in the sample of non-zero precipitation data, and not only one value per block (like GEV distribution) or only values above a given threshold (as in GPD distribution), thereby

reducing uncertainty. It doesn't require the choice of GPD threshold and has the advantage of being compliant with extreme value theory (EVT) in both the lower and upper tails. Another advantage is its ability to model all non-zero precipitation. This has applications in water resources management for urban water supplies, hydropower, flood forecast, and irrigation systems. As an example, marginal distributions for the entire range of non-zero precipitation are required in stochastic generators (see an example of application in [Evin et al., 2018](#); [Viviroli et al., 2022](#)). Simulated precipitation from the generators is used as input to conceptual hydrological models for subsequent flood modeling and risk assessment. Other practical applications are in the evaluation of numerical weather simulations or investigation of the climatology of precipitation events as outlined by [Blanchet et al. \(2019\)](#). On the contrary, the assumption of independence in estimation is more likely to be violated in the case of EGPD which uses all the information compared to the GEV and GPD. A common approach to reducing the dependence is to apply temporal declustering, e.g. by taking one-third of the data as done in [Naveau et al. \(2016\)](#); [Le Gall et al. \(2022\)](#); [Haruna et al. \(2022\)](#).

The goal of this article is to use the EGPD to build IDF curves for the non-zero precipitation intensities, based on the three outlined approaches, *i.e.*, scale-invariance, the general formulation of [Koutsoyiannis et al. \(1998\)](#) and the data-driven approach. Due to the marked seasonality of precipitation in the study area, we consider a seasonally-based analysis, with winter (Dec-Jan-Feb), spring (Mar-Apr-May) summer (Jun-Jul-Aug), and autumn (Sep-Oct-Nov). We use a split-sampling cross-validation framework, based on some well-chosen criteria to compare and select the best model. To the best of our knowledge, this study is the first to present an objective comparison of the three outlined IDF curve modeling approaches. Furthermore, while several applications of the EGPD have been presented in the literature (e.g. [Evin et al., 2018](#); [Blanchet et al., 2019](#); [Tencaliec et al., 2020](#); [Rivoire et al., 2021](#); [Le Gall et al., 2022](#); [Haruna et al., 2022](#)), our study presents the first attempt to use this distribution in IDF curves development. In addition to the common application of IDF curves for return level estimation, IDF curves based on all the non-zero precipitation data (modeled with EGPD) will allow for a more comprehensive validation of stochastic weather generators. Finally, apart from the traditional IDF curves (see [Eicher and Krejci, 1996](#)), as far as we know, this study presents the first attempt to model IDF curves that consider linking parameters with duration over the whole of Switzerland.

The rest of the paper is organized as follows: Section [VI.1.2](#) introduces the data and the area under study, and Section [VI.1.3](#) presents the EGPD, the models for the IDF curves, and the evaluation framework. Section [VI.1.4](#) presents the results and discussion and finally, Section [VI.1.5](#) draws the conclusions and gives relevant perspectives.

VI.1.2 Data and area under study

The study area is Switzerland, a small country by size with an area of 41,285 km². Despite its relatively small size, it however presents a complex topography with elevations ranging from 191 m to 4127 m above mean sea level. Around 30% of the area is located above the elevation of 1500 m above sea level. This results in marked spatial variability both in intensity and occurrence of precipitation.

Point precipitation data from a total of 81 stations, with a minimum record length of 20 years, are available for this study. They are spread across Switzerland and their location is shown in [Fig. VI.1](#). Out of this total, 71 stations belong to the SwissMetNet of the Swiss Federal Office for Meteorology and Climatology (MeteoSwiss) while 10 belong to the canton of Lucerne, a partner network of MeteoSwiss. The precipitation data is measured with a tipping-bucket gauge of 0.1 mm depth resolution at a sampling resolution of 10 minutes. Most of the stations at high altitudes are shielded from wind and the tipping gauge is heated in order to account for snow. As a result, heating-related losses can result in up to 24% underestimation compared to the measurements us-

ing an electronic weighting system (Savina et al., 2012). The sample data has a variable length ranging from a minimum of 20 years to a maximum of 40 years from 1981 to 2020. The stations are located at elevations ranging from a minimum of 203 m, an average of 952.4 m, and a maximum of 3294 m.

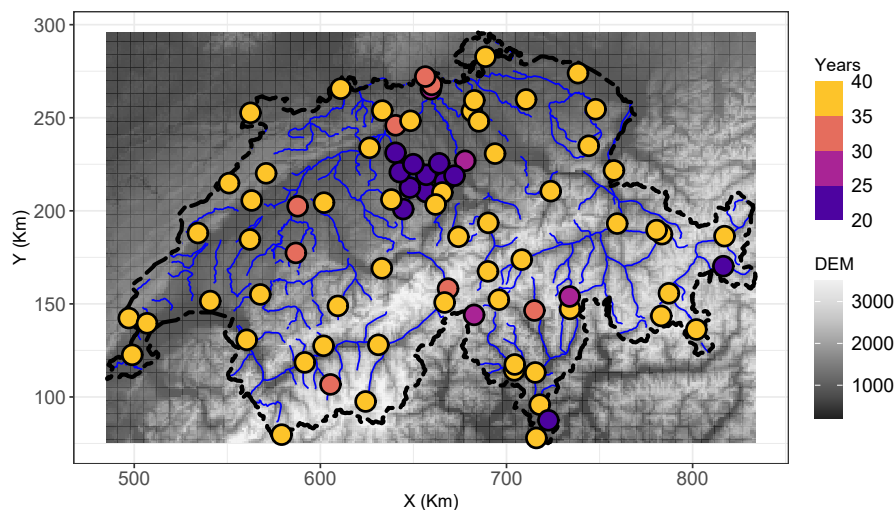


Figure VI.1: Map of Switzerland showing the location of the 81 stations. The color of the points indicates the length of the precipitation data in years. The background color shows the elevation above sea level in meters.

The climate of Switzerland is influenced by the Alps, the Atlantic Ocean, as well as the Mediterranean sea (Sodemann and Zubler, 2009; Giannakaki and Martius, 2015). As a result, precipitation is characterized by seasonality and spatial variability both in intensity and occurrence (Sevruk, 1997; Sevruk et al., 1998; Frei and Schär, 1998; Molnar and Burlando, 2008; Isotta et al., 2014). Annual precipitation is largest in the Alps and its rims, along the Jura (in the northwest), and in the Ticino (south of the Alps). In these areas, annual sums are generally larger than 2000 mm. On the other hand, the inner valleys (Rhône and Inn) receive the least annual precipitation (less than 700 mm). Summer is the main season of precipitation all over Switzerland, except in Ticino where autumn is the main season. For all regions, winter receives the least precipitation. In terms of heavy precipitation (see Panziera et al., 2018), the spatial distribution depends on the accumulation duration. For short-duration accumulations (e.g. 1 hour), the heaviest precipitation occurs in summer everywhere. It is largest in Ticino, Jura, and the northern rim, where maximum summer hourly intensities can reach up to 30 mm/hr. For higher accumulations (1 day and greater), Ticino receives the heaviest precipitation, where a maximum of more than 130 mm over 24 hours can be expected in autumn. For the other regions, the heaviest precipitation almost always occurs in summer.

Due to this marked seasonality, we consider a seasonally-based analysis. We divide the data into four seasons of three months each. Winter includes Dec-Jan-Feb, Spring Mar-Apr-May, Summer Jun-Jul-Aug while Autumn includes Sep-Oct-Nov. A similar seasonal approach was used in previous studies in this area (see Molnar and Burlando, 2008; Fukutome et al., 2015; Evin et al., 2018; Haruna et al., 2022),.

VI.1.3 Methodology

In this section, we start by presenting the distribution for the non-zero precipitation intensities, we then present the various IDF models, and finally the inference strategy to estimate the parameters.

A flowchart for the framework is shown in Figure VI.2.

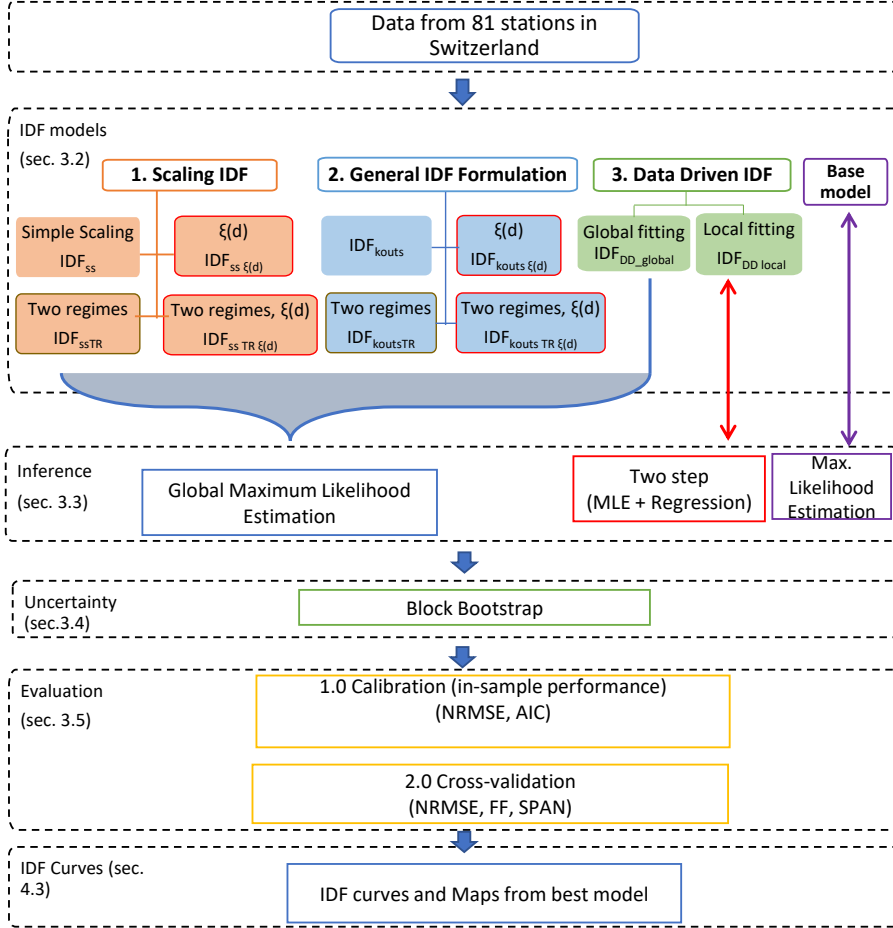


Figure VI.2: Flowchart of the framework for the study

VI.1.3.1 Marginal distribution for non-zero precipitation intensities

We define the random variable I to represent non-zero precipitation intensities. We intend to find a distribution that will allow us to model the entire distribution of I , i.e. both its small, medium, and extreme values. A natural choice would be the Gamma distribution, but it has been reported to underestimate extreme precipitation (Katz et al., 2002).

When looking at only the extremes, the extreme value theory (Coles, 2001) tells us that, the probability that the excesses of I above a high threshold u are less than or equal to a large value i can be approximated by the Generalized Pareto Distribution (GPD) given as:

$$\mathbb{P}(I - u \leq i | I > u) \longrightarrow H_{\xi} \left(\frac{i}{\sigma} \right), \quad (\text{VI.1})$$

with the cumulative distribution function (CDF)

$$H_{\xi} \left(\frac{i}{\sigma} \right) = \begin{cases} 1 - (1 + \xi \frac{i}{\sigma})_+^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - \exp(-\frac{i}{\sigma}) & \text{if } \xi = 0 \end{cases}, \quad (\text{VI.2})$$

where $a_+ = \max(a, 0)$, $\sigma > 0$ is the scale parameter, and $-\infty < \xi < +\infty$ is the GP shape parameter that controls the upper tail of the distribution. The bounded case (short-tailed) is when $\xi < 0$, the exponential case (light-tailed) when $\xi = 0$, and the unbounded case (heavy-tailed) when $\xi > 0$.

The GPD has a key advantage over the GEV since it makes use of all the observations above the threshold u , compared to the GEV case where only the largest observation within a block, usually a season or year, is retained. However, some major issues remain. First, there is yet no unified method for the choice of the optimum threshold u over which excesses are GP distributed. Secondly, the observations that are below the chosen threshold u , although precious and not easy to come by, are discarded and entirely not utilized. Lastly, the question of modeling the whole range of non-zero precipitation which has applications in water resource management remains.

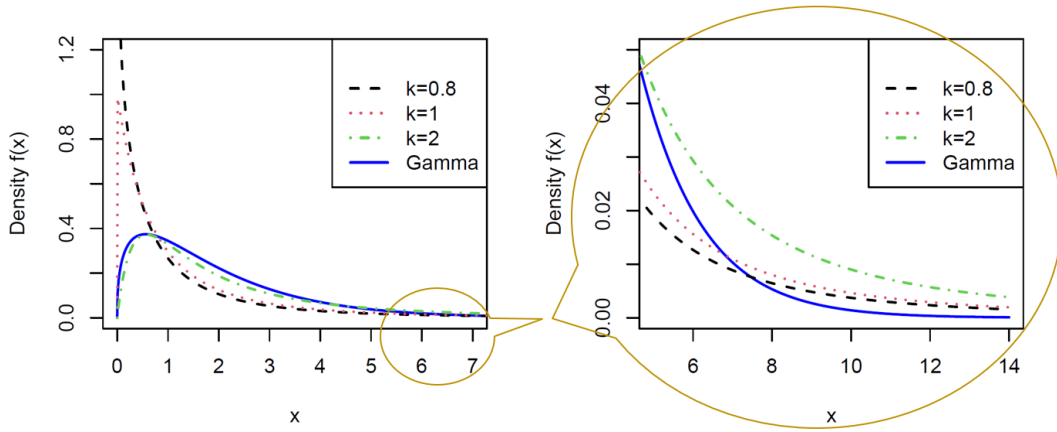


Figure VI.3: Density of the EGPD for $G(v) = v^k$ for $\sigma = 1$, $\xi = 0.5$ and $\kappa = 0.8, 1$ and 2 (adapted from Naveau et al. (2016)). The density of gamma distribution with both shape and scale = 1.4 is shown in blue color. The right panel zooms on the right tail. Indeed the gamma tail decays much faster while having a similar shape in the bulk of the distribution for the case of $\kappa = 2$. The case where $\kappa = 1$ is the exact GP case.

Many parametric, semi-parametric, and non-parametric models have been proposed in the literature as extensions to the GPD to model the full distribution of I , and not only its upper tail as given in Equation VI.1 (Scarrot and MacDonald, 2012, see review in). In particular, Naveau et al. (2016) proposed a family of models they called the extended generalized Pareto distributions (EGPD) that are defined as:

$$\mathbb{P}(I \leq i) = G \left[H_{\xi} \left(\frac{i}{\sigma} \right) \right], \quad (\text{VI.3})$$

where G is a continuous CDF that is also defined on $[0,1]$, with constraints given in Naveau et al. (2016) to ensure that the distribution of I remains GP in its upper tail, a gamma-like lower tail, and a smooth transition in between. Four parametric families of G were proposed by the authors, the simplest of which is defined as $G(v) = v^k$. The CDF of this model is defined as:

$$\mathbb{P}(I \leq i) = \left[H_{\xi} \left(\frac{i}{\sigma} \right) \right]^k. \quad (\text{VI.4})$$

The model thus has three parameters. $k > 0$ controls the lower tail, $\sigma > 0$ is the scale parameter, and $\xi \geq 0$ controls the upper tail. $H_\xi\left(\frac{i}{\sigma}\right)$ is as defined in Eq. VI.2. We exclude the bounded case ($\xi < 0$) since we are dealing with precipitation. We thus assume precipitation to be either light-tailed ($\xi = 0$) or heavy-tailed ($\xi > 0$). The density of the model, alongside that of the Gamma distribution, is illustrated in Figure VI.3. With just one additional parameter, κ , the distribution is parsimonious and able to adequately model the full range of non-zero precipitation (see applications in Evin et al., 2018; Le Gall et al., 2022; Haruna et al., 2022). We thus use this model in the rest of the paper.

VI.1.3.2 IDF models

We define the random variable I_d as the average non-zero precipitation intensity over the duration d . It is described by the CDF, $F_d(i)$, such that $F_d(i) = \mathbb{P}(I_d < i)$. The exceedance frequency is defined as $p_d(i) = 1 - F_d(i)$. The return period of any non-zero intensity i , as a function of p_d is given by $T(I_d \geq i) = \frac{1}{p_d \times \delta_d}$, with δ_d the average number of non-zero precipitation intensities of duration d per year. We estimate δ_d based on the long-term average of the non-zero precipitation intensities per year. Consequently, the T -year return level over duration d , $i(T, d)$, is defined as the $(1 - \frac{1}{T \times \delta_d})$ quantile of F_d .

Accordingly, IDF relationship is a mathematical function $(T, d) \mapsto i(T, d)$ that relates non-zero precipitation intensity i with its duration d , and the frequency of exceedance p_d (or rather the return level T). In this article, the CDF of I_d , $F_d(i)$ is defined by the EGPD presented in Section VI.1.3.1. All the different formulations considered here simply differ by how they define this mathematical relationship between i , T and d , while taking $F_d(i)$ as an EGPD model.

In the following subsections, we present the different IDF-EGPD models based on the three outlined approaches, *i.e.*, scale-invariance, the general formulation of Koutsoyiannis et al. (1998), and data-driven approaches. For sake of simplicity, we drop the "EGPD" term and simply refer to the IDF models as $\text{IDF}_{\text{modelname}}$, where the subscript "modelname" refers to the approach used to build the model.

For all the models, the IDF curve, corresponding to the $(1 - \frac{1}{T \times \delta_d})$ quantile of the EGPD is defined in Eq. VI.5 as:

$$i(T, d) = \frac{\sigma_d}{\xi_d} \left\{ \left(1 - \left[1 - \frac{1}{T \times \delta_d} \right]^{\frac{1}{\kappa_d}} \right)^{-\xi_d} - 1 \right\}, \quad (\text{VI.5})$$

where $\kappa_d > 0$, $\sigma_d > 0$ and $\xi_d \geq 0$ are the three EGPD parameters for the duration d , T is the return period in years, δ_d is the average number of non-zero precipitation intensities per year for the duration d . The choice of the model determines whether each of the three parameters; κ , σ , and ξ varies with d or not, and the form of the relationship.

We consider thirteen durations, *i.e.*, $d = 30$ min, 40 min, 1, 2, 3, 6, 10, 12, 16, 18, 24, 48, and 72 hours. We consider durations up to 72 hours (3 days) because according to Froidevaux et al. (2015), precipitation accumulations from 0 to 3 days before an event are the most relevant for triggering floods of high magnitudes in Switzerland. The intermediate durations are meant to ensure a good spread in a logarithmic scale and for a later comparison of our quantiles with those provided by MeteoSwiss (not in this paper). We use a fixed window to aggregate the data from the gauge resolution of 10 mins, to the various durations. For instance, the 24hr intensities correspond to amounts accumulated from 00h00 to 24h00 UTC of every day, divided by 24. We chose this time window because, in the study area, convective events are mostly in the evening. This will allow us to differentiate between the first and last 12 hours starting from 00 hrs. The choice of a fixed window over a moving window will likely result in omitting the highest intensities in each

duration. However, since we are using all the non-zero precipitation intensities, using a moving window will result in significant dependence in the time series.

VI.1.3.2.a Scaling IDF model: Scale invariance in the strict sense of [Gupta and Waymire \(1990\)](#) refers to the property where the probability distribution of I_d can be inferred from the distribution of I_{d_0} at the reference duration d_0 through:

$$I_d \stackrel{dist}{=} C_\lambda I_{d_0}, \quad (\text{VI.6})$$

where the parameter C_λ determines the type of scaling; simple-scaling or multi-scaling. For our case, the reference duration d_0 is taken as 1 hour.

A weaker assumption, the so-called "wide sense scaling" ([Gupta and Waymire, 1990](#)), is when the scaling is in the moments according to:

$$\mathbb{E} [I_d^q] = \left(\frac{d}{d_0} \right)^{-k(q)} \mathbb{E} [I_{d_0}^q], \quad (\text{VI.7})$$

where q is the order of the moment, $k(q)$ is called the moment scaling function, d_0 is the reference duration. Moment scaling analysis as described by [Gupta and Waymire \(1990\)](#) is used to determine the type of scaling.

When the parameter in Eq. VI.6, $C_\lambda = \left(\frac{d}{d_0} \right)^{-H}$, i.e, a scalar that depends only on the ratio of the scales, we have "strict sense simple-scaling". This is expressed in Eq. VI.8.

$$I_d \stackrel{dist}{=} \left(\frac{d}{d_0} \right)^{-H} I_{d_0}, \quad (\text{VI.8})$$

where $0 < H < 1$ is the scaling exponent. The other variables retain their meanings.

Wide sense simple-scaling is when the moment scaling function in Eq. VI.7 is linear in q , i.e. $k(q) = Hq$, as expressed in Eq. VI.9.

$$\mathbb{E} [I_d^q] = \left(\frac{d}{d_0} \right)^{-Hq} \mathbb{E} [I_{d_0}^q]. \quad (\text{VI.9})$$

It can be shown that, under the strict sense simple-scaling, only one parameter of the EGPD is scaling, which is σ , whereas κ and ξ are independent of duration. For the rest of the paper, we drop the term "strict sense", and simply use "simple scaling" for convenience. Accordingly, the simple-scaling EGPD model, IDF_{ss} , is defined such that: $\kappa_d = \kappa_{d_0}$, $\xi_d = \xi_{d_0}$, and σ_d is a power law given as:

$$\sigma_d = \left(\frac{d}{d_0} \right)^{-H} \sigma_{d_0}, \quad (\text{VI.10})$$

where $\kappa_{d_0} > 0$, $\sigma_{d_0} > 0$ and $\xi_{d_0} \geq 0$, are the parameters of the reference duration $d_0 = 1$ hour, and $0 < H < 1$ is the scaling exponent. The inference method for the parameters is described in section VI.1.3.3.

An important issue is the existence of multiple scaling regimes in precipitation. This means that different scaling behaviors (scaling exponents) exist for different ranges of duration. IDF curves have to be modeled considering the existence of this change in scaling (e.g. [Yu et al., 2004](#); [Bougadis and Adamowski, 2006](#); [Courty et al., 2019](#)). To illustrate this behavior, we consider the case of a station at Robbia in Graubünder in winter. We fit the EGPD to the data of each of the 13 durations separately by maximum likelihood. We then inspected how the estimated scale parameter (σ) varies with duration. The result is given in Fig. VI.4a. Here, a single power law

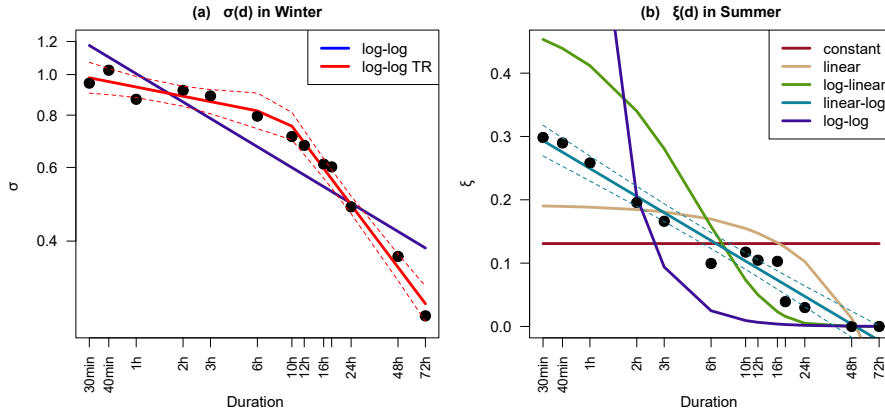


Figure VI.4: Illustration of **a)** Break in scaling of the σ parameter in winter at a station, Robbia in Graubünden. The points colored black are the estimated σ for each duration separately. The models in blue and red are given in Eq. VI.10 and VI.11 respectively. **b)** Dependence of ξ on duration in summer at a station in Zurich. The black colored points are the estimated ξ for each duration separately. The lines are the fitted linear models whose equations are given in A.2.1. The broken lines represent the 95% confidence interval for the fitted models. The plots of the complete EGPD parameters with duration at these two stations are also given in Appendix A.2.1.2 for the four seasons.

(log-log given in Eq. VI.10) in blue is not enough to explain the scaling. To account for this break in the scaling relationship, we define the two-regime simple-scaling EGPD IDF model, IDF_{ss_TR} as:

$$\sigma_d = \begin{cases} \left(\frac{d}{d_0}\right)^{-H_1} \sigma_{d_0} & \text{if } d \leq K \\ \left(\frac{d}{d_0}\right)^{-H_2} \sigma_{d_0} \times K^{H_2-H_1} & \text{if } d > K \end{cases} \quad (\text{VI.11})$$

An illustration of this equation is in Figure VI.4a (log-log TR in red). Where K is the duration of the scaling break, $\sigma_{d_0} > 0$ is the scale parameter for the reference duration $d_0 = 1$ and σ_d is continuous in $d = K$. The other parameters, $0 < H_1 < 1$ and $0 < H_2 < 1$ are the scaling exponents of the first and second regimes. The other two EGPD parameters $\kappa_d = \kappa_{d_0}$, $\xi_d = \xi_{d_0}$, remain independent of durations. Hence for this model, a total of six parameters have to be estimated, *i.e.*, κ_{d_0} , σ_{d_0} , ξ_{d_0} , H_1 , H_2 and K .

Lastly, although the simple-scaling EGPD model imposes a constraint on the linkage of ξ with d , *i.e.*, $\xi_d = \xi_{d_0}$, we however notice some of the stations to show an apparent variation of ξ with respect to d . For illustration, we consider a station in Zurich in summer. We fit EGPD to the data of each of the 13 durations separately by maximum likelihood. We then inspected how the estimated shape parameter (ξ) varies with duration and modeled the relationship through a linear-log form as expressed in Eq. VI.12. Fig. VI.4b illustrates this and how the linear-log model fits the points correctly:

$$\xi_d = a_\xi + b_\xi \log(d), \quad (\text{VI.12})$$

where a_ξ and b_ξ are the intercepts and slopes respectively. This leads to two additional IDF models, with $\xi = f(d)$, namely:

- $IDF_{ss_xi(d)}$: an extension of the basic simple-scaling model IDF_{ss} , to allow ξ to vary with d according to Eq. VI.12.

- $\text{IDF}_{ss_TR_ξ(d)}$: an extension of the two-regime simple-scaling model IDF_{ss_TR} , to allow $ξ$ to vary with d according to Eq. VI.12.

VI.1.3.2.b General IDF formulation: Koutsoyiannis et al. (1998) proposed a general formulation for the different traditional formulations of the IDF curves in the literature. He showed that all of them can be simplified into the form:

$$i(T, d) = \frac{a(T)}{b(d)}, \quad (\text{VI.13})$$

where $b(d) = (d + \theta)^H$. The parameter $\theta > 0$ is the duration offset, and $0 < H < 1$ is the duration exponent. $a(T)$ is the $(1 - \frac{1}{T})$ quantile of the re-scaled intensities $I_d b(d)$. $a(T)$ is independent of d and completely determined by the statistical model considered for I_d , in our case, the EGPD.

This formulation has the key advantage of being a separable function of return levels $a(T)$, and duration $b(d)$ that is consistent with both probabilistic theories and the physical constraints of scaling across duration. Menabde et al. (1999) showed that this formulation is the same as the scale-invariant model if θ is set to zero.

When applied to the EGPD, IDF_{kouts} is defined such that: $\kappa_d = \kappa_{d_0}$, $\sigma_d = \left(\frac{d+\theta}{d_0+\theta}\right)^{-H} \sigma_{d_0}$, $\xi_d = \xi_{d_0}$. Five parameters, $\kappa_{d_0} > 0$, $\sigma_{d_0} > 0$, $\xi_{d_0} \geq 0$, $\theta > 0$ and $0 < H < 1$ have to be inferred. The reference duration here is taken as $d_0 = 1 - \theta$.

Following the same arguments discussed in Section VI.1.3.2.a regarding the existence of a break in the scaling relationship, and the dependence of $ξ$ with d , we propose three extensions to this model:

- IDF_{kouts_TR} : Allowing for a break in the scaling regime. This model is defined as:

$$\sigma_d = \begin{cases} \left(\frac{d+\theta}{d_0+\theta}\right)^{-H_1} \sigma_{d_0} & \text{if } d \leq K \\ \left(\frac{d}{d_0}\right)^{-H_2} \sigma_{d_0} \times K^{H_2-H_1} & \text{if } d > K \end{cases}, \quad (\text{VI.14})$$

where K is the duration of the scaling break, $\sigma_{d_0} > 0$ is the scale parameter for the reference duration $d_0 = 1 - \theta$ and σ_d is continuous in $d = K$. The other parameters, $0 < H_1 < 1$ and $0 < H_2 < 1$ are the scaling exponents of the first and second regimes. The other two EGPD parameters $\kappa_d = \kappa_{d_0}$, $\xi_d = \xi_{d_0}$, remain independent of durations. Hence for this model, a total of six parameters have to be estimated, *i.e.*, κ_{d_0} , σ_{d_0} , ξ_{d_0} , H_1 , H_2 and K .

- $\text{IDF}_{kouts_ξ(d)}$: an extension of the basic model IDF_{kouts} , to allow $ξ$ to vary with d according to Eq. VI.12.
- $\text{IDF}_{kouts_TR_ξ(d)}$: an extension of the two-regime model IDF_{kouts_TR} , to allow $ξ$ to vary with d according to Eq. VI.12.

VI.1.3.2.c Data-Driven IDF model: The scaling theory and the specific form of Eq. VI.13 impose particular functions for the relation between the scale parameter, σ of the EGPD with respect to duration, d . However, in the case of the data-driven models, the expression of the relationship for each of the three EGPD parameters is empirically determined by the data itself. To guide our choice of the appropriate functional relationship, we inspected how each locally estimated EGPD parameter varies with duration. Fig. VI.4 gives an example for the σ and $ξ$ parameters at two stations. We finally settled on the following functions to model the three parameters with respect to duration:

$$\kappa_d = \begin{cases} \exp[a_\kappa + b_{1,\kappa} \log(d)] & \text{if } d \leq K_\kappa \\ \exp[a_\kappa + b_{2,\kappa} \log(d) + (b_{1,\kappa} - b_{2,\kappa}) \log(K_\kappa)] & \text{if } d > K_\kappa \end{cases}, \quad (\text{VI.15})$$

$$\sigma_d = \begin{cases} \exp[a_\sigma + b_{1,\sigma} \log(d)] & \text{if } d \leq K_\sigma \\ \exp[a_\sigma + b_{2,\sigma} \log(d) + (b_{1,\sigma} - b_{2,\sigma}) \log(K_\sigma)] & \text{if } d > K_\sigma \end{cases}. \quad (\text{VI.16})$$

For the first two parameters, κ and σ , the function is a continuous two-linear piece-wise model in log space. K_* is the duration of the break-point (σ continuous for $d = K_*$). a_* , $b_{1,*}$, $b_{2,*}$ are the intercepts and slopes of the first and second lines respectively. In the case of ξ , the function was already given in Eq. VI.12.

Note that, by keeping κ and σ independent of duration, and using either $\sigma_d = \left(\frac{d}{d_0}\right)^{-H} \sigma_{d_0}$ or $\sigma_d = \left(\frac{d+\theta}{d_0+\theta}\right)^{-H} \sigma_{d_0}$, the simple-scaling or the general formulations of Koutsoyiannis et al. (1998) presented in Section VI.1.3.2.a and VI.1.3.2.b respectively can be obtained from this data-driven approach.

We consider two IDF models in this class, $\text{IDF}_{DD_{global}}$ and $\text{IDF}_{DD_{local}}$, both impose the same type of functional relationships (Eq. VI.15, VI.16, VI.12), but simply differ in the way the regression parameters are estimated. The inference strategy is explained in detail in section VI.1.3.3.

The different models compared in this study are summarized in Table VI.1.

Table VI.1: Summary of the IDF models that are compared in this study.

Model	No. of Params	Name of Approach	Inference Method	Ref. Section
1 IDF_{ss}	4	Simple-scaling	Global MLE ^a	VI.1.3.2.a
2 $\text{IDF}_{ss_{TR}}$	6	Simple-scaling	" "	VI.1.3.2.a
3 $\text{IDF}_{ss_{\xi}(d)}$	5	Extension of Simple-scaling	" "	VI.1.3.2.a
4 $\text{IDF}_{ss_{TR_{\xi}(d)}}$	7	Extension of Simple-scaling	" "	VI.1.3.2.a
5 IDF_{kouts}	5	Koutsoyiannis et al. (1998)	" "	VI.1.3.2.b
6 $\text{IDF}_{kouts_{TR}}$	7	Ext. of Koutsoyiannis et al. (1998)	" "	VI.1.3.2.b
7 $\text{IDF}_{kouts_{\xi}(d)}$	6	Ext. of Koutsoyiannis et al. (1998)	" "	VI.1.3.2.b
8 $\text{IDF}_{kouts_{TR_{\xi}(d)}}$	8	Ext. of Koutsoyiannis et al. (1998)	" "	VI.1.3.2.b
9 $\text{IDF}_{DD_{global}}$	10	Data-driven	" "	VI.1.3.2.c
10 $\text{IDF}_{DD_{local}}$	10	Data-driven	Two-step (MLE + Regr)	VI.1.3.2.c

^a Maximum Likelihood Estimation

VI.1.3.3 Inference

In this section, we describe the inference methods to estimate the parameters of the ten (10) IDF models presented in Table VI.1.

As a prerequisite to an objective comparison, the same estimation strategy has to be employed for all the models. For this reason, we use a global maximum likelihood estimation for all the models (as done by Blanchet et al., 2016), which we describe in the next paragraph. An exception is the case of only one model, the $\text{IDF}_{DD_{local}}$, which involves a two-step method. In this case, first,

we fit, for a given station, the EGPD on the data of each duration separately (using maximum likelihood estimation). Secondly, we fit for each fitted parameter, the chosen regression model as a function of duration. The parameters in Eq. VI.15 and VI.16 are estimated by segmented regression (i.e regression model with break-points), while those of Eq. VI.12 by ordinary least squares (OLS). The segmented regression we use here is based on the algorithm of Muggeo (2003), and with the functions that are implemented in the **R** segmented package. Details of the algorithm can be found in Muggeo (2003).

We now come back to the global maximum likelihood for the other nine (9) IDF models. This method involves pooling, for each station, all the data from the thirteen durations to estimate the model parameters. The duration d is used as a covariate. We note here that by pooling all the data, we made the hypothesis of independence between the intensities of the different time steps and different durations (since we use independence likelihood (as in Blanchet et al., 2016)). This hypothesis is difficult to justify given that we consider all the non-zero intensities. Note that by taking fixed windows rather than moving windows, we have reduced the dependence between the different time steps. Previous studies considered similar approaches to reduce temporal dependence between different time steps, by taking a fraction of the data (Naveau et al., 2016; Le Gall et al., 2022; Haruna et al., 2022)). The dependence between the intensities of different duration e.g. intensities of 1 hour vs those of 2 hours, however, remains. While it might be possible to account for this dependence using different strategies proposed in the literature (Davison et al., 2012; Seville et al., 2017, e.g.), we believe this requires a complete task of its own and will likely complicate the optimization of the models. This is in addition to the possibility of performance deterioration due to dependence structure misspecification. We take this as a limitation of the present study and a topic for future research.

The log-likelihood (ll) that is maximized here (given in Eq. VI.17) takes left censoring into account. The importance of using left censoring in fitting precipitation data by maximum likelihood has been pointed out by Naveau et al. (2016), and they showed that better performance is obtained by taking it into account.

$$ll_{EGPD}(\kappa_d, \sigma_d, \xi_d) = ll_{censored}(\kappa_d, \sigma_d, \xi_d) + ll_{uncensored}(\kappa_d, \sigma_d, \xi_d), \quad (\text{VI.17})$$

where $ll_{censored}$ and $ll_{uncensored}$ are the contributions of the censored and uncensored data, given in Eq. VI.18 and VI.19 respectively, as

$$ll_{censored}(\kappa_d, \sigma_d, \xi_d) = \sum_d \sum_{j:i_d < c_d} \kappa_d \log \left[1 - \left(1 + \frac{\xi_d c_d}{\sigma_d} \right)^{-\frac{1}{\xi_d}} \right], \quad (\text{VI.18})$$

$$ll_{uncensored}(\kappa_d, \sigma_d, \xi_d) = \sum_d \sum_{j:i_d \geq c_d} \log \kappa_d - \sum_d \sum_{j:i_d \geq c_d} \log \sigma_d - \sum_d \sum_{j:i_d \geq c_d} \left[1 + \frac{\xi_d i_{d,j}}{\sigma_d} \right]^{1 + \frac{1}{\xi_d}} + \sum_d \sum_{j:i_d \geq c_d} \left[1 - \left(\left(1 + \frac{\xi_d i_{d,j}}{\sigma_d} \right)^{-\frac{1}{\xi_d}} \right) \right]^{[\kappa_d - 1]} \quad (\text{VI.19})$$

where d ranges over the 13 durations and j ranges over the number of time steps in the data of duration d . $\kappa_d > 0$, $\sigma_d > 0$, $\xi_d \geq 0$ are the EGPD parameters for duration d . $i_{d,j}$ is the precipitation intensity for the duration d and time step j . $c_d \geq 0$ is the left censoring threshold applied to the data of duration d . Many authors have taken this censoring approach into account but they usually take a uniform threshold value for all the stations (e.g. Tencaliec et al. (2020) used 2 mm for daily precipitation). Here we didn't find the use of a common threshold over the 81 stations sufficient. We had to select, for each station and duration, the lower threshold c that minimizes the Normalized Root Mean Square Error (NRMSE) of Eq. VI.20 in Section VI.1.3.5.

In both Eqs. VI.18 and VI.19, the choice of the IDF model specifies the function linking the EPGD parameters to duration. For instance, in the case of the simple-scaling, $IDF_{DD_{ss}}$, $\kappa_d = \kappa_{d_0} \sigma_d = \left(\frac{d}{d_0}\right)^{-H} \sigma_{d_0}$ $\xi_d = \xi_{d_0}$.

For the cases where two linear models are fitted (see Eq. VI.11, VI.14, VI.15, VI.16, VI.12), we use segmented regression to estimate the regression parameters. We then use the fitted parameters as initial values in the optimization of the likelihood function.

We note here that other estimation methods were used in other studies for the simple-scaling approach and the general IDF formulation of Koutsoyiannis et al. (1998), besides the global MLE. In the case of the simple-scaling models, a two-step procedure, where the scaling exponent in Eq. VI.8 is first obtained through moment scaling analysis, then all re-scaled intensities from all the durations are pooled to fit the IDF model using MLE for example, (see Nhat et al., 2008; Panthou et al., 2014; Innocenti et al., 2017). For the general IDF formulation of Koutsoyiannis et al. (1998), the authors proposed two different estimation strategies; the so-called 'robust estimation', and the 'one-step least square method'. The robust estimation is a two-step procedure that involves the estimation of the parameters of $b(d)$, and then those of $a(T)$ (see Eq. VI.13), through the minimization of the Kruskal-Wallis statistic. The one-step least square method involves the joint estimation of all the parameters of Eq. VI.13 that minimizes the squared error of the observed and modeled quantiles from the IDF model. Despite this, we only use the global MLE method in order to objectively compare the models based on the same inference strategy. In addition, it has the advantage of being a one-step estimation procedure and is better suited than least squares for non-gaussian distribution estimation.

VI.1.3.4 Uncertainty Estimation

We employ a block bootstrap approach for the uncertainty estimation in all the models considered in this study. The principle of the block bootstrap involves sampling with replacement, all the data contained in a block of a given size B , R number of times. Here, for computational reasons, we use $R = 500$. We then use the percentile method to estimate the 95% confidence interval (CI). The choice of the appropriate block size B is a delicate issue in the scientific community, but one common way is to choose B large enough to ensure that the temporal dependence in the data is maintained. A block size B too small will underestimate the uncertainty. Here, we check the autocorrelation in the seasonal data (result not shown), and saw that it does not decrease after a lag of more than 1 week. Hence we choose $B = 2$ weeks. To maintain the dependence between data of different durations d , we always ensured that all the data of the different durations d contained in the same block $B = 2$ weeks are sampled together. A summary of the block bootstrapping is summarized in the steps below:

For each station s and season:

1. Aggregate the data to intensities of the 13 duration d , store the data in a matrix of n by d . Where n is the number of observations and $d = 13$. We call this matrix, M_{orig} .
2. Sample with replacement, $\frac{n}{B}$ blocks from the row of matrix M_{orig} in the previous step to form the bootstrap matrix M_{boot} . Both M_{orig} and M_{boot} have the same dimensions. By sampling from the rows of M_{orig} , we keep the data of the different durations d together, and hence the dependence structure.
3. Fit the IDF model on the data in M_{boot} and estimate the intended return level.
4. Repeat step 2 to 3 a total of 500 times ($R = 500$) to obtain the bootstrap distribution of the return level.

5. Obtain the 95% CI of the intended return level by percentile method. This is done by taking the empirical 0.025 and 0.975 quantiles of the bootstrap distribution of the return level obtained in step 4.

We note here that to ensure all the models use the same bootstrap matrix M_{boot} , we use "set.seed" in **R** to keep track of the random number generation in the sampling of step 2. Examples of the CI obtained with this approach are later given in Figure VI.10.

VI.1.3.5 Evaluation framework

We evaluate the performance of the models in two aspects. First, in calibration, that is how well a given model predicts the data that was used in training it. Secondly, we evaluate their predictive performance in a cross-validation framework.

VI.1.3.5.a Calibration: We use two criteria to evaluate the performance of the models in calibration (in-sample performance). They are itemized below:

- Normalized Root Mean Square Error (NRMSE): The normalization, which here is done by the mean, allows the comparison of intensities of different duration across different stations. For each station s , and duration d , we compute the NRMSE over the non-zero precipitation intensities as:

$$\text{NRMSE}_s(d) = \frac{\left\{ \frac{1}{n_s(d)} \sum_{j=1}^{n_s(d)} (r_{s,T_j}(d) - \hat{r}_{s,T_j}(d))^2 \right\}^{1/2}}{\overline{r_s(d)}}, \quad (\text{VI.20})$$

where $\text{NRMSE}_s(d)$ denotes the score computed at station s , and duration d , $n_s(d)$ is the number of non-zero precipitation intensities for duration d , $r_{s,T_j}(d)$ is the empirical quantile with return period $T_j = \frac{n_s(d)+1}{j \times \delta_d}$, δ_d the average number of non-zero precipitations of duration d per year, $\hat{r}_{s,T_j}(d)$ is the corresponding T_j -year return level estimated from the fitted model. The denominator is the average precipitation at site s and duration d , given as $\frac{1}{n_s(d)} \sum_{j=1}^{n_s(d)} r_{s,T_j}(d)$. The best model according to this criteria is the one with the lowest $\text{NRMSE}_s(d)$.

- Akaike Information Criteria (AIC) (Akaike, 1974): This information criterion rewards the goodness of fit, as measured by the likelihood, but also penalizes the additional number of parameters to be estimated. It is computed as:

$$\text{AIC} = -2 \log(L) + 2p, \quad (\text{VI.21})$$

where L is the maximized likelihood from Equation VI.17 and p is the number of parameters to be estimated (see Table VI.1). The lower the AIC, the better the model.

VI.1.3.5.b Cross-validation: We follow a split sampling procedure in a cross-validation framework. For each station s , we divide the 10 min precipitation intensities into two equal sub-samples of the same length but on different years that are randomly chosen. We then aggregate the data into intensities of various duration, $d = 30$ min, 40 min, 1, 2, 3, 6, 10, 12, 16, 18, 24, 48, and 72 hours. Then we fit each of the 10 IDF models.

We then evaluate the performance of the models fitted on sub-sample 1 on the observations in sub-sample 2 and *vice versa*, by computing the relevant cross-validation criteria. We repeat this procedure 10 times.

In the following, we present three different criteria we use to evaluate the models. These criteria have been used in several studies to evaluate and compare competing models, (see Garavaglia et al., 2011; Renard et al., 2013; Blanchet et al., 2015; Evin et al., 2016; Haruna et al., 2022).

- The Robustness criteria, SPAN, measures the stability of the estimate of a high return level when the training data is changed. It is computed as:

$$\text{SPAN}_{s,T}(d) = \frac{2 \left| \hat{r}_{s,T}^{(1)}(d) - \hat{r}_{s,T}^{(2)}(d) \right|}{\left(\hat{r}_{s,T}^{(1)}(d) + \hat{r}_{s,T}^{(2)}(d) \right)}, \quad (\text{VI.22})$$

where $\hat{r}_{s,T}^{(1)}(d)$ and $\hat{r}_{s,T}^{(2)}(d)$ are the T -year return levels estimated from sub-sample 1 and 2 respectively at station s and duration d . A SPAN of 0.5 means that the absolute difference between the two return levels is half of their average.

For each duration d , a regional score over all the N stations ($N = 81$) is computed as $\text{SPAN}_{\text{reg},T}(d) = 1 - \frac{1}{N} \sum_{s=1}^N \text{SPAN}_{s,T}(d)$ and a perfect model in terms of robustness according to this criteria should have $\text{SPAN}_{\text{reg},T}(d) = 1$.

- The reliability of the model fitted on sub-sample 1 in predicting the maxima in sub-sample 2 and *vice versa* is measured by the FF criteria:

$$\text{FF}_s^{(12)}(d) = \left[\hat{F}_s^{(1)}(d) \left(\max_s^{(2)}(d) \right) \right]^{n_s^{(2)}(d)}, \quad (\text{VI.23})$$

where $\text{FF}_s^{(12)}(d)$ is the cross-validation criteria computed at station s , and duration d , by predicting the probability of the maximum value in sub-sample 2, of sample size $n_s^{(2)}(d)$ using the model $\hat{F}_s^{(1)}(d)$ fitted on the sub-sample 1. $\text{FF}_s^{(21)}(d)$ is computed symmetrically.

For a given duration, [Renard et al. \(2013\)](#) and [Blanchet et al. \(2015\)](#) showed that each $\text{FF}_s^{(12)}(d)$ at a station should be a realization of a uniform distribution. So the difference in the area, *diff* between a theoretical uniform distribution and that of the N set of $\text{FF}_s^{(12)}(d)$ (computed over the N stations), should be close to zero. $\text{FF}_{\text{reg}}(d)$ at the regional scale, given as $1 - \text{diff}$, should therefore take a value of 1 for a reliable model and 0 for a completely unreliable model; the lower the value the less reliable the model is.

- The reliability/accuracy of the model in predicting the entire observations in cross-validation is measured by the NRMSE_CV.

$$\text{NRMSE_CV}_s^{(12)}(d) = \frac{\left\{ \frac{1}{n_s^{(2)}(d)} \sum_{j=1}^{n_s^{(2)}(d)} \left(r_{s,T_j}^{(2)}(d) - \hat{r}_{s,T_j}^{(1)}(d) \right)^2 \right\}^{1/2}}{\overline{r_s^{(2)}(d)}}, \quad (\text{VI.24})$$

where $\text{NRMSE_CV}_s^{(12)}(d)$ is the score computed at station s , and duration d , $n_s^{(2)}(d)$ is the sample size, $r_{s,T_j}^{(2)}(d)$ is the empirical quantile with return period T_j in sub-sample 2, $\hat{r}_{s,T_j}^{(1)}(d)$ is the corresponding T_j return level estimated from $\hat{F}_s^{(1)}(d)$. The denominator is the average daily precipitation in sub-sample 2 at site s given as

$$\frac{1}{n_s^{(2)}(d)} \sum_{j=1}^{n_s^{(2)}(d)} r_{s,T_j}^{(2)}(d).$$

Finally, for each duration d , the regional score computed over the N stations is given as $\text{NRMSE_CV}_{\text{reg}}^{(12)}(d) = 1 - \frac{1}{N} \sum_{s=1}^N \text{NRMSE_CV}_s^{(12)}(d)$.

$\text{NRMSE_CV}_{\text{reg}}^{(21)}(d)$ is computed in similar way. $\text{NRMSE_CV}_{\text{reg}} = 1$ means a perfect model and indicates that there is a complete agreement between all the empirical and theoretical quantiles for all return periods. The closer the value is to 1, the more accurate the model is.

VI.1.4 Results and discussion

We present the results in the following order: first we investigate the appropriateness of the EGPD to fit the data of each duration. Then we present the results of the comparison of the IDF models in calibration, and then in cross-validation. Finally, we show some IDF curves modeled with the best model.

VI.1.4.1 Assessment of EGPD goodness of fit

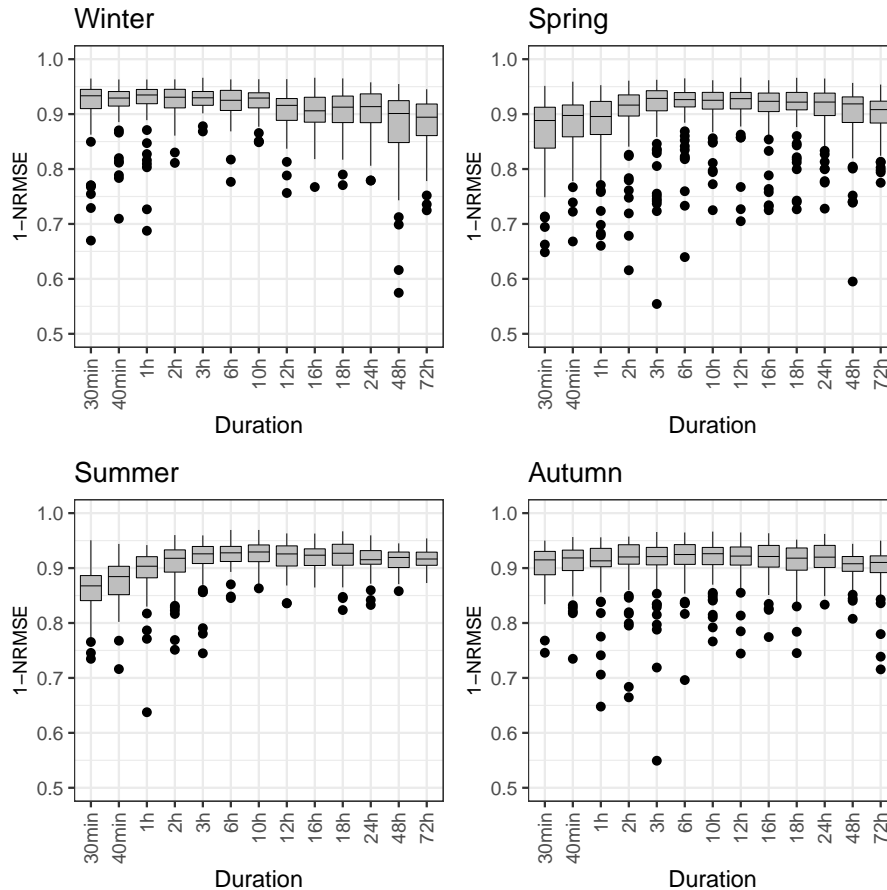


Figure VI.5: Boxplots of $1 - \text{NRMSE}_s(d)$ versus duration for the base EGPD model, i.e., fitted on data of each duration separately. Each boxplot contains 81 points, with each point corresponding to one station.

The first issue is to investigate whether EGPD is an appropriate model for the precipitation data at hand. To check this, we fitted the model at each station and for each duration, independently. We call this EGPD model fitted on each data separately as the "base" model. We then assess the quality of the resulting fits by computing the NRMSE given in Eq. VI.20. The seasonal boxplots of the score for each duration are shown in Fig. VI.5. The higher the score, the better the model.

In spring and summer, the quality of the fit is less good for durations lower than 2 hours. In winter, on the other hand, the fit is less good for $d = 48$ and 72 hours. Overall, more than 74% of the scores fall above 0.9 and 96% above 0.8. We, therefore, consider the EGPD to be a reasonable model for the data.

The fitted shape parameter ξ with respect to duration is shown on Fig. VI.6. Each boxplot contains 81 values, one for each station. We can observe strong dependence of this parameter on

duration, especially in summer. For this season, while 75% of the stations have a $\xi > 0.17$ for $d = 1\text{hr}$, only 25% have $\xi > 0.06$ at $d = 24\text{hr}$. In winter, however, the dependence is not very strong, as judged by the large variability of the boxplots.

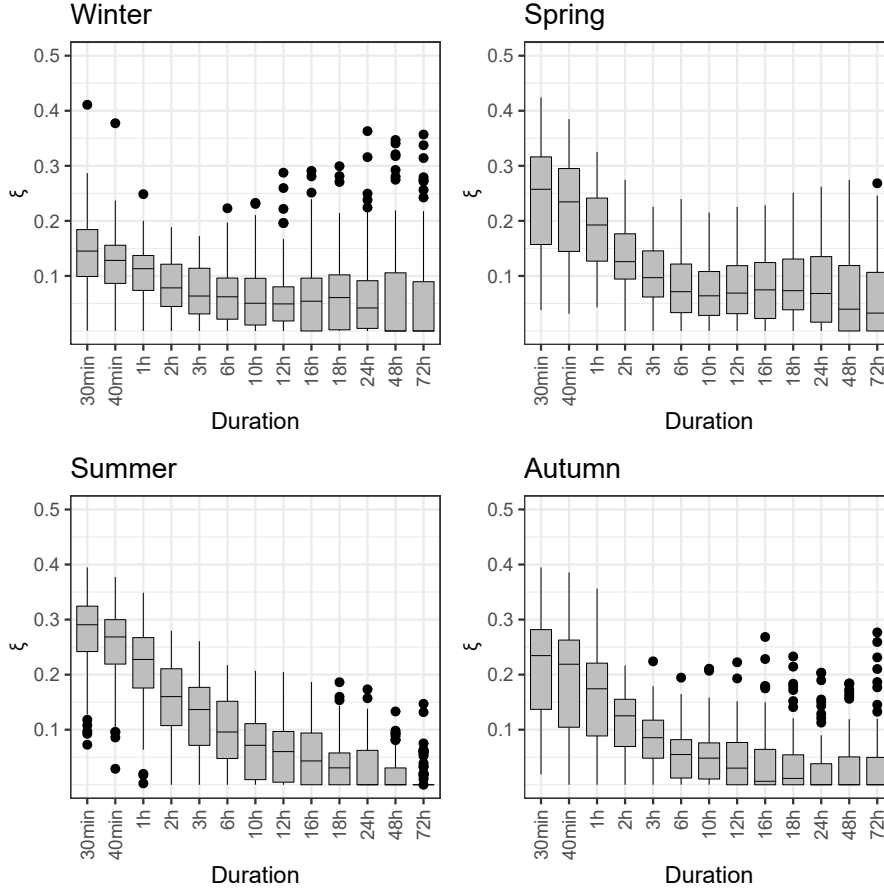


Figure VI.6: Boxplots of the fitted ξ versus duration obtained with the base model. Each boxplot contains 81 points, with each point corresponding to one station.

VI.1.4.2 Comparison of models

Results of the model comparison are presented under two frameworks, first in calibration, and then secondly in cross-validation based on split sampling.

VI.1.4.2.a Evaluation in calibration: Figure VI.7 presents the seasonal boxplots of the 1-NRMSE for the 10 IDF models and the base model. Each of the boxplots contains 1053 points, summarizing the score over 81 stations and 13 durations. In the case of the base model (in yellow), the scores are the same as those in Fig. VI.5, but here we merge the scores for all the durations together.

For all seasons, the two data-driven IDF models, $IDF_{DD_{local}}$ and $IDF_{DD_{global}}$ always show the best performance compared to the others. When looking at the two, the $IDF_{DD_{global}}$ generally outperforms the $IDF_{DD_{local}}$. This means that the global fitting of the model improves the estimation performance compared to the simple interpolation of the locally estimated parameters.

Comparing the IDF_{ss} and the IDF_{kouts} (white vs red boxplots), the results show that for all seasons, the IDF_{kouts} has a better performance compared to the IDF_{ss} . Recall that the two models differ by the additional parameter θ in the former to account for curvature for short durations.

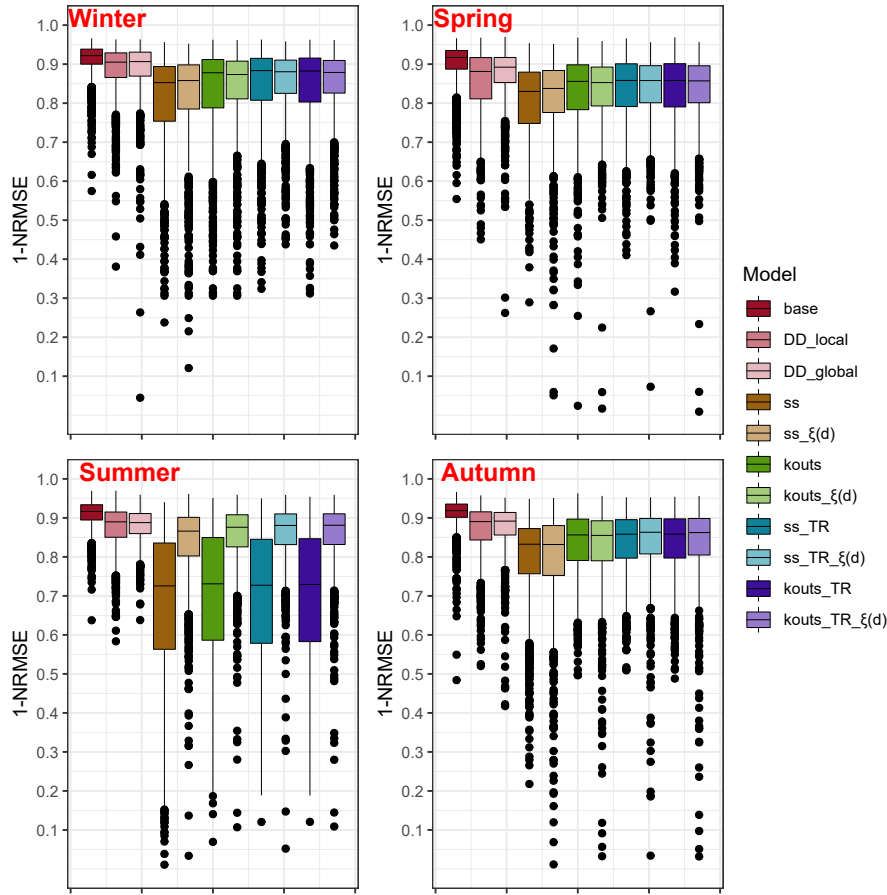


Figure VI.7: Boxplots of the (1-NRMSE) in calibration. Each boxplot contains 1053 points, each point corresponding to one station and duration.

Allowing for $\xi = f(d)$ (models with subscript $_{\xi}(d)$) increased the performance of the models mainly in summer, where all the models without this addition showed very poor performance. For the other seasons, the gain in performance is not as pronounced.

Lastly, the models allowing for scaling break (those with subscript $_{TR}$), show improved performance compared to those with the single regime for all the seasons, except summer (*e.g.* IDF_{ss} vs $IDF_{ss_{TR}}$, *i.e.* the white and violet boxplots).

We note here that we used the NRMSE (Figure VI.7) to measure the accuracy over all non-zero intensities. This is because we are using the EGPD model, which is supposed to model correctly all the non-zero intensities. However, we also computed the NRMSE on extremes only, which we define as the exceedances over the 98% quantile overall intensities (including zeros). The normalization was done by the average of the exceedances. The result (not shown) maintains the same performance ranking order of the models as in Figure VI.7.

The result in Figure VI.7 as measured by the NRMSE assesses the in-sample accuracy of the models but doesn't reward parsimony in terms of the number of model parameters. A natural question to ask is whether the additional performance is worth the additional complexity. To answer this, we compute the AIC for each model at each of the 81 stations. Each time, we rank the models from the best (rank = 1, smallest AIC) to the least (rank = 10, largest AIC). Figure VI.8 shows results in the form of stacked-bar chars for the four seasons. The horizontal axis shows the rank from 1 to 10, and the vertical axis shows the percentage of stations over which a model is ranked. For instance, in winter, $IDF_{DD_{global}}$ is the best (rank = 1) in 82.7% of the stations, while the base model is the best in 17.3% of the stations.

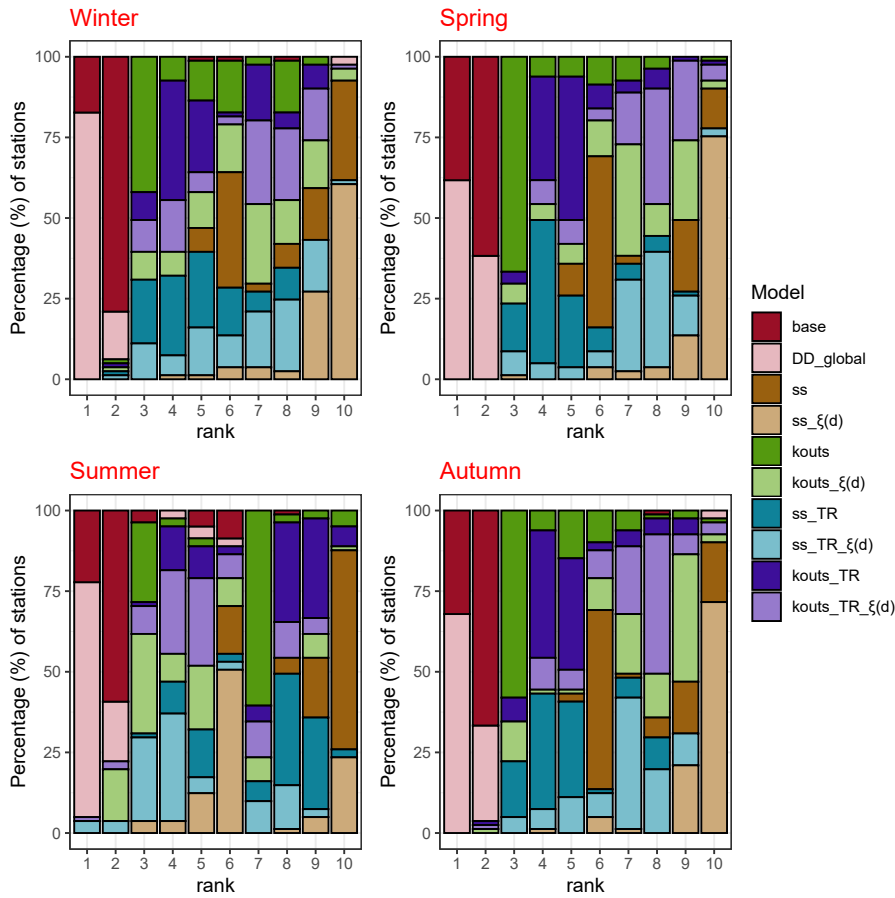


Figure VI.8: AIC results for the four seasons. Each panel shows the percentage of stations (y-axis) over which a model is ranked (x-axis) according to the AIC criteria. The ranks are such that '1' is the best and '10' is the worst.

The result shows that for all the seasons, the 1st and 2nd rank is almost exclusively shared between the IDF_{DD_global} and the base model. To summarize the results, Table VI.2 gives the ranking of the models for the four seasons. The model with the highest percentage is selected for each rank (most likely model at each rank). For all the seasons, the Data-driven (IDF_{DD_global}) is the best, the base model following behind. The results also revealed the relevance of the models allowing for shape parameters to vary with duration only in summer. For the other three seasons, however, the gain in performance is not worth the additional parameter modeling of the shape parameter as a function of duration. Comparing the models with a constant shape parameter, while the IDF_{kouts} is always the most parsimonious, the simple-scaling (IDF_{ss}) remained the worst.

In summary, the IDF_{DD_global} which had the best in-sample performance among the IDF models (as measured by the NRMSE in Figure VI.7), remained the best across all the seasons based on the AIC criteria. Although our focus has been to compare the IDF models that consider the linkage of parameters with duration, we still included the base model in the comparison of the AIC. The result here showed that the base model (with 39 parameters) is less parsimonious compared to the IDF_{DD_global} for most of the stations.

Finally, it wasn't possible to consider the IDF_{DD_local} here, because, unlike the 10 models that are based on maximum likelihood estimation, this model involves a two-step estimation process. The first is a maximum likelihood followed by a second step to estimate the link between the parameters and duration through least squares and segmented regressions. Nevertheless, this is not an issue because this model is based on the same principle as the IDF_{DD_global} and simply differs by

Table VI.2: Summary of the AIC ranking of the models in the four seasons. The model with rank 1 is the best, i.e it has the smallest AIC for most of the stations in that season. A rank of 10 indicates the worst-performing model.

Model	No. of Parameters	Winter	Spring	Summer	Autumn
base	39	2	2	2	2
$IDF_{DD_{global}}$	10	1	1	1	1
IDF_{ss}	4	6	6	10	6
$IDF_{ss_{\xi}(d)}$	5	10	10	6	10
IDF_{kouts}	5	3	3	7	3
$IDF_{kouts_{\xi}(d)}$	6	9	7	3	9
$IDF_{ss_{TR}}$	6	5	4	8	5
$IDF_{ss_{TR}_{\xi}(d)}$	7	8	9	4	7
$IDF_{kouts_{TR}}$	7	4	5	9	4
$IDF_{kouts_{TR}_{\xi}(d)}$	8	7	8	5	8

the estimation method. Thus, comparing the two $IDF_{DD_{global}}$ vs $IDF_{DD_{local}}$ can be seen as comparing two estimation methods within the same model, and according to the NRMSE results in Figure VI.7, better performance is achieved with the $IDF_{DD_{global}}$ compared to the $IDF_{DD_{local}}$.

VI.1.4.2.b Evaluation in cross-validation: The split-sampling procedure allows for the comparison of the models in a cross-validation framework. We use three regional criteria: NRMSE_CV, FF, and SPAN (see Section VI.1.3.5.b), to enable the comparison of the models based on their predictive capabilities. We want to select a model, which in addition to being able to fit the data used to train it, is able to perform reliably and robustly in the presence of new data.

In the following, we present the results in three paragraphs, first according to the reliability/accuracy of the model in predicting all the observations as measured by NRMSE_CV, then the reliability in predicting the maxima as measured by the FF criterion, and lastly, the robustness of the model in predicting the 100-year return level as measured by SPAN100. Figure VI.9 presents the results for the four seasons. For all the criteria, the model with a regional score equal to 1 is the best model.

For all seasons, the NRMSE_CV shows the data-driven models, specifically the $IDF_{DD_{global}}$, to be the most accurate/reliable in predicting the entire observations compared to the other models. In winter, however, the difference in the performance of the models is not very clear. Looking at the summer results, the models without accounting for $\xi = f(d)$ always have the worst performance.

In terms of the FF criterion, the best performance in predicting the maxima in winter is shown by the $IDF_{ss_{TR}}$ model. In fact, all the models with no allowance for $\xi = f(d)$ happen to be the most reliable models in this season. The converse is however true in the case of the remaining seasons. In summer, while $IDF_{DD_{local}}$ is the best model, $IDF_{ss_{\xi}(d)}$ is the best in spring and autumn.

The robustness criteria, SPAN100 shows the models with no allowance for $\xi = f(d)$ to be the most robust models. An exception to this is in summer, where the $IDF_{DD_{global}}$ model is the most robust model. Also, higher robustness is found for the models not accounting for $\xi = f(d)$ compared to their counterparts, for example, IDF_{ss} vs $IDF_{ss_{\xi}(d)}$. This is despite the fact that the former performs poorly in calibration, and is the least performing according to the other cross-validation criteria of reliability. This confirms the previous comments of Garavaglia et al. (2011) that a robust model can completely fail to model/predict the data. Hence the robustness criteria

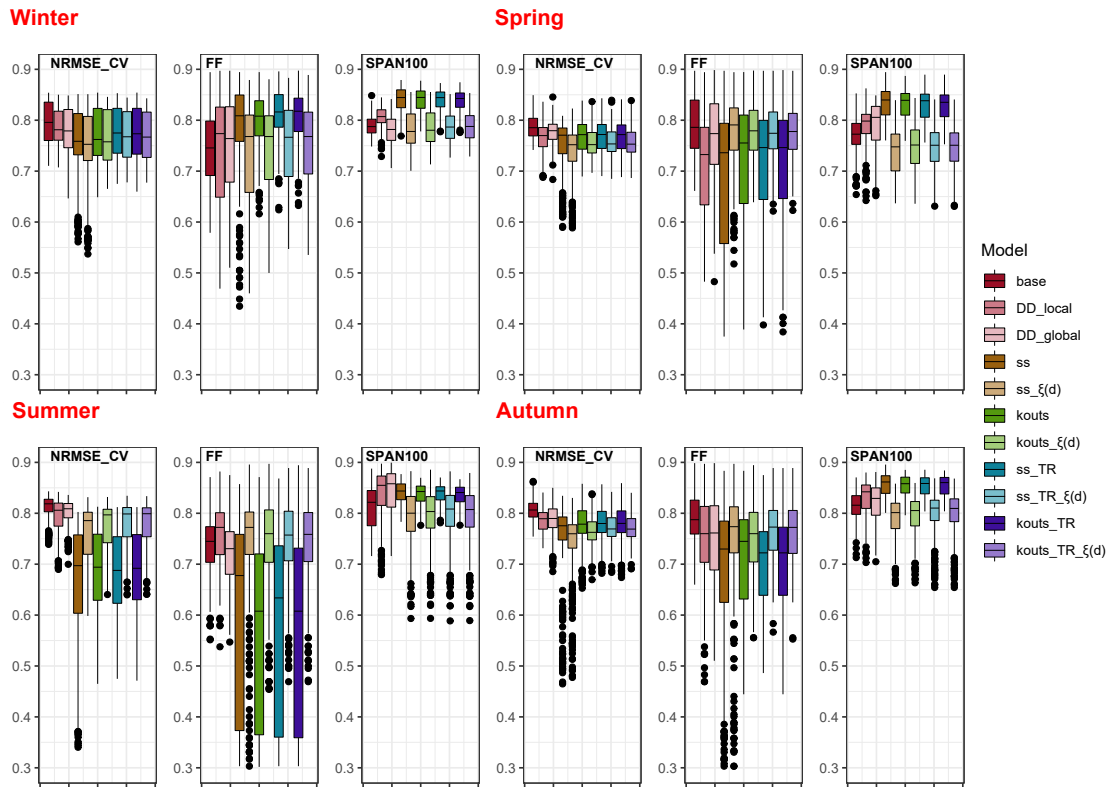


Figure VI.9: Boxplots of the regional cross-validation criteria, NRMSE_CV, FF, and SPAN. For the first two criteria, each boxplot contains 2×130 points, corresponding to one regional score for each of the 13 durations and 10 repetitions of the split sampling. For the SPAN, each boxplot contains 130 points. The optimal value for each criterion is equal to 1.

should only be used alongside other reliability criteria, such that the most robust model is only selected among models of similar reliability.

To summarize the results, the best IDF model should perform well in calibration, and should not be very sensitive to the data used to train it. In calibration, the data-driven model $IDF_{DD_{global}}$ showed the best performance compared to all the other nine models, it also remains accurate and reliable at predicting the entire observations in the split-sampling cross-validation (as measured by the NRMSE_CV), especially in summer. This is an important feature since we are interested in the complete range of intensities. Finally, it generally showed more robustness compared to the other models of similar reliability.

VI.1.4.3 IDF curves

Figure VI.10a shows the IDF curves from two models, IDF_{ss} and $IDF_{DD_{global}}$, along with their 95% CI in summer, at a station in Zurich which is located in the Northeast of Switzerland. In this region, summer is the main season of heavy precipitation. As a reminder, the IDF_{ss} allows scaling only in the scale parameter, σ of the EGPD, the other two parameters (κ and ξ), are independent of duration. The $IDF_{DD_{global}}$ on the other hand allows each of the three parameters to vary with duration. The curves are for return periods $T = 2, 5, 10, 50,$ and 100 years, while points are the empirical levels for $T = 2, 5,$ and 10 years. The IDF_{ss} performed poorly at predicting the empirical quantiles. The curves modeled by the $IDF_{DD_{global}}$ on the other hand are in agreement with the empirical levels. Similar IDF curves for autumn are shown in Fig. VI.10b for a station in Locarno which is located in the Ticino area in the south of Switzerland. The Ticino area is subject

to the heaviest precipitation compared to the other regions in Switzerland. Again, the $IDF_{DD_{global}}$ is able to model the empirical levels correctly for both the short and long durations.

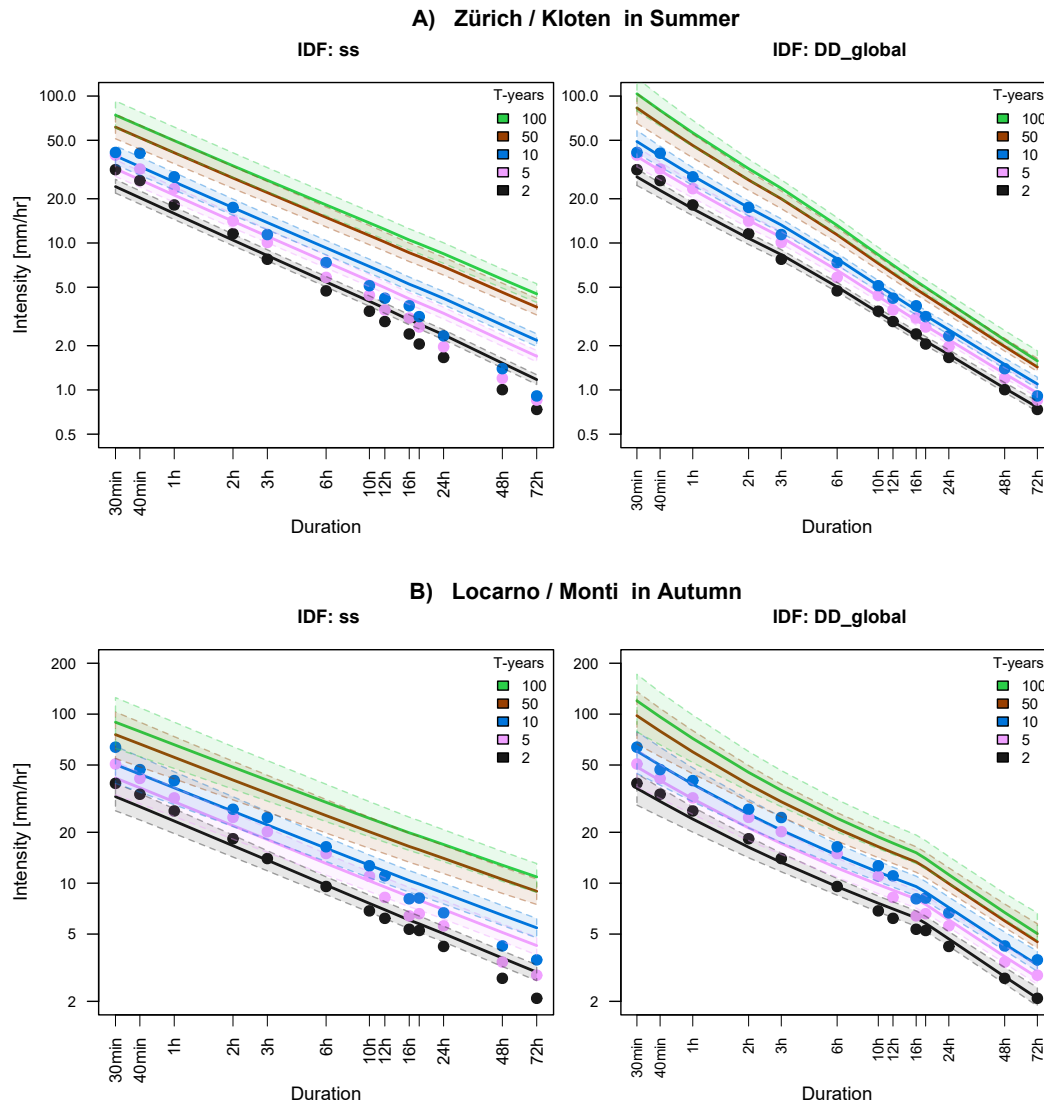


Figure VI.10: Simple-scaling (IDF_{ss}) and data-driven ($IDF_{DD_{global}}$) curves **a**) in summer at a station in Zurich (North-east). **b**) in autumn at a station in Locarno (Ticino area in the south). The curves are for the return periods $T = 2, 5, 10, 40,$ and 100 years. The points are the empirical quantiles corresponding to $T = 2, 5,$ and 10 years. The envelopes represent the 95% confidence bounds obtained by block bootstrap.

The 95% confidence bounds for the return periods $T = 2, 5, 10, 50,$ and 100 years in both stations are shown in Figure VI.11 along with the empirical levels for $T = 2, 5,$ and 10 years (broken lines). For comparison sake, we also include those of the base model, i.e where EGPD is fitted to the data of each duration separately (without linkage of parameters to duration). It can be seen that narrower confidence intervals are obtained with the simple scaling (IDF_{ss}) and data-driven $IDF_{DD_{global}}$ compared to the base model. This is because the base model uses less amount of data in its inference compared to the other two models where all the data of the 13 durations are pooled together in the estimation. It is expected that the width of the CI would be narrower as the available data for estimation increases. The IDF_{ss} has six parameters less than

those of the $IDF_{DD_{global}}$, so a natural question is whether the CI will be wider in the latter model. Indeed for levels where both models predicted the empirical level correctly ($d = 1$ hour in the case of Locarno, second row of Figure VI.11), the CI is a bit narrower in the case of the IDF_{ss} . But looking at the same station for $d = 24$ hours, the bounds are wider, in addition to the bias in predicting the empirical level. Of course, a narrower CI is only relevant if it contains the empirical level.

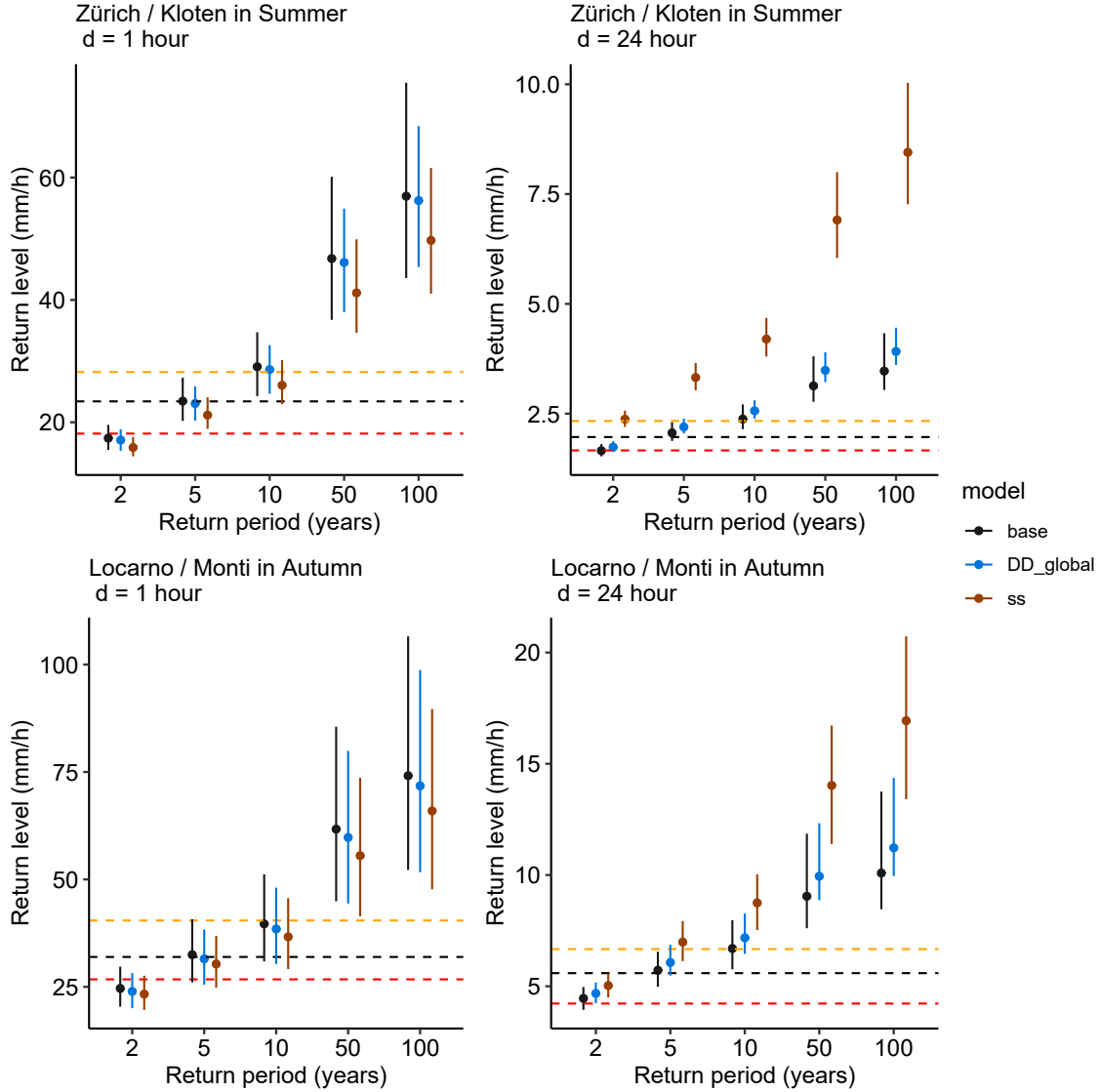


Figure VI.11: Return level estimates (points) and their 95% CI bounds (lines) for return levels $T = 2, 5, 10, 50,$ and 100 years estimated with the simple-scaling (IDF_{ss}), data-driven ($IDF_{DD_{global}}$) and the base model (no linkage of EGPD parameters with duration). The confidence bounds were obtained using block bootstrap. The top row is for a station in Zurich (northeast) in summer for $d = 1$ hour and 24 hours. The second row is for a station in Locarno (Ticino area in the south) in autumn for $d = 1$ hour and 24 hours. In all cases, the broken lines in red, black, and yellow are the empirical levels for $T = 2, 5,$ and 10 years respectively.

In Fig. VI.10, the curves of the simple-scaling model (IDF_{ss}) are not parallel. This behavior resulted from the definition of IDF models for non-zero precipitation in Eq. VI.5. From this equation, we see that the T -year return level is defined as $(1 - \frac{1}{T \times \delta_d})$. The term δ_d , representing

the average number of non-zero precipitation per year varies across the durations leading to a non-constant slope for the different curves.

We finally show, in Fig VI.12 and VI.13 respectively, the seasonal 100-year return level maps for $d = 1\text{hr}$ and 24h . The levels in the maps can be interpreted as the levels which are expected to be exceeded every 100 seasons. The levels were obtained with the best performing model, *i.e.* $\text{IDF}_{DD_{\text{global}}}$. Looking at the return levels for $d = 1\text{hr}$ (Fig. VI.12), we see that the levels in winter are the lowest, with no specific spatial pattern or variability. In spring, the levels in the north and Ticino starts to increase. Summer has the highest levels, and similar levels are obtained all along the plateau in the north. In autumn, while the levels in the north are comparable to those in winter, those in Ticino are comparable to those in summer. A different spatial pattern is however observed for the 100-year return level for $d = 24\text{hr}$. Specifically for summer, the levels in the plateau are lower than those along the northern alpine rim (Prealps). The exhibited spatial pattern of the levels produced by this model is similar to those observed in earlier studies (see Fukutome et al. (2015) for the hourly, and Haruna et al. (2022) for the 24hr precipitation).

VI.1.4.4 Discussion

In the following paragraphs, we briefly discuss some of our choices in terms of the functional forms of the data-driven models, taking into account the varying shape parameter with duration, and the issue of scaling break in the data.

First, for the data-driven models, we limited our choice of functional relationships to simple parametric models, specifically to piece-wise linear models. Other choices would be possible such as smooth regression splines (e.g. Youngman, 2019, 2020). This choice has its advantage and drawback. The advantage is that the splines are able to automatically adjust to fit any form of relationship. The main drawback is that it is inherently non-parametric, and so the mapping of the IDF models, to allow predictions at ungauged locations, is not directly possible. One can only map the three EGPD parameters for a particular duration. For instance, for 13 durations, this means $3 \times 13 = 39$ maps. For our choice of linear functions, 10 parameters are able to describe the IDF curves at each station, and hence 10 maps for the whole area under study.

Regarding the variation of the shape parameter ξ with respect to duration d , some earlier studies did observe or discuss it (Veneziano et al., 2007; Ulrich et al., 2021, e.g.). They however did not model it, either due to the weak form of the relationship, or because the IDF model did not allow for it. Here, especially, in summer, we found very strong dependence, and the results have shown that taking it into account is invaluable. For the other seasons, however, the AIC results showed that it is more parsimonious to let the parameter be independent of duration. We note that the strength of the variation of the shape parameter with duration is different in the four seasons. In our case, it is strongest in summer and weakest in winter. A possible explanation is that the underlying precipitation formation mechanism responsible for the formation of short and long-duration precipitation is different in the four seasons. For instance, in winter, the same frontal system is responsible for both short and long-duration events, and hence the variation of the shape parameter with duration is weak. In summer where the linkage is strong, different systems are responsible for the short-duration intensities (e.g. due to short convective events) and long-lasting duration intensities (e.g. due to frontal systems). The other seasons (spring and autumn) present a mixture of the two behaviors.

Finally, we observed a break in the scaling relationship of the EGPD scale parameter. The existence of scaling regimes has been investigated by Fraedrich and Larnder (1993) and they linked it to atmospheric processes such as the structure of frontal systems, the diurnal cycle, the seasonal periodicity, and climatic variability. Hence, break points in the regime observed at stations could be explained by the possible transition from one precipitation system to another. For example, from short convective events at short duration to frontal systems at longer duration. For breaks

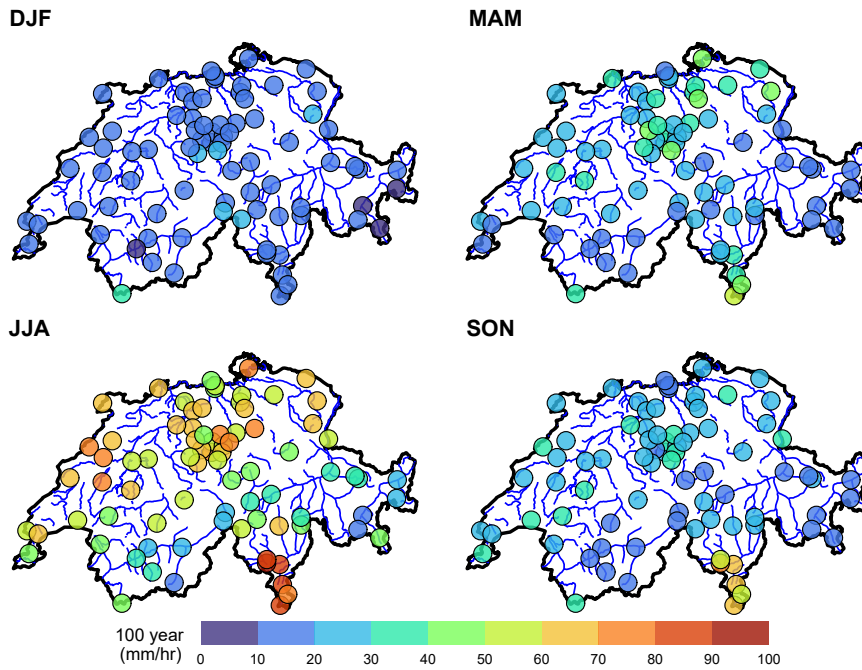


Figure VI.12: Map of Switzerland showing the seasonal 100-year return level in mm/hr for $d = 1$ hr. Levels predicted with $IDF_{DD_{global}^*}$.

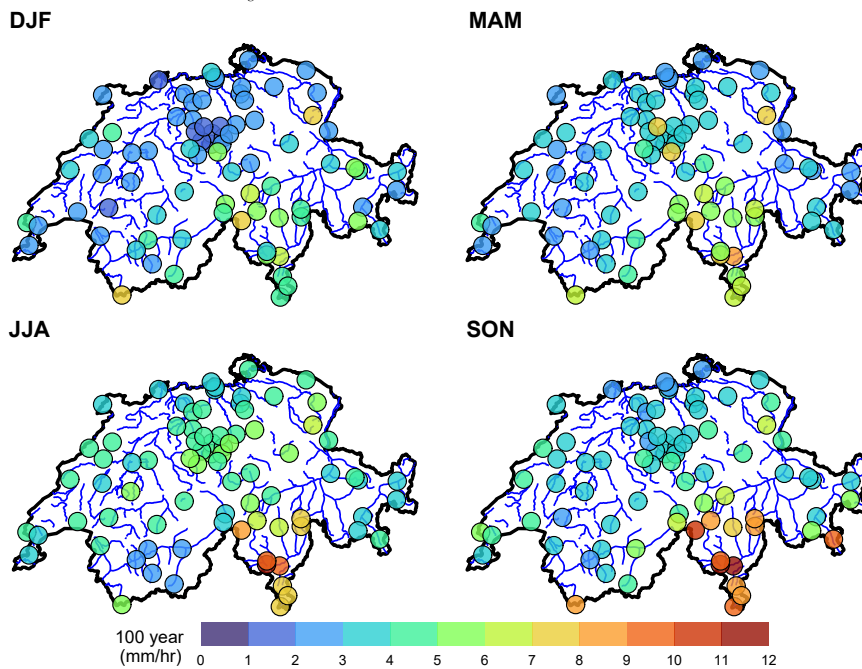


Figure VI.13: Map of Switzerland showing the seasonal 100-year return level in mm/hr for $d = 24$ hr. Levels predicted with $IDF_{DD_{global}^*}$.

at very short durations in the range of around 1 hour, however, it is generally attributed to being artificial, imposed by the resolution of the gauge (measurement precision errors). Despite this, the identification of scaling breaks involves some level of uncertainty depending on the method used. For instance, [Paschalis \(2013\)](#) tried to identify scaling breaks in Switzerland using the same 10 mins datasets as in our case. He used spectral density analysis based on the Fourier transform, and another based on wavelet decomposition. While both methods agree on the existence of scaling breaks, the exact times of the breaks were different in both. The author also mentioned that no clear seasonal or regional pattern was observed. In our case, we rely on the estimated EGPD scale parameter to identify the scaling breaks. Other authors relied on moment analysis (plot of moment vs duration in log-log scale for different orders) to identify breaks in scaling ([Bougaard and Adamowski, 2006](#); [Nhat et al., 2008](#)). In reference to this, we do not rule out the possible effect of the EGPD model on the observed scaling breaks. A detailed study of its own might be required to study and characterize/interpret the scaling breaks in the study area since any detailed interpretation might be premature.

VI.1.5 Conclusions

Our aim in this paper was to build IDF curves using all the non-zero precipitation data in Switzerland. To achieve this, we used the EGPD model as the parametric model for the precipitation intensities. The literature presents various approaches to link the different durations together in IDF curves. We considered three of these approaches to build the IDF curves while using the EGPD as the parametric model. The first approach is based on the scale invariance theory, where IDF curves are built based on the scaling behavior of precipitation. The second approach is based on the general IDF formulation of [Koutsoyiannis et al. \(1998\)](#), which generalizes the various traditional IDF formulations. The last approach is called the data-driven approach, where each parameter can vary with duration, and the form of the relationship is empirically determined by the data at hand.

We started from these three approaches and added some extensions to account for scaling break and varying shape parameter with duration. We ended up with a total of ten IDF models. We then compared them, first in calibration, and then in a split-sample cross-validation approach.

The results showed that, given the EGPD as the parametric model, the data-driven IDF-EGPD, particularly the $IDF_{DD_{global}}$, is the best model for the data at hand. The AIC also showed this model to be more parsimonious compared to the base model where no linkage of parameter with duration is considered. The IDF curves based on simple-scaling and the general formulation of [Koutsoyiannis et al. \(1998\)](#), did not perform as efficiently even with the added extensions in terms of scaling break and in the way the shape parameter varies with duration. The fact that the simple-scaling IDF models performed poorly in summer confirms the previous findings of [Molnar and Burlando \(2008\)](#) and [Paschalis \(2013\)](#) that in Switzerland, precipitation in summer shows multiscaling behavior.

Although our work focused on Switzerland, the data-driven approach can be applied everywhere, especially for regions where high-resolution data are available. This is because the approach considers the empirical relationship between the model parameters and duration, and is not constrained by the hypothesis of scaling across durations. Our result also showed that it is possible to model the linkage of the shape parameter with duration in IDF curve modeling. Moreover, the study highlights the need to explore in detail the empirical relationship of model parameters with duration prior to the application of any of the widely used IDF construction approaches. Finally, we showed that it is possible to use the EGPD for IDF curve modeling.

In terms of perspectives, it would be interesting to produce maps of the parameters to allow for predictions at ungauged sites. This could be achieved by simple interpolation of the local IDF parameters as done by [Blanchet et al. \(2016\)](#), or through quantile regression methods ([Ouali and Cannon, 2018](#)), or by global estimation using spatial covariates (e.g. [Ulrich et al., 2020](#)). Another

possibility is to use a regionalization technique, such as the method of [Hosking and Wallis \(2005\)](#) and then interpolate the index flood to allow predictions at the ungauged sites ([Mascaro, 2020](#), e.g.).

It should be mentioned that throughout this work, we estimated the IDF models through the independence likelihood, thus omitting the correlation between different timesteps and durations. In the case of annual maximum series (GEV), [Nadarajah et al. \(1998\)](#) has modeled the dependence between the different time steps using multivariate extreme value distributions (MEVD), and [Tyrallis and Langousis \(2019\)](#) followed suit by using max-stable processes. Later, [Jurado et al. \(2020\)](#), investigated the impact of accounting for this dependence in extremes and showed that there is little gain in performance, in addition to the added complexity of using max-stable processes. Since all these authors focused on the distributions that are based on extreme data only, an interesting perspective will be to investigate the effect of accounting for the dependence in IDF curve modeling for the case of the EGPD that uses all the non-zero data

Lastly, consideration of the effect of climate change in building IDF curves is invaluable. For instance, [Cheng and AghaKouchak \(2014\)](#) showed that by neglecting nonstationarity in modeling IDF curves, there could be up to 60% underestimation of extreme precipitation, especially for short durations. It would therefore be interesting to model the curves while accounting for a warmer climate (e.g. [Mirhosseini et al., 2013](#); [Cheng and AghaKouchak, 2014](#); [Ragno et al., 2018](#); [Ouarda et al., 2019](#); [Kristvik et al., 2019](#)).

VI.2 Catchment IDF curves for hydrological applications in Switzerland

In the previous section (**Paper 2**), we identified the data-driven IDF model, denoted as $IDF_{DD_{global}}$, as the best model in the study area. This model allows linking the parameters with duration based on empirically determined parametric relationships. It allows departure from simple scaling relationships and the presence of scaling breaks. In this section, we will apply this model to construct IDF curves at the catchment level for operational purposes in Switzerland. The reader is referred back to Section VI.1.3.2.c for the model formulation and Section VI.1.3.3 for the inference strategy. In the remainder of the section, for the sake of simplicity, we will drop the subscript in $IDF_{DD_{global}}$ and refer to the model simply as IDF model.

A total of 24 large catchments, shown in Figure VI.14, are considered. They are catchments of river Aare, Rhine, Limamat, Saane, and Rhône, with areas ranging from 900 km² to 35,000 km². These catchments were provided by the current EXCH project, specifically, the Swiss Federal Office of Energy (SFOE), and are the same ones used in the previous phase of EXAR project (Andres et al., 2021).

VI.2.1 Data

We consider mean areal precipitation at hourly resolution for each of the 24 catchments, covering a period of 90 years from 1930 to 2019. Similar to the catchments, the data was provided by the project and a detailed description of how the data was obtained can be found in EXAR report (Andres et al., 2021). To summarize, daily mean annual precipitation for each catchment was obtained through spatial interpolation of daily precipitation observed at rain gauge stations using Thiessen polygons. The daily mean areal precipitation was then disaggregated using the method of fragments (Wójcik and Buishand, 2003) to obtain the pseudo-observations of hourly mean areal precipitation. The daily rain gauge stations considered for this purpose were the same as those used in Evin et al. (2018).

VI.2.2 IDF curves

Figure VI.15 shows the seasonal IDF curves for 4 out of the 24 catchments. Each column corresponds to a season, while each row corresponds to a catchment. Their areas are 17,601, 34,970, 3,219 and 10,303 km² (in that order from top to bottom). The IDF curves for the remaining catchments can be found in Appendix A.2.3. The curves represent the modeled return levels for $T \in \{2, 5, 10, 30, 100\}$ years, while the points are the corresponding empirical return levels. It can be seen that there is generally a good agreement between the empirical and modeled return levels for all the seasons. The flexibility of the adopted IDF model allows the curves to have different forms depending on the season, climate, and catchment area.

Figure VI.16 shows the maps of summer 100-year return level in mm/day for 24 hr duration. The levels are obtained from the IDF curves shown in Figure VI.15. The 100-year return level ranges from 75 mm (e.g. Rapperswil-Auenstein) to 150 mm in the case of small catchments (e.g. Reichenau). The Aare catchment at Beznau (17,600 km²) has nearly the same 100-year level as the Rhine catchment at Birsfelden (34,970 km²). The catchments with the largest levels are the Lavey of the Rhône and the Reichenau of the Rhine River.

The IDF curves we built provide critical information for designing infrastructure in the analyzed catchments, such as stormwater drainage systems, reservoirs, and flood protection measures. Additionally, the presented maps offer valuable insights into extreme precipitation and can guide emergency response planning, infrastructure design for extreme events, and the assessment of potential flood damages. Decision-makers can use these maps to identify regions at higher risk of

VI.2. Catchment IDF curves for hydrological applications in Switzerland

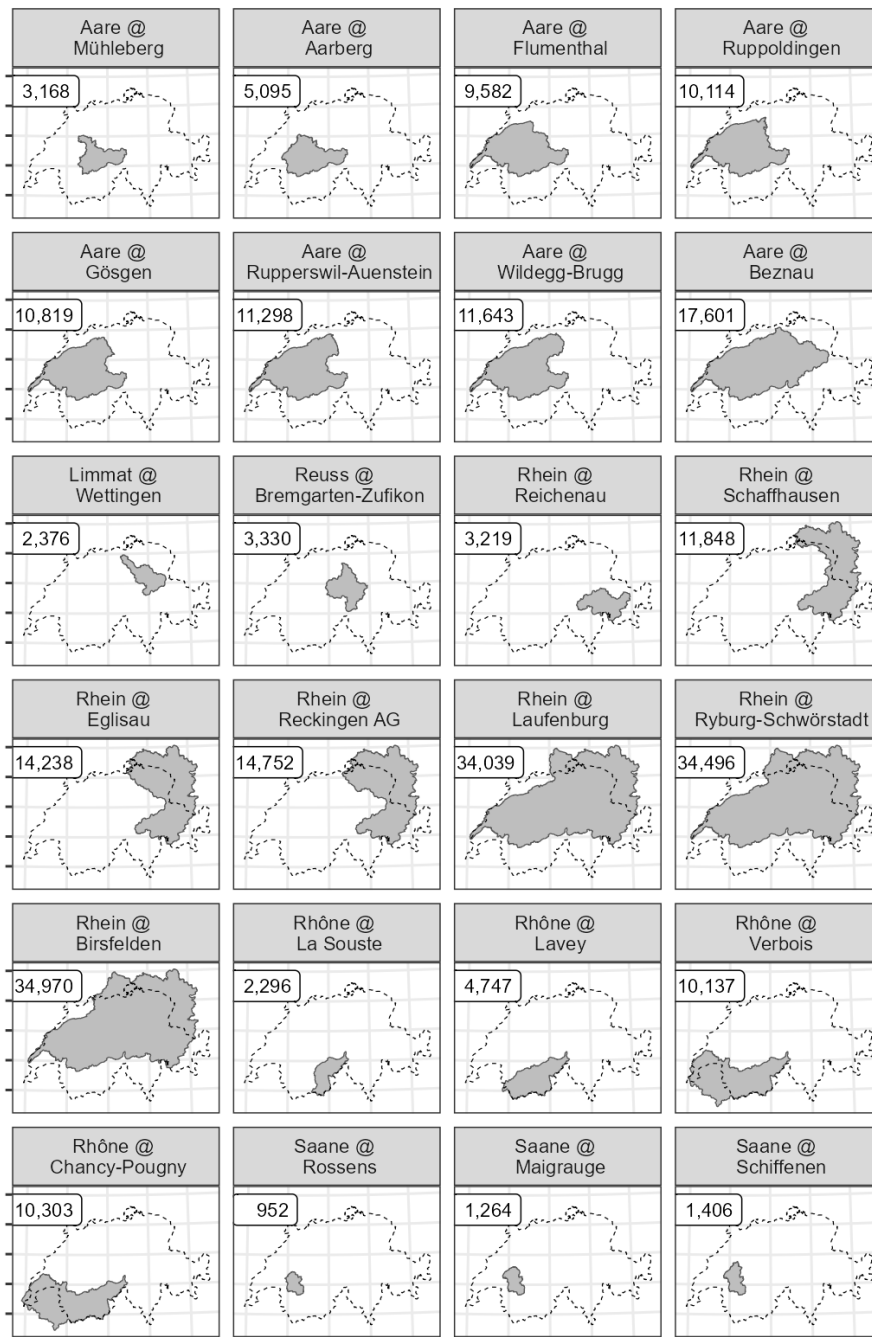


Figure VI.14: Map of the 24 hydrological catchments. The dashed-black line shows the border of Switzerland. The catchments are arranged according to the river they drain and the number in the black box represents the area of each catchment in km².

extreme precipitation and prioritize mitigation efforts accordingly.

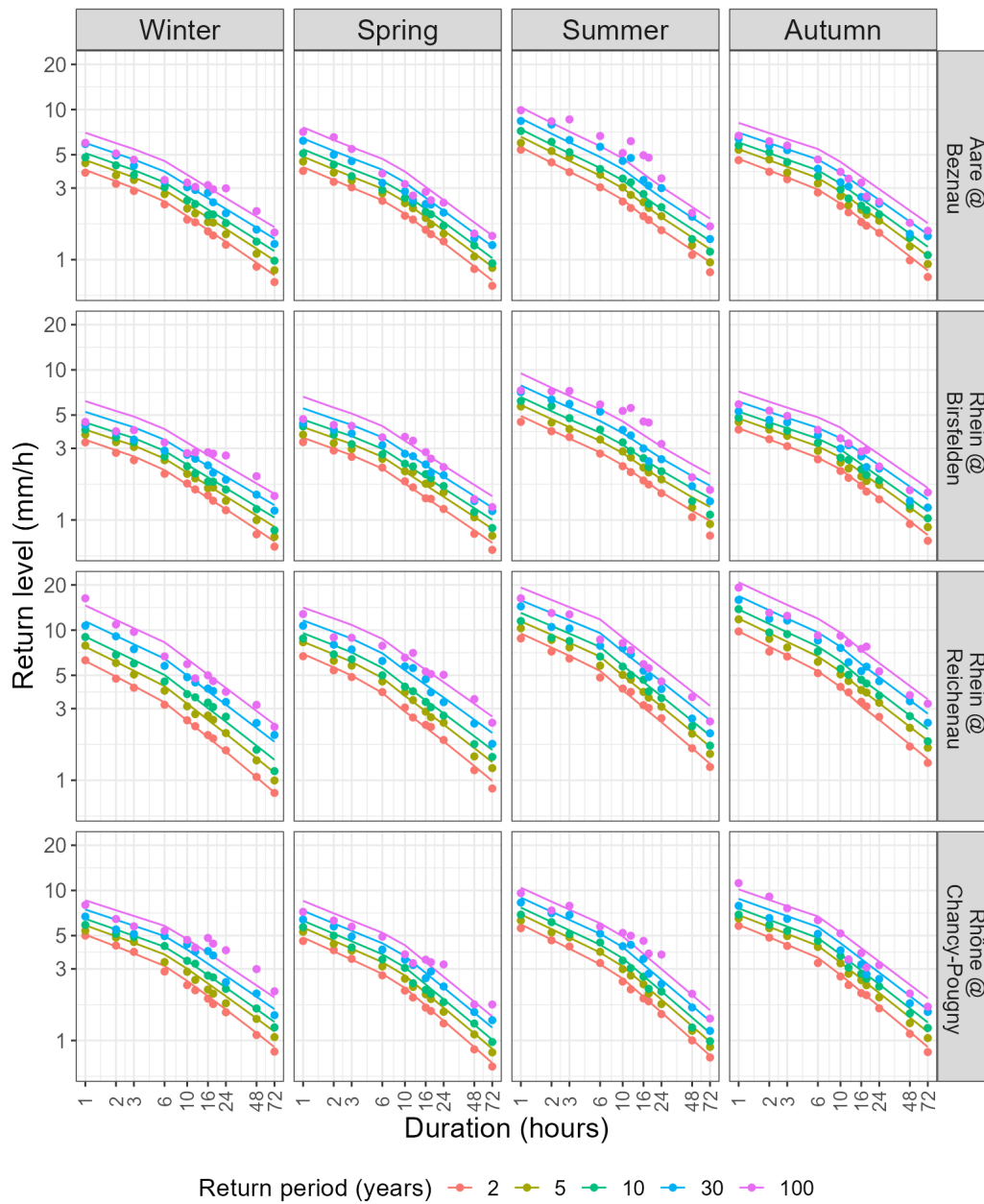


Figure VI.15: Seasonal IDF curves for 4 out of the 24 large catchments. The lines are the modeled return levels, while the points are the corresponding empirical return levels. Both are colored according to the return period for $T \in \{2, 5, 10, 30, 100\}$ years.

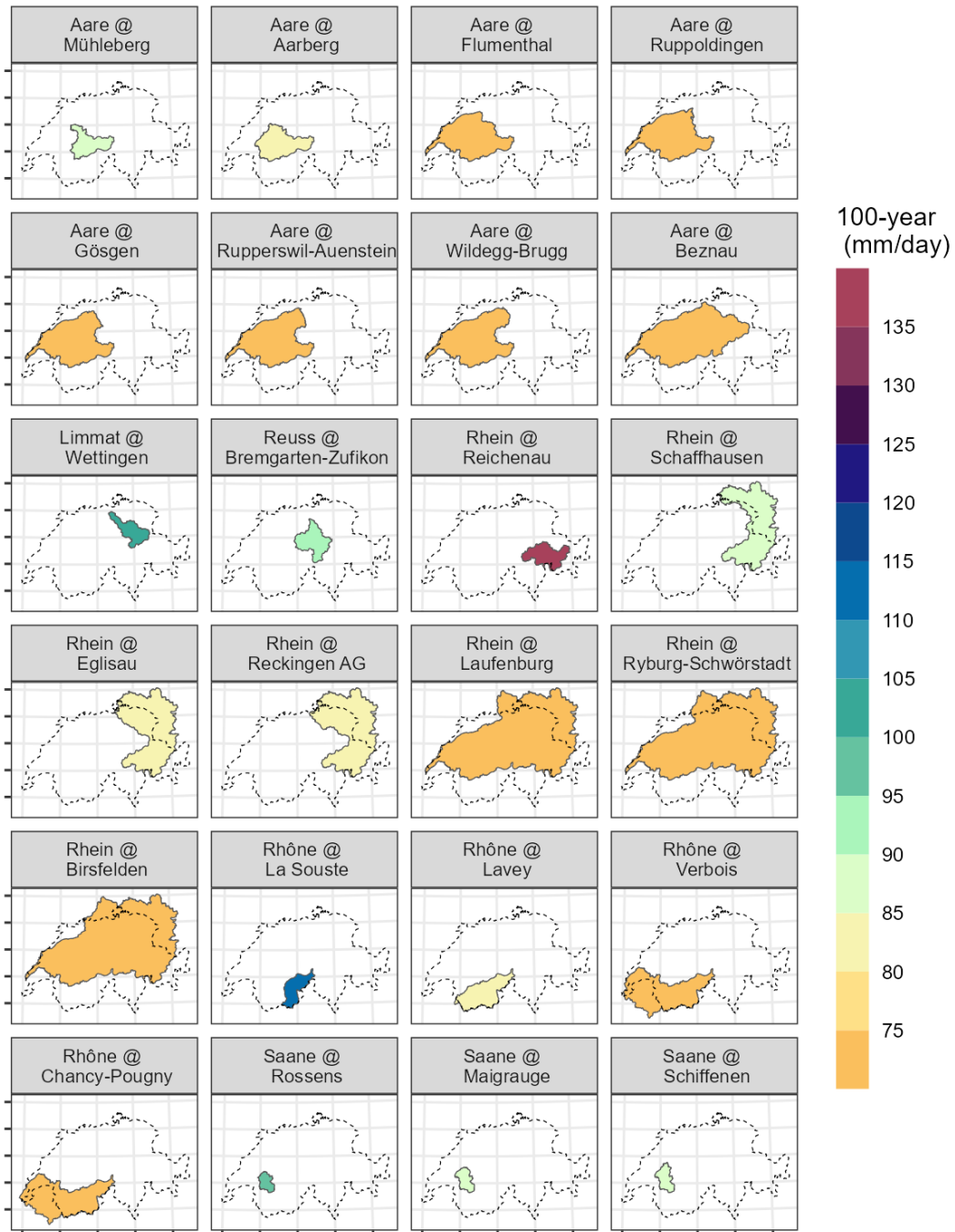


Figure VI.16: 100-year return level of summer mean areal precipitation for 24 hr duration.

VI.3 Summary

To sum up, the main take-home message from this chapter is summarized below:

Question 3: What is the best model of Intensity-Duration-Frequency (IDF) relationships for the full range of non-zero precipitation intensities in a topographically complex area?

- Models of IDF relationships were developed using all non-zero precipitation intensities. The three-parameter EGPD was used as the model for the precipitation intensities.
- Three approaches to modeling IDF curves, along with some extensions to account for scaling break and varying shape parameters with duration, were considered.
- The first approach relies on the scale invariance of precipitation, the second is based on the general IDF formulation of Koutsoyiannis et al. [Koutsoyiannis et al. \(1998\)](#), and the third is called the data-driven approach, where each parameter can vary with duration based on empirically determined relationships.
- The IDF models were compared in calibration and cross-validation, and the model based on the data-driven approach was shown to be the best performing. It was used to build catchment-level IDF curves for hydrological applications in Switzerland.
- As a limitation, the independence likelihood was used to estimate the model parameters, thereby neglecting the serial correlation and the correlation between data from different durations. It will be interesting to investigate the effect of this assumption.

VIII

Modeling of Intensity-Duration-Area-Frequency (IDAF) Relationships in Switzerland

Chapter overview

In the last chapter, we developed a model of Intensity-Duration-Frequency (IDF) relationships in Switzerland using all the non-zero precipitation intensities. This chapter addresses Question 4 by extending the IDF relationships to account for the spatial extent of precipitation, *i.e.* area (A), by modeling the Intensity-Duration-Area-Frequency (IDAF) relationships including all the non-zero precipitation intensities.

Contents

VII.1	Evaluation of the gridded datasets	114
VII.1.1	Comparison of the empirical values at daily time scales and above	114
VII.1.2	Comparison of return level estimates	118
VII.1.3	Conclusion and discussion	119
VII.2	Paper 3: Modeling Areal Precipitation Hazard: A Data-driven Approach to Model Intensity-Duration-Area-Frequency Relationships using the Full Range of Non-Zero Precipitation in Switzerland	120
VII.2.1	Introduction	121
VII.2.2	Study area and Data	123
VII.2.3	Methodology	125
VII.2.4	Results and Discussion	132
VII.2.5	Conclusions	140
VII.3	Summary	142

VII.1 Evaluation of the gridded datasets

Building models of Intensity-Duration-Area-Frequency (IDAF) relationships requires information on areal precipitation, and to obtain this, precipitation data is required everywhere in the study area. This information is usually obtained from either station interpolation (e.g. [Panthou et al., 2014](#)) or more commonly from radar (e.g. [Overeem et al., 2010](#)), or radar reanalysis data (e.g. [Overeem et al., 2010](#); [Mélèse et al., 2019](#); [Blanchet and Mélèse, 2020](#)). As outlined in Section II.3, there are two gridded datasets in Switzerland, Combiprecip (CPC) and RhiresD. CPC is a radar-reanalysis product at hourly resolution and RhiresD results from the spatial interpolation of rain gauge data at the daily resolution. Both datasets are at a spatial resolution of 1 km by 1 km.

CPC is the only gridded dataset available at the sub-daily timescale in our study area, and so it is the product we use to build the IDAF models. However, before using the data, it is essential to evaluate the time series from CPC in comparison with those from the rain gauge network as well as the other gridded data set, RhiresD. The evaluation is done through a point-to-pixel analysis that involves comparing the time series from a gauge to the time series from the pixel underlying the gauge.

We consider hourly data from 71 gauge locations and in each case, the data from the gauge is considered the "truth". These 71 stations have no missing data from 2005 to 2020 (the period of overlap of the three datasets). The comparison is performed in two steps: in the first step, we compare the two time series using some chosen criteria and in the second step, we fit EPGD to both time series and compare the 20-year return level estimate. The comparison of the datasets at the daily timescales and above is shown in this section, while the comparison results for sub-daily time scales (applicable for the CPC data only) is presented later in [Paper 3](#) (see Section VII.2.4.1).

VII.1.1 Comparison of the empirical values at daily time scales and above

Recall from Section II.3.2.2 that RhiresD contains daily precipitation totals. The value recorded on day D corresponds to accumulations from 06:00 UTC of day D to 06:00 UTC of day D+1. For a fair comparison, we aggregate both CPC and the raingauge data using the same time window.

VII.1.1.1 Criteria on all observations

Following the work of [Zambrano-Bigiarini et al. \(2017\)](#), we use the three sub-components of the Kling-Gupta-Efficiency (KGE) criterion ([Kling et al., 2012](#)) to compare a given gridded dataset (CPC or RhiresD) to the reference dataset (rain gauges). The criterion is computed from:

$$\text{KGE} = 1 - \sqrt{(r - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2}, \text{ with}$$

$$\begin{cases} r = \frac{\text{Cov}(i_{grid}, i_{gauge})}{\sigma_{grid}\sigma_{gauge}} \\ \beta = \frac{\mu_{grid}}{\mu_{gauge}} \\ \gamma = \frac{\text{CV}_{grid}}{\text{CV}_{gauge}} \end{cases},$$

where r is the Pearson correlation coefficient that measures the linear correlation between the gridded data (i_{grid}) and the gauge data (i_{gauge}). Cov is the co-variance between the two time series and σ denotes the standard deviation. β evaluates the bias between the two time series (the tendency of gridded data to under- or overestimate the gauge data), with μ the mean. γ is the variability ratio, that is the ratio between the coefficient of variations (CV) of the two time series. It measures the under or over-dispersion of gridded data compared to the gauge. For a perfect match between the two time series, r , β , γ , and KGE should be equal to 1.

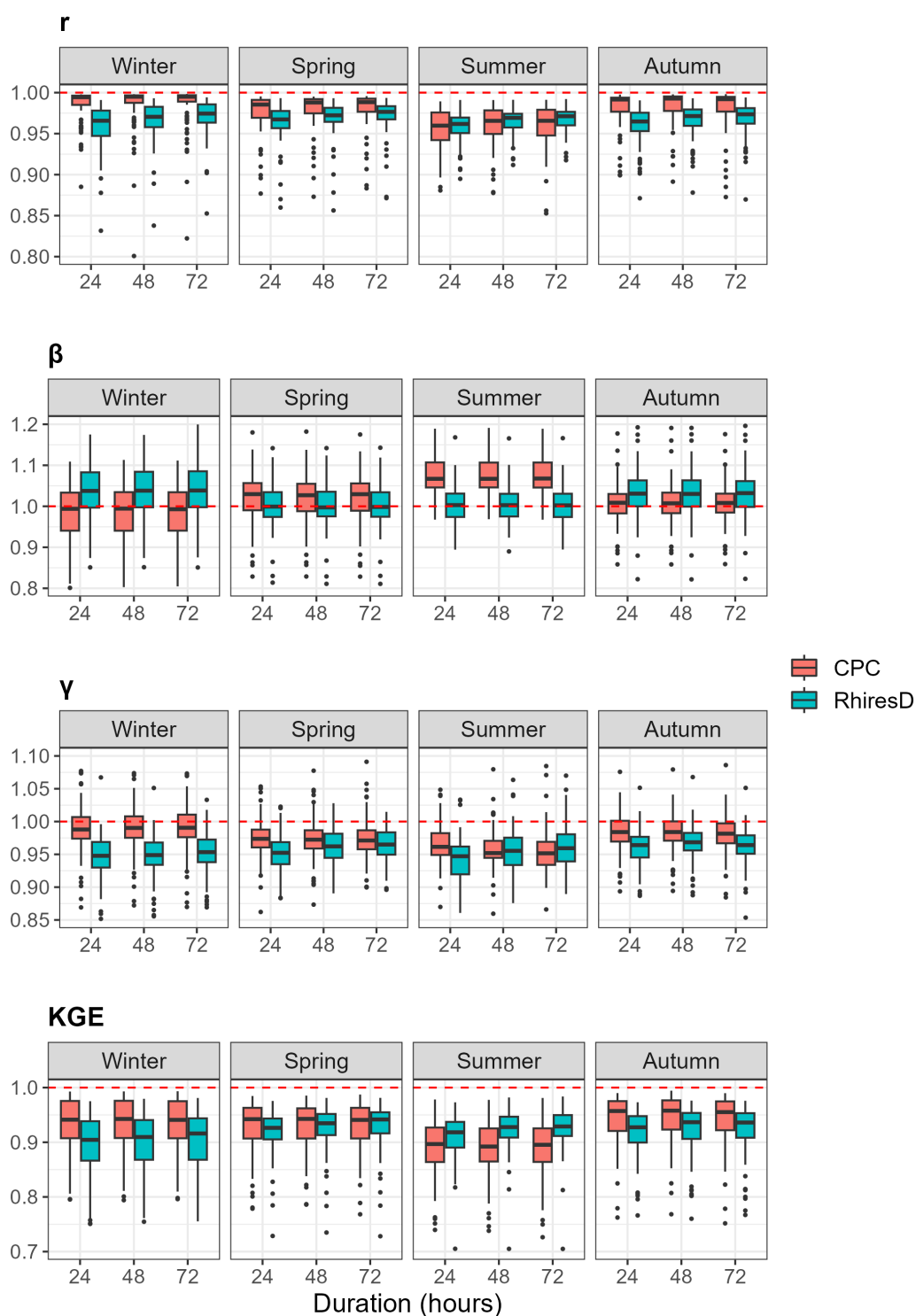


Figure VII.1: Seasonal boxplots of the KGE and its sub-components. Each boxplot contains 71 points, 1 point for each pair of rain gauge and the underlying pixel.

The seasonal boxplot of the KGE and its sub-components are shown in Figure VII.1. Starting from the top row, the correlation coefficient r shows that there is a good temporal correlation between the two gridded data sets and the rain gauge data for all seasons and durations (median > 0.95). For all seasons, the correlation increases with the aggregation duration. CPC exhibits its

lowest correlation in summer, while RhiresD shows its lowest correlation in winter. Overall, however, except for summer, CPC has a better correlation with the gauge data compared to RhiresD. In the second row, β shows that CPC has a higher tendency toward overestimation of the data in spring and summer compared to RhiresD, while RhiresD has a higher overestimation tendency in autumn and winter. There is no difference in the durations since the bias is based on the seasonal average. The dispersion bias γ in the third row shows that both datasets generally have a negative dispersion bias. Nonetheless, RhiresD has a higher negative dispersion bias compared to CPC for all seasons and durations. Finally, KGE which summarizes the three components shows that CPC has a better performance compared to RhiresD in spring, and autumn, while the latter is better in summer. In spring, CPC has a higher median, although the largest variability.

VII.1.1.2 Criteria on extremes

Next, we evaluate the ability of CPC and RhiresD to correctly detect extreme precipitation as measured by the gauge. Extremes here are defined as the exceedances of 99.5% quantile over the whole sample (zeros included). We compute three scores based on Table VII.1 similar to Panziera et al. (2018).

	gauge \geq 99.5th perc.	gauge $<$ 99.5th perc.
gridded data \geq 99.5th perc.	a	b
gridded data $<$ 99.5th perc.	c	d

Table VII.1: Contingency table to compare gridded data and rain gauge data

The scores are given below:

1. Bias of hits: This score measures the bias in extreme precipitation totals recorded on the days that are detected as extreme in both datasets (gridded and gauge). If extGrid_i indicates extremes detected by the gridded data and extGauge_i indicates extremes detected by the gauge for $i = 1, 2, \dots, a$, then the bias of hits is computed as:

$$\text{Bias} = \frac{\sum_{j=1}^a \text{extGrid}_j}{\sum_{j=1}^a \text{extGauge}_j}.$$

For a perfect agreement, Bias should be equal to 1.

2. Probability of detection (POD). This score computes the ability of the gridded data to classify events as extremes, given that they are also extremes according to the gauge. POD is computed from:

$$\text{POD} = \frac{a}{a + c}.$$

For a perfect agreement, POD should be equal to 1.

3. False alarm ratio (FAR): This score measures the rate at which the gridded data classify events as extremes when they are not extremes according to the gauge. FAR is computed from:

$$\text{FAR} = \frac{b}{a + b}.$$

For a perfect agreement, FAR should be equal to 0.

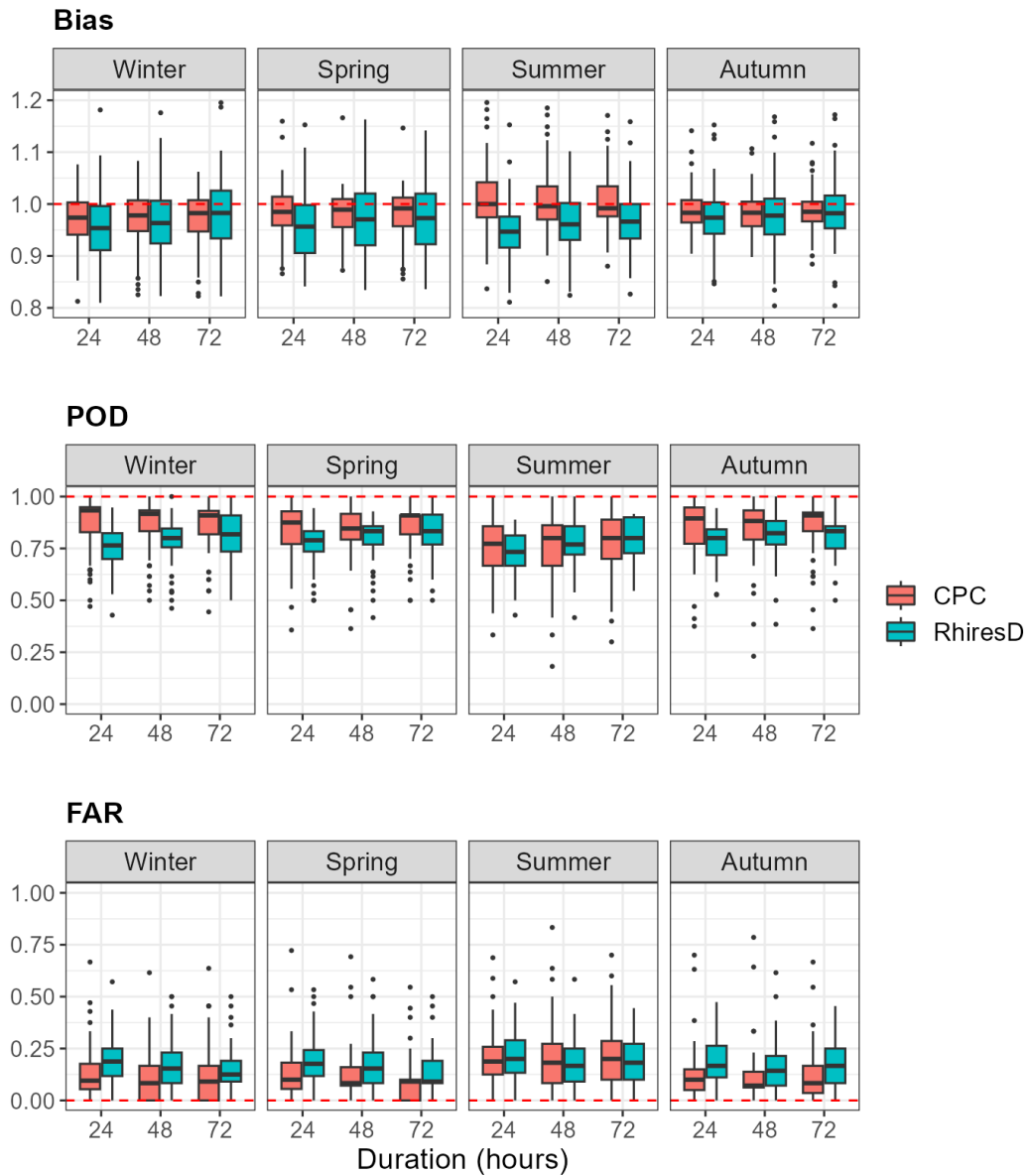


Figure VII.2: Seasonal boxplots of the KGE and its subcomponents. Each boxplot contains 71 points, 1 point for each pair of rain gauge and the underlying pixel.

Figure VII.2 shows the seasonal boxplots of the three scores. Starting from the top row, the bias of hits shows that CPC has a lower bias compared to RhiresD for all seasons and durations. The median of the bias of CPC is nearly equal to one, which indicates that the data is almost unbiased in terms of the extreme precipitation totals. Moving to the second row, the boxplot of the POD shows that CPC has a better score compared to RhiresD for all seasons except summer. The median of the score for CPC ranges from 0.7 to 0.99, which means that 70% to 99% of the gauge extreme events are correctly classified as extremes by the CPC. The boxplots of FAR in the third row similarly show a better performance of CPC compared to RhiresD for all seasons except summer.

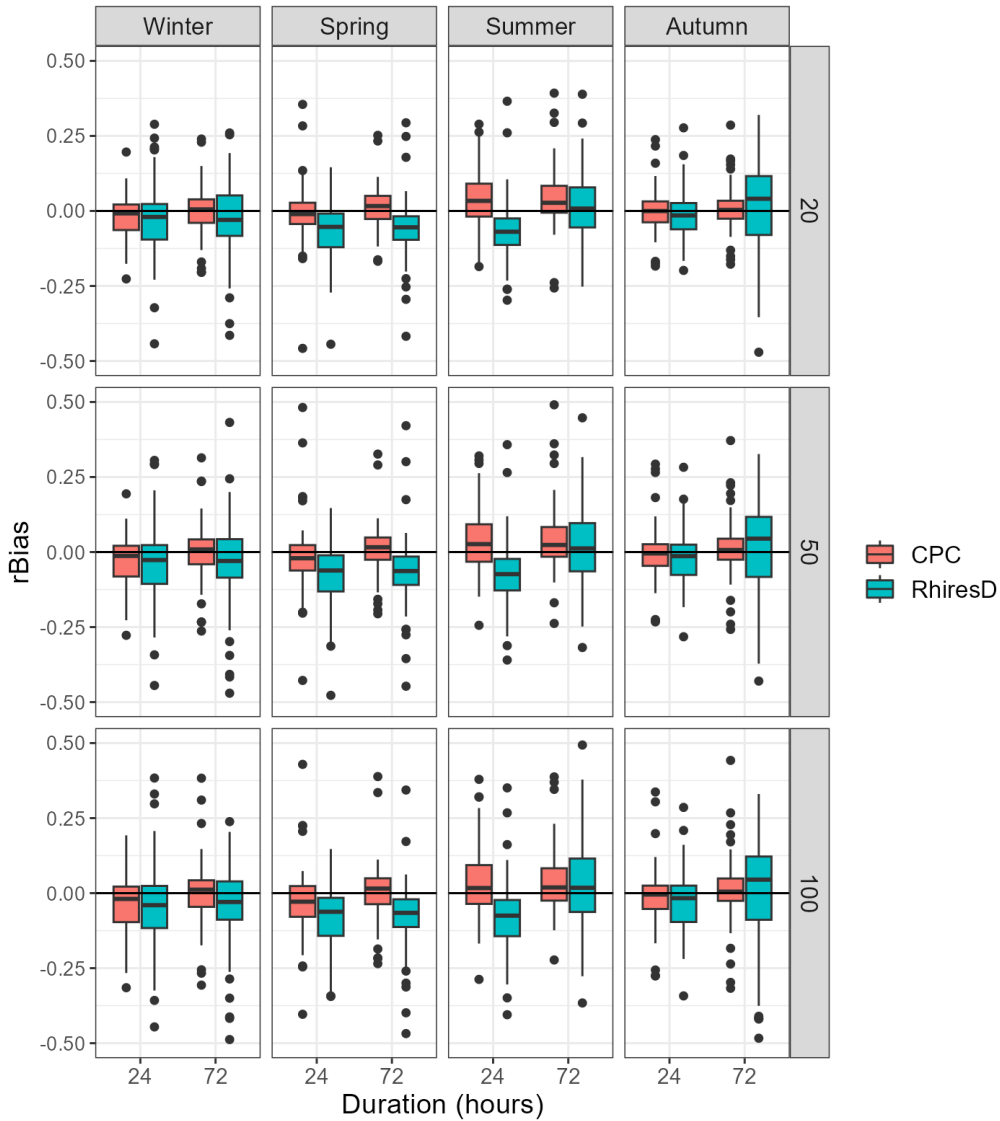


Figure VII.3: Seasonal boxplots of relative bias ($rBias$) of a T -year return level estimate for $T \in \{20, 50, 100\}$ years. Each column corresponds to a season, and each row corresponds to a return period. The boxplots contain 71 points, 1 point for each pair of rain gauge and the underlying pixel.

VII.1.2 Comparison of return level estimates

In the final step, we compare the T -year return level estimates, for $T \in \{20, 50, 100\}$, from the gridded data to those from the gauge stations. We fit EGPD to each dataset, estimate the T -year return level, and compute the relative bias ($rBias$) from:

$$rBias_T = \frac{\hat{i}_{T,grid} - \hat{i}_{T,gauge}}{\hat{i}_{T,gauge}},$$

where $\hat{i}_{T,grid}$ and $\hat{i}_{T,gauge}$ are the T -year return level estimates from the gridded data and the gauge data respectively.

Figure VII.3 shows the seasonal boxplot of $rBias$ in the T -year return level estimates. A positive bias indicates that the estimates from the gridded data are higher than the gauge estimate. For all seasons and return periods, CPC shows lower bias in the return level estimates compared

to RhiresD. The median of the bias from CPC is slightly positive in summer which signifies a tendency towards having higher estimates with CPC compared to the gauge.

VII.1.3 Conclusion and discussion

This section was aimed at evaluating CPC data, in comparison to daily data from the rain gauge and RhiresD, before using it in modeling the IDAF relationships. We comment here that, whilst CPC is corrected using the rain gauge data, some differences still remain between the two, mainly due to the nugget effect in the variogram model, and the convection control scheme in summer (Sideris et al., 2014a).

In general, from all the criteria considered, CPC showed a better agreement with the gauge data compared to RhiresD, especially for extremes. The better agreement of CPC might result from the effective resolution of the two datasets. CPC fields can be seen as areal estimates over a 1 km by 1 km grid. RhiresD, although on a 1 km by 1 km grid, has an effective resolution in the order of 15–20 km (MeteoSwiss, 2021), and therefore can be seen as areal estimates over a 250–400 km². Consequently, RhiresD fields are expected to be much smoother than those from CPC. This is seen from the bias of hits (the top row in Figure VII.2) where RhiresD shows a higher tendency to underestimate the extreme precipitation totals compared to CPC.

In the remainder of the thesis, only CPC is used to build the IDAF models since it is beyond the scope of the thesis to develop a new gridded product for our topographically complex study area. Moreover, it is the only product at the sub-daily temporal resolution. It also brings the required spatial information needed for modeling IDAF relationships which cannot be obtained from rain gauges due to their limited spatial representativity. We however acknowledge its obvious limitations, such as the limited length (17 years), conditional bias, and the non-homogeneity of the series due to radar upgrades and the evolution of the number of radars. These limitations will undoubtedly contribute to the uncertainties related to our analysis and results.

VII.2 Paper 3: Modeling Areal Precipitation Hazard: A Data-driven Approach to Model Intensity-Duration-Area-Frequency Relationships using the Full Range of Non-Zero Precipitation in Switzerland

Abubakar Haruna¹, Juliette Blanchet² and Anne-Catherine Favre¹

¹University Grenoble Alpes, Grenoble INP, CNRS, IRD, IGE, Grenoble, France,

²University Grenoble Alpes, CNRS, IRD, Grenoble INP, IGE, Grenoble, France.

Article under review at Water Resources Research (WRR)

<https://doi.org/10.22541/essoar.169111775.53035997/v1>

Preamble to the paper

Recall that in Chapter VI, we built Intensity-Duration-Frequency (IDF) models in the study area and used the selected model to construct catchment-level IDF curves for operational use (see section VI.2). A major drawback of IDF relationships is that they do not explain the relationship in space. In other words, they do not account for the spatial extents (*i.e.* area). This paper aims to develop models for Intensity-Duration-Area-Frequency (IDAF) relationships, which extend conventional IDF relationships to account for the areal extent of precipitation. These models can be used for quantifying areal rainfall hazards, characterizing storms, and developing early warning systems. The IDAF relationships are developed using all non-zero precipitation intensities, not just the extreme ones. The article uses the three-parameter extended generalized Pareto distribution (EGPD) to model the precipitation intensities and 17 years of data from the CombiPrecip radar-reanalysis product, which merges radar and rain gauge data in an operational setting using geostatistics. The article starts by reviewing and identifying a flexible and robust method of building IDAF models that is suitable for topographically complex locations. The method is based on a data-driven approach that links EGPD parameters with duration and area based on empirically determined parametric relationships. After evaluating the model's performance using split-sampling cross-validation, it is applied to construct IDF and Intensity-Area-Frequency (IAF) curves and to characterize and assess extreme areal precipitation hazards in Switzerland, the study area.

Abstract

Intensity-Duration-Area-Frequency (IDAF) models provide the mathematical link between precipitation intensities (I), durations (D), areas (A), and frequency of occurrence (F). They play a critical role in hydrological design, areal rainfall hazard quantification, storm characterization, and early warning system development. IDAF models extend the conventional Intensity-Duration-Frequency (IDF) models by accounting for the spatial extent of precipitation (*i.e.*, the area). In this study, we develop IDAF models using the entire non-zero precipitation intensities, not only the extremes. We use the extended generalized Pareto distribution (EGPD) to model the precipitation intensities. To build the IDAF models, we adopt a data-driven approach that allows the linkage of EGPD parameters with duration and area, based on empirically determined parametric relationships. The inference of model parameters is done using a global maximum likelihood estimation, and uncertainties are assessed by the bootstrap method. The study area is Switzerland, a topographically complex region of 42,000 km² with regional precipitation variability and clear seasonality. The study utilizes 17-years of data from CombiPrecip, a radar-reanalysis product developed by geostatistically merging radar and rain gauge data in an operational setting. We build the IDAF models for the spatio-temporal range of 1 to 72 hours and 1 to 1,089 km² at each pixel in the study area. To the best of our knowledge, our study is the first attempt to use the EGPD in IDAF curve modeling. It discusses the use and limitations of CombiPrecip in extreme value analysis and highlights the challenges of modeling areal precipitation in a complex topographical environment.

VII.2.1 Introduction

In the face of escalating threats posed by climate change and increasingly volatile weather patterns, understanding and predicting extreme precipitation is necessary, now more than ever, in safeguarding communities and infrastructure. One of the key factors driving flood generation is the spatial aggregate of precipitation over a given area, rather than just the precipitation at a specific point location. This is because watersheds and river basins integrate the precipitation falling over their respective areas, leading to the accumulation of runoff and subsequent flood generation. Additionally, extreme precipitation events, manifesting at varying scales, contribute differently to flood dynamics. Short and small-scale intense precipitation may induce rapid, localized flash flooding, whereas longer and larger-scale precipitation events can lead to sustained fluvial flooding (Sikorska et al., 2015). However, the interactions and synergies between these scales are crucial in shaping the overall flood risk landscape. As a consequence, it is vital to consider multiple spatio-temporal scales in the modeling of extreme precipitation. This will enhance our ability to better predict and manage the impacts on communities and infrastructure, ensuring their resilience in an ever-changing climate.

Intensity Duration Area Frequency (IDAF) curves summarize the main statistical characteristics of extreme precipitation (return level, return period, duration, and area.) They provide the mathematical link between precipitation intensities (I), durations (D), areas (A), and frequency of occurrence (F). They are useful tools for engineers and hydrologists in hydrological design (see Bertini et al., 2020, for example), quantification of areal rainfall hazard (Overeem et al., 2010; Panthou et al., 2014; Mélése et al., 2019; Zhao et al., 2023), storm characterization (Ramos et al., 2005; Ceresetti et al., 2012; Blanchet and Mélése, 2020), and development of early warning systems (Panziera et al., 2016). IDAF models extend the well-known Intensity Duration Frequency curves (IDF) by incorporating the spatial extent of precipitation (*i.e.*, the area).

IDAF curves are commonly built by coupling IDF models and a coefficient, the areal reduction

factor (ARF) that transforms point rainfall of a given duration and return period to areal return levels of the corresponding duration and return period. Applications of the ARF-based IDAF models can be found in the literature, for example, [De Michele et al. \(2001\)](#) derived an ARF formulation based on the concept of dynamic scaling of rainfall and used it to model IDAF curves in Milan. Later, [Ceresetti et al. \(2012\)](#) used the ARF of [De Michele et al. \(2001\)](#) to model IDAF curves for storm severity assessment in southern France. [Panthou et al. \(2014\)](#) also used the same ARF formulation to characterize areal rainfall in West Africa. [Ramos et al. \(2005\)](#) used an empirical ARF formulation to model IDAF curves for storm severity assessment in Marseille. [Bertini et al. \(2020\)](#) used another empirical ARF formulation to build IDAF curves and used it to design a dam in Italy. [Mélèse et al. \(2019\)](#) and [Blanchet and Mélèse \(2020\)](#) used an extension of the ARF formulation of [De Michele et al. \(2001\)](#) to build IDAF curves respectively for areal hazards and storm severity assessment in southern France. The extension was to cope with the significant spatio-temporal variability in the mountainous area

Beyond the ARF-based IDAF curves modeling approach, [Overeem et al. \(2010\)](#) proposed a purely data-driven approach to model IDAF curves. This involves modeling the parameters of the statistical distribution of the precipitation intensities as a function of duration and area. The type of relationship is empirically determined from the data, with no underlying physical hypothesis such as spatial correlation (as done in [Rodriguez-Iturbe and Mejía, 1974](#)) or scaling (as done in [De Michele et al., 2001](#)). As highlighted by [Mélèse et al. \(2019\)](#), this approach has the advantage of being flexible and applicable in cases where the assumptions of the analytical ARF formulations cannot be verified.

In spite of the chosen method of building the IDAF curves, whether ARF-based or purely data-driven, the previously cited works have one thing in common; the precipitation data they used and by extension, the underlying parametric distribution. To elaborate more, all the authors used only extreme data in the form of block maxima and, as the distribution, the generalized extreme value (GEV) distribution ([Overeem et al., 2010](#); [Ceresetti et al., 2012](#); [Panthou et al., 2014](#)) or its special case, the Gumbel distribution ([Nhat et al., 2007](#); [Mélèse et al., 2019](#); [Blanchet and Mélèse, 2020](#); [Bertini et al., 2020](#)), or lognormal distribution ([De Michele et al., 2011](#)). A rare application of generalized Pareto distribution (GPD) for threshold excesses is found in [Zhao et al. \(2023\)](#) for IDAF curve modeling. A major drawback of such approaches is the inefficient use of the data since only one value is retained in a block (usually the maximum in a year) or the excesses of a threshold (a tiny fraction of the data), and all the other data in the block is discarded. This can result in significant uncertainty in estimation, especially in cases where the length of the data series is not sufficiently long. The problem of short record length is more apparent with radar and radar reanalysis products, which are usually used in IDAF curve modeling ([Overeem et al., 2010](#); [Mélèse et al., 2019](#); [Blanchet and Mélèse, 2020](#); [Zhao et al., 2023](#)), due to the required spatial information they provide.

To address this issue, our approach here is to make efficient use of information by including all the non-zero precipitation intensities, instead of only the block maxima, in modeling the IDAF curves. We then use the extended generalized Pareto distribution (EGPD) of [Naveau et al. \(2016\)](#) as the parametric model for the intensities. This distribution is compliant with extreme value theory in both tails (an advantage over the gamma distribution), it models the entire distribution of non-zero precipitation and does not require the choice of the threshold as in the generalized Pareto distribution (GPD). It has been shown in many applications to be able of adequately modeling precipitation ([Naveau et al., 2016](#); [Evin et al., 2018](#); [Le Gall et al., 2022](#); [Haruna et al., 2022, 2023b](#)). In particular, [Haruna et al. \(2023b\)](#) showed that it is possible to model IDF curves (without Area) with the EGPD, and we intend to extend their work to model IDAF curves with the EGPD. To our best knowledge, this is the first time the EGPD is used in modeling IDAF curves.

Modeling IDAF curves using all the non-zero data has two potential advantages. First, by using all the non-zero data, estimation uncertainty is expected to reduce, resulting in more accurate

predictions. Secondly, in addition to having IDAF curves, we will have robust marginal distributions for the entire non-zero precipitation that can be used in stochastic weather generators for simulations, or verification of weather and climate models.

We apply the model in Switzerland, a topographically complex location with seasonality, regional variability, and multiple precipitation regimes. Following the work of [Mélèse et al. \(2019\)](#) which underscores the complex spatio-temporal variability of precipitation in mountainous areas, we use the more flexible data-driven method of [Overeem et al. \(2010\)](#) to model the IDAF curves.

The data and study area are presented in Section VII.2.2. The EGPD, the methodology for building the IDAF curves, and the method for uncertainty assessment are explained in Section VII.2.3. Results on the goodness-of-fit of the model and areal rainfall hazard assessment in the study area are presented and discussed in Section VII.2.4. Finally, conclusions and perspectives are given in Section VII.2.5

VII.2.2 Study area and Data

VII.2.2.1 Study Area

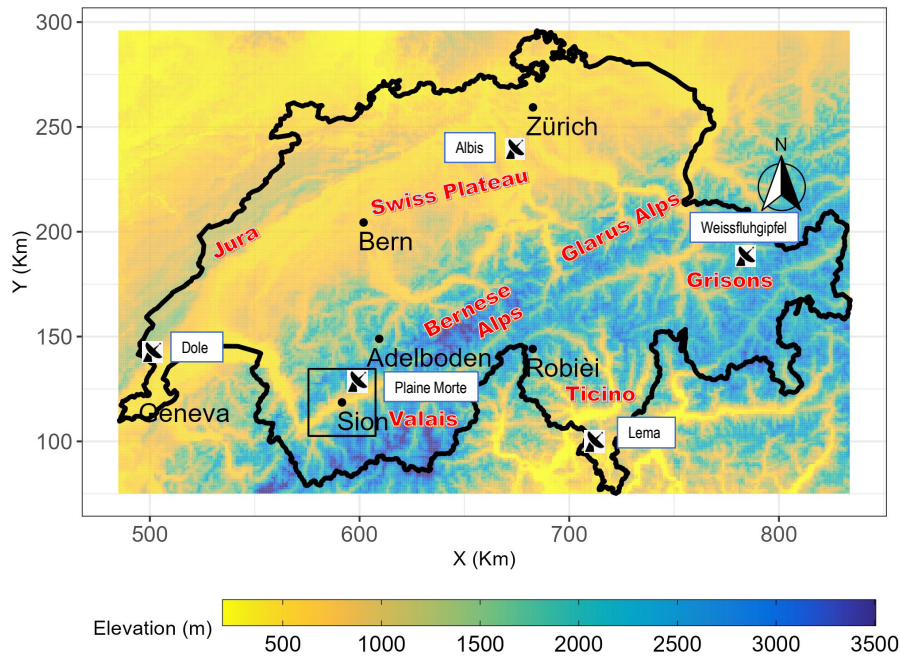


Figure VII.4: Map of Switzerland, the study area. The background color denotes the elevation (meters) above mean sea level. The Radar symbols show the location of the five radars in Switzerland, with their names in the white boxes. The name of some cities is shown in black and the name of some mountains and regions are shown in red. The black-colored square shows exemplarily the maximum extent of the rectangular window used for data aggregation, i.e. 1089 km².

Our study focuses on Switzerland, a country covering 41,285 km². Despite its small size, Switzerland exhibits a complex topography, ranging in elevation from 191 to 4,127 m above mean sea level. Figure VII.4 shows the map of the study area. Approximately 30% of the land is situated above 1,500 m elevation, resulting in pronounced spatial variability in both the intensity and occurrence of precipitation. The climate of Switzerland is influenced by multiple factors, such as the Alps, the Atlantic Ocean, and the Mediterranean Sea, and these contribute to the seasonal and spatial variability of precipitation, as documented in previous studies ([Sodemann and Zubler,](#)

2009; Giannakaki and Martius, 2015; Scherrer et al., 2016a). Precipitation patterns show distinct regional differences, with the highest annual sums exceeding 2,000 mm in the Alps, the Jura region (northwest), and the Ticino region (south of the Alps). Conversely, the inner valleys such as the Rhône and Inn receive the lowest annual precipitation, less than 700 mm. Summer is the primary season for precipitation throughout Switzerland, except in Ticino, where autumn dominates. Conversely, winter experiences the least amount of precipitation across all regions. In terms of heavy precipitation, defined as the average seasonal maxima, the spatial distribution varies according to accumulation duration (Panziera et al., 2018). For short-duration accumulations (*e.g.*, 1 hr), the heaviest precipitation occurs in summer across the entire country, with maximum intensities reaching up to 30 mm/hr in Ticino, Jura, and the northern rim. For longer accumulations (1 day and more), Ticino receives the most intense precipitation, with autumn experiencing maximum 24 hr totals exceeding 130 mm. In other regions, heavy precipitation predominantly occurs during summer.

VII.2.2.2 CombiPrecip

CombiPrecip (CPC) is a radar-reanalysis product resulting from the geostatistical merging of radar and rain gauge in an operational setting (Sideris et al., 2014a). It combines the high accuracy of rain gauge with the high spatial coverage of radar. The geostatistical merging is through co-kriging with external drift, where the rain gauge data is treated as the primary source, and the radar data as the external drift. Information from rain gauge comes from more than 250 automatic stations at 10 minutes resolution, and that from the radar comes from five polarimetric C-band Doppler radars that are suitably located to provide the reliable coverage required in the topographically complex area (see Figure VII.4). Since CPC is produced operationally, only rain gauge data within Switzerland are used in the algorithm. As a result, an algorithm for the treatment of extrapolation is used in which some radar pixels outside the Switzerland border are used as virtual rain gauges in the merging. Additionally, a convection control scheme is implemented to overcome the limited representativeness of rain gauges during convection events, especially in summer (see Sideris et al., 2014b, for details)

The data from both the rain gauge and radar are subjected to substantial quality control before being employed in the CPC algorithm. The gauge data is checked to ensure that recorded values are within climatologically physical limits, they are consistent with those from nearby gauges, they satisfy inter-parameter consistency, and variability tests (MeteoSwiss, 2017). Treatment of the radar data (Germann et al., 2006) involves clutter elimination through a robust algorithm designed for this purpose, visibility correction resulting from orthographic shielding, correction for vertical profile of reflectivity, and bias correction. This is in addition to an automatic hardware calibration of the radars to check the stability/accuracy of the components and a tailored operational scan strategy (20 elevation sweeps every five minutes) crucial in mountainous regions such as Switzerland (Germann et al., 2015).

CPC is available at hourly temporal resolution and a spatial grid of 1 km by 1 km and extends 100-150 km beyond the borders of Switzerland. It has been available since 2005, and 17 years of data are available for this study, from 1st January 2005 to 31st December 2021. It has been used in several applications in Switzerland for extreme value analysis (Panziera et al., 2016), climatological studies (Panziera et al., 2018), meteorological forcing of hydrological model (Andres et al., 2016), and has been evaluated in several aspects (Gabella et al., 2017; Panziera et al., 2018; Gugerli et al., 2020). Known limitations of CPC involve the limited length of the data, non-homogeneity of the series due to radar upgrades and evolution of the number of radars, and conditional bias (MeteoSwiss, 2017). Despite these limitations, it is the only sub-daily gridded data available in the study area, and producing a gridded product is beyond the scope of the present study. We note that these limitations are not unique to CPC alone, but common to other radar and radar-reanalysis

products, and notwithstanding, they have been used in IDAF modeling (e.g. [Overeem et al., 2010](#); [Mélèse et al., 2019](#); [Blanchet and Mélèse, 2020](#)), or extreme value analysis ([Durrans et al., 2002](#); [Allen and DeGaetano, 2005](#); [Wright et al., 2014](#); [Goudenhoofd et al., 2017](#); [Panziera et al., 2018](#)). This is due to the detailed spatial representativeness they provide, especially in mountainous areas, which is practically unobtainable with rain gauge networks alone.

VII.2.3 Methodology

VII.2.3.1 Space-time aggregation of the data

The total area of Switzerland is 41,285 km², and so we have hourly time series of precipitation at 41,285 CPC pixels, each of size 1 km². We take each time series and stratify it into four seasons, with winter (Dec-Jan-Feb), spring (Mar-Apr-May), summer (Jun-Jul-Aug), and autumn (Sep-Oct-Nov). This seasonal approach is done to account for the pronounced seasonality in the study area, as done in several studies in the same area ([Molnar and Burlando, 2008](#); [Fukutome et al., 2015](#); [Panziera et al., 2018](#); [Evin et al., 2018](#); [Haruna et al., 2022, 2023b](#)). To produce the areal precipitation for use in modeling the IDAF relationships, we aggregate the data into 10 spatial scales (area) and 10 temporal scales (duration). The area includes 1, 9, 25, 49, 81, 169, 289, 529, 729, and 1,089 km², centered around each pixel in the study domain, which corresponds to squares of sides 1, 3, 5, 7, 9, 13, 17, 23, 27 and 33 km. Since CPC is available beyond the borders, it allows us to have spatially aggregated rainfall everywhere in Switzerland (including the pixels close to the border). We comment here that the choice of the squared area is for simplicity and convenience since the CPC data is originally in this geometry. Other choices are possible such as circular or elliptical shapes as discussed in [Mélèse et al. \(2019\)](#). The durations include 1, 2, 3, 6, 10, 12, 16, 24, 48 and 72 hr. We consider durations up to 72 hr (3 days) because according to [Froidevaux et al. \(2015\)](#) these time scales are the most relevant for flood-triggering precipitation accumulations in Switzerland. The intermediate area and durations are meant to ensure a good spread on a logarithmic scale. At the end of the aggregation, we have a total of 100 time series of areal precipitation, each for a pair (D, A). Unlike in the case where only block maxima will be used for modeling the IDAF relationships, here, we retain and use all the non-zero precipitation intensities in modeling the IDAF relationships. Although we have the areal precipitation at all the pixels, for computational reasons (an average of 260,000 non-zero observations in summer, at each pixel location), we fit the IDAF model only at a subset of the pixels, by considering every second and third pixel along the latitude and longitude respectively. This results in a total of 7,056 pixels.

VII.2.3.2 Marginal Distribution for Non-Zero Precipitation Intensities

We use the three-parameter EGPD of [Naveau et al. \(2016\)](#) as the marginal distribution for the non-zero rainfall intensities in the IDAF model. The model is an extension of the classical GPD (which applies only to the excesses of a chosen threshold) to model the entire distribution of precipitation intensities (the low, medium, and extremes). The first advantage of EGPD is that since it is an extension of GPD, it is compliant with extreme value theory, so it behaves like the GPD in the upper tail of the distribution, *i.e.* the same shape parameter (see [Tencaliec et al., 2020](#), for demonstration). Secondly, since it makes use of all the non-zero precipitation data, one does not need to worry about the delicate issue of threshold selection that is known with the GPD. Finally, it models the whole range of non-zero precipitation, which has several practical applications in cases where the interest is not only in the largest values but in the medium and low values as well (*e.g.* in simulation frameworks or climatological studies).

We define the random variable I to represent non-zero rainfall intensities. We assume that it follows the EGPD whose cumulative distribution function (CDF) is defined as:

$$\mathbb{P}(I \leq i) = \left[H_{\xi} \left(\frac{i}{\sigma} \right) \right]^{\kappa}, \quad (\text{VII.1})$$

with

$$H_{\xi} \left(\frac{i}{\sigma} \right) = \begin{cases} 1 - (1 + \xi \frac{i}{\sigma})_+^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - \exp(-\frac{i}{\sigma}) & \text{if } \xi = 0 \end{cases}, \quad (\text{VII.2})$$

where $a_+ = \max(a, 0)$, $\sigma > 0$ is the scale parameter, and $\xi \geq 0$ is the shape parameter that controls the upper tail of the distribution. The flexibility parameter, $\kappa > 0$ controls the lower tail. With the addition of only one parameter, κ , compared to the GPD, the distribution is able to model the full range of non-zero precipitation (see applications in [Evin et al., 2018](#); [Le Gall et al., 2022](#); [Haruna et al., 2022, 2023b](#)).

VII.2.3.3 EGPD-IDAF Model

Our assumption is that the random variable of non-zero precipitation intensities for any duration D and area A , $I(D, A)$ follows the EGPD, *i.e.*:

$$I(D, A) \sim \text{EGPD}[\kappa(D, A), \sigma(D, A), \xi(D, A)], \quad (\text{VII.3})$$

where $\kappa(D, A) > 0$, $\sigma(D, A) > 0$ and $\xi(D, A) \geq 0$ are the three EGPD parameters for the duration D and area A .

Let $F_{D,A}(i)$ be the CDF of $I(D, A)$, such that $F_{D,A}(i) = \mathbb{P}(I_{D,A} < i)$, then IDAF curve, which is the T -year return level over duration D and area A is defined by the quantile function of $F_{D,A}$, *i.e.*:

$$i(T, D, A) = \frac{\sigma(D, A)}{\xi(D, A)} \left\{ \left(1 - \left[1 - \frac{1}{T \times \delta(D, A)} \right]^{\frac{1}{\kappa(D, A)}} \right)^{-\xi(D, A)} - 1 \right\}, \quad (\text{VII.4})$$

where $\kappa(D, A) > 0$, $\sigma(D, A) > 0$ and $\xi(D, A) \geq 0$ are the three EGPD parameters for the duration D and area A . T is the return period in years, $\delta(D, A)$ is the average number of non-zero precipitation intensities for duration D and area A per year. We estimate $\delta_{D,A}$ based on the long-term average of the non-zero precipitation intensities per year.

As already highlighted in Section [VII.2.1](#), we use the data-driven approach of [Overeem et al. \(2010\)](#) to model the IDAF relationships. The approach involves empirically finding the appropriate regression model to explain the relationship between each of the three EGPD parameters as a function of duration and area. We will now explain our methodology to determine the appropriate regression model.

We begin by considering each pixel and fitting EGPD separately to the 100 aggregated time series of scales (D , A) at that pixel location. We then examine how the three EGPD parameters change with A and D . To model the relationships, we test various regression models using A , D , their transformations; $\log(A)$, $\log(D)$, \sqrt{A} , \sqrt{D} , as well as some interactions terms. To avoid having a different regression model at each pixel, we compare competing models regionally by assessing their predictive performance in cross-validation. In the end, we retain the following regression models for the EGPD parameters:

$$\begin{aligned} \log[\kappa(D, A)] &= \beta_{0,\kappa} + \beta_{1,\kappa}A + \beta_{2,\kappa}D + \beta_{3,\kappa} \log(A) + \beta_{4,\kappa} \sqrt{D} + \beta_{5,\kappa} \sqrt{D} \log(A) + \beta_{6,\kappa} D \log(A) \\ \log[\sigma(D, A)] &= \beta_{0,\sigma} + \beta_{1,\sigma}A + \beta_{2,\sigma}D + \beta_{3,\sigma} \log(A) + \beta_{4,\sigma} \sqrt{D} + \beta_{5,\sigma} \sqrt{D} \log(A) + \beta_{6,\sigma} D \log(A), \\ \xi(D, A) &= \beta_{0,\xi} + \beta_{1,\xi}D + \beta_{2,\xi} \log(A) + \beta_{3,\xi} \sqrt{D} + \beta_{4,\xi} \sqrt{D} \log(A) + \beta_{5,\xi} D \log(A) \end{aligned} \quad (\text{VII.5})$$

where D is in hours and A is in km^2 . $\beta_{i,*}$ for $i = 0, 1, \dots, 6$ are the regression coefficients. The scale (σ) and flexibility parameter (κ) both have a log link transformation because of their positive support. They both have seven regression parameters (β_i for $i = 0, 1, \dots, 6$). The shape parameter ξ has six parameters, making a total of 20 parameters for the complete EGPD-IDAF model for each season and pixel location. We note here that while the number of parameters might appear large, the model is still parsimonious compared to fitting EGPD separately for each time series of (D, A) , which amounts to a total of 300 parameters (three (3) EGPD parameters for the 100-time series in our case). In the result Section we will show additional performance comparisons between the 20-parameter EGPD-IDAF model, and the 300-parameter base model. In addition to this, the relative complexity of the model (in terms of parameterization), highlights the inherent difficulty of modeling areal precipitation in mountainous regions, where areal rainfall is less homogeneous in space compared to relatively flat regions. A similar attempt to model IDAF curves in southern France (Massif Central) by M el ese et al. (2019) highlights similar complexity.

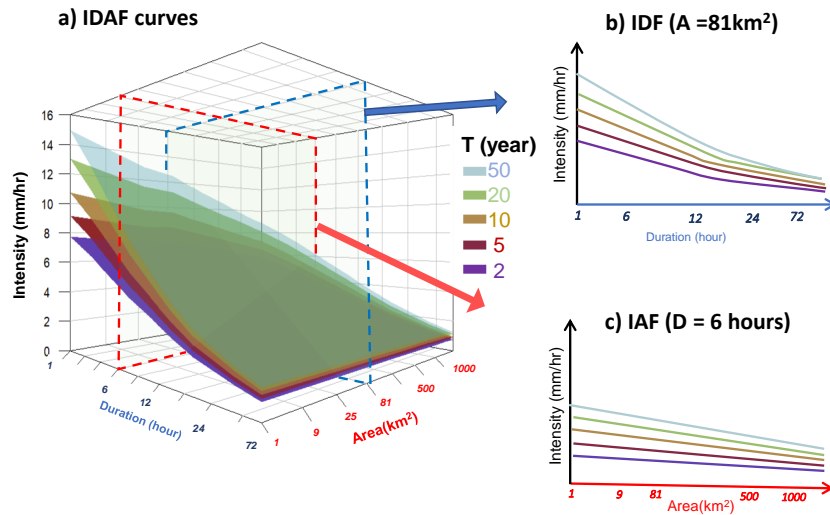


Figure VII.5: a) Conceptual illustration of IDAF curves in 3-dimension. IDF curves for $A = 81 \text{ km}^2$ (shown in panel b) are obtained by cutting a plane on the IDAF curves in panel a at $A = 81 \text{ km}^2$ (red-colored broken lines). The IAF curves on panel c) are obtained by cutting a plane at $D = 6 \text{ hr}$ on panel a (blue-colored broken lines).

To conclude this section, we illustrate a conceptual plot of IDAF curves in Figure VII.5. A plot of IDAF curves is 3-dimensional (Figure VII.5a), with Intensity (I) along the vertical axis, duration (D) along the horizontal axis, and area (A) along the third axis which is perpendicular to the other two axes. For each specific return period (e.g. 2-year, 10-year, or 50-year), a curve is shown to visualize how the intensity changes across A and D . However, a much simpler approach is to decouple the 3-dimensional plot into two sub-plots, each in 2-dimension. The first one shows how the intensities of specific return periods change across durations for a fixed area, i.e. IDF curves (Figure VII.5b), and the second one, a plot of Intensity-Area-Frequency (IAF) curves, shows how the intensities change across areas for a fixed duration (Figure VII.5c).

VII.2.3.4 Model Estimation

Let us call θ the vector of 20 regression parameters of the EGPD-IDAF model to be estimated at a given pixel location. We estimate θ by maximizing the censored log-likelihood of the EGPD-

IDAF model, which is given by:

$$l(\boldsymbol{\theta}) = \sum_A \sum_D \sum_{j:i(D,A,j) < C(D,A)} \log F_{D,A} [C(D,A)] + \sum_A \sum_D \sum_{j:i(D,A,j) \geq C(D,A)} \log f_{D,A} [i(D,A,j)], \quad (\text{VII.6})$$

where $\boldsymbol{\theta}$ is the vector of the 20 regression parameters to be estimated. $F_{D,A}$ and $f_{D,A}$ are the CDF and PDF of the EGPD associated with (D, A) , $i_{(D,A,j)}$ is the precipitation intensity for (D, A) and time step j . $C_{(D,A)} \geq 0$ is the left censoring threshold applied to the data of (D, A) . The log-likelihood is finally expressed in Equation VII.7 as:

$$\begin{aligned} l(\boldsymbol{\theta}) = & \sum_A \sum_D \sum_{j:i(D,A,j) < C(D,A)} \kappa(D,A) \log \left[1 - \left(1 + \frac{\xi(D,A)C(D,A)}{\sigma(D,A)} \right)^{-\frac{1}{\xi(D,A)}} \right] + \\ & \sum_A \sum_D \sum_{j:i(D,A,j) \geq C(D,A)} \log \kappa(D,A) - \sum_A \sum_D \sum_{j:i(D,A,j) \geq C(D,A)} \log \sigma(D,A) - \\ & \sum_A \sum_D \sum_{j:i(D,A,j) \geq C(D,A)} \left[1 + \frac{\xi(D,A)i(D,A,j)}{\sigma(D,A)} \right]^{[1+\frac{1}{\xi(D,A)}]} + \\ & \sum_A \sum_D \sum_{j:i(D,A,j) \geq C(D,A)} \left[1 - \left(\left(1 + \frac{\xi(D,A)i(D,A,j)}{\sigma(D,A)} \right)^{-\frac{1}{\xi(D,A)}} \right) \right]^{[\kappa(D,A)-1]}, \end{aligned} \quad (\text{VII.7})$$

where $\kappa(D,A) > 0$, $\sigma(D,A) > 0$ and $\xi(D,A) \geq 0$, are the EGPD parameters for (D,A) and the other variables retain their earlier definitions.

We highlight two important points here: the use of censored likelihood, and the use of the independence likelihood. First, the use of censored likelihood is done to reduce the influence of the small intensities on the MLE. Without censoring, the smaller intensities influence the parameter estimation, thereby resulting in a large overestimation of the upper tail shape parameter (ξ). This is likely due to the insufficient flexibility of the three-parameter EGPD model to adequately model the left tail of the distribution or the associated uncertainty in the instrumental recording of very small intensities. A usual censoring approach is to apply a uniform censoring threshold (*e.g.* 2 mm for all the daily data, or 0.5 mm for all the hourly intensities), but as highlighted by [Haruna et al. \(2023b\)](#), this is not usually sufficient, and so, an appropriate censoring threshold has to be obtained for each time series. We follow their footsteps and estimate a threshold, for each time series of (D, A) that minimizes the squared error between the modeled and empirical quantiles (see Equation VII.9). This approach usually results in an adequate fit of the model.

The second point is on the use of independence likelihood, which assumes independence in the data. This assumption is unlikely to hold given that we have three levels of dependence in the data; serial dependence within time series of the same (D, A) , dependence between time series of different durations (*e.g.* time series of 1 hr and 1 km², *versus* time series of 2 hr and 1 km²), and lastly the dependence between time series of different spatial scales (*e.g.* time series of 1 hr and 1 km², *versus* time series of 1 hr and 3 km²). Despite these, since our target is on the marginal (univariate) return levels, the violation of the independence assumption is unlikely to induce bias in our estimates ([Sebille et al., 2017](#)). Additionally, within the framework of modeling IDF curves using generalized extreme value (GEV) distribution, [Jurado et al. \(2020\)](#) showed that little gain in performance is achieved by explicitly modeling the dependence between the data of different durations, in addition to the added complexity. Since their application is with GEV rather than EGPD, an interesting perspective is to investigate this effect with the EGPD. Here we retain the independence assumption to avoid additional complexity to our model which already has 20 free parameters. Notwithstanding, we apply temporal declustering to reduce the serial dependence

in our time series (as done in [Naveau et al. \(2016\)](#); [Le Gall et al. \(2022\)](#); [Haruna et al. \(2022, 2023b\)](#)). To achieve this, we retain every 3rd observation in the 1 hr time series, and every 4th, 5th, 8th, 10th, 12th, 16th, 24th, 48th, 72nd, respectively in the time series of 2, 3, 6, 10, 12, 16, 24, 48, and 72 hr. Finally, to avoid underestimating uncertainties in our model, which is one of the main consequences of the independence assumption, we resort to block-bootstrapping for uncertainty assessment (see Section [VII.2.3.5](#)).

VII.2.3.5 Uncertainty Assessment

In order to assess uncertainty in the EGPD-IDAF model, we use the block bootstrap approach ([Kunsch, 1989](#)). The principle of the block bootstrap involves dividing the time series into blocks of consecutive observations. Resamples are then generated by randomly selecting blocks with replacements and concatenating them to create a bootstrap sample. By preserving the block structure, the block bootstrap can capture the dependence structure of the original data. This approach is suitable for uncertainty estimation in our case, where we made the independence assumption in the likelihood estimation of the parameters. The block bootstrap method was used for uncertainty estimation by [Overeem et al. \(2010\)](#) in IDAF curves modeling, and by [Overeem et al. \(2009\)](#); [Haruna et al. \(2023b\)](#) in IDF curves modeling.

To apply the block bootstrap approach, we take the seasonal time series at each pixel and estimate the uncertainty by following the outlined steps below:

1. Aggregate the time series into the 10 durations and 10 areas, resulting in a total of 100-time series, each for a pair of duration and area (D, A). Decluster each of the series according to the declustering procedure explained in Section [VII.2.3.4](#). We call this sample M_{orig} .
2. Randomly select blocks of size 2 weeks with replacement, G times, to form the resampled time series (M_{boot}). Both M_{orig} and M_{boot} have the same dimensions. The block bootstrapping ensures that we keep the data of the different durations D and areas A together, and hence the dependence structure. We use a block size of 2 weeks, beyond which the autocorrelation in the data does not decrease, as done in [Haruna et al. \(2023b\)](#) for the same study area in the case of IDF curve modeling.
3. Fit the EGPD-IDAF model to the data in M_{boot} and estimate the intended return levels.
4. Repeat steps 2 to 3 a total of 300 times to obtain the bootstrap distribution of the return levels. Finally, compute the 95% Confidence Interval (CI) of the return levels by the percentile method. This is done by taking the empirical 0.025 and 0.975 quantiles of the bootstrap distribution of the return levels obtained in step 4.

As a measure of model precision, we compute the normalized width of the 95% CI of a T -year return level estimate ([Shehu and Haberlandt, 2023](#)). For a given pixel location s , it is computed from:

$$\text{n95CI}_{\text{width},s} = \frac{r_{T,97.5\%} - r_{T,2.5\%}}{\bar{r}_T}, \quad (\text{VII.8})$$

where $r_{T,p\%}$ is the $p\%$ quantile of the 300 bootstrap estimates of the T -year return level (r_T) and \bar{r}_T denotes the average of the 300 estimates. The normalization is to enable the comparison of uncertainty width across intensities of different scales and return periods.

VII.2.3.6 Goodness-of-fit of the EGPD IDAF model

To assess the goodness of fit of the EGPD-IDAF model, we compute the normalized root-mean-square error (NRMSE) and the normalized bias (NBias) at each pixel s and spatio-temporal scale (D, A). The criteria are computed by comparing the estimated return levels from the model to the empirical levels computed using the Weibull plotting position defined as $\frac{j}{n+1}$ with j being the rank (from largest to smallest) and n is the sample size. The normalization allows comparison of the score across intensities of different spatio-temporal scales (D, A). For a given pixel s , the two criteria are given as:

$$\text{NRMSE}_s = \frac{\left\{ \frac{1}{n_s} \sum_{j=1}^{n_s} (\hat{r}_{s,T_j} - r_{s,T_j})^2 \right\}^{1/2}}{\bar{r}_s}, \quad (\text{VII.9})$$

$$\text{NBias}_s = \frac{\frac{1}{n_s} \sum_{j=1}^{n_s} (\hat{r}_{s,T_j} - r_{s,T_j})}{\bar{r}_s}, \quad (\text{VII.10})$$

where n_s is the sample size, r_{s,T_j} is the empirical quantile with return period $T_j = \frac{n_s+1}{j \times \delta}$, δ is the average number of non-zero precipitations for (D, A) per year, \hat{r}_{s,T_j} is the corresponding T_j return level estimated from the EGPD-IDAF model. The denominator is the mean of non-zero precipitation of (D, A) at pixel s computed from $\frac{1}{n_s} \sum_{j=1}^{n_s} r_{s,T_j}$.

NRMSE measures the accuracy of the model in predicting the empirical quantiles. A good model should have NRMSE = 0, and the smaller the score, the better the model. NBias measure the ability of the model to avoid systematic underestimation (NBias < 0) or overestimation (NBias > 0) of the empirical quantiles. NBias = 0 means an unbiased model.

VII.2.3.7 Cross validation

A natural question to ask is whether the EGPD-IDAF model which links the EGPD parameters with duration and area is a better model, in terms of some performance indicators, compared to fitting the EGPD model separately to each time series of spatio-temporal scale (D, A). The two models will henceforth be referred to as the global model and the base model, respectively. To answer this, we compare the two models in a split-sample cross-validation framework. We will start by describing the cross-validation framework, and then introduce the criteria for measuring the performance.

In the split sampling cross-validation, we consider each pixel and divide the time series into two subsamples of the same length but on different randomly chosen years. We consider the first sub-sample, aggregate the data into the 10 durations and 10 areas, and fit the two competing models, *i.e.*, the base model and the global model. We then assess how the two models perform on the second sub-sample (validation sample). A good predictive model should perform well in the data not used in training it. We do the same on the second sub-sample (use it as the training sample, and the first sub-sample as the validation sample). Since the split sampling is done randomly, we repeat the procedure 40 times to address sampling bias. We apply the same procedure to all the pixels in the study area. We then select the method that has the best regional performance (average of the scores over all the pixels.)

We use some well-chosen predictive performance criteria to measure the performance of the models. The criteria have seen wide applications in the literature (see [Garavaglia et al., 2011](#); [Renard et al., 2013](#); [Blanchet et al., 2015](#); [Evin et al., 2016](#); [Haruna et al., 2022, 2023b](#)). We give a brief overview of the criteria, while details can be found in the cited references.

- **Robustness:** The Robustness criteria, SPAN, measures the ability of a model to give similar estimates of a high return level when data from two different calibration periods are used to

train the model (Garavaglia et al., 2011). At a given pixel (s) and for a spatio-temporal scale (D, A), SPAN is computed as:

$$\text{SPAN}_{s,T} = \frac{2 \left| \hat{r}_{s,T}^{(1)} - \hat{r}_{s,T}^{(2)} \right|}{\left(\hat{r}_{s,T}^{(1)} + \hat{r}_{s,T}^{(2)} \right)}, \quad (\text{VII.11})$$

where $\hat{r}_{s,T}^{(1)}$ and $\hat{r}_{s,T}^{(2)}$ are the T -year return levels estimated from sub-sample 1 and 2 respectively at pixel s . A SPAN of 0.5 means that the absolute difference between the two return levels is half of their average.

A regional value of SPAN, over Switzerland, is computed as $\text{SPAN}_{\text{reg},T} = 1 - \frac{1}{N} \sum_{s=1}^N \text{SPAN}_{s,T}$, where $N = 7,056$ is the total number of pixels. A perfectly robust model should have $\text{SPAN}_{\text{reg},T} = 1$.

- Reliability in predicting the maximum value: At a given pixel (s) and for a given spatio-temporal scale (D, A), the reliability of the model fitted on sub-sample 1 in predicting the maxima in sub-sample 2 and *vice versa* is measured by the FF criteria as follows:

$$\text{FF}_s^{(12)} = \left[\hat{F}_s^{(1)} \left(\max_s^{(2)} \right) \right]^{n_s^{(2)}}, \quad (\text{VII.12})$$

where $\text{FF}_s^{(12)}$ is the cross-validation criteria computed at pixel s , by predicting the probability of the maximum value in sub-sample 2, of sample size $n_s^{(2)}$ using the model fitted on the sub-sample 1. $\text{FF}_s^{(21)}$ is computed symmetrically.

As noted by Renard et al. (2013) and Blanchet et al. (2015), if the fitted model is a good estimate of the true distribution of the data, $\text{FF}_s^{(12)}$ should be a realization of a uniform distribution. Hence, the difference in the area, noted *diff*, between a theoretical uniform distribution and that of the $N = 7,056$ values of $\text{FF}_s^{(12)}$ (computed over the N pixels), should be close to zero. FF_{reg} at the regional scale, given as $1 - \text{diff}$, should therefore take a value of 1 for a reliable model and 0 for a completely unreliable model; the lower the value the less reliable the model is.

- The reliability/accuracy over the entire observations: While the previous reliability score (FF), and SPAN focus on extremes only, it is important that the model is also reliable in the bulk of the distribution, especially given that we use the EGPD. To measure the reliability of a model in predicting all the observations in cross-validation, we use the normalized root mean square error (NRMSE_CV), which is expressed as:

$$\text{NRMSE_CV}_s^{(12)} = \frac{\left\{ \frac{1}{n_s^{(2)}} \sum_{j=1}^{n_s^{(2)}} \left(r_{s,T_j}^{(2)} - \hat{r}_{s,T_j}^{(1)} \right)^2 \right\}^{1/2}}{r_s^{(2)}}, \quad (\text{VII.13})$$

where NRMSE_CV_s^{12} is the score computed at pixel s , $n_s^{(2)}$ is the sample size of the second sub-sample, $r_{s,T_j}^{(2)}$ is the empirical quantile with return period $T_j = \frac{n_s+1}{j \times \delta}$, δ is the average number of non-zero precipitations for (D, A) per year in sub-sample 2, $\hat{r}_{s,T_j}^{(1)}$ is the corresponding T_j return level estimated from the model fitted on sub-sample 1. The denominator is the mean of non-zero precipitation in sub-sample 2 at pixel s computed as $\frac{1}{n_s^{(2)}} \sum_{j=1}^{n_s^{(2)}} r_{s,T_j}^{(2)}$.

Similar to the other criteria, the regional score for each spatio-temporal scale (D,A), computed over the N pixels, is given as $\text{NRMSE_CV}_{\text{reg}}^{(12)} = 1 - \frac{1}{N} \sum_{s=1}^N \text{NRMSE_CV}_s^{(12)}$. The

other score, $\text{NRMSE_CV}_{\text{reg}}^{(21)}$ is computed symmetrically. $\text{NRMSE_CV}_{\text{reg}} = 1$ indicates a perfectly accurate model (the model accurately predicts the empirical return levels).

VII.2.4 Results and Discussion

VII.2.4.1 Validation of CPC data

We start with a validation/quality check on the CPC by comparing the time series with those from the rain gauge through a point-to-pixel analysis. This involves comparing the time series from a gauge to the time series from a CPC pixel at the location of the gauge. We consider 71 gauge locations and in each case, the data from the gauge is considered the "truth". These 71 stations have no missing data from 2005 to 2020 (the period of overlap of both datasets). The comparison is in two steps, in the first step, we compare the two time series using some chosen criteria and in the second step, we fit EGPD to both time series and compare the 20-year return level estimate. The result of the comparison is presented in the following subsections.

VII.2.4.1.a Comparison on the empirical values

i) Criteria on all observations

Following the work of [Zambrano-Bigiarini et al. \(2017\)](#), we use the three sub-components of the Kling-Gupta-Efficiency (KGE) criterion ([Kling et al., 2012](#)) to compare the two datasets (see Equation A.15 in Appendix A.3.1). The first is the bias (the tendency of CPC to under or overestimate the gauge data). The second is the variability ratio, which measures the under or over-dispersion of CPC data compared to the gauge. The third component measures the linear correlation between the two time series. For a perfect match between the gauge and CPC, all the criteria should be equal to 1.

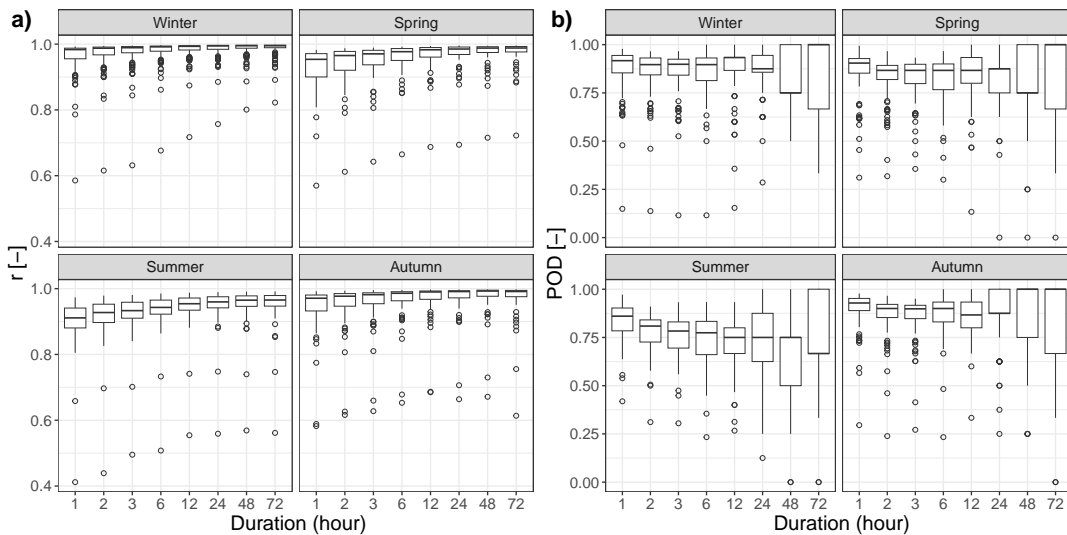


Figure VII.6: Boxplots of linear correlation (a) and probability of detection (POD) (b) for the four seasons. Each boxplot contains 71 points, 1 point for each pair of rain gauge and the underlying CPC pixel.

The boxplot of the correlation coefficient is shown in Figure VII.6a for the four seasons and eight aggregation durations (1, 2, 3, 6, 12, 24, 48, and 72 hr). Generally, there is a good temporal correlation between the two data sets for all seasons and durations (median > 0.9). For all seasons, the correlation increases with the aggregation duration. Summer generally exhibits the lowest correlation irrespective of the duration, due to the localized and isolated nature of convective

events that are likely to be missed by the rain gauge. The bias and variability scores are given in Appendix A.3.1. There is generally a tendency toward overestimation of the data (Figure A.11a) by the CPC for all seasons (median > 0), again the bias is more pronounced in summer compared to the other seasons. Lastly, the dispersion bias (Figure A.11b) is generally negative with a median of 5% for all seasons

ii) Criteria on extremes

Next, we evaluate the ability of CPC to correctly detect extreme precipitation as measured by the gauge. Extremes here are defined as the exceedances of 99.5% quantile over the whole sample (zeros included). We compute three criteria similar to Panziera et al. (2018). The first criterion measures the bias in extreme precipitation totals. The second criterion computes the probability of detecting extremes (POD), *i.e.*, the ability of CPC to classify events as extremes, given that they are also extremes according to the gauge. Lastly, we compute the false alarm ratio (FAR), which measures the rate at which CPC classifies events as extremes when they are not extremes according to the gauge. For a perfect agreement, bias should be equal to 0, POD should be equal to 1, and FAR should be equal to 0.

Figure VII.6b shows the seasonal POD scores. The median of the score ranges from 0.7 to 0.99, which means that 70% to 99% of the gauge extreme events are correctly classified as extremes by the CPC. Again, summer shows the lowest values compared to the other seasons. Surprisingly, the median score tends to decrease with duration. The same trend is observed in the case of FAR in Figure A.12b in Appendix A.3.1. The median of the bias in the extremes precipitation totals (Figure A.12a) is less than 5% for all cases. Summer in this case has the lowest bias but shows the most spread in the case of short durations.

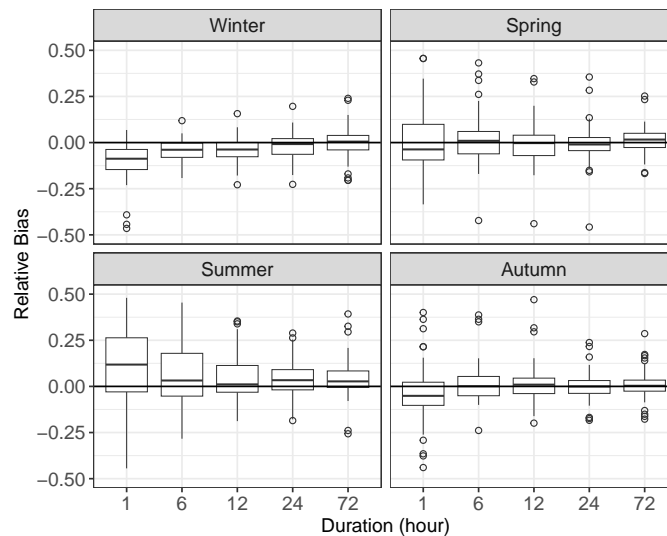


Figure VII.7: Boxplots of relative bias in a 20-yr return level estimate for the four seasons. Each boxplot contains 71 points, 1 point for each pair of rain gauge and the underlying CPC pixel.

VII.2.4.1.b Comparison of return level estimates In the final phase, we compare the 20-year return level estimates from the two datasets. We fit EGPD to each dataset and estimate the 20-year return level. Figure VII.7 shows the relative bias in the 20-year return level estimates. A positive bias indicates that the CPC estimates are higher than the gauge estimate. In general, the bias for durations greater than 6 hr is close to zero. For the 1 hr duration, however, there is a tendency to have lower estimates with CPC for all seasons, except summer which shows the opposite.

In conclusion, the result of the comparison shown in these sections is aimed at checking and validating the CPC data before using it in modeling the IDAF relationships. In spite of the no-

ticeable disagreements, there is generally, a good agreement between the two datasets, given the inherent uncertainties in both databases (gauge *versus* radar reanalysis). Although CPC is corrected using the rain gauge data, some differences still remain between the two, mainly due to the nugget effect in the variogram model, and the convection control scheme in summer (Sideris et al., 2014a). As emphasized in Section VII.2.2.2, it is beyond the scope of this study to develop a new gridded dataset for this topographically complex study area. CPC presents the only dataset at the sub-daily temporal resolution in the study area. It brings the required spatial information needed for modeling IDAF which cannot be obtained from rain gauges due to their limited spatial representativity. In the remainder of the article, only CPC is used to build the IDAF models.

VII.2.4.2 EGPD parameters as a function of Duration and Area

The purpose of this section is to show the complex relationship that exists between the EGPD parameters and duration D , and area A . Moreover, it aims to showcase that the EGPD-IDAF model is flexible enough to adequately capture this complexity.

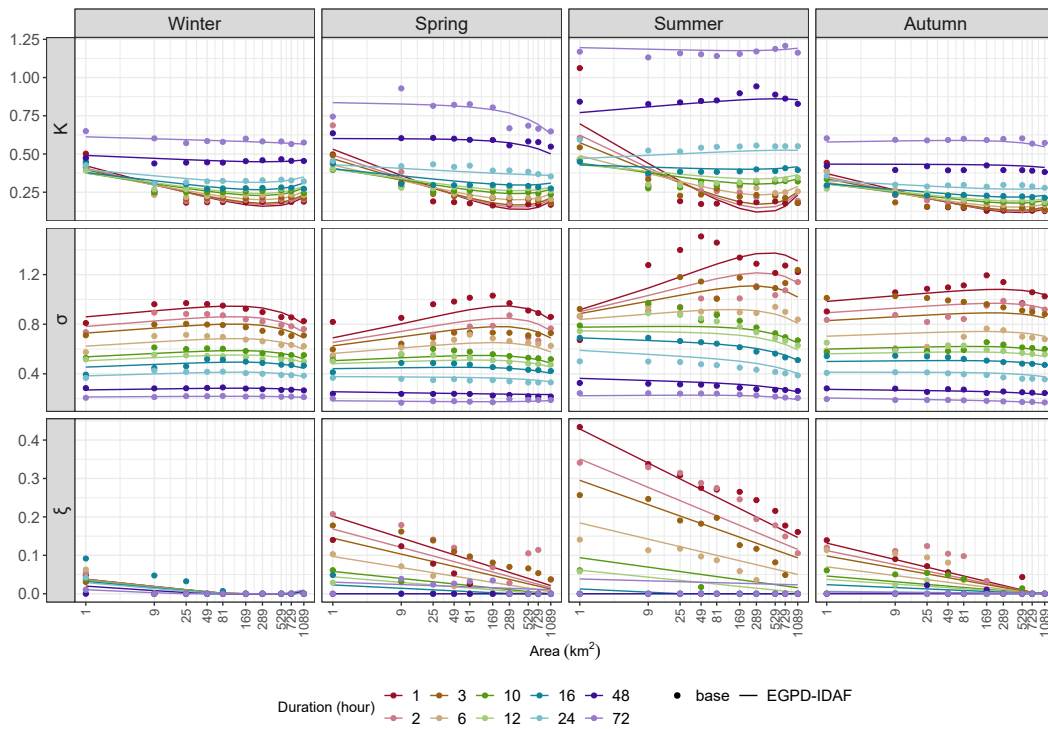


Figure VII.8: EGPD parameters as a function of duration D and area A at a pixel located in Adelboden (elevation of 1354 m, see Figure VII.4), for the four seasons (columns). The first row is for $\kappa(D,A)$, the second row is for $\sigma(D,A)$, and the last row is for $\xi(D,A)$. In each panel, the lines represent the modeled relationship using the EGPD-IDAF model, and the points show the parameter estimates using the base model. The lines and points are colored by duration.

As an illustration, we focus on a single pixel located at an elevation of 1,351 m in Adelboden, west of the Bernese Alps (see Figure VII.4). The estimated EGPD parameters as a function of D and A for the four seasons are shown in Figure VII.8. In each panel, the lines represent the modeled relationship using the EGPD-IDAF model, and the points show the parameter estimates using the base model. It can be observed from the figure that there is a clear season-dependent relationship of the parameters with D and A . We will focus on winter and summer since the other two seasons present behavior in between the two.

Starting with the top row, the flexibility parameter κ that controls the bulk and lower tail of the distribution shows a clear relationship with both D and A . For large A , it shows a positive monotonic relationship with D , while for small A , it shows a non-monotonic relationship, decreasing and then increasing with D . This non-monotonic relationship with D was also observed by Haruna et al. (2023b) while modeling IDF curves in the study area using rain gauge data. Next, looking at the middle row, the scale parameter σ decreases with an increase in D for all A in both seasons. It however shows a non-monotonic relationship with A , which also varies with D . Finally, in the bottom row, the upper tail shape parameter ξ shows a season-dependent relationship with D and A . The strongest relationship is observed in summer, where it decreases with both D and A . While it shows exponential tail ($\xi \approx 0$) for $D > 24$ hr irrespective of A , it shows a heavy tail ($\xi > 0.1$) for $D = 1$ hr even at $A = 1089$ km². In winter, however, it shows an exponential tail for all D and A .

We highlight here that the pattern of relationship observed at this pixel location is not general all over Switzerland, and our aim is just to illustrate the complexity of the relationship by focusing on this pixel. For instance, for some locations, σ can show a positive-monotonic relationship with A for all D . The shape parameter ξ can remain positive for all D and A in winter increases with D , or increases with A . This intricate relationship of the parameters with D and A underscores the difficulty and complexity of modeling relationships of areal precipitation in topographically complex locations, due to the regional heterogeneity of the rainfall process. Despite this, the EGPD-IDAF model adequately captures the relationships as seen in Figure VII.8, and so the proposed regression models in Equation VII.5 are flexible enough to capture the observed trends in the points corresponding to the base model estimates. In the next section, we will present the results of the goodness of fit of the EGPD-IDAF model at all the pixel locations in Switzerland.

VII.2.4.3 Goodness of fit of the EGPD-IDAF model

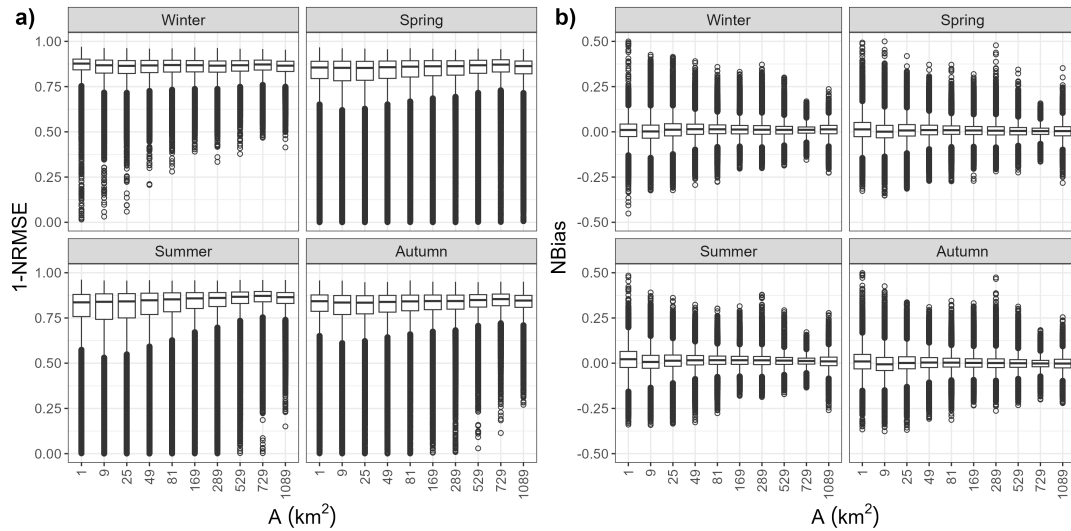


Figure VII.9: Goodness of fit of the EGPD-IDAF model. a) Boxplots of $(1 - \text{NRMSE})$ and b) NBias for the four seasons. Each boxplot contains 7,056 by 10 points, 1 point each for a pixel and a spatio-temporal scale (D, A).

We fitted the 20-parameter EGPD-IDAF at each pixel and for each of the four seasons and assessed the goodness of fit of the model using NRMSE and NBias (see Section VII.2.4.3). The normalization allows comparison of the score across intensities of different spatio-temporal scales (D, A). Figure VII.9 shows the results for the two criteria. In both figures, each of the four panels shows the score for a given season. The results are shown as a function of area (A), and so each

boxplot contains the results of 7,056 pixels for the 10 aggregation durations of a given A . Figure VII.9a shows the result for 1 - NRMSE and so the ideal score is 1. For all seasons, the median of the score is greater than 0.8 and the score gets better as A increases, possibly because as we aggregate the process over larger spatial domains, the variability decreases and the fit of the model gets better. While the score is relatively the same across seasons, summer shows slightly lower scores for smaller A (as seen from the width of the boxplot). These smaller scales in summer largely correspond to those experiencing more intense and skewed rainfall due the convective events. As such the shape parameter is heavy, and so the fit becomes more difficult. In Figure VII.9b, the median of the NBias remains close to zero which means that the model does not consistently overestimate or under-estimate the empirical quantiles. As with the other score, the variability around zero decreases as A increases.

These two scores show that the model can adequately reproduce the areal precipitation across durations in the study area. It shows good predictive performance as judged by the NRMSE, and doesn't show a systematic tendency to overestimate or underestimate the empirical values (NBias).

VII.2.4.4 Comparison of the EGPD-IDAF model with the base model

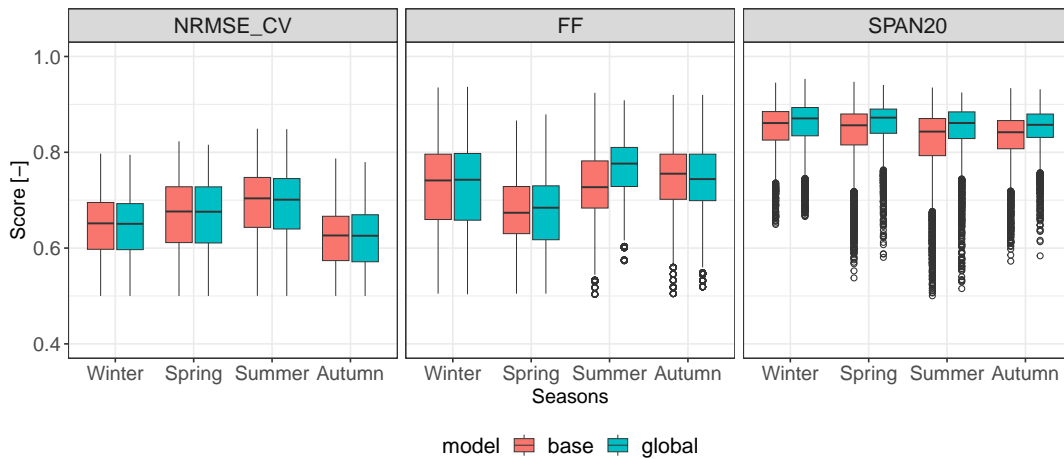


Figure VII.10: Boxplots of the cross-validation criteria for the four seasons. Each boxplot contains $2 \times 100 \times 40$ points for NRMSE_CV and FF (2 regional scores (i.e. $FF_{reg}^{(12)}$ and $FF_{reg}^{(21)}$) for each pair of (D, A) , and 40 resamplings). In the case of SPAN20, each boxplot contains 100×40 points (1 regional score for each pair of (D, A) and 40 resamplings).

VII.2.4.4.a Cross validation results The result of the split sampling cross-validation for the comparison of the EGPD-IDAF model (global model) and the base model is shown in Figure VII.10. This Figure shows the three cross-validation scores (NRMSE_CV, FF, and SPAN20), one panel for each criterion. As a reminder, the 20-parameter global model allows the linkage of the EGPD parameters with duration and area, the base model fits a separate EGPD model to each of the 100-time series of spatio-temporal scales (D, A) . The best model in each case has a score of 1. Starting with the first panel from the left, NRMSE_CV is nearly the same for both models, which means that both models have the same accuracy in predicting the whole non-zero precipitation. Next, the FF criterion also shows similar performance by the two models. A noticeable exception is in summer, where the global model shows better performance. Hence, according to this criterion, while the models have similar reliability in predicting the maximum value, the global model is slightly better in summer. Finally, SPAN20 shows that a better performance is obtained with the global model for all seasons compared to the base model. This means that the global model gives a more stable estimate of a 20-year return level when the calibration sample is changed.

In summary, both models have similar reliability in their predictive ability (NRMSE_CV and FF), however, the global model is more robust in 20-year return level estimations (SPAN20). The robustness of the global model can be explained by the fact that the model is trained with much more data (all the 100-time series are pooled in the parameter estimation), compared to the base model.

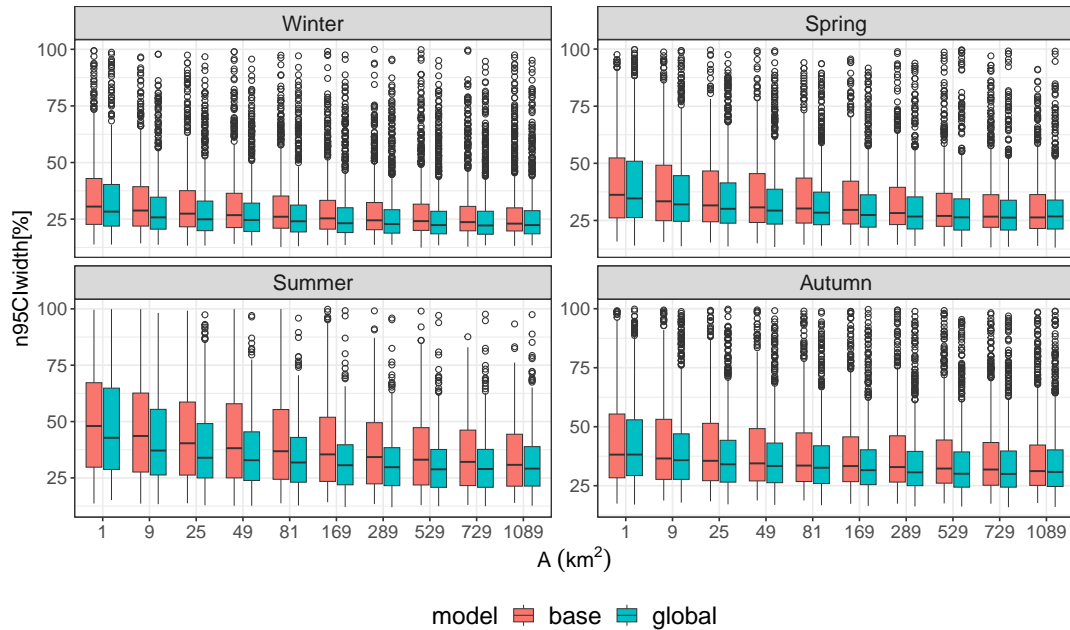


Figure VII.11: Boxplots of $n95CI_{width}(\%)$ for a 50-year return level estimate, using the base model and the global model. Each boxplot contains $7,056 \times 10$ points (7,056 pixels, 10 durations).

VII.2.4.4.b Uncertainty Since the two models have similar predictive performance, we also go a step further to compare the models in terms of their uncertainty estimates. While a good model should give correct predictions, the uncertainty of the prediction should not be too large. Figure VII.11 shows the $n95CI_{width}(\%)$ (Equation VII.8) of a 50-year return level estimate with both the global and base model. The smaller the score, the better the preciseness of the model (less uncertainty). Each panel in this Figure shows the result for a given season. The results are shown as a function of area (A), and so each boxplot contains the results of 7,056 pixels for the 10 aggregation durations of a given A . For all seasons, the global model has the smallest values of the $n95CI_{width}$ as seen from the median and width of the boxplots. The lower values of the global model mean less uncertainty compared to the base model, which can be explained by the fact that the global model is trained with more data, and this translates to less uncertainty (narrower confidence intervals). Two more comments can be made from Figure VII.11. First, for all seasons, the uncertainty decreases with A , which can be a result of the smoothing effect due to spatial averaging. Secondly, some inter-seasonal differences are noticeable, with summer (winter) having the highest (lowest) uncertainty. A possible explanation is that since more extremes are observed in summer (especially at sub-daily time scales), the uncertainty is expected to be larger. For a given return period, the magnitude of the return levels in summer at the small scale is larger compared to the other seasons, and so will the uncertainty.

To conclude, the results shown so far demonstrate that the modeled EGPD-IDAF can be used in the study area. It has adequate goodness of fit, is reliable and robust in prediction, and has relatively low uncertainty in estimation. With this validation, we will now proceed to showcase

examples of IDAF curves constructed from the EGPD-IDAF model at some pixel locations in the next section.

VII.2.4.5 IDAF curves

Figure VII.12 shows an application of the EGPD-IDAF model to build summer IDF and IAF curves at the pixel located in Adelboden. This pixel has been introduced in Section VII.2.4.2 and is shown in Figure VII.4. Starting with the top row (Figure VII.12a), IDF curves are shown in the case of four aggregation areas *i.e.* $A = 1, 25, 529,$ and $1,089 \text{ km}^2$ (1 column each). In each column, the colored lines represent the T -year return level estimate across duration for $T = 2, 10,$ and 20 years. The corresponding empirical estimates are shown by the colored points. It can be seen that the EGPD-IDAF correctly predicts the observation as they are within the 95% CI (shown by the colored envelopes). We also see that the uncertainty (indicated by the width of the bounds) is higher for shorter durations. Finally, irrespective of the spatial scale (A), the return levels decrease as the duration increases.

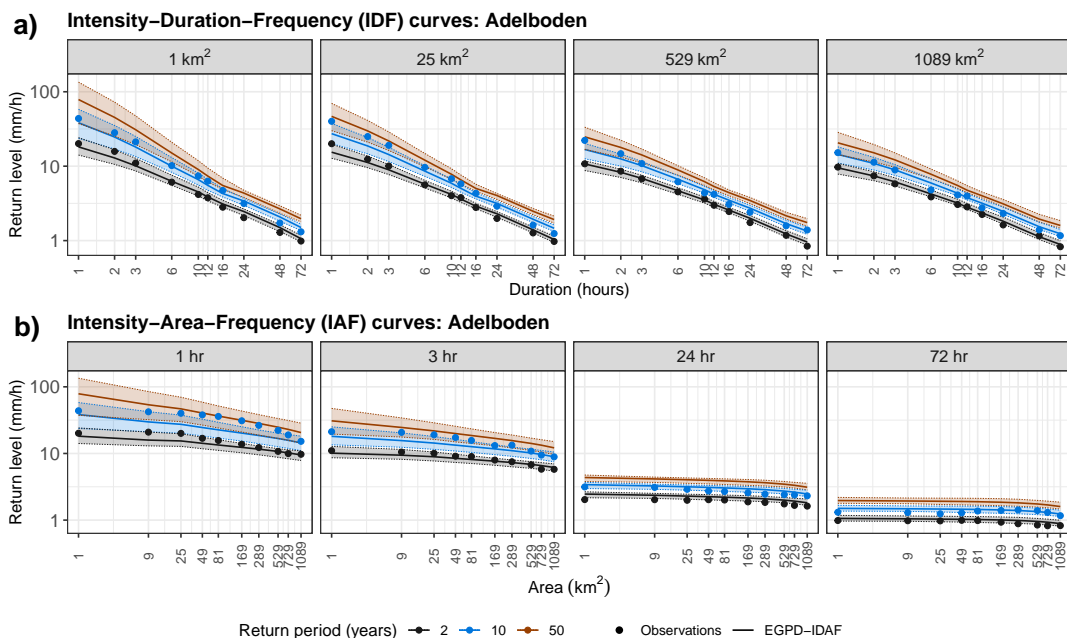


Figure VII.12: Application of the fitted EGPD-IDAF model at a pixel location in Adelboden for the summer season. The top row (a) shows some IDF curves for four spatial scales (one per column). The bottom row (b) shows the IAF curves for four temporal scales (one per column). The lines and the points show the modeled and empirical levels respectively, colored by their return periods. The colored envelopes are the 95% confidence intervals of the model estimates obtained by block bootstrap. The 50-year empirical values are not shown due to the short record length of the data

The second row (Figure VII.12b) shows the IAF curves for four temporal scales, $D = 1, 3, 24,$ and 72 hr . While the model shows an adequate performance for longer durations ($D \geq 24 \text{ hr}$), the fit is less good in the case of shorter durations, especially for higher return periods. Looking at the IAF curves for short durations, we see that the return levels tend to decrease with an increase in the spatial scale. For longer durations, however, the return levels have nearly the same magnitude (flat IAF curves) irrespective of the spatial scale. A possible explanation is that at short durations, the rainfall events are more localized (typical of convective events) and so the magnitude decreases due to spatial averaging. For longer durations, however, the rainfall is more homogeneous in space

(typical of frontal events), with no significant variations in rainfall intensity, leading to similar marginal distributions for the considered areal rainfall.

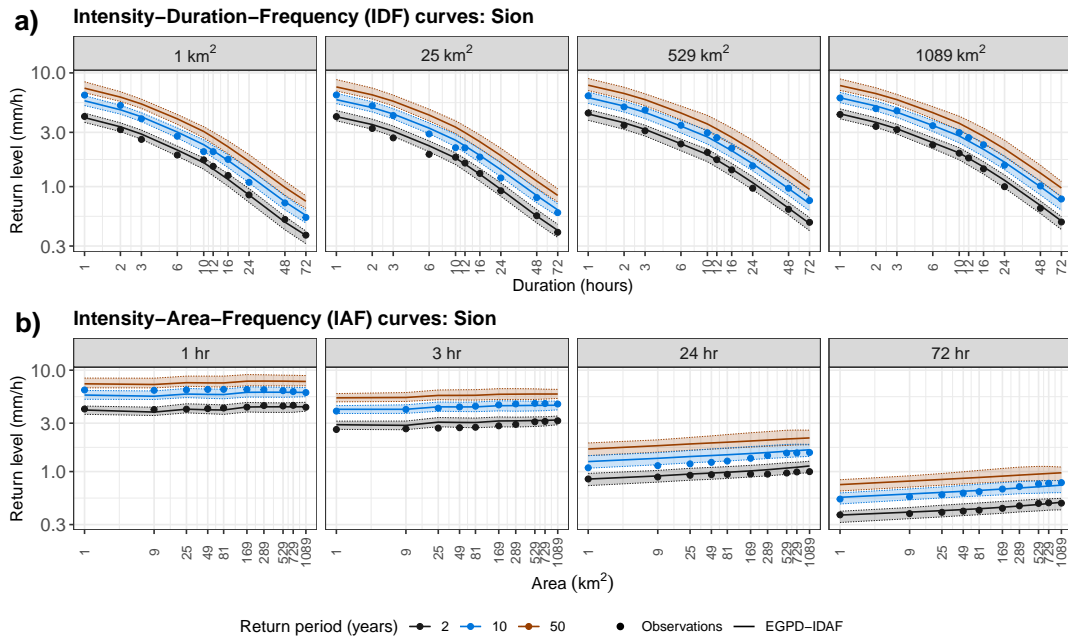


Figure VII.13: Same as Figure VII.12 but for autumn at a location in Sion in the Canton of Valais (see Figure VII.4).

To explore the regional and seasonal variability of the IDAF relationships, Figure VII.13 shows the autumnal IDF curves (top row) and IAF curves (bottom row) at a location in Sion, in the inner valleys, southwestern Switzerland. This location is at a relatively low elevation of 482 m and experiences low-intensity rainfall due to the shielding effect of the Alps on both sides. Remarkably, the IDF and IAF curves at this pixel location exhibit a distinctive behavior, diverging from the conventional trend of decreasing return levels with increasing spatial scales. The IAF curves (bottom row) highlight this feature. It can be seen that the IAF curves for 1 hr are nearly flat, and the IAF curves for $D \geq 24$ hr have positive slopes. A plausible explanation of this behavior is that rainfall, of short and long duration, is less intense at the pixel location compared to its neighborhood locations, which are at a higher altitude (see Figure 6 to 8 of Panziera et al. (2018)). As a consequence, more intense rainfall is observed as the rainfall is spatially aggregated around the pixel location. Figure VII.4 shows that a spatial window of 1,089 km², centered around the pixel (elevation of 480 m), extends well beyond the valley into the Bernese alpine slopes (elevation up to 2,400 m). This seasonal and regional variability highlights the complexity of modeling areal precipitation in the study area due to the complex topography.

VII.2.4.6 Areal rainfall hazard in Switzerland

In this last section, we use the EGPD-IDAF model to assess areal rainfall hazards in the study area. We investigate the 20-year return level for two spatio-temporal scales, specifically the scales ($D = 1$ hr, $A = 1$ km²) and ($D = 24$ hr, $A = 1,089$ km²). The corresponding maps of the seasonal 20-year return level are shown in Figure VII.14a and Figure VII.14b respectively. For the scale ($D = 1$ hr, $A = 1$ km²), we observe that the highest return levels occur during the summer months, while the lowest values are observed in winter. This can be attributed to the prevalence of convective rainfall during summer. We also see significant regional variability across all seasons, particularly during summer. The Ticino region in the south of the Alps, the Bernese Alps in the north, and

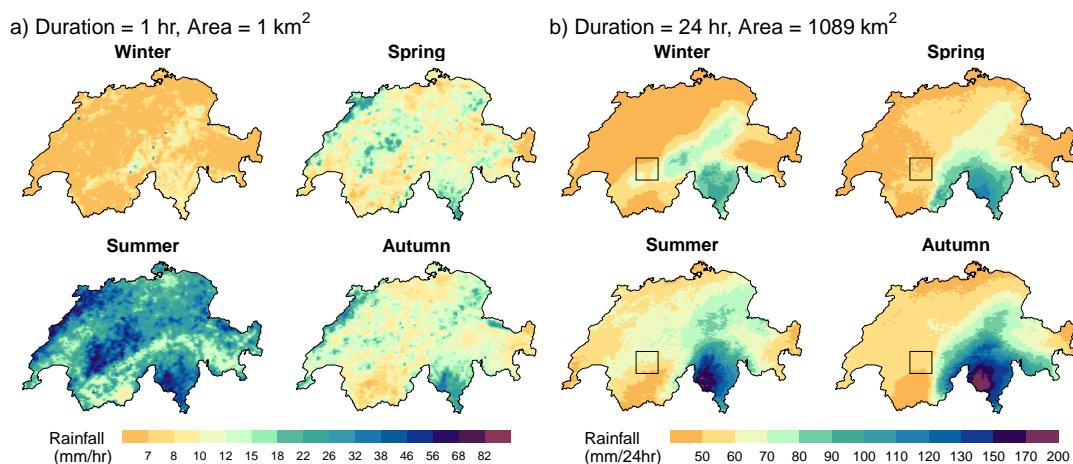


Figure VII.14: Map of seasonal 20-year return level obtained with the EGPD-IDAF model for the spatio-temporal scales a) ($D = 1$ hr, $A = 1$ km²) and b) ($D = 24$ hr, $A = 1,089$ km²). The black-colored square in b) shows exemplarily the maximum extent of the rectangular window used for data aggregation, i.e. 1089 km².

the Jura Mountains consistently exhibit the highest return levels. Conversely, the inner valleys in Valais and Grison, due to their location between mountains, depict the lowest values as they are shielded from both directions.

Moving to the scale ($D = 24$ hr, $A = 1,089$ km²), we see a shift in the seasonal and regional variability of the 20-year return level. The black colored square shows the spatial coverage of $A = 1,089$ km², centered around a pixel in Adelboden. The map in Figure VII.14b shows that the largest values are observed in Ticino, regardless of the season. The Ticino region consistently exhibits the highest levels of extreme precipitation in Switzerland. In the north of the Alps, the plateau displays lower levels compared to the pre-Alps (along the Glarus Alps). These results emphasize the influence of spatio-temporal scale on the seasonality and regional patterns of rainfall hazard in Switzerland. Smaller scales show a higher hazard during summer, while larger scales demonstrate a higher hazard during autumn, particularly in the Ticino region. It is important to note that the Ticino region consistently remains at a higher hazard of extreme precipitation, irrespective of the scale. Conversely, the inner valleys in Valais and Grison exhibit lower susceptibility to extreme precipitation events.

In conclusion, this result provides insights into the seasonal and regional patterns of rainfall hazards in Switzerland, highlighting the importance of considering spatio-temporal scales when assessing extreme precipitation hazards. It is important to note that while this assessment focuses on the hazard of extreme precipitation, it is essential to consider other factors such as exposure and vulnerabilities at specific locations to fully evaluate the overall risk.

VII.2.5 Conclusions

This paper focused on modeling the relationship of extreme precipitation across duration and area through Intensity-Duration-Frequency (IDAF) curves in Switzerland. We proposed a novel approach to model IDAF curves, by using all the non-zero (low, medium, and extremes) precipitation

data, instead of only the extremes. To build the IDAF curves, we used the EGPD as the parametric distribution for the precipitation intensities. The EGPD has the key advantage of adequately modeling the entire distribution of non-zero precipitation while being compliant with extreme value theory in both tails. We followed the footsteps of [Overeem et al. \(2010\)](#) to model the IDAF curves through a data-driven approach. This approach involves modeling the EGPD parameters as a function of area and duration, with the form of the relationship being empirically determined from the data. We used 17 years of data from the radar-reanalysis product, CombiPrecip (CPC) ([Sideris et al., 2014a](#)) to build the EGPD-IDAF model at each pixel location in the study area.

We used the model to assess areal rainfall hazard for some spatio-temporal scales in Switzerland. More than any region, the results showed that Ticino, located south of the Alps, is the most exposed to extreme precipitation for all the scales considered. Overall, the result provided insights into the seasonal and regional patterns of rainfall hazards in Switzerland, highlighting the importance of considering multiple spatio-temporal scales when assessing extreme precipitation hazards. We comment here that although we used the EGPD-IDAF model for areal rainfall hazard assessment, it can be used in several applications, such as the design of hydraulic structures ([Bertini et al., 2020](#)), or the determination of thresholds for use in early-warning systems ([Panziera et al., 2016](#)). Another potential application is that since the EGPD models the whole distribution of non-zero precipitation, not only the upper tail, it can be used as marginal distribution in stochastic weather generators for areal rainfall generation. The model will provide for a robust marginal distribution, given the quantity of data used to train it.

Additional results through a point-to-pixel comparison showed that both CPC and rain gauge data provided similar return level estimates, especially for longer durations. While this can be seen as a sort of validation of the CPC in extreme value analysis, the inferred return levels have to be interpreted with caution, mainly due to the limited length of the data. Notwithstanding, our work still provided a framework for further analysis in the presence of longer time series, *e.g.* from simulated series using weather generators.

Some perspectives for the present work involve using splines to model the relationships in the EGPD-IDAF model rather than regression forms. A possibility is to use Generalized Additive Models (GAMs) as implemented in [Youngman \(2020\)](#), or its extension that uses censored likelihood as used in [Haruna et al. \(2022\)](#). While splines can be promising due to their flexibility, a likely drawback is the enormous computational time required for inference of the model when using the EGPD, which uses all non-zero data. Our experience in [Haruna et al. \(2022\)](#) shows that the model requires significant time before convergence. The problem will be more complicated in this case where 100 time series is used and for more than 7,000 pixels. Another avenue for further research involves developing an Areal-Reduction-Factor (ARF)-based IDAF model and comparing it with the data-driven approach used in this model. While empirical (*e.g.* [Mineo et al., 2018](#)) and analytical (*e.g.* [De Michele et al., 2001](#)) ARF formulations exist in the literature (see [Svensson and Jones, 2010](#), for a review), our suggestion is to empirically develop an ARF model that works in the study area. This is because a previous research by [Mélèse et al. \(2019\)](#) showed that in mountainous regions, ARF formulations can exhibit unusual behavior (*e.g.* increasing value of ARF with an increase in Area, or $ARF > 1$). Finally, from an inference point of view, it will be interesting to explicitly account for dependence in the likelihood of the EGPD model (Equation VII.7). Beyond addressing the potential of likelihood misspecification, it will allow the possibility to estimate the conditional probability of observing an extreme event of a particular spatio-temporal scale, given that an extreme of another scale has been observed. This kind of information is invaluable in practice for risk management and planning.

VII.3 Summary

To sum up, the main take-home message from this chapter is summarized below:

Question 4: What is an appropriate model of Intensity-Duration-Area-Frequency (IDAF) relationships for the full range of non-zero precipitation intensities in a topographically complex area?

- We develop models of IDAF relationships using all non-zero precipitation intensities. The three-parameter EGPD was used as the model for the precipitation intensities.
- Following the work of [Overeem et al. \(2010\)](#), we developed a data-driven approach, where each parameter is modeled as a function of duration and area based on empirically determined relationships.
- We applied the model to construct IDF and IAF curves and to characterize and assess extreme areal precipitation hazards in Switzerland. Additionally, the marginal distribution from the IDAF model can be employed in simulation settings to feed stochastic weather generators for areal rainfall generation.
- The proposed parametric relationships in the IDAF model could be used in other regions with similar climates, we however expect them to be less complex in regions with lesser precipitation variability.
- As a limitation, the independence likelihood was used to estimate the model parameters, it will be interesting to explicitly model the dependence and to investigate the potential effect of the independence assumption.
- Another perspective is to develop an Areal-Reduction-Factor (ARF)-based IDAF model and compare it with the data-driven approach we considered.

VIII

Conclusions and perspectives

Chapter overview

This chapter presents the general conclusions and some relevant avenues for future research to address some of the limitations of the thesis.

Contents

VIII.1 Conclusion	144
VIII.2 Perspectives	145
VIII.2.1 Modeling in a non-stationarity context	145
VIII.2.2 Comparing EGPD-IDAF models with those from other distributions	147
VIII.2.3 Regional IDAF model	148
VIII.2.4 Accounting for dependence in the IDF and IDAF relationships	149

VIII.1 Conclusion

The goal of the thesis was to develop a model of IDAF relationships of non-zero precipitation applicable everywhere in Switzerland. To achieve this aim, four research questions were raised and answered accordingly in Chapter IV through Chapter VII. The main results are summarized below:

- The first question, "**What is the most suitable and parsimonious probability distribution to model the entire range of non-zero precipitation intensities in a topographically complex area?**" was addressed in Chapter IV. We considered three parametric families of the EGPD proposed by Naveau et al. (2016) and compared them using seasonal daily precipitation in the study area. The three-parameter model based on power law for the bulk of the distribution appeared to be the parsimonious model and was retained for subsequent application in the thesis. The inference procedure we adopted is based on maximum likelihood estimation due to its flexibility in accommodating covariates in regression settings. We however discovered that a station-specific threshold choice in the left-censored maximum likelihood estimation is needed to ensure adequate model performance, as a uniform threshold may result in poor performance at some stations. We then applied the chosen model to characterize extreme daily and hourly precipitation in Switzerland. In particular, the magnitude and spatial patterns of the seasonal return levels obtained are in agreement with those in the literature.
- Chapter V answered the second question, "**What is the most effective regionalization method to improve the robustness and reliability of daily precipitation estimates in a topographically complex area?**". To answer the question, we developed regional models based on three regionalization methods and we compared them based on their ability to improve the robustness and reliability of daily precipitation estimates using the EGPD. The results showed that all the regional models offered improvement compared to the local application of the EGPD. Of the regional models, the one based on ROI and the other based on GAM appeared to compete hand in hand in temporal cross-validation and were further compared in spatial cross-validation. The spatial cross-validation results showed that the ROI coupled with thin plate spline interpolation emerged as the best model for both gauged and ungauged locations. The maps of seasonal 100-year return levels of daily precipitation using the model revealed a clear seasonality and spatial pattern. Ticino, located south of the Alps, is subjected to the highest levels, especially in autumn, where up to 400 mm can be expected.
- The third question raised was "**What is the best model of Intensity-Duration-Frequency (IDF) relationships for the full range of non-zero precipitation intensities in a topographically complex area?**". We tackled this question in Chapter VI by building models of IDF relationships using all non-zero precipitation intensities with the three-parameter EGPD as the model for the precipitation intensities. We considered three approaches to modeling IDF curves, along with their extensions to account for potential breaks in the scaling relationship of precipitation and varying shape parameters with duration. We then built IDF models and compared them in calibration and cross-validation. The model based on the data-driven approach was shown to have the best performance and was applied to build catchment-level IDF curves for hydrological applications in Switzerland. The seasonal maps of the 100-year return level obtained with the IDF model revealed that the seasonal and regional patterns depend on the considered accumulation duration. For short durations (e.g. 1 hr), the highest levels are almost exclusively observed in summer, while for the daily scale, the highest levels are observed during autumn, particularly in Ticino. Additionally,

the marginal distributions derived from the model are intended to be used in a stochastic weather generator currently developed in the EXCH project.

- Finally, Chapter VII answered the last question which is "**What is an appropriate model of Intensity-Duration-Area-Frequency (IDAF) relationships for the full range of non-zero precipitation intensities in a topographically complex area?**". We developed models of IDAF relationships using all non-zero precipitation intensities with the three-parameter EGPD as the model for the precipitation intensities. We applied the model to construct IDF and IAF curves and to characterize and assess extreme areal precipitation hazards in Switzerland. Overall, the results provided insights into the seasonal and regional patterns of rainfall hazards in Switzerland, highlighting the importance of considering multiple spatio-temporal scales when assessing extreme precipitation hazards. The results showed that Ticino is the most exposed to extreme precipitation for all the scales considered.

VIII.2 Perspectives

Despite the results obtained, there are still avenues for further research to address the limitations of the thesis. These perspectives are outlined in the following sections:

VIII.2.1 Modeling in a non-stationarity context

Throughout the thesis, we assumed a stationary climate, mainly due to the requirement of the EXCH project. However, since Switzerland has experienced a warming trend almost double the global average (Scherrer et al., 2016b), it will be necessary to investigate the observed and future trends of extreme precipitation in the study area, and more generally in mountainous regions and to adapt the IDF/IDAF relationships accordingly.

VIII.2.1.1 Observed and future trends

Several studies were performed to investigate the observed changes in both mean and heavy precipitation in Switzerland. Such studies include the work of Widmann and Schär (1997) and Schmidli et al. (2002) for the 1901-1990 period, Schmidli and Frei (2005), for the 1901–2000 interval, Begert et al. (2005) for the duration covering 1864–2000, Masson and Frei (2016) for the 1901–2008 period. Their conclusions on the trend and significance varied depending on the dataset and the period considered. However, the most recent work for the 1901–2013 period by Scherrer et al. (2016a) reported a significant positive trend in the case of annual and winter precipitation, while no significant trends were observed in spring, summer, and autumn. Regarding heavy precipitation, Scherrer et al. (2016b) reported a positive trend in both the intensity and frequency of heavy precipitation throughout Switzerland during the 1901–2014 period. The trends were observed both at the annual and seasonal scales. All regions displayed a positive trend, with the inner Alps showing the lowest. In terms of frequency, the trends were larger in the northern part of the country compared to the south.

In addition to the observed trend above, based on climate change scenarios in Switzerland (CH2018, 2018), the positive trend in heavy precipitation is expected to continue throughout the century, especially in winter, where less snow is expected in favor of liquid precipitation. In addition, heavy precipitation at the hourly scale in summer is also expected to increase at a rate of 6–7% per degree warming (Clausius-Clapeyron scaling rate). A recent work by Vergara-Temprado et al. (2021) showed that even at sub-hourly scales, heavy precipitation in Switzerland will increase

according to the Clausius-Clapeyron scaling rate. The mean precipitation on the other hand is likely to increase in winter and decrease in summer. A weak increase in northern Switzerland is likely during spring, whereas no clear changes are expected in autumn.

The changes are not particular to Switzerland, but general to other regions in the European Alps. For instance, [Ménégoz et al. \(2020\)](#) also reported a significant increase in annual and winter precipitation and a reduction in summer precipitation in the northwestern Alps. Heavy seasonal precipitation is also observed to have increased at the annual and seasonal scale. In addition, most of the Alpine regions are projected to experience an increase in the intensity of extreme precipitation events in all seasons except summer. This increase will be most significant and widespread during the fall, especially in the northern Alpine region ([Gobiet et al., 2014](#)).

Based on the findings mentioned above, an interesting perspective is to model the trends using the EGPD and to account for the non-stationarity in building the IDF and IDAF relationships. This is elaborated more in the following sections:

VIII.2.1.2 Modeling trend with the EGPD

A starting point could be to use the EGPD to model the non-stationarity and to explore the physical mechanism associated with the trends. While GEV is widely used for modeling non-stationarity (e.g. [Cheng and AghaKouchak, 2014](#); [Blanchet et al., 2021](#); [Zhao et al., 2023](#)), the EGPD could be an interesting choice since it models both the bulk and the upper tail. This is particularly useful because the same distribution could be used to explore potential trends and their associated drivers in both the bulk and the extremes, which may differ. One possibility to model the non-stationarity using the EGPD is to model the parameters as parametric functions of time or other climatic covariates such as global mean temperature. This can be achieved by using the generalized linear model (GLM) for instance. The non-stationary EGPD model within the GLM framework can be written as:

$$\eta_{\gamma}(\gamma) = \beta_{0,\gamma} + \sum_{j=1}^{m_{\gamma}} \beta_{j,\gamma} x_{j,\gamma}, \quad (\text{VIII.1})$$

where $\gamma \in \{\kappa_t, \sigma_t, \xi_t, p_{0_t}\}$. The subscript t indicates that each parameter is time-dependent. p_0 is the probability of zeros required for return level estimate, $\eta_{\gamma}(\cdot)$ is the link function (e.g. log link for κ_t and σ_t , identity for ξ_t , and logit for p_{0_t}), $(\beta_{0,\gamma}, \dots, \beta_{m_{\gamma},\gamma})$ are model parameters to describe the trends of the parameter γ_t , $x_{j,\gamma}(j = 1, \dots, m_{\gamma})$ are the time-dependent covariates for the parameter γ introduced to explain the non-stationarity, and m_{γ} is the number of covariates for the parameter γ .

VIII.2.1.3 Modeling trend in IDF relationships

The observed and future trends in the intensity and frequency of heavy precipitation at daily and sub-daily time scales in the study area have broad implications for the application of the IDF curves developed under the stationary assumption because a structure designed under this assumption is likely to fail more frequently than intended. It is thus an interesting perspective to consider non-stationarity in the IDF relationships. To the best of our knowledge, no work has been done to account for non-stationarity in IDF curves in Switzerland. A review of approaches to account for non-stationarity in IDF relationships can be found in the literature (e.g. [Martel et al., 2021](#); [Yan et al., 2021](#)). In summary, the approaches can be described as follows:

- First, the IDF relationships developed under the stationarity assumption can be updated by a simple percentage increase (e.g. 30% in Belgium as reported in [Martel et al., 2021](#)), or

a percentage increase based on the Clausius-Clapeyron relationship. The latter is given in [Martel et al. \(2021\)](#) as:

$$i_{fut} = i_{ref} \times \left(\frac{100 + R_{sc}}{100} \right)^{\Delta T},$$

where i_{fut} and i_{ref} denote the projected future and reference period precipitation intensities respectively, R_{sc} is the precipitation scaling factor (%) based on the Clausius-Clapeyron relationship and ΔT is the projected change in local temperature (°C).

- The second way is to develop covariate-based IDF curves. In this case, the EGPD parameters in the IDF relationship are modeled as parametric functions of time or other physical covariates related to urbanization, temperature, global warming, etc. (see [Yan et al., 2021](#)). This can be achieved by using the generalized linear model (GLM) for instance (e.g. [Cheng and AghaKouchak, 2014](#); [Agilan and Umamahesh, 2016](#); [Ouarda et al., 2019](#)). In addition to the duration dependence of the EGPD parameters, they are also modeled as functions of time-dependent covariates.
- The third approach is to consider climate model-based future IDF curves (e.g. [Agilan and Umamahesh, 2016](#); [Zhao et al., 2022](#)). These future IDF curves are developed by performing frequency analysis directly on projected high-resolution precipitation data obtained from Regional Climate Models (RCMs) or Global Climate Models (GCMs). The future IDF curves as well as their related changes can then be evaluated.

In any case, each approach has its drawbacks. The percentage increase method adopts the same percentage irrespective of the duration and frequency of the precipitation. The established relationship between the model parameters and the explanatory variables in the covariate-based IDF curves is assumed to remain unchanged when extrapolating in the future. The climate model-based future IDF curves on the other hand involve precipitation simulations with various levels of uncertainties ranging from the emission scenario, model structure, internal climate variability, downscaling method, and post-processing of climate simulations.

VIII.2.1.4 Modeling trend in IDAF relationships

Several studies have shown that a warming climate is likely to result in a potential shift in the spatial organization and spatial dependence of storms (e.g. [Wasko et al., 2016](#); [Blanchet et al., 2018](#); [Matte et al., 2022](#)). This has the potential to increase the spatial concentration and the intensity of precipitation which could increase the severity of flooding. Modeling the trend in the IDAF relationships can help investigate potential shifts in the structure of precipitation, the spatial organization of storms, and the potential physical drivers. Similar to the case of IDF relationships, non-stationarity in IDAF models can be modeled by letting the EGPD parameters in the IDAF relationships depend on time or time-related covariates using the GLM framework. A circular aggregation window could be employed to better approximate storm shapes. A main difficulty however is that for our study, the observed gridded data at the hourly scale is limited in length to only a few years (≈ 17 years). It will be impossible, therefore, to reliably investigate any trend using the data set. Another possibility is to use *RhiresD* which covers a much longer period (≈ 60 years). However, this dataset is available at the daily scale so no insight will be obtained for the sub-daily time scale which is related to the intense convective events.

VIII.2.2 Comparing EGPD-IDAF models with those from other distributions

In this thesis, EGPD was used as the distribution of non-zero precipitation intensities to model the IDF and IDAF relationships. The motivation for using the EGPD was to make efficient use of information thereby reducing estimation uncertainty. This decision was also motivated by the need

to have a robust marginal distribution to be applied in the stochastic weather generator developed during the EXCH project. An avenue for further research is to make an objective comparison of the performance of the EGPD and other distributions, such as GEV or GPD, or the recently proposed meta-statistical extreme value (MEV) distribution (Marani and Ignaccolo, 2015) in modeling extreme precipitation, such as daily data.

The MEV distribution, in particular, and its variants (see Section III.1.2.3) have become increasingly popular in hydrological applications (e.g. Schellander et al., 2019; Gründemann et al., 2023) because it does not require the asymptotic assumption, and it uses more data compared to GEV and GPD. Recent comparison has been done to compare the MEV and the classical EVT distributions such as GEV and GPD and the former has been shown to yield more accurate estimates especially when the length of the available record is short (Marani and Ignaccolo, 2015) less uncertainty (Marani and Ignaccolo, 2015; Zorzetto et al., 2016; Marra et al., 2018, 2019), more robust (Schellander et al., 2019) and more spatially coherent estimates (Gründemann et al., 2023). However, except for the work of Milojevic et al. (2023) which reported better performance of EGPD over MEV at high return level estimates from short records, to the best of our knowledge, no work has been done to comprehensively compare EGPD and MEV.

A further step would be to build IDF and IDAF models using the GEV/GPD/MEV and to make a performance comparison with the EGPD. The evaluation framework and criteria used in this thesis (Section III.5) can be applied to compare these distributions. Additionally, the criterion proposed by Gründemann et al. (2023) to measure the heaviness of the tail of the distribution can also be employed. This will allow a thorough evaluation of the advantages and potential drawbacks of using the EGPD when the interest is only in the extremes. In any case, the EGPD has an edge over the GPD/GEV/MEV distributions since it models the entire non-zero precipitation (low, medium, and extremes), while the latter distributions only model the extremes.

VIII.2.3 Regional IDAF model

The IDAF model we developed in Chapter VII modeled the EGPD parameters as functions of duration (D) and area (A), and each model was separately fitted at each pixel in Switzerland. That is:

$$I(D,A) \sim \text{EGPD}[\kappa(D,A), \sigma(D,A), \xi(D,A)], \quad (\text{VIII.2})$$

where $\kappa(D,A) > 0$, $\sigma(D,A) > 0$ and $\xi(D,A) \geq 0$ are the three EGPD parameters for the duration D and area A .

The proposed model has a total of 20 parameters, with 7 each for σ and κ , and 6 for ξ (see Section VII.2.3.3). Since we fitted the model at each of the 7,000 pixels in Switzerland, we ended up with $(20 \times 7,000)$ parameters. This is obviously a large number of parameters, so one possible solution is to model the EGPD parameters as functions of geographical covariates such as longitude, latitude, and elevation. This way, Equation VIII.2 becomes:

$$I(D,A,\mathbf{x}) \sim \text{EGPD}[\kappa(D,A,\mathbf{x}), \sigma(D,A,\mathbf{x}), \xi(D,A,\mathbf{x})], \quad (\text{VIII.3})$$

where \mathbf{x} denotes a vector of geographical covariates, and each EGPD parameter $\alpha \in \{\kappa, \sigma, \xi\}$ depends on some form of \mathbf{x} . One possibility is to extend the relationships in Equation VII.5 to model each parameter as a smooth function of \mathbf{x} , that is:

$$\eta_\alpha(\alpha(D,A,\mathbf{x})) = g_\alpha(D,A) + f_\alpha(\mathbf{x},D,A), \quad (\text{VIII.4})$$

where α is a given parameter, $\eta_\alpha(\cdot)$ is a link function, $g_\alpha(D,A)$ is a function of D and A (see Equation VII.5) and $f_\alpha(\mathbf{x},D,A)$ is a smooth function, modeled with GAM, allowing for the potential interaction between the vector of covariates \mathbf{x} , D , and A .

The model can be estimated by pooling all the data of different durations and areas in the study area to estimate the parameters of the model. Using the independence likelihood will impose an assumption of independence between neighboring pixels in addition to the dependence between intensities of different durations and areas. This is however unlikely to be an issue since various authors (e.g. [Chavez-Demoulin and Davison, 2005](#); [Davison et al., 2012](#); [Zheng et al., 2015](#)) suggest that spatial dependencies can be neglected for the estimation of point-wise return levels. Anyways, explicitly modeling the dependence will largely inflate the number of parameters to be estimated. Even with the independence assumption, a potential difficulty is that since EGPD uses all non-zero precipitation data, the model will most likely require a lot of computational time, and a robust optimization algorithm will be necessary to estimate a large number of parameters. Nonetheless, Equation [VIII.3](#) means that only one model will be estimated for the entire study area, with far fewer parameters compared to fitting a separate model at each pixel.

VIII.2.4 Accounting for dependence in the IDF and IDAF relationships

Since our interest was on the marginal (univariate) return levels, the IDF and IDAF models developed in this study were based on the hypothesis of independence between intensities of different durations (e.g. time series of 1 hr *versus* time series of 2 hr) in the IDF models (Equation [VIII.5](#)) and independence between intensities of different durations and different areas (e.g. time series of 1 hr and 1 km², *versus* time series of 2 hr and 9 km²) in the IDAF models (Equation [VIII.6](#)):

$$l(\boldsymbol{\theta}) = \prod_{D \in \mathcal{D}} \prod_{j \in \mathcal{J}} f_D [i(D, j)], \quad (\text{VIII.5})$$

$$l(\boldsymbol{\theta}) = \prod_{A \in \mathcal{A}} \prod_{D \in \mathcal{D}} \prod_{j \in \mathcal{J}} f_{D,A} [i(D, A, j)], \quad (\text{VIII.6})$$

where f_* is the PDF of the EGPD and $\boldsymbol{\theta}$ is the vector of parameters to be estimated, \mathcal{D} and \mathcal{J} are respectively the considered sets of duration and area, and $i(D, A, j)$ is the non-zero precipitation intensities for duration D , area A and time step j .

This independence assumption neglects the correlation between intensities of different durations and different areas. Previous studies investigated this issue in the case of IDF curve modeling and the annual maximum series (GEV). For instance, [Nadarajah et al. \(1998\)](#) modeled the dependence between the different durations using multivariate extreme value distributions (MEVD), and [Tyalis and Langousis \(2019\)](#) followed suit by using max-stable processes. Later, [Jurado et al. \(2020\)](#) investigated the impact of accounting for this dependence in extremes using max-stable processes and showed that there is little gain in performance for return level estimates, in addition to the added complexity of using max-stable processes.

However, all these studies only focused on IDF models and marginal distributions based on extreme data. To our knowledge, modeling dependence with EGPD, which includes all the non-zero precipitation data has not been done yet. It would therefore be interesting to explicitly model the dependence in IDF and IDAF curve modeling for the case of the EGPD and to investigate the potential gain in the inferred return levels.

A starting point in the case of the IDF models (the same procedure applies to the IDAF model), could be to follow the footsteps of [Tyalis and Langousis \(2019\)](#) and [Jurado et al. \(2020\)](#) by modeling the dependence using the Brown-Resnick max-stable model ([Padoan et al., 2010](#)). According to this approach, the joint distribution function of any pair of unit-Frechét variables $Z(D_i)$, $Z(D_j)$ drawn from a Brown Resnick process is given by:

$$\mathbb{P}[Z(D_i) \leq z_i, Z(D_j) \leq z_j] = \exp \left[-\frac{1}{z_i} \Phi \left(\frac{\sqrt{\gamma(h)}}{2} + \frac{1}{\sqrt{\gamma(h)}} \log \frac{z_j}{z_i} \right) - \frac{1}{z_j} \Phi \left(\frac{\sqrt{\gamma(h)}}{2} + \frac{1}{\sqrt{\gamma(h)}} \log \frac{z_i}{z_j} \right) \right], \quad (\text{VIII.7})$$

where Φ denotes the standard normal distribution function, $\gamma^2(h) = 2 \left(\frac{h}{\rho} \right)^\tau$, with $0 < \tau \leq 2$ the smooth parameter and $\rho > 0$ the range parameter of the semivariogram. h is a measure of the distance between duration pairs (D_i, D_j) and following [Jurado et al. \(2020\)](#), $h = \log_2 \left(\frac{D_j}{D_i} \right)$ with $j > i$.

To proceed, the EGPD distribution of each non-zero precipitation data for duration D is transformed into unit Fréchet using

$$z(D) = -\frac{1}{\log\{F_D(i(D))\}}, \quad (\text{VIII.8})$$

where $F_D(\cdot)$ is the EGPD CDF. To link $i(D)$ to $z(D)$, the relationships in Equation [VI.12](#), to [VI.15](#) can be employed. The model thus has a total of 12 parameters, 10 for the IDF formulation in Section [VI.1.3.2.c](#) and two, α and ρ , for the variogram that models the asymptotic dependence. The estimation can be done using pairwise likelihood as shown in [Jurado et al. \(2020\)](#), after which the IDF curves can be obtained from Equation [III.14](#). In addition, the extremal coefficient that gives information on the dependence at extreme levels between D_i and D_j can be obtained from $\Psi(h) = 2\Phi[\gamma(h)/2]^{1/2}$. In general $1 \leq \Psi(h) \leq 2$, for complete dependence $\Psi(h) = 1$, and complete independence $\Psi(h) = 2$.

A potential difficulty in applying this approach is that the EGPD considers all the non-zero precipitation intensities. This means that the number of non-zero observations in the EGPD case decreases with duration, For instance, in a 20-year record, there are $(20 \times 365 \times 24 \times \text{the fraction of wet hours})$ of hourly non-zero observations, while there is $(20 \times 365 \times \text{the fraction of wet days})$ of 24 hr non-zero precipitation. This is unlike the annual maxima series where there are the same number of observations regardless of the duration. The open question is then how to account for this difference while applying Equation [VIII.7](#) in the case of the EGPD. Furthermore, in the case of IDF curves the distance between D_i and D_j is obtained from $h = \log_2 \left(\frac{D_j}{D_i} \right)$ with $j > i$, another difficulty in the case of IDAF model is in the definition of the distance between (D, A) and (D', A') .

In any case, by explicitly accounting for the dependence in the likelihood, we can avoid potential model misspecification that might occur when using Equations [VIII.5](#) and [VIII.6](#). Moreover, such an approach enables us to estimate the conditional probability of observing an extreme event of a particular spatio-temporal scale, given that an extreme of a different scale has already been observed. This information can be extremely valuable in practice, especially for issuing alerts.

A

Appendix

Contents

A.1	Appendix of Chapter III	152
A.1.1	GPD	152
A.2	Appendix of Chapter VI	153
A.2.1	EGPD parameters vs duration	153
A.2.2	Variables and their meaning	158
A.2.3	Catchment IDF curves	159
A.3	Appendix of Chapter VII	164
A.3.1	Comparison criteria	164

A.1 Appendix of Chapter III

A.1.1 GPD

A.1.1.1 Justification of the GPD

Now that we know that GPD is the model for the excesses, we will now show the justification as outlined in Coles (2001). First, the conditional distribution of the excess of u can be written as:

$$F_u(x) = \mathbb{P}(X - u \leq x \mid X > u) = \frac{F(x+u) - F(u)}{1 - F(u)}, \quad (\text{A.1})$$

for $0 \leq x < x_F - u$, x_F being the right endpoint of F . The conditional distribution of the tail, $1 - F_u(x)$ in turn can be written as:

$$\mathbb{P}(X - u > x \mid X > u) = \frac{1 - F(x+u)}{1 - F(u)} \text{ for } x > 0. \quad (\text{A.2})$$

Since F is not always known, we seek to find the approximation of F_u within the EVT framework. Let us recall from Section III.1.1.1 that if the random variable X has the CDF F , then for large n , the distribution of the block maxima of X , M_n can be expressed approximated by the GEV, that is:

$$F^n(x) \approx \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}, \quad (\text{A.3})$$

with $\mu \in \mathbb{R}$ the location, $\sigma > 0$ the scale and $\xi > 0$ the shape parameter.

By taking the logarithm of both sides, applying Taylor expansion on $\log F(x)$ for large values of x , Equation A.3 becomes:

$$1 - F(u) \approx \frac{1}{n} \left[1 + \xi \left(\frac{u - \mu}{\sigma} \right) \right]^{-1/\xi}. \quad (\text{A.4})$$

Similarly, for $x > 0$,

$$1 - F(u+x) \approx \frac{1}{n} \left[1 + \xi \left(\frac{u+x - \mu}{\sigma} \right) \right]^{-1/\xi}. \quad (\text{A.5})$$

By substituting Equation A.4 and A.5 in Equation A.2, simplifying, and writing $\mathbb{P}(X - u \leq x \mid X > u) = 1 - \mathbb{P}(X - u > x \mid X > u)$, the conditional distribution of excesses above a threshold in A.1 can be expressed as:

$$\mathbb{P}(X - u \leq x \mid X > u) = 1 - \left(1 + \frac{\xi x}{\sigma_u} \right)^{-1/\xi} \text{ for } \xi > 0,$$

where

$$\sigma_u = \sigma + \xi(u - \mu).$$

This shows that the GEV and the GPD are related, such that $\sigma_u = \sigma + \xi(u - \mu)$ with both distributions having the same shape parameter ξ . μ and σ are respectively the location and scale parameters of the GEV.

A.1.1.2 GPD Return Level Estimation

We now proceed to the important question of return level estimation. To do this, we need to recall from Equation III.3, that the CDF of the excesses $X - u$, conditioned on $X > u$ is approximated by the GPD. By substituting $y = x + u$, and recalling that the nonexceedance probability $\mathbb{P}(X - u > x|X > u) = 1 - \mathbb{P}(X - u \leq x|X > u)$, the equation becomes:

$$\mathbb{P}(X > y|X > u) = \begin{cases} \left[1 + \xi \frac{(y-u)}{\sigma_u}\right]_+^{-1/\xi} & \text{if } \xi \neq 0 \\ \exp\left[-\frac{(y-u)}{\sigma_u}\right] & \text{if } \xi = 0 \end{cases}. \quad (\text{A.6})$$

The left hand side of the equation, $\mathbb{P}(X > y|X > u)$ can be expressed as $\mathbb{P}(X > y|X > u) = \frac{\mathbb{P}(X > y \cap X > u)}{\mathbb{P}(X > u)} = \frac{\mathbb{P}(X > y)}{\mathbb{P}(X > u)}$. Substituting this in Equation A.6, we have

$$\mathbb{P}(X > y) = \begin{cases} \mathbb{P}(X > u) \left[\left(1 + \xi \frac{y-u}{\sigma_u}\right)_+^{-1/\xi} \right] & \text{if } \xi \neq 0 \\ \mathbb{P}(X > u) \left\{ \exp\left[-\frac{(y-u)}{\sigma_u}\right] \right\} & \text{if } \xi = 0 \end{cases}. \quad (\text{A.7})$$

By putting $\mathbb{P}(X > y) = \frac{1}{m}$, the level that is exceeded on average once every m observations is obtained by writing the equation in terms of y_m , that is

$$y_m = \begin{cases} u + \frac{\sigma_u}{\xi} \left[(m\zeta_u)^\xi - 1 \right] & \text{if } \xi \neq 0, \\ u + \sigma_u \log(m\zeta_u) & \text{if } \xi = 0 \end{cases}, \quad (\text{A.8})$$

with $\zeta_u = \mathbb{P}(X > u)$.

Since our target is to associate y to a given return period T in years, we need to substitute $m = T \times n_y$. This means that there are T years, with each year having n_y number of observations. Accordingly, The T -year return level using the GPD is obtained from:

$$y_T = \begin{cases} u + \frac{\sigma_u}{\xi} \left[(Tn_y\zeta_u)^\xi - 1 \right] & \text{if } \xi \neq 0 \\ u + \sigma_u \log(Tn_y\zeta_u), & \text{if } \xi = 0 \end{cases}. \quad (\text{A.9})$$

$\zeta_u = \mathbb{P}(X > u)$ is estimated from the sample proportion of the points exceeding u , *i.e.* $\frac{k}{n}$, with k the number of exceedances of u and n the total number of observations.

A.2 Appendix of Chapter VI

A.2.1 EGPD parameters vs duration

A.2.1.1 Parametric models for the EGPD shape parameter vs duration

- constant:

$$\xi_d = \frac{1}{n} \sum_{j=1}^n \xi_{d,n} \quad (\text{A.10})$$

- linear:

$$\xi_d = a_\xi + b_\xi d \quad (\text{A.11})$$

- log-linear:

$$\log(\xi_d) = a_\xi + b_\xi d \quad (\text{A.12})$$

- linear - log:

$$\xi_d = a_\xi + b_\xi \log(d) \quad (\text{A.13})$$

- log-log:

$$\log(\xi_d) = a_\xi + b_\xi \log(d) \quad (\text{A.14})$$

A.2.1.2 Plots of the EGPD parameters vs duration

In this section, we explore the relationship between the three EGPD parameters with respect to duration. Two stations located at Zurich (KLO) and Robbia in Graubünden (ROB) are given as examples. For each season, the fitted parameter from the base model is plotted against the duration d .

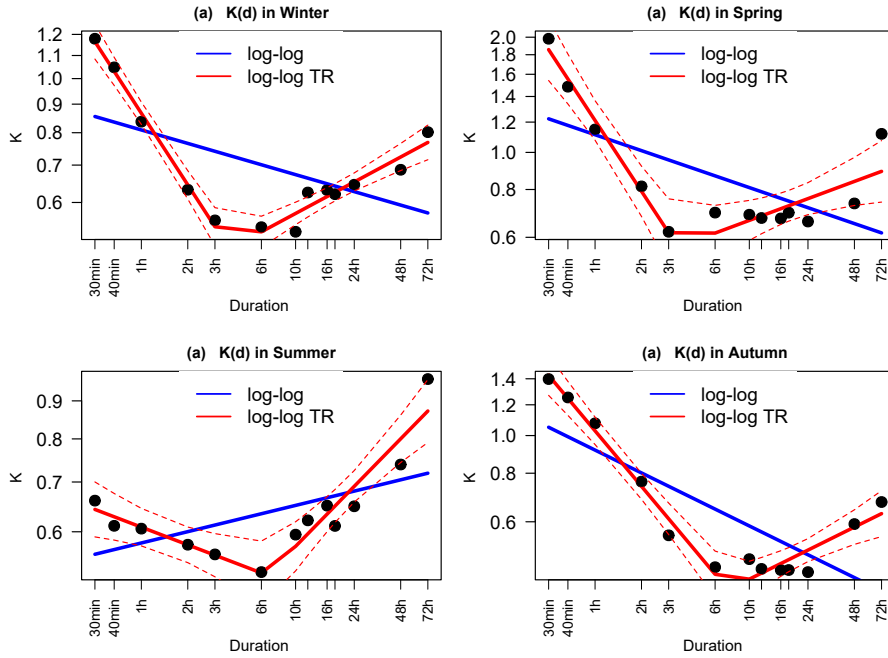


Figure A.1: Relationship between the EGPD κ parameter and duration d for a station in **Zurich** (KLO). To obtain this, we fitted EGPD to the data of each of the 13 durations separately by maximum likelihood. The black points are the fitted κ . The broken lines in red represent the 95% confidence interval for the fitted model (log-log TR) see Eqn. VI.11.

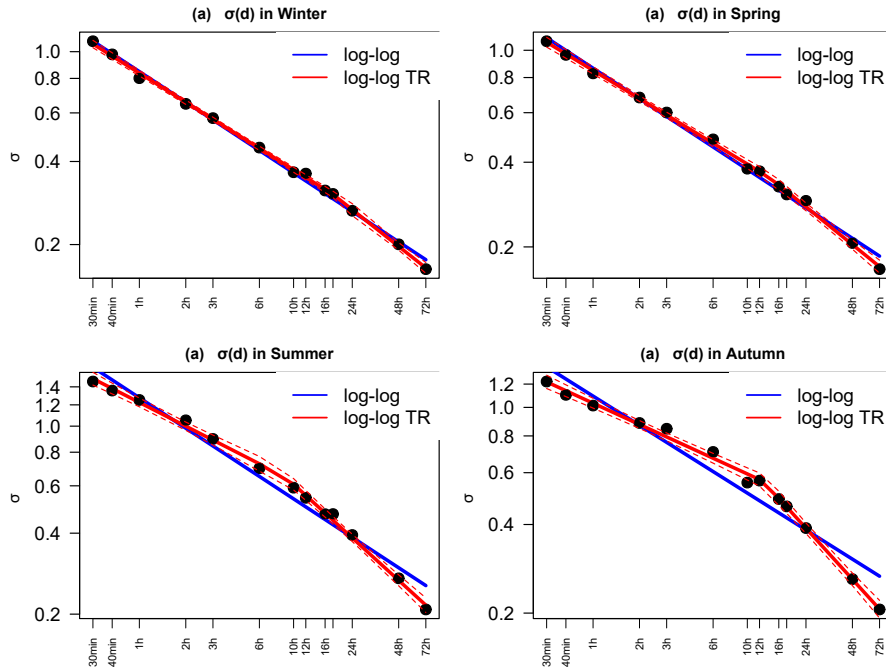


Figure A.2: Same as in Figure A.1, but for the EGD scale parameter σ .

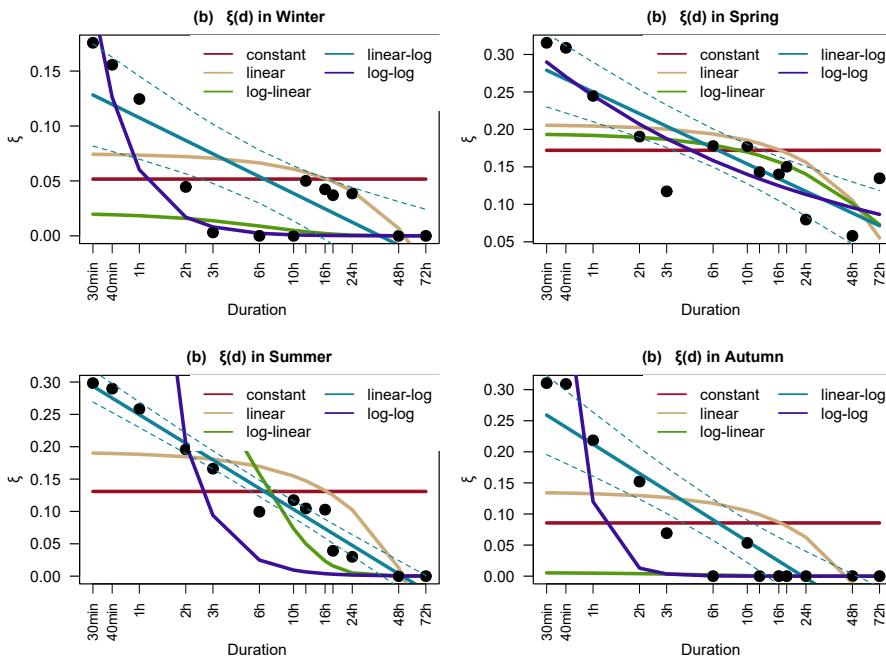


Figure A.3: Same as in Figure A.1, but for the EGD shape parameter ξ . The equations for the fitted models are given in A.2.1. The broken lines represent the 95% confidence interval for the linear-log model.

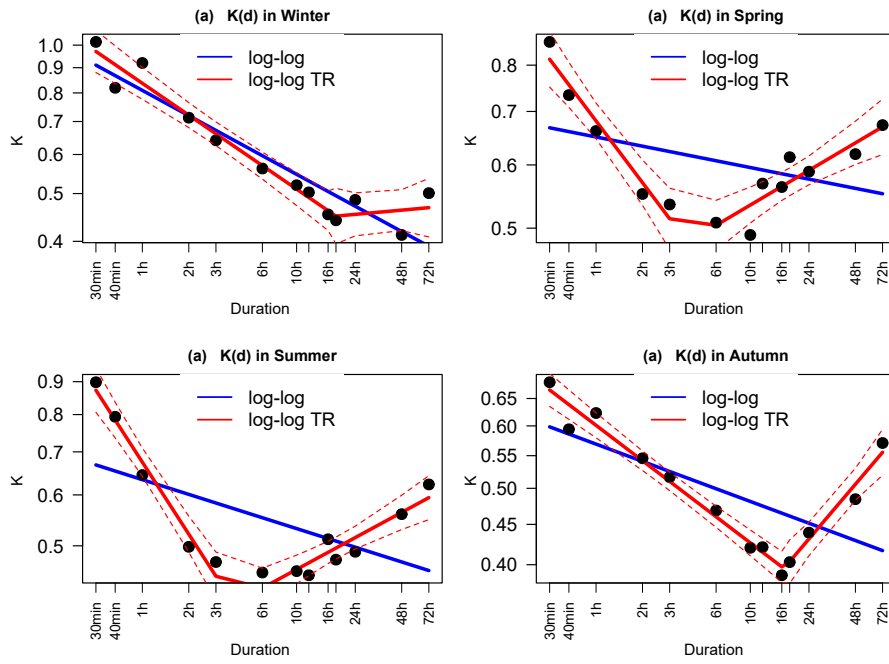


Figure A.4: Same as in Figure A.1, but for a station at *Robbia in Graubünder (ROB)*

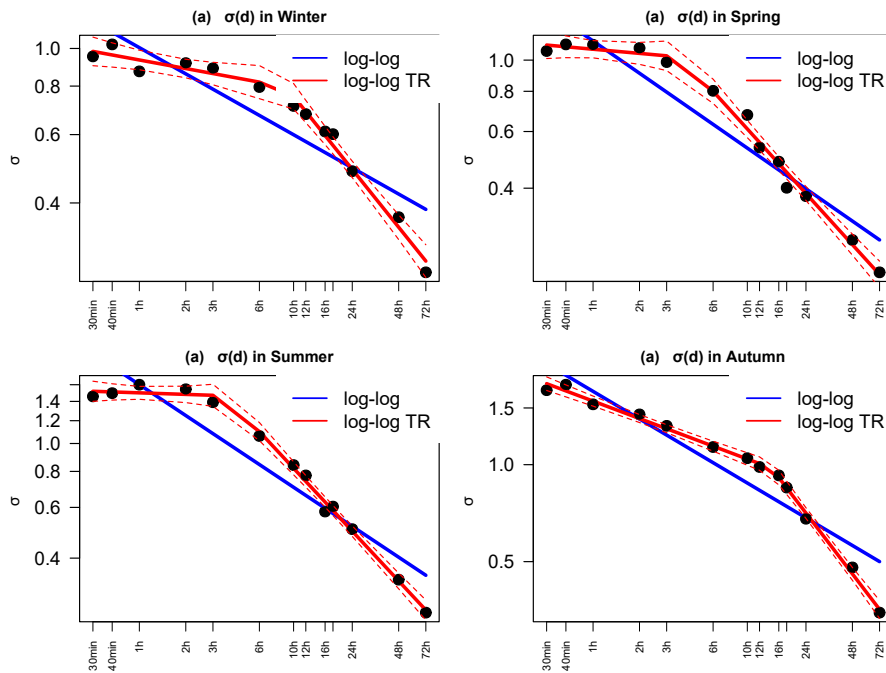


Figure A.5: Same as in Figure A.4, but for the EGPD scale parameter σ .

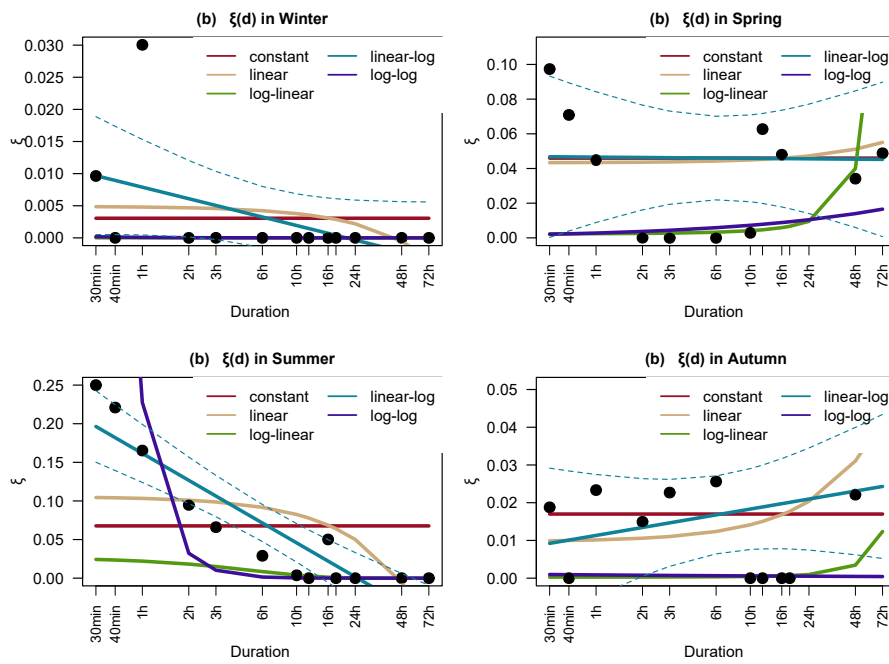


Figure A.6: Same as in Figure A.4, but for the EGPD shape parameter ξ . The equations for the fitted models are given in A.2.1. The broken lines represent the 95% confidence interval for the linear-log model.

A.2.2 Variables and their meaning

Table A.1: List of variables and their meaning

	Variable	Meaning	Range	Unit
1	H	Power-law scaling exponent	$0 < H < 1$	-
2	θ	Duration offset	$\theta > 0$	hour
3	d_0	Reference duration	$d_0 = 1$ in Simple scaling $d_0 = 1 - \theta$ in General IDF formulation	hour
4	κ_{d_0}	EGPD kappa parameter for d_0	$\kappa_{d_0} > 0$	-
5	σ_{d_0}	EGPD scale parameter for d_0	$\sigma_{d_0} > 0$	mm/hr
6	ξ_{d_0}	EGPD shape parameter for d_0	$\xi_{d_0} \geq 0$	-
7	K, K_*	Duration of scaling break	$K > 0, K_* > 0$	hr
8	H_1, H_2	1st and 2nd slopes in two-scaling regimes	$0 < H_1 < 1, 0 < H_2 < 1$	-

A.2.3 Catchment IDF curves

This section presents the catchment level IDF curves modeled in section [VI.2](#).

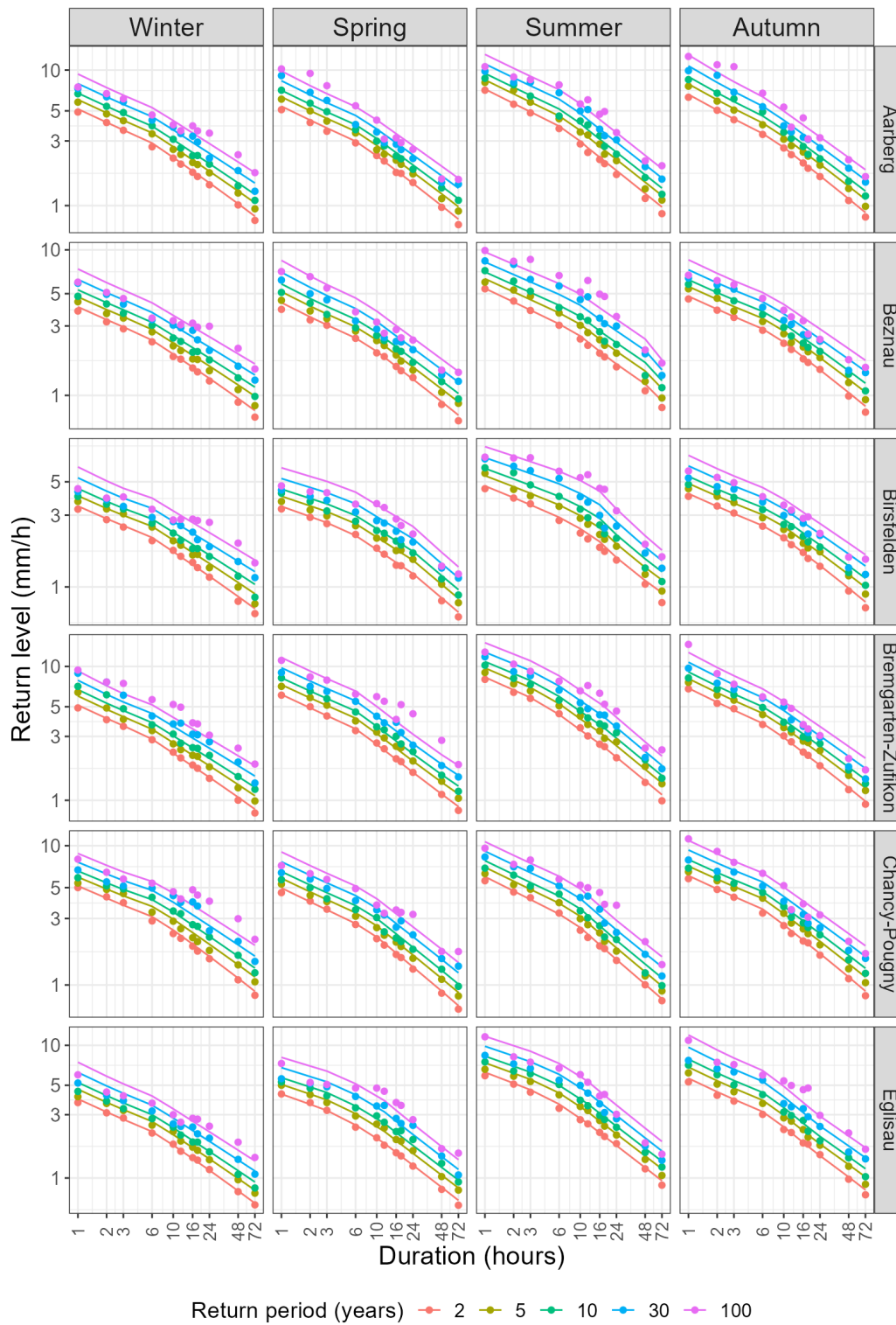


Figure A.7: Seasonal IDF curves for 6 out of the 24 large catchments. The lines are the model return levels, while the points are the corresponding empirical return levels. Both are colored according to the return period for $T = 2, 5, 10, 30,$ and 100 years.

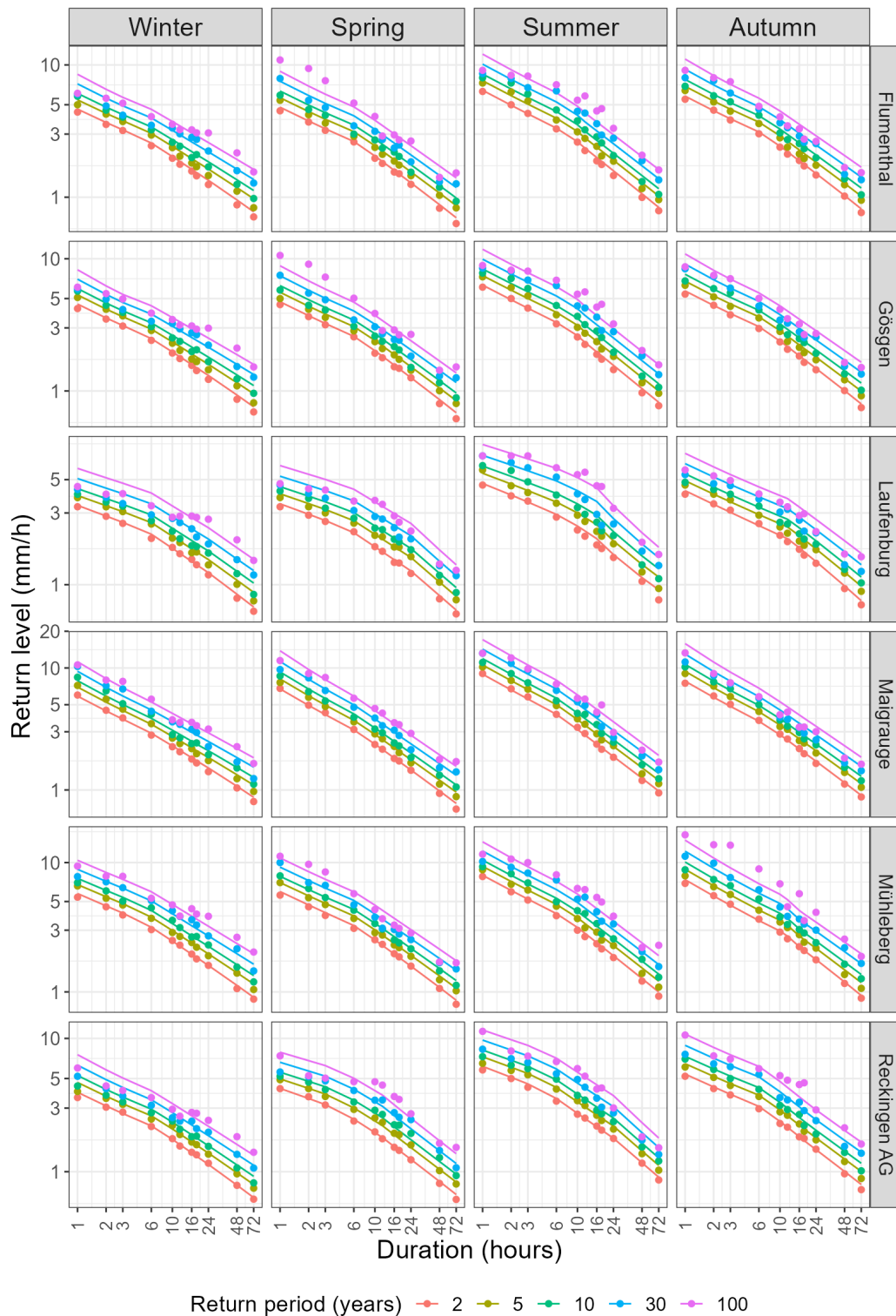


Figure A.8: Seasonal IDF curves for 6 out of the 24 large catchments. The lines are the model return levels, while the points are the corresponding empirical return levels. Both are colored according to the return period for $T = 2, 5, 10, 30,$ and 100 years.

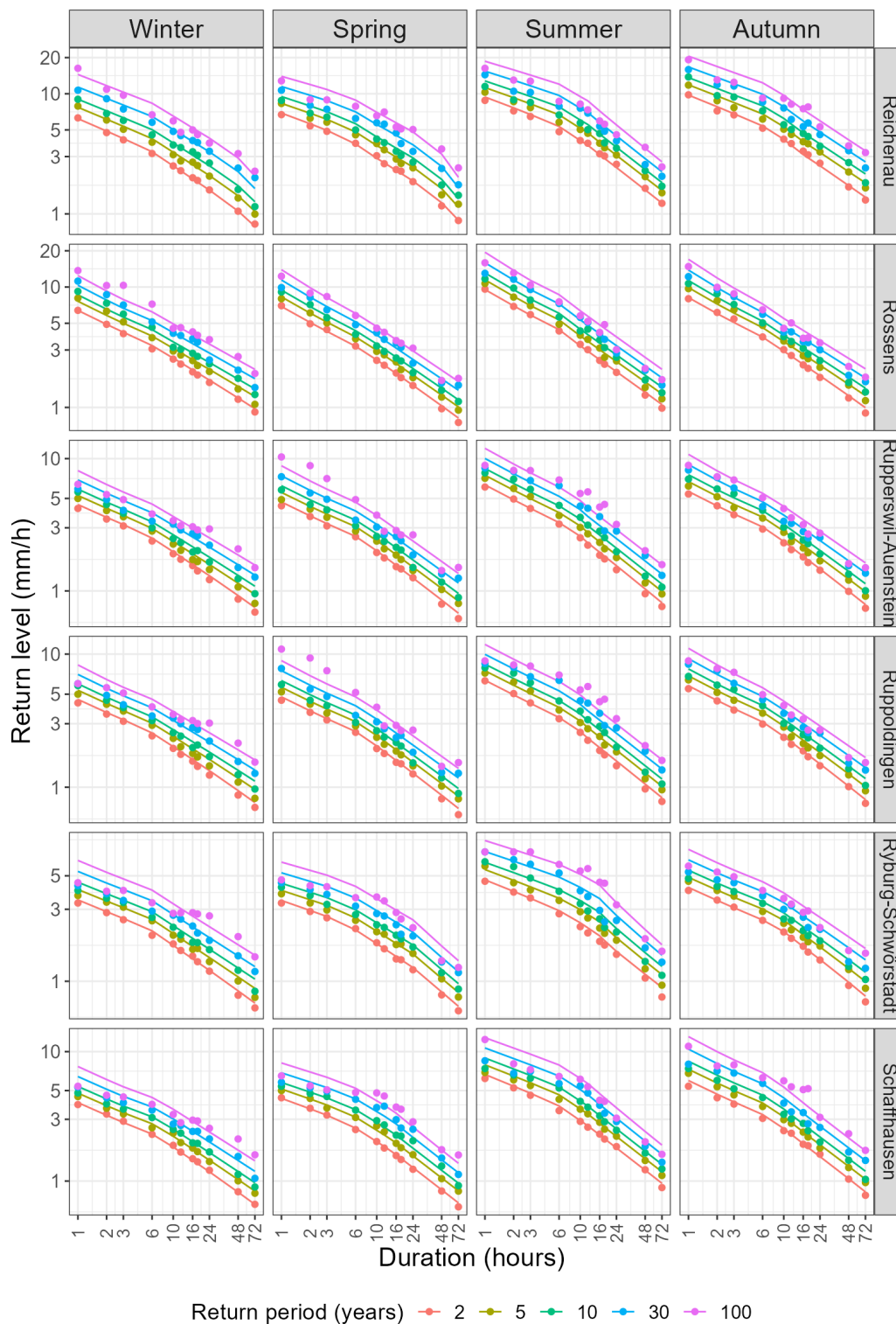


Figure A.9: Seasonal IDF curves for 6 out of the 24 large catchments. The lines are the model return levels, while the points are the corresponding empirical return levels. Both are colored according to the return period for $T = 2, 5, 10, 30,$ and 100 years.

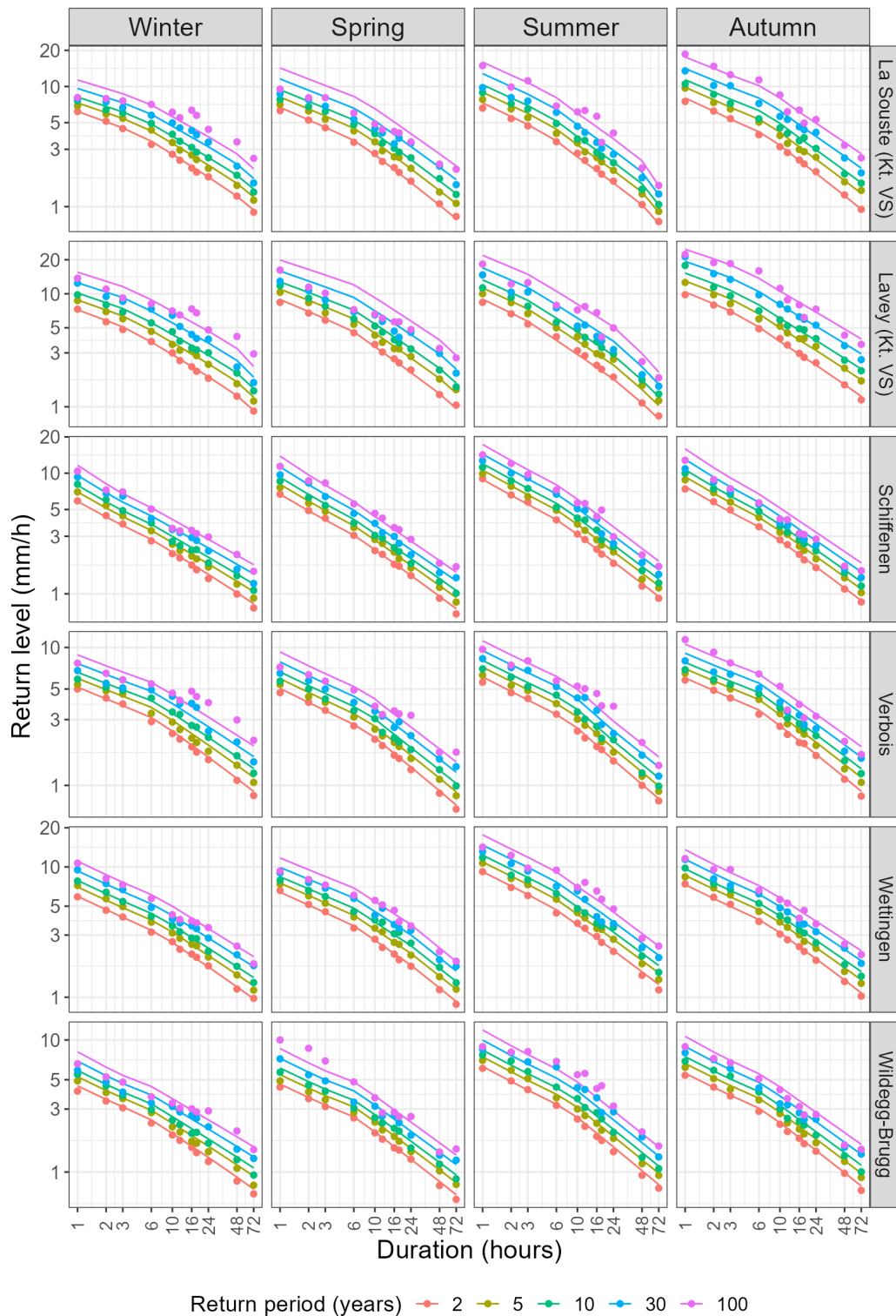


Figure A.10: Seasonal IDF curves for 6 out of the 24 large catchments. The lines are the model return levels, while the points are the corresponding empirical return levels. Both are colored according to the return period for $T = 2, 5, 10, 30,$ and 100 years.

A.3 Appendix of Chapter VII

A.3.1 Comparison criteria

The Kling-Gupta Efficiency (KGE) (Kling et al., 2012) is computed from:

$$KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2}, \quad (A.15)$$

where $r = \frac{\text{Cov}(i_{CPC}, i_{Gauge})}{\sigma_{CPC}^2 \sigma_{Gauge}^2}$ is the Pearson correlation coefficient between the CPC data (i_{CPC}) and the station data (i_{Gauge}), Cov is the co-variance between the two time series and σ denotes the standard deviation. $\beta = \frac{\mu_{CPC}}{\mu_{Gauge}}$, evaluates the bias between the two time series, with μ being the mean. $\gamma = \frac{CV_{CPC}}{CV_{Gauge}}$ is the variability ratio, that is the ratio between the coefficient of variations of the two time series.

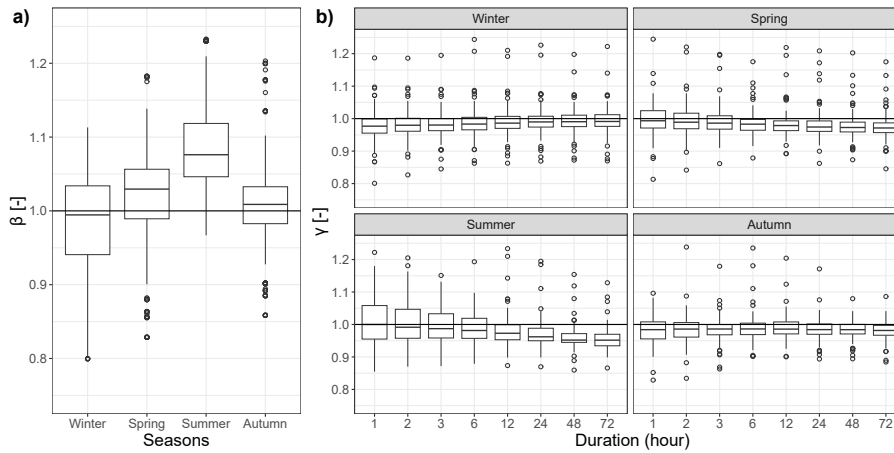


Figure A.11: a): Boxplots of bias (β) for the four seasons. b): Boxplots of variability ratio (γ) for the four seasons. Each boxplot contains 71 points, 1 point for each pair of gauge and the underlying CPC pixel.

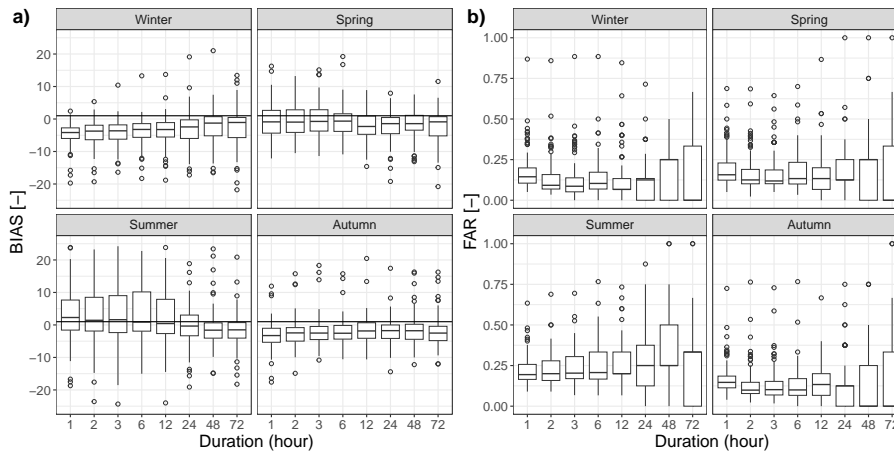


Figure A.12: a): Boxplots of the bias in extreme precipitation totals for the four seasons. b): Boxplots of the false alarm ratio (FAR) for the four seasons. Each boxplot contains 71 points, 1 point for each pair of gauges and the underlying CPC pixel.

Bibliography

- Agbazo, M. N., Koto N’Gobi, G., Kounouhewa, B., Alamou, E., Afouda, A., and Akpo, A. (2016). Estimation of IDF Curves of Extreme Rainfall by Simple Scaling in Northern Oueme Valley, Benin Republic (West Africa). *Earth Sciences Research Journal*, 20(1):1–7.
- Agilan, V. and Umamahesh, N. (2016). Is the covariate based non-stationary rainfall IDF curve capable of encompassing future rainfall changes? *Journal of Hydrology*, 541:1441–1455.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723. Conference Name: IEEE Transactions on Automatic Control.
- Allen, R. J. and DeGaetano, A. T. (2005). Considerations for the use of radar-derived precipitation estimates in determining return intervals for extreme areal precipitation amounts. *Journal of Hydrology*, 315(1-4):203–219.
- Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *The annals of mathematical statistics*, pages 193–212.
- Andres, N., Lieberherr, G., Sideris, I. V., Jordan, F., and Zappa, M. (2016). From calibration to real-time operations: an assessment of three precipitation benchmarks for a Swiss river system. *Meteorological Applications*, 23(3):448–461. _eprint: <https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/met.1569>.
- Andres, N., Steeb, N., Badoux, A., and Hegg, C. E. (2021). Grundlagen Extremhochwasser Aare: Hauptbericht Projekt EXAR, Methodik und Resultate . Technical report, WSL Berichte 104, WSL, Birmensdorf. <https://www.wsl.ch/de/projekte/exar-1.html> (last access: 23 October 2023).
- Bacchi, B. and Ranzi, R. (1996). On the derivation of the areal reduction factor of storms. *Atmospheric Research*, 42(1-4):123–135.
- Balkema, A. A. and De Haan, L. (1974). Residual life time at great age. *The Annals of probability*, 2(5):792–804.
- Begert, M., Schlegel, T., and Kirchhofer, W. (2005). Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000. *International Journal of Climatology*, 25(1):65–80. _eprint: <https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/joc.1118>.
- Behrens, C. N., Lopes, H. F., and Gamerman, D. (2004). Bayesian analysis of extreme events with threshold estimation. *Statistical modelling*, 4(3):227–244.
- Ben-Zvi, A. (2009). Rainfall intensity–duration–frequency relationships derived from large partial duration series. *Journal of Hydrology*, 367(1-2):104–114.
- Bernard, M. M. (1932). Formulas for rainfall intensities of long duration. *Transactions of the American Society of Civil Engineers*, 96(1):592–606. Publisher: American Society of Civil Engineers.

References

- Bertini, C., Buonora, L., Ridolfi, E., Russo, F., and Napolitano, F. (2020). On the Use of Satellite Rainfall Data to Design a Dam in an Ungauged Site. *Water*, 12(11):3028. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.
- Blanchet, J., Aly, C., Vischel, T., Panthou, G., Sané, Y., and Diop Kane, M. (2018). Trend in the Co-occurrence of extreme daily rainfall in west africa since 1950. *Journal of Geophysical Research: Atmospheres*, 123(3):1536–1551.
- Blanchet, J., Blanc, A., and Creutin, J.-D. (2021). Explaining recent trends in extreme precipitation in the Southwestern Alps by changes in atmospheric influences. *Weather and Climate Extremes*, 33:100356.
- Blanchet, J., Ceresetti, D., Molinié, G., and Creutin, J.-D. (2016). A regional GEV scale-invariant framework for Intensity–Duration–Frequency analysis. *Journal of Hydrology*, 540:82–95.
- Blanchet, J. and Lehning, M. (2010). Mapping snow depth return levels: smooth spatial modeling versus station interpolation. *Hydrology and Earth System Sciences*, 14(12):2527–2544.
- Blanchet, J. and Mélése, V. (2020). A Bayesian Framework for the Multiscale Assessment of Storm Severity and Related Uncertainties. *Journal of Hydrometeorology*, 21(1):109–122. Publisher: American Meteorological Society Section: Journal of Hydrometeorology.
- Blanchet, J., Paquet, E., Vaittinada Ayar, P., and Penot, D. (2019). Mapping rainfall hazard based on rain gauge data: an objective cross-validation framework for model selection. *Hydrology and Earth System Sciences*, 23(2):829–849.
- Blanchet, J., Touati, J., Lawrence, D., Garavaglia, F., and Paquet, E. (2015). Evaluation of a compound distribution based on weather pattern subsampling for extreme rainfall in Norway. *Nat. Hazards Earth Syst. Sci.*, page 15.
- Bougadis, J. and Adamowski, K. (2006). Scaling model of a rainfall intensity-duration-frequency relationship. *Hydrological Processes*, 20(17):3747–3757.
- Burlando, P. and Rosso, R. (1996). Scaling and multiscaling models of depth-duration-frequency curves for storm precipitation. *Journal of Hydrology*, 187(1):45–64.
- Burn, D. H. (1990). Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resources Research*, 26(10):2257–2265.
- Buytaert, W., Celleri, R., Willems, P., De Bievre, B., and Wyseure, G. (2006). Spatial and temporal rainfall variability in mountainous areas: A case study from the south Ecuadorian Andes. *Journal of hydrology*, 329(3-4):413–421.
- Cabras, S., Castellanos, M. E., et al. (2009). An objective Bayesian approach for threshold estimation in the peaks over the threshold model. In *The 7th International Workshop on Objective Bayesian Methodology*, volume 1, pages 11–11.
- Carreau, J. and Bengio, Y. (2009). A hybrid Pareto model for asymmetric fat-tailed data: the univariate case. *Extremes*, 12(1):53–76.
- Carreau, J., Naveau, P., and Neppel, L. (2017). Partitioning into hazard subregions for regional peaks-over-threshold modeling of heavy precipitation. *Water Resources Research*, 53(5):4407–4426.

-
- Carreau, J., Neppel, L., Arnaud, P., and Cantet, P. (2013). Extreme Rainfall Analysis at Ungauged Sites in the South of France : Comparison of Three Approaches. *Journal de la société française de statistique*, 154(2):119–138.
- Carrer, N. L. and Gaetan, C. (2022). Distributional regression models for Extended Generalized Pareto distributions. *arXiv preprint arXiv:2209.04660*.
- Carreteras, M. T. (1987). Cálculo hidrometeorológico de caudales máximos en pequeñas cuencas naturales. *Textos de la Direccion General de Carreteras*, page 1987.
- Ceresetti, D., Anquetin, S., Molinié, G., Leblois, E., and Creutin, J.-D. (2012). Multiscale Evaluation of Extreme Rainfall Event Predictions Using Severity Diagrams. *Weather and Forecasting*, 27(1):174–188. Publisher: American Meteorological Society Section: Weather and Forecasting.
- CH2018 (2018). CH2018-Climate Scenarios for Switzerland-Technical Report. Technical report, National Centre for Climate Services, Zurich (Switzerland). ISBN: 978-3-9525031-4-0.
- Chandan, B., Lucia, B., Thierry, C., James, F., and Roman, L. (2023). Natural catastrophes and inflation in 2022: a perfect storm. last access: 06-11-2023.
- Chang, K. B., Lai, S. H., and Othman, F. (2016). Comparison of annual maximum and partial duration series for derivation of rainfall intensity-duration-frequency relationships in peninsular malaysia. *Journal of Hydrologic Engineering*, 21(1):05015013.
- Chavez-Demoulin, V. and Davison, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(1):207–222.
- Cheng, L. and AghaKouchak, A. (2014). Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate. *Scientific reports*, 4(1):1–6.
- Chow, V. (1962). Hydrologic determination of waterway areas for drainage structures in small drainage basins, *Engrg. Experimental Station, Univ. of Illinois, Urbana I*, 11.
- Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer series in statistics. Springer, London Berlin Heidelberg.
- Cooley, D., Nychka, D., and Naveau, P. (2007). Bayesian Spatial Modeling of Extreme Precipitation Return Levels. *Journal of the American Statistical Association*, 102(479):824–840.
- Courty, L. G., Wilby, R. L., Hillier, J. K., and Slater, L. J. (2019). Intensity-duration-frequency curves at the global scale. *Environmental Research Letters*, 14(8):084045. Publisher: IOP Publishing.
- Cramer, H. (1928). On the composition of elementary errors. *Scandinavian Actuarial Journal*, 1:141–180.
- Cunnane, C. (1988). Methods and merits of regional flood frequency analysis. *Journal of Hydrology*, 100(1-3):269–290.
- Darling, D. A. (1957). The kolmogorov-smirnov, cramer-von mises tests. *The annals of mathematical statistics*, 28(4):823–838.
- Das, S. (2017). Performance of region-of-influence approach of frequency analysis of extreme rainfall in monsoon climate conditions. *International Journal of Climatology*, 37:612–623.

References

- Das, S. (2019). Extreme rainfall estimation at ungauged sites: Comparison between region-of-influence approach of regional analysis and spatial interpolation technique. *International Journal of Climatology*, 39(1):407–423. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/joc.5819>.
- Davies, D. L. and Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2):224–227.
- Davison, A. C., Padoan, S. A., and Ribatet, M. (2012). Statistical Modeling of Spatial Extremes. *Statistical Science*, 27(2).
- de Carvalho, M., Pereira, S., Pereira, P., and Bermudez, P. d. Z. (2021). An Extreme Value Bayesian Lasso for the Conditional Left and Right Tails. *arXiv:2010.07164 [stat]*. arXiv: 2010.07164.
- De Michele, C., Kottegoda, N. T., and Rosso, R. (2001). The derivation of areal reduction factor of storm rainfall from its scaling properties. *Water Resources Research*, 37(12):3247–3252.
- De Michele, C., Zenoni, E., Pecora, S., and Rosso, R. (2011). Analytical derivation of rain intensity–duration–area–frequency relationships from event maxima. *Journal of Hydrology*, 399(3):385–393.
- Deidda, R., Hellies, M., and Langousis, A. (2021). A critical analysis of the shortcomings in spatial frequency analysis of rainfall extremes based on homogeneous regions and a comparison with a hierarchical boundaryless approach. *Stochastic Environmental Research and Risk Assessment*.
- do Nascimento, F. F., Gamerman, D., and Lopes, H. F. (2012). A semiparametric Bayesian approach to extreme value estimation. *Statistics and Computing*, 22:661–675.
- Durrans, S. R., Julian, L. T., and Yekta, M. (2002). Estimation of Depth-Area Relationships using Radar-Rainfall Data. *Journal of Hydrologic Engineering*, 7(5):356–367. Publisher: American Society of Civil Engineers.
- Eicher, C. and Krejci, V. (1996). A new rainfall data system for urban hydrology in Switzerland. *Atmospheric Research*, 42(1):177–198. Closing the gap between theory and practice in urban rainfall applications.
- Evin, G., Blanchet, J., Paquet, E., Garavaglia, F., and Penot, D. (2016). A regional model for extreme rainfall based on weather patterns subsampling. *Journal of Hydrology*, 541:1185–1198.
- Evin, G., Favre, A.-C., and Hingray, B. (2018). Stochastic generation of multi-site daily precipitation focusing on extreme events. *Hydrol. Earth Syst. Sci.*, page 18.
- Falkensteiner, M.-A., Schellander, H., Ehrensperger, G., and Hell, T. (2023). Accounting for seasonality in the metastatistical extreme value distribution. *Weather and Climate Extremes*, page 100601.
- Fauer, F. S., Ulrich, J., Jurado, O. E., and Rust, H. W. (2021). Flexible and Consistent Quantile Estimation for Intensity-Duration-Frequency Curves. preprint, Hydrometeorology/Modelling approaches.
- Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical proceedings of the Cambridge philosophical society*, volume 24, pages 180–190. Cambridge University Press.

-
- Flinch, A. (2022). Florida faces grim reality: Hurricane Ian is deadliest storm in state since 1935. last access: 06-11-2023.
- FOEN (2008). The Floods of 2005 in Switzerland. Technical report, Federal office of environment.
- Fraedrich, K. and Larnder, C. (1993). Scaling regimes of composite rainfall time series. *Tellus A*, 45(4):289–298.
- Frei, C., Davies, H. C., Gurtz, J., and Schär, C. (2000). Climate dynamics and extreme precipitation and flood events in Central Europe. *Integrated Assessment*, 1:281–300.
- Frei, C. and Schär, C. (1998). A precipitation climatology of the Alps from high-resolution rain-gauge observations. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 18(8):873–900.
- Frigessi, A., Ola, H., and Havard, R. (2002). A Dynamic Mixture Model for Unsupervised Tail Estimation without Threshold Selection. *Extremes*, 5(3):219–235.
- Froidevaux, P., Schwanbeck, J., Weingartner, R., Chevalier, C., and Martius, O. (2015). Flood triggering in Switzerland: the role of daily to monthly preceding precipitation. *Hydrology and Earth System Sciences*, 19(9):3903–3924.
- Fukutome, S., Liniger, M. A., and Süveges, M. (2015). Automatic threshold and run parameter selection: a climatology for extreme hourly precipitation in Switzerland. *Theoretical and Applied Climatology*, 120(3-4):403–416.
- Fukutome, S., Schindler, A., and Capobianco, A. (2018). MeteoSwiss extreme value analyses: User manual and documentation. Technical report, Federal Office of Meteorology and Climatology, MeteoSwiss.
- Gabella, M., Speirs, P., Hamann, U., Germann, U., and Berne, A. (2017). Measurement of Precipitation in the Alps Using Dual-Polarization C-Band Ground-Based Radars, the GPM Spaceborne Ku-Band Radar, and Rain Gauges. *Remote Sensing*, 9(11):1147. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.
- Gamet, P. and Jalbert, J. (2022). A flexible extended generalized Pareto distribution for tail estimation. *Environmetrics*, 33(6):e2744.
- Garavaglia, F., Lang, M., Paquet, E., Gailhard, J., Garçon, R., and Renard, B. (2011). Reliability and robustness of rainfall compound distribution model based on weather pattern sub-sampling. *Hydrol. Earth Syst. Sci.*, page 15.
- Gaál, L. and Kysely, J. (2009). Comparison of region-of-influence methods for estimating high quantiles of precipitation in a dense dataset in the Czech Republic. *Hydrology and Earth System Sciences*, 13(11):2203–2219.
- Gaál, L., Kysely, J., and Szolgay, J. (2008). Region-of-influence approach to a frequency analysis of heavy precipitation in Slovakia. *Hydrology and Earth System Sciences*, 12(3):825–839.
- Germann, U., Boscacci, M., Clementi, L., Gabella, M., Hering, A., Sartori, M., Sideris, I. V., and Calpini, B. (2022). Weather Radar in Complex Orography. *Remote Sensing*, 14(3):503. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.
- Germann, U., Boscacci, M., Gabella, M., and Sartori, M. (2015). Peak performance: Radar design for precipitation in the Swiss Alps. In *Meteorological Technology International*,., pages 42–45, United kingdom. UKIP Medai and Events.
-

References

- Germann, U., Galli, G., Boscazzi, M., and Bolliger, M. (2006). Radar precipitation measurement in a mountainous region. *Quarterly Journal of the Royal Meteorological Society*, 132(618):1669–1692.
- Ghanmi, H., Bargaoui, Z., and Mallet, C. (2016). Estimation of intensity-duration-frequency relationships according to the property of scale invariance and regionalization analysis in a mediterranean coastal area. *Journal of Hydrology*, 541:38–49.
- Giannakaki, P. and Martius, O. (2015). Synoptic-scale flow structures associated with extreme precipitation events in northern Switzerland. *Int. J. Climatol.*, 36(6):2497–2515.
- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, pages 423–453.
- Gobiet, A., Kotlarski, S., Beniston, M., Heinrich, G., Rajczak, J., and Stoffel, M. (2014). 21st century climate change in the European Alps—A review. *Science of The Total Environment*, 493:1138–1151.
- Gottardi, F., Obled, C., Gailhard, J., and Paquet, E. (2012). Statistical reanalysis of precipitation fields based on ground network data and weather patterns: Application over French mountains. *Journal of Hydrology*, page 14.
- Goudenhoofd, E., Delobbe, L., and Willems, P. (2017). Regional frequency analysis of extreme rainfall in Belgium based on radar estimates. *Hydrology and Earth System Sciences*, 21(10):5385–5399. Publisher: Copernicus GmbH.
- Gründemann, G. J., Zorzetto, E., Beck, H. E., Schleiss, M., van de Giesen, N., Marani, M., and van der Ent, R. J. (2023). Extreme precipitation return levels for multiple durations on a global scale. *Journal of Hydrology*, 621:129558.
- Gugerli, R., Gabella, M., Huss, M., and Salzmann, N. (2020). Can Weather Radars Be Used to Estimate Snow Accumulation on Alpine Glaciers? An Evaluation Based on Glaciological Surveys. *Journal of Hydrometeorology*, 21(12):2943–2962. Publisher: American Meteorological Society Section: Journal of Hydrometeorology.
- Gugerli, R., Salzmann, N., Huss, M., and Desilets, D. (2019). Continuous and autonomous snow water equivalent measurements by a cosmic ray sensor on an alpine glacier. *The Cryosphere*, 13(12):3413–3434. Publisher: Copernicus GmbH.
- Gupta, V. K. and Waymire, E. (1990). Multiscaling properties of spatial rainfall and river flow distributions. *Journal of Geophysical Research: Atmospheres*, 95(D3):1999–2009.
- Gupta, V. K. and Waymire, E. C. (1993). A Statistical Analysis of Mesoscale Rainfall as a Random Cascade. *Journal of Applied Meteorology and Climatology*, 32(2):251–267.
- Haan, L. and Ferreira, A. (2006). *Extreme value theory: an introduction*, volume 3. Springer.
- Halkidi, M. and Vazirgiannis, M. (2001). Clustering validity assessment: finding the optimal partitioning of a data set. In *Proceedings 2001 IEEE International Conference on Data Mining*, pages 187–194, San Jose, CA, USA. IEEE Comput. Soc.
- Harris, D., Seed, A., Menabde, M., and Austin, G. (1997). Factors affecting multiscaling analysis of rainfall time series. *Nonlinear Processes in Geophysics*, 4(3):137–156.

-
- Haruna, A., Blanchet, J., and Favre, A.-C. (2022). Performance-based comparison of regionalization methods to improve the at-site estimates of daily precipitation. *Hydrology and Earth System Sciences*, 26(10):2797–2811.
- Haruna, A., Blanchet, J., and Favre, A.-C. (2023a). Modeling Areal Precipitation Hazard: A Data-driven Approach to Model Intensity-Duration-Area-Frequency Relationships using the Full Range of Non-Zero Precipitation in Switzerland. *Authorea Preprints*.
- Haruna, A., Blanchet, J., and Favre, A.-C. (2023b). Modeling Intensity-Duration-Frequency Curves for the Whole Range of Non-Zero Precipitation: A Comparison of Models. *Water Resources Research*, 59(6):e2022WR033362.
- Hastie, T. and Tibshirani, R. (1986). Generalized Additive Models. *Statistical Science*, 1(3):14.
- Hosking, J. R. M. and Wallis, J. R. (1988). The effect of intersite dependence on regional flood frequency analysis. *Water Resources Research*, 24(4):588–600.
- Hosking, J. R. M. and Wallis, J. R. (2005). *Regional frequency analysis: an approach based on L-moments*. Cambridge University Press, Cambridge ; New York.
- Innocenti, S., Mailhot, A., and Frigon, A. (2017). Simple scaling of extreme precipitation in North America. *Hydrology and Earth System Sciences*, 21(11):5823–5846. Publisher: Copernicus GmbH.
- IPCC (2023). Summary for Policymakers. Technical report, Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland.
- Isotta, F. A., Frei, C., Weilguni, V., Perčec Tadić, M., Lassègues, P., Rudolf, B., Pavan, V., Cacciamani, C., Antolini, G., Ratto, S. M., Munari, M., Micheletti, S., Bonati, V., Lussana, C., Ronchi, C., Panettieri, E., Marigo, G., and Vertačnik, G. (2014). The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data. *International Journal of Climatology*, 34(5):1657–1675.
- Jordan, A., Krüger, F., and Lerch, S. (2018). Evaluating probabilistic forecasts with scoringRules. *arXiv:1709.04743 [stat]*. arXiv: 1709.04743.
- Jurado, O. E., Ulrich, J., Scheibel, M., and Rust, H. W. (2020). Evaluating the Performance of a Max-Stable Process for Estimating Intensity-Duration-Frequency Curves. *Water*, 12(12):3314. Number: 12 Publisher: Multidisciplinary Digital Publishing Institute.
- Katz, R. W., Parlange, M. B., and Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, 25(8-12):1287–1304.
- Kaufman, L. and Rousseeuw, P. J. (2005). *Finding groups in data: an introduction to cluster analysis*. Wiley series in probability and mathematical statistics. Wiley, Hoboken, N.J.
- Kling, H., Fuchs, M., and Paulin, M. (2012). Runoff conditions in the upper danube basin under an ensemble of climate change scenarios. *Journal of hydrology*, 424:264–277.
- Kolmogoroff, A. (1941). Confidence limits for an unknown distribution function. *The annals of mathematical statistics*, 12(4):461–463.
- Koutsoyiannis, D., Kozonis, D., and Manetas, A. (1998). A mathematical framework for studying rainfall intensity-duration-frequency relationships. *Journal of Hydrology*, 206(1-2):118–135.

- Koutsoyiannis, D. and Xanthopoulos, T. (1999). *Engineering Hydrology, Edition 3*. National Technical University of Athens, Athens. (in Greek).
- Kristvik, E., Johannessen, B. G., and Muthanna, T. M. (2019). Temporal downscaling of IDF curves applied to future performance of local stormwater measures. *Sustainability*, 11(5):1231.
- Kunsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, pages 1217–1241.
- Kysely, J., Gaál, L., and Picek, J. (2011). Comparison of regional and at-site approaches to modelling probabilities of heavy precipitation. *International Journal of Climatology*, 31(10):1457–1472.
- Köplin, N., Schädler, B., Viviroli, D., and Weingartner, R. (2014). Seasonality and magnitude of floods in Switzerland under future climate change. *Hydrological Processes*, 28(4):2567–2578.
- Langousis, A., Mamalakis, A., Puliga, M., and Deidda, R. (2016). Threshold detection for the generalized Pareto distribution: Review of representative methods and application to the NOAA NCDC daily rainfall database. *Water Resources Research*, 52(4):2659–2681. Publisher: John Wiley & Sons, Ltd.
- Langousis, A. and Veneziano, D. (2007). Intensity-duration-frequency curves from scaling representations of rainfall. *Water Resources Research*, 43(2). Publisher: John Wiley & Sons, Ltd.
- Le Gall, P., Favre, A.-C., Naveau, P., and Prieur, C. (2022). Improved regional frequency analysis of rainfall data. *Weather and Climate Extremes*, 36:100456.
- Leonard, M., Metcalfe, A., and Lambert, M. (2008). Frequency analysis of rainfall and streamflow extremes accounting for seasonal and climatic partitions. *Journal of Hydrology*, 348(1):135–147.
- Lima, M. I. P. d. (1998). *Multifractals and the temporal structure of rainfall*. PhD Thesis, Wageningen Agricultural University, Wageningen, Netherlands.
- MacDonald, A., Scarrott, C. J., Lee, D., Darlow, B., Reale, M., and Russell, G. (2011). A flexible extreme value mixture model. *Computational Statistics & Data Analysis*, 55(6):2137–2157.
- Madsen, H., Rosbjerg, D., and Harremoës, P. (1995). Application of the Bayesian approach in regional analysis of extreme rainfalls. *Stochastic Hydrology and Hydraulics*, 9(1):77–88.
- Malekinezhad, H. and Zare-Garizi, A. (2014). Regional frequency analysis of daily rainfall extremes using L-moments approach. *Atmosfera*, 27(4):411–427.
- Marani, M. and Ignaccolo, M. (2015). A metastatistical approach to rainfall extremes. *Advances in Water Resources*, 79:121–126.
- Marra, F., Nikolopoulos, E. I., Anagnostou, E. N., and Morin, E. (2018). Metastatistical Extreme Value analysis of hourly rainfall from short records: Estimation of high quantiles and impact of measurement errors. *Advances in Water Resources*, 117:27–39.
- Marra, F., Zoccatelli, D., Armon, M., and Morin, E. (2019). A simplified MEV formulation to model extremes emerging from multiple nonstationary underlying processes. *Advances in Water Resources*, 127:280–290.
- Martel, J.-L., Brissette, F. P., Lucas-Picher, P., Troin, M., and Arsenault, R. (2021). Climate change and rainfall intensity–duration–frequency curves: Overview of science and guidelines for adaptation. *Journal of Hydrologic Engineering*, 26(10):03121001.

-
- Mascaro, G. (2020). Comparison of local, regional, and scaling models for rainfall intensity–duration–frequency analysis. *Journal of applied meteorology and climatology*, 59(9):1519–1536.
- Masson, D. and Frei, C. (2016). Long-term variations and trends of mesoscale precipitation in the Alps: recalculation and update for 1901–2008. *International Journal of Climatology*, 36(1):492–500.
- Matte, D., Christensen, J. H., and Ozturk, T. (2022). Spatial extent of precipitation events: when big is getting bigger. *Climate Dynamics*, 58(5-6):1861–1875.
- Mèlèse, V., Blanchet, J., and Creutin, J.-D. (2019). A Regional Scale-Invariant Extreme Value Model of Rainfall Intensity-Duration-Area-Frequency Relationships. *Water Resources Research*, 55(7):5539–5558.
- Menabde, M., Seed, A., and Pegram, G. (1999). A simple scaling model for extreme rainfall. *Water Resources Research*, 35(1):335–339.
- MeteoSwiss (2017). Documentation of MeteoSwiss Grid-Data Products Hourly Precipitation Estimation through Rain-Gauge and Radar: CombiPrecip.
- MeteoSwiss (2021). Daily Precipitation (final analysis): RhiresD - MeteoSwiss. https://www.meteoswiss.admin.ch/content/dam/meteoswiss/de/service-und-publicationen/produkt/raeumliche-daten-niederschlag/doc/ProdDoc_RhiresD.pdf.
- MeteoSwiss (2023). The climate of Switzerland. <https://www.meteoswiss.admin.ch/services-and-publications/applications/ext/climate-norm-maps-public.html> [Accessed: 2023-09-13].
- Meylan, P., Favre, A.-C., and Musy, A. (2012). *Predictive hydrology: a frequency analysis approach*. CRC Press.
- Milojevic, T., Blanchet, J., and Lehning, M. (2023). Determining return levels of extreme daily precipitation, reservoir inflow, and dry spells. *Frontiers in Water*, 5:1141786.
- Mineo, C., Ridolfi, E., Napolitano, F., and Russo, F. (2018). The areal reduction factor: A new analytical expression for the Lazio Region in central Italy. *Journal of Hydrology*, 560:471–479.
- Mirhosseini, G., Srivastava, P., and Stefanova, L. (2013). The impact of climate change on rainfall intensity–duration–frequency (idf) curves in alabama. *Regional Environmental Change*, 13(1):25–33.
- Molnar, P. and Burlando, P. (2008). Variability in the scale properties of high-resolution precipitation data in the Alpine climate of Switzerland. *Water Resources Research*, 44(10).
- Muggeo, V. M. (2003). Estimating regression models with unknown break-points. *Statistics in medicine*, 22(19):3055–3071.
- Mèlèse, V., Blanchet, J., and Molinié, G. (2018). Uncertainty estimation of Intensity–Duration–Frequency relationships: A regional analysis. *Journal of Hydrology*, 558:579–591.
- Ménégoz, M., Valla, E., Jourdain, N. C., Blanchet, J., Beaumet, J., Wilhelm, B., Gallée, H., Fettweis, X., Morin, S., and Anquetin, S. (2020). Contrasting seasonal changes in total and intense precipitation in the European Alps from 1903 to 2010. *Hydrology and Earth System Sciences*, 24(11):5355–5377.
-

References

- Nadarajah, S., Anderson, C., and Tawn, J. (1998). Ordered multivariate extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(2):473–496.
- Naveau, P., Huser, R., Ribereau, P., and Hannart, A. (2016). Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research*, 52(4):2753–2769.
- Naveau, P., Toreti, A., Smith, I., and Xoplaki, E. (2014). A fast nonparametric spatio-temporal regression scheme for generalized Pareto distributed heavy precipitation. *Water Resources Research*, 50(5):4011–4017.
- Nhat, L. M., Tachikawa, Y., Sayama, T., and Takara, K. (2007). A simple scaling characteristics of rainfall in time and space to derive intensity duration frequency relationships. *Proceedings of Hydraulic Engineering*, 51:73–78.
- Nhat, L. M., Tachikawa, Y., Sayama, T., and Takara, K. (2008). Estimation of sub-hourly and hourly IDF curves using scaling properties of rainfall at gauged site in Asian Pacific Region. *Annual of Disaster Prevention Research Institute. B*, 51(B):63–73.
- Northrop, P., Attalides, N., and Jonathan, P. (2017). Cross-validators extreme value threshold selection and uncertainty with application to ocean storm severity. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66(1):93–120. arXiv: 1504.06653.
- Ouali, D. and Cannon, A. (2018). Estimation of rainfall intensity–duration–frequency curves at ungauged locations using quantile regression methods. *Stochastic Environmental Research and Risk Assessment*, 32(10):2821–2836.
- Ouarda, T. B., Yousef, L. A., and Charron, C. (2019). Non-stationary intensity-duration-frequency curves integrating information concerning teleconnections and climate change. *International Journal of Climatology*, 39(4):2306–2323.
- Over, T. M. (1995). *Modeling space-time rainfall at the mesoscale using random cascades*. PhD Thesis, University of Colorado, Boulder.
- Overeem, A., Buishand, A., and Holleman, I. (2008). Rainfall depth-duration-frequency curves and their uncertainties. *Journal of Hydrology*, 348(1-2):124–134.
- Overeem, A., Buishand, T. A., and Holleman, I. (2009). Extreme rainfall analysis and estimation of depth-duration-frequency curves using weather radar. *Water Resources Research*, 45(10).
_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2009WR007869>.
- Overeem, A., Buishand, T. A., Holleman, I., and Uijlenhoet, R. (2010). Extreme value modeling of areal rainfall from weather radar. *Water Resources Research*, 46(9).
- Padoan, S. A., Ribatet, M., and Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105(489):263–277.
- Panthou, G., Vischel, T., Lebel, T., Quantin, G., and Molinié, G. (2014). Characterising the space–time structure of rainfall in the Sahel with a view to estimating IDAF curves. *Hydrology and Earth System Sciences*, 18(12):5093–5107.
- Panziera, L., Gabella, M., Germann, U., and Martius, O. (2018). A 12-year radar-based climatology of daily and sub-daily extreme precipitation over the Swiss Alps. *International Journal of Climatology*, 38(10):3749–3769.

-
- Panziera, L., Gabella, M., Zanini, S., Hering, A., Germann, U., and Berne, A. (2016). A radar-based regional extreme rainfall analysis to derive the thresholds for a novel automatic alert system in Switzerland. *Hydrology and Earth System Sciences*, 20(6):2317–2332. Publisher: Copernicus GmbH.
- Papastathopoulos, I. and Tawn, J. A. (2013). Extended generalised Pareto models for tail estimation. *Journal of Statistical Planning and Inference*, 143(1):131–143.
- Paschalis, A. (2013). *Modelling the Space-Time Structure of Precipitation and its Impact on Basin Response*. PhD thesis, ETH Zurich.
- Pfister, L., Brönnimann, S., Schwander, M., Isotta, F. A., Horton, P., and Rohr, C. (2020). Statistical reconstruction of daily precipitation and temperature fields in Switzerland back to 1864. *Climate of the Past*, 16(2):663–678. Publisher: Copernicus GmbH.
- Pickands III, J. (1975). Statistical inference using extreme order statistics. *the Annals of Statistics*, pages 119–131.
- Ragno, E., AghaKouchak, A., Love, C. A., Cheng, L., Vahedifard, F., and Lima, C. H. (2018). Quantifying changes in future intensity-duration-frequency curves using multimodel ensemble simulations. *Water Resources Research*, 54(3):1751–1764.
- Ramos, M. H., Creutin, J.-D., and Leblois, E. (2005). Visualization of storm severity. *Journal of Hydrology*, 315(1):295–307.
- Raynaud, D., Hingray, B., Evin, G., Favre, A.-C., and Chardon, J. (2020). Assessment of meteorological extremes using a synoptic weather generator and a downscaling model based on analogues. *Hydrology and Earth System Sciences*, 24(9):4339–4352. Publisher: Copernicus GmbH.
- Renard, B., Kochanek, K., Lang, M., Garavaglia, F., Paquet, E., Neppel, L., Najib, K., Carreau, J., Arnaud, P., Aubert, Y., Borch, F., Soubeyrou, J.-M., Jourdain, S., Veysseire, J.-M., Sauquet, E., Cipriani, T., and Auffray, A. (2013). Data-based comparison of frequency analysis methods: A general framework. *Water Resources Research*, 49(2):825–843.
- Ribereau, P., Guillou, A., and Naveau, P. (2008). Estimating return levels from maxima of non-stationary random sequences using the Generalized PWM method. *Nonlinear Processes in Geophysics*, 15(6):1033–1039.
- Ritschel, C., Ulbrich, U., N vir, P., and Rust, H. W. (2017). Precipitation extremes on multiple timescales—bartlett—lewis rectangular pulse model and intensity—duration—frequency curves. *Hydrology and Earth System Sciences*, 21(12):6501–6517.
- Rivoire, P., Martius, O., and Naveau, P. (2021). A Comparison of Moderate and Extreme ERA-5 Daily Precipitation With Two Observational Data Sets. *Earth and Space Science*, 8(4):e2020EA001633.
- Rodriguez-Iturbe, I. and Mej a, J. M. (1974). On the transformation of point rainfall to areal rainfall. *Water Resources Research*, 10(4):729–735.
- Roksv g, T., Lutz, J., Grinde, L., Dyr dal, A. V., and Thorarinsdottir, T. L. (2021). Consistent intensity-duration-frequency curves by post-processing of estimated Bayesian posterior quantiles. *Journal of Hydrology*, 603:127000.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.
-

- Sane, Y., Panthou, G., Bodian, A., Vischel, T., Lebel, T., Dacosta, H., Quantin, G., Wilcox, C., Ndiaye, O., Diongue-Niang, A., and Diop Kane, M. (2018). Intensity–duration–frequency (IDF) rainfall curves in Senegal. *Natural Hazards and Earth System Sciences*, 18(7):1849–1866. Publisher: Copernicus GmbH.
- Savina, M., Schäppi, B., Molnar, P., Burlando, P., and Sevruk, B. (2012). Comparison of a tipping-bucket and electronic weighing precipitation gage for snowfall. *Atmospheric Research*, 103:45–51.
- Scarrot, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT–Statistical Journal*, 10(1):33–60.
- Schellander, H., Lieb, A., and Hell, T. (2019). Error structure of metastatistical and generalized extreme value distributions for modeling extreme rainfall in Austria. *Earth and Space Science*, 6(9):1616–1632.
- Scherrer, S. C., Begert, M., Croci-Maspoli, M., and Appenzeller, C. (2016a). Long series of Swiss seasonal precipitation: regionalization, trends and influence of large-scale flow. *International Journal of Climatology*, 36(11):3673–3689.
- Scherrer, S. C., Fischer, E. M., Posselt, R., Liniger, M. A., Croci-Maspoli, M., and Knutti, R. (2016b). Emerging trends in heavy precipitation and hot temperature extremes in Switzerland. *Journal of Geophysical Research: Atmospheres*, 121(6):2626–2637.
- Schertzer, D. and Lovejoy, S. (1987). Physical Modeling and Analysis of Rain and Clouds by Anisotropic Scaling Multiplicative Processes. *Journal of Geophysical Research*, 92:22.
- Schmidli, J. and Frei, C. (2005). Trends of heavy precipitation and wet and dry spells in Switzerland during the 20th century. *International Journal of Climatology*, 25(6):753–771. [_eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/joc.1179](https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/joc.1179).
- Schmidli, J., Schmutz, C., Frei, C., Wanner, H., and Schär, C. (2002). Mesoscale precipitation variability in the region of the European Alps during the 20th century. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 22(9):1049–1074.
- Sebille, Q., Fougères, A.-L., and Mercadier, C. (2017). Modeling extreme rainfall A comparative study of spatial extreme value models. *Spatial Statistics*, 21:187–208.
- Sevruk, B. (1997). Regional Dependency of Precipitation-Altitude Relationship in the Swiss Alps. In *Climatic change at high elevation sites*, pages 123–137. Springer.
- Sevruk, B., Matokova-Sadlonova, K., and Toskano, L. (1998). Topography effects on small-scale precipitation variability in the Swiss pre-Alps. *IAHS Publications-Series of Proceedings and Reports-Intern Assoc Hydrological Sciences*, 248:8.
- Shehu, B. and Haberlandt, U. (2023). Uncertainty estimation of regionalised depth–duration–frequency curves in Germany. *Hydrology and Earth System Sciences*, 27(10):2075–2097.
- Sherman, C. W. (1931). Frequency and intensity of excessive rainfalls at Boston, Massachusetts. *Transactions of the American Society of Civil Engineers*, 95(1):951–960. Publisher: American Society of Civil Engineers.
- Sideris, I., Gabella, M., Sassi, M., and Germann, U. (2014a). The CombiPrecip experience: development and operation of a real-time radar-raingauge combination scheme in Switzerland. In *2014 International Weather Radar and Hydrology Symposium*, pages 1–10.

-
- Sideris, I. V., Gabella, M., Erdin, R., and Germann, U. (2014b). Real-time radar–rain-gauge merging using spatio-temporal co-kriging with external drift in the alpine terrain of Switzerland. *Quarterly Journal of the Royal Meteorological Society*, 140(680):1097–1111. _eprint: <https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.2188>.
- Sikorska, A. E., Viviroli, D., and Seibert, J. (2015). Flood-type classification in mountainous catchments using crisp and fuzzy decision trees. *Water Resources Research*, 51(10):7959–7976.
- Sivapalan, M. and Blöschl, G. (1998). Transformation of point rainfall to areal rainfall: Intensity-duration-frequency curves. *Journal of Hydrology*, 204(1-4):150–167.
- Sminorv, N. V. (1944). Approximate laws of distribution of random variables from empirical data. *Uspekhi Matematicheskikh Nauk*, 10:179–206.
- Sodemann, H. and Zubler, E. (2009). Seasonal and inter-annual variability of the moisture sources for Alpine precipitation during 1995-2002. *International Journal of Climatology*, pages 947–961.
- Stucki, P., Rickli, R., Brönnimann, S., Martius, O., Wanner, H., Grebner, D., and Luterbacher, J. (2012). Weather patterns and hydro-climatological precursors of extreme floods in Switzerland since 1868. *Meteorologische Zeitschrift*, 21(6):531–550.
- Svensson, C. and Jones, D. (2010). Review of methods for deriving areal reduction factors. *Journal of Flood Risk Management*, 3(3):232–245. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1753-318X.2010.01075.x>.
- Tancredi, A., Anderson, C., and O’Hagan, A. (2006). Accounting for threshold uncertainty in extreme value estimation. *Extremes*, 9:87–106.
- Tencaliec, P., Favre, A., Naveau, P., Prieur, C., and Nicolet, G. (2020). Flexible semiparametric generalized Pareto modeling of the entire range of rainfall amount. *Environmetrics*, 31(2).
- Tyralis, H. and Langousis, A. (2019). Estimation of intensity–duration–frequency curves using max-stable processes. *Stochastic Environmental Research and Risk Assessment*, 33(1):239–252.
- Ulrich, J., Fauer, F. S., and Rust, H. W. (2021). Modeling seasonal variations of extreme rainfall on different timescales in Germany. *Hydrology and Earth System Sciences*, 25(12):6133–6149.
- Ulrich, J., Jurado, O. E., Peter, M., Scheibel, M., and Rust, H. W. (2020). Estimating IDF Curves Consistently over Durations with Spatial Covariates. *Water*, 12(11):3119.
- Umbricht, A., Fukutome, S., Liniger, M. A., Frei, C., and Appenzeller, C. (2013). Seasonal Variation of Daily Extreme Precipitation in Switzerland. Technical Report 97, Scientific Report MeteoSwiss.
- Van de Vyver, H. (2018). A multiscaling-based intensity–duration–frequency model for extreme precipitation. *Hydrological Processes*, 32(11):1635–1647. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.11516>.
- Van de Vyver, H. and Demarée, G. R. (2010). Construction of Intensity–Duration–Frequency (IDF) curves for precipitation at Lubumbashi, Congo, under the hypothesis of inadequate data. *Hydrological Sciences Journal*, 55(4):555–564.
-

References

- Veneziano, D. and Langousis, A. (2005). The areal reduction factor: A multifractal analysis. *Water Resources Research*, 41(7). [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2004WR003765](https://onlinelibrary.wiley.com/doi/pdf/10.1029/2004WR003765).
- Veneziano, D. and Lepore, C. (2012). The scaling of temporal rainfall. *Water Resources Research*, 48(8).
- Veneziano, D., Lepore, C., Langousis, A., and Furcolo, P. (2007). Marginal methods of intensity-duration-frequency estimation in scaling and nonscaling rainfall. *Water Resources Research*, 43(10).
- Vergara-Temprado, J., Ban, N., and Schär, C. (2021). Extreme sub-hourly precipitation intensities scale close to the Clausius-Clapeyron rate over Europe. *Geophysical Research Letters*, 48(3):e2020GL089506.
- Viviroli, D., Sikorska-Senoner, A. E., Evin, G., Staudinger, M., Kauzlaric, M., Chardon, J., Favre, A.-C., Hingray, B., Nicolet, G., Raynaud, D., et al. (2022). Comprehensive space-time hydrometeorological simulations for estimating very rare floods at multiple sites in a large river basin. *Natural Hazards and Earth System Sciences*, 22(9):2891–2920.
- Wasko, C., Sharma, A., and Westra, S. (2016). Reduced spatial extent of extreme storms at higher temperatures. *Geophysical Research Letters*, 43(8):4026–4032.
- Widmann, M. and Schär, C. (1997). A principal component and long-term trend analysis of daily precipitation in Switzerland. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 17(12):1333–1356.
- Willems, P. (2000). Compound intensity/duration/frequency-relationships of extreme precipitation for two seasons and two storm types. *Journal of Hydrology*, 233(1):189–205.
- Wójcik, R. and Buishand, T. A. (2003). Simulation of 6-hourly rainfall and temperature by two resampling schemes. *Journal of Hydrology*, 273(1-4):69–80.
- Wood, S. N. (2017). *Generalized additive models: an introduction with R*. CRC press.
- Wood, S. N., Pya, N., and Säfken, B. (2016). Smoothing Parameter and Model Selection for General Smooth Models. *Journal of the American Statistical Association*, 111(516):1548–1563.
- Wright, D. B., Smith, J. A., and Baeck, M. L. (2014). Flood frequency analysis using radar rainfall fields and stochastic storm transposition. *Water Resources Research*, 50(2):1592–1615. [_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2013WR014224](https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2013WR014224).
- Yan, L., Xiong, L., Jiang, C., Zhang, M., Wang, D., and Xu, C.-Y. (2021). Updating intensity-duration-frequency curves for urban infrastructure design under a changing environment. *Wiley Interdisciplinary Reviews: Water*, 8(3):e1519.
- Youngman, B. D. (2019). Generalized Additive Models for Exceedances of High Thresholds With an Application to Return Level Estimation for U.S. Wind Gusts. *Journal of the American Statistical Association*, 114(528):1865–1879.
- Youngman, B. D. (2020). evgam: An R package for Generalized Additive Extreme Value Models. *arXiv:2003.04067 [stat]*. arXiv: 2003.04067.
- Yu, P.-S., Yang, T.-C., and Lin, C.-S. (2004). Regional rainfall intensity formulas based on scaling property of rainfall. *Journal of Hydrology*, 295(1):108–123.

- Zambrano-Bigiarini, M., Nauditt, A., Birkel, C., Verbist, K., and Ribbe, L. (2017). Temporal and spatial evaluation of satellite-based rainfall estimates across the complex topographical and climatic gradients of Chile. *Hydrology and Earth System Sciences*, 21(2):1295–1320. Publisher: Copernicus GmbH.
- Zhao, W., Abhishek, Kinouchi, T., Ang, R., and Zhuang, Q. (2022). A framework for quantifying climate-informed heavy rainfall change: Implications for adaptation strategies. *Science of The Total Environment*, 835:155553.
- Zhao, W., Abhishek, Takhellambam, B. S., Zhang, J., Zhao, Y., and Kinouchi, T. (2023). Spatiotemporal variability of current and future sub-daily rainfall in Japan using state-of-the-art high-quality datasets. *Water Resources Research*, page e2022WR034305.
- Zheng, F., Thibaud, E., Leonard, M., and Westra, S. (2015). Assessing the performance of the independence method in modeling spatial extreme rainfall. *Water Resources Research*, 51(9):7744–7758.
- Zorzetto, E., Botter, G., and Marani, M. (2016). On the emergence of rainfall extremes from ordinary events. *Geophysical Research Letters*, 43(15):8076–8082.