

DOTTORATO DI RICERCA IN
INGEGNERIA CIVILE, CHIMICA, AMBIENTALE E DEI MATERIALI

Ciclo XXXI

08/A1 - Idraulica, Idrologia, Costruzioni Idrauliche e Marittime

ICAR/02 - Costruzioni Idrauliche e Marittime e Idrologia

CATCHMENT SIMILARITY AND SPATIAL
CORRELATION: ADDED VALUE AND IMPACTS ON
HYDROLOGICAL PREDICTIONS IN UNGAUGED BASINS

Presentata da:
Simone Persiano

Supervisore:
Prof. Armando Brath

Co-supervisore:
Prof. Attilio Castellarin

Coordinatore del Dottorato:
Prof. Luca Vittuari

To my Family.
To my Friends.

*But of all the water's secrets,
he saw today only a single one-one that struck his soul.
He saw that this water flowed and flowed,
it was constantly flowing, and yet it was always there;
it was always eternally the same and yet new at every moment!
Oh, to be able to grasp this, to understand this!*

Hermann Hesse, *Siddharta*, 1922.

Abstract

The present research work focuses on the regionalisation of hydrometric information (i.e. transferring empirical information on streamflow regime from neighbouring catchments to the catchment of interest), which is widely used for retrieving accurate estimates of hydrological design variables (e.g. flood flows, mean annual streamflow, low-flow indices, etc.) in ungauged or scarcely gauged basins. The literature reports on several statistical regionalisation methods, which are characterised by different ways of accounting for hydrological similarity between catchments and spatial correlation (or cross-correlation, or intersite correlation) among the hydrological observations collected at different stream-gauges. This Thesis aims at deepening our understanding on the added value and impacts of catchment similarity and spatial correlation on the prediction of flood quantiles and flow-duration curves in ungauged river cross-sections by presenting the results of a three-fold study.

First, we consider the reference procedure for design flood estimation in Triveneto, North-eastern Italy, which assumes the entire study area to be a single hydrologically homogeneous region. Our analyses, based on an updated database of annual maximum floods, confirm the outcomes of previous studies, that is Triveneto cannot be regarded as homogeneous in terms of flood frequency regime; our study also highlights the need for an update of the reference procedure for design flood estimation in the study area. To this aim, we show that a focused-pooling approach, which delineates homogeneous pooling-groups of sites for any given target site by referring to selected geomorphoclimatic descriptors which are particularly relevant for describing regional flood frequency, leads to regional samples characterised by significantly improved homogeneity and, therefore, more reliable design flood estimates.

Although focused pooling is capable of properly exploiting catchment similarity, the general approach does not consider the effects associated with spatial correlation among streamflow series. Nevertheless, all regional datasets of annual sequences of flood flows present some degree of cross-correlation between observed series. Its effects on the accuracy of regional prediction are not well studied yet. Therefore, the second part of our study addresses this important issue by considering two regionalisation procedures that do

consider intersite correlation explicitly, although in two radically different ways. These are the Generalized Least Squares, GLS, and a geostatistical method (i.e. Top-kriging, TK). Recent studies show that TK outperforms GLS for predicting empirical flood quantiles, but they also speculate that the presence of intersite correlation might affect the accuracy of these methods in predicting true flood quantiles. To better understand this aspect, we applied GLS and TK for predicting flood quantiles in a homogeneous pooling-group of sites in Triveneto under different cross-correlation scenarios through a Monte Carlo simulation experiment. Our analyses clearly show that, for both methods, an increasing degree of spatial correlation among the flood sequences results in an increasing masking-effect on the true flooding potential. Moreover, we confirm that TK significantly outperforms GLS when they both assume flood quantiles to scale with drainage area alone, yet, we clearly point out that both methodologies (GLS and TK) significantly improve their accuracy and reliability when flood quantiles are regressed against several catchment descriptors, leading to rather similar overall prediction performances.

In the third and last part of our study, we compare regression methods and geostatistical methods for predicting flow-duration curves in a large and heterogeneous study region, the Danube river basin. In particular, we show that multi-regression models are not a viable regionalisation procedure, while geostatistical models provide much more accurate predictions of flow-duration curves over large and hydrologically heterogeneous study areas.

In summary, all the analyses confirmed the added value for statistical regionalisation of properly handling hydrological heterogeneity, also highlighting the pivotal role played by intersite correlation in observed streamflow time-series.

Contents

Premise	1
1 Prediction of hydrological variables in ungauged basins	7
1.1 Hydrological design variables considered in the Thesis	7
1.1.1 Design floods	7
1.1.2 Flow-duration curves	10
1.2 Predictions in ungauged basins	12
1.2.1 The IAHS-PUB (Predictions in Ungauged Basins) international sci- entific initiative	12
1.2.2 Regionalisation of the hydrological information	13
Homogeneous regions and catchment similarity	13
Spatial correlation	16
1.2.3 Generalities on possible statistical regionalisation methods	17
Regression methods	17
Index-flow methods	19
Geostatistical methods	21
1.3 Research questions	22
2 Statistical methods considered in this Dissertation	25
2.1 Regression methods	25
2.1.1 Ordinary Least Squares (OLS)	26
2.1.2 Weighted Least Squares (WLS)	27
2.1.3 Generalized Least Squares (GLS)	28
2.2 Index-flood methods	29
2.2.1 VAPI project	29
2.2.2 Region of Influence approach	32
2.3 Geostatistical methods	35
2.3.1 Top-kriging	35
2.3.2 Total Negative Deviation Top-kriging	39

3	Design-flood regionalisation: the added value of catchment similarity	43
3.1	Introduction	43
3.2	Study area and database	44
3.3	Test of the accuracy of the VAPI project	48
3.4	Test of the viability of a focused-pooling approach	50
3.4.1	Climate and scale controls on regional flood frequency distribution	50
3.4.2	Application of the Region of Influence approach	53
3.5	Results and discussion	56
3.6	Concluding remarks	58
4	Design-flood regionalisation: the impact of spatial correlation	59
4.1	Introduction	59
4.2	Study area	61
4.2.1	Flood frequency regime of the study region	61
4.2.2	Cross-correlation structure of the study region	64
4.3	Preliminary application of GLS and TK in the real-world study region	66
4.3.1	Application of Generalized Least Squares (GLS)	66
4.3.2	Application of Top-kriging (TK)	68
4.3.3	Performance metrics	69
4.3.4	Results and discussion	70
4.4	Application of GLS and TK to cross-correlated realisations of the homogeneous region	72
4.4.1	Monte Carlo simulation framework	72
4.4.2	Cross-correlated realisations of the region	73
4.4.3	Empirical flood quantiles	75
4.4.4	Prediction of flood quantiles in ungauged sites: application of GLS and TK in cross-validation	75
4.4.5	Results and discussion	75
4.5	Concluding remarks	80
5	Prediction of flow-duration curves in ungauged sites across large geographical areas: the potential of geostatistical approaches	83
5.1	Introduction	83
5.2	Study area and database	84
5.3	Relationships between streamflow indices and catchment descriptors	87
5.3.1	Correlation analysis	87

5.3.2	Multi-regression models	89
5.3.3	Results and discussion	89
5.4	Top-kriging interpolation of flow-duration curves	91
5.4.1	Implementation of TNDTK to the Danube region	91
5.4.2	Cross-validation procedures	92
5.4.3	Results and discussion	94
	Viability of geostatistical prediction of FDCs over large geographical regions	94
	Indicators of the reliability of interpolated FDCs over large areas .	100
5.5	Concluding remarks	104
Conclusions		105
Appendix A Regional frequency analysis with L-moments		109
A.1	Moments	109
A.1.1	Moments	109
A.1.2	Dimensionless moments	110
A.1.3	Sample moments	110
A.2	L-moments	111
A.2.1	Probability weighted moments	111
A.2.2	L-moments	112
A.2.3	L-moment ratios	112
A.2.4	Sample L-moments	113
A.3	Hosking and Wallis test	114
A.4	Choice of a frequency distribution	116
Bibliography		119
Acknowledgements		131

List of Figures

1.1	Example of at-site flood frequency analysis for a gauged site having 50 station-years of data. Green circles represent the empirical sample (i.e. empirical return periods evaluated with the Weibull plotting position), the black line the GEV distribution fitted with the L-moments approach, while red and blue dots indicate the maximum return period for which the at-site estimate can be considered reliable according to Cunnane (1987) and Jakob et al. (1999), respectively.	10
1.2	Example of empirical POR-FDC (black line) and AFDCs (grey lines) for a given site. The observed streamflow data are reported in log-scale versus the corresponding (dimensionless) duration.	12
1.3	Different approaches for identifying hydrologically homogeneous regions: (a) geographically contiguous homogeneous regions, (b) non-contiguous homogeneous regions, (c) hydrologic neighbourhood. This figure is an adaptation of the analogous figure reported in Pugliese (2016) (see also Ouarda et al., 2001).	14
2.1	Compartments of the former SIMN, considered as the basis for identifying independent homogeneous (sub)regions in the VAPI project.	32
2.2	Discretisation of two nested catchments. Smaller points represent the refined grid for smaller areas. This figure is an adaptation of the analogous figure in Pugliese (2016).	37
2.3	Effect of (a) catchment size, and (b, c) nesting on the kriging weights λ_j (red numbers) as estimated by Top-kriging. The centre of the target river basin is indicated as x_0 . This figure is an adaptation of the analogous figures reported in Skøien et al. (2006) and Pugliese (2016).	39
2.4	Sketch of the Total Negative Deviation (TND).	40
2.5	Scheme of the index-flow framework for estimating flow-duration curves with the TNDTK approach.	42

3.1	Location of the 76 Triveneto gauged sites (triangles indicate artificial reservoirs, circles the other gauging stations) considered in the study: the grey colour scale depicts the record length of the corresponding AMS series of peak flow discharges. Red triangles indicate the ungauged dams considered in Sec. 3.4.2; grey lines represent catchments boundaries.	44
3.2	Histogram of the station-years of data per year available in the updated AMS database considered in the study.	45
3.3	Comparison between the regional growth curve (TCEV distribution, black curve) proposed for Triveneto by Villi and Bacchi (2001) and the (dimensionless) empirical regional samples collected for the overall period (i.e. 1913-2013; green dots) and for three different sub-periods: PRE1970 (i.e. 1913-1970, blue dots), PRE1970-VAPI (i.e. 1913-1970 for only the stations considered in the VAPI; blue circles) and POST1970 (i.e. 1971-2013; red dots).	49
3.4	Catchment characteristics of the Triveneto data set (black points). Grey points represent the Austrian, Slovakian and Italian data sets considered in Salinas et al. (2014). Black-dashed lines identify the subdivision into the six subsets based on the 20% and 80% quantiles of the catchments descriptor values for the European data sets.	51
3.5	L-moment ratio diagrams for the subsets defined by intermediate area (a, c) and medium MAP (b, d). Each point represents a record length weighted moving average (WMA) of L-Cv (a, b) or L-Ck (c, d) against the corresponding values of L-Cs and the colour intensity is proportional to the catchment descriptor of interest (area or MAP); WMAs are computed for either 35 Triveneto basins (black-border) or 70 European basins (no-border) having similar area or MAP value.	52
3.6	Comparison between the dimensional growth curve proposed by the VAPI project (black line) and the dimensional growth curve associated with the RoI pooling-group of sites (green line) for Pieve di Cadore. The choice of the GLO distribution for the RoI sample was made by applying the L-moment approach proposed by Hosking and Wallis (see Hosking and Wallis, 1993, 1997). Green dots and blue circles indicate the empirical growth curve for the RoI sample and for the AMS recorded at Pieve di Cadore, respectively.	55

4.1	Homogeneous pooling-group of 20 catchments in Triveneto: orange points and black lines represent the gauging stations and their drainage area, respectively.	62
4.2	Nested structure of the selected RoI pooling-group. For each pair of catchments, the percentage of overlapping area (relative to the bigger of the two) is represented in blue colour scale (the higher the percentage, the deeper the blue).	62
4.3	Regional growth curve (LP3 distribution, blue curve) obtained by applying the L-moments algorithm (i.e. Hosking and Wallis, 1993, 1997) to the regional sample (grey points). The true dimensionless regional quantiles of interest (i.e. return periods $T = 10, 30, 50, 100$ years) are represented with blue points.	63
4.4	Cross-correlation structure of the groups of AMS included in (a) the selected homogeneous pooling-group, and (b) the entire Triveneto region analysed in Chapter 3. Correlations between all pairs of sites in the normal space as a function of distances between their centroids are indicated with circles (diameter reflects the size of simultaneous observations between the two selected catchments), while the solid line depicts the adopted theoretical model (i.e. Tasker and Stedinger, 1989). For the evaluation of the cross-correlation structure, all the pairs of basins having less than 10 concurrent observations were neglected. Blue-filled circles in panel (a) indicate pairs of nested catchments.	65
4.5	Comparison of the absolute error (AE) between empirical and predicted 100-year flood quantiles resulting from (a) 1v- GLS and s-TK, and (b) mv-GLS and ed-TK.	70
4.6	Outcomes of the application (LOOCV scheme) of 1v-GLS (red circles), mv-GLS (red dots), s-TK (blue-bordered triangles) and ed-TK (blue-filled triangles) for predicting at-site 100-year flood quantiles in the study area. Prediction performances are in terms of (a) RBIAS, (b) RMSNE, and (c) RNSE.	71

4.7 Generation of 1000 realisations of the region for each of the three levels of regional cross-correlation (i.e. $\bar{\rho} = 0.2, 0.6, 0.8$). Panels (a), (b) and (c) represent the dispersion of the synthetic series of empirical estimators (grey points) around the corresponding cross-correlation model (black solid line) for $\bar{\rho} = 0.2, 0.6, 0.8$, in this order; the dashed lines indicate the 90% confidence band (upper and lower lines represent 95th and 5th quantiles, respectively). Box-plots in panel (d) show the dispersion of the average correlation values for each of the 1000 realisations generated for each of the three cross-correlation scenarios ($\bar{\rho} = 0.2, 0.6, 0.8$, in this order). 74

4.8 Prediction performance of 1v-GLS (red-bordered box-plots), mv-GLS (red-filled box-plots), s-TK (blue-bordered box-plots) and ed-TK (blue-filled box-plots) in estimating (LOOCV scheme) the empirical estimates of flood quantiles for the three degrees of cross-correlation (i.e. $\bar{\rho} = 0.2, 0.6, 0.8$): box-plots represent the distribution of 1000 values of each metric (i.e. RBIAS, RMSNE, and RNSE) computed for each realisation of the region and for selected return periods (i.e. $T = 10, 30, 50, 100$ years). 76

4.9 Prediction performance of 1v-GLS (red-bordered box-plots), mv-GLS (red-filled box-plots), s-TK (blue-bordered box-plots) and ed-TK (blue-filled box-plots) in estimating (LOOCV scheme) the true flood quantiles for the three degrees of cross-correlation (i.e. $\bar{\rho} = 0.2, 0.6, 0.8$): box-plots represent the distribution of 1000 values of each metric (i.e. RBIAS, RMSNE, and RNSE) computed for each realisation of the region and for selected return periods (i.e. $T = 10, 30, 50, 100$ years). 77

5.1 511 streamgauges considered in the Danube region: blue open circles and red solid dots represent high-quality data (DQ1, 138 gauges) and low-quality data (DQ2, 373 gauges), respectively. 85

5.2 Spearman correlation between streamflow regime indices and catchment descriptors of the basins in the Danube region. Colour and radius of each circle is proportional to the value of the empirical correlation coefficient (see colour scale); numbers indicate p -value of the null hypothesis (i.e. absence of correlation between the two variables). (a) 511 basins (DQ1+DQ2); and (b) 138 basins (DQ1), generated with R-package `corrplot` (Wei and Simko, 2016). 87

5.3	Unit mean annual streamflow (MAS/Area) as a function of basin area for high- (DQ1, blue open circles) and low-quality (DQ2, red solid dots) measurement points in the Danube region. The points outside the highlighted interval ($0.0015\text{--}0.08\text{ m}^3\text{ s}^{-1}\text{ km}^{-2}$; dashed lines) identify extremely discordant sites.	92
5.4	Top-kriging interpolation of mean annual streamflow (MAS) values in cross-validation: empirical (x-axis) vs predicted (y-axis) MAS and Nash-Sutcliffe efficiency for log-transformed (LNSE) and natural (NSE) streamflows. See Sec. 5.4.2 for the three different resampling strategies used in cross-validation: LPOCV-1 (a, d), LPOCV- $1/3$ (b, e) and LPOCV- $1/2$ (c, f). The LPOCV-1 cross-validated predictions of MAS for the six DQ1 gauges associated with the worst prediction of dimensional FDCs are highlighted (red solid dots).	95
5.5	Top-kriging interpolation of standardised FDCs (each empirical curve is standardised by local mean annual streamflow) in cross-validation: empirical (x-axis) vs predicted (y-axis) dimensionless streamflow quantiles and overall NSE for log-transformed (LNSE) and natural (NSE) streamflows.	96
5.6	Top-kriging interpolation of dimensional FDCs in cross-validation: empirical (x-axis) vs predicted (y-axis) dimensionless streamflow quantiles and overall Nash-Sutcliffe efficiency for log-transformed (LNSE) and natural (NSE) streamflows.	96
5.7	Cross-validation of predicted dimensional FDCs: box-plots of LNSE (red) and NSE (blue) values computed for all (a) DQ1 and (b) DQ1+DQ2 measurement points for three different resampling strategies used in the study (see Sec. 5.4.2). Each box shows 25 th , 50 th (i.e. median) and 75 th percentiles; whiskers indicate the most extreme data points that are no more than 1.5 times the inter-quartile range (difference between 75 th and 25 th percentiles) from the box; outlying values are indicated as circles.	98
5.8	Cross-validation of predicted dimensional FDCs: LNSE values computed across all (a) DQ1 and (b) DQ1+DQ2 measurement points as a function of duration; different curves refer to the three different resampling strategies.	99
5.9	Observed and predicted dimensional FDCs for the two catchments having (a) the best and (b) the worst performances in terms of LNSE in LPOCV-1 for DQ1 measurement points.	99

5.10 Prediction variance and local cross-validation LNSE for Danube region elementary catchments: local LNSE values obtained in cross-validation (LPOCV-1 sampling strategy) at (a) 137 DQ1 streamgauges and (b) 497 DQ1+DQ2 streamgauges are colour-coded; kriging variance is also illustrated. 101

5.11 Standardised kriging variance for TNDTK interpolation procedure as a function of LNSE for (a) DQ1 and (b) DQ1+DQ2 subsets. LNSE values smaller than 0.7 are omitted. Dashed (red) lines represent the rolling mean computed with a rolling window of 30 catchments. 102

5.12 Top-kriging interpolation in cross-validation, empirical (x-axes) vs predicted (y-axes): (a) mean annual streamflow (MAS), (b) dimensionless flow-duration curves (FDCs), (c) dimensional FDCs. Predictions refer to 360 DQ2 catchments and are based on observations collected at 137 DQ1 measuring points. 103

List of Tables

2.1	Estimated values of the parameters of the growth curves for the VAPI compartments in Northern Italy.	31
3.1	Catchment characteristics for the 76 river basins included in the Triveneto dataset. For each catchment, the table provides the following information: record length of the AMS series of peak flow discharges n and its average value (i.e. mean annual flood, MAF), drainage area, mean annual precipitation (MAP), minimum and mean elevation (H_{\min} and H_{mean} , respectively), longitude X_g and latitude Y_g of the catchment centroid in the WGS84 /UTM 32 N coordinate system.	46
3.2	Artificial reservoirs for which the RoI approach was applied to estimate the 100-year flood. For each considered dam, the catchment characteristics used for the definition of the RoI pooling-group are reported together with the number of station-years of data included in the RoI sample n_{RoI} and the corresponding heterogeneity measures H_1 and H_2	56
5.1	Empirical values of a selection of streamflow indices and catchment descriptors for the 511 gauged basins in the Danube region.	86
5.2	Effective multi-regression models (i.e. adjusted R-squared $R_{\text{adj}}^2 > 0.50$) identified for the DQ1 class gauges in the Danube region.	90

Premise

The knowledge of hydrological design variables is of paramount importance in many practical engineering applications. Runoff hydrographs, low- and high-flow indices, flow-duration curves are just some of the hydrological variables which provide useful information for different management purposes, such as drought and flood risk management, design of dams, levees and water supply systems, feasibility studies for the construction of new hydropower production facilities, etc. As gauging stations are heterogeneously and sparsely distributed in space, one of the most common tasks in Hydrology is to produce an accurate estimation of the hydrological design streamflow of interest (i.e. high flows, low flows, streamflow indices, flow-duration curves) at ungauged or scarcely gauged river cross-sections. This topic received and continues to receive a great deal of attention from the scientific community; a prominent example is the Predictions in Ungauged Basins (PUB) initiative, promoted by the International Association of Hydrological Sciences (IAHS) for the decade 2003-2012, with the primary aim of reducing uncertainty in hydrological predictions in ungauged basins (see Sivapalan et al., 2003; Blöschl et al., 2013).

The task of estimating hydrological design variables at ungauged catchments is frequently addressed by means of regional analysis, or statistical regionalisation, which consists in transferring to the ungauged (or scarcely gauged) target site the hydrological information collected at gauged sites which are supposed to be hydrologically similar to the target one (see e.g. Hosking and Wallis, 1993, and the text book on the subject from the same authors in 1997). For this reason, regional analysis requires the proper identification of homogeneous regions consisting of hydrologically similar catchments in terms of the processes controlling dynamics and regime of the hydrological variable of interest. The literature reports a significant evolution of the concept of homogeneous region over the last decades: the traditional idea of contiguous and geographically identifiable regions (see e.g. NERC, 1975) has been gradually replaced with the more general idea of homogeneous groups of basins with similar hydrological behaviour, which may or may not be geographically close to each other (see e.g. Acreman and Wiltshire, 1989; Burn, 1990; Ouarda et al., 2001). Hydrological similarity between catchments is a fundamental

prerequisite for several regionalisation procedures, overlooking which they may result in totally unreliable regional estimates of the hydrological signature of interest.

Another important aspect is the potential effect of spatial correlation (i.e. cross-correlation, or intersite correlation) of streamflow series on the accuracy of regional predictions; cross-correlation reduces the actual hydrological information content of a region, impacting the prediction uncertainty of regional models and the assessment of the homogeneity of the region itself (see e.g. Matalas and Langbein, 1962; Stedinger, 1983; Hosking and Wallis, 1988; Stedinger and Lu, 1995; Hosking and Wallis, 1997; Madsen and Rosbjerg, 1997; Madsen et al., 2002; Castellarin et al., 2008).

The present Thesis specifically addresses hydrological similarity and intersite correlation, given their pivotal role in all practical applications of statistical regionalisation of hydrological information. In particular, the study reported herein focuses on the regionalisation of hydrometric information, meaning the regionalisation of streamflow information and statistics, with particular reference to two useful hydrological design variables: design flood and flow-duration curves. The design flood is the flood quantile associated with a given non-exceedance probability over an extended period of time (usually expressed in terms of return period T , and measured in years, the so called T -year flood), for a given river cross-section. The prediction of the design flood is required for the design of dams' spillways, culverts, levees, and for reservoir management, river restoration and risk management (see e.g. Blöschl et al., 2013). Each specific application is associated with a specific return period T : for instance, levees and dams are designed with reference to $T = 100 \div 200$ years and $T = 1000 \div 5000$ years, respectively (see the EU Floods Directive, i.e. Directive 2007/60/EC, transposed in Italy by means of the Legislative Decree 49/2010). On the other hand, a flow-duration curve (FDC) represents the frequency with which a given streamflow is equaled or exceeded over an historical period of time for the river cross-section of interest (see e.g. Vogel and Fennessey, 1994). FDCs provide a simple and compact view of the historical variability of streamflows, reflecting climate conditions and the hydrogeological characteristics of the catchment itself (see e.g. Castellarin, 2014; Westerberg et al., 2016). For this reason, FDCs are routinely used for addressing water resources management problems such as hydropower feasibility studies, classification of streamflow regimes, design of water supply systems, irrigation planning and management, etc. (see e.g. Vogel and Fennessey, 1995; Yaeger et al., 2012).

For the regionalisation of hydrometric information, the literature reports several statistical methods, all assuming the hydrological variable of interest to be a random variable, but at the same time differing significantly in the way they identify groups of hydrologically similar catchments and account for spatial correlation among the variables. In

general, statistical methods are classified in regression-based methods, index-flow methods and geostatistical methods (see e.g. Blöschl et al., 2013). Regression-based methods relate the hydrological variable of interest to observable catchment and climate characteristics; they can require the preliminary identification of a homogeneous region and may or may not account for the presence of unequal record lengths from site to site and cross-correlation among concurrent streamflows between sites (see e.g. Thomas and Benson, 1970; Tasker, 1980; Stedinger and Tasker, 1985; Tasker and Stedinger, 1989). Index-flow methods (see e.g. Dalrymple, 1960; Hosking and Wallis, 1997) estimate the hydrological design variable of interest as the product between an index-flow (i.e. scale factor, depending only on the specific target site) and a dimensionless quantile (i.e. growth factor, which is unique within the given homogeneous region). Finally, geostatistical methods have the peculiarity of not requiring the preliminary identification of homogeneous regions and, differently from the above mentioned methods, exploit the spatial correlation of runoff signatures: they assume the hydrological signature of interest in the ungauged catchment to be a weighted mean of the hydrological signatures in the neighbouring catchments, where the weights account for the spatial correlation of the signatures and the relative locations of the catchments (see e.g. De Marsily, 1986; Chokmani and Ouarda, 2004; Skøien et al., 2006; Skøien and Blöschl, 2007). Despite their remarkable differences in identifying homogeneous regions and accounting for spatial correlation, these methods are widely and successfully used for the prediction of several hydrological design variables in various geographical and climatic contexts. With the aim of deepening our understanding on the added value and impacts of catchment similarity and spatial correlation on the prediction of flood flows and flow-duration curves in ungauged basins, this Thesis considers all three of the types of statistical regionalisation methods in different contexts, as described below.

The first part of the Thesis considers the reference procedure for design flood estimation in Italy, available from the Italian CNR (*Consiglio Nazionale delle Ricerche*, National Research Council) research project VAPI (*VALutazione delle PIene*), which developed an index-flood regional model based upon geographically contiguous regions identified with reference to administrative borders and upon flood data collected up to the 1980s and the 1990s, depending on the specific area. In particular, we focus on Triveneto, a broad mountainous geographical area in North-eastern Italy which counts numerous dams that routinely undergo hydrologic and hydraulic risk assessments, with the aim of verifying accuracy and reliability of the VAPI project for the region (see Villi and Bacchi, 2001). As the VAPI project for Triveneto is based upon AMS of peak discharges that were collected up to the 1980s, we refer to an updated AMS database in order to evaluate the potential of developing an updated reference procedure for design flood estimation in Triveneto.

In particular, we consider an updated index-flood method, which is based on a focused-pooling approach (i.e. Region of Influence, RoI; see e.g. Burn, 1990), and which identifies homogeneous pooling-groups of sites for a given target site, by accounting for the control of geomorphological and climatic characteristics of a given catchment on flood frequency regime.

Concerning the impact of cross-correlation on regional predictions of flood flows, in addition to the above mentioned index-flood procedures (i.e. VAPI project, RoI approach), we test the viability of regression-based methods and geostatistical methods for predicting the design flood in ungauged sites in Triveneto. In particular, we consider Generalized Least Squares (GLS; i.e. Stedinger and Tasker, 1985; Tasker and Stedinger, 1989), which is the reference procedure for estimating streamflow characteristics in ungauged catchments in the USA, and Top-kriging (TK; i.e. Skøien et al., 2006), procedures which are widely applied in several contexts, but are not commonly used in Italy. As it will be discussed in the Thesis, both procedures are intimately dependent on the cross-correlation structure of flood sequences in the study region, an important aspect which is not addressed by the VAPI project and the RoI approach. GLS and TK are completely different in the ways they treat cross-correlation among flows: while GLS accounts for sampling variability and cross-correlation among concurrent streamflows in developing a regional regression model, TK explicitly exploits spatial correlation and is increasingly used for predicting several streamflow indices and hydrological signatures, flood quantiles included, in ungauged sites (see e.g. Castiglioni et al., 2009; Archfield et al., 2013; Pugliese et al., 2014, 2016). In this context, a recent study by Archfield et al. (2013) in the South-eastern USA highlighted that when the aim is the prediction of empirical flood quantiles in an ungauged site, TK is likely to result in better predictive models than GLS, but if the goal is to predict the true flood quantile, based on a limited set of observations, then the effect of spatial correlation could mask the real flood magnitude. As recognised by the authors themselves, the analyses performed in Archfield et al. (2013), being entirely based on empirical data, cannot address the fundamental science question of understanding which technique, between GLS and TK, is better suited for predicting the true unknown flood quantiles in ungauged catchments when the observed flood sequences are affected by cross-correlation. To this aim, we refer to a homogeneous pooling-group of catchments in Triveneto, for which we generate a total of 3000 realisations under different cross-correlation scenarios and evaluate the performances of GLS and TK in predicting the known theoretical values and the sample estimates of flood quantiles in a leave-one-out cross-validation scheme. In particular, our Monte Carlo experiment is aimed at confirming what one could speculate, that is that TK has better efficiencies in predicting the sample estimator of flood quantiles, while

GLS is a better predictor of the true and unknown regional quantiles (looking behind the cross-correlation).

Finally, we focus on the use of geostatistical techniques for predicting flow-duration curves (FDCs) in ungauged basins over large and hydrologically heterogeneous geographical areas. Indeed, compared to regional regression models, whose accuracy is generally unsatisfactory for large and highly heterogeneous study regions, geostatistical procedures have been shown to provide highly reliable predictions of streamflow point indices over large study areas, such as low flows (see e.g. Castiglioni et al., 2011; Parajka et al., 2015) and floods (see e.g. Archfield et al., 2013). Recently, Pugliese et al. (2014) introduced the Total Negative Deviation Top-kriging (TNDTK), which uses TK in an index-flow framework for predicting the entire FDC in ungauged sites. TNDTK has been shown to be reliable for predicting FDCs both in Europe (i.e. Pugliese et al., 2014) and in the USA (i.e. Pugliese et al., 2016). However, it has never been applied over large geographical areas, which can be characterised by significant heterogeneities in terms of streamflow regimes. For this reason, we perform a cross-validation of TNDTK in the Danube region, the largest watershed in Europe, with the aim of evaluating the performance of the procedure in estimating FDCs in ungauged basins and discussing the uncertainty of the interpolation and its dependence on the existing streamgauging network density. This research activity was carried out within a research project with the Joint Research Centre of the European Commission (DG JRC), with the aim of generating a GIS (Geographic Information System) layer reporting the predicted streamflow regime (FDCs) for about 4000 prediction nodes within the watershed of the Danube River.

The Thesis is structured as follows.

Chapter 1 defines the hydrological design variables considered in the study (i.e. design floods, flow-duration curves) and provides the reader with an overview of statistical models for the prediction of runoff signatures in ungauged basins, highlighting the way in which they address the issues of hydrological similarity and spatial correlation.

Chapter 2 deepens the statistical approaches which were applied for the purposes of this Thesis, distinguishing between regression-based methods (Ordinary Least Squares, Weighted Least Squares, and Generalized Least Squares; see Sec. 2.1), index-flow methods (i.e. index-flood methods in this context: VAPI project and Region of Influence approach; see Sec. 2.2), and geostatistical methods (Top-kriging and Total Negative Deviation Top-kriging; see Sec. 2.3).

Chapter 3 reports the assessment of the VAPI project for estimating design flood in ungauged catchments in Triveneto and the updated reference procedure based on the Region of Influence approach, including a detailed analysis of the climate and scale controls

on regional flood frequency distribution.

Chapter 4 compares GLS and TK for predicting flood quantiles in ungauged sites. The section focuses on the Triveneto region and it first compares the procedures for a real-world application which considers a hydrologically homogeneous pooling-group of alpine catchments. Second, the analysis focuses on cross-correlated realisations of the homogeneous region generated through a Monte Carlo simulation framework and aims at understanding which technique, between GLS and TK, is better suited for predicting the true unknown flood quantiles in ungauged catchments when the observed flood sequences are affected by cross-correlation.

Chapter 5 shifts the focus from relatively small hydrologically homogeneous regions to large heterogeneous areas, and from floods to the entire streamflow regime, from high flows to low flows. The chapter presents the statistical regionalisation of flow-duration curves in the Danube region, deepening our understanding on accuracy and applicability of TNDTK over a large region and the dependence of uncertainty of predicted flow-duration curves at ungauged sites on the streamgauging network density.

Chapter 1

Prediction of hydrological variables in ungauged basins

1.1 Hydrological design variables considered in the Thesis

The knowledge of hydrological variables is of paramount importance in many practical applications in the context of integrated water resources and flood and drought risk management (Blöschl et al., 2013). Annual and seasonal runoff, flow-duration curves, low-flow indices, flood quantiles, runoff hydrographs characterise the spatial and temporal distribution of water resources in a river basin, and are fundamental pieces of information for different management purposes, such as water allocation and supply, hydropower production, drought and flood management, irrigation, ecological purposes (e.g. environmental flows for ecological stream health), design of dams and levees, etc. This dissertation focuses on design floods and flow-duration curves, whose definitions and usefulness are illustrated in the following paragraphs.

1.1.1 Design floods

The design flood is generally defined as the flood quantile associated with a given non-exceedance probability $F_X(x) = P[x \leq X]$, usually expressed in terms of return period $T(x)$, which denotes the average recurrence interval over an extended period of time (often expressed in years).

In general, the concept of return period, first introduced by Fuller (1914), gives an easily understandable indication about the rareness of an extreme event, and relies on the basic hypotheses of stationarity (i.e. extreme events arise from a stationary distri-

bution) and independence (i.e. extreme events are independent from each other), which are commonly assumed as necessary conditions in conventional frequency analysis in Hydrology (see e.g. Chow et al., 1988). To this aim, the Annual Maximum Series (AMS) and the Peaks Over Threshold (POT; otherwise known as Partial Duration Series, PDS) approaches are usually adopted for the preliminary selection of independent data. The advantage of the AMS is that flood events can be extracted easily and can be considered independent, whereas the POT approach, considering all the events above a certain threshold level, enables one to select more data points, but requires the definition of a criteria to identify independent consecutive events (see e.g. Madsen et al., 1997, for a detailed comparison of the two approaches). In the present Dissertation, we refer to the AMS approach, which is the most common approach in flood frequency analysis. Volpi et al. (2015) recently demonstrated that, although independence is usually invoked, the concept of return period can be applied also in the case of time-dependent processes (e.g. rainfall processes), and, when they are strongly affected by autocorrelation, one can refer to the concept of equivalent return period (see Volpi et al., 2015, for further details).

The conventional definition of return period is as follows:

$$T(x) = \frac{1}{P[x > X]} = \frac{1}{1 - P[x \leq X]} = \frac{1}{1 - F_X(x)} \quad (1.1)$$

where $P[x > X]$ indicates the exceedance probability of the given value x , and $F_X(x) = P[x \leq X]$ represents the cumulative distribution function of the annual maxima of flood peak discharges. In other words, the design flood associated with a given return period T corresponds with the flood quantile $x = x(T)$.

The prediction of flood quantiles is needed for the design of dams' spillways, culverts, levees, reservoir management, river restoration and risk management (see e.g. Blöschl et al., 2013), where each specific application is associated with a specific return period T . For instance, for assessing the risk associated with levee systems the 100 ÷ 200 years flood quantiles are generally considered, whereas the design flood for dams is associated with $T = 1000 \div 5000$ years (see the EU Floods Directive, i.e. Directive 2007/60/EC, transposed in Italy by means of the Legislative Decree 49/2010).

Clearly, the estimation of the T -year design flood requires the preliminary evaluation of the cumulative distribution function and of the corresponding probability density function (i.e. flood frequency distribution). For a gauged site, where an annual maximum series (AMS) of flood peaks discharges is available, a local (i.e. at-site) flood frequency analysis can be performed to this aim:

1. ranking the observed AMS series, $Q_{AMS} = \{Q_{AMS,1}, Q_{AMS,2}, \dots, Q_{AMS,i}, \dots, Q_{AMS,n}\}$ (where n indicates the sample size) in ascending order, and

2. plotting each ordered AMS observation $Q_{AMS,i}$ versus its corresponding empirical non-exceedance probability F_i , computed with a proper plotting position, e.g. the Weibull plotting position, which is quantile and probability unbiased (Stedinger et al., 1993):

$$F_i = \frac{i}{n+1}, \quad i \in [1, n] \quad (1.2)$$

3. estimating the corresponding empirical return period by applying Eq. (1.1).

In order to estimate the required flood quantile, the empirical cumulative distribution function of annual floods needs to be fitted with a proper theoretical distribution. The selection of a suitable distribution and the estimation of its parameters are crucial steps in flood frequency analysis, and depends on many factors such as tradition, modeller expertise, complexity of study, legislative requirements, and data availability (see Castellarin et al., 2012). In European countries, national guidelines for flood frequency estimation preferentially recommend the two-parameter distributions of Gumbel (GUM), Gamma (GAM), Two-Parameter Log-Normal (LN2) and Exponential (EXP), and the three-parameter distributions Generalized Logistic (GLO), Generalized Extreme Value (GEV), Three-Parameter Log-Normal (LN3) and Pearson type III (PE3) (see e.g. Castellarin et al., 2012; Salinas et al., 2014). Other common distributions are Log-Pearson type III (LP3), which represents the reference distribution in the USA, and the four-parameter distribution Two-Component Extreme Value (TCEV), representing the reference distribution in Italy. Concerning the parameters estimation procedures, the most commonly used include methods of moments, L-moments, maximum likelihood, and Bayesian methods (see e.g. Grimaldi et al., 2011; see also Appendix A for a concise description of moments and L-moments).

In performing an at-site flood frequency analysis, hydrologists have to consider that the reliability of the T -year flood quantile depends on the sample size of the AMS of flood peak discharges which is available at the gauged site of interest. In this regard, Cunnane (1987) suggests that a number of station-years of annual maxima $n \geq T/2$ is required to get a reliable estimation of the T -year flood quantile (e.g. at least 50 station-years of data are necessary for reliably estimating the 100-year flood), while the more recent and restrictive $5T$ -rule proposed by Jakob et al. (1999) indicates that $n \geq 5T$ (e.g. at least 500 station-years of data are required for $T = 100$ yr). Figure 1.1 shows an example of at-site flood frequency analysis, highlighting the remarkable difference between the rules proposed by Cunnane (1987) and Jakob et al. (1999).

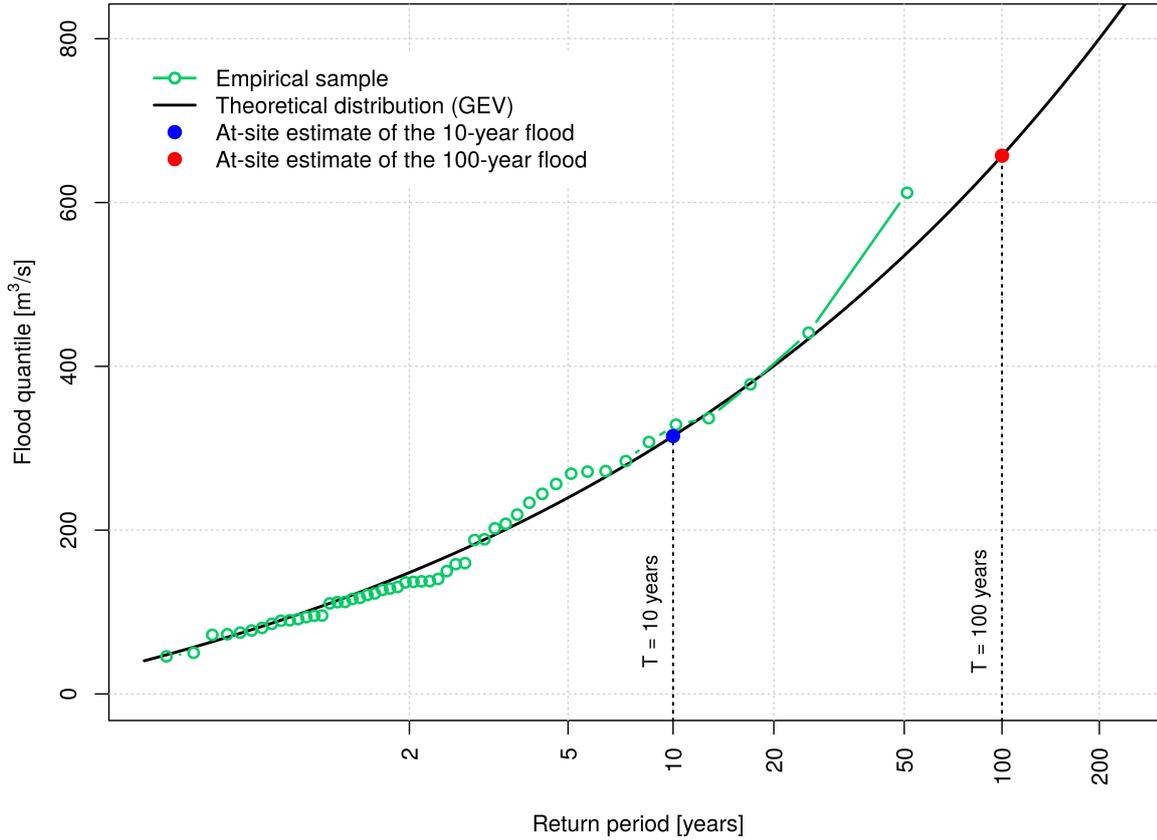


Figure 1.1: Example of at-site flood frequency analysis for a gauged site having 50 station-years of data. Green circles represent the empirical sample (i.e. empirical return periods evaluated with the Weibull plotting position), the black line the GEV distribution fitted with the L-moments approach, while red and blue dots indicate the maximum return period for which the at-site estimate can be considered reliable according to Cunnane (1987) and Jakob et al. (1999), respectively.

Given the limited sample size of the AMS of flood peak discharges available around the world (some gauged sites can reach a hundred station-years of data, at best), the task of estimating the design flood for a given target site is often addressed by means of regional flood frequency analysis (RFFA), by transferring to the target site the hydrological information available at other gauged river cross-sections. RFFA, which is useful also for the estimation at ungauged basins, is described in Sec. 1.2.

1.1.2 Flow-duration curves

A flow-duration curve (FDC) is a graphical representation of the frequency (i.e. percentage of time, or duration) with which a given streamflow is equaled or exceeded over an historical period of time at a given river basin (see e.g. Vogel and Fennessey, 1994). Providing a simple and compact view of the historical variability of streamflows, from high flows to low flows, an FDC is a key signature of the hydrological behaviour of a given

catchment. In particular, the shape of an FDC reflects climate conditions and the hydrogeological characteristics (i.e. size, morphology, permeability) of the catchment itself (see e.g. Castellarin, 2014; Westerberg et al., 2016). For this reason, FDCs are routinely used for addressing water resources management problems such as hydropower feasibility studies, classification of streamflow regimes, design of water supply systems, irrigation planning and management, definition of environmental flows, habitat suitability studies, etc. (see e.g. Vogel and Fennessey, 1995; Yaeger et al., 2012).

The empirical FDC for a gauged site can be obtained from daily runoff data as follows:

1. ranking the observed runoff series, $Q = \{Q_1, Q_2, \dots, Q_i, \dots, Q_n\}$ (where n indicates the sample size) in ascending order, and
2. plotting each ordered observation Q_i versus its corresponding duration d_i , which can be dimensional (e.g. in days) or dimensionless (i.e. fractional, percentage), and coincides with an estimate of the exceedance probability of the i th observation in the sorted sample; as seen also for flood data, the Weibull plotting position can be used to this aim (see Eq. (1.2)).

In general, empirical FDCs can be constructed with reference to two different and complementary representations, depending on the reference period of time (see Vogel and Fennessey, 1994):

- period-of-record FDCs (POR-FDCs), estimated on the basis of the entire runoff record; POR-FDCs provide a long-term representation of the streamflow regime and are useful tools, e.g., for assessing the long-term hydropower potential of a given site, or for patching and extending streamflow data (i.e. Hughes and Smakhtin, 1996; Smakhtin and Masse, 2000);
- annual FDCs (AFDCs), estimated year-wise (i.e. for each year of record), can be used for quantifying the streamflow regime in a typical hydrological year, or in a particularly wet or dry year (see Vogel and Fennessey, 1994).

Figure 1.2 shows the POR-FDC and AFDCs (one AFDC for each year of record) for an example site, highlighting the complementarity of the two representations, which, together, offer a perspective of the inter-annual variability of the streamflow regime for the considered site.

As seen above, the construction of empirical FDCs requires the availability of streamflow data at the river cross-section of interest. However, it is common the need of estimating FDCs at sites which lack streamflow observations (i.e. estimating FDCs at ungauged

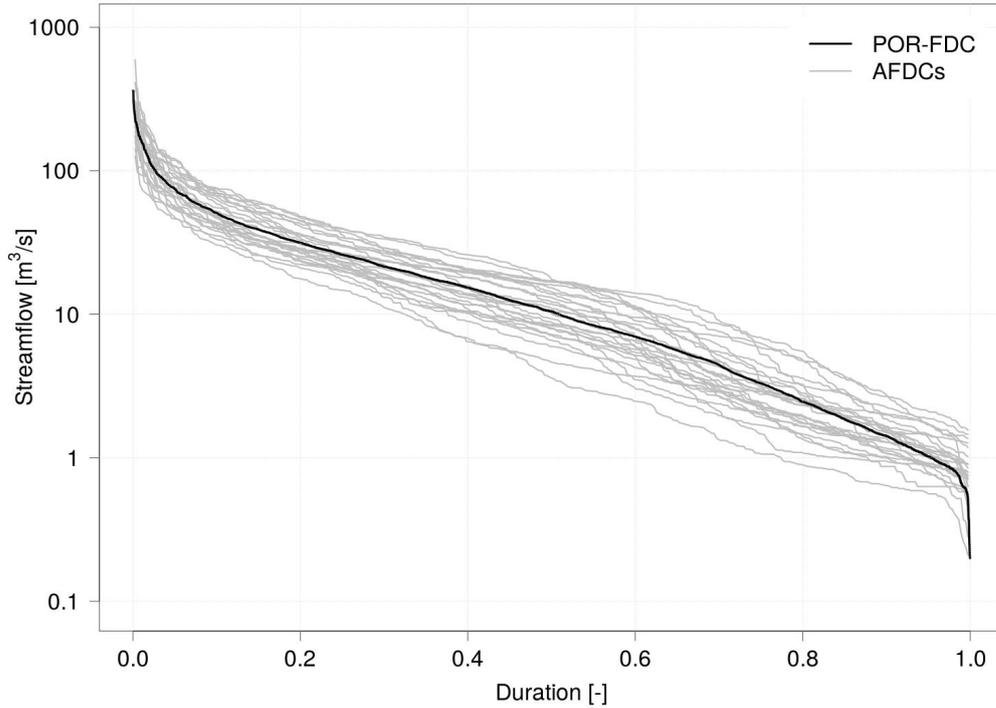


Figure 1.2: Example of empirical POR-FDC (black line) and AFDCs (grey lines) for a given site. The observed streamflow data are reported in log-scale versus the corresponding (dimensionless) duration.

sites, or enhancing empirical FDCs constructed for streamgauges where streamflow observations are limited). To this aim, the literature reports the successful application of a variety of procedures based on the regionalisation of the hydrological information (see e.g. Fennessey and Vogel, 1990; Castellarin et al., 2004a, 2013; Ganora et al., 2009; Pugliese et al., 2014). Because of the widespread use of FDCs in water resources engineering, the prediction of FDCs at ungauged sites was one of the main objectives of the IAHS-PUB initiative (see Sivapalan et al., 2003; Blöschl et al., 2013) described in the following section.

1.2 Predictions in ungauged basins

1.2.1 The IAHS-PUB (Predictions in Ungauged Basins) international scientific initiative

The vast majority of catchments around the world are ungauged (i.e. absence of hydrometric observations) or scarcely gauged (i.e. the available hydrometric observations are not accurate, intermittent and/or not sufficient for estimating the hydrological variable of interest), and, therefore, the only chance to estimate hydrological design variables (e.g. design floods, and flow-duration curves) at these sites is using hydrometric informa-

tion or hydrological knowledge gained elsewhere. This topic has been much debated and has received increasing attention by the scientific community in recent decades, leading to the Predictions in Ungauged Basins (PUB) initiative, promoted by the International Association of Hydrological Sciences (IAHS) for the decade 2003-2012, with the primary aim of reducing uncertainty in hydrological predictions in ungauged basins (see Sivapalan et al., 2003; Blöschl et al., 2013). In particular, Blöschl et al. (2013) identified two fundamentally different types of methods for estimating hydrological variables in ungauged basins: (1) statistical methods (i.e. the hydrological signature of interest is assumed to be a random variable), and (2) process-based methods (i.e. based on some combination of balance equations of mass, momentum and energy). This Thesis focuses on statistical methods, which are the subject of the brief general overview reported in this section.

1.2.2 Regionalisation of the hydrological information

In order to estimate hydrological design variables at ungauged or scarcely gauged river cross-sections, hydrologists usually resort to regional analysis (i.e. regionalisation methods), which consists in identifying homogeneous regions in terms of the hydrological variable of interest (e.g. annual runoff, seasonal runoff, floods, flow-duration curves, etc.), and then transferring the hydrological information from the gauged sites to the ungauged (or scarcely gauged) target site. Concerning floods, for instance, regional flood frequency analysis (see e.g. Hosking and Wallis, 1993, 1997) consists in collecting flood data from gauged basins which are supposed to be hydrologically similar to the target basin (i.e. basins which belong to the same homogeneous region) in terms of flood frequency regime.

Homogeneous regions and catchment similarity

In order to perform a regional analysis, hydrological similarity between catchments is a fundamental requirement. In general, catchments can be considered hydrologically similar "*if they filter the climate variability in a similar fashion, as expressed by their (scaled) hydrological signatures*" (Blöschl et al., 2013). In other words, the concept of homogeneity is related to the regularity of the processes leading to the signature of interest, and permits the simplifying assumption that there is a unique relationship between predictors and the signature for a given group of sites (Blöschl et al., 2013). Since catchments are complex systems and only a partial knowledge of hydrological processes is actually available, the concept of hydrological similarity between catchments usually refers to the similarity between their main characteristics: as runoff is the result of the interplay of geomorphological and climatic characteristics of the catchments, the assumption is that catchments having similar geomorphological and climatic characteristics will have also a

similar hydrological behaviour (Blöschl et al., 2013). The identification of homogeneous regions in terms of the hydrological signature of interest is therefore a crucial aspect in regional analysis. In particular, the concept of homogeneous regions has evolved significantly over time (see Fig. 1.3).

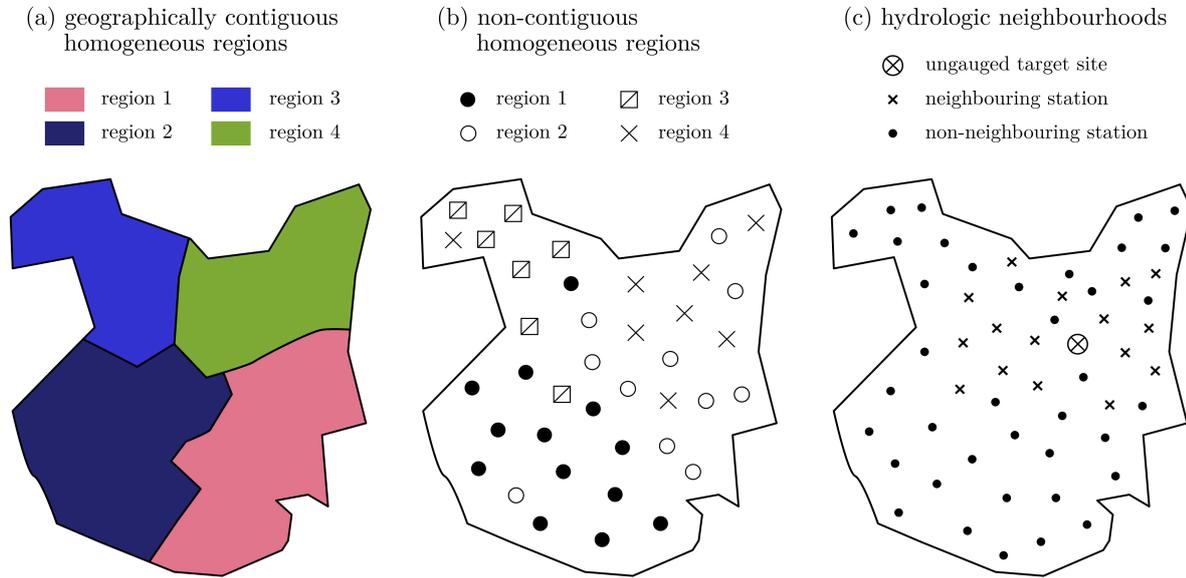


Figure 1.3: Different approaches for identifying hydrologically homogeneous regions: (a) geographically contiguous homogeneous regions, (b) non-contiguous homogeneous regions, (c) hydrologic neighbourhood. This figure is an adaptation of the analogous figure reported in Pugliese (2016) (see also Ouarda et al., 2001).

The traditional approach (see Fig. 1.3(a)) consists in identifying geographically contiguous regions (see e.g. NERC, 1975). Fixed contiguous regions were used, for instance, for developing the reference procedure for design flood estimation in Italy (i.e. VAPI project; see Sec. 2.2.1 for a detailed description). Exploiting spatial proximity, this approach can be advantageous in homogeneous areas with smoothly varying catchment characteristics (Blöschl et al., 2013). Nevertheless, generally, spatial proximity alone does not imply hydrological similarity: close catchments can be very different e.g. in terms of geological characteristics, and, on the other hand, two catchments can be hydrologically similar even if they are far apart from each other.

In line with this consideration, a more advanced approach (see Fig. 1.3(b)) refers to non-contiguous groups of sites which are selected on the basis of climate and catchment characteristics without using spatial proximity as a similarity measure. Thus, this approach allows a greater flexibility in including catchments which are spatially scattered but hydrologically similar. Multivariate statistical methods, and cluster analysis (see e.g. Burn, 1989) in particular, are commonly used for delineating non-contiguous regions. In cluster analysis, a clustering algorithm (see e.g. the Ward hierarchical algorithm; Ward,

1963) is used for automatically classifying catchments into similar groups on the basis of catchment and climate characteristics, whose selection and relative weighting represent a crucial aspect. Moreover, differently from contiguous regions (where ungauged catchments are allocated to the corresponding region according to their geographical location), for non-contiguous regions, allocation rules (e.g. discriminant analysis, classification trees, etc.) need to be defined on the basis of the catchment characteristics which are available for the ungauged site (Blöschl et al., 2013).

A more straightforward approach for the identification of homogeneous regions consists in delineating the hydrologic neighbourhood for the given target site (see Fig. 1.3(c)). From this perspective, the Region of Influence (RoI) approach, introduced by Burn (1990) and Zrinji and Burn (1994) and further refined with the addition of a hierarchical feature by Zrinji and Burn (1996), delinates pooling-groups of sites for a given target site referring to a minimum amount of information in terms of climatic and geomorphological descriptors. A more detailed description of this approach, which is typically used in regional flood frequency analysis in an index-flood framework, can be found in Sec. 2.2.2.

It is important to highlight that the identification of homogeneous pooling-groups of sites is characterised by a "*trade-off between hydrological homogeneity and the size of the group*" (Blöschl et al., 2013). Indeed, the improvement of the reliability of the estimate for the given target site associated with a larger pooling-group (i.e. higher sample size of the corresponding regional sample) is effective only if the group is really homogeneous: as pooling-groups are never truly homogeneous, an increase in the size of the group usually corresponds to a decrease in homogeneity. Different methods were therefore developed for optimising the delineation of the pooling-groups of sites (see e.g. Reed et al., 1999).

With regards to the identification of homogeneous regions, the literature on regional flood frequency analysis proposes different homogeneity tests for the index-flood method (see Sec. 1.2.3). For example, the test proposed by Dalrymple (1960) (see also Chow, 1964) assesses flood homogeneity by analysing the variability of the coefficient of variation and/or the coefficient of skewness of the AMS of flood peak discharges recorded across multiple sites. Other important contributes regarding homogeneity tests can be found, e.g., in Lettenmaier et al. (1987), Stedinger and Lu (1995) and Hosking and Wallis (1993, 1997). In particular, Hosking and Wallis (1993, 1997) introduced a heterogeneity measure and a procedure for selecting the most suitable regional parent distribution, both based on L-moments statistics (Hosking, 1990; see also Appendix A). A comparison of the power of the most common homogeneity tests is reported in Viglione et al. (2007), while Castellarin et al. (2008) show how their performances can be affected by the cross-correlation among sites (i.e. spatial correlation), which represents another crucial aspect in regional analysis.

Spatial correlation

The different finite samples of observed streamflows at gauges which are close in space, usually temporally overlap to each other (i.e. concurrent flows are recorded from different streamgauges). Therefore, as temporal variations of streamflows are spatially correlated, also streamflow series are usually correlated in space (i.e. cross-correlated), due to e.g. extreme rainfall events affecting different streamgauges in the same region.

The main effect of spatial correlation concerns the reduction of the actual hydrological information content of a region: classical studies (i.e. Matalas and Langbein, 1962; Stedinger, 1983) theoretically derived this reduction, and quantified the associated increase in the uncertainty of the regional empirical estimators of streamflow statistics. The effect of spatial correlation among concurrent streamflows in regional analysis has been addressed by several other studies in the literature. Hosking and Wallis (1988) showed that intersite correlation increases the variance of regional flood statistics by impacting the prediction uncertainty of regional flood frequency models (not their bias). Rosbjerg (2007) demonstrated the importance of including the cross-correlation of flood peaks for properly quantifying the uncertainty of flood quantiles regional estimates. Further studies (i.e. Hosking and Wallis, 1997; Madsen and Rosbjerg, 1997; Madsen et al., 2002) showed that cross-correlation may also impact the assessment of the homogeneity of the region, which is the fundamental hypothesis of the index-flood procedures (i.e. Dalrymple, 1960) and a fundamental requirement for performing an effective regional estimation of flood quantiles (see e.g. Lettenmaier et al., 1987; Stedinger and Lu, 1995). In this regard, Castellarin et al. (2008) quantified the loss of performance of the homogeneity tests proposed by Hosking and Wallis (1993, 1997) due to the presence of intersite correlation, which reduces the power of the test by masking the actual degree of heterogeneity of the region (e.g. heterogeneous pooling-groups of cross-correlated sites may be considered as possibly homogeneous by the standard statistical test).

Other studies tackled the problem of accounting for, or exploiting, the presence of spatial correlation when predicting flood quantiles in ungauged basins (i.e. by means of regression methods, or spatial interpolation methods). Concerning regression methods, one example is certainly the Generalized Least Squares regression (GLS; i.e. Stedinger and Tasker, 1985; Tasker and Stedinger, 1989, see also Sec. 2.1.3 for a detailed description of the method), which accounts for cross-correlation of flood peaks between sites. Another significant example is the group of procedures (i.e. spatial interpolation methods) which were developed in the last decade and that explicitly exploit spatial correlation. Spatial interpolation methods have been shown to be effective for predicting several hydrological measures in ungauged sites. For instance, the map-correlation method introduced by

Archfield and Vogel (2010) suggests that, for estimating daily streamflows at an ungauged catchment, the selection of the most correlated gauge leads to better performances than the selection of the nearest one. Moreover, Top-kriging (i.e. Skøien et al., 2006, see also Sec. 2.3.1 for a detailed description of the method) is increasingly used for predicting several streamflow indices in ungauged sites. A recent study by Archfield et al. (2013) in the South-eastern USA shows that Top-kriging resulted in outperforming GLS regression for predicting empirical flood quantiles in ungauged catchments, highlighting an important distinction between the treatments of spatial correlation when using regression-based or spatial interpolation methods to estimate flood quantiles at ungauged locations; nevertheless, the authors themselves underline that if the goal is to predict the true unknown flood quantile, based on a limited set of observations, then the effect of spatial correlation could mask the flood magnitude. This still represents an open issue, which will be investigated in the present dissertation.

It is therefore clear that homogeneity and spatial correlation are two important and much debated aspects in regional analysis. A more detailed overview on the different statistical methods available in literature and on the way in which they address catchments similarity and spatial correlation is reported in the following section.

1.2.3 Generalities on possible statistical regionalisation methods

The scientific literature reports on several possible approaches for transferring available data from gauged to ungauged sites. These approaches can be grouped into two main categories: statistical and process-based (Blöschl et al., 2013). The present dissertation focuses on statistical methods, which assume the hydrological signature of interest to be a random variable. Statistical methods can differ in the way they (1) formulate the model between data and catchment and climate characteristics, (2) estimate model parameters, (3) group catchments and (4) account for spatial correlation among the variables (Cunnane, 1988). The following sections offer a synthetic overview of the available statistical methods, distinguishing between regression methods, index-flow methods, and geostatistical methods.

Regression methods

Regression methods relate the hydrological variable of interest (e.g. flood quantile, flow-duration quantiles), \hat{y} , or a given transformation of it (e.g. logarithmic), to observable catchment and climate characteristics (e.g. drainage area, mean annual precipitation,

etc.), x_i , or a given transformation of them (e.g. logarithmic):

$$\hat{y} = \beta_0 + \sum_{i=1}^p \beta_i x_i + \epsilon \quad (1.3)$$

where p indicates the number of catchment characteristics, β_i are the model parameters (i.e. regression coefficients), and ϵ is the total error (which, in turn, can be expressed as $\epsilon = \delta + \eta$, where δ and η are the model error and the sampling error, respectively). In general, the relationship between the hydrological variable of interest and catchment and climate characteristics is non-linear, but is often approximated with transformed (e.g. log-transformed) variables (see e.g. Thomas and Benson, 1970; Pandey and Nguyen, 1999; Griffis and Stedinger, 2007a). The literature reports several techniques for estimating the model parameters for linear models. The simplest method is Ordinary Least Squares (OLS). Although unbiased, OLS has the limit of lumping sampling and model errors into a single error term, which is supposed to have mean equal to zero and in which the errors are uncorrelated. Therefore, when sampling errors vary from site to site (i.e. short flood peak series are associated with larger sampling errors), OLS is inefficient and can lead to overestimated predictive errors. The Weighted Least Squares (WLS; see e.g. Tasker, 1980) procedure is able to overcome this problem, as it accounts for the sampling error introduced by unequal record lengths, but it does not deal with the presence of correlated sampling errors in neighbouring catchments (which are typically impacted by the same storms). Generalized Least Squares (GLS; see Stedinger and Tasker, 1985; Tasker and Stedinger, 1989) regression represent an extension of WLS for accounting for cross-correlation of flood peaks between sites. A more detailed description of the regression methods applied in the present dissertation (i.e. Ordinary Least Squares and Generalized Least Squares) can be found in Sec. 2.1.

Regression analyses can be performed for the entire domain of interest (i.e. global regressions) or to different regions (or pooling-groups of sites), to which apply separate regression models (i.e. regional regressions).

With regards to the estimation of design flood, the regression approach assumes the presence of a relationship between the T -year flood quantile of interest (or the parameters of the distribution function of flood peaks) and catchment and climate characteristics (see e.g. Thomas and Benson, 1970). An application of least squares regression for estimating flood quantiles for four different regions in the USA (where GLS is the reference procedure) is presented in Thomas and Benson (1970), while Laio et al. (2011) report an application of GLS for predicting the moments of AMS of flood peak discharges in the Piemonte region (Italy).

Concerning the estimation of flow-durations curves (FDCs), regional regression ap-

proaches usually model streamflow quantiles independently of each other. The methods consists in two phases: (1) regionalisation of a certain number of empirical runoff quantiles through a series of multi-regression models, and (2) analytical or graphical interpolation of the regional estimates of streamflow quantiles (see e.g. Franchini and Suppo, 1996; Smakhtin, 2001; Shu and Ouarda, 2012). Quantile regression methods generally do not make assumptions about the distribution or shape of the FDC (nevertheless, there are exceptions, as in Franchini and Suppo, 1996) and avoid the normalisation of FDC and the resulting use of regional dimensionless FDCs. Generally, the method may produce smooth and continuous FDC predictions when a sufficient number of quantiles are regionalised; however, the regression of a large number of streamflow quantiles implies the identification of a large number of multi-regression models. Moreover, the application of regression models for the estimation in ungauged basins may result in inconsistent estimates of the streamflow quantiles, as this prediction strategy might not preserve the monotone relationship between streamflow and duration. In order to avoid such inconsistent results, Archfield (2009) and Archfield et al. (2010) developed a recursive regression approach.

Index-flow methods

Index-flow methods assume that the hydrological variable X , within a statistically homogeneous region, has the same frequency distribution $F(X' = X/\mu)$, apart from a scale factor $\mu(X)$, called index-flow, which is usually the at-site mean of the probability distribution (acknowledge that there are other choices as well: the Flood Estimation Handbook published in the UK by the Institute of Hydrology, 1999, recommends using the median). This means that in a homogeneous region, the T -year quantile for the variable $X(T)$ can be estimated as the product between the index-flow $\mu(X)$ and the dimensionless quantile $X'(T)$ (growth factor):

$$X(T) = X'(T) \mu(X) \tag{1.4}$$

Index-flow methods can be applied to different runoff signatures, from flood quantiles to flow-duration curves. Concerning the estimation of flood quantiles, the index-flood method (Dalrymple, 1960; Hosking and Wallis, 1997) estimates the T -year flood as the product of an index-flood (often defined as the mean or median of the AMS of peak discharges; i.e. mean annual flood, MAF) and a growth factor, which describes the relationship (i.e. growth curve) between the dimensionless flood and the return period T . When observed annual maximum flood data are not available for the site of interest, index-flood can be estimated by means of regression models, or more complex methods such as geostatistical procedures or process-based methods (see e.g. Bocchiola et al., 2003).

The choice of the model for estimating index-flood is a crucial aspect, as its efficacy affects the efficacy of the overall index-flood method. Index-flood methods are a commonly used approach in regional flood frequency analysis and represent the framework on the basis of which the VAPI project (i.e. the reference procedure for estimating design floods in Italy; see Sec. 2.2.1) was originally developed. Rosbjerg (2007) found that quantile estimates from the index-flood method have slightly less uncertainty than quantile regressions.

Concerning flow-duration curves, index-flow methods proposed in the literature can be broadly grouped in two classes, that is (1) parametric (i.e. regionalisation of the parameters of the analytical expression representing the standardised, or dimensionless, FDCs), and (2) non-parametric (i.e. the FDCs scaled with an index-flow are assumed to be the same in all the catchments in the region). In both cases, the index-flow for the ungauged sites needs to be estimated, typically, as the mean annual streamflow (MAS; see e.g. Smakhtin et al., 1997; Ganora et al., 2009) or the median annual runoff. In parametric approaches, the models used for representing the standardised FDCs are parametrised and the parameters are regionalised through regression techniques (see e.g. Le Bouillier and Waylen, 1993a,b; Castellarin et al., 2004a, 2007). Parametric approaches generally consist in three steps: (1) choice of a suitable frequency distribution as parent distribution for a specific region, (2) estimation of the parameters of the distribution on a local basis for the gauged sites located in the pooling-group using the streamflow observations, (3) identification of regional regression models for predicting the distribution parameters on the basis of geomorphological and climatic characteristics of the basin. Despite the practical need to limit the number of parameters, literature (see e.g. Le Bouillier and Waylen, 1993a,b; Castellarin et al., 2004a, 2007) shows that the daily regime of daily streamflow may not be accurately described by theoretical distributions with less than four parameters. Differently from parametric methods, non-parametric methods require two steps: (1) identification of pooling-groups for gauged sites which can be assumed homogeneous in terms of the scaled FDC, and (2) definition of an allocation rule to assign ungauged sites to a group. For instance, in the construction of regional dimensionless FDCs described in Ganora et al. (2009), the pooling-group of gauged sites is identified through cluster analysis using a distance metric which expresses the dissimilarity between pairs of curves, and the FDC of each cluster corresponds to the mean normalised duration curve. In summary, parametric and non-parametric methods are complementary. Modeling the entire FDC, parametric methods have the advantage of producing streamflow estimates which can be associated with any duration of the FDC, but have the limitation of requiring the regionalisation of three or more parameters. On the other hand, non-parametric methods do not have the need to fit a distribution function, but the identification of homogeneous

pooling-groups becomes more important.

Geostatistical methods

Recent studies have shown that geostatistical techniques, which have been originally adopted for the spatial interpolation of point data (see e.g. De Marsily, 1986), can be effectively applied for regionalising hydrometric information (see e.g. Skøien et al., 2006; Skøien and Blöschl, 2007; Chokmani and Ouarda, 2004). Differently from the above mentioned methods, geostatistical methods exploit the spatial correlation of runoff signatures: they assume the hydrological signature of interest in the ungauged catchment to be a weighted mean of the hydrological signatures in the neighbouring catchments, whose weights are estimated considering (1) the spatial correlation of the signature themselves, and (2) the relative locations of the catchments. The main advantage of geostatistical methods is that they are best (i.e. the mean squared error is a minimum) linear (i.e. the estimate is a weighted mean of the data in the area) unbiased (i.e. the mean expected error is zero) estimators (BLUE; see Journel and Huijbregts, 1978). In particular, kriging techniques (i.e. methods for optimising the estimation of variables which are spatially distributed and measured at a network of points) perform the spatial interpolation as follows:

$$\hat{Z}(x_0) = \sum_{j=1}^n \lambda_j Z(x_j) \quad (1.5)$$

where $\hat{Z}(x_0)$ is the prediction of the variable of interest Z at location x_0 , $Z(x_j)$ is the observed value at the point x_j (where $i = 1, 2, \dots, n$), and λ_j represents the corresponding weighting coefficient. For the estimation of weights, kriging considers spatial correlation and configuration of the observation through variogram models fitted to experimental variograms, which describe the spatial correlation structure of the sample data (i.e. express the semivariance between observations as a function of distance and direction of pairs of sampling locations; see e.g. Cressie, 1993). In practical applications, theoretical variogram models (see e.g. Cressie, 1993, for the most commonly used models) are fitted to experimental variograms to ensure a positive-definite covariance matrix.

A significant advantage of geostatistical approaches is that they do not require the delineation of homogeneous pooling-groups of sites, which is a critical but necessary phase for the application of the vast majority of traditional regionalisation approaches (Grimaldi et al., 2011). Moreover, they provide a continuous representation of the runoff signature of interest in the physiographic space (i.e. PSBI, otherwise known as Canonical kriging, CK; see e.g. Chokmani and Ouarda, 2004; Castiglioni et al., 2009, 2011) or along the stream network (i.e. Topological kriging, or Top-kriging; see e.g. Skøien et al., 2006).

Compared to regional regression models (see e.g. Blöschl et al., 2013), whose accuracy is generally unsatisfactory for large and highly heterogeneous study regions, geostatistical procedures have been shown to provide highly reliable predictions of streamflow indices over large study areas, such as FDCs (see e.g. Pugliese et al., 2016), low flows (see e.g. Castiglioni et al., 2011; Parajka et al., 2015), floods (see e.g. Archfield et al., 2013), or the entire streamflow regime (see e.g. Farmer, 2016).

Concerning the estimation of flood quantiles in ungauged sites, an increasingly used technique is topological kriging (or Top-kriging, TK; see Skøien et al., 2006), which interpolates the runoff signature of interest along the stream network by taking the area and the nested structure of catchments into account. The method, originally tested for the prediction of specific 100-year flood for two Austrian regions (Skøien et al., 2006), was shown to provide more plausible and accurate estimates than Ordinary kriging; Top-kriging also provides estimates of the uncertainty, which were found to be smallest on the main stream and to gradually increase moving towards the headwaters. In a more recent study conducted by Archfield et al. (2013) on 61 streamgauges in the South-eastern USA, Top-kriging was shown to outperform, particularly for large catchments, Canonical kriging and GLS regression for predicting 10-, 50-, 100- and 500-year floods.

Geostatistical approaches are increasingly used also for the regionalisation of flow-duration curves. Chokmani and Ouarda (2004) and Castiglioni et al. (2009) use the physiographic space-based interpolation (PSBI, or Canonical kriging, CK), for spatially interpolating the characteristics of FDCs in terms of geomorphoclimatic characteristics. The study by Castiglioni et al. (2009), in particular, estimates FDC quantiles with PSBI by applying a three-dimensional kriging technique for interpolating long-term dimensionless FDCs in the physiographic space. FDC quantiles can be estimated also with Top-kriging (see e.g. Skøien and Blöschl, 2007). In particular, Pugliese et al. (2014, 2016) propose a method (i.e. Total Negative Deviation Top-kriging, TNDTK) for using Top-kriging in an index-flow framework to predict continuous FDCs at ungauged locations, overcoming the limit of modelling streamflow quantiles independently of each other. TNDTK was recently shown to be useful also for the local enhancement of macro-scale rainfall-runoff simulations (see Pugliese et al., 2018). A detailed description of Top-kriging and TNDTK can be found in Sec. 2.3.

1.3 Research questions

As introduced in the previous sections, the identification of pooling-groups of sites that are truly hydrologically homogeneous and how best to handle and, possibly, exploit

spatial correlation in statistical regionalisation are two crucial and, to some extent, open issues. This Thesis aims at deepening our knowledge and understanding with a specific reference to the regionalisation of flood quantiles (i.e. regional flood frequency analysis) and flow-duration curves.

The first part of this Thesis focuses on the reference procedure for design flood estimation in Italy (i.e. VAPI project), which is founded on geographically contiguous regions identified with reference to administrative borders and on AMS series of peak flow discharges collected up to the 1980s and the 1990s, depending on the specific area. In particular, we focus on Triveneto, a broad mountainous geographical area in North-eastern Italy regarded as a unique homogeneous region (i.e. unique growth curve) for operational purposes by Villi and Bacchi (2001), who also pointed out the possible presence of a not fully homogeneous behaviour. For this reason, the first research question we pose is:

Does the reference procedure for design flood estimation in Triveneto, which is founded on geographically contiguous homogeneous regions, need to be updated? If yes, does a focused-pooling approach (i.e. RoI approach), accounting for climatic and scale controls on flood frequency regime, provide more reliable estimates? To this aim, which are the most important drivers controlling the regional frequency regime of flood flows in Triveneto?

A further step regards the treatment of spatial correlation, which cannot be properly addressed by the VAPI project, nor the RoI approach, since neither method does consider it. Other techniques explicitly consider or exploit spatial correlation: GLS and Top-kriging, which are not commonly used in Italy. While GLS develops a regional regression model by accounting for sampling variability and cross-correlation among concurrent streamflows, Top-kriging produces its estimates by exploiting the spatial correlation structure of the region. In particular, a recent study (Archfield et al., 2013) on a set of 61 gauged basins located across the South-eastern USA shows that Top-kriging outperforms GLS, highlighting that when the aim is the prediction of empirical flood quantiles in an ungauged catchment, Top-kriging is likely to result in better predictive models than GLS. Nevertheless, the authors also point out that if the goal is to predict the true unknown flood quantile based on a limited set of observations, then the effect of spatial correlation could mask the flood magnitude. In this context, a further aim of this Thesis is to investigate added value and impact of spatial correlation in predicting flood quantiles in ungauged sites, addressing the unsolved issue raised by Archfield et al. (2013). Specifically, for a simplified situation of nested catchments in a homogeneous region, the second research question reads:

Which technique, between GLS and Top-kriging, is better suited for predicting the true

unknown flood quantiles in ungauged sites when the observed flood sequences are affected by cross-correlation?

Finally, we focus on the use of geostatistical techniques for predicting flow-duration curves in ungauged basins. As introduced in the previous section, a powerful tool to this aim is TNDTK, which was shown to be reliable for predicting flow-duration curves both in Europe (i.e. Pugliese et al., 2014) and in the USA (i.e. Pugliese et al., 2016). However, this geostatistical procedure has never been applied to large geographical areas, which can be characterised by significant heterogeneities in terms of streamflow regimes. For this reason, we perform an extensive sensitive analysis of TNDTK through the application of the procedure to a very large and hydrologically heterogeneous region, the Danube watershed, asking:

Do geostatistical procedures, such as TNDTK, provide reliable estimates of streamflow regime (i.e. flow-duration curves) over very large, and therefore rather heterogeneous, geographical areas? Moreover, is it possible to characterise the uncertainty of their predictions at ungauged sites?

Chapter 2

Statistical methods considered in this Dissertation

After the overview of statistical models for the prediction of runoff signatures in ungauged basins reported in Chapter 1, this chapter presents a more detailed description of the statistical approaches which were applied for the purposes of this dissertation, always distinguishing between regression methods (Ordinary Least Squares, Weighted Least Squares, and Generalized Least Squares; see Sec. 2.1), index-flow methods (i.e. index-flood methods in this context: VAPI project and Region of Influence approach; see Sec. 2.2), and geostatistical methods (Top-kriging and Total Negative Deviation Top-kriging; see Sec. 2.3).

2.1 Regression methods

As introduced in Sec. 1.2.3, the general multi-variate linear regression for estimating streamflow characteristics can be expressed as:

$$\hat{y}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_k x_{ik} + \epsilon_i \quad (2.1)$$

where \hat{y}_i is the estimate of the streamflow characteristic (i.e. dependent variable) of interest at gauge $i = 1, 2, \dots, n$, x_{ik} are basin characteristics (i.e. independent variables; k being the number of basin characteristics), $\beta_0, \beta_1, \beta_2, \beta_3, \beta_k$ are the regression parameters, and ϵ_i indicates the total error. In particular, the total error can be expressed as $\epsilon_i = \delta_i + \eta_i$, where δ_i is the model error (i.e. identifies the lack of fit of the model itself), and η_i is the time-sampling error (i.e. considers that the dependent variable itself is an estimate, often obtained from a limited sample size at each gauge). In general, η_i values from gauges close together are often correlated, as there is a mutual temporal overlap between finite samples

of observed streamflows at the different gauges, and temporal variations of streamflows are spatially correlated (Eng et al., 2009).

Equation (2.1) can be written also in matrix notation:

$$\hat{\mathbf{Y}} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2.2)$$

where $\hat{\mathbf{Y}}$ is the $(n \times 1)$ vector of streamflow characteristics at n sites, \mathbf{X} is the $(n \times k)$ matrix of $(k - 1)$ catchment characteristics augmented with a column of 1, $\boldsymbol{\beta}$ is the $(k \times 1)$ vector of regression parameters, and $\boldsymbol{\epsilon}$ is the $(n \times 1)$ vector of total errors. In particular, $\boldsymbol{\epsilon}$ is supposed to be a random variable with zero mean and variance equal to σ_{ϵ}^2 .

In general, the independent and dependent variables can be transformed to obtain a linear relationship between $\hat{\mathbf{Y}}$ and \mathbf{X} values: the most common transformations include logarithms (base 10, or natural) and addition or subtraction of a constant value. Several studies report the reduction to linear additive forms by means of a log-transformation of both sides of equation (see e.g. Thomas and Benson, 1970; Pandey and Nguyen, 1999; Griffis and Stedinger, 2007a; Laio et al., 2011).

The general expression which indicates the estimation of regression parameters is given by:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \boldsymbol{\Lambda}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Lambda}^{-1} \hat{\mathbf{Y}} \quad (2.3)$$

where \mathbf{X}^T indicates the transpose of matrix \mathbf{X} , and $\boldsymbol{\Lambda}^{-1}$ is the inverse of the weighting matrix $\boldsymbol{\Lambda}$. In particular, $\boldsymbol{\Lambda}$ can be constructed in different ways, depending on the specific approach: Ordinary Least Squares (OLS), Weighted Least Squares (WLS), and Generalized Least Squares (GLS). Once estimated, $\hat{\boldsymbol{\beta}}$ can be used to compute the regression estimates $\hat{\mathbf{Y}}$ (i.e. \hat{y}_i at each i th gauge).

2.1.1 Ordinary Least Squares (OLS)

Ordinary Least Squares (OLS) method represents the easiest approach for estimating regression parameters. In particular, in the OLS approach the weighting matrix $\boldsymbol{\Lambda}$ is equal to the identity matrix \mathbf{I} (see e.g. Montgomery et al., 2001), and the resulting expression for estimating $\hat{\boldsymbol{\beta}}$ is the following:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{Y}} \quad (2.4)$$

The OLS approach is suitable for applications in which δ_i values are independent and have the same variance, i.e. (1) no variation in the precision of calculated dependent variables among different gauges and (2) absence of correlation among concurrent streamflows.

2.1.2 Weighted Least Squares (WLS)

Because records from different gauges have different record lengths, the precision of the estimated streamflow characteristics varies, meaning that different δ_i values have different variances. In cases in which different record lengths (but no correlation) are present, Weighted Least Squares (WLS) approach is recommended. In WLS (see e.g. Tasker, 1980), the regression parameters are computed as shown in Eq. (2.3), where the weighting matrix $\mathbf{\Lambda}$, corresponding to the covariance matrix $\mathbf{\Lambda}_{WLS}$, assigns larger weights to gauges having more reliable estimates of streamflow statistics.

In particular, Tasker (1980) provides a method for estimating $\mathbf{\Lambda}_{WLS}$ for streamflow characteristics which are computed from a Log-Pearson type III (LP3) frequency analysis (see e.g. Bulletin 17B of the Interagency Advisory Committee on Water Data, 1982).

$$\hat{\Lambda}_{WLS,ij} = \begin{cases} \sigma_{\delta}^2 + c_1 \left(\frac{1}{m_i} \right) & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad (2.5)$$

where

$$c_1 = \max \left[0, \bar{\sigma}^2 \left(1 + \frac{\bar{K}^2}{2} (1 + 0.75 \bar{G}^2) + \bar{K} \bar{G} \right) \right] \quad (2.6)$$

$$\sigma_{\delta}^2 = \max \left[0, \sigma_{OLS}^2 - c_1 \left(\frac{1}{n} \sum_{p=1}^n \frac{1}{m_p} \right) \right] \quad (2.7)$$

where σ_{δ}^2 is the model-error variance, m_i is the record length for the i th gauge, σ_{OLS}^2 is the observed mean square error (MSE) of estimate using OLS approach, \bar{K} is the arithmetic average of the LP3 deviates for all gauges in the regression, and \bar{G} is the arithmetic average of the skew values at all gauges (either at-site skew g , or weighted skew G_w). In particular, LP3 deviates are function of probability of exceedance (i.e. return period T) and at-site skew g (see Bulletin 17B of the Interagency Advisory Committee on Water Data, 1982), and $\bar{\sigma}$ is the arithmetic average of standard deviation of the annual time series at each gauged site against the corresponding basin characteristics (Tasker and Stedinger, 1989).

As reported in the Bulletin 17B of the Interagency Advisory Committee on Water Data (1982), the weighted skew $G_{w,i}$ for the i th gauged site can be computed as:

$$G_{w,i} = \omega_i g_i + (1 - \omega_i) G_{R,i} \quad (2.8)$$

where $G_{R,i}$ is the regional skew estimate for the i th gauged site, and

$$\omega_i = \frac{MSE(G_R)}{MSE(g_i) + MSE(G_R)} \quad (2.9)$$

where $MSE(g_i)$ is the estimated mean square error of the skew value at the i th gauged site, and $MSE(G_R)$ is the estimated mean square error of the regional skew values. Several methods are available for the determination of G_R values, and also an alternative approach for computing σ_g^2 is reported in Stedinger and Tasker (1986). The method illustrated in this section is the one reported in the User's Guide to the Weighted-Multiple Linear Regression Program (WREG) (i.e. Eng et al., 2009), which is the software used by the US Geological Survey for estimating streamflow characteristics at ungauged basins in the United States.

2.1.3 Generalized Least Squares (GLS)

In addition to the issue of different record lengths between different stream gauges, another important aspect is that concurrent streamflows observed at different gauges in a region are often cross-correlated. If not properly represented in a regional analysis, cross-correlation affects the precision of the regression parameters, and the estimators of precision are inaccurate. For this reason, in order to account for both correlated streamflows and time-sampling errors, Stedinger and Tasker (1985) and Tasker and Stedinger (1989) introduced the Generalized Least Squares (GLS) method, which improves the representation of the overall regression error ϵ , by assuming it as the sum of the sampling error η for the estimates of the streamflow statistics (e.g. flood statistics), and the modelling error δ in modelling the true index-flows (e.g. index-floods) across catchments.

The regression parameters in GLS are computed by setting $\mathbf{\Lambda} = \mathbf{\Lambda}_{GLS}$ in Eq. (2.3), where $\mathbf{\Lambda}_{GLS}$ contains the estimates of the covariances of ϵ_i among gauged sites. In particular, the main diagonal elements of $\mathbf{\Lambda}_{GLS}$ include a part associated with the model error δ_i and all elements include the effect of the time-sampling error η_i . For streamflow characteristics computed from a LP3 frequency analysis, Tasker and Stedinger (1989) propose to estimate $\mathbf{\Lambda}_{GLS}$ as follows:

$$\hat{\Lambda}_{GLS,ij} = \begin{cases} \sigma_{\delta_i}^2 + \frac{\sigma_i^2}{m_i} \left[1 + K_i G_i + 0.5 K_i^2 (1 + 0.75 G_i^2) \right] & \text{if } i = j \\ \frac{\hat{\rho}_{ij} \sigma_i \sigma_j m_{ij}}{m_i m_j} \left[1 + 0.5 K_i G_i + 0.5 K_j G_j + 0.5 K_i K_j (\hat{\rho}_{ij} + 0.75 G_i G_j) \right] & \text{if } i \neq j \end{cases} \quad (2.10)$$

where i and j are indices of the gauged sites in the region of interest, G_i and G_j are the corresponding skew values (i.e. equal to either at-site skews g , or weighted skew G_w), m_i and m_j are the corresponding record lengths, m_{ij} is the concurrent record length, and $\hat{\rho}_{ij}$ is an estimated value for the cross-correlation of time series of streamflow values used to calculate the streamflow characteristics at gauges i and j .

As sample estimates of ρ_{ij} are imprecise due to the short record lengths of observed flows, the resulting $\hat{\mathbf{\Lambda}}_{GLS}$ cannot often be inverted. To overcome this problem, values of the cross-correlation are usually estimated approximately by referring to the non-linear relationship introduced by Tasker and Stedinger (1989), which is useful for smoothing the sample correlations as function of distance between gauges:

$$\hat{\rho}_{ij} = \theta^{\frac{d_{ij}}{1+\alpha d_{ij}}} = e^{\frac{\ln(\theta) d_{ij}}{1+\alpha d_{ij}}} \quad (2.11)$$

where d_{ij} is the distance between gauges i and j (expressed in miles), and θ and α are the dimensionless model parameters estimated from data. In particular, $\hat{\rho}_{ij}$ is a convex, monotonically decreasing function of d_{ij} when $0 < \theta < 1$ and $\alpha > 0$.

As reported in Stedinger and Tasker (1985), the $\hat{\boldsymbol{\beta}}$ values of Eq. (2.3) for GLS and the $\sigma_{\delta_i}^2$ values in Eq. (2.10) are jointly determined by iteratively searching for a non-negative solution to the following equation:

$$\left(\hat{\mathbf{Y}} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)^T \mathbf{\Lambda}_{GLS}^{-1} \left(\hat{\mathbf{Y}} - \mathbf{X}\hat{\boldsymbol{\beta}}\right) = n - (k + 1) \quad (2.12)$$

As Eq. (2.10) does not account for the error associated with estimating G , Griffis and Stedinger (2007b) introduced a modified version of $\mathbf{\Lambda}_{GLS}$, named $\mathbf{\Lambda}_{GLS,skew}$, which accounts for the uncertainty in the skew estimates. For the estimation of $\mathbf{\Lambda}_{GLS,skew}$ values, Griffis and Stedinger (2007b) consider additional terms, such as the partial derivatives for the gauges (calculated from the approximation for K given in Kite, 1975, 1976), the covariance between the skew values at the different gauged sites, which, in turn, depends on the correlation $\rho_{g_i g_j}$ between skew values (estimated by Martins and Stedinger, 2002) and on the variances of the skew values at gauges i and j (estimated by Griffis and Stedinger, 2009). Further details concerning the computation of $\mathbf{\Lambda}_{GLS,skew}$ can be found in Griffis and Stedinger (2007b) and Eng et al. (2009).

A great value of the GLS approach is being a best linear unbiased estimator (BLUE). Also for this reason, GLS regressions are used in several hydrological applications and represent the standard procedure for the estimation of streamflow characteristics (e.g. flood statistics) in ungauged basins in the United States.

2.2 Index-flood methods

2.2.1 VAPI project

The VAPI (*VALutazione delle PIene*) project is a large-scale research program carried out in the 1990s by the National Group for the Prevention of Hydrogeological Dis-

asters (GNDCI, *Gruppo Nazionale per la Difesa dalle Catastrofi Idrogeologiche*) of the Italian National Research Council (CNR, *Consiglio Nazionale delle Ricerche*), with the aim of defining a uniform methodology for design flood estimation in Italy. The resulting regionalisation model, valid under the hypothesis of stationarity and spatial-temporal independence of the observations, is based on a modified version of the index-flood approach, arranged hierarchically into three distinct levels of analysis (corresponding to the coefficient of skewness, the coefficient of variation and index-flood) and based on the Two-Component Extreme Value (TCEV) distribution (see e.g. Rossi et al., 1984; Fiorentino et al., 1987a,b).

The two components of the TCEV reflect the two different populations which are supposed to generate the population of annual maximum value: the first one caused by ordinary events and the second one caused by extreme events. The cumulative distribution function of the TCEV is expressed as:

$$F_X(x) = \exp \left[-\Lambda_1 \exp\left(-\frac{x}{\theta_1}\right) - \Lambda_2 \exp\left(-\frac{x}{\theta_2}\right) \right] \quad (2.13)$$

where the parameters Λ_1 and Λ_2 (with $\Lambda_1 \geq \Lambda_2 > 0$) are the shape parameters and represent the mean number of annual events of the ordinary and the outlying components, respectively, and θ_1 and θ_2 (with $\theta_2 > \theta_1 > 0$) are the scale parameters of the ordinary and the extraordinary components, respectively.

In practice, TCEV is commonly applied with reference to the standardised variable $Y = X/\theta_1 - \ln(\theta_1)$, whose cumulative distribution function is given by:

$$F_Y(y) = \exp \left[-\exp(-y) - \Lambda_* \exp\left(-\frac{y}{\theta_*}\right) \right] \quad (2.14)$$

where the shape parameter $\theta_* = \theta_2/\theta_1$ and the scale parameter $\Lambda_* = \Lambda_2/\Lambda_1^{1/\theta_*}$ depend only on the coefficient of skewness γ of the distribution; moreover, the coefficient of variation Cv of the TCEV distribution depends on the parameters Λ_* , θ_* and Λ_1 .

Furthermore, the dimensionless reduced variable $K = X/\mu(X)$ is introduced, where $\mu(X) = \theta_1 \eta$ is the expected value of the variable X, and η is function of Λ_1 , Λ_* and θ_* . The cumulative distribution function of K is expressed as:

$$F_K(k) = \exp \left[-\Lambda_1 \exp(-k\eta) - \Lambda_1^{1/\theta_*} \Lambda_* \exp\left(-\frac{k\eta}{\theta_*}\right) \right] \quad (2.15)$$

Note that both the standardised variable Y and the dimensionless reduced variable K are not analytically invertible and, therefore, the use of asymptotic forms is necessary for expressing Y and K as functions of the return period T . In particular, the function $K_T = K(T)$ identifies the growth curve and depends on the parameters Λ_* , θ_* and Λ_1 .

In summary, the estimation of the regional parameters of the TCEV is performed with a hierarchical procedure, based on three successive levels:

1. the first level of regionalisation refers to the standardised variable Y and consists in the research of regions having a constant value of γ (i.e. constant values of Λ_* and θ_* , which are estimated with the maximum likelihood method);
2. the second level of regionalisation refers to the dimensionless reduced variable K and requires the research of sub-regions having constant Cv (i.e. constant Λ_1 , in addition to constant Λ_* and θ_*);
3. the third level of regionalisation consists in the identification of other sub-regions (which do not necessarily coincide with the regions and sub-regions identified at the above mentioned levels), for which empirical relations for estimating the index-flood need to be defined (i.e. mono- and multi-regression models with geomorphologic and climatic catchment descriptors).

By means of the hierarchical procedure described above, the VAPI project analysed the Italian territory referring to the compartments of the former Italian National Hydrographic and Mareographic Service (SIMN, *Servizio Idrografico e Mareografico Nazionale*), each one studied independently from the others (see Fig. 2.1).

Different researchers, from both public institutions and private companies, have been involved in the VAPI project. Concerning Northern Italy, for instance, the most significant contributions were from Brath et al. (1995) and Brath et al. (1997) for the compartments of Parma and Genova, Franchini and Galeati (1996) for Bologna, and Villi and Bacchi (2001) for Venezia. Each of these studies report the estimated values of the parameters Λ_* , θ_* , Λ_1 and η for the considered compartments (see Table 2.1).

Table 2.1: Estimated values of the parameters of the growth curves for the VAPI compartments in Northern Italy.

Compartment	Regions	Λ_*	θ_*	Λ_1	η
Genova	unique	0.30	4.90	9.74	4.11
	region 1	1.21	3.28	24.74	6.51
Parma	region 2	0.22	7.33	16.31	4.80
	region 3	0.56	5.21	13.68	5.55
	region 4	0.13	1.34	9.39	2.97
Bologna	unique	0.75	2.51	9.50	4.22
Venezia	unique	0.89	2.02	15.68	4.56



Figure 2.1: Compartments of the former SIMN, considered as the basis for identifying independent homogeneous (sub)regions in the VAPI project.

As shown in Tab. 2.1, the compartment of Parma was subdivided into four different sub-regions, whereas Genova, Bologna and Venezia were regarded as single homogeneous regions in terms of flood frequency regime. For each region, also the sub-regions for estimating the index-flood were identified. For instance, for estimating index-flood in the homogeneous region of Venezia (i.e. Triveneto area), Villi and Bacchi (2001) identified different regression models with catchment area for the different main river basins in the region.

All the regions and sub-regions identified in the VAPI project are geographically contiguous, reflecting the first approach reported in Fig. 1.3 for the determination of homogeneous regions.

2.2.2 Region of Influence approach

In the context of index-flood methods, the Region of Influence (RoI) approach (see Burn, 1990; Zrinji and Burn, 1994, 1996) is useful for delineating pooling-groups of sites for the given target site (i.e. hydrologic neighbourhood; see Fig. 1.3(c)), referring to geomorphological and climatic characteristics of the catchments. Starting from the sug-

gestion of Acreman and Wiltshire (1987) that a station lying on the boundary between two homogeneous regions can be regarded as a partial member in both regions, the RoI approach is founded on the idea that *"there is no need of defining boundaries between regions but rather each site can have its own "region" consisting of those stations that are sufficiently similar to the site of interest"* (Burn, 1990).

The RoI approach measures the similarity among basins by means of a weighted Euclidean distance in the M -dimensional space defined by a set of M geomorphological and climatic descriptor of the catchments, which are considered to influence the regional flood frequency regime of the basin. The distance metric (i.e. dissimilarity index) between the target and the gauged catchments is usually measured by the root mean square difference of all the catchment and climate characteristics:

$$D_{ij} = \left[\sum_{m=1}^M w_m (X_{m,i} - X_{m,j})^2 \right]^{1/2} \quad (2.16)$$

where D_{ij} is the Euclidean distance from site i to site j , $X_{m,i}$ the standardised value of the m th attribute for site i , and w_m a weight reflecting the relative importance of the m th attribute. As catchment descriptors are usually expressed in terms of different units and scales, in order to make them comparable, the characteristics are usually standardised by their sample standard deviation, calculated using the values observed for all the sites in the study area. A screening process (e.g. multivariate analysis) can be used for selecting the attributes that are most indicative of similarity and, therefore, which can be included in the computation of the Euclidean distance. Characteristics like catchment area, mean annual precipitation (MAP), elevation are typically used for the identification of pooling-groups. Moreover, as expressed in Eq. (2.16), catchment characteristics can be weighted to give preference to some of them, if a prior knowledge about the most important hydrological controls is available (see e.g. Kjeldsen and Jones, 2009).

In particular, Burn (1990) defined a threshold value useful for defining a cutoff for the inclusion of stations into the RoI pooling-group for the given target site (i.e. any station having a distance which exceeds the threshold is not included in the RoI):

$$I_i = \{j : D_{ij} \leq \theta_i\} \quad (2.17)$$

where I_i defines the set of stations in the RoI pooling-group for site i , and θ_i is the threshold for site i , reflecting the target size of the pooling-group. Moreover, Burn (1990) introduced a weighting function η_{ij} which reflects the relative closeness (in the M -dimensional attribute space) to the target site of each station in the RoI, and can be useful for pooling the information from all stations in the RoI for the target site. Burn (1990)

proposed different options for defining the threshold value θ_i and the weighting function η_{ij} in order to maximise the information return resulting from spatial data transfer.

As introduced in Sec. 1.2.2, the effective identification of pooling-group of sites is governed by the homogeneity of the group (see e.g. Lettenmaier et al., 1987; Stedinger and Lu, 1995) and its target size (see e.g. the $5T$ -rule by Jakob et al., 1999), which are both fundamental requirements in order to perform a reliable estimation of the T -year flood quantile. In the search of the optimum size of the pooling-group, it is necessary to consider that, while a higher size avoids undue extrapolations (e.g. the $5T$ -rule suggests that a pooling-group should contain at least $5T$ station-years of data for accurately estimating the T -year flood quantile), an excessively large pooling-group affects homogeneity (i.e. homogeneity usually decreases as the group size increases) and can unduly reduce the influence of each individual site (i.e. negative effect on the estimation of floods) (see also Castellarin et al., 2001).

To take this fundamental requirements into account, Zrinji and Burn (1994) explicitly incorporated a homogeneity test in the process for selecting the pooling-group for the given ungauged target site. In particular, being N the number of sites in the study area, the RoI for a given target site can be defined by organising the dissimilarity indices in an $(N \times N)$ lower triangular matrix, having all zeros on the main diagonal. The identification of the RoI pooling-group starts by including the most similar basin to the site of interest. Then, at each step of the procedure, the next most similar site is added to the RoI, and a homogeneity test is performed in order to assess the degree of homogeneity of the pooling-group. The procedure ends when at least one of the two following conditions is reached: (1) the inclusion of another site leads to an unacceptable level of heterogeneity, or (2) the target size is reached in the case of a particularly homogeneous pooling-group of sites. As suggested in Zrinji and Burn (1994), in the sequential addition of new sites to the RoI, the first site which results in an unacceptable heterogeneity of the overall pooling-group is discarded and the process continues by considering the next most similar sites until one of the two above mentioned conditions is reached. The possibility of skipping one site was introduced by Zrinji and Burn (1994) to avoid a premature ending of the process due to one unusual site (see also Castellarin et al., 2001).

Although the framework of the procedure allows the use of alternative homogeneity tests, the homogeneity test proposed by Hosking and Wallis (1993) (see also Appendix A.3) is commonly used for assessing the heterogeneity of the RoI pooling-groups (see e.g. Zrinji and Burn, 1994, 1996; Castellarin et al., 2001). The Hosking and Wallis test assesses the homogeneity of a group of basins at three different levels, by focusing on three measures of dispersion H_k (where $k = 1, 2, 3$) for different orders of the

sample L-moment ratios (see Hosking, 1990, for an explanation of L-moments; see also Appendix A.2). In particular, H_1 refers to L-Cv (i.e. L-coefficient of variation), H_2 to the sum of L-Cv and L-Cs (i.e. L-coefficient of skewness), and H_3 to the sum of L-Cs and L-Ck (i.e. L-coefficient of kurtosis). Hosking and Wallis (1993) suggest to regard a group of sites as "acceptably homogeneous" if $H_k < 1$, "possibly heterogeneous" if $1 \leq H_k < 2$, and "definitely heterogeneous" if $H_k \geq 2$. Moreover, Hosking and Wallis (1993) observed that higher-order L-moments tend to be more homogeneous in space than the lower-order ones.

On this basis, Zrinji and Burn (1996) proposed a refinement of the RoI approach by introducing a hierarchical feature, which follows the three different levels of the homogeneity test introduced by Hosking and Wallis (1993). In particular, the three different homogeneous RoI pooling-groups for the same target site that can be identified by referring to H_1 , H_2 and H_3 are hierarchically nested. This implies that, for instance, the RoI defined with reference to H_1 is a subset of the pooling-group defined using H_3 as the homogeneity criterion. For this reason, the incorporation of the hierarchical approach in the RoI framework results in the identification of a set of different RoI pooling-groups for the same target site: this approach accounts for the different spatial similarity of different orders of moments and is therefore able to lead to improved parameter estimation of the pooled growth curve. As suggested in Zrinji and Burn (1996), in order to estimate the parameters of the pooled growth curve, the approach based on PWM (i.e. Probability Weighted Moments, introduced by Greenwood et al., 1979) can be used. The literature reports several studies relative to the estimation of parameters for the most commonly used distributions by using PWM (see e.g. Hosking J. R. M., 1985, for the GEV distribution) or L-moments (see e.g. Hosking and Wallis, 1997). Further details on L-moments and the Hosking and Wallis test are provided in Appendix A.

2.3 Geostatistical methods

2.3.1 Top-kriging

Top-kriging (or topological kriging) is a powerful geostatistical procedure developed by Skøien et al. (2006) for the prediction of hydrological variables. Like all kriging approaches, Top-kriging produces predictions of hydrological variables at ungauged sites with a linear combination of the empirical information collected at neighbouring gauging stations. Through this method, the unknown value of the streamflow index of interest at prediction location x_0 , $Z(x_0)$, can be estimated as a weighted average of the regionalised

variable, measured within the neighbourhood:

$$Z(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (2.18)$$

where λ_i is the kriging weight for the empirical value $Z(x_i)$ at location x_i , and n is the number of neighbouring stations used for interpolation. Kriging weights λ_i can be found by solving the typical ordinary kriging linear system (see Eq. (2.19a)) with the constraint of unbiased estimation (see Eq. (2.19b)):

$$\sum_{j=1}^n \gamma_{i,j} \lambda_j + \theta = \gamma_{0,i} \quad i = 1, \dots, n \quad (2.19a)$$

$$\sum_{j=1}^n \lambda_j = 1 \quad (2.19b)$$

where θ is the Lagrange parameter and $\gamma_{i,j}$ is the semi-variance between catchments i and j (Isaaks and Srivastava, 1990). The semi-variance, or variogram, represents the spatial variability of the regionalised variable Z .

Differently from any other kriging method, Top-kriging considers the variable defined over a non-zero support S (i.e. the catchment drainage area A ; Cressie, 1993; Skøien et al., 2006). In particular, the point variable $Z(x_0)$ is averaged over the drainage area A to obtain the spatially average variable $Z_r(A)$:

$$Z_r(A) = \frac{1}{A} \int_A Z(x) dx \quad (2.20)$$

In this way, the kriging system of Eq. (2.19) remains the same, but the semi-variances between the measurements need to be obtained by regularisation, i.e. smoothing the point variogram over the support area. In particular, considering two measurements with catchment area A_i and A_j , respectively, the regularisation consists in assuming the existence of a point variogram $\gamma(h)$, where $h = |x_i - x_j|$ represents the euclidean distance (evaluated on a horizontal plane) between two generic position vectors x_i and x_j within the corresponding catchments, and evaluating the semi-variance $\gamma_{i,j}^r$ between the two measurements as:

$$\gamma_{i,j}^r = \frac{1}{A_i A_j} \int_{A_i A_j} \int_{A_i A_j} \gamma(|x_i - x_j|) dx_i dx_j - \frac{1}{2} \left\{ \frac{1}{A_i^2} \iint_{A_i} \gamma(|x_i - x_j|) dx_i dx_j + \frac{1}{A_j^2} \iint_{A_j} \gamma(|x_i - x_j|) dx_i dx_j \right\} \quad (2.21)$$

As can be seen, Eq. (2.21) is composed of two parts: the first one integrates all the variance between the two catchments, while the second one subtracts the averaged variance within the catchments (i.e. representing the smoothing effect of the support,

which indicates that the variance of the averaged variable decreases as the support area increases; Skøien et al., 2006). In this way, Eq. (2.21) can be used for evaluating the variogram of the averaged variable from the point variogram. Then, $\gamma_{i,j}^r$ can be inserted into the kriging matrix of Eq. (2.19) and the kriging system can be solved to compute the weights λ_i .

The integration shown in Eq. (2.21) is performed over the catchment area that drains to a particular location on the stream network (i.e. the outlet of the target catchment), therefore enabling hydrologists to perform geo-statistics on stream network (Skøien et al., 2006). In particular, the catchment area is discretised by a grid and the integrals in Eq. (2.21) are replaced by sums. In order to avoid numerical problems that are likely to flaw the results, the discretisation has to be performed so that the grid is the same for each catchment every time it is discretised. Specifically, experiments indicate that regularly gridded discretisations give numerically better results for the same number of discretisation points (Skøien, 2014). As the size of areas can differ by several orders of magnitude, usually an adaptive grid is used, where the resolution scales with the area: the discretisation consists in starting with a coarse grid covering the region of interest, and refining it for a certain area until a requested minimum number of points is present (Skøien, 2014). This method is also useful for assuring that points used to discretise a large support will be reused when discretising smaller nested catchments (e.g. sub-catchments within larger catchments Skøien, 2014, see also Fig. 2.2).

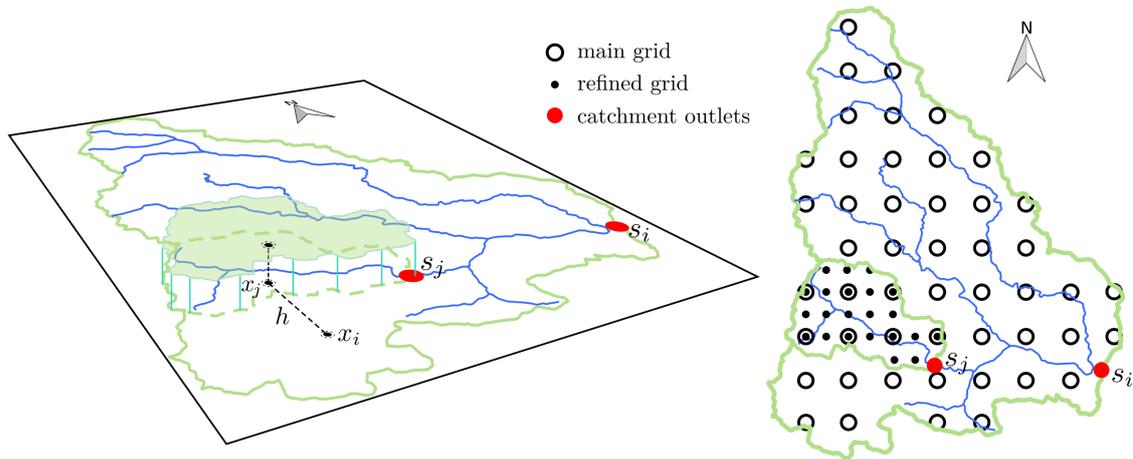


Figure 2.2: Discretisation of two nested catchments. Smaller points represent the refined grid for smaller areas. This figure is an adaptation of the analogous figure in Pugliese (2016).

Another important aspect which Top-kriging accounts for is the possible presence of a nugget effect (i.e. discontinuity close to the origin, representing the variability at scales much smaller than the distance between measurements) in the point variogram. Many variables of interest in hydrological applications, such as streamflow data, are likely to

have a nugget effect. As the direct regularisation with Eq. (2.21) would make the nugget vanish (even for small catchments), Skøien et al. (2006) suggest to regularise the nugget separately, by considering the nugget variance as the variance of a spatially independent random variable, and then adding the regularised nugget effect to the regularised variogram of Eq. (2.21). More details about the regularisation of the nugget effect can be found in Skøien et al. (2006).

A remarkable advantage of Top-kriging is that (like any other kriging method) it provides an estimate of the kriging variance σ_R^2 , which represents the uncertainty of the estimates at any location:

$$\sigma_R^2 = \sum_{j=1}^n \gamma_{j,0} \lambda_j + \theta \quad (2.22)$$

where $\gamma_{j,0}$ is the gamma value between the target catchment and the neighbouring catchments.

The merits of Top-kriging over traditional kriging methods (i.e. Ordinary kriging) are illustrated in Fig. 2.3, which highlights the influence of catchment area and the nested structure of catchments on the evaluation of the weights λ_j . In all the three examples, the neighbouring catchments have the same centre-to-centre distance to the target catchment. While in Ordinary kriging this would imply the same weights λ_j for all the neighbouring catchments, Top-kriging weights them differently. In particular, Fig. 2.3(a) shows that Top-kriging assigns larger weights to larger catchments (which are regarded as the most certain, or having the least biased measurement in comparison to the mean); moreover, for identical catchment areas and centre-to-centre distances, a sub-catchment of the target catchment gets a larger weight (see Fig. 2.3(b)), and, for the reverse case, more weight is attached to the catchment into which the target catchment drains, although the neighbours have the same areas and the same centre-to-centre distances to the target catchment (see Fig. 2.3(c)) (Skøien et al., 2006).

Thanks to its important advantages, Top-kriging has been shown to be particularly successful in predicting a wide spectrum of point streamflow indices and variables in various geographical and climatic contexts: low flows (Castiglioni et al., 2011; Laaha et al., 2014), high flows and floods (Merz et al., 2008; Archfield et al., 2013), stream temperature (Laaha et al., 2013), habitat suitability indices (Ceola and Pugliese, 2014), and daily streamflow series (Skøien and Blöschl, 2007; Vormoor et al., 2011; Parajka et al., 2015; de Lavenne et al., 2016; Farmer, 2016).

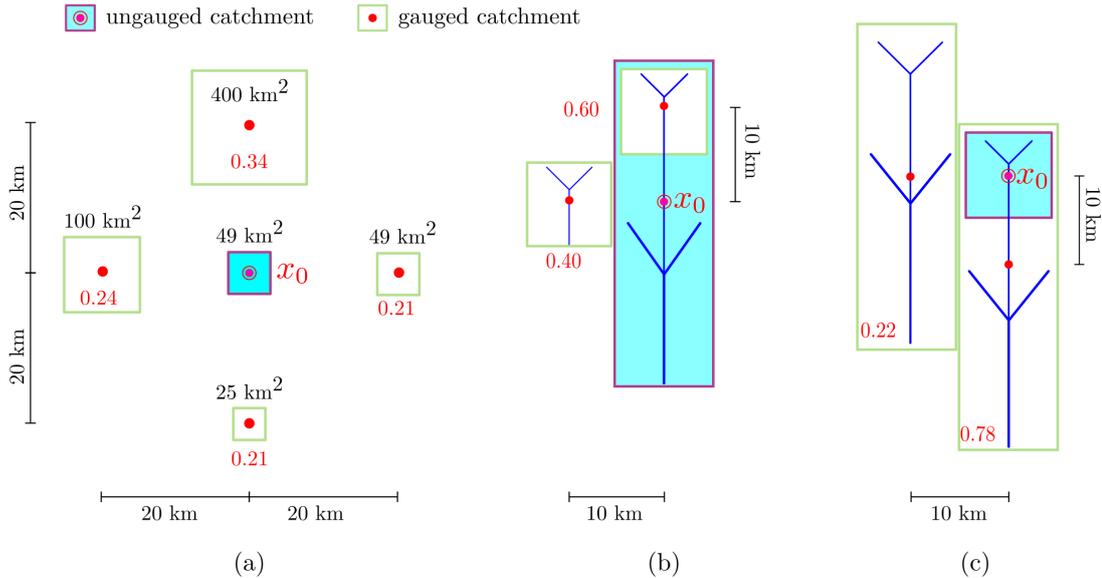


Figure 2.3: Effect of (a) catchment size, and (b, c) nesting on the kriging weights λ_j (red numbers) as estimated by Top-kriging. The centre of the target river basin is indicated as x_0 . This figure is an adaptation of the analogous figures reported in Skøien et al. (2006) and Pugliese (2016).

2.3.2 Total Negative Deviation Top-kriging

Traditional applications of Top-kriging mainly focus on the prediction of point stream-flow indices (e.g. flood quantiles, low-flow indices, etc.). In order to estimate the entire FDC in ungauged sites (i.e. not separately interpolating the different FDC quantiles), Pugliese et al. (2014) proposed a method for using Top-kriging in an index-flow framework. This method, named Total Negative Deviation Top-kriging (TNDTK; see Pugliese et al., 2014), interpolates the entire FDC (therefore ensuring its monotonicity), by reducing the dimensionality of the problem by identifying a unique index of site-specific FDCs. This is accomplished by first standardising the empirical FDCs at site x , $\Psi(x, d)$, for some reference value, $Q^*(x)$:

$$\psi(x, d) = \frac{\Psi(x, d)}{Q^*(x)} \quad (2.23)$$

where $\psi(x, d)$ denotes the dimensionless FDC, and d is a specific duration. Pugliese et al. (2014) identify an overall point index, named Total Negative Deviation (TND), that effectively summarizes the entire curve: TND is derived by integrating the area between the lower limb of the FDC and the reference streamflow value Q^* (see Fig. 2.4).

Empirical TND values are computed as:

$$TND(x) = \sum_{i=1}^m |q_i(x) - 1| \delta_i \quad (2.24)$$

where $q_i = \frac{Q_i}{Q^*}$ represents the i th empirical dimensionless quantile standardised for the selected reference value Q^* , δ_i is half of the frequency interval between the $(i + 1)$ th and $(i - 1)$ th quantile and the summation involves only the m standardised quantiles lower than 1 (i.e. negative deviation). The equality between a given streamflow value and the reference value Q^* is represented by an horizontal dashed line in Fig. 2.4, i.e. the threshold given by the equation $\frac{Q}{Q^*} = 1$. The range of the summation, m , in Eq. (2.24) is a function of the maximum duration d_{max} , which is itself a function of that sample with minimum length across gauged sites in the study region.

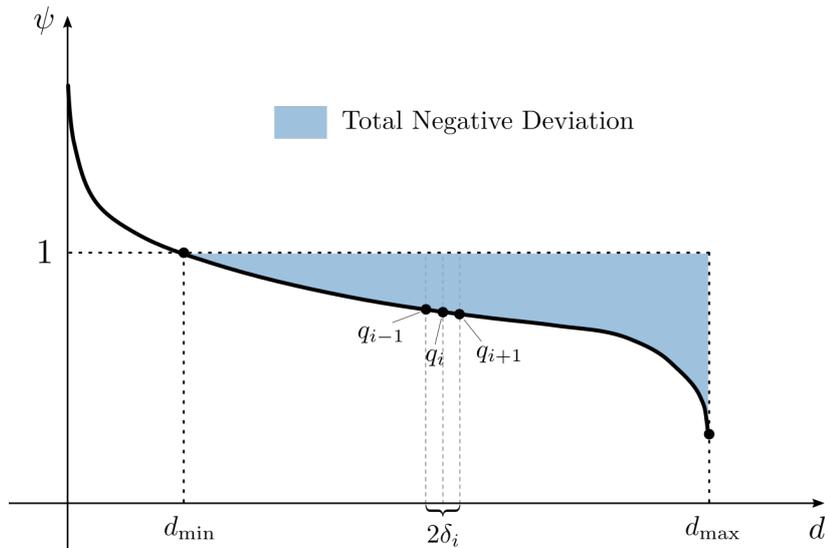


Figure 2.4: Sketch of the Total Negative Deviation (TND).

Even though TND in Fig. 2.4 does not describe the portion of the curve associated with low durations (high flows), it is very informative on the shape of the FDC, which is controlled by climatic and geomorphologic characteristics of the catchment. In particular, larger TND values are associated with catchments which are dominated by rapidly responding near-surface runoff processes (i.e. steeper FDC slopes), while smaller TNDs are associated with catchments where slower responding runoff generation processes prevail (i.e. less steep FDCs).

Having calculated empirical TNDs, Pugliese et al. (2014) use the TNDs as a regionalised variable to develop site-specific weighting schemes. The same weights, derived through the solution of the linear kriging system of Eq. (2.19), are used for a batch prediction of the continuous, dimensionless FDC for the ungauged site x_0 :

$$\hat{\psi}(x_0, d) = \sum_{j=1}^n \lambda_j \psi(x_j, d) \quad \forall d \in (0, 1) \quad (2.25)$$

where λ_j , with $j = 1, \dots, n$, are the weights resulting from the kriging interpolation of TNDs, $\psi(x_j, d)$ is the dimensionless empirical FDC at the donor site x_j , and $\hat{\psi}(x_0, d)$

is the predicted dimensionless FDC. It is worth highlighting that the computation of the linear kriging system (see Eq. (2.19)) depends on n , the number of neighbouring sites on which to base the spatial interpolation. In particular, Pugliese et al. (2014, 2016) suggest to limit to $n = 6$ the size of the kriging neighbourhood when interpolating streamflow indices, and standardised FDCs in particular.

If a reliable model for predicting Q^* at the ungauged site x_0 can be developed, the prediction of the dimensional FDC, $\hat{\Psi}(x_0, d)$, is obtained as:

$$\hat{\Psi}(x_0, d) = \hat{Q}^*(x_0) \hat{\psi}(x_0, d) \quad \forall d \in (0, 1) \quad (2.26)$$

where $\hat{Q}^*(x_0)$ is the prediction of Q^* at the ungauged site x_0 and $\hat{\psi}(x_0, d)$ has the same meaning as in Eq. (2.25).

In this regard, Pugliese et al. (2014) suggest two alternative procedures for standardising empirical FDCs (i.e. choice of Q^* ; see Eq. (2.23)): the reference value Q^* can be set equal to the mean annual streamflow (MAS; the corresponding TND value is named TND_1), or to a reference streamflow (i.e. MAP^*) equal to the catchment area A times the mean annual precipitation MAP (the corresponding TND value is named TND_2). In particular, the standardisation with MAS (i.e. TND_1) represents the traditional application of an index-flow approach to regionalise FDCs (see Castellarin et al., 2004b; Ganora et al., 2009), where an appropriate regional model for predicting the index-flow in ungauged basins (e.g. multiregression models, or Top-kriging itself) is needed. However, setting up a regional model for index-flood (i.e. MAS) is a critical and delicate step in the regionalisation procedure (see e.g. Brath et al., 2001b; Castellarin et al., 2004a), and its efficacy affects the efficacy of the overall index-flood method. For this reason, MAP^* standardisation (i.e. TND_2) can represent a valid alternative, as it simply requires catchment area to be known and MAP to be estimated, where the uncertainty associated with predictions of MAP is generally significantly smaller than the uncertainty associated with predictions of MAS for ungauged sites (i.e. large availability of raingauges and accuracy of geostatistical procedure for interpolating point observation; see e.g. Brath et al., 2001b; Castellarin et al., 2004a).

Figure 2.5 reports a scheme of the TNDTK approach applied by standardising the empirical FDCs with MAS, and estimating MAS in the ungauged basin with Top-kriging. This specific procedure, which was shown to be reliable for the prediction of FDCs in ungauged sites (see e.g. Pugliese et al., 2016), is the same which will be implemented in Chapter 5.

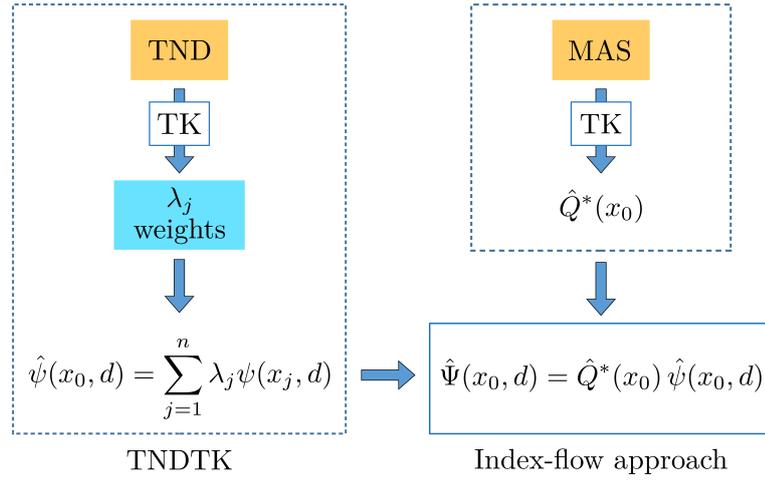


Figure 2.5: Scheme of the index-flow framework for estimating flow-duration curves with the TNDTK approach.

Further details on TNDTK (i.e. cross-validations in different geomorphological and climatic regions, sensitivity analyses, etc.) can be found, e.g., in Pugliese et al. (2014, 2016) and Kim et al. (2017). TNDTK was recently shown to be useful also for the local enhancement of macro-scale rainfall–runoff simulations (see Pugliese et al., 2018).

Chapter 3

Design-flood regionalisation: the added value of catchment similarity

3.1 Introduction

This chapter focuses on the estimation of design flood at ungauged or sparsely gauged river cross-sections in Triveneto, a broad mountainous geographical area in North-eastern Italy which counts the presence of numerous dams which routinely undergo hydrologic and hydraulic risk assessment. Given the growing concern about the possible effects of climate change (see e.g. Wilby et al., 2008; Fowler and Wilby, 2010) on flood frequency regime, Triveneto was the subject of a research agreement between the Italian *Direzione Generale Dighe del Ministero delle Infrastrutture e dei Trasporti (DG Dighe)* and the department DICAM of the University of Bologna, with the aim of verifying accuracy and reliability of the VAPI project, the reference procedure for design flood estimation in Italy (see Sec. 2.2.1 for details; see also Villi and Bacchi, 2001, for the VAPI report for Triveneto).

As the VAPI project for Triveneto is based upon AMS of peak discharges that were collected up to the 1980s, and Villi and Bacchi (2001) themselves warned about the not fully homogeneous behaviour of the unique growth curve proposed for the region, we referred to an updated AMS database in order to evaluate the potential of developing an updated reference procedure for design flood estimation in Triveneto. In particular, we considered a focused-pooling approach (i.e. Region of Influence, RoI; see Burn, 1990), which accounts for the control of geomorphological and climatic characteristics on flood frequency regime.

Part of the analyses presented in this chapter has been published in Persiano et al. (2016) and Persiano et al. (2018).

3.2 Study area and database

Triveneto is a broad geographical area in North-eastern Italy which includes the Italian Eastern Alps and consists of the administrative districts of Trentino-Alto Adige, Veneto and Friuli-Venezia Giulia. Due to its orography, the study area counts numerous dams that routinely undergo hydrologic and hydraulic risk assessments, which consist in updating the estimate of the design flood and verifying that it can safely flow over the dam spillways. The reference procedure for design flood estimation in Triveneto is available from the Italian GNDICI-CNR research project VAPI (see Sec. 2.2.1 for details), which developed an index-flood regional model based upon AMS of peak discharges that were collected up to the 1980s, mainly referring (except for some specific integrations based on more recent observations) to the AMS reported in the Publication n.17 of the former Italian National Hydrographic and Mareographic Service (SIMN, *Servizio Idrografico e Mareografico Nazionale*), which covers a period closed in 1970. Hence, the need of verifying the viability of the VAPI project for Triveneto in light of the observations collected in the last four decades.

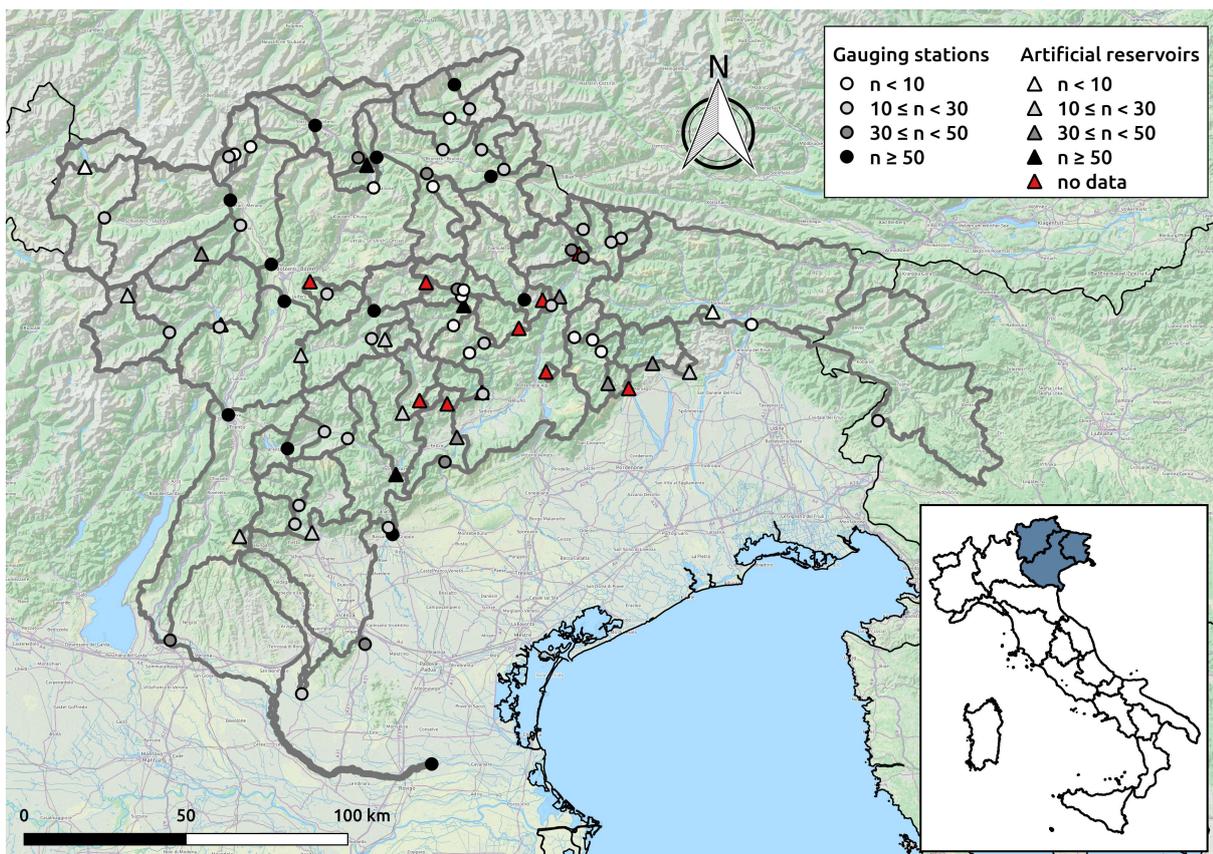


Figure 3.1: Location of the 76 Triveneto gauged sites (triangles indicate artificial reservoirs, circles the other gauging stations) considered in the study: the grey colour scale depicts the record length of the corresponding AMS series of peak flow discharges. Red triangles indicate the ungauged dams considered in Sec. 3.4.2; grey lines represent catchments boundaries.

To this aim, we compiled an updated AMS database, including historical data together with more recent data for the last 40 years. In particular, we integrated the dataset published in the Publication n.17 and in Villi and Bacchi (2001) with the AMS of peak flow discharges reported in the several technical reports about hydrologic and hydraulic risk assessment for the artificial reservoirs analysed within the research agreement with *DG Digne*. In particular, technical reports include also inflow flood series for artificial reservoirs, reconstructed by the dams' managing companies by means of the mass balance equation on the basis of outflow data. In summary, the update consisted of both (1) including more recent observations for stations already considered in the previous database, and (2) including river cross-sections which were not available before. More detailed information about the compilation of the database can be found in the reports we produced within the research agreement (see e.g. DICAM-UniBo, 2015). The resulting detailed AMS database includes 76 catchments (of which 18 associated with artificial reservoirs) located in Triveneto (see Fig. 3.1), with annual flood sequences spanning from 1913 to 2013, and minimum, mean and maximum lengths of observed AMS equal to 5, 31 and 87 years, in this order.

The consistency of the updated dataset (2433 station-years of data) is shown in Fig. 3.2, which reports the histogram of the number of station-years of data per year, highlighting the significance of the update, especially for the period after 1970 (see vertical red line in Fig. 3.2).

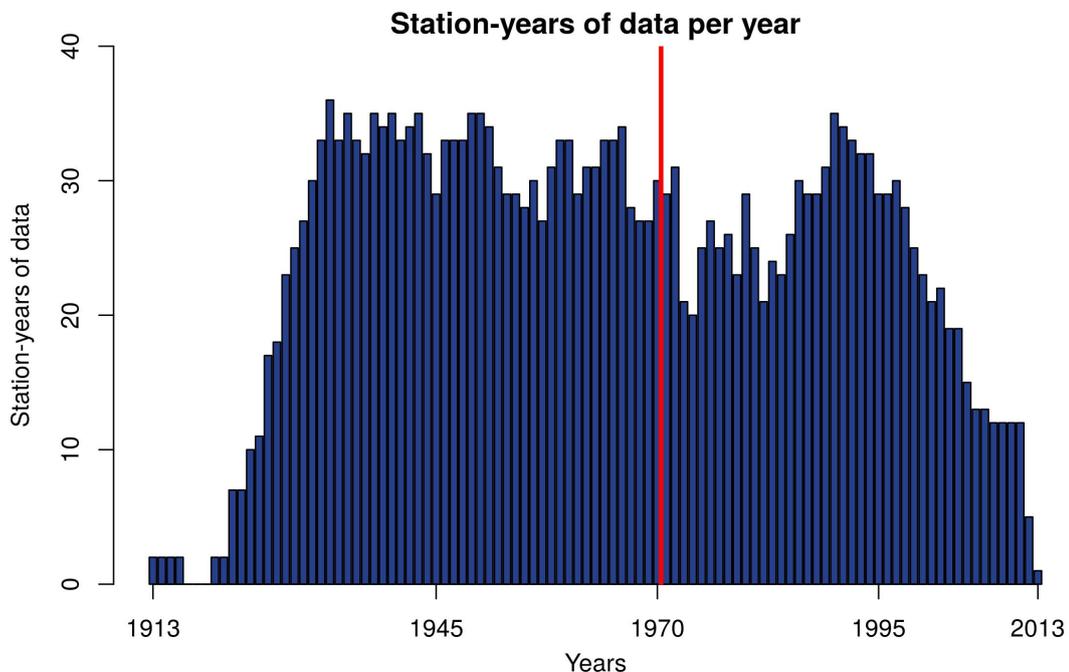


Figure 3.2: Histogram of the station-years of data per year available in the updated AMS database considered in the study.

Table 3.1 reports a summary of the dataset for the 76 gauged sites in Triveneto, including catchment descriptors, such as drainage area, mean annual precipitation (MAP), minimum and mean catchment elevation, longitude and latitude of the catchment centroid. Information about elevation were retrieved from the 10-meter resolution DEM provided by the italian National Institute of Geophysics and Volcanology (INGV, *Istituto Nazionale di Geofisica e Vulcanologia*), which was useful also for the definition of catchment boundaries, performed by using the free and open source software QGIS with the GRASS (Geographic Resources Analysis Support System) plugin. Concerning MAP, the average value for each catchment was computed from the dataset of total mean annual precipitation for the period 1960-1990 provided by the Joint Research Centre (JRC) of the European Commission in collaboration with the European Food Safety Authority (EFSA) (see Hijmans et al., 2005, for a detailed description of the dataset).

Table 3.1: Catchment characteristics for the 76 river basins included in the Triveneto dataset. For each catchment, the table provides the following information: record length of the AMS series of peak flow discharges n and its average value (i.e. mean annual flood, MAF), drainage area, mean annual precipitation (MAP), minimum and mean elevation (H_{\min} and H_{mean} , respectively), longitude X_g and latitude Y_g of the catchment centroid in the WGS84 /UTM 32 N coordinate system.

No.	Station name	n	MAF [$\text{m}^3 \text{s}^{-1}$]	Area [km^2]	MAP [mm]	H_{\min} [m a.s.l.]	H_{mean} [m a.s.l.]	X_g [m]	Y_g [m]
C01	Adige a Boara Pisani	70	857.3	11954.0	915.3	16	1498	676128.1	5143205.6
C02	Adige a Bronzolo	87	759.3	6926.0	974.8	220	1785	684017.2	5176198.5
C03	Adige a Pescantina	41	904.6	10957.0	924.6	73	1616	677109.6	5154336.4
C04	Adige a Ponte d'Adige	86	331.9	2642.0	989.1	236	1922	642951.8	5169470.9
C05	Diga di San Valentino alla Muta	8	68.9	176.0	1209.2	1487	2364	621844.7	5185761.1
C06	Adige a Spondigna	27	49.6	685.0	1074.4	884	2209	619881.1	5172496.6
C07	Adige a Tel	81	154.3	1675.0	1046.0	492	2110	630931.8	5169176.4
C08	Adige a Trento	75	1031.2	9763.0	938.8	160	1707	679463.6	5163787.4
C09	Agno-Guà a Cologna Veneta	11	148.9	260.0	880.4	17	405	681696.8	5050000.9
C10	Piave Ansiei ad Auronzo	32	40.0	206.5	998.8	833	1791	754712.5	5161189.5
C11	Astico alla Diga di Leda	27	233.7	520.0	807.2	232	1201	684324.9	5083195.7
C12	Astico a Forni di Val d'Astico	16	71.2	135.1	792.6	308	1178	679016.3	5086345.1
C13	Aurino a Cadipietra	74	51.8	159.0	1189.2	1045	2165	734389.8	5213288.9
C14	Aurino a Caminata	8	98.0	419.2	1163.6	848	2109	729658.8	5207591.7
C15	Aurino a San Giorgio	29	148.4	597.0	1129.5	819	2035	726556.0	5204291.8
C16	Avisio a Pezzè di Moena-Soraga	65	29.6	211.4	1007.9	1192	2044	710436.9	5147836.8
C17	Avisio a Stramentizzo	29	204.5	728.0	910.6	773	1829	702254.2	5135292.2
C18	Bacchiglione a Montegaldella	46	268.1	1384.0	881.2	17	636	689826.3	5068462.7
C19	Boite a Perarolo	23	81.8	398.8	979.3	552	1748	742343.8	5155481.7
C20	Boite a Vodo	74	82.9	323.0	989.8	835	1834	740057.9	5157769.6
C21	Brenta a Barziza (Bassano)	55	681.7	1567.0	854.8	107	1245	702187.7	5104110.6
C22	Brenta a Borgo Valsugana Brolo	25	22.2	218.0	820.5	376	948	678655.8	5098631.0
C23	Brenta a Levico	50	14.8	121.0	816.4	429	891	674504.8	5096645.4
C24	Brenta a Ospedaletto	26	58.2	465.0	827.1	295	1200	686082.8	5103323.5
C25	Brenta a Sarson	28	701.5	1564.4	854.2	113	1250	702134.6	5104264.6
C26	Cellina alla Diga di Barcis	46	400.8	392.0	1019.5	382	1305	770830.0	5130642.1

Continued on next page

Table 3.1 – Continued from previous page

No.	Station name	n	MAF [m ³ s ⁻¹]	Area [km ²]	MAP [mm]	H _{min} [m a.s.l.]	H _{mean} [m a.s.l.]	X _g [m]	Y _g [m]
C27	Cellina a Stich Lesis	5	45.3	41.2	1012.4	652	1421	776343.9	5132036.7
C28	Cimoliana a Cimolais	6	116.6	83.0	1003.1	687	1666	767876.9	5139652.7
C29	Cismon alla Diga del Corlo	71	261.8	628.0	877.9	253	1377	712574.9	5113386.4
C30	Cordevole alla Diga di Alleghe	68	66.9	246.8	966.6	965	1861	728058.5	5149323.4
C31	Cordevole a Caprile	17	33.1	223.7	972.2	977	1893	727547.0	5149893.0
C32	Cordevole a Ponte Alto	8	149.0	571.2	945.4	557	1747	726998.1	5141140.3
C33	Cordevole a Ponte Ghirlo	5	85.2	415.3	951.4	750	1815	725831.0	5144624.9
C34	Cordevole a Saviner	42	26.8	109.7	982.5	1028	1911	724942.0	5152676.0
C35	Diga di Tul	24	65.3	24.0	1206.1	282	711	798772.6	5127717.7
C36	Ega a Ponte Nova	17	11.3	118.8	826.3	865	1639	694208.0	5141631.8
C37	Fiorentina a Sottorovei	9	22.1	56.4	972.9	1073	1856	735600.9	5148893.3
C38	Gadera a Mantana	38	66.0	384.8	979.5	938	1878	723385.3	5168274.8
C39	Isarco a Bressanone	9	158.9	740.6	998.8	562	1819	687473.0	5195174.5
C40	Isarco a Fortezza	55	140.6	680.0	1011.7	720	1863	686215.3	5196921.7
C41	Isarco a Prà di Sopra	32	104.5	652.0	1015.1	764	1883	685956.0	5197280.5
C42	Isonzo a Canale	10	1332.0	1357.0	1298.2	98	884	867660.8	5126464.8
C43	Diga di Speccheri	12	18.1	13.8	776.5	743	1338	668099.3	5070253.2
C44	Diga di Ponte Racli	40	454.5	217.6	1115.9	303	952	788075.1	5132709.0
C45	Mis a Ponte Sant'Antonio	12	86.6	114.0	891.8	363	1292	731080.7	5120310.6
C46	Noce a Ponte Rovina	21	69.2	383.3	958.9	784	2140	629769.1	5131720.9
C47	Noce a Tassullo-Dermulo	23	156.5	1056.0	856.7	401	1763	644171.0	5137403.1
C48	Piave Padola a Ponte Padola	22	11.2	58.5	1049.6	1194	1834	763959.7	5171199.1
C49	Passirio a Belprato	7	49.1	53.8	1134.9	1547	2369	658266.0	5182335.7
C50	Passirio a Moso	6	10.4	181.0	1104.8	917	2245	661797.3	5189212.5
C51	Piave Busche traversa	34	979.0	3224.1	976.4	232	1414	747915.6	5138899.5
C52	Piave a Pieve di Cadore (diga)	49	287.1	818.0	1013.7	649	1646	765728.5	5160964.8
C53	Piave a Ponte Cordevole	19	20.6	63.0	1061.1	1023	1679	781818.1	5163964.1
C54	Piave a Ponte della Lasta	43	104.7	357.0	1043.0	805	1669	773660.0	5166521.0
C55	Piave a Presenaio	28	47.4	142.0	1061.7	951	1691	779893.8	5167131.1
C56	Piave Segusino	48	913.3	3333.0	971.7	195	1370	746101.2	5135945.0
C57	Plan a Plan	14	12.1	44.0	1145.3	1596	2413	657638.9	5181891.9
C58	Posina a Stancari	13	94.0	116.0	807.9	299	1102	676214.8	5076870.4
C59	Ridanna a Vipiteno	50	70.1	206.0	1010.5	939	1941	675749.9	5195334.5
C60	Rienza a Monguelfo	74	19.3	273.0	1013.8	1099	1879	743619.6	5174924.5
C61	Rienza a Vandoies	79	191.8	1923.0	1034.5	720	1863	728237.9	5188109.7
C62	Rio Anterselva a Bagni Salomone	25	14.0	83.2	1112.8	1089	2031	738426.0	5194645.4
C63	Rio Casies a Colle	28	16.3	116.6	1084.3	1196	1962	745717.4	5190657.4
C64	Rio Missiaga	19	3.4	4.7	938.0	1081	1710	739951.0	5131928.1
C65	Rio Riva a Seghe di Riva	28	53.0	91.2	1278.6	1525	2387	735218.5	5203242.9
C66	San Vigilio a Longega	6	5.6	104.0	1004.9	1025	1904	727927.5	5172589.5
C67	Tagliamento a Pioverno	9	1472.8	1884.8	1175.3	227	1163	809953.4	5153675.7
C68	Travignolo a Sottosassa	15	66.0	103.0	941.1	1197	1967	711039.0	5132116.1
C69	Val Settimana a Stalli Nucci	5	47.5	51.6	1016.4	688	1492	774895.3	5137461.3
C70	Rio Valsura a Lana	24	49.5	282.0	888.8	385	1913	649618.3	5154803.0
C71	Diga di Ambiesta	6	33.1	9.1	1190.9	470	868	803884.7	5144004.9
C72	Diga di Forte Buso	22	68.1	66.3	964.1	1407	2019	712909.5	5132574.4
C73	Diga di Santa Giustina	21	257.3	1050.0	854.8	514	1765	644187.1	5137467.2
C74	Diga di Val Schener	17	142.8	203.0	885.1	538	1513	718720.5	5119203.4
C75	Diga di Careser	25	7.7	10.4	1469.9	2535	2968	631232.1	5144698.7
C76	Diga di Zoccolo	34	29.3	181.2	940.2	1112	2160	644948.5	5151708.4

3.3 Test of the accuracy of the VAPI project

As introduced in Sec. 2.2.1, Villi and Bacchi (2001) regarded Triveneto (i.e. the former SIMN compartment of Venezia) as a unique homogeneous region in terms of flood frequency regime, for which they proposed a unique TCEV growth curve (see also Tab. 2.1) for operational purposes, warning at the same time about the possible presence of a not fully homogeneous behaviour of the curve itself. In light of this consideration, we tested the viability of the VAPI growth curve for the study region, referring to the updated AMS database compiled in the present study.

In order to assess the reliability of the growth curve proposed by Villi and Bacchi (2001), the most straightforward way was a visual comparison between the TCEV growth curve itself and the empirical regional sample contained in the updated AMS database (see Fig. 3.3). An empirical regional sample consists of all the dimensionless (i.e. standardised with the corresponding MAF) AMS available for the study area in a selected period; to allow the graphical representation, the empirical regional sample is ranked in ascending order and each value is plotted versus the corresponding empirical non-exceedance probability computed with the Weibull plotting position (see also Sec 1.1.1). In particular, we referred to different empirical regional samples by selecting three different periods: the overall period available in the updated dataset (i.e. 1913-2013; 2433 station-years of data), and the two sub-samples PRE1970 (all years up to 1970, included; i.e. 1913-1970; 1421 station-years of data) and POST1970 (all years after 1970; i.e. 1971-2013; 983 station-years of data). As PRE1970 sub-sample includes also stations which were not considered in the VAPI project, we also referred to a PRE1970-VAPI subset (587 station-years of data), including only the stations which were actually considered in Villi and Bacchi (2001), while taking into account that for the period before 1970 the updated dataset includes some additional data also for stations already considered in the VAPI. To guarantee the reliability of the analysis, sub-records having record length lower than 5 were escluded from each sub-sample (i.e. PRE1970, PRE1970-VAPI, POST1970).

Figure 3.3 shows that the total sample and PRE1970 differ considerably from the VAPI growth curve. This behaviour can be explained by considering that the AMS database includes much more stations and observations than the one considered in Villi and Bacchi (2001). Indeed, if we limit the analysis to the PRE1970-VAPI, we can observe a good overlap with the VAPI growth curve, except for the three highest values in the sub-sample. At the same time, an unexpected behaviour is noticed with reference to POST1970, which results in overlapping with the VAPI growth curve. In this context, the analyses about changes in flood frequency regime for Triveneto reported in Persiano et al. (2018) (i.e. application of non-parametric stastical tests with 5% significance level to inspect possible

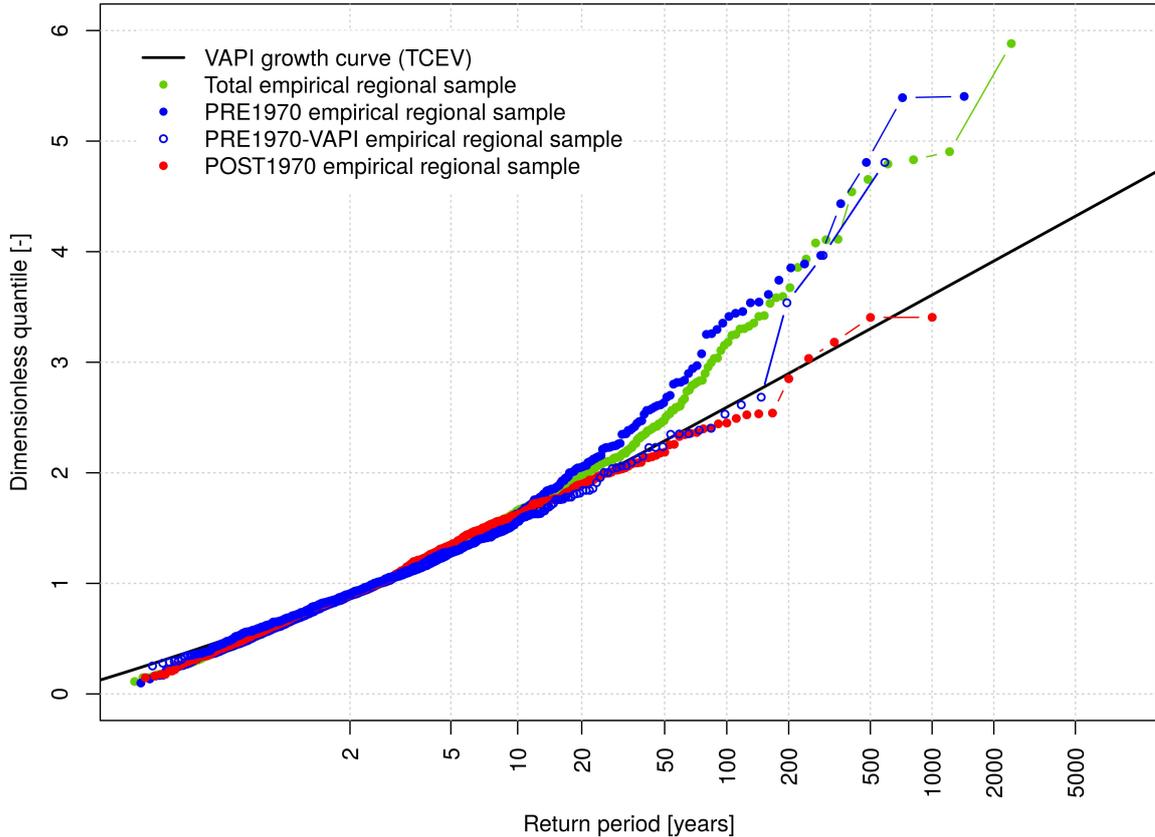


Figure 3.3: Comparison between the regional growth curve (TCEV distribution, black curve) proposed for Triveneto by Villi and Bacchi (2001) and the (dimensionless) empirical regional samples collected for the overall period (i.e. 1913-2013; green dots) and for three different sub-periods: PRE1970 (i.e. 1913-1970, blue dots), PRE1970-VAPI (i.e. 1913-1970 for only the stations considered in the VAPI; blue circles) and POST1970 (i.e. 1971-2013; red dots).

trends as well as abrupt changes in the intensity of annual maximum floods) highlight the spatial heterogeneity of the few statistically significant changes detected, which seem not to be related to climate change. These resultings suggest the need of reconsidering the VAPI growth curve proposed by Villi and Bacchi (2001).

As homogeneity is a fundamental requirement to guarantee the reliability of the regionalisation procedure, a further step was the assessment of the actual heterogeneity degree for each sample. To this aim, we referred to the Hosking and Wallis test (see Hosking and Wallis, 1993), which suggests to regard a group of sites as "acceptably homogeneous" if $H_k < 1$, "possibly heterogeneous" if $1 \leq H_k < 2$, and "definitely heterogeneous" if $H_k \geq 2$, where H_k ($k = 1, 2, 3$) are measures of dispersion for different orders of the sample L-moment ratios (see also Hosking, 1990, and Appendix A). In particular, we considered H_1 (related to the L-coefficient of variation) and H_2 (related to the sum of L-coefficient of variation and L-coefficient of skewness). The test indicates for the study region a high

heterogeneity degree for the overall Triveneto sample ($H_1 \simeq 10$ and $H_2 \simeq 5$), and also for the three sub-samples shown in Fig. 3.3 ($H_1 \simeq 4$ and $H_2 \simeq 3$ for PRE1970, $H_1 \simeq 3$ and $H_2 \simeq 2$ for PRE1970-VAPI, $H_1 \simeq 9$ and $H_2 \simeq 3$ for POST1970). Therefore, even the POST1970 sub-sample, which overlaps the VAPI growth curve, is not reliable for stating the validity of the VAPI growth curve after 1970.

In summary, the analyses highlight that the dimensionless growth curve in Fig. 3.3 does not refer to a homogeneous regional sample: although the VAPI project considered as a single homogeneous region for operational purposes, the significant update done in this Thesis confirms the warning of Villi and Bacchi (2001) about the possible presence of a not fully homogeneous behaviour for the study area, showing that the hypothesis of homogeneity does not hold for Triveneto.

3.4 Test of the viability of a focused-pooling approach

The analyses performed in the previous section highlighted the need of updating the reference procedure for design flood estimation in Triveneto (i.e. VAPI project), founded on geographically contiguous homogeneous regions. Given the high heterogeneity degree for the study area, we tested the viability of a focused-pooling approach (i.e. RoI approach; see Burn, 1990; Zrinji and Burn, 1994, 1996), useful for delineating pooling-groups of sites for a given target site accounting for climatic and geomorphologic controls on flood frequency regime. In order to properly select the catchment characteristics to be considered, a preliminary step was the inspection of climatic and scale controls on flood frequency regime in Triveneto, checking their consistency with the controls that were recently found in Europe (see Salinas et al., 2014).

3.4.1 Climate and scale controls on regional flood frequency distribution

The selection of the most suitable regional parent distribution is a key aspect in any regional flood frequency analysis. The scientific literature recommends using the L-moment ratio diagrams for addressing this task (see e.g. Hosking and Wallis, 1993, 1997). Because of the unavailability or high uncertainty of sample statistics for ungauged or poorly gauged regions, many studies have focused on the relationships between sample L-moments and catchment descriptors. In particular, a recent study by Salinas et al. (2014) on a data set of annual maximum series (AMS) of peak flow from a total of 813 catchments (in Austria, Italy and Slovakia) shows the great importance of incorporating into regional models information on mean annual precipitation (MAP) and basin area as

surrogates of climate and scale controls (see also Blöschl and Sivapalan, 1997; Padi et al., 2011).

In this context, we inspected climatic and scale controls on flood frequency regime in Triveneto, checking whether they are similar to the controls that were recently found in Europe by Salinas et al. (2014). To corroborate the significance of this analysis, it is important to highlight that the catchments included in the AMS database considered in this Thesis have geomorphologic and climatic characteristics similar to the catchments considered in Salinas et al. (2014), in whose study, however, Triveneto is represented only by a limited number of catchments (less than 10). To reduce the effects of sampling variability when estimating higher order L-moments (see e.g. Viglione, 2010), the information from each site was weighted proportionally to the site record length (Hosking and Wallis, 1993) during all regionalisation phases. For comparison purposes, the data set was divided into the same six subsets identified in Salinas et al. (2014): smaller (area $< 55 \text{ km}^2$), intermediate ($55 \text{ km}^2 \leq \text{area} \leq 730 \text{ km}^2$) and larger (area $> 730 \text{ km}^2$) catchments, drier (MAP $< 860 \text{ mm yr}^{-1}$), medium ($860 \text{ mm yr}^{-1} \leq \text{MAP} \leq 1420 \text{ mm yr}^{-1}$) and wetter (MAP $> 1420 \text{ mm yr}^{-1}$) catchments. It is evident that the majority of the Triveneto catchments (black points in Fig. 3.4) belong to the central subset.

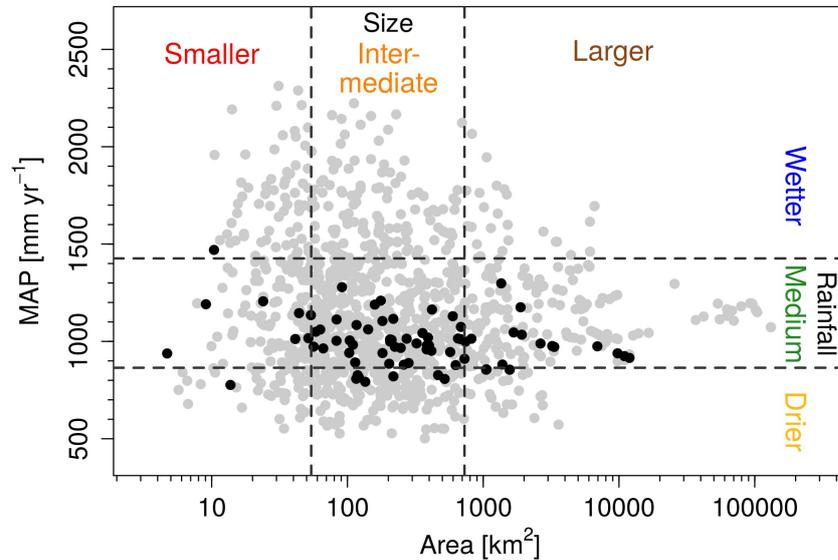


Figure 3.4: Catchment characteristics of the Triveneto data set (black points). Grey points represent the Austrian, Slovakian and Italian data sets considered in Salinas et al. (2014). Black-dashed lines identify the subdivision into the six subsets based on the 20% and 80% quantiles of the catchments descriptor values for the European data sets.

Therefore, in order to verify the impact of catchment size and MAP on the L-moment ratios for the Triveneto regional parent distribution, we considered only the two subsets associated with intermediate site and medium MAP. For each one of them, the two L-moment ratio diagrams L-Cv–L-Cs and L-Ck–L-Cs were plotted, including the theoret-

ical lines for the most common two- and three-parameter distributions, respectively (see Fig. 3.5). Similarly to Salinas et al. (2014), the points drawn in the diagrams represent the record length weighted moving average (WMA) values of the corresponding sample L-moments and their colour intensity is proportional to the mean value of the descriptor not used for the stratification. WMA was computed for a window including the 35 most similar catchments in terms of the evaluated descriptor (i.e. area or MAP).

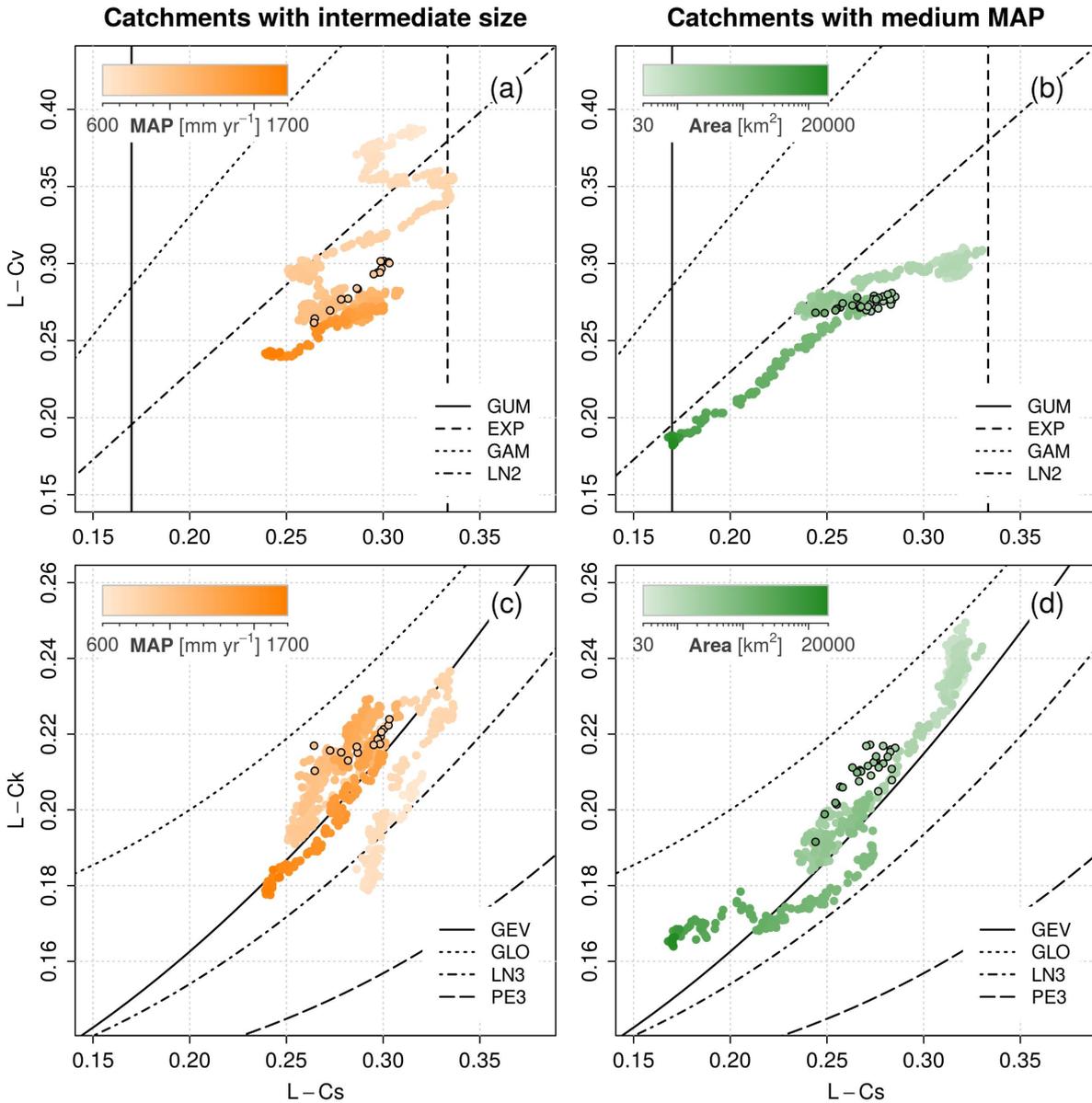


Figure 3.5: L-moment ratio diagrams for the subsets defined by intermediate area (a, c) and medium MAP (b, d). Each point represents a record length weighted moving average (WMA) of L-Cv (a, b) or L-Ck (c, d) against the corresponding values of L-Cs and the colour intensity is proportional to the catchment descriptor of interest (area or MAP); WMAs are computed for either 35 Triveneto basins (black-border) or 70 European basins (no-border) having similar area or MAP value.

Figure 3.5 clearly shows the superimposition of Triveneto WMAs to WMAs obtained by Salinas et al. (2014). Figures 3.5(a) and 3.5(b) confirm that sample L-Cv and L-Cs are poorly described by the most common two-parameter distributions. On the other hand, Fig. 3.5(c) shows how medium-sized (intermediate) catchments associate with high to medium MAP are well described by the Generalized Extreme Value (GEV) distribution, and Fig. 3.5(d) clearly reports the same behaviour for medium MAP catchments. The strong control of area and MAP on regional L-moments (i.e. WMAs) is also confirmed by the intensity of gradation, which increases with decreasing L-moment ratios values for all the considered cases: average L-moment ratio values are higher for smaller than larger catchments and for drier than wetter ones. In other words, the regional flood frequency regime of medium-sized catchments associated with high to medium MAP is well described by the GEV distribution and average L-moment ratio values are higher for smaller than larger catchments and for drier than wetter ones.

3.4.2 Application of the Region of Influence approach

Once proved the influence of climatic and scale characteristics on the regional flood frequency regime in Triveneto, we tested and applied for the study area the Region of Influence (RoI) approach (see Burn, 1990; Zrinji and Burn, 1994, 1996, see also Sec. 2.2.2 for a detailed description of the approach), which replaces the idea of homogeneous regions consisting of contiguous and geographically identifiable regions with the most general idea of homogeneous groups of basins with similar hydrological behaviour, which may or may not be geographically close to each other. The RoI approach delineates homogeneous pooling-groups for a given target site referring to a minimum amount of information in terms of climatic and geomorphological descriptors of the basins that strongly influence the flood frequency regime at regional scale. In this study, we considered the following descriptors: area and MAP (whose strong influence on the flood frequency regime has been confirmed for the study area, see Sec. 3.4.1), mean and minimum elevation, and latitude and longitude of catchment centroid. In particular, it is very important to consider catchment elevation as a descriptor of the orographic effect on the flood frequency regime (see e.g. Allamano et al., 2009). The Euclidean distance in the 6-dimensional descriptor space was used as measure of hydrological dissimilarity (see Eq. (2.16)), where the same weight has been assigned to each of the mentioned descriptors.

The RoI approach arranges catchments according to their dissimilarity with the target site, pooling together only the most similar ones. The number of catchments to include in the RoI is determined by considering the return period of the design flood T : according to the $5T$ -rule proposed by Jakob et al. (1999), a number of station-years of annual

maxima $n \geq 5T$ is required to get a reliable estimation of the T -year flood (T -quantile). Therefore the RoI sample consists of a pooling-group of catchments that are the most hydrologically similar to the target site and whose overall station-years of data sums up to $5T$ (i.e. 500 station-years of data for $T = 100$ years). In the sequential addition of new sites to the RoI, a homogeneity test is necessary for assessing the heterogeneity degree of the pooling-group. To this aim, we referred to the Hosking and Wallis test (Hosking and Wallis, 1993), consistently with what suggested in Zrinji and Burn (1996) and Castellarin et al. (2001). In particular, Zrinji and Burn (1994) and Castellarin et al. (2001) (see Sec. 2.2.2 for further details) allow to discard the first site which results in an unacceptable heterogeneity of the overall pooling-group. In the present study, given the high heterogeneity degree found for Triveneto (i.e. $H_1 \simeq 10$, see Sec. 3.3), to avoid a premature ending of the process, we decided to allow to discard more than one site, whenever a significant increase in the heterogeneity degree was detected. Moreover, for the same reason, in this context we accepted pooling-groups characterised by H_1 values up to 2.5, but only if associated with $H_2 < 2$. After having identified the RoI pooling-group of sites for the given target site, the L-moment approach proposed by Hosking and Wallis (1993, 1997) (see also Appendix A.3) was applied in order to select the most suitable regional parent distribution and the corresponding dimensionless T -year flood quantile.

The RoI approach was indeed applied within an index-flood framework (see Sec. 1.2.3): the T -year design flood was computed as the product between the dimensionless T -year flood quantile (evaluated with L-moment approach proposed by Hosking and Wallis, 1993, 1997, on the basis of the homogeneous pooling-group identified with the RoI approach) and the index-flood. The index-flood was computed as the average value of the AMS available for the target site if at least 5 observations were available; otherwise multi-regression models (Brath et al., 2001a, see also Sec. 1.2.3) can be built by referring to the RoI pooling-group itself. When the identification of an effective regression model is not possible, scaling laws available in the literature for the study area or the drainage-area ratio method can be used (see Villi and Bacchi, 2001 for Triveneto; see also Farmer and Vogel, 2013; Farmer et al., 2015).

Within the research agreement with *DG Dighe*, we applied the RoI approach to estimate the design flood for several artificial reservoirs (both gauged and ungauged) in Triveneto. We report herein an application example for the Pieve di Cadore dam (*Diga di Pieve di Cadore*, see station C52 in Tab. 3.1 and reservoir R24 in Tab. 3.2), for which a 49 AMS (from 1951 to 1999) has been compiled from inflow series to the artificial reservoir. The RoI sample consists of 16 catchments, for a total of 567 observations, which enable one to get a reliable prediction of the 100-year quantile according to the $5T$ -rule. We

can observe the significant advantage of the RoI approach in terms of homogeneity: the heterogeneity associated with the RoI sample ($H_1 \simeq 1.26$) is much lower than that of the total Triveneto sample ($H_1 \simeq 10$). Therefore the corresponding RoI growth curve (green line in in Fig. 3.6) is characteristic of a regional parent distribution that is suitable for predicting flood quantiles up to $T = 100$ years for Pieve di Cadore. Figure 3.6 also highlights the significant difference with the VAPI growth curve (black line in in Fig. 3.6), which, as well as being associated with a high heterogeneity degree, results in underestimating the design flood for $T > 10$ years if compared to the more reliable RoI growth curve. Note that Fig. 3.6 reports the dimensional growth curves for Pieve di Cadore, which were obtained by multiplying the dimensionless growth curves for the index-flood estimated as the mean value ($287.1 \text{ m}^3 \text{ s}^{-1}$) of the 49 AMS available for the target site.

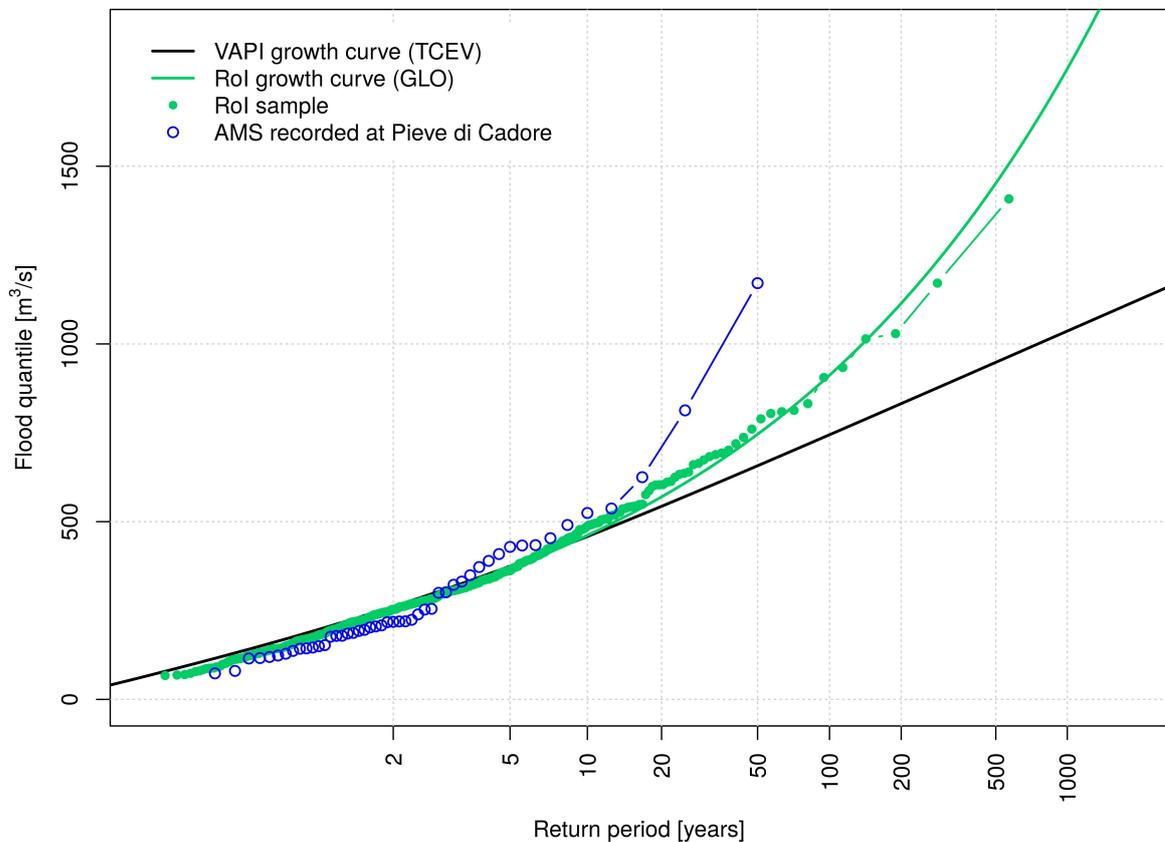


Figure 3.6: Comparison between the dimensional growth curve proposed by the VAPI project (black line) and the dimensional growth curve associated with the RoI pooling-group of sites (green line) for Pieve di Cadore. The choice of the GLO distribution for the RoI sample was made by applying the L-moment approach proposed by Hosking and Wallis (see Hosking and Wallis, 1993, 1997). Green dots and blue circles indicate the empirical growth curve for the RoI sample and for the AMS recorded at Pieve di Cadore, respectively.

The application of the RoI approach for the 24 artificial reservoirs in Triveneto which were analysed within the research agreement with *DG Dighe* is summarised in Tab. 3.2,

which reports the number of station-years of data included in the RoI sample (n_{RoI}) and values of H_1 and H_2 , together with the catchment descriptors considered for the identification of the RoI pooling-groups. The artificial reservoirs analysed in the study included both gauged and ungauged cases; for gauged cases, information about the AMS available for the target site is reported in Tab. 3.1.

Table 3.2: Artificial reservoirs for which the RoI approach was applied to estimate the 100-year flood. For each considered dam, the catchment characteristics used for the definition of the RoI pooling-group are reported together with the number of station-years of data included in the RoI sample n_{RoI} and the corresponding heterogeneity measures H_1 and H_2 .

No.	Station name	n_{RoI}	H_1	H_2	Area [km ²]	MAP [mm]	H_{\min} [m a.s.l.]	H_{mean} [m a.s.l.]	X_g [m]	Y_g [m]
R01	Astico alla Diga di Leda	538	1.75	0.06	520.0	807.2	232	1200	684324.9	5083195.7
R02	Avisio a Stramentizzo	521	2.30	0.59	728.0	910.6	773	1831	702254.2	5135292.2
R03	Caorame (Piave) a La Stua (diga)	501	2.10	0.44	27.5	903.9	699	1569	728781.3	5114947.9
R04	Cellina a Ravedis (diga)	541	0.76	-0.21	445.0	1031.4	292	1255	771768.5	5129679.6
R05	Cellina alla Diga di Barcis	499	1.97	0.70	392.0	1019.5	382	1304	770830.0	5130642.1
R06	Cismon alla Diga del Corlo	534	1.95	0.09	628.0	877.9	253	1375	712574.9	5113386.4
R07	Diga di Ambiesta	507	1.44	0.99	9.1	1190.9	463	870	803884.7	5144004.9
R08	Diga di Fedai	442	1.46	0.31	8.2	1160.5	2028	2454	720633.6	5148370.3
R09	Diga di Forte Buso	508	1.83	0.52	66.3	964.1	1407	2021	712909.5	5132574.4
R10	Diga di Mis	519	1.19	-0.45	108.0	890.3	398	1297	731123.2	5120213.2
R11	Diga di Pontesei	542	1.83	0.47	151.6	919.4	744	1590	742322.7	5139052.6
R12	Diga di San Valentino alla Muta	535	1.69	-0.51	176.0	1209.2	1487	2365	621844.7	5185761.1
R13	Diga di Santa Caterina	550	1.14	0.51	225.0	993.8	814	1750	755652.9	5161193.6
R14	Diga di Santa Giustina	529	2.21	0.61	1050.0	854.8	514	1765	644187.1	5137467.2
R15	Diga di Speccheri	512	2.48	0.69	13.8	776.5	764	1343	668099.3	5070253.2
R16	Diga di Tul	441	2.10	1.70	24.0	1206.1	272	706	798772.6	5127717.7
R17	Diga di Val d'Ega	515	2.22	0.15	154.0	812.5	544	1539	692656.6	5142657.7
R18	Diga di Val Gallina	480	2.47	0.27	14.4	949.0	646	1246	758291.5	5123578.4
R19	Diga di Val Noana	529	1.92	-0.18	31.1	878.3	969	1545	723109.4	5114872.1
R20	Diga di Val Schener	747	0.52	-0.78	203.0	885.1	538	1512	718720.5	5119203.4
R21	Diga di Valle di Cadore	510	1.74	0.81	380.2	981.7	699	1775	741768.5	5155916.9
R22	Diga Ponte Racli	441	2.12	1.70	217.6	1115.9	303	949	788075.1	5132709.0
R23	Isarco a Fortezza	477	1.59	1.75	680.0	1011.7	720	1862	686215.3	5196921.7
R24	Piave a Pieve di Cadore (diga)	567	1.26	1.05	818.0	1013.7	649	1645	765728.5	5160964.8

3.5 Results and discussion

As shown in Sec. 3.3, our update of Triveneto AMS dataset showed that the TCEV growth curve proposed by Villi and Bacchi (2001) in the VAPI project for Triveneto needs to be reconsidered, as it is not properly representative of the updated empirical regional

sample (see Fig. 3.3). Moreover our analyses falsified the main assumption of the VAPI project for Triveneto, showing the high heterogeneity degree (i.e. $H_1 \simeq 10$) for the study region, which cannot be regarded as homogeneous in terms of flood frequency regime.

The analyses reported in Sec. 3.4.1 confirmed the value of including physiographic and climatic information, such as drainage area and mean annual precipitation (MAP), in statistical regionalisation. In particular, climatic and scale controls on flood frequency regime in Triveneto resulted to be similar to the controls that were recently found in Europe by Salinas et al. (2014): the regional flood frequency regime of medium-sized catchments associated with high to medium MAP is well described by the GEV distribution and average L-moment ratio values are higher for smaller than larger catchments and for drier than wetter ones.

Once proved the influence of climatic and scale characteristics on the regional flood frequency regime in the study area, we tested the Region of Influence (RoI) approach as possible candidate for updating the reference procedure for design flood estimation in Triveneto. Sec. 3.4.2 provides a practical example for Pieve di Cadore (see Fig. 3.6), for which the RoI approach has led to a pooling-group of sites with a number of station-years of data ($n_{RoI} = 567$) suitable for predicting flood quantiles up to $T = 100$ years, and a much smaller heterogeneity ($H_1 \simeq 1.26$ and $H_2 \simeq 1.05$) than the overall regional sample available for the entire study area ($H_1 \simeq 10$ and $H_2 \simeq 5$). The same behaviour was detected for all the 24 artificial reservoirs analysed in Triveneto within the research agreement with *DG Dighe*. Table 3.1 reports H_1 values spanning from 0.52 to 2.48, H_2 from -0.78 to 1.75, and n_{RoI} from 441 to 747. The best case is represented by *Diga di Val Schener* (reservoir R20; $n_{RoI} = 747$, $H_1 \simeq 0.52$, $H_2 \simeq -0.78$), which resulted in being "acceptably homogeneous" in the strict sense according to the Hosking and Wallis test, and which will be used as reference homogeneous region for the analyses performed in Chapter 4. Even if in many cases the identified RoI pooling-groups reported in Tab. 3.1 cannot be regarded as homogeneous in the strict sense according to the Hosking and Wallis test (see e.g. artificial reservoirs R02, R03, etc.), the improvement in terms of heterogeneity degree is significant if compared to the one of the total Triveneto sample (i.e. $H_1 \simeq 10$). These findings prove the effectiveness of the RoI approach in estimating the T -year flood the study area, highlighting that focused-pooling approaches accounting for the control of geomorphological and climatic characteristics on flood frequency regime are preferable to approaches based on geographically contiguous regions, especially in highly heterogeneous areas. In particular, the RoI approach represents a valid candidate for updating the reference procedure for design flood estimation in Triveneto.

3.6 Concluding remarks

This chapter focused on the first research question reported in Sec. 1.3. The need of updating the reference procedure for design flood estimation in Triveneto (i.e. VAPI project, see Villi and Bacchi, 2001) was highlighted: the updated AMS dataset for the study region showed that Triveneto is not homogeneous in terms of flood frequency regime, falsifying the main assumption of the VAPI project.

Therefore, we developed an updated reference procedure by using the Region of Influence (RoI) approach, which delineates homogeneous pooling-groups of sites for any given target site referring to selected climatic and geomorphological descriptors which were shown to be particularly relevant for describing regional flood frequency. In particular, catchment descriptors as mean annual precipitation (MAP), basin area, mean and minimum elevation and catchment location can be used as surrogates of climate and scale controls. We observed that the regional samples obtained through the RoI approach are characterised by homogeneity degrees which are much higher than the Triveneto one. Focused-pooling approaches are therefore preferable to approaches based on geographically contiguous regions, especially in highly heterogeneous study areas. This behaviour confirms the added value of explicitly accounting for catchment similarity (i.e. control of geomorphological and climatic characteristics on flood frequency regime) in statistical regionalisation.

Chapter 4

Design-flood regionalisation: the impact of spatial correlation

4.1 Introduction

While Chapter 3 highlighted the added value of including physiographic and climatic information for properly identifying hydrologically homogeneous regions, the present chapter addresses another fundamental issue in regional flood frequency analysis: the impact of spatial correlation among concurrent flood flows on regional predictions. As illustrated in Sec. 1.2.2, the effect of spatial correlation in regional flood frequency analysis has been widely investigated in the literature (see e.g. Matalas and Langbein, 1962; Stedinger, 1983; Hosking and Wallis, 1988; Stedinger and Lu, 1995; Hosking and Wallis, 1997; Madsen and Rosbjerg, 1997; Madsen et al., 2002; Castellarin et al., 2008), and mainly concerns the reduction of the actual hydrological information content of the region, impacting the prediction uncertainty of the classical index-flood approaches (not their bias) and the assessment of the homogeneity of the region itself.

For this reason, the presence of spatial correlation is a very important issue when predicting flood quantiles in ungauged basins, and the literature itself reports statistical methods which have completely different ways of treating cross-correlation among flows. For instance, Generalized Least Squares (GLS; i.e. Stedinger and Tasker, 1985; Tasker and Stedinger, 1989, see also Sec. 2.1.3), which is the reference procedure for estimating streamflow characteristics in ungauged catchments in the USA, accounts for sampling variability and cross-correlation among concurrent streamflows in developing a regional (multi-)regression model. On the other hand, in the last decade geostatistical approaches which explicitly exploit spatial correlation (e.g. Canonical kriging, CK, see Ouarda et al., 2001; Chokmani and Ouarda, 2004; Top-kriging, TK, see Skøien et al., 2006, see also

Sec. 2.3.1) have been shown to be effective for predicting several streamflow indices and hydrological signatures in ungauged sites (see e.g. Castiglioni et al., 2009; Archfield et al., 2013; Pugliese et al., 2014, 2016).

In this context, a recent study by Archfield et al. (2013) compared the performances of GLS, CK and TK in predicting empirical estimators of flood quantiles in ungauged sites for a set of 61 gauged basins located in the Flint River basin in the South-eastern USA. Their study highlighted that when the aim is the prediction of empirical flood quantiles in an ungauged site, TK is likely to result in better predictive models than GLS. Nevertheless, Archfield et al. (2013) also pointed out that, being entirely based on empirical data, their analysis cannot address the fundamental science problem of understanding which technique is better suited for predicting the true unknown flood quantiles in ungauged catchments when the observed flood sequences are affected by cross-correlation: for a limited set of cross-correlated streamflow observations, the true flooding potential at a given site could be masked. Despite this aspect is crucial, its impact on the accuracy of flood quantiles predictions has never been formally investigated in the literature.

In this context, this chapter investigates the impact of spatial correlation in flood data on the prediction accuracy of both GLS and TK, the analysis of which is of particular interest as neither of the two procedures are commonly used in Italy. To this aim, we referred to a simplified situation consisting of a homogeneous region with nested catchments: we considered a dataset of 20 AMS of peak flow discharges collected in Triveneto which may be regarded as possibly homogeneous in terms of flood frequency regime (see Hosking and Wallis, 1997; Castellarin et al., 2008). Considering this regional dataset, we first repeated the same exercise performed in Archfield et al. (2013), to see to what extent the results obtained then hold here. Second, based on the empirical characteristics of this regional dataset (i.e. mainly regional flood frequency regime and spatial correlation structure), we designed a Monte Carlo simulation framework to generate 1000 realisations of the homogeneous region for three different levels of regional cross-correlation (i.e. 3000 realisations in total). For each realisation of the region and level of cross-correlation, we applied GLS and TK to obtain predictions of at-site flood quantiles for return periods T equal to 10, 30, 50 and 100 years in a leave-one-out cross-validation (LOOCV) scheme. The performances of cross-validated GLS and TK were then quantified relative to the prediction of (a) known theoretical values and (b) sample estimates of flood quantiles at each and every site in the realisations. The objective was to understand, between the two different ways of incorporating information on the cross-correlation structure of the data featured by GLS and TK, which is the most effective one for estimating (a) the true statistic of interest and (b) its empirical estimator. In particular, from the knowledge of

the structure of the models (see Sec. 2.1.3 for GLS, and Sec. 2.3.1 for TK), we expect TK to have better efficiencies in predicting the empirical estimator of the flood quantiles, while GLS would theoretically better predict the true (unknown) values (looking behind the cross-correlation).

It is important to highlight that, given the considerable nesting of catchments in the study area, the ungauged sites considered within the LOOCV scheme are likely to have other gauges upstream or downstream within the same catchment. For this reason, our analyses do not refer to ungauged catchments in the strict sense and could therefore partly favour TK relative to GLS (in which the impact of cross-correlation was considered very explicitly for ungauged catchments in the strict sense by Stedinger and Tasker, 1985). The better investigation of this aspect, as well as of the behaviour of GLS and TK in a heterogeneous region, is suggested for future studies.

4.2 Study area

4.2.1 Flood frequency regime of the study region

The objective of this chapter was investigated by referring to a possibly homogeneous group of catchments which was identified in the Italian Eastern Alps using the regionalisation approach based on L-moments (Hosking and Wallis, 1993, 1997, see Appendix A) and the Region of Influence approach (RoI, see e.g. Burn, 1990; Zrinji and Burn, 1996) to the detailed AMS database presented in Chapter 3. Consistently with the analyses presented in Chapter 3, the selected homogeneous pooling-group (see Fig. 4.1) was delineated by referring to the climatic and geomorphological descriptors available for the study area (i.e. basin area, mean annual precipitation, mean and minimum elevation, latitude and longitude of catchment centroid). In particular, among the RoI pooling-groups of basins identified in Tab. 3.2, we considered the best one in terms of number of station-years of data and homogeneity (see artificial reservoir R20, *Diga di Val Schener*, in Tab. 3.2). The identified homogeneous group consists of 20 annual sequences (AMS) of flood flows with an average sample size approximately equal to 35, for a total of 747 station-years of data, which ensure a reliable estimation of flood quantiles associated with return periods T slightly higher than 100 years (i.e. $5T$ -rule proposed by Jakob et al., 1999).

The nested structure of the selected RoI pooling-group is shown in Fig. 4.2, which reports the percentage of overlapping area between all possible pairs of catchments. Figure 4.2 highlights the considerable nesting of the catchments for the study area: except for catchment C38 (i.e. *Gadera a Mantana*, see Tab. 3.1), all the other catchments in the RoI pooling-group overlap with at least another one from the same group.

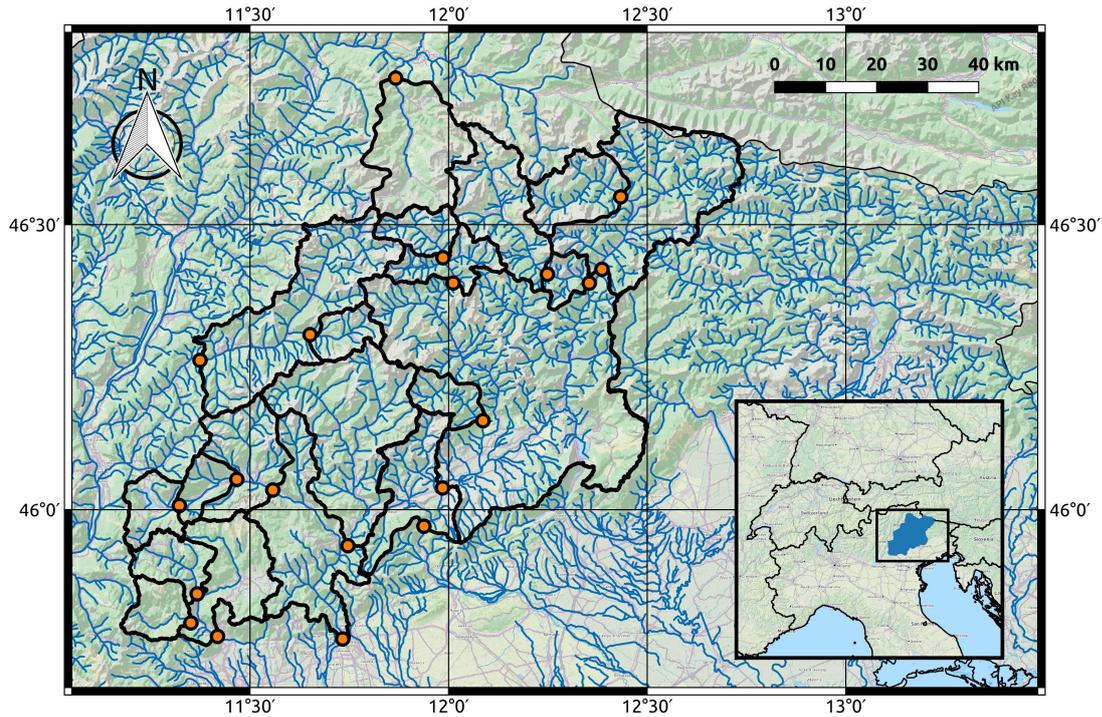


Figure 4.1: Homogeneous pooling-group of 20 catchments in Triveneto: orange points and black lines represent the gauging stations and their drainage area, respectively.

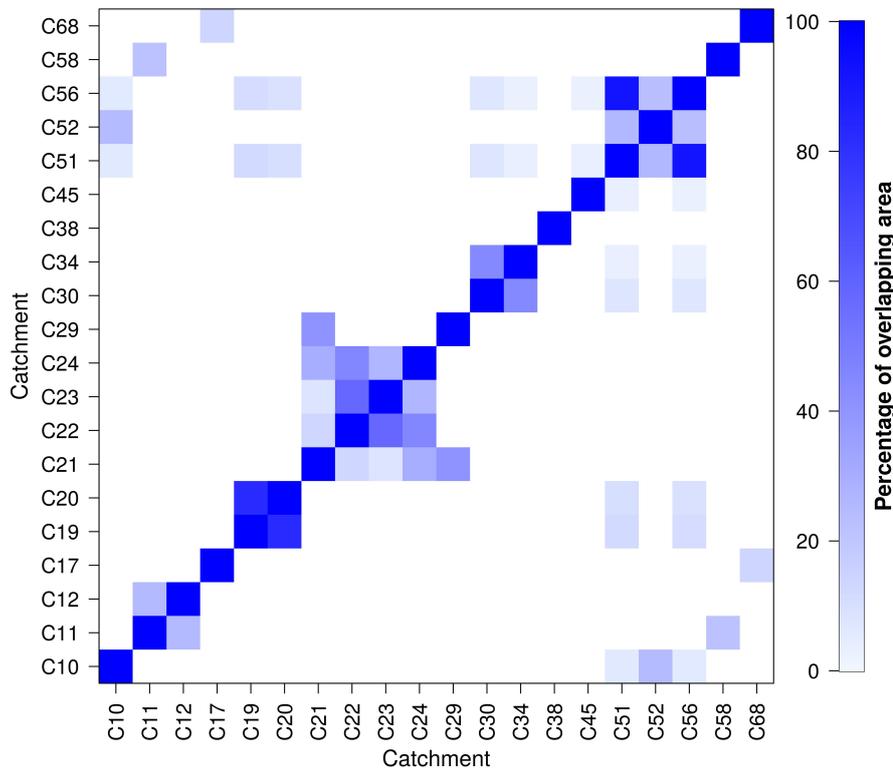


Figure 4.2: Nested structure of the selected RoI pooling-group. For each pair of catchments, the percentage of overlapping area (relative to the bigger of the two) is represented in blue colour scale (the higher the percentage, the deeper the blue).

The application of the L-moments algorithm (i.e. Hosking and Wallis, 1993, 1997) (see also Appendix A) to this pooling-group of sites returned an acceptable degree of homogeneity (i.e. heterogeneity measures H_1 , H_2 and H_3 equal to 0.52, -0.78, and -1.06 in this order, see Hosking and Wallis, 1993) and showed that the LP3 (Log-Pearson type III) family is a suitable parent for representing the flood frequency distribution within an index-flood framework (i.e. Dalrymple, 1960). In particular, LP3 was fitted by referring to the following procedure: (1) dimensionless flood data (i.e. annual flood sequences divided by the corresponding sample means) were log10-transformed, (2) L-moment ratios were estimated in log-space, (3) a Pearson type III distribution was fitted in log-space with L-moments approach (i.e. Hosking and Wallis, 1997), (4) finally, the back-transformation from log-space to real-space was applied to obtain an LP3 distribution.

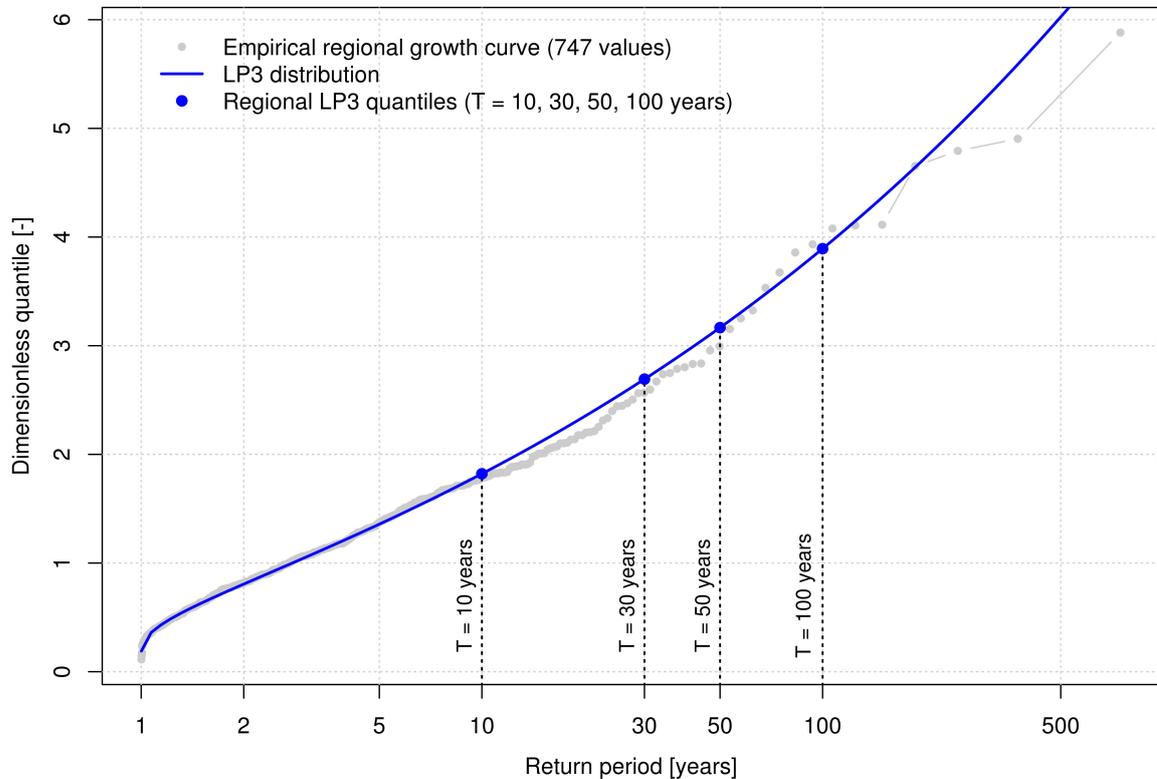


Figure 4.3: Regional growth curve (LP3 distribution, blue curve) obtained by applying the L-moments algorithm (i.e. Hosking and Wallis, 1993, 1997) to the regional sample (grey points). The true dimensionless regional quantiles of interest (i.e. return periods $T = 10, 30, 50, 100$ years) are represented with blue points.

The L-moments estimates of LP3 regional parameters in the log-space, μ^R , σ^R and γ^R , were adopted as parameters of the theoretical parent distribution, and the corresponding regional dimensionless quantiles q_T^R (with selected return periods $T = 10, 30, 50, 100$ years) were set to be the true theoretical dimensionless regional quantiles in the Monte Carlo simulation experiment (see Fig. 4.3). Note that, while the LP3 distribution

is the reference parent distribution in the USA (see e.g. Bulletin 17B of the Interagency Advisory Committee on Water Data, 1982), it is not considered in Europe (see e.g. Castellarin et al., 2012), where the GEV (Generalized Extreme Value) distribution is generally a common choice. Nevertheless, for the purposes of the present study, the choice of the LP3 distribution enabled us to correctly apply the GLS procedure, which is indeed developed with reference to LP3 (see e.g. Tasker and Stedinger, 1989; Farmer, 2017, see also Sec. 2.1.3).

4.2.2 Cross-correlation structure of the study region

In order to characterise the cross-correlation structure of the study region, we first applied quantile-quantile transformation to each series in order to obtain standard normal distributions from the original annual flood sequences divided by their sample mean (i.e. mean annual flood, MAF). Then, using the sample correlations between all pairs of sites i and j in the normal space, we computed the cross-correlation estimators $\hat{\rho}_{ij}$ and developed a model of ρ_{ij} as a function of distance $d_{i,j}$ between the centroids of any two basins, ignoring cases where basins were nested and where they were not (see Fig. 4.4(a)). For the evaluation of the cross-correlation structure, all the pairs of basins having less than 10 concurrent observations were neglected. The relationship between cross-correlation and distance between catchment pairs was modelled through the non-linear regression model of Tasker and Stedinger (1989) already introduced in Eq. (2.11):

$$\hat{\rho}_{ij} = \theta^{\frac{d_{ij}}{1+\alpha d_{ij}}} = e^{\frac{\ln(\theta) d_{ij}}{1+\alpha d_{ij}}}$$

where d_{ij} is the distance (expressed in miles) between centroids for catchments i and j , and $0 < \theta < 1$ and $\alpha > 0$ are the dimensionless model parameters estimated from data. In the study region, $\bar{\rho}$ was found to be 0.6 (i.e. $\theta \simeq 0.8$, $\alpha \simeq 0.4$).

Figure 4.4(a) shows that cross-correlation in the study region is almost constant for larger distances between catchment centroids. Although, generally, one could expect cross-correlation to decay more rapidly with distance, the specific behaviour observed for the study region can be explained by its homogeneity and its limited spatial extension (i.e. the maximum distance between catchment centroids is about 120 km). Indeed, Fig. 4.4(b) highlights that for the entire Triveneto region, which is wider (i.e. distances are within 230 km) and highly heterogeneous (see Chapter 3), the decay of cross-correlation with distance is more rapid (i.e. $\theta \simeq 0.97$, $\alpha \simeq 0.01$).

The identification of the cross-correlation structure of the homogeneous study region in the standard-normal space was preparatory for the generation of the different cross-correlated realisations of the region described in Sec. 4.4.

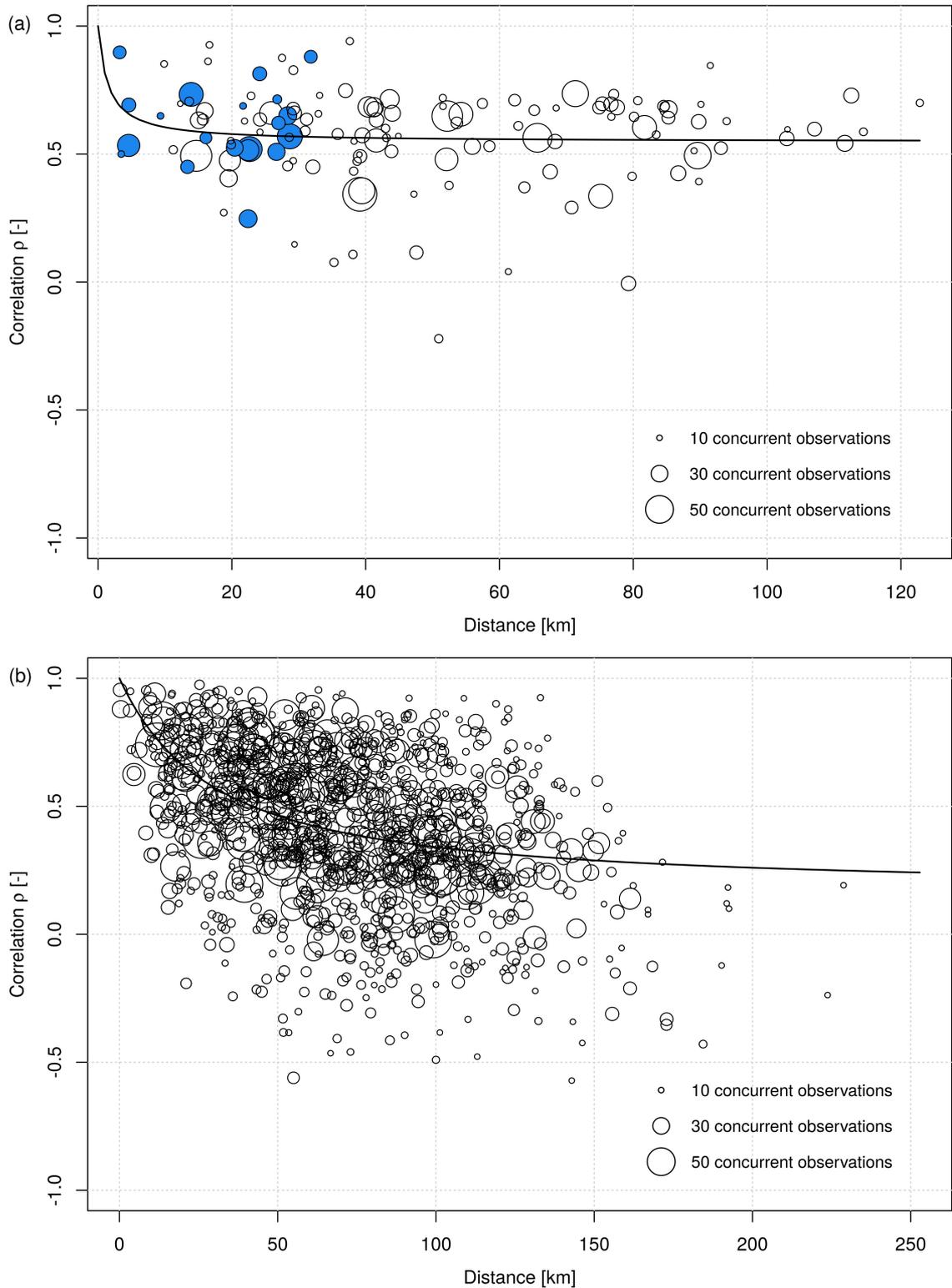


Figure 4.4: Cross-correlation structure of the groups of AMS included in (a) the selected homogeneous pooling-group, and (b) the entire Triveneto region analysed in Chapter 3. Correlations between all pairs of sites in the normal space as a function of distances between their centroids are indicated with circles (diameter reflects the size of simultaneous observations between the two selected catchments), while the solid line depicts the adopted theoretical model (i.e. Tasker and Stedinger, 1989). For the evaluation of the cross-correlation structure, all the pairs of basins having less than 10 concurrent observations were neglected. Blue-filled circles in panel (a) indicate pairs of nested catchments.

4.3 Preliminary application of GLS and TK in the real-world study region

The preliminary analyses conducted in this chapter considered the above mentioned regional dataset and repeated the same exercise performed in Archfield et al. (2013), to see to what extent the results obtained in the South-eastern USA hold in our study area.

For this purpose, we performed an application of GLS and TK aimed at predicting the at-site empirical flood quantiles for selected return periods T (i.e. $T=10, 30, 50, 100$ years). In particular, empirical quantiles were computed by combining at-site and regional information: we referred to the LP3 distribution (i.e. no uncertainty on the model) and estimated the LP3 parameters in log-space μ and σ locally by using the L-moments method (i.e. Hosking and Wallis, 1997) (i.e. uncertainty from sampling error only on the moments of order 1 and 2); the shape parameter in log-space γ was set equal to γ^R (i.e. true shape parameter in log-space), instead (i.e. no uncertainty on the third order moment). GLS and TK were then applied to the sample flood quantiles in order to predict flood quantiles at ungauged locations using a leave-one-out cross-validation (LOOCV) scheme (i.e. simulating ungauged conditions at each and every site in the study region).

4.3.1 Application of Generalized Least Squares (GLS)

The GLS analysis was carried out by using the R-package WREG (i.e. Farmer, 2017; see also Eng et al., 2009) in R (R Core Team, 2016). As illustrated in Sec. 2.1, in order to apply the GLS procedure, estimates of standard deviation s_i , regional skew estimate G_{Ri} for the gauge i , weighted skew $\tilde{G}_{w,i}$, and LP3 distribution standard deviate K_i for each i th site are required (see also Griffis and Stedinger, 2007b). These variables were computed for the log-transformed samples, as follows:

- standard deviations s_i were evaluated as at-site sample standard deviations;
- as isoline maps for regional skew are not available in the study area, we used the same value of regional skew G_R for each and every catchment in the study region (i.e. assumption of homogeneity); this value was computed as the sample skew of the log-transformed regional dimensionless sample;
- weighted skews $\tilde{G}_{w,i}$ for each site were computed as indicated in Eq. (2.8), combining the at-site sample skews g_i with the regional skew G_R ; weights were computed as indicated in Eq. (2.9), evaluating the estimated mean square errors $MSE(g_i)$ and

$MSE(G_R)$ in accordance with Bulletin 17B of the Interagency Advisory Committee on Water Data (1982);

- standard deviates K_i were computed as functions of non exceedance probability (i.e. return period T) and weighted skew $\tilde{G}_{w,i}$, as indicated in Bulletin 17B of the Interagency Advisory Committee on Water Data (1982).

Moreover, an estimate of the cross-correlation structure is required, in terms of the parameters θ and α of the model introduced by Tasker and Stedinger (1989) (see Eq. (2.11)). In the present study, the cross-correlation model for the application of GLS was evaluated with reference to the distances between catchment centroids computed with the Haversine approximation, and the model parameters θ and α were estimated by OLS procedure.

GLS regressions employ different catchment descriptors, e.g. drainage area, precipitation, elevation, etc. In the present study, we decided to implement a GLS regional model in two different ways:

- monivariate-GLS (hereinafter referred to as 1v-GLS): the GLS quantile regression analysis is performed by considering drainage area A only:

$$Q_{T_i} = a_T A_i^{b_T} \quad (4.1)$$

where Q_{T_i} is the T -year flood for site i , A_i is the drainage area for site i , and a_T and b_T are the GLS parameters;

- multivariate-GLS (hereinafter referred to as mv-GLS): the GLS quantile regression analysis takes more catchment descriptors into account. In the context of the present study, we referred to drainage area A , mean annual precipitation MAP, latitude of catchment centroid Y_g , mean elevation H_{mean} , and minimum elevation H_{min} , which were selected by performing a preliminary stepwise log-linear regression analysis (see Draper and Smith, 1981; Weisberg, 1985; Chambers, 1992) aimed at identifying the best possible descriptors (i.e. adjusted R-squared $R_{adj}^2 \simeq 0.90$ and normally distributed standardised residuals) for estimating the mean annual flood (MAF) in the study area. The resulting multivariate-GLS model can be described with the following equation:

$$Q_{T_i} = a_T A_i^{b_T} MAP_i^{c_T} Y_{g_i}^{d_T} H_{mean_i}^{e_T} H_{min_i}^{f_T} \quad (4.2)$$

where a_T , b_T , c_T , d_T , e_T and f_T are the GLS parameters.

In the present study, both Eq. (4.1) and Eq. (4.2) were reduced to linear additive forms by means of a log-transformation of both sides of equation (see e.g. Thomas and Benson, 1970; Pandey and Nguyen, 1999; Griffis and Stedinger, 2007a; Laio et al., 2011).

Both the 1v- and mv-GLS applications were performed in a LOOCV scheme, by removing one site in turn from the dataset and referring to the other 19 sites for estimating: (1) the cross-correlation structure (i.e. fitting of the model of Tasker and Stedinger, 1989), (2) the GLS parameters, and (3) the flood quantiles Q_T at the discarded site.

4.3.2 Application of Top-kriging (TK)

In order to perform an ungauged application of TK for each realisation of the region, in accordance with previous studies (see e.g. Pugliese et al., 2014, 2016), the first step was the use of OLS for identifying a regional power-law model between flood quantile Q_T and basin area. The OLS estimates were then used to standardise LP3 quantiles (i.e. Q_T with $T = 10, 30, 50, 100$ years) at all sites: this fundamental step is necessary as TK directly handles drainage area as a key variable of the model itself.

In this regard, to be coherent with the aforementioned application of GLS, we opted for two different types of OLS regional power-law:

- monivariate-OLS (1v-OLS; i.e. considering only drainage area), resulting in a standard application of TK (hereinafter referred to as s-TK);
- multivariate-OLS (mv-OLS; i.e. considering the five descriptors identified via stepwise regression analysis; see e.g. Eq. (4.2)), resulting in a TK with external drift (hereinafter referred to as ed-TK; see e.g. Laaha et al., 2013, for the use of ed-TK for predicting stream temperatures).

The TK interpolation was then applied by fitting the sample variogram of the standardised quantiles with the five-parameter fractal-exponential model suggested in Skøien et al. (2006) through a modified version of Weighted Least Squares (WLS) regression (see Cressie, 1993). The fitted variogram model was then used to compute the kriging weights, referring to the 6 closest neighbouring stations (in line with recent studies conducted for flow-duration curves, see e.g. Pugliese et al., 2014, 2016). The standardised quantiles were then predicted site by site by using Eq. (2.18). Finally, the TK estimates of the standardised quantiles were combined with the regression estimates resulting from 1v-OLS and mv-OLS to obtain the s-TK and ed-TK estimated T-year value at each site.

Both the s- and ed-TK analyses were performed in a LOOCV scheme, by removing one site in turn from the dataset and referring to the remaining 19 sites for estimating:

(1) the OLS regional power-law useful for standardising flood quantiles, (2) the variogram (i.e. five-parameter fractal-exponential model suggested in Skøien et al., 2006), and (3) the flood quantiles Q_T at the discarded site.

4.3.3 Performance metrics

The performance of the considered versions of GLS (i.e. 1v-GLS and mv-GLS) and TK (i.e. s-TK and ed-TK) in predicting at-site flood quantiles Q_T (i.e. $T = 10, 30, 50, 100$ years) in a LOOCV scheme were evaluated with reference to different metrics.

Consistently with Archfield et al. (2013), we firstly looked at the at-site absolute errors, computed as follows:

$$AE_i = |\hat{x}_i - x_i| \quad (4.3)$$

where AE_i indicates the absolute error for site i , \hat{x}_i the estimated variable at site i , and x_i the observed value of the variable of interest (i.e. Q_T , with $T = 10, 30, 50, 100$ years) at site i . As defined in Eq. (4.3), absolute errors are useful for evaluating the performance site-by-site.

For this reason, to better evaluate the overall performance of GLS and TK in estimating Q_T (i.e. $T = 10, 30, 50, 100$ years) in the study region, we considered also two error measures, Relative Bias (RBIAS) and Root Mean Square Normalized Error (RMSNE):

$$RBIAS = \frac{1}{N} \sum_{j=1}^N \left(\frac{\hat{x}_i - x_i}{x_i} \right) \quad (4.4)$$

$$RMSNE = \left[\frac{1}{N} \sum_{j=1}^N \left(\frac{\hat{x}_i - x_i}{x_i} \right)^2 \right]^{1/2} \quad (4.5)$$

and a performance measure, Relative Nash-Sutcliffe Efficiency (RNSE):

$$RNSE = 1 - \frac{\sum_{j=1}^N \left(\frac{\hat{x}_i - x_i}{x_i} \right)^2}{\sum_{j=1}^N \left(\frac{\hat{x}_i - \bar{x}}{\bar{x}} \right)^2} \quad (4.6)$$

where N is the number of river cross-sections, \hat{x}_i the estimated variable at site i , x_i the observed value of the variable at site i , and \bar{x} its mean value. The choice of relative (i.e. RBIAS, RNSE) and normalised (i.e. RMSNE) measures was done in order to quantify errors and performance, regardless of the size of the drainage area of each catchment. For a meaningful interpretation of results, note that the lower the value of RBIAS and

RMNSE the better the performance, whereas the higher the value of RNSE the better the performance of the model (i.e. RNSE can span from $-\infty$ to 1, where 1 indicates the perfect model). In particular, RNSE is a rescaling of RMSNE: while RMSNE is an average fractional error, RNSE represents the fraction of variability explained.

4.3.4 Results and discussion

The evaluation of AEs was useful mainly for assessing the validity for our study area of the results found by Archfield et al. (2013) in the South-eastern USA. At the same time, the application of mv-GLS and ed-TK performed in this Thesis represents a further extension of the analyses implemented in Archfield et al. (2013), who considered only the univariate versions (i.e. dependence on drainage area only) of GLS and TK. Fig. 4.5 shows AEs obtained site-by-site by applying (a) 1v-GLS and s-TK, and (b) mv-GLS and ed-TK, for estimating the 100-year flood.

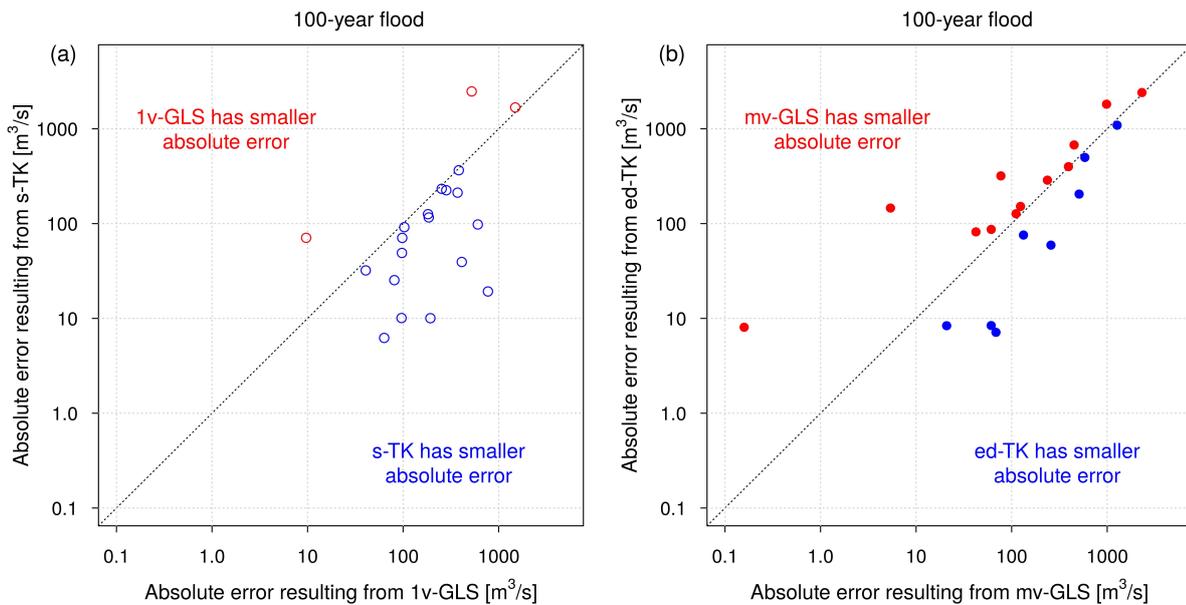


Figure 4.5: Comparison of the absolute error (AE) between empirical and predicted 100-year flood quantiles resulting from (a) 1v-GLS and s-TK, and (b) mv-GLS and ed-TK.

Figure 4.5(a) highlights that s-TK results in smaller AEs than 1v-GLS for 17 out of 20 sites, and confirms what already observed in Archfield et al. (2013). At the same time, Fig. 4.5(b) shows that, if more than one significant catchment descriptors (besides drainage area) are taken into account, mv-GLS tends to have smaller AEs (see x-axis); in particular mv-GLS has smaller AEs than ed-TK in 12 of the 20 sites. The fact that 1v-GLS has lower performance than 1v-TK, but mv-GLS does about the same as mv-TK is explanatory of the fact that the mean annual flood (MAF) at each station depends

upon more than just drainage area, and that 1v-TK seems to be able to capture some of that, whereas 1v-GLS would not. AE diagrams for 10-, 30-, 50-year flood show analogous results and are not reported here for the sake of conciseness.

Still regarding the prediction of the 100-year flood quantiles, results in terms of RBIAS, RMSNE and RNSE (see Fig. 4.6) highlight that mv-GLS, s-TK, and ed-TK outperform 1v-GLS, which shows higher RBIAS and RMSNE, and lower RNSE (even under 0.6). At the same time, mv-GLS, s-TK, and ed-TK show similar performance in terms of RMSNE and RNSE. Concerning RBIAS, Fig. 4.6(a) shows that all the considered methodologies resulted to be positively biased in the present application. As concluded also from the analysis of AEs, this behaviour seems to be related with the ability of the considered methods to reproduce the mean annual flood (MAF). In particular, the smallest bias is associated with ed-TK. Analogous results were observed also for 10-, 30- and 50-year flood quantiles.

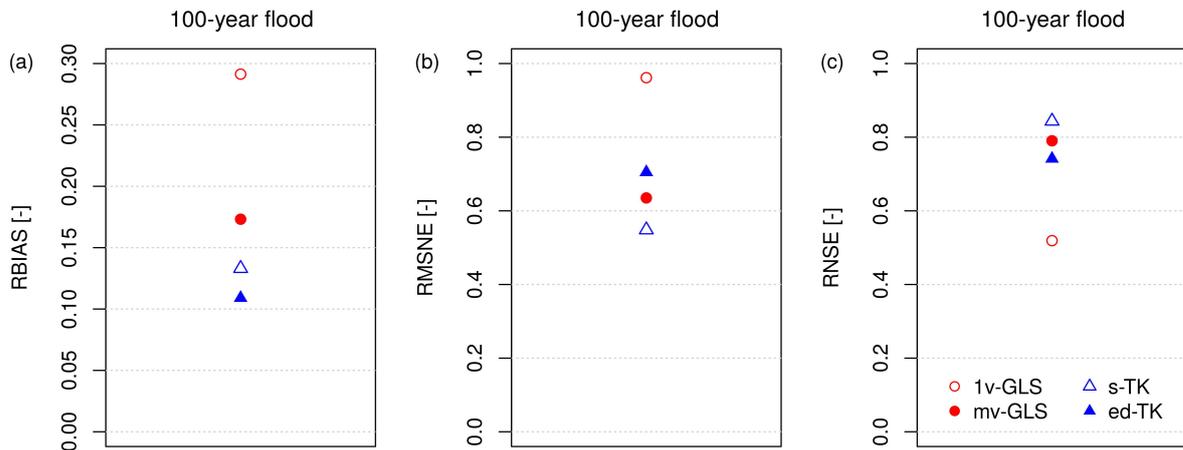


Figure 4.6: Outcomes of the application (LOOCV scheme) of 1v-GLS (red circles), mv-GLS (red dots), s-TK (blue-bordered triangles) and ed-TK (blue-filled triangles) for predicting at-site 100-year flood quantiles in the study area. Prediction performances are in terms of (a) RBIAS, (b) RMSNE, and (c) RNSE.

In summary, the preliminary analyses performed for the study area highlighted a behaviour of 1v-GLS and s-TK which is consistent with the results in Archfield et al. (2013) for a set of 61 gauged basins located across the South-eastern USA, wherein many basins were nested, as in the present study area. The outstanding performances of s-TK compared to 1v-GLS are expected as, in performing its estimates, TK refers to n neighbouring sites and therefore is implicitly able to take some climate and geomorphological similarities between catchments into account, especially in regions with preponderance of nested catchments, as in the case considered here. Indeed, the preliminary application performed here suggests that the inclusion of more catchment descriptors in the GLS analysis (i.e.

mv-GLS) can lead to significantly improved performances (i.e. similar to s-TK). Smaller differences are obtained moving from s-TK to ed-TK.

Although informative, this preliminary analysis does not address the main science question of the present chapter, which is specifically tackled through the Monte Carlo simulation experiment described in the following Sec. 4.4.

4.4 Application of GLS and TK to cross-correlated realisations of the homogeneous region

4.4.1 Monte Carlo simulation framework

In order to assess the behaviour of GLS and TK under different cross-correlation scenarios, we implemented a Monte Carlo simulation experiment for generating cross-correlated realisations of the homogeneous region. The simulation framework and set of analyses can be summarised as follows:

1. we focused on the selected real-world study region to mimic regional flood frequency regime and controls of relevant catchment descriptors, as well as spatial correlation structure of flood flows, and define true flood quantiles at each and every site in the region referring to a unique regional parent distribution;
2. we generated 1000 cross-correlated realisations of the region for three different degrees of regional cross-correlation; each realisation consists of 20 concurrent sequences for 20 sites of 35 annual floods (see Sec. 4.4.2 for the Monte Carlo simulation algorithm);
3. we applied the L-moments algorithm (i.e. Hosking and Wallis, 1993, 1997) for predicting at-site flood quantiles associated with selected return periods (i.e. $T = 10, 30, 50, 100$ years) at each and every site in each realisation of the region (see Sec. 4.4.3);
4. we then referred to at-site flood quantiles for predicting flood quantiles in ungauged locations by means of 1v-GLS, mv-GLS, s-TK and ed-TK within a leave-one-out cross-validation (LOOCV) procedure for each realisation of the region (see Sec. 4.4.4);
5. we finally compared TK and GLS cross-validated predictions with at-site and known true flood in terms of the performance metrics illustrated in Sec. 4.3.3 (i.e. RBIAS, RMSNE, RNSE; see Sec. 4.4.5 for the results).

4.4.2 Cross-correlated realisations of the region

The objective of the study was investigated by referring to cross-correlated realisations of the region generated in a Monte Carlo framework. To this aim, we mimicked the main characteristics of the real-world homogeneous study region in terms of its flood frequency regime: (1) unique regional parent distribution of annual flood flows (see Sec. 4.2.1) and similar geomorphological and climatic controls on the annual flood; (2) cross-correlation structure (see Sec. 4.2.2).

L-moments regional estimates of LP3 regional parameters, μ^R , σ^R and γ^R , presented in Sec. 4.2.1 were adopted as parameters of the theoretical parent distribution for each realisation, and the corresponding regional dimensionless quantiles q_T^R (with arbitrarily selected return periods $T = 10, 30, 50, 100$ years) were selected as true dimensionless regional quantiles (see Fig. 4.3). Note that the corresponding true dimensional quantiles at each i th site, $Q_{T,i}^R$, were obtained by multiplying q_T^R by the local empirical mean annual flood, MAF_i (i.e. mean value of the observed AMS at the i th site). No effect of the cross-correlation structure was therefore considered on the mean annual flood.

The cross-correlation structure of the study region was modelled through the non-linear model of Tasker and Stedinger (1989) (see Eq. (2.11)). This relationship was then used for generating cross-correlated sequences of annual floods at the 20 sites in the region. Together with the model fitted to our regional case study (see Fig. 4.4(a)), which corresponds to an average cross-correlation between all pairs, $\bar{\rho}$, equal to 0.6 (i.e. $\theta \simeq 0.8$, $\alpha \simeq 0.4$), we also considered two alternative models that adopt similar decaying laws between correlation and distance, but describe a lower and a higher spatial correlation scenarios, resulting in $\bar{\rho}$ equal to 0.2 ($\theta \simeq 0.3$, $\alpha \simeq 0.7$) and 0.8 ($\theta \simeq 0.9$, $\alpha \simeq 0.3$), respectively.

For each cross-correlation scenario considered in our study (i.e. $\bar{\rho} = 0.2, 0.6, 0.8$), 1000 realisations of the study region were compiled by generating random annual flood sequences. Flood sequences were simulated by first generating cross-correlated sequences from a multivariate standard-normal distribution, and then by back-transforming the standard normal variate to dimensionless LP3 flows through quantile-quantile transformations. Finally, to obtain the synthetic dimensional annual sequence of flood flows at the i th site for any given realisation, the synthetic dimensionless annual sequence at the i th site was multiplied by the local empirical mean annual flood, MAF_i (i.e. we did not model any synthetic generation of the MAF , in accordance with the hypothesis of neglecting the effect of cross-correlation on the mean annual flood). Each realisation of the region consists of 20 overlapping annual sequences with record length equal to 35 years (average record length in the real-world dataset). Note that, due to sampling variability

(i.e. limited record length), the mean value of each generated dimensionless series is not exactly equal to 1; for this reason, the resulting dimensional series for the i th site shows an actual mean value which is slightly different from the local empirical mean annual flood (MAF_i), and varies from realisation to realisation of the region.

Figure 4.7 illustrates the cross-correlation structure in the standard-normal space as results from an empirical analysis of the 1000 realisations of the region that we generated for each cross-correlation scenario: (a) $\bar{\rho} = 0.2$, (b) $\bar{\rho} = 0.6$, (c) $\bar{\rho} = 0.8$. In particular, Fig. 4.7(d) highlights the effectiveness of the Monte Carlo procedure implemented in the study: the dispersion of the average correlation values for each of the 1000 realisations generated for each scenario results to be centered on the corresponding expected cross-correlation value (i.e. median values of each box-plot are consistent with the imposed $\bar{\rho}$ values).

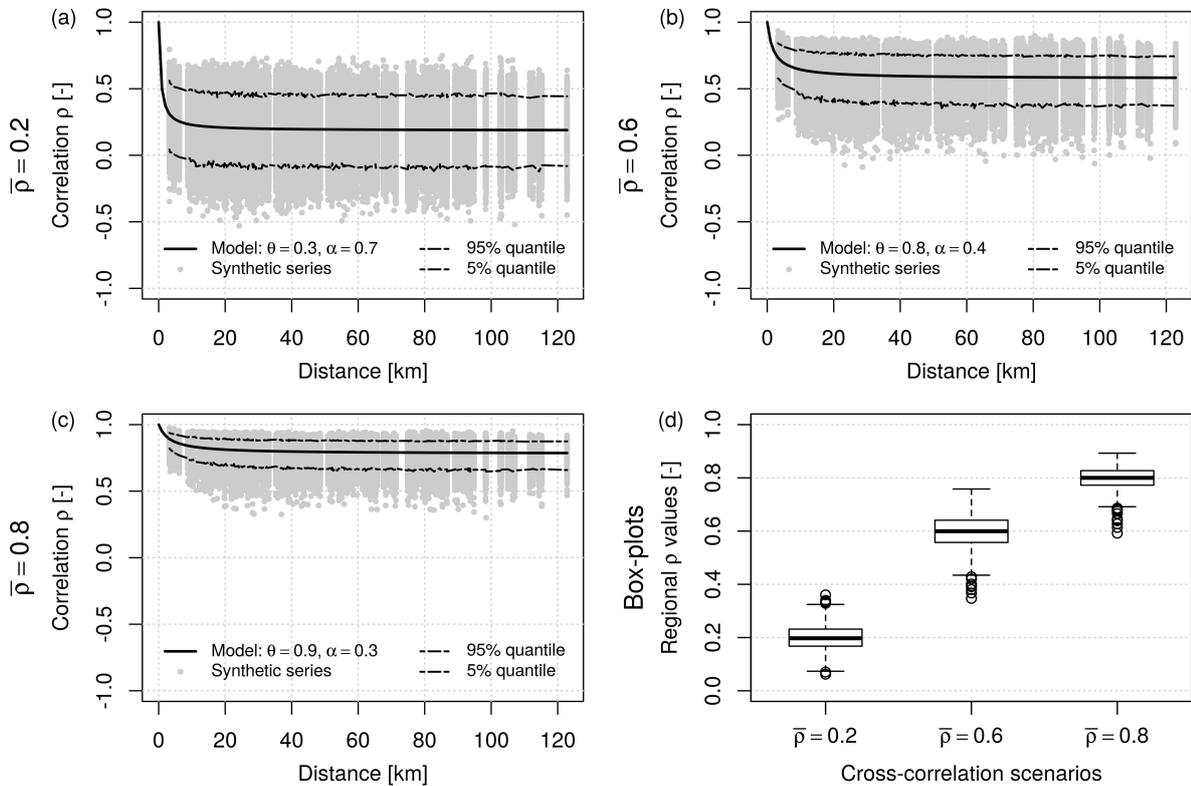


Figure 4.7: Generation of 1000 realisations of the region for each of the three levels of regional cross-correlation (i.e. $\bar{\rho} = 0.2, 0.6, 0.8$). Panels (a), (b) and (c) represent the dispersion of the synthetic series of empirical estimators (grey points) around the corresponding cross-correlation model (black solid line) for $\bar{\rho} = 0.2, 0.6, 0.8$, in this order; the dashed lines indicate the 90% confidence band (upper and lower lines represent 95th and 5th quantiles, respectively). Box-plots in panel (d) show the dispersion of the average correlation values for each of the 1000 realisations generated for each of the three cross-correlation scenarios ($\bar{\rho} = 0.2, 0.6, 0.8$, in this order).

4.4.3 Empirical flood quantiles

For each realisation and each site in the region, empirical flood quantiles were then computed for the selected return periods T (i.e. $T = 10, 30, 50, 100$ years). Quantiles estimators were computed by combining at-site and regional information. In particular, we referred to the LP3 distribution (i.e. no uncertainty on the model selection) and estimated the LP3 parameters μ and σ locally by using the L-moments method (i.e. Hosking and Wallis, 1997) (i.e. uncertainty from sampling error only on the moments of order 1 and 2); while the shape parameter γ was set equal to γ^R (true shape parameter, i.e. no uncertainty on the third order moment).

4.4.4 Prediction of flood quantiles in ungauged sites: application of GLS and TK in cross-validation

GLS and TK were applied to the locally estimated (dimensional) empirical flood quantiles for any of the realisations of the region that we generated as described above in order to predict flood quantiles at ungauged locations. Ungauged conditions at each and every site in each realisation of the region were simulated through a leave-one-out cross-validation (LOOCV) scheme. For each realisation in each cross-correlation scenario, we applied 1v-GLS, mv-GLS, s-TK and ed-TK with exactly the same settings and LOOCV scheme illustrated in Sec. 4.3.1 and Sec. 4.3.2. Finally, the TK and GLS cross-validated predictions were compared with at-site and known true flood quantiles in terms of RBIAS, RMSNE, RNSE (see Sec. 4.3.3).

4.4.5 Results and discussion

The efficiencies of GLS and TK in estimating the empirical estimates of flood quantiles (see Fig. 4.8) and the true flood quantiles (see Fig. 4.9) are represented by means of box-plots. Each box-plot indicates the dispersion of the 1000 values (i.e. one value for each realisation of the region) of the selected performance measure (panels (a), (b), and (c) refers to RBIAS, RMSNE, and RNSE, in this order) for the selected method (i.e. 1v-GLS, mv-GLS, s-TK, ed-TK), return period T (i.e. $T = 10, 30, 50, 100$ years) and cross-correlation scenario $\bar{\rho}$ (i.e. $\bar{\rho} = 0.2, 0.6, 0.8$). In order to allow a much easier and direct visual comparison between the results obtained for empirical estimates of flood quantiles and true flood quantiles, y-axes in the corresponding panels of Fig. 4.8 and Fig. 4.9 are reported with the same scale.

Concerning the efficiencies in estimating the empirical estimates of flood quantiles, Fig. 4.8 shows that TK results in generally better predictions of empirical quantiles (i.e.

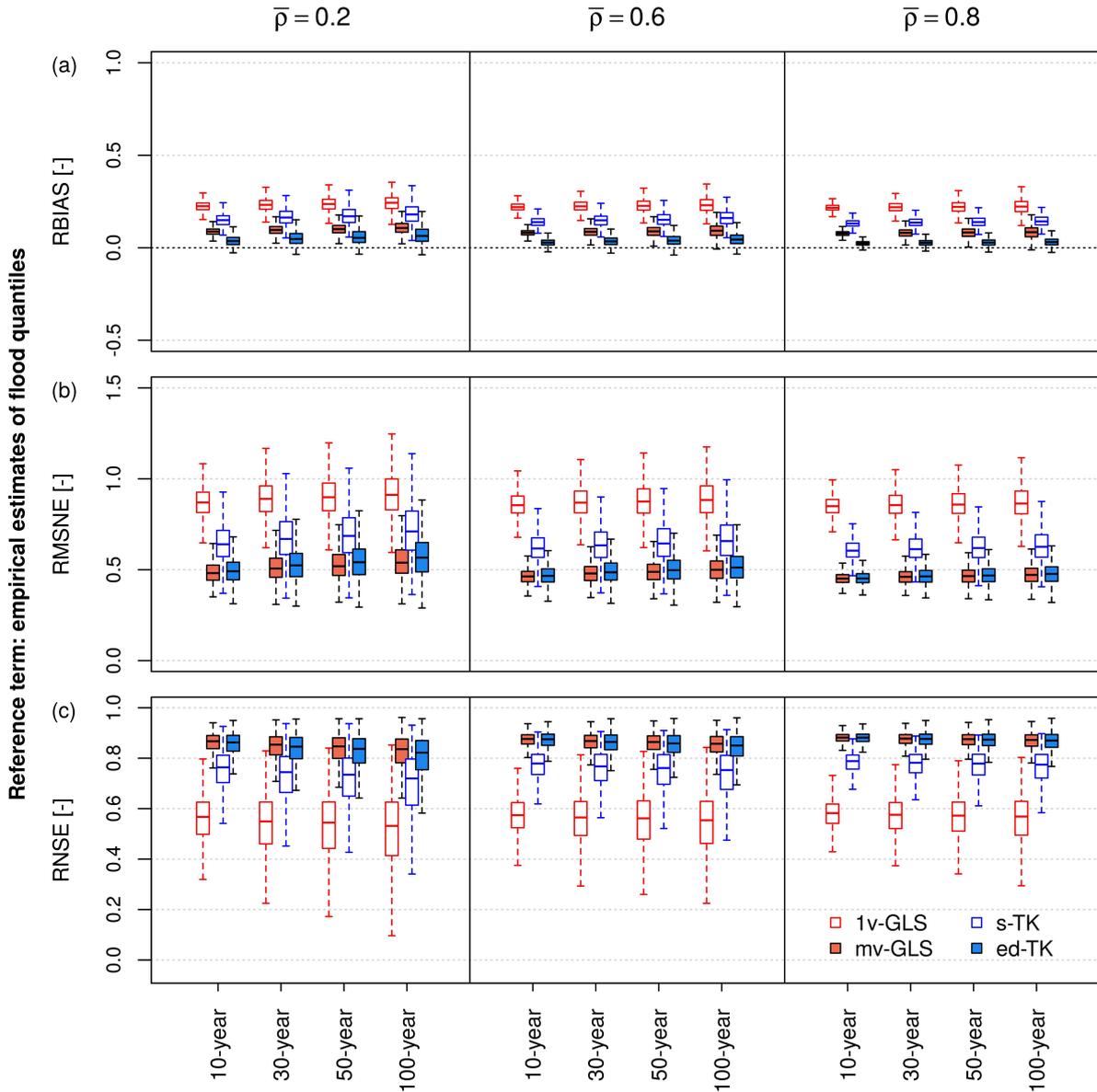


Figure 4.8: Prediction performance of 1v-GLS (red-bordered box-plots), mv-GLS (red-filled box-plots), s-TK (blue-bordered box-plots) and ed-TK (blue-filled box-plots) in estimating (LOOCV scheme) the empirical estimates of flood quantiles for the three degrees of cross-correlation (i.e. $\bar{\rho} = 0.2, 0.6, 0.8$): box-plots represent the distribution of 1000 values of each metric (i.e. RBIAS, RMSNE, and RNSE) computed for each realisation of the region and for selected return periods (i.e. $T = 10, 30, 50, 100$ years).

lower RBIAS, lower RMSNE, higher RNSE) than the corresponding version of GLS: s-TK outperforms 1v-GLS, and ed-TK shows similar median RMNSE and median RNSE, but smaller RBIAS than mv-GLS. Moreover, the comparison between 1v-GLS versus mv-GLS, and s-TK versus ed-TK confirms what observed in our preliminary analysis of the real-world study area (see Sec. 4.3.4): the inclusion of more catchment descriptors in the regression analysis (i.e. mv-GLS; mv-OLS for ed-TK) leads to significantly improved per-

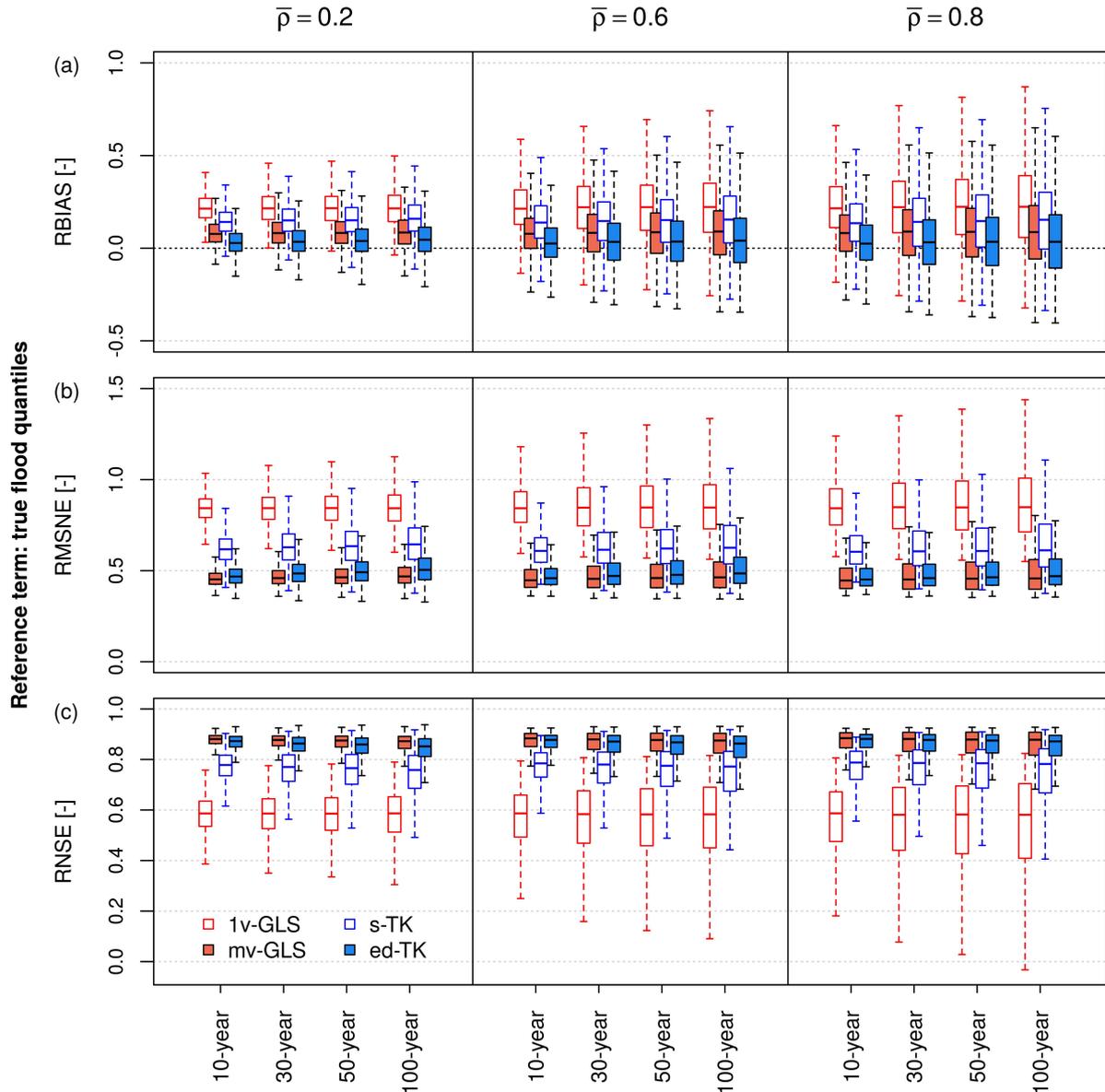


Figure 4.9: Prediction performance of 1v-GLS (red-bordered box-plots), mv-GLS (red-filled box-plots), s-TK (blue-bordered box-plots) and ed-TK (blue-filled box-plots) in estimating (LOOCV scheme) the true flood quantiles for the three degrees of cross-correlation (i.e. $\bar{\rho} = 0.2, 0.6, 0.8$): box-plots represent the distribution of 1000 values of each metric (i.e. RBIAS, RMSNE, and RNSE) computed for each realisation of the region and for selected return periods (i.e. $T = 10, 30, 50, 100$ years).

formances; in particular, the weak performances of 1v-GLS can be explained with the fact that drainage area alone is not enough for fully describing MAF (and therefore dimensional flood quantiles) in the study region. Moreover, the similar behaviour of mv-GLS and mv-TK can be explained with the fact that they both use the important physiographic information needed to explain variations in the mean flood. In this regard, as already observed for the real-world study area (see Sec. 4.3.4), all the considered procedures are

positively biased, with median RBIAS values which are similar to the corresponding values obtained for the real-world application (see Fig. 4.6 for the 100-year flood quantiles), and RBIAS significantly decreasing from 1v-GLS to ed-TK.

Another aspect shown in Fig. 4.8 is the weak dependence on T : for each $\bar{\rho}$ and considered method, we observed that the higher the T value, the lower the performance (i.e. the higher the bias or uncertainty). This behaviour is totally expected and confirms that estimates of flood quantiles associated with lower probability of occurrence (i.e. higher return period T) are affected by higher uncertainties.

Finally, we observed a dependence of the performance on the average regional cross-correlation $\bar{\rho}$: for a given method and return period T , the higher the $\bar{\rho}$ the better the median of the performance and the lower the uncertainty, with much less dispersed performance. This behaviour is totally expected, especially for TK, which explicitly exploits cross-correlation in predicting flood quantiles.

With regards to the efficiencies of GLS and TK in estimating the true flood quantiles, Fig. 4.9 confirms the trend already observed for empirical flood quantiles: s-TK, mv-GLS, and ed-TK outperform 1v-GLS in predicting true flood quantiles (i.e. lower RBIAS, lower RMSNE, higher RNSE), and, in particular, mv-GLS and ed-TK perform similarly, with slightly less biased predictions for ed-TK. This confirms that incorporating multivariate regression in mv-GLS and ed-TK improves significantly the prediction performances. Moreover, the weak dependence on T is confirmed: the higher the T value, the higher the uncertainty.

Notwithstanding the above mentioned similarities in the relative behaviour of the methods, a significant difference is present in the extent of the performances: all the considered methods show much lower performance in estimating true flood quantiles than empirical quantiles. The only exception is observed for $\bar{\rho} = 0.2$, for which, even if a higher dispersion of RBIAS is observed in predicting true flood quantiles (see panels (a) in Fig. 4.8 and Fig. 4.9), RMNSE and RNSE indicate a better behaviour of all the considered methods in predicting true flood quantiles than their empirical estimates. This can be explained with the fact that the presence of cross-correlation masks the true flooding potential in the region, and none of the considered methods is able to effectively overcome this effect, unless the regional average cross-correlation is very limited (i.e. see results for $\bar{\rho} = 0.2$). In this regard, other important indications come from the dependence on $\bar{\rho}$, which is in this case the opposite to what observed for empirical flood quantiles: for estimating true flood quantiles, we observed that the higher the $\bar{\rho}$ the lower the performance. In particular, as the spatial correlation has the effect of masking the flood magnitude for a limited set of flood-flow observations, and TK exploits cross-correlation to performs its

estimates, the decreasing performances and increasing uncertainty of s-TK and ed-TK with increasing $\bar{\rho}$ are expected. On the other hand, the increasing uncertainty of GLS with increasing $\bar{\rho}$ can be explained as follows: even if GLS is able to look behind cross-correlation, we could expect a residual masking-effect with increasing levels of spatial correlation.

In summary, the Monte Carlo experiment performed in this chapter enabled us to address the unsolved issue raised by Archfield et al. (2013) regarding the ability of GLS and TK in predicting the true unknown flood quantiles in ungauged sites when the observed flood sequences are affected by cross-correlation. At the same time, it is important to consider the extremely simplified situation which was represented in our analyses: we referred to an acceptably homogeneous region with nested catchments. Indeed, as already stated in the introduction to this chapter (see Sec. 4.1), given the considerable nesting of catchments in the study area, the LOOCV scheme performed in our analyses refers to ungauged sites which are likely to have other gauges upstream or downstream within the same catchment, and therefore are not ungauged catchments in the strict sense. As Steindinger and Tasker (1985) considered GLS very explicitly for ungauged catchments in the strict sense (no nested sites), the analyses performed in this Thesis could have benefited TK rather than GLS. The better investigation of this aspect is suggested for future studies, anyway bearing in mind that the total absence of nesting could be penalising for TK (see Sec. 2.3.1, and Fig. 2.3 in particular, for further details). Another important aspect is related to the homogeneity of our study region: we referred to a homogeneous region for the sake of simplicity, but in general homogeneity is not a fundamental assumption for the application of GLS and TK, which can indeed be applied to heterogeneous areas. For this reason, future studies could consider the application of a Monte Carlo experiment similar to the one considered here, but referring to a heterogeneous region (i.e. true flood quantiles for each site referring to different distributions).

Another important issue is represented by the mean annual flood: the Monte Carlo experiment implemented in this study (see Sec. 4.4.2) considered a direct effect of cross-correlation only on the synthetic dimensionless quantiles, but no effect on the mean annual floods (i.e. we referred to the mean value of the observed AMS at the i th site). In this context, future studies could consider the generation of synthetic mean annual floods, which may include the effect of spatial correlation.

Further issues could be addressed by future studies. Indeed, as explained in Sec. 4.3.1 and Sec. 4.3.2, we applied GLS and TK in a LOOCV scheme by removing one site in turn and referring to the remaining 19 sites. In this procedure, the kriging weights for TK were computed referring to the 6 closest neighbouring stations to the target one.

Although that was in line with Pugliese et al. (2014, 2016), a sensitivity analysis could be interesting for evaluating the effect of different numbers of closest neighbouring stations in the application of TK. Moreover, we considered overlapping annual sequences with record length equal to 35 years for every site, without investigating the sensitivity of the methods to different record lengths: sample sizes can vary widely and, while GLS takes this variation into account, TK does not. Finally, we considered the L-moments approach (Hosking and Wallis, 1993, 1997) to fit the LP3 distribution, but we should consider that the GLS procedure implements another approach for fitting LP3 (see Sec. 4.3.1), and this difference could in part affect the results.

Despite the above mentioned constraints given by the assumptions of our analyses, we believe that the study presented in this chapter represents an important contribute for understanding the behaviour of GLS and TK in predicting empirical and true flood quantiles when observed flood sequences are affected by cross-correlation.

4.5 Concluding remarks

The study reported in this chapter addressed the important issue raised by Archfield et al. (2013) of understanding which technique between GLS and TK is better suited for predicting the true unknown flood quantiles in ungauged sites when the observed flood sequences are affected by cross-correlation.

The preliminary LOOCV analyses performed over the real-world study area (i.e. homogeneous region in Triveneto consisting of 20 nested catchments) highlighted that the behaviour of the univariate (i.e. function of the drainage area only) versions of GLS (i.e. 1v-GLS) and TK (i.e. s-TK) applied for predicting the 100-year flood is consistent with the results reported in Archfield et al. (2013). The outstanding performances of s-TK compared to 1v-GLS are totally expected: referring to n neighbouring sites, TK is implicitly able to take some climate and geomorphological similarities between catchments into account, especially in regions with preponderance of nested catchments. On the other hand, the inclusion of more catchment descriptors in the analysis (i.e. mv-GLS, ed-TK) can lead to significantly improved performances, especially for GLS.

Although informative, this preliminary analysis did not address the unsolved issue raised by Archfield et al. (2013) either. In order to shed some light on which is better suited between GLS and TK for predicting the true unknown flood quantiles in ungauged sites, we performed the Monte Carlo simulation experiment described in Sec. 4.4. The application of 1v-GLS, s-TK, mv-GLS and ed-TK in a LOOCV scheme for predicting at-site flood quantiles (with return periods T equal to 10, 30, 50, 100 years) for the three

different cross-correlation scenarios provided us with some significant information about the ability of GLS and TK to predict empirical estimates of flood quantiles and true flood quantiles. First, consistently with what already seen for the real-world region, we observed that the multivariate versions of both GLS and TK are better suited than the corresponding univariate versions (i.e. functions of the drainage area only) for predicting both empirical and true flood quantiles. Moreover, the analyses highlighted an analogous dependence of performances of GLS and TK on the degree of cross-correlation: the higher the regional average cross-correlation, the higher the performance in predicting empirical estimates of flood quantiles and the lower the performance in predicting true flood quantiles. These findings highlighted that the presence of cross-correlation in the region introduces a masking-effect on the flood magnitude for both GLS and TK. This behaviour is totally expected for TK (which explicitly exploits spatial correlation in performing its estimates), whereas for GLS, which should be able to look behind cross-correlation, can be explained by the presence of a residual masking-effect with increasing levels of spatial correlation. In particular, the multivariate versions of GLS and TK show very similar performances: the application of mv-GLS or ed-TK is almost equivalent when significant catchment descriptors are available for describing mean annual flood; for practical estimation in cases like this, one could consider the application of a model-averaging approach between the two candidate models. On the contrary, when only a univariate analysis with drainage area can be performed, the application of TK is recommended, even in the presence of high degrees of spatial correlation.

The findings outlined are valid for the simplified situation which was investigated: homogeneous region with nested catchments and no effect of cross-correlation on mean annual flood. Further analyses regarding heterogeneous regions, no nested catchments, the dependence of mean annual flood on cross-correlation, as well as sensitivity analyses of the results to the presence of different record lengths in the AMS series, are suggested for future studies.

Chapter 5

Prediction of flow-duration curves in ungauged sites across large geographical areas: the potential of geostatistical approaches

5.1 Introduction

Thanks to the increasing accessibility of global datasets on soil, land-cover, morphology, and weather forcing, and enhanced computing capacity, progressively accurate macro-scale rainfall-runoff models have been developed over the last decade (see e.g. Collischonn et al., 2007; de Paiva et al., 2013; Bierkens et al., 2015). Their output can be open-access and freely distributed, providing extremely useful hydrological information across data scarce regions (e.g. Pechlivanidis and Arheimer, 2015) for the implementation of trans-boundary policies for water resources management (e.g. de Roo et al., 2012) or flood-risk mitigation (de Paiva et al., 2013; Sampson et al., 2015; Falter et al., 2016). However, the local performances are highly variable (see e.g. de Paiva et al., 2013; Donnelly et al., 2016), reflecting the quality of macro-scale input data and the adequacy of the conceptual scheme to accurately represent peculiar hydrological processes that locally drive the rainfall-runoff transformation.

On the other hand, geostatistical procedures, such as Top-kriging (TK), have been shown to provide reliable predictions of streamflow points indices over large study areas (see e.g. Castiglioni et al., 2011; Parajka et al., 2015 for low flows, and Archfield et al., 2013 for floods), especially if compared to regional regression models, whose accuracy is generally unsatisfactory for large and highly heterogeneous study regions. The Total

Negative Deviation Top-kriging (TNDTK) recently introduced by Pugliese et al. (2014) has been shown to be reliable for predicting the entire streamflow regime (i.e. flow-duration curves, FDCs) both in Europe (i.e. Pugliese et al., 2014) and in the USA (i.e. Pugliese et al., 2016), but it has never been applied over large geographical areas, which can be characterised by significant heterogeneities in terms of streamflow regimes.

In this context, the present chapter focuses on the use of TNDTK for FDCs in ungauged basins in the Danube region, the largest watershed in Europe. The analyses were developed within a research project with the Joint Research Centre of the European Commission (DG JRC), with the aim of generating a GIS (Geographic Information System) layer reporting the predicted streamflow regime (i.e. FDCs) for about 4000 prediction nodes within the watershed of the Danube River. The DG JRC provided streamflow indices, empirical period-of-record FDCs, and a set of catchment descriptors for a compilation of 511 discharge measurement stations across the Danube river basin.

We first conducted a comprehensive exploration of the relationships between streamflow regime descriptors and the characteristics of the basins. The identified relationships were used to develop multi-regression models for predicting the streamflow indices of interest and for quantifying their predictive accuracy. Subsequently, we interpolated the streamflow regime over the whole Danube river basin, using TNDTK. Then, we discussed the performance of TNDTK in the Danube region and the uncertainty of the interpolation.

The analyses presented in this chapter have been recently published in Castellarin et al. (2018).

5.2 Study area and database

The database used for these analyses was compiled by the DG JRC and consists of 511 streamgauges across the Danube Basin (see Fig. 5.1). For each streamgauge, the DG JRC database provides:

- streamflow indices, computed from the time series of streamflow at each gauge: Mean Annual Streamflow (MAS) and 15 streamflow quantiles Q_d associated with durations d of 1, 5, 10, 20, 30, 40, 50, 60, 70, 75, 80, 90, 95, 97 and 99.7%;
- a set of catchment descriptors: basin area [km^2]; minimum, maximum and mean basin elevation (H_{\min} , H_{\max} and H_{mean} , in this order) [m a.s.l.]; maximum and minimum average daily temperature (T_{\max} and T_{\min} , respectively) [$^{\circ}\text{C}$]; mean annual precipitation (MAP, rainfall) [mm]; mean annual potential evapotranspiration (ET_0) [mm]; mean annual number of rainy days (N_{rd}) [–]; population density for the years 1980, 1990, 2000 and 2005 [inhab km^{-2}]; mean of population densities [inhab km^{-2}];

fractions of cropland, grassland, shrub, bare soil, forest, water, urban, fertilised cropland and fertilised grassland within the total basin area [–].

Moreover, the DG JRC database classifies streamflow data into two different categories:

- high quality (DQ1, blue open circles in Fig. 5.1): gauging stations with a precise positioning along the stream that are unique in their elementary sub-basin (i.e. portion of basin directly drained by a river stretch, between two confluences, or from the headwater to the first confluence);
- low quality (DQ2, red solid dots in Fig. 5.1): cases in which more streamgauges are present in a single elementary basin, hence potentially affected by imprecise positioning along the stream.

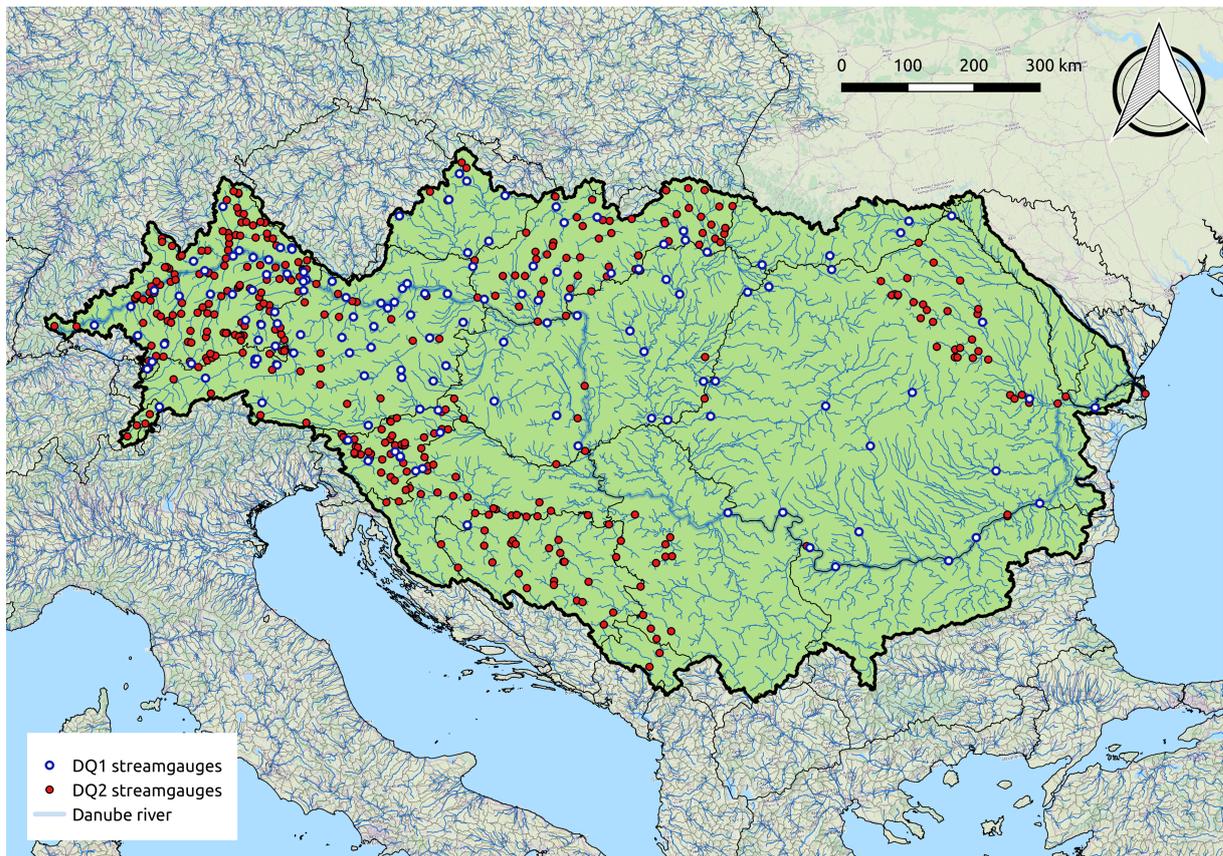


Figure 5.1: 511 streamgauges considered in the Danube region: blue open circles and red solid dots represent high-quality data (DQ1, 138 gauges) and low-quality data (DQ2, 373 gauges), respectively.

The preliminary analyses performed in the present case study considered a selection of streamflow indices and all catchment descriptors reported in the DG JRC database;

a summary of their empirical values for the 511 gauged sites is reported in Tab. 5.1. In particular, concerning streamflow indices, we referred to 95th, 50th and 1st unit streamflow percentiles (i.e. $Q_{1\%}/\text{Area}$, $Q_{50\%}/\text{Area}$ and $Q_{95\%}/\text{Area}$) [$\text{m}^3 \text{s}^{-1} \text{km}^{-2}$]; FDC slope between 70th and 30th streamflow percentiles at logarithmic scale (see e.g. Yaeger et al., 2012) [-]; TND: empirical Total Negative Deviation (a metric of empirical FDCs shape as defined in Pugliese et al., 2014; see Sec. 2.3.2 for further details). Together with the above mentioned streamgauges, the DG JRC identifies 4381 prediction nodes over the Danube region, for which we performed the prediction of FDCs described herein. Note that the analyses performed in this chapter considered streamflow indices without applying any correction able to account for anthropic effects on the streamflow regime.

Table 5.1: Empirical values of a selection of streamflow indices and catchment descriptors for the 511 gauged basins in the Danube region.

	$Q_{95\%}/\text{Area}$ [$\text{m}^3 \text{s}^{-1} \text{km}^{-2}$]	$Q_{50\%}/\text{Area}$ [$\text{m}^3 \text{s}^{-1} \text{km}^{-2}$]	$Q_{1\%}/\text{Area}$ [$\text{m}^3 \text{s}^{-1} \text{km}^{-2}$]	FDC slope [-]	TND [-]	Area [km^2]	H_{\min} [m a.s.l.]
Minimum	0.000026	0.000054	0.000501	0.53	0.83	100.0	-78
25 th %ile	0.001505	0.004248	0.033688	1.43	1.57	298.5	151
Median	0.003582	0.008394	0.054243	1.90	1.85	814.6	315
Mean	0.004456	0.011060	0.087061	1.96	1.84	22352.8	322
75 th %ile	0.006257	0.015250	0.096959	2.34	2.13	4371.5	439
Maximum	0.024340	0.142900	5.036927	6.84	3.10	802032.1	1711
	H_{\max} [m a.s.l.]	H_{mean} [m a.s.l.]	T_{\max} [°C]	T_{\min} [°C]	MAP [mm]	ET_0 [mm]	N_{rd} [-]
Minimum	101	88	2.46	-5.79	447.0	417.6	69
25 th %ile	1025	476	11.21	2.24	699.0	621.9	112
Median	1704	623	12.53	3.69	875.5	672.3	129
Mean	1822	742	12.21	3.25	943.2	665.6	128
75 th %ile	2492	949	13.48	4.5	1113.1	715.0	142
Maximum	3873	2434	17.61	7.25	2033.1	943.6	201
	Population density 1980 [inhab km^{-2}]	Population density 1990 [inhab km^{-2}]	Population density 2000 [inhab km^{-2}]	Population density 2005 [inhab km^{-2}]	Population density mean [inhab km^{-2}]	Cropland	Grassland
Minimum	1.3	1.3	1.3	1.5	1.3	0.00%	0.00%
25 th %ile	48.2	50.4	52.4	52.8	51.0	0.13%	0.00%
Median	76.7	80.9	82.7	84.2	81.7	3.46%	0.56%
Mean	87.7	92.1	94.9	95.5	92.6	7.28%	3.39%
75 th %ile	109.3	114.4	120.3	119.8	115.3	10.91%	4.43%
Maximum	656.8	722.9	645.9	692.1	679.4	36.32%	41.60%
	Shrub	Bare soil	Forest	Water	Urban	Fertilised cropland	Fertilised grassland
Minimum	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
25 th %ile	0.00%	0.00%	39.16%	0.00%	0.00%	4.20%	10.32%
Median	0.00%	0.00%	55.85%	0.00%	0.39%	14.24%	14.05%
Mean	0.11%	0.00%	55.10%	0.39%	0.75%	17.64%	15.34%
75 th %ile	0.00%	0.00%	73.45%	0.02%	1.17%	28.98%	18.22%
Maximum	6.12%	0.00%	99.94%	13.52%	10.61%	66.36%	56.49%

5.3 Relationships between streamflow indices and catchment descriptors

5.3.1 Correlation analysis

The presence of statistically significant correlation between streamflow indices and catchment descriptors was assessed using Pearson and Spearman (rank) correlation coefficients. The Spearman correlations for all 511 gauges (DQ1+DQ2) and for the 138 DQ1 gauges in the Danube region are represented in Fig. 5.2, where size and colours of dots illustrate the empirical correlation coefficients between streamflow regime indices and catchment descriptors, and numbers indicate the p -values associated with the null hypothesis of no correlation between two variables, obtained with the R-function `cor.test` of the package `corrplot` (Wei and Simko, 2016). The results of Pearson correlation (not reported here for conciseness) show slightly lower absolute values of correlation coefficients and generally higher p -values, as expected, since Pearson correlation quantifies the degree of linear dependence between pairs of observations.

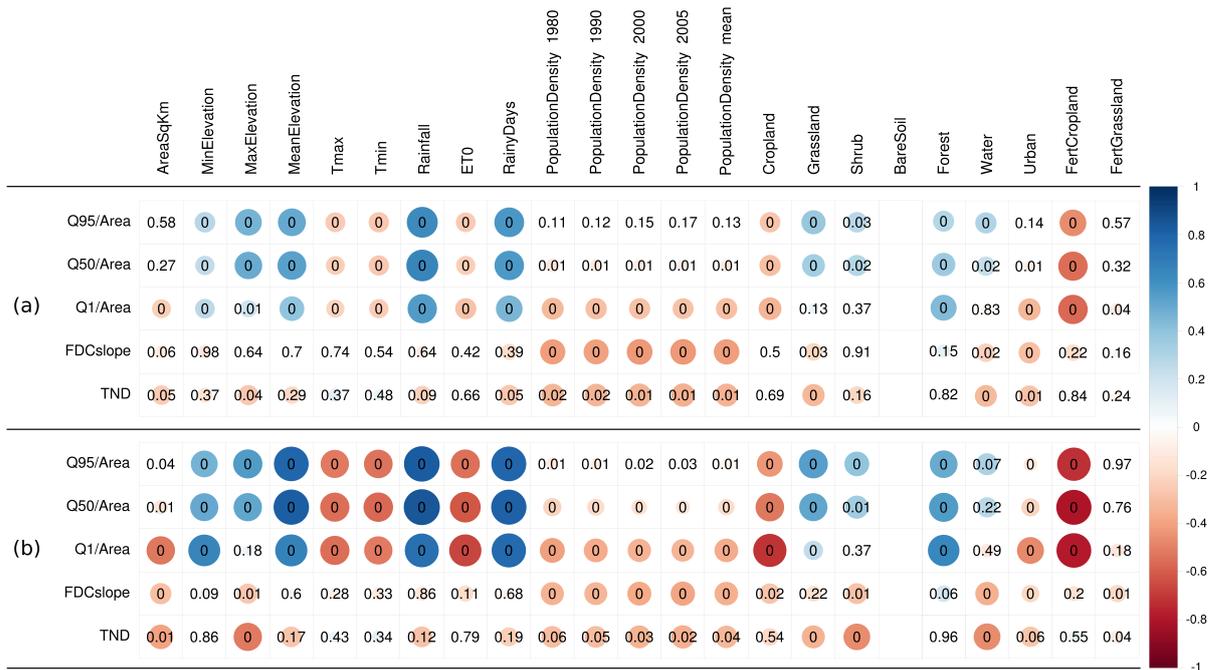


Figure 5.2: Spearman correlation between streamflow regime indices and catchment descriptors of the basins in the Danube region. Colour and radius of each circle is proportional to the value of the empirical correlation coefficient (see colour scale); numbers indicate p -value of the null hypothesis (i.e. absence of correlation between the two variables). (a) 511 basins (DQ1+DQ2); and (b) 138 basins (DQ1), generated with R-package `corrplot` (Wei and Simko, 2016).

As expected, Fig. 5.2 highlights that the correlations become stronger if the analysis

is limited to high-quality data (DQ1 basins). In particular, 95th and 50th unit streamflow quantiles show significant correlations with annual rainfall and number of rainy days, and lower positive correlations can also be observed with reference to maximum and mean catchment elevation, and fractions of total basin area characterised by grassland, bare soil, forest and water. The positive correlation found between $Q_{95\%}/\text{Area}$ (low-flow index) and forested area seems to be in contradiction with empirical evidences (see e.g. Brown et al., 2005), which show lower surface water availability during low-flow periods for catchments with fully developed forested areas (i.e. higher evapotranspiration due to deeper root zones characterising forests and tree plantations relative to field and crops). However, the literature (see e.g. Calder, 1998; Moore and Heilman, 2011) shows that the relation between basin forest cover and streamflow is not univocal.

The strong inverse correlations of streamflow quantiles with mean daily maximum and minimum temperature, annual potential evapotranspiration and fraction of fertilised cropland, are all expected due to the inverse correlation between runoff production and potential evapotranspiration. Significant but weaker inverse correlations were found between unit streamflow percentiles and cropland. The 1st percentile of unit daily streamflows ($Q_{1\%}/\text{Area}$, representing high flow) shows a weak negative correlation with population density (years 1980, 1990, 2000, 2005 and mean value), which is more unexpected. A possible explanation could be the positive correlation between population density and catchment area (large cities are usually found in the lower parts of the rivers), combined with the well-known negative correlation between catchment size and unit flood associated with a low exceedance probability (see e.g. regional envelope curve of flood flows: Castellarin et al., 2005; Castellarin, 2007), which is also found in the Danube region.

Concerning the selected descriptors of FDC shape (i.e. slope and TND, the smaller the value the flatter the curve in both cases), we found significant positive correlation with mean daily maximum and minimum temperature and annual potential evapotranspiration, while negative correlations were observed with the fraction of water and urbanised area. All positive correlations listed above were expected: usually higher temperature and evapotranspiration correspond to more arid climates, where river basin water storage is reduced, causing steeper FDCs; the larger the presence of inland water bodies (e.g. lakes) the larger the natural capability to retard and dampen flood peaks, the flatter the curve. The negative correlation between urbanised area and FDC slope or TND could be analogous to what was observed between high-flow regime and population density; that is, the larger the catchment, the larger the percentage of urbanised areas (big cities and large urbanised area tend to cluster in floodplains, see e.g. Di Baldassarre et al., 2013) the flatter the FDC, due to the increased capability of the catchment to store water (see

Castellarin et al., 2013).

There are some additional statistically significant dependencies, which are particularly pronounced if we limit our attention to DQ1 data (Fig. 5.2(b)). For instance, the analysis points out a significant positive correlation between all unit streamflow percentiles (i.e. $Q_{1\%}/\text{Area}$; $Q_{50\%}/\text{Area}$ and $Q_{95\%}/\text{Area}$) and minimum catchment elevation; it also highlights an inverse correlation between $Q_{1\%}/\text{Area}$ and population density, and a significant inverse correlation between FDC slope and TND and maximum catchment elevation, population density and fractions of grassland, shrub and water. This is a sensible result as flatter flow-duration curves are associated with higher capability of the catchment to temporarily store water volumes; and this capability generally increases with increasing elevation (winter snow-pack), presence of waterbodies, or size of the catchment.

5.3.2 Multi-regression models

We used the above correlation analysis as a basis to identify log-linear multi-regression models for predicting a given dependent variable (i.e. a streamflow index) using catchment descriptors. This was done by applying a stepwise OLS regression analysis (see Draper and Smith, 1981; Weisberg, 1985; Chambers, 1992) using the R-function `lm` in R (R Core Team, 2016). We excluded multi-regression models associated with an adjusted R-squared¹ $R_{\text{adj}}^2 \leq 0.50$.

5.3.3 Results and discussion

We were only able to derive acceptable (i.e. $R_{\text{adj}}^2 > 0.50$) regression models for streamflow indices $Q_{50\%}$ (median discharge), $Q_{1\%}$ (1st percentile) and $Q_{95\%}$ (95th percentile), using DQ1 class gauges only. Table 5.2 shows the details of all the acceptable models.

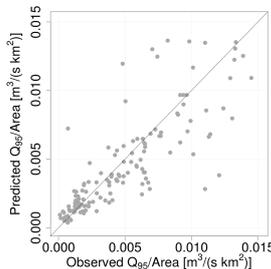
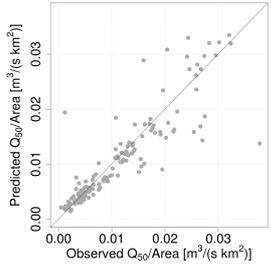
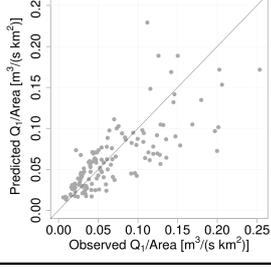
The multi-regression model analysis highlights the following aspects:

- it is not possible to identify effective (i.e. $R_{\text{adj}}^2 > 0.50$) multi-regression models for all streamflow indices of interest in the Danube region;
- high- and low-flow percentiles (i.e. $Q_{1\%}$ and $Q_{95\%}$, respectively) are more difficult to predict than indices of typical streamflow conditions (i.e. $Q_{50\%}$);

¹The adjusted R-squared is defined as $R_{\text{adj}}^2 = 1 - (1 - R^2) \frac{n-1}{n-p-1}$, where p is the total number of explanatory variables (i.e. predictors; constant term excluded), and n is the sample size. Representing an unbiased (or less biased) estimator of the population R^2 , R_{adj}^2 is more appropriate than the observed sample R^2 (which is positively biased) when evaluating model fit and comparing alternative models in the feature selection stage of model building.

- including lower quality streamflow data (DQ2 basins) has a negative impact on model performance;
- multi-regression models are characterised by rather limited accuracy despite significant correlations between predictands and some of the predictors; and
- indices of the FDC shape (i.e. FDC slope and TND) cannot be effectively regressed against any of the available catchment descriptors (predictors).

Table 5.2: Effective multi-regression models (i.e. adjusted R-squared $R_{adj}^2 > 0.50$) identified for the DQ1 class gauges in the Danube region.

No.	Model	R_{adj}^2	Scatterplot
1	$\frac{Q_{95\%}}{\text{Area}} = 1.03 \cdot 10^{-12} N_{rd}^{3.46} H_{max}^{0.71}$	0.73	
2	$\frac{Q_{50\%}}{\text{Area}} = 3.74 \cdot 10^{-10} MAP^{0.97} H_{mean}^{0.53} H_{max}^{0.34} N_{rd}^{0.92}$	0.80	
3	$\frac{Q_{1\%}}{\text{Area}} = 2.40 N_{rd}^{2.19} ET_0^{-2.22}$	0.66	

In other words, multi-regression models are not capable of accurately representing streamflow quantiles across all durations (from high-flow to low-flow quantiles) and study area (i.e. high- and low-quality gauges, DQ1+DQ2). Moreover, the unsupervised stepwise regression procedure used in the analysis does not select any catchment descriptor (i.e. predictor) associated with anthropogenic pressure or human presence in the catchment as an explanatory variable in any of the models.

This makes rather evident that resorting to macro-scale multi-regression models is not a viable approach for predicting the streamflow regime in ungauged basins located

in the Danube region. For this reason, we interpolated the empirical FDCs over the stream network of the Danube basin using the geostatistical method recently proposed by Pugliese et al. (2014, 2016), briefly outlined in the next section.

5.4 Top-kriging interpolation of flow-duration curves

5.4.1 Implementation of TNDTK to the Danube region

We applied the procedure presented above to the entire Danube region. All analyses were carried out by applying the R-package `rtop` (Skøien, 2014). We selected the mean annual streamflow (MAS) as the reference streamflow value for standardising empirical FDCs across the study region. The MAS values are available from the database as long-term average daily discharges. Concerning DQ1 basins located in the Danube region, the minimum value, 25th percentile, median, mean value, 75th percentile and maximum value of empirical MAS are equal to 0.640, 5.90, 28.7, 527,184 and 6380 m³/s, respectively. Figure 5.3 illustrates the values of MAS standardised by catchment area as a function of basin area for the study region. Based on values illustrated in Fig. 5.3 and some preliminary TNDTK runs, we regarded as highly discordant all values of MAS/Area outside the interval 0.0015–0.08 m³ s⁻¹ km⁻². All basins with empirical MAS/Area values falling outside this interval were therefore excluded from further analyses. We can observe that all but one of the 14 discarded basins are associated with low-quality (DQ2) streamgauges, which further highlights the low reliability of these outlying values.

Therefore, as the Danube region includes a large number of low-quality measurement points (i.e. DQ2 streamgauges, see Fig. 5.1), we decided to perform all analyses twice, first by focusing only on high-quality data (i.e. DQ1 measuring points, or 137 catchments) and then by considering low- and high-quality data combined (i.e. DQ1+DQ2 measuring points, or 497 catchments). Note that, in order to properly investigate the effect of data quality on the reliability of the estimates, the analysis of the DQ1+DQ2 dataset was performed without adopting any specific weighting scheme relative to data quality (i.e. high- and low-quality data were treated in the same way). The results of both analyses are reported in the figures in double-panel layouts. Top-kriging has been applied by fitting the sample variogram of the empirical TND values with a 5-parameter fractal-exponential model (for details, see Skøien et al., 2006) through a modified version of weighted least squares regression (WLS; Cressie 1993; for details, see also the neutral WLS method in `rtop`, Skøien 2014). The fitted variogram model was then used to evaluate the kriging weights for all ungauged sub-basins, based on the n closest neighbouring gauges. Standardised FDCs were then predicted at locations of interest through Eq. (2.25). After

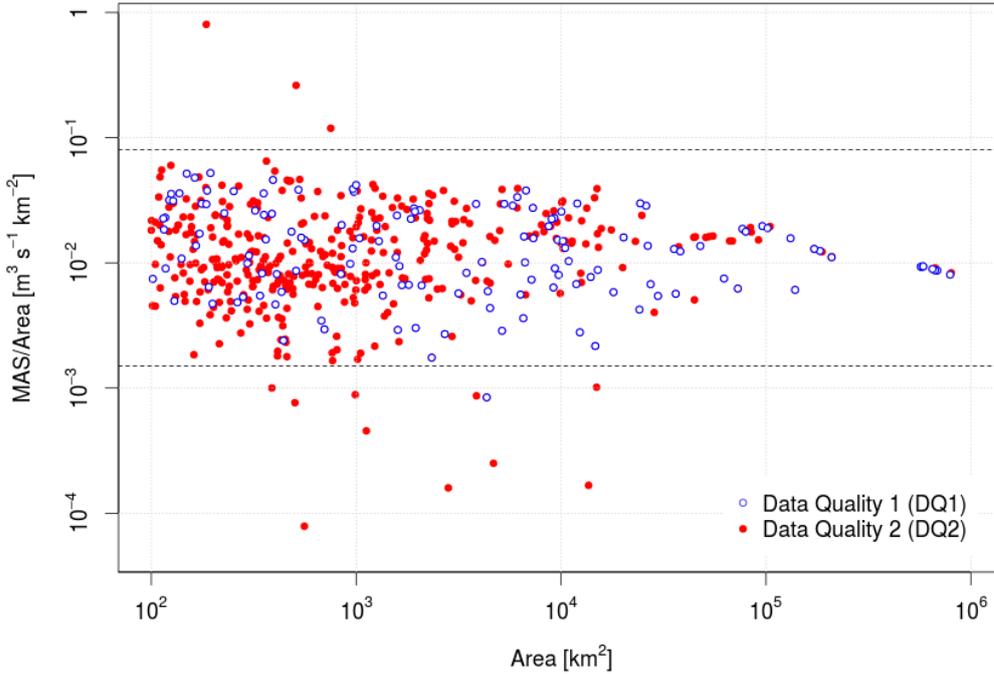


Figure 5.3: Unit mean annual streamflow (MAS/Area) as a function of basin area for high- (DQ1, blue open circles) and low-quality (DQ2, red solid dots) measurement points in the Danube region. The points outside the highlighted interval (0.0015–0.08 m³ s⁻¹ km⁻²; dashed lines) identify extremely discordant sites.

a preliminary sensitivity analysis, we set $n = 6$ in line with previous studies, suggesting to limit the size of the kriging neighbourhood when interpolating streamflow indices, and standardised FDCs in particular (see e.g. Pugliese et al., 2014, 2016). The prediction of dimensional FDCs at locations of interest via Eq. (2.26) requires prediction of the local MAS value, which we achieved via a traditional application of Top-kriging that uses the same settings listed above (i.e. a modified exponential variogram fitted via WLS regression, neighbourhood size $n = 6$).

5.4.2 Cross-validation procedures

The reliability and robustness of (a) Top-kriging for predicting MAS values, and (b) TNDTK for predicting FDCs in ungauged basins, were quantitatively assessed by means of three different validation strategies, useful also for better understanding the dependence of the prediction performance on the spatial density of the empirical data. In particular, three leave-p-out cross-validation procedures (LPOCVs) were performed, in which p coincides with 1 site (LPOCV-1), one third of the sites (LPOCV- $\frac{1}{3}$) and one half of the sites (LPOCV- $\frac{1}{2}$). All three resampling procedures simulate ungauged conditions at each and every site belonging to the network of N measuring points:

- LPOCV-1 drops, in turn, one site at a time and performs the prediction of the streamflow indices of interest in that very site on the basis of the remaining $N - 1$ measuring points;
- LPOCV- $1/3$ (or LPOCV- $1/2$) randomly subdivides the N gauged sites into three (or two) subsets and predicts the streamflow indices of interest in all sites belonging to one subset on the basis of the data available at the remaining $2/3$ (or $1/2$) sites.

LPOCV-1, LPOCV- $1/3$ and LPOCV- $1/2$ were applied for both DQ1 and DQ1+DQ2 subsets. Finally, we combined LPOCV predictions of MAS and dimensionless FDCs by using Eq. (2.26) to obtain cross-validation predictions of dimensional FDCs at each gauging site in the Danube region.

The performances were evaluated in terms of Nash-Sutcliffe efficiency between empirical and predicted log-transformed (LNSE) and natural (NSE) values. The general definition of NSE and LNSE is given by the following equations:

$$NSE = 1 - \frac{\sum_{j=1}^N (\hat{x}_i - x_i)^2}{\sum_{j=1}^N (\hat{x}_i - \bar{x})^2} \quad (5.1)$$

$$LNSE = 1 - \frac{\sum_{j=1}^N (\ln(\hat{x}_i) - \ln(x_i))^2}{\sum_{j=1}^N (\ln(\hat{x}_i) - \ln(\bar{x}))^2} \quad (5.2)$$

where N is the number of observations, \hat{x}_i the estimated variable at site i , x_i the observed value of the variable at site i , and \bar{x} its mean value. NSE and LNSE can range from $-\infty$ to 1, where an efficiency of 1 corresponds to a perfect match of modeled discharge to the observed data, and an efficiency of 0 indicates that the model predictions are as accurate as the mean of the observed data. An efficiency less than zero ($NSE < 0$) occurs when the observed mean is a better predictor than the model or, in other words, when the residual variance (described by the numerator), is larger than the data variance (denominator). Essentially, the closer the model efficiency is to 1, the more accurate the model is. Threshold values to indicate a model of sufficient quality have been suggested between 0.5 and 0.65.

Concerning the prediction of MAS, we quantified the regional accuracy in terms of regional Nash-Sutcliffe efficiency between empirical and predicted log-transformed (LNSE) and natural (NSE) MAS values; concerning the prediction of dimensionless and dimensional FDCs we computed LNSE and NSE values either globally (i.e. assessing overall

LNSE and NSE values across all sites and durations, or across all sites but duration-wise) and locally (i.e. at each gauge on the basis of the 15 interpolated streamflow quantiles). Note that the comparison between LNSE and NSE values is important for better understanding the efficiencies of TNDTK for low flows (LNSE) and high flows (NSE).

5.4.3 Results and discussion

Figures 5.4, 5.5 and 5.6 present, in a similar fashion, the results obtained relative to MAS, dimensionless FDCs and dimensional FDCs (dimensionless and dimensional curves are described through 15 streamflow quantiles). Scatter diagrams distinguish between DQ1 and DQ1+DQ2 subsets and report empirical values vs predictions for the three different resampling strategies used in the study.

Concerning cross-validated predictions of dimensional FDCs, Fig. 5.7 reports the distributions of local LNSE and NSE values for both DQ1 and DQ1+DQ2 subsets and all resampling strategies, while Fig. 5.8 illustrates LNSE values computed across all DQ1 (or DQ1+DQ2) sites as a function of duration and resampling strategy. Figure 5.9 shows the comparison between observed and interpolated FDCs for the two gauges having the best and the worst performances in terms of LNSE values for DQ1 - LPOCV-1.

Given the paramount importance of FDCs for many water-resources management applications (see e.g. Vogel and Fennessey, 1995; Yaeger et al., 2012), the accuracy of interpolated FDCs needs to be properly assessed.

Viability of geostatistical prediction of FDCs over large geographical regions

Figures 5.4, 5.5, 5.6, 5.7 and 5.8 illustrate an overall good agreement between empirical indices of streamflow regimes and their predictions for all three resampling strategies used in cross-validation. In particular, the scatter diagrams between empirical and predicted MAS values in Fig. 5.4 highlight a very good agreement between observed and predicted values, with the majority of points falling in the vicinity of the one-to-one line; as a result, LNSE and NSE values are rather high both for high-quality data (DQ1) and high- and low-quality data (DQ1+DQ2). The overall prediction performance appears almost independent on the resampling strategy, and the detriment of cross-validation predictions remains limited when moving from LPOCV-1 to LPOCV- $\frac{1}{3}$, or to LPOCV- $\frac{1}{2}$. It is worth stressing here that the three cross-validation procedures base all predictions only on 136, 68 and 45 measurement points if we consider the DQ1 subset, and a significantly larger number of streamgauges (i.e. 496, 248 and 165) when DQ1+DQ2 subset is considered. As Top-kriging is a geostatistical procedure, its prediction performance should increase with the density of the gauging network. This effect is visible when looking at the LNSEs and

NSEs obtained for a single data subset, where LNSE and NSE values slightly decrease for DQ1 (or DQ1+DQ2) when moving from LPOCV-1 to LPOCV- $1/3$, and LPOCV- $1/2$. Yet, the same consideration does not hold across datasets, that is, when comparing the results of the same resampling strategy for DQ1 and DQ1+DQ2. The higher number of streamgauges included in the DQ1+DQ2 subset does not result in better MAS prediction due to the lower quality of the streamflow data collected at the additional measuring points. Figure 5.5 illustrates the performance of TNDTK for predicting dimensionless FDCs. These scatterplots show an excellent agreement between predictions and empirical data. Overall LNSE and NSE values are well above 0.8 for both subsets and all three resampling strategies. As for the results for MAS predictions, changing the resampling strategy shows a very limited impact on predicted dimensionless FDCs. Including additional streamflow data of lower quality (i.e. DQ2 basins) does not have any significant effect on predicted MAS values and dimensionless FDCs, and therefore the empirical streamflow regime is captured equally well by DQ1 and DQ2 subsets.

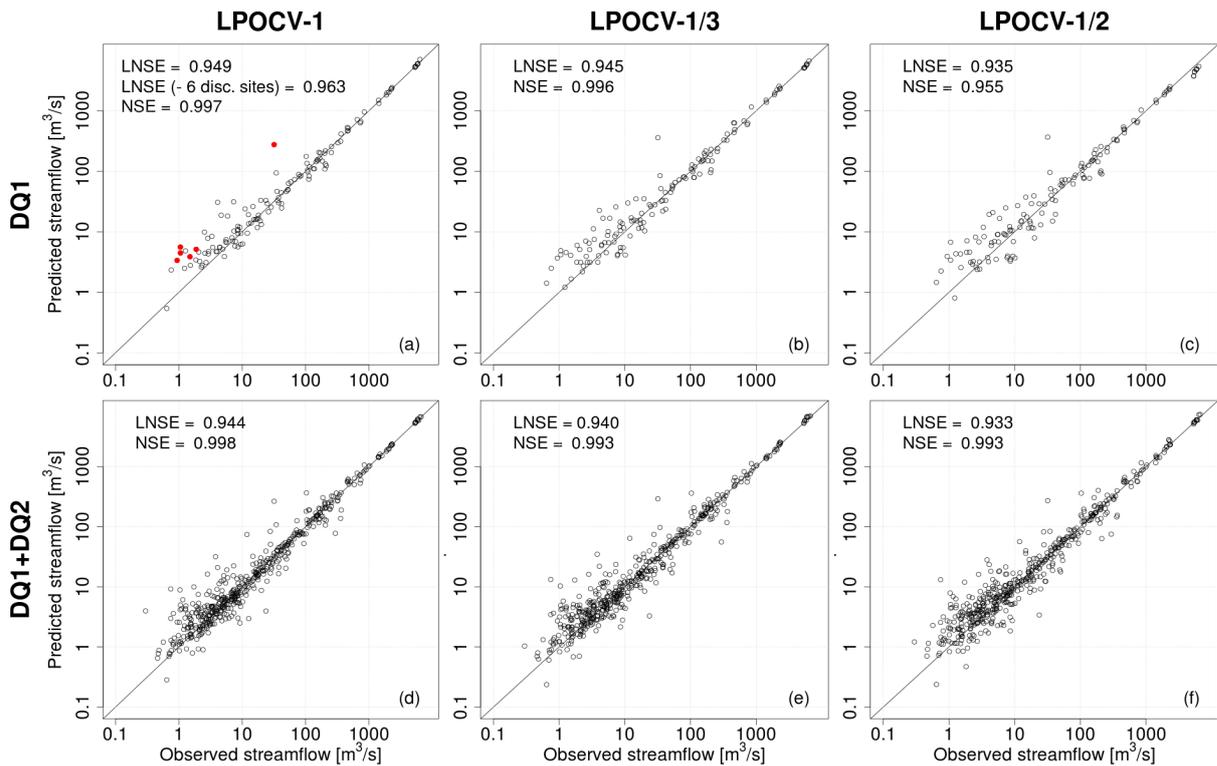


Figure 5.4: Top-kriging interpolation of mean annual streamflow (MAS) values in cross-validation: empirical (x-axis) vs predicted (y-axis) MAS and Nash-Sutcliffe efficiency for log-transformed (LNSE) and natural (NSE) streamflows. See Sec. 5.4.2 for the three different resampling strategies used in cross-validation: LPOCV-1 (a, d), LPOCV- $1/3$ (b, e) and LPOCV- $1/2$ (c, f). The LPOCV-1 cross-validated predictions of MAS for the six DQ1 gauges associated with the worst prediction of dimensional FDCs are highlighted (red solid dots).

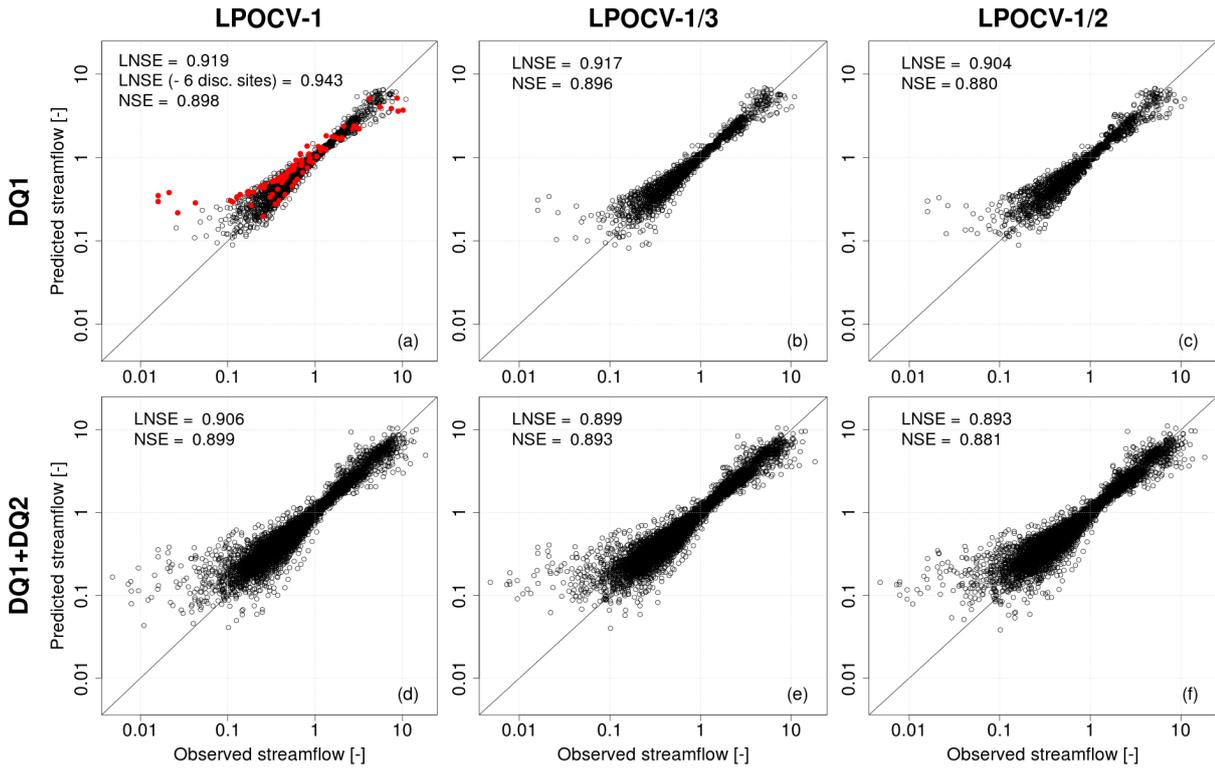


Figure 5.5: Top-kriging interpolation of standardised FDCs (each empirical curve is standardised by local mean annual streamflow) in cross-validation: empirical (x-axis) vs predicted (y-axis) dimensionless streamflow quantiles and overall NSE for log-transformed (LNSE) and natural (NSE) streamflows.

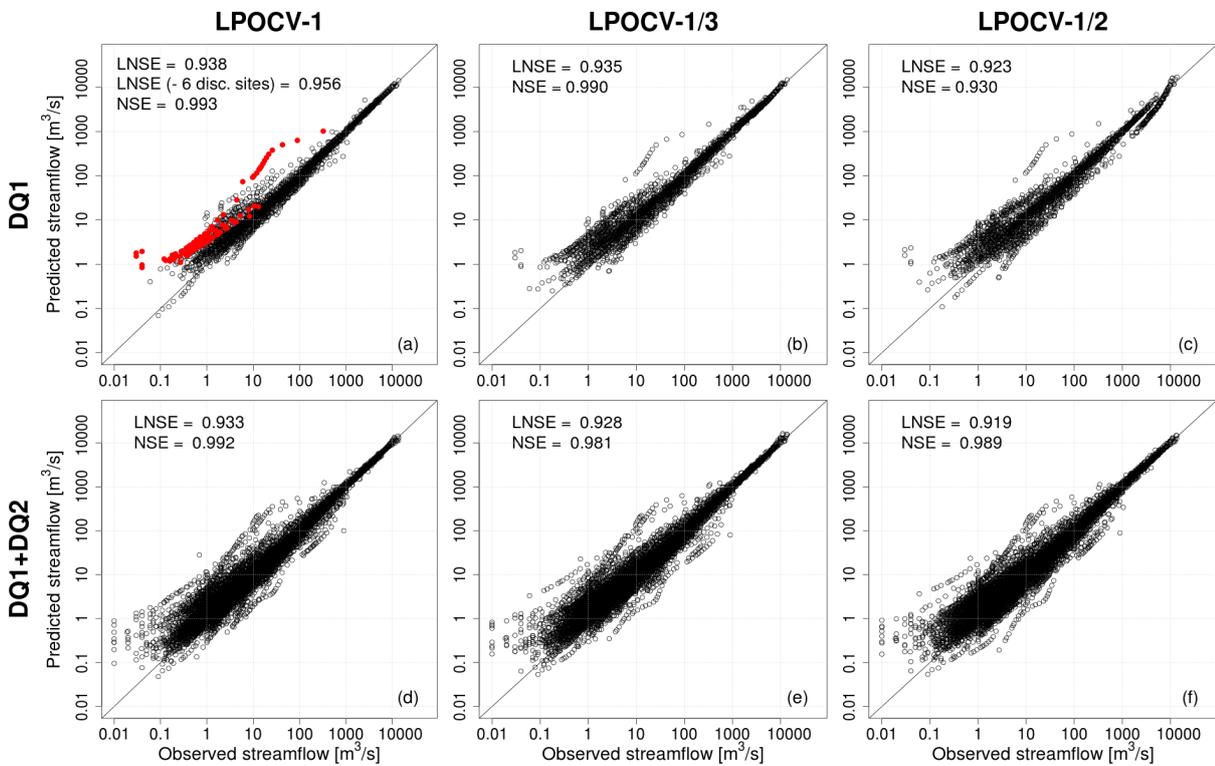


Figure 5.6: Top-kriging interpolation of dimensional FDCs in cross-validation: empirical (x-axis) vs predicted (y-axis) dimensionless streamflow quantiles and overall Nash-Sutcliffe efficiency for log-transformed (LNSE) and natural (NSE) streamflows.

Figure 5.6 shows the relationship between empirical and predicted FDCs in a similar fashion to Fig. 5.5. The cross-validation exercise shows outstanding performance, with overall LNSE and NSE values above 0.9, and the detriment of prediction performance associated with the reduction of gauging network density is, again, very limited. In fact, the scatterplots of Fig. 5.6 show that the overall LNSE values might be significantly impacted by a very limited number of dimensionless FDCs that are poorly predicted. To further discuss this point, LPOCV-1 panels in Fig. 5.4, 5.5 and 5.6 highlight (in red) the predictions of MAS, dimensionless and dimensional FDCs obtained for six DQ1 gauges associated with very poor prediction of dimensional FDCs (i.e. the six predicted FDCs are associated with the lowest at-site LNSE values). Closer inspection reveals that these six gauging points are all located in areas where the station density is high, and therefore the low performance should not be attributed to the lack of hydrological information. Figure 5.4 reveals that poor predictions in terms of dimensional FDCs are mainly associated with poor prediction of MAS, and that five out of six catchments are associated with low or very low empirical values of MAS. In fact, five out of six discordant sites are headwater catchments, for which Top-kriging has been already shown to be less effective than for medium to large catchments (see e.g. Castiglioni et al., 2011; Laaha et al., 2014), and whose mean annual streamflow is likely to be altered by e.g. manmade diversions. The same consideration (i.e. significantly altered streamflow regime) may apply also to larger catchments.

Aside from a small number of peculiar sites, Fig. 5.4, 5.5 and 5.6 show a generalised excellent agreement between empirical and predicted streamflow indices and flow-duration curves.

Figure 5.7 details the local prediction performances through a box-plot representation of the distributions of at-site LNSEs and NSEs between empirical and predicted dimensional FDCs (LNSE and NSE values are computed on the basis of 15 streamflow quantiles; box-plots are truncated at LNSE = 0 and NSE = 0, respectively). It can be seen that, for both DQ1 and DQ1+DQ2 datasets and all three resampling strategies, more than 50% of the predictions are associated with at-site LNSE and NSE values that are above 0.8; in almost all cases 75% of predicted FDCs correspond to LNSE and NSE values in excess of 0.5 (the one exception is for LNSEs in LPOCV- $\frac{1}{2}$ for the DQ1 dataset). Figure 5.7 also shows that there are outlying sites with very low, and sometimes negative, LNSE and NSE values; in particular, negative LNSE (NSE) values are obtained in a number of cases that varies from a minimum of 10.9% (12.3%) to a maximum of 16.3% (16.1%) of sites, corresponding to DQ1+DQ2-LPOCV- $\frac{1}{3}$ (DQ1+DQ2-LPOCV-1) and DQ1-LPOCV- $\frac{1}{2}$ (DQ1-LPOCV- $\frac{1}{3}$), respectively. All these cases are associated with

poor predictions of MAS (see also Fig. 5.4). The box-plots of Fig. 5.7 clearly illustrate the decrease in prediction performance associated with the three considered resampling procedures (and the corresponding reduction of gauging network resolution), which is more evident if results are analysed on an at-site basis relative to the overall performance illustrated in Fig. 5.6.

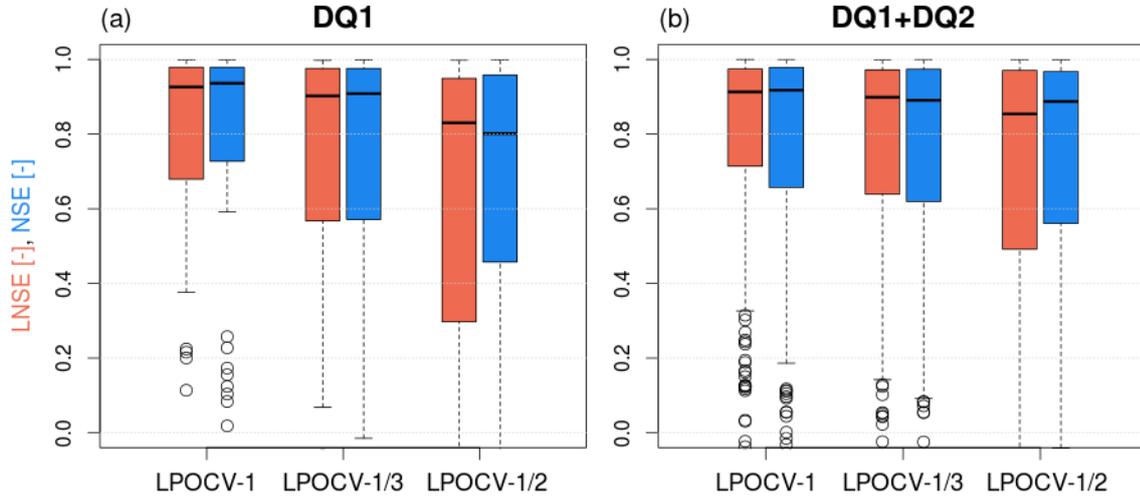


Figure 5.7: Cross-validation of predicted dimensional FDCs: box-plots of LNSE (red) and NSE (blue) values computed for all (a) DQ1 and (b) DQ1+DQ2 measurement points for three different resampling strategies used in the study (see Sec. 5.4.2). Each box shows 25th, 50th (i.e. median) and 75th percentiles; whiskers indicate the most extreme data points that are no more than 1.5 times the inter-quartile range (difference between 75th and 25th percentiles) from the box; outlying values are indicated as circles.

Finally, the LNSE values computed by comparing duration-wise predicted and empirical streamflow quantiles across all sites for the 15 durations considered in the study (Fig. 5.8) indicate very good performance in all cases, and slightly decreasing for increasing durations but generally well above 0.9 and above 0.85 for all durations and both subsets DQ1+DQ2 and DQ1. The results are similar in terms of NSE values, but are not reported here for the sake of conciseness. Figure 5.8 confirms the limited impact of reducing the gauging network density through the different resampling strategies; it indicates a high robustness of TNDTK and shows a limited dependence of prediction performance on duration. A slightly worse performance can be noted in the low-flow section of the curves, which was expected. The TNDTK approach features a homogeneous prediction accuracy across all durations, differently from conventional quantile regression techniques (see e.g. Castellarin et al., 2013; Blöschl et al., 2013), whose application to the study area resulted in significantly lower efficiencies (see Tab. 5.2: LNSEs for 95th, 50th and 1st percentiles for DQ1 class gauges).

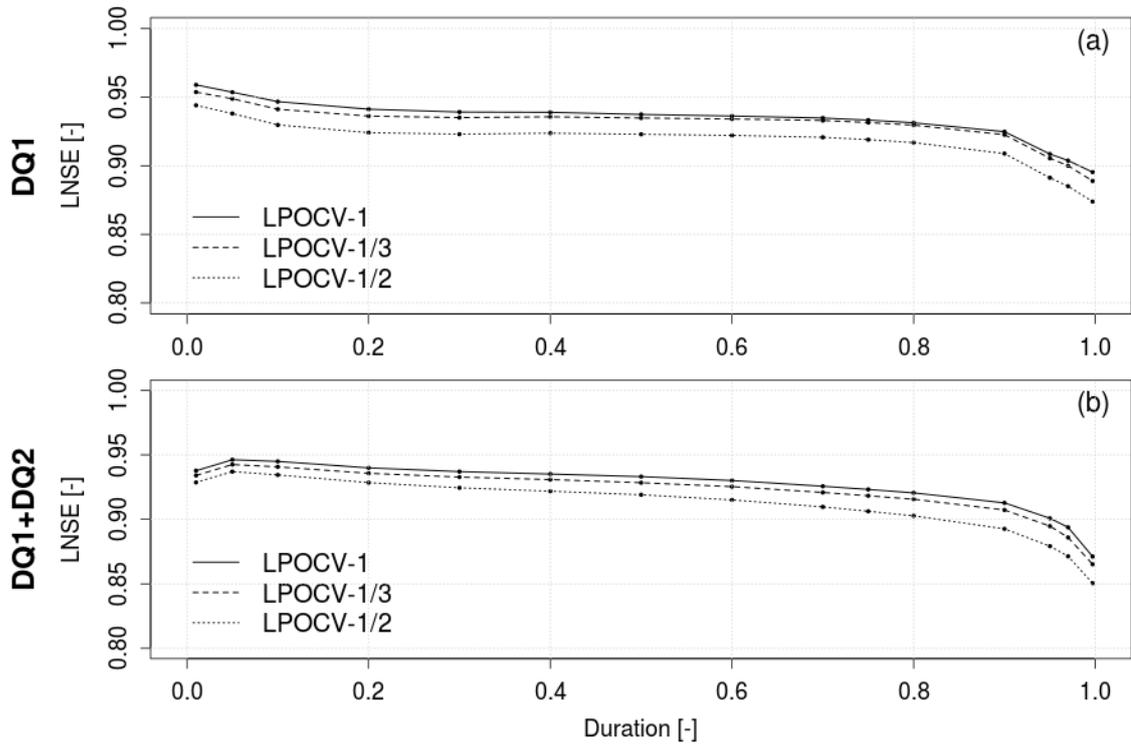


Figure 5.8: Cross-validation of predicted dimensional FDCs: LNSE values computed across all (a) DQ1 and (b) DQ1+DQ2 measurement points as a function of duration; different curves refer to the three different resampling strategies.

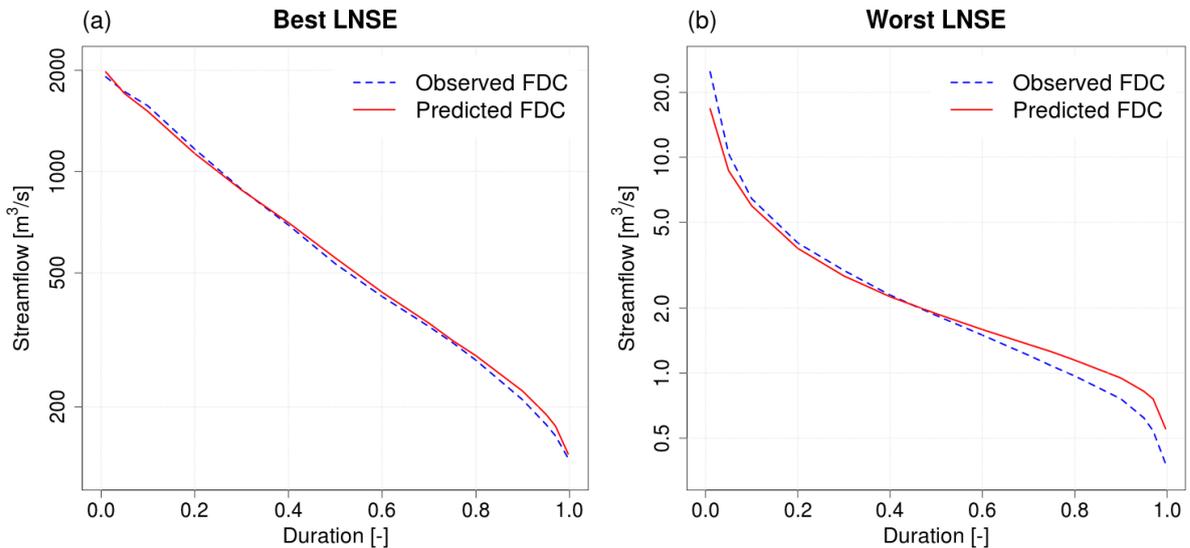


Figure 5.9: Observed and predicted dimensional FDCs for the two catchments having (a) the best and (b) the worst performances in terms of LNSE in LPOCV-1 for DQ1 measurement points.

It is worth emphasising that the overall LNSE and NSE values are 0.923 and 0.930, respectively, for 137 interpolated FDCs which were predicted in cross-validation on the basis of 45 measuring points (i.e. less than one gauge per 17 500 km² in the study area), which proves the effectiveness of TNDTK for the interpolation of FDCs over large regions (Pugliese et al., 2016).

Figures 5.4, 5.5, 5.6 and 5.7 do not show significant differences between efficiencies computed in terms of LNSE or NSE, meaning that TNDTK performances for high and low flows are equivalent. In particular, discharge values reported on y-axis in Fig. 5.9 allow us to confirm that TNDTK performs best for larger catchments, while performances get lower for smaller catchments, especially headwater ones (see e.g. Castiglioni et al., 2011; Laaha et al., 2014). In both cases (best and worst LNSE), comparison between the lower tails of observed and predicted FDCs confirms that TNDTK tends to overestimate low flows (see Pugliese et al., 2016).

Indicators of the reliability of interpolated FDCs over large areas

Given the similarity between results in terms of LNSE and NSE, we decided to present the assessment of the reliability of interpolated FDCs by referring to LNSE values only. The maps in Fig. 5.10 highlight the LPOCV-1 for the gauged elementary sub-catchments with an at-site efficiency of cross-validated FDCs (LNSE) lower than 0, between 0 and 0.7 or higher than 0.7. As expected, we can observe that the best performances are typically obtained for nested catchments, large and very large Danube sub-catchments and nodes where station density is higher, while lower performances are associated with headwater catchments located in low station-density areas. It would be extremely useful if statements on the expected accuracy were attached to all interpolated FDCs; unfortunately, LNSE values for cross-validated flow-duration curves are available only for gauged elementary catchments (see Fig. 5.10). A possible measure of prediction uncertainty is the kriging variance (i.e. estimate of the interpolation error), which can be derived for any kriging interpolation scheme and, as such, is an output of each Top-kriging application. This statistic is a combination of model uncertainty and configuration of observation locations; so that lower kriging variances are expected for large prediction catchments that are surrounded by several streamgauges, whereas higher variances are expected for prediction nodes located in data scarce subareas and in upstream catchments.

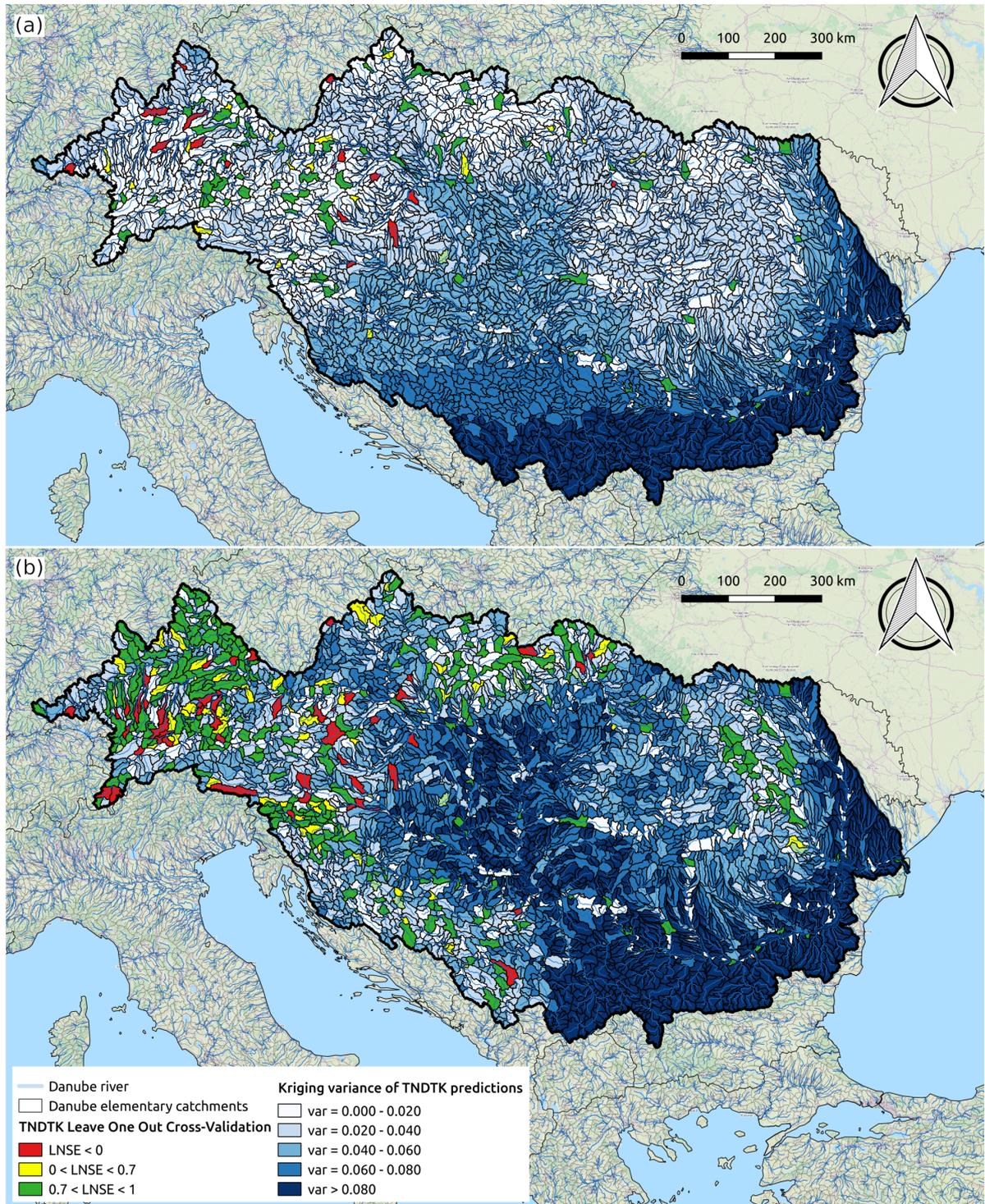


Figure 5.10: Prediction variance and local cross-validation LNSE for Danube region elementary catchments: local LNSE values obtained in cross-validation (LPOCV-1 sampling strategy) at (a) 137 DQ1 streamgauges and (b) 497 DQ1+DQ2 streamgauges are colour-coded; kriging variance is also illustrated.

Figure 5.11 illustrates standardised prediction variances (y-axis) resulting from Top-kriging interpolation of empirical TND values as a function of LNSE values of cross-validated FDCs (LPOCV-1 cross-validation, x-axis). Standardisation of kriging variances was performed by dividing each value by 0.073, which is the maximum kriging variance computed for the study region and refers to the DQ1+DQ2 dataset. Figure 5.11 also reports the rolling mean for a subset of 30 catchments.

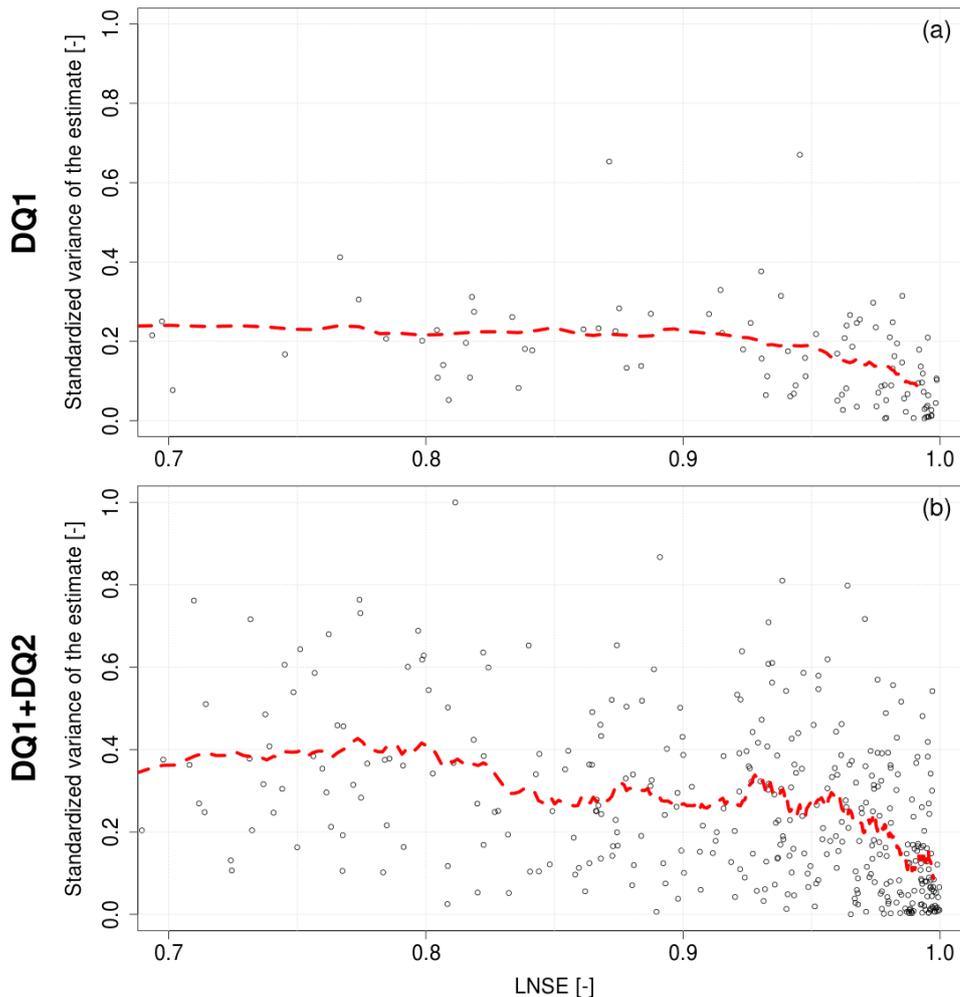


Figure 5.11: Standardised kriging variance for TNDTK interpolation procedure as a function of LNSE for (a) DQ1 and (b) DQ1+DQ2 subsets. LNSE values smaller than 0.7 are omitted. Dashed (red) lines represent the rolling mean computed with a rolling window of 30 catchments.

Figure 5.11 confirms that higher LNSE values are associated with lower kriging variances; the relationship is clearer for the DQ1+DQ2 subset due to the larger sample size, but it is visible also for DQ1. Moreover, despite the larger number of gauges, the DQ1+DQ2 subset is associated with higher kriging variances relative to DQ1, which is yet another indication of the higher uncertainty and noise of the streamflow information

coming from DQ2 streamgauges (see Fig. 5.11).

Therefore, kriging variance can be used as a proxy for uncertainty of predicted FDCs. Kriging variance is graphically illustrated in Fig. 5.10 for each ungauged elementary catchment belonging to the Danube region using a colour scale (the darker the colour blue, the higher the variance). It is evident that, in both cases, prediction variance tends to be lower where station density is higher. The comparison between the two maps points out that integrating the gauging network with DQ2 streamgauges may enable one to locally reduce the prediction variance (see e.g. the North-eastern portion of the Danube region). Nevertheless, a weighted average of the kriging variance that weights the information proportionally to the size of the considered elementary catchment is equal to 0.042 for DQ1 subset and to 0.060 for the DQ1+DQ2 subset, and therefore significantly larger for the latter subset. This is consistent with what is reported in Fig. 5.10, which shows that kriging variance for DQ1+DQ2 is significantly larger than for DQ1 in the central portion of the Danube region. Therefore, adding catchments with less accurate streamflow data (DQ2 subset, see Fig. 5.1) negatively impacts the capability of the geostatistical interpolation procedure to represent the streamflow regime in the central portion of the study region.

This behaviour is effectively illustrated in Fig. 5.12, which refers to 360 DQ2 catchments and shows scatterplots of empirical vs geostatistically predicted MAS values (Top-kriging), together with dimensional and dimensionless FDCs (TNDTK procedure). These geostatistical predictions are entirely based on the data collected at 137 DQ1 measuring points. As illustrated in Fig. 5.12, the overall performance is analogous to performances illustrated in Fig. 5.4, 5.5 and 5.6.

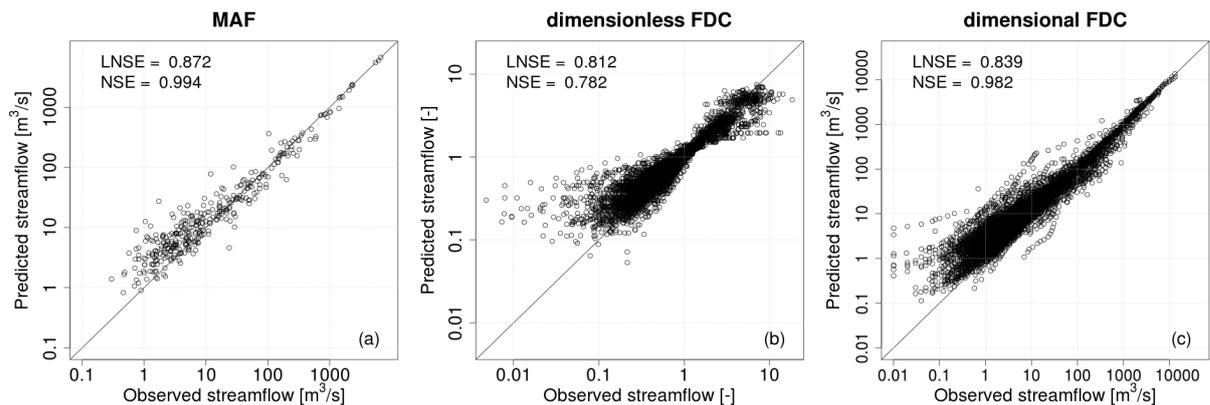


Figure 5.12: Top-kriging interpolation in cross-validation, empirical (x-axes) vs predicted (y-axes): (a) mean annual streamflow (MAS), (b) dimensionless flow-duration curves (FDCs), (c) dimensional FDCs. Predictions refer to 360 DQ2 catchments and are based on observations collected at 137 DQ1 measuring points.

The slight decrease in terms of prediction accuracy relative to DQ1+DQ2-LPOCV- $1/2$

(i.e. panel (f) in Fig. 5.4, 5.5 and 5.6) is to be expected, and results on the one hand from the reduction of the available empirical data on which interpolation is based (e.g. the ratio between measuring and prediction points is equal to $137/360 = 0.38$ in this case, while it is 1 for DQ1+DQ2-LPOCV- $1/2$), and on the other hand from the poorer quality of streamflow data collected at DQ2 measuring points, which has been highlighted above.

On the basis of these considerations, we decided to use the DQ1 subset to predict FDCs over the whole study region (i.e. 4381 prediction nodes), and we used the kriging variance as an indicator of prediction uncertainty.

5.5 Concluding remarks

This chapter presented the statistical regionalisation of streamflow regimes in the Danube region, the largest watershed in Europe, with the aim of testing regional regression models and a geostatistical method for predicting the entire streamflow regime (i.e. flow-duration curves) in ungauged basins over large and hydrologically heterogeneous geographical areas.

The comprehensive exploration of the relationships between streamflow regime descriptors and the characteristics of the basins highlighted that, even if streamflow indices are significantly correlated to catchment characteristics within the Danube region, their prediction using multi-regression models may not be satisfactory. A much improved regionalisation of empirical flow-duration curves has been obtained for more than 4000 sub-basins in the Danube river basin by using the Total Negative Deviation Top-kriging method (TNDTK; see Pugliese et al., 2014, 2016), which was shown to be an effective and accurate interpolation technique across the entire study region. Although the spatial density of streamgauging network affects the estimation variance of interpolation, it has been proven that the regionalisation becomes more accurate when low-quality measurements are discarded. The maps of streamflow quantiles presented herein may be useful for the evaluation of water resources availability at ungauged locations, and as a benchmark for the development of hydrological macro-scale models.

Conclusions

This Thesis focuses on the estimation of hydrological design variables in ungauged or scarcely gauged basins and performs a threefold analysis, addressing three relevant issues in statistical regionalisation of hydrometric information: the value of catchment similarity in the regionalisation of flood flows, the impact of spatial correlation on regional predictions of flood flows, and the potential of geostatistical approaches for predicting flow-duration curves in ungauged sites across large geographical areas.

First, concerning the value of catchment similarity in regionalisation of flood flows, we focused on Triveneto, a broad mountainous geographical area in North-eastern Italy, for which the reference procedure for design flood estimation in ungauged or scarcely gauged basins is available from the Italian CNR research project VAPI (Villi and Bacchi, 2001). In particular, the VAPI project considered Triveneto as a single homogeneous region and developed a regional model using annual maximum series (AMS) of peak discharges that were collected up to the 1980s. Our analyses considered a very detailed AMS database for 76 catchments located in the study area, including historical data together with more recent data for the last 40 years (data spanning from 1913 to 2013). Considering this significantly updated database, the aim was to test the viability of the unique growth curve proposed for operational purposes by Villi and Bacchi (2001), who at the same time warned about the possible presence of a not fully homogeneous behaviour of the region itself. Our study confirms this last warning, showing that the hypothesis of homogeneity does not hold for Triveneto and highlighting the need of updating the reference procedure for design flood estimation in the study area. To this aim, we dispensed with the hypothesis of a single homogeneous region, and the more general concept of geographically contiguous homogeneous regions altogether, and referred to a focused-pooling approach (i.e. Region of Influence, RoI Burn, 1990; Zrinji and Burn, 1994, 1996), which delineates homogeneous pooling-groups of sites for any given target site referring to selected climatic and geomorphological descriptors which result to be relevant for describing regional flood frequency. In this context, we confirmed the value of including in statistical regionalisation physiographic and climatic information, such as drainage area and mean annual precipitation, showing the similarity of climatic and scale controls on flood frequency regime in

Triveneto with the controls recently found in Europe by Salinas et al. (2014). The test of the RoI approach for the estimation of the 100-year flood in more than 20 artificial reservoirs in the study area highlighted the significant improvement in terms of heterogeneity degree for the RoI pooling-groups compared to the entire Triveneto region, and confirmed the RoI approach to be a valid candidate for updating the reference procedure for design flood estimation in the study area. These analyses highlighted the added value of explicitly accounting for catchment similarity in statistical regionalisation: focused-pooling approaches accounting for the control of geomorphological and climatic characteristics on flood frequency regime are preferable to approaches based on geographically contiguous regions, especially in highly heterogeneous areas.

Second, we addressed the impact of spatial correlation on regional predictions of flood flows by referring to regional procedures proposed in the literature, such as some regression methods and geostatistical methods, that do address the presence of spatial correlation among streamflow series. For this reason, we considered Generalized Least Squares (GLS; i.e. Stedinger and Tasker, 1985; Tasker and Stedinger, 1989) and Top-kriging (TK; i.e. Skøien et al., 2006), procedures which are widely applied in several climatic and geographical contexts, but not commonly used in Italy. The preliminary cross-validated application of GLS and TK performed over a homogeneous region in Triveneto consisting of 20 nested catchments highlighted that the behaviour of the monovariate (i.e. function of the drainage area only) versions of GLS and TK applied for predicting the 100-year flood is consistent with the results reported in Archfield et al. (2013) for the South-eastern USA: TK outperforms GLS in predicting empirical estimates of flood quantiles. This result is expected as, referring to n neighbouring sites, TK implicitly accounts for some climate and geomorphological similarities between catchments, especially in regions with preponderance of nested catchments. On the other hand, we showed that the inclusion of more catchment descriptors in the analysis (multivariate versions of GLS and TK) can lead to significantly improved performances, especially for GLS. Moreover, the Monte Carlo simulation experiment performed for the same study area enabled us to address the important issue raised by Archfield et al. (2013) of understanding which technique between GLS and TK is better suited for predicting the true unknown flood quantiles in ungauged catchments when the observed flood sequences are affected by cross-correlation. The application of monovariate and multivariate versions of GLS and TK for predicting at-site flood quantiles (with return periods T equal to 10, 30, 50, 100 years) for three different cross-correlation scenarios highlighted that the presence of cross-correlation in the region introduces a masking-effect on the flood magnitude for both GLS and TK, and that this masking-effect increases with increasing degrees of cross-correlation. This behaviour

is totally expected for TK (which explicitly exploits spatial correlation in performing its estimates), whereas for GLS, which should be able to look behind cross-correlation, can be explained by the presence of a residual masking-effect with increasing levels of spatial correlation. In particular, the multivariate versions of GLS and TK show very similar performances: their performance is almost equivalent when significant catchment descriptors for describing mean annual flood are found. On the contrary, when only a univariate analysis with drainage area can be performed, the application of TK is recommended, even in the presence of high degrees of spatial correlation. These findings are valid for the simplified situation of homogeneous region with nested catchments and no effect of cross-correlation on mean annual flood. Further analyses able to overcome the simplifying hypotheses adopted are suggested for future studies.

Finally, we focused on the potential of geostatistical approaches for predicting flow-duration curves in ungauged sites across large geographical areas. Indeed, the above mentioned analyses highlighted the reliability of TK for predicting flood quantiles in ungauged sites in the study area. In general, geostatistical procedures have been shown to provide highly reliable predictions of streamflow indices over large study areas, especially if compared to regional regression models, whose accuracy is generally unsatisfactory for large study regions, which can be characterised by significant heterogeneities in terms of streamflow regimes. In this context, we focused on the use of Total Negative Deviation Top-kriging (TNDTK; i.e. Pugliese et al., 2014) for predicting flow-duration curves (FDCs) in ungauged basins in the Danube region, the largest watershed in Europe, within a research agreement with the Joint Research Centre of the European Commission (DG JRC), with the aim of generating a GIS (Geographic Information System) layer reporting the predicted streamflow regime (FDCs) for about 4000 prediction nodes within the watershed of the Danube River. Our analyses highlighted that, even though streamflow indices are significantly correlated to catchment characteristics within the study region, their prediction using multi-regression models may be not completely satisfactory. The cross-validated application of TNDTK for about 500 discharge measurement stations across the Danube river basin showed the effectiveness of TNDTK for predicting FDCs in ungauged catchments for the entire region. Moreover, although the spatial density of streamgauging network affects the estimation variance of interpolation, we proved that the regionalisation becomes more accurate when low-quality measurements are discarded. In particular, we showed that kriging variance can be used as a proxy for uncertainty of predicted FDCs. The maps of streamflow quantiles resulted from the analyses are useful for the evaluation of water resources availability at ungauged locations, and as a benchmark for the development of hydrological macro-scale models.

In conclusion, this Thesis provides an important contribute for better undestanding the added value and impacts of catchment similarity and spatial correlation on the prediction of flood quantiles and flow-duration curves in ungauged basins. In particular, the analyses performed in this study and presented herein confirm the added value of including climatic and geomorphological descriptors in the identification of homogeneous pooling-groups of sites, especially for highly heterogeneous areas. Concerning large geographical areas, which can be characterised by significant heterogeneities in terms of streamflow regimes, geostatistical procedures were shown to provide much more reliable estimates than regression methods for predicting flow-duration curves. At the same time, with regards to a small homogeneous study region, geostatistical procedures were shown to be more effective in predicting both empirical estimates of flood quantiles and true flood quantiles than regression methods accounting for spatial correlation but scaling flood quantiles with drainage area alone. The corresponding versions accounting for more significant descriptors than drainage area provide more reliable estimates and are characterised by much more similar efficiencies, yet they are affected by an increasing masking-effect on the real flood magnitude with increasing levels of cross-correlation in the region. In this context, other relevant topics could be addressed in future studies, with particular reference to the comparison of GLS and TK in heterogeneous regions with different levels of nesting between catchments.

Appendix A

Regional frequency analysis with L-moments

A.1 Moments

Let X be a random variable taking values which are real numbers, $F(x)$ its cumulative distribution function, and, if $F(x)$ is differentiable, $f(x) = \frac{d}{dx}F(x)$ its probability density function. The shape of a probability distribution is traditionally described by the moments of the distribution.

A.1.1 Moments

The moment of order 1 is the expected value of the random variable X :

$$\mu = E(X) = \int_{-\infty}^{\infty} x dF(x) = \int_{-\infty}^{\infty} x f(x) dx \quad (\text{A.1})$$

The higher moments are defined as:

$$\mu_r = E(X - \mu)^r, \quad r = 2, 3, \dots \quad (\text{A.2})$$

The moment of order 1 (typically the mean) represents the centre of location of the distribution. The dispersion of the distribution about its centre is measured by the standard deviation:

$$\sigma = \mu_2^{1/2} = \{E(X - \mu)^2\}^{1/2} \quad (\text{A.3})$$

or the variance:

$$\sigma^2 = \mu_2 = \text{var}(X) \quad (\text{A.4})$$

A.1.2 Dimensionless moments

Usually, in order to compare samples from different gauging stations, dimensionless moments are used. The coefficient of variation Cv expresses the dispersion of a distribution as a proportion to the mean:

$$\text{Cv} = \sigma/\mu \quad (\text{A.5})$$

Dimensionless higher moments $\mu_r/\mu_2^{r/2}$ are also used. In particular, the coefficient of skewness Cs (or γ) is defined as:

$$\text{Cs} = \gamma = \mu_3/\mu_2^{3/2} \quad (\text{A.6})$$

while the coefficient of kurtosis Ck (or κ) is:

$$\text{Ck} = \kappa = \mu_4/\mu_2^2 \quad (\text{A.7})$$

A.1.3 Sample moments

Moments have been defined for a probability distribution, but in practise must often be estimated from a finite sample $x_1, x_2, \dots, x_i, \dots, x_n$. Therefore, sample estimates of moments (i.e. sample moments) need to be computed.

The natural estimator of μ is the sample mean:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (\text{A.8})$$

and the higher sample moments are defined as follows:

$$m_r = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^r \quad (\text{A.9})$$

Sample moments are reasonable estimates of moments μ_r , but are not unbiased. Therefore, unbiased estimators are often used. For example the unbiased estimate of sample variance can be obtained as:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (\text{A.10})$$

Sample estimates of the coefficient of variation and the coefficient of skewness can be obtained as follows:

$$\hat{\text{Cv}} = \frac{s}{\bar{x}} \quad (\text{A.11})$$

$$\hat{\gamma} = g = \frac{\frac{n}{(n-1)(n-2)} \sum_{i=1}^n (x_i - \bar{x})^3}{s^3} \quad (\text{A.12})$$

where the sample standard deviation, $s = \sqrt{s^2}$, is an estimator of σ but is not unbiased. Moreover, the estimators of skewness g and kurtosis k can be severely biased, having algebraic bounds which depend on the sample size.

Inferences based on sample moments of skew distributions are therefore likely to be very unreliable. A more satisfactory set of measures of distributional shape is obtained from L-moments (Hosking, 1990; see also Hosking and Wallis, 1993, 1997), which are described in App. A.2.

A.2 L-moments

L-moments are an alternative system of describing the shapes of probability distributions. Historically, they arise as modifications of the probability weighted moments (PWMs) of Greenwood et al. (1979).

A.2.1 Probability weighted moments

Given the random variable X with cumulative distribution function $F(x) = P[X \leq x]$, PWMs are defined as:

$$M_{p,r,s} = E[X^p \{F(X)\}^r \{1 - F(X)\}^s] \quad (\text{A.13})$$

where p, r, s are real numbers.

Particularly useful cases of PWMs are $\alpha_r = M_{1,0,r}$ and $\beta_r = M_{1,r,0}$:

$$\alpha_r = \int_0^1 X \{1 - F(X)\}^r dF(x) \quad (\text{A.14})$$

$$\beta_r = \int_0^1 X \{F(X)\}^r dF(x) \quad (\text{A.15})$$

While conventional moments (see App. A.1) involve successively higher powers of X , PWMs involve successively higher powers of $F(X)$ or $1 - F(X)$ and may be regarded as integrals of X weighted by the polynomials $\{F(X)\}^r$ or $\{1 - F(X)\}^r$. This means that the relationships between PWMs and the parameters of distributions have a simpler

mathematical form than the relationships between conventional moments and the parameters themselves. PWMs allow to compute parameters through linear combinations, while conventional moments usually requires iterative methods. For this reason, α_r and β_r have been used as the basis of methods for estimating parameters of distributions in several studies (see e.g. Landwehr et al., 1987, and the other studies cited in Hosking and Wallis, 1997). However, they are difficult to interpret directly as measures of the scale and shape of a probability distribution.

A.2.2 L-moments

L-moments can be expressed as linear combinations of PWMs:

$$\lambda_1 = \alpha_0 = \beta_0 \tag{A.16}$$

$$\lambda_2 = \alpha_0 - 2\alpha_1 = 2\beta_1 - \beta_0 \tag{A.17}$$

$$\lambda_3 = \alpha_0 - 6\alpha_1 + 6\alpha_2 = 6\beta_2 - 6\beta_1 + \beta_0 \tag{A.18}$$

$$\lambda_4 = \alpha_0 - 12\alpha_1 + 30\alpha_2 - 20\alpha_3 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \tag{A.19}$$

and in general:

$$\lambda_{r+1} = (-1)^r \sum_{k=0}^r p_{r,k}^* \alpha_k = \sum_{k=0}^r p_{r,k}^* \beta_k \tag{A.20}$$

where

$$p_{r,k}^* = (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} = \frac{(-1)^{r-k} (r+k)!}{(k!)^2 (r-k)!} \tag{A.21}$$

L-moments have the great advantage of being interpreted directly as measures of the scale and shape of a probability distribution. In particular, λ_1 , defined as L-location ($-\infty < \lambda_1 < +\infty$), represents the expected value of the distribution (i.e. a measure of central tendency of the distribution), while λ_2 , defined as L-scale ($\lambda_2 \geq 0$), is a measure of variability of the random variable. Moreover, L-moments are less affected than PWMs by sampling variability and distortion due to exponentiation.

A.2.3 L-moment ratios

As seen for conventional moments (see App. A.1), it is convenient to define dimensionless versions of L-moments, called L-moment ratios. L-moment ratios are obtained

by dividing the higher-order L-moments by the scale measure λ_2 :

$$\tau_r = \frac{\lambda_r}{\lambda_2}, \quad r = 3, 4, \dots \quad (\text{A.22})$$

where $|\tau_r| < 1$ for $r \geq 3$.

L-moment ratios measure the shape of a distribution independently of its scale of measurement. Similarly to dimensionless conventional moments, τ_3 is named L-skewness (L-coefficient of skewness, or coefficient of L-skewness, L-Cs), and τ_4 is named L-kurtosis (L-coefficient of kurtosis, or coefficient of L-kurtosis, L-Ck).

Moreover, the L-coefficient of variation (or coefficient of L-variation) L-Cv is defined as follows:

$$\tau = \tau_2 = \frac{\lambda_2}{\lambda_1} \quad (\text{A.23})$$

whose meaning is the same of the conventional coefficient of variation Cv. For a distribution with positive values, $0 \leq \tau_2 < 1$.

The L-moments λ_1 and λ_2 , L-Cv (i.e. τ) and the L-moment ratios τ_3 and τ_4 are the most useful quantities for summarising probability distributions (see e.g. Hosking and Wallis, 1997).

A.2.4 Sample L-moments

As for conventional moments, L-moments are defined for a probability distribution, but in practise need to be estimated from a finite sample. Estimation is based on a sample of size n , arranged in ascending order. Let $x_{1:n} \leq x_{2:n} \leq \dots \leq x_{n:n}$ be the ordered sample.

An unbiased estimator of the PWM β_r (see Landwehr et al., 1987) is given by:

$$b_r = \frac{1}{n} \binom{n-1}{r}^{-1} \sum_{j=r+1}^n \binom{j-1}{r} x_{j:n} \quad (\text{A.24})$$

which can be written also as follows:

$$b_r = \frac{1}{n} \sum_{j=r+1}^n \frac{(j-1)(j-2)\dots(j-r)}{(n-1)(n-2)\dots(n-r)} x_{j:n} \quad (\text{A.25})$$

Analogously to what seen for conventional moments, the sample L-moments are defined by:

$$l_1 = b_0 \quad (\text{A.26})$$

$$l_2 = 2b_1 - b_0 \quad (\text{A.27})$$

$$l_3 = 6 b_2 - 6 b_1 + b_0 \quad (\text{A.28})$$

$$l_4 = 20 b_3 - 30 b_2 + 12 b_1 - b_0 \quad (\text{A.29})$$

and in general:

$$l_{r+1} = \sum_{k=0}^r p_{r,k}^* b_k, \quad r = 0, 1, \dots, n-1 \quad (\text{A.30})$$

The sample L-moment l_r is an unbiased estimator of λ_r .

Analogously to conventional moments, the sample L-moment ratios are defined by:

$$t_r = \frac{l_r}{l_2} \quad (\text{A.31})$$

and the sample L-Cv as follows:

$$t = t_2 = \frac{l_2}{l_1} \quad (\text{A.32})$$

t_r and t are natural estimators of τ_r and τ , respectively.

These estimators are not unbiased, but their biases are very small in moderate or large samples. This characteristic makes them the most used estimators, especially in cases where the distribution from which the sample is drawn is already identified.

A.3 Hosking and Wallis test

Once a region has been defined, it is desirable to assess its meaningfulness. This involves testing if the proposed region may be accepted as being homogeneous, and if two or more homogeneous regions are sufficiently similar that they should be combined into a single region. The hypothesis of homogeneity is that the at-site frequency distributions are the same except for a site-specific scale factor. A test of this hypothesis is naturally based on whether the data at the N sites in the region are consistent with the relation between the at-site frequency distributions. The test is most conveniently constructed as a statistical significance test of the similarity of appropriately chosen statistics computed from the at-site data (Hosking and Wallis, 1997).

In this context, Hosking and Wallis (1993) (see also Hosking and Wallis, 1997) proposed a homogeneity test based on L-moments (see Hosking, 1990; see also App. A.2 for a concise description of L-moments). The aim is to estimate the degree of heterogeneity in a group of sites and to assess whether the sites might reasonably be treated as a homogeneous region. Specifically, the heterogeneity measure compares the between-site variations in

sample L-moments for the group of sites with what would be expected for a homogenous region (Hosking and Wallis, 1997).

In particular, the Hosking and Wallis test assesses the homogeneity of a group of basins at three different levels by focusing on three measures of dispersion for different orders of the sample L-moments ratios:

1. a measure of dispersion for the L-Cv:

$$V_1 = \frac{\sum_{i=1}^N n_i (t_{2(i)} - \bar{t}_2)^2}{\sum_{i=1}^N n_i} \quad (\text{A.33})$$

2. a measure of dispersion for both the L-Cv and the L-Cs in the L-Cv – L-Cs space:

$$V_2 = \frac{\sum_{i=1}^N n_i [(t_{2(i)} - \bar{t}_2)^2 + (t_{3(i)} - \bar{t}_3)^2]^{1/2}}{\sum_{i=1}^N n_i} \quad (\text{A.34})$$

3. a measure of dispersion for both the L-Cs and the L-Ck in the L-Cs – L-Ck space:

$$V_3 = \frac{\sum_{i=1}^N n_i [(t_{3(i)} - \bar{t}_3)^2 + (t_{4(i)} - \bar{t}_4)^2]^{1/2}}{\sum_{i=1}^N n_i} \quad (\text{A.35})$$

where \bar{t}_2 , \bar{t}_3 and \bar{t}_4 are the weighted average L-moments ratios L-Cv, L-Cs and L-Ck, in this order (i.e. they are weighted proportionally to the sites' record length; e.g. $\bar{t}_2 = \sum_{i=1}^N n_i t_{2(i)} / \sum_{i=1}^N n_i$); $t_{2(i)}$, $t_{3(i)}$, $t_{4(i)}$ and n_i are the values of L-Cv, L-Cs, L-Ck and the sample size for site i ; N is the number of sites in the pooling-group.

The underlying concept of the test is to compare the sample variability of the L-moment ratios to the variation that would be expected in a homogeneous group of sites. The expected mean value μ_{V_k} and standard deviation σ_{V_k} of these dispersion measures for a homogeneous group are assessed through repeated simulations, by generating homogeneous groups of basins having the same record lengths as those of the observed data (Hosking and Wallis, 1993). The heterogeneity measures are then evaluated with the following expression:

$$H_k = \frac{V_k - \mu_{V_k}}{\sigma_{V_k}}, \quad \text{for } k = 1, 2, 3 \quad (\text{A.36})$$

Hosking and Wallis (1993) suggest that a group of sites may be regarded as:

- "acceptably homogeneous" if $H_k < 1$;
- "possibly heterogeneous" if $1 \leq H_k < 2$;

- "definitely heterogeneous" if $H_k \geq 2$.

Moreover, Hosking and Wallis (1993) observe that higher-order L-moments tend to be more homogeneous in space than the lower-order ones. Therefore, the Hosking and Wallis test presents a hierarchical feature.

A.4 Choice of a frequency distribution

In regional frequency analysis a single frequency distribution is fitted to data from several sites. In general, the chosen distribution does not need to be the distribution which gives the closest approximation to the observed data, as there is no guarantee that future values will match those of the past. For this reason, it is preferable to use a robust approach based on a distribution that will yield reasonably accurate quantile estimates even when the true at-site frequency distributions deviate from the fitted regional frequency distribution (Hosking and Wallis, 1997).

Many families of distributions might be candidates to be fitted to a regional dataset. Their suitability can be evaluated by considering their ability to reproduce features of the data which are of particular importance in a given application, i.e. upper (or lower) bound of the distribution, upper (or lower) tail of the distribution, shape of the body of the distribution, exact zero values. It can happen that several distributions fit the data adequately; in this case, the best choice is the distribution that is most robust, i.e. most capable of giving good quantile estimates even though future data values may come from a distribution somewhat different from the fitted distribution (Hosking and Wallis, 1997).

To this aim, Hosking and Wallis (1993) recommended a method for the selection of the parent distribution based on the use of L-moments (see Hosking, 1990; see also App. A.2 for a concise description of L-moments). In particular, given a homogeneous group of sites (see App. A.3 for a concise description of the Hosking and Wallis test), the aim is to choose the candidate distribution which gives the best fit to the data. Assuming an acceptably homogeneous region, the L-moment ratios of the sites are well summarised by the regional average; the scatter of the at-site L-moment ratios about the regional average represents sampling variability. For this reason, Hosking and Wallis (1993) propose a goodness-of-fit test which judges the goodness of fit by considering how well the L-Cs and L-Ck of the fitted distribution match the regional average L-Cs and L-Ck of the observed data. The use of fifth- or higher-order L-moments was found to be not necessary.

The procedure proposed by Hosking and Wallis (1993) (see also Hosking and Wallis, 1997) reads as follows:

1. assume a region consisting of N sites, with site i having record length n_i , and sample L-moment ratios $t_{2(i)}$, $t_{3(i)}$ and $t_{4(i)}$; denote by \bar{t}_2 , \bar{t}_3 and \bar{t}_4 the regional average L-Cv, L-Cs, L-Ck, weighted proportionally to the sites' record length:

$$\bar{t}_k = \frac{\sum_{i=1}^N n_i t_{k(i)}}{\sum_{i=1}^N n_i} \quad k = 2, 3, 4 \quad (\text{A.37})$$

2. assemble a set of candidate three-parameters distributions (e.g. GLO, GEV, LN3, PE3, etc.);
3. fit each distribution to the regional average L-moment ratios 1, \bar{t}_2 and \bar{t}_3 ; denote by τ_4^{DIST} the L-Ck of the fitted distribution, where DIST may refer to any of the candidate distributions;
4. fit a kappa distribution to the regional average L-moment ratios 1, \bar{t}_2 , \bar{t}_3 and \bar{t}_4 ;
5. simulate a large number N_{sim} (e.g. $N_{sim} = 500, 1000$) of realisations of a region with N sites, each having the above mentioned kappa distribution as its frequency distribution; the simulated regions are homogeneous and have no cross-correlation or serial correlation, while sites have the same record lengths as their real-world counterparts;
6. for the m th simulated region, compute the regional average L-Cs $t_3^{[m]}$ and L-Ck $t_4^{[m]}$;
7. calculate the bias of \bar{t}_4 :

$$B_4 = \frac{1}{N_{sim}} \sum_{m=1}^{N_{sim}} (t_4^{[m]} - \bar{t}_4) \quad (\text{A.38})$$

and the standard deviation of \bar{t}_4 :

$$\sigma_4 = \left[\frac{1}{N_{sim} - 1} \left\{ \sum_{m=1}^{N_{sim}} (t_4^{[m]} - \bar{t}_4)^2 - N_{sim} B_4^2 \right\} \right]^{1/2} \quad (\text{A.39})$$

8. for each candidate distribution, compute the goodness-of-fit measure Z^{DIST} , as follows:

$$Z^{\text{DIST}} = \frac{\tau_4^{\text{DIST}} - \bar{t}_4 + B_4}{\sigma_4} \quad (\text{A.40})$$

9. declare the fit to be adequate if Z^{DIST} is sufficiently close to zero; a reasonable criterion for a small value is $|Z^{\text{DIST}}| \leq 1.64$.

As stated by Hosking and Wallis (1997), the criterion $|Z^{\text{DIST}}| \leq 1.64$ is arbitrary and corresponds to acceptance of the hypothesised distribution at a confidence level of 90% when the Z statistic has approximately a standard normal distribution. As the assumptions necessary for Z to be standard normal include two that are unlikely to be exactly satisfied in practise (i.e. exactly homogeneous region and no intersite dependence), the criterion is a rough indicator of goodness of fit and is not recommended as a formal test.

For an acceptably homogeneous region, all the candidate distributions for which the goodness-of-fit measure satisfies the above mentioned criterion are flagged as "acceptable". If the corresponding growth curves are approximately equal, then each one of them is adequate and one can choose the most robust to the misspecification of the region. On the other hand, if the growth curves are not approximately equal, there is a problem of scarcity of data and models show differences which are statistically insignificant but operationally important. In this case, robustness is very important and, rather than choose a three-parameter distribution, it may be better to use the four-parameter kappa or five-parameter Wakeby distributions, which are more robust to misspecification of the frequency distribution of a homogeneous region (Hosking and Wallis, 1997). Another case is when none of the candidate distributions is accepted by the Z criterion; this can occur when the number of sites in the region or the at-site record lengths are large. If the L-moment ratios diagram shows a regional average (\bar{t}_3, \bar{t}_4) point which falls between two or more distributions having approximately equal growth curves, then there is superabundance of data: two or more models display differences that are statistically significant but operationally not important. In this case, any of the operationally equivalent distributions can be reconsidered as "acceptable". On the other hand, if the regional average point in the L-moment ratios diagram falls above or below all the considered distributions, then three-parameter distributions are not acceptable and more general kappa or Wakeby distributions should be used (Hosking and Wallis, 1997).

If the region is not acceptably homogeneous, there is no reason to suppose that a single regional distribution gives a good fit to data site by site. At the same time, referring to a single distribution can still yield much more accurate quantile estimates than fitting separate at-site distributions. In this case, kappa and Wakeby distributions are recommendable choices, considering their robustness to moderate heterogeneity in the at-site frequency distributions. Finally, when the region is heterogeneous, Hosking and Wallis (1997) recommend the use of the Wakeby distribution.

Bibliography

- Acreman, M. C. and Wiltshire, S. E. (1987). Identification of regions for regional flood frequency analysis (abstract). *Eos, Transactions American Geophysical Union*, 68(44).
- Acreman, M. C. and Wiltshire, S. E. (1989). The regions are dead: long live the regions. methods of identifying and dispensing with regions for flood frequency analysis. In Roald, L., Nordseth, K., and Hassel, K. A., editors, *FRIENDS in Hydrology*. IAHS Publication 187.
- Allamano, P., Claps, P., and Laio, F. (2009). An analytical model of the effects of catchment elevation on the flood frequency distribution. *Water Resources Research*, 45(1).
- Archfield, S. (2009). *Estimation of continuous daily streamflow at ungaged locations in southern New England*. PhD thesis, Tufts University.
- Archfield, S. A., Pugliese, A., Castellarin, A., Skøien, J. O., and Kiang, J. E. (2013). Topological and canonical kriging for design flood prediction in ungauged catchments: an improvement over a traditional regional regression approach? *Hydrology and Earth System Sciences*, 17(4):1575–1588.
- Archfield, S. A. and Vogel, R. M. (2010). Map correlation method: Selection of a reference streamgage to estimate daily streamflow at ungauged catchments. *Water Resources Research*, 46(10).
- Archfield, S. A., Vogel, R. M., Steeves, P. A., Brandt, S. L., Weiskel, P. K., and Garabedian, S. P. (2010). *The Massachusetts Sustainable-Yield Estimator: A decision-support tool to assess water availability at ungauged stream locations in Massachusetts*. US Department of the Interior, US Geological Survey.
- Bierkens, M. F. P., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., Roo, A., Döll, P., Drost, N., Famiglietti, J. S., Flörke, M., Gochis, D. J., Houser, P., Hut, R., Keune, J., Kollet, S., Maxwell, R. M., Reager, J. T., Samaniego, L., Sudicky, E., Sutanudjaja, E. H., Giesen, N., Winsemius, H., and Wood, E. F. (2015). Hyper-resolution global hydrological modelling: what is next? *Hydrological Processes*, 29(2):310–320.
- Blöschl, G. and Sivapalan, M. (1997). Process controls on regional flood frequency: coefficient of variation and basin scale. *Water Resources Research*, 33:2967–2980.

- Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H. (2013). *Runoff prediction in ungauged basins: synthesis across processes, places and scales*. Cambridge University Press.
- Bocchiola, D., De Michele, C., and Rosso, R. (2003). Review of recent advances in index flood estimation. *Journal of Hydrology*, 7(3):283–296.
- Brath, A., Castellarin, A., Franchini, M., and Galeati, G. (2001a). Estimating the index flood using indirect methods. *Hydrological Sciences Journal*, 46(3):399–418.
- Brath, A., Castellarin, A., and Montanari, A. (2001b). At-site and regional assessment of the possible presence of non-stationarity in extreme rainfall in northern Italy. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, 26(9):705–710.
- Brath, A., De Michele, C., Galeati, G., and Rosso, R. (1997). Una metodologia per l'identificazione di regioni omogenee nel regime di piena. Applicazione all'Italia nord-occidentale. *L'Acqua*, 1:17–26.
- Brath, A., De Michele, C., and Rosso, R. (1995). Metodologie di valutazione del rischio idrologico nei corsi d'acqua padani e liguri, con particolare riferimento alla stima della piena indice. *Rapporto interno, Politecnico di Milano*.
- Burn, D. H. (1989). Cluster analysis as applied to regional flood frequency. *Journal of Water Resources Planning and Management*, 115(5):567–582.
- Burn, D. H. (1990). Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resources Research*, 26(10):2257–2265.
- Calder, I. R. (1998). Water use by forests, limits and controls. *Tree Physiology*, 18(8–9):625–631.
- Castellarin, A. (2007). Probabilistic envelope curves for design flood estimation at ungauged sites. *Water Resources Research*, 43(4).
- Castellarin, A. (2014). Regional prediction of flow-duration curves using a three-dimensional kriging. *Journal of Hydrology*, 513:179–191.
- Castellarin, A., Botter, G., Hughes, D. A., Liu, S., Ouara, T. B. M. J., Parajka, J., Post, M., Sivapalan, M., Spence, C., Viglione, A., and Vogel, R. (2013). *Prediction of flow duration curves in ungauged basins*, chapter 7, pages 135–162. Cambridge University Press.
- Castellarin, A., Burn, D. H., and Brath, A. (2001). Assessing the effectiveness of hydrological similarity measures for regional flood frequency analysis. *Journal of Hydrology*, 241(3–4):270–285.
- Castellarin, A., Burn, D. H., and Brath, A. (2008). Homogeneity testing: How homogeneous do heterogeneous cross-correlated regions seem? *Journal of Hydrology*, 360(1–4):67–76.

- Castellarin, A., Camorani, G., and Brath, A. (2007). Predicting annual and long-term flow-duration curves in ungauged basins. *Advances in Water Resources*, 30(4):937–953.
- Castellarin, A., Galeati, G., Brandimarte, L., Montanari, A., and Brath, A. (2004a). Regional flow-duration curves: reliability for ungauged basins. *Advances in Water Resources*, 27(10):953–965.
- Castellarin, A., Kohnová, S., Gaal, L., Fleig, A., Salinas, J. L., Toumazis, A., Kjeldsen, T., and MacDonald, N. (2012). *Review of applied statistical methods for flood frequency analysis in Europe: WG2 of COST Action ES0901*. (NERC) Centre for Ecology and Hydrology.
- Castellarin, A., Persiano, S., Pugliese, A., Aloe, A., Skøien, J. O., and Pistocchi, A. (2018). Prediction of streamflow regimes over large geographical areas: interpolated flow–duration curves for the Danube region. *Hydrological Sciences Journal*, 63(6):845–861.
- Castellarin, A., Vogel, R. M., and Brath, A. (2004b). A stochastic index flow model of flow duration curves. *Water Resources Research*, 40(3).
- Castellarin, A., Vogel, R. M., and Matalas, N. C. (2005). Probabilistic behavior of a regional envelope curve. *Water Resources Research*, 41(6).
- Castiglioni, S., Castellarin, A., and Montanari, A. (2009). Prediction of low-flow indices in ungauged basins through physiographical space-based interpolation. *Journal of Hydrology*, 378:272 – 280.
- Castiglioni, S., Castellarin, A., Montanari, A., Skøien, J. O., Laaha, G., and Blöschl, G. (2011). Smooth regional estimation of low-flow indices: physiographical space based interpolation and top-kriging. *Hydrology and Earth System Sciences*, 15(3):715–727.
- Ceola, S. and Pugliese, A. (2014). Regional prediction of basin-scale brown trout habitat suitability. *Proceedings of the International Association of Hydrological Sciences*, 364:26–31.
- Chambers, J. M. (1992). Linear models. chapter 4. In Chambers, J. M. and Hastie, T. J., editors, *Statistical models*, Pacific Grove, CA. S. Wadsworth & Brooks/Cole.
- Chokmani, K. and Ouarda, T. B. M. J. (2004). Physiographical space-based kriging for regional flood frequency estimation at ungauged sites. *Water Resour. Res.*, 40(12):W12514–.
- Chow, V. (1964). *Handbook of Applied Hydrology*. McGraw-Hill Book Company, New York.
- Chow, V., Maidment, D. R., and Mays, L. W. (1988). *Applied Hydrology*. McGraw-Hill Book Company, New York.

- Collischonn, W., Allasia, D., Da Silva, B. C., and Tucci, C. E. M. (2007). The mgb-iph model for large-scale rainfall—runoff modelling. *Hydrological Sciences Journal*, 52(5):878–895.
- Cressie, N. A. C. (1993). *Statistics for spatial data*. Wiley series in probability and mathematical statistics: Applied probability and statistics. J. Wiley.
- Cunnane, C. (1987). Review of statistical models for flood frequency estimation. In Singh, V., editor, *Hydrologic frequency modeling*, pages 49–95.
- Cunnane, C. (1988). Methods and merits of regional flood frequency analysis. *Journal of Hydrology*, 100(1):269–290.
- Dalrymple, T. (1960). *Flood Frequency Analysis*. US Geological Survey, Water Supply Paper, 1543 A.
- de Lavenne, A., Skøien, J. O., Cudennec, C., Curie, F., and Moatar, F. (2016). Transferring measured discharge time series: Large-scale comparison of Top-kriging to geomorphology-based inverse modeling. *Water Resources Research*, 52(7):5555–5576.
- De Marsily, G. (1986). *Quantitative Hydrogeology: Groundwater Hydrology for Engineers*. Academic Press.
- de Paiva, R. C. D., Buarque, D. C., Collischonn, W., Bonnet, M.-P., Frappart, F., Calmant, S., and Bulhões Mendes, C. A. (2013). Large-scale hydrologic and hydrodynamic modeling of the Amazon River basin. *Water Resources Research*, 49(3):1226–1243.
- de Roo, A., Burek, P., Gentile, A., Udias, A., Bouraoui, F., Aloe, A., Bianchi, A., La Notte, A., Kuik, O., Elorza Tenreiro, J., Vandecasteele, I., Mubareka, S., Baranzelli, C., van der Perk, M., Lavalle, C., and Bidoglio, G. (2012). *A multi-criteria optimisation of scenarios for the protection of water resources in Europe: support to the EU blueprint to safeguard Europe’s waters*, volume 25552 of *EUR – Scientific and Technical Research*. Publications Office of the European Union, Luxembourg.
- Di Baldassarre, G., Kooy, M., Kemerink, J. S., and Brandimarte, L. (2013). Towards understanding the dynamic behaviour of floodplains as human–water systems. *Hydrology and Earth System Sciences*, 17(8):3235–3244.
- DICAM-UniBo (2015). Rapporto sintetico riguardante l’esame critico degli studi idrologici sui bacini delle Alpi Orientali e dell’Appennino Settentrionale - Secondi 10 casi significativi, Convenzione di ricerca tra Ministero delle Infrastrutture e dei Trasporti (Direzione Generale per le Dighe e le Infrastrutture Idriche ed Elettriche) e DICAM. Technical report, University of Bologna, Italy.
- Donnelly, C., Andersson, J. C. M., and Arheimer, B. (2016). Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe. *Hydrological Sciences Journal*, 61(2):255–273.

-
- Draper, N. R. and Smith, H. (1981). *Applied regression analysis*. John Wiley & Sons, Inc., New York, USA, 2nd edition.
- Eng, K., Chen, Y., and Kiang, J. (2009). *User's Guide to the Weighted-Multiple-Linear Regression Program (WREG Version 1.0)*. Techniques and methods. U.S. Geological Survey.
- Falter, D., Dung, N., Vorogushyn, S., Schröter, K., Hundecha, Y., Kreibich, H., Apel, H., Theisselmann, F., and Merz, B. (2016). Continuous, large-scale simulation model for flood risk assessments: proof-of-concept. *Journal of Flood Risk Management*, 9(1):3–21.
- Farmer, W. H. (2016). Ordinary kriging as a tool to estimate historical daily streamflow records. *Hydrology and Earth System Sciences*, 20(7):2721–2735.
- Farmer, W. H. (2017). WREG: USGS WREG v. 2.02. R package version 2.02. *GitHub repository*: <https://github.com/USGS-R/WREG>.
- Farmer, W. H., Over, T. M., and Vogel, R. M. (2015). Multiple regression and inverse moments improve the characterization of the spatial scaling behavior of daily streamflows in the Southeast United States. *Water Resources Research*, 51:1775–1796.
- Farmer, W. H. and Vogel, R. M. (2013). Performance-weighted methods for estimating monthly streamflow at ungauged sites. *Journal of Hydrology*, 477:240–250.
- Fennessey, N. and Vogel, R. M. (1990). Regional flow-duration curves for ungauged sites in massachusetts. *Journal of Water Resources Planning and Management*, 116(4):530–549.
- Fiorentino, M., Arora, K., and Singh, V. (1987a). The two-component extreme value distribution for flood frequency analysis: derivation of a new estimation method. *Stochastic Hydrology and Hydraulics*, 1:199–208.
- Fiorentino, M., Gabriele, S., Rossi, F., and Versace, P. (1987b). Hierarchical Approach for Regional Flood Frequency Analysis. In Singh, V., editor, *Regional Flood Frequency Analysis*, pages 35–49. Springer Netherlands, Dordrecht.
- Fowler, H. J. and Wilby, R. L. (2010). Detecting changes in seasonal precipitation extremes using regional climate model projections: Implications for managing fluvial flood risk. *Water Resources Research*, 46(3).
- Franchini, M. and Galeati, G. (1996). Analisi regionale dei massimi annuali delle portate al colmo per la regione Romagna-Marche. *L'Energia Elettrica*, 73(3):200–212.
- Franchini, M. and Suppo, M. (1996). Regional analysis of flow duration curves for a limestone region. *Water resources management*, 10(3):199–218.
- Fuller, W. (1914). Flood flows. *Transactions of the American Society of Civil Engineers*, 77(1):564–617.

- Ganora, D., Claps, P., Laio, F., and Viglione, A. (2009). An approach to estimate non-parametric flow duration curves in ungauged basins. *Water Resources Research*, 45(10).
- Greenwood, J. A., Landwehr, J. M., Matalas, N. C., and Wallis, J. R. (1979). Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research*, 15(5):1049–1054.
- Griffis, V. W. and Stedinger, J. R. (2007a). Log-Pearson Type 3 distribution and its application in Flood Frequency Analysis. I: Distribution characteristics. *Journal of Hydrologic Engineering*, 12(5):482–491.
- Griffis, V. W. and Stedinger, J. R. (2007b). The use of GLS regression in regional hydrologic analyses. *Journal of Hydrology*, 344(1-2):82–95.
- Griffis, V. W. and Stedinger, J. R. (2009). Log-Pearson Type 3 distribution and its application in Flood Frequency Analysis. III: sample skew and weighted skew estimators. *Journal of Hydrologic Engineering*, 14(2):121–130.
- Grimaldi, S., Kao, S.-C., Castellarin, A., Papalexiou, S.-M., Viglione, A., Laio, F., Aksoy, H., and Gedikli, A. (2011). 2.18 – Statistical Hydrology. In *Treatise on Water Science*, pages 479–517. Elsevier.
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25:1965–1978.
- Hosking, J. R. M. (1990). L-moments: analysis and estimation of distributions using linear combination of order statistics. *Journal of the Royal Statistical Society, Series B*, 52(1):105–124.
- 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. (1993). Some statistics useful in regional frequency analysis. *Water Resources Research*, 29(2):271–281.
- Hosking, J. R. M. and Wallis, J. R. (1997). *Regional frequency analysis: an approach based on L-moments*. Cambridge University Press.
- Hosking J. R. M., Wallis J. R., W. E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27(3):251–261.
- Hughes, D. A. and Smakhtin, V. (1996). Daily flow time series patching or extension: a spatial interpolation approach based on flow duration curves. *Hydrological Sciences Journal*, 41(6):851–871.
- Institute of Hydrology (1999). *Flood Estimation Handbook*. Wallingford, UK.

- Interagency Advisory Committee on Water Data (1982). *Guidelines for determining flood-flow frequency*. Bulletin 17B of the Hydrology Subcommittee. Office of Water Data Coordination: U.S. Geological Survey, Reston, Va.
- Isaaks, E. H. and Srivastava, R. M. (1990). *Applied Geostatistics*. OUP USA.
- Jakob, D., Reed, D., and Robson, A. (1999). Choosing a pooling-group. In *Flood Estimation Handbook*, volume 3. Wallingford, UK.
- Journel, A. G. and Huijbregts, C. J. (1978). *Mining geostatistics*. Academic Press, London, UK.
- Kim, D., Jung, I. W., and Chun, J. A. (2017). A comparative assessment of rainfall–runoff modelling against regional flow duration curves for ungauged catchments. *Hydrology and Earth System Sciences*, 21(11):5647–5661.
- Kite, G. W. (1975). Confidence limits for design events. *Water Resources Research*, 11(1):48–53.
- Kite, G. W. (1976). Reply to comment on confidence limits for design events. *Water Resources Research*, 12(4).
- Kjeldsen, T. R. and Jones, D. A. (2009). An exploratory analysis of error components in hydrological regression modeling. *Water Resources Research*, 45(2).
- Laaha, G., Skøien, J. O., and Blöschl, G. (2014). Spatial prediction on river networks: comparison of top-kriging with regional regression. *Hydrological Processes*, 28(2):315–324.
- Laaha, G., Skøien, J. O., Nobilis, F., and Blöschl, G. (2013). Spatial Prediction of Stream Temperatures Using Top-Kriging with an External Drift. *Environmental Modeling and Assessment*, pages 1–13.
- Laio, F., Ganora, D., Claps, P., and Galeati, G. (2011). Spatially smooth regional estimation of the flood frequency curve (with uncertainty). *Journal of Hydrology*, 408(1–2):67–77.
- Landwehr, J. M., Matalas, N. C., and Wallis, J. R. (1987). Probability-weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Water Resources Research*, 15:1055–1064.
- Le Boutillier, D. W. and Waylen, P. R. (1993a). Regional variations in flow-duration curves for rivers in British Columbia, Canada. *Physical Geography*, 14(4):359–378.
- Le Boutillier, D. W. and Waylen, P. R. (1993b). A stochastic model of flow duration curves. *Water Resources Research*, 29(10):3535–3541.
- Lettenmaier, D. P., Wallis, J. R., and Wood, E. F. (1987). Effect of regional heterogeneity on flood frequency estimation. *Water Resources Research*, 23(2):313–323.

- Madsen, H., Mikkelsen, P. S., Rosbjerg, D., and Harremoës, P. (2002). Regional estimation of rainfall intensity-duration-frequency curves using generalized least squares regression of partial duration series statistics. *Water Resources Research*, 38(11):21–1–21–11.
- Madsen, H., Pearson, C. P., and Rosbjerg, D. (1997). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 2. Regional modeling. *Water Resources Research*, 33(4):759.
- Madsen, H. and Rosbjerg, D. (1997). Generalized least squares and empirical bayes estimation in regional partial duration series index-flood modeling. *Water Resources Research*, 33(4):771–781.
- Martins, E. S. and Stedinger, J. R. (2002). Cross correlations among estimators of shape. *Water Resources Research*, 38(11):34–1–34–7.
- Matalas, N. C. and Langbein, W. B. (1962). Information content of the mean. *Journal of Geophysical Research*, 67(9):3441–3448.
- Merz, R., Blöschl, G., and Humer, G. (2008). National flood discharge mapping in Austria. *Natural Hazards*, (1):53–72.
- Montgomery, D. C., Peck, E. A., and Vining, G. G. (2001). *Introduction to linear regression analysis*. Wiley series in probability and statistics. Wiley, New York, USA, 3rd edition.
- Moore, G. W. and Heilman, J. L. (2011). Proposed principles governing how vegetation changes affect transpiration. *Ecohydrology*, 3(4):351–358.
- NERC (1975). *Flood Studies Report*. NERC (Natural Environment Research Council), London, UK.
- Ouarda, T. B., Girard, C., Cavadias, G. S., and Bobée, B. (2001). Regional flood frequency estimation with canonical correlation analysis. *Journal of Hydrology*, 254(1):157–173.
- Padi, P., Di Baldassarre, G., and Castellarin, A. (2011). Floodplain management in Africa: Large scale analysis of flood data. *Physics and Chemistry of the Earth*, 36:292–298.
- Pandey, G. and Nguyen, V.-T.-V. (1999). A comparative study of regression based methods in regional flood frequency analysis. *Journal of Hydrology*, 225(1-2):92–101.
- Parajka, J., Merz, R., Skøien, J. O., and Viglione, A. (2015). The role of station density for predicting daily runoff by top-kriging interpolation in Austria. *Journal of Hydrology and Hydromechanics*, 63.
- Pechlivanidis, I. G. and Arheimer, B. (2015). Large-scale hydrological modelling by using modified PUB recommendations: the India-HYPE case. *Hydrology and Earth System Sciences*, 19(11):4559–4579.

- Persiano, S., Castellarin, A., Salinas, J. L., Domeneghetti, A., and Brath, A. (2016). Climate, orography and scale controls on flood frequency in triveneto (Italy). In *IAHS-AISH Proceedings and Reports*, volume 373.
- Persiano, S., Domeneghetti, A., Galeati, G., Castellarin, A., and Brath, A. (2018). Influenza dei fattori climatici e topografici sul regime di frequenza delle piene nel Triveneto. *L'Acqua*, 2:61–70.
- Pugliese, A. (2016). *Tecniche innovative a base geostatistica per la stima dei deflussi idrici superficiali in bacini non strumentati*. PhD thesis, Alma Mater Studiorum, University of Bologna.
- Pugliese, A., Castellarin, A., and Brath, A. (2014). Geostatistical prediction of flow–duration curves in an index-flow framework. *Hydrology and Earth System Sciences*, 18(9):3801–3816.
- Pugliese, A., Farmer, W. H., Castellarin, A., Archfield, S. A., and Vogel, R. M. (2016). Regional flow duration curves: Geostatistical techniques versus multivariate regression. *Advances in Water Resources*, 96:11–22.
- Pugliese, A., Persiano, S., Bagli, S., Mazzoli, P., Parajka, J., Arheimer, B., Capell, R., Montanari, A., Blöschl, G., and Castellarin, A. (2018). A geostatistical data-assimilation technique for enhancing macro-scale rainfall-runoff simulations. *Hydrology and Earth System Sciences*, 22(9).
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Reed, D., Jakob, D., Robson, A., Faulkner, D., and Stewart, E. (1999). Regional frequency analysis: A new vocabulary. In Gottschalk, L., Olivry, J., Reed, D., and Rosbjerg, D., editors, *Hydrological Extremes: Understanding, Predicting, Mitigating (Proceedings, Birmingham Symposium, July 1999)*, pages 237–243. IAHS Publication 255.
- Rosbjerg, D. (2007). Regional flood frequency analysis. In Vasiliev, O. F., van Gelder, P. H. A. J. M., Plate, E. J., and Bolgov, M. V., editors, *Extreme Hydrological Events: New Concepts for Security*, volume 78, page 151–171.
- Rossi, F., Fiorentino, M., and Versace, P. (1984). Two component extreme value distribution for flood frequency analysis. *Water Resources Research*, 20:847–856.
- Salinas, J. L., Castellarin, A., Kohnová, S., and Kjeldsen, T. R. (2014). Regional parent flood frequency distributions in Europe – Part 2: Climate and scale controls. *Hydrology and Earth System Sciences*, 18(11):4391–4401.
- Sampson, C. C., Smith, A. M., Bates, P. D., Neal, J. C., Alfieri, L., and Freer, J. E. (2015). A high-resolution global flood hazard model. *Water Resources Research*, 51(9):7358–7381.

- Shu, C. and Ouarda, T. B. M. J. (2012). Improved methods for daily streamflow estimates at ungauged sites. *Water Resources Research*, 48(2).
- Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J. J., Mendiondo, E. M., O'connell, P. E., et al. (2003). Iahs decade on predictions in ungauged basins (pub), 2003–2012: Shaping an exciting future for the hydrological sciences. *Hydrological sciences journal*, 48(6):857–880.
- Skøien, J. O. (2014). *rtop: Interpolation of data with variable spatial support*. R package version 0.3-45.
- Skøien, J. O. and Blöschl, G. (2007). Spatiotemporal topological kriging of runoff time series. *Water Resources Research*, 43(9).
- Skøien, J. O., Merz, R., and Blöschl, G. (2006). Top-kriging - geostatistics on stream networks. *Hydrology and Earth System Sciences*, 10(2):277–287.
- Smakhtin, V. (2001). Low flow hydrology: a review. *Journal of hydrology*, 240(3):147–186.
- Smakhtin, V., Hughes, D., and Creuse-Naudin, E. (1997). Regionalization of daily flow characteristics in part of the Eastern Cape, South Africa. *Hydrological Sciences Journal*, 42(6):919–936.
- Smakhtin, V. Y. and Masse, B. (2000). Continuous daily hydrograph simulation using duration curves of a precipitation index. *Hydrological Processes*, 14(6):1083–1100.
- Stedinger, J. R. (1983). Estimating a regional flood frequency distribution. *Water Resources Research*, 19(2):503–510.
- Stedinger, J. R. and Lu, L. H. (1995). Appraisal of regional and index flood quantile estimators. *Stochastic Hydrology and Hydraulics*, 9(1):49–75.
- Stedinger, J. R. and Tasker, G. D. (1985). Regional Hydrologic Analysis: 1. Ordinary, Weighted, and Generalized Least Squares Compared. *Water Resources Research*, 21(9):1421–1432.
- Stedinger, J. R. and Tasker, G. D. (1986). Regional Hydrologic Analysis: 2. Model-error estimators, estimation of sigma and log-Pearson Type 3 distributions. *Water Resources Research*, 22(10):1487–1499.
- Stedinger, J. R., Vogel, R. M., and Foufoula-Georgiou, E. (1993). Frequency analysis of extreme events. In Maidment, D. R., editor, *Handbook of Hydrology*, chapter 18, pages 18.11–18.66. McGraw-Hill, New York.
- Tasker, G. D. (1980). Hydrologic regression with weighted least squares. *Water Resources Research*, 16(6):1107–1113.
- Tasker, G. D. and Stedinger, J. R. (1989). An operational GLS model for hydrologic regression. *Journal of Hydrology*, 111(1-4):361–375.

- Thomas, D. and Benson, M. (1970). Generalization of streamflow characteristics from drainage-basin characteristics. *US Geological Survey, Water Supply Paper, 1975*.
- Viglione, A. (2010). Confidence intervals for the coefficient of L-variation in hydrological applications. *Hydrology and Earth System Sciences*, 14(11):2229–2242.
- Viglione, A., Laio, F., and Claps, P. (2007). A comparison of homogeneity tests for regional frequency analysis. *Water Resources Research*, 43(3).
- Villi, V. and Bacchi, B. (2001). Padova-Brescia, Italia,.
- Vogel, R. M. and Fennessey, N. M. (1994). Flow-duration curves I: New interpretation and confidence intervals. *Journal of Water Resources Planning and Management*, 120(4):485–504.
- Vogel, R. M. and Fennessey, N. M. (1995). Flow-duration curves II: A review of applications in water resources planning. *JAWRA Journal of the American Water Resources Association*, 31(6):1029–1039.
- Volpi, E., Fiori, A., Grimaldi, S., Lombardo, F., and Koutsoyiannis, D. (2015). One hundred years of return period: Strengths and limitations. *Water Resources Research*, 51(10):8570–8585.
- Vormoor, K., Skaugen, T., Langsholt, E., Diekkrüger, B., and Skøien, J. O. (2011). Geostatistical regionalization of daily runoff forecasts in Norway. *International Journal of River Basin Management*, 9(1):3–15.
- Ward, J. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58:236–244.
- Wei, T. and Simko, V. (2016). *corrplot: visualization of a correlation matrix*. R-package version 0.77.
- Weisberg, S. (1985). *Applied linear regression*. John Wiley & Sons, Inc., New York, USA, 2nd edition.
- Westerberg, I. K., Wagener, T., Coxon, G., McMillan, H. K., Castellarin, A., Montanari, A., and Freer, J. (2016). Uncertainty in hydrological signatures for gauged and ungauged catchments. *Water Resources Research*, 52(3):1847–1865.
- Wilby, R. L., Beven, K. J., and Reynard, N. S. (2008). Climate change and fluvial flood risk in the uk: more of the same? *Hydrological Processes*, 22(14):2511–2523.
- Yaeger, M., Coopersmith, E., Ye, S., Cheng, L., Viglione, A., and Sivapalan, M. (2012). Exploring the physical controls of regional patterns of flow duration curves – part 4: A synthesis of empirical analysis, process modeling and catchment classification. *Hydrology and Earth System Sciences*, 16(11):4483–4498.

Zrinji, Z. and Burn, D. H. (1994). Flood frequency analysis for ungauged sites using a region of influence approach. *Journal of Hydrology*, 153(1).

Zrinji, Z. and Burn, D. H. (1996). Regional flood frequency with hierarchical region of influence. *Journal of Water Resources Planning and Management*, 122(4):245–252.

Acknowledgements

First of all, I would like to express my sincere gratitude to my two Supervisors, Prof. Armando Brath and Prof. Attilio Castellarin, for believing in me and giving me the opportunity to work in the world of Research, with this Thesis and many other interesting projects. In particular, Attilio really encouraged me throughout the past three years, resulting in a friendly and wise guidance. Special thanks to the "Alessi", Dr. Alessio Domeneghetti and Dr. Alessio Pugliese, for their great patience in supporting me with my thousands of questions and doubts; I spent formative and funny times in their office, which was always open for me. I would like to thank also all the other colleagues in our great Department: Alberto, Elena, Cristiana, Marco, Andrea, Emanuele, Serena, Sara, Francesca, Matteo, Umberto, Federica, Iuliia, Mattia, Lorenzo, Giada. Each one of them, in their own way, played a part in making the last three years positive and formative, in a very friendly and collaborative environment.

Many thanks are owed to Prof. Alberto Viglione and Dr. José Luis Salinas, for the warm and friendly hospitality and valuable and stimulating supervision during my stay at TU Wien, Austria. I would like to express my gratitude to the Vienna Doctoral Programme organisation and all its members: warm thanks to Bori, David, Julia, Lovrenc, Marlies, Miriam, Thomas, Tommaso and all the other DKs. That was my first experience abroad, but, thanks to them all, I felt like I was home. Special thanks are owed also to Prof. Murugesu Sivapalan and Prof. Jerry Russell Stedinger, with whom I had the great honour of sharing the desk during my time at TU Wien, always trying to catch even a small part of their knowledge and advice.

Thanks are due to all the co-authors of the scientific contributes I published during my PhD. Particular thanks to Giorgio Galeati, for the wise suggestions regarding flood frequency analysis in Triveneto, and Jon Olav Skøien, for the support regarding the R-package `rtop` he developed for Top-kriging. Many thanks also to William H. Farmer for his suggestions and advice about the R-package `WREG`.

Last but not least, I warmly thank my Family and my Friends, for being a peerless support during these years and all my life in general, unique "traveling companions" without whom the "travel" would make no sense. Meanwhile, Forza Bologna!

