ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA

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

Ciclo XXXIII

08/A1 - Idraulica, Idrologia, Costruzioni Idrauliche e Marittime
ICAR/02 - Costruzioni Idrauliche e Marittime e Idrologia

INNOVATIVE METHODOLOGIES FOR ENHANCING THE REGIONALISATION OF RAINFALL-RUNOFF MODEL PARAMETERS

Presentata da:
Mattia Neri

Supervisore:
Prof.ssa Elena Toth

Coordinatore del Dottorato:
Prof. Luca Vittuari

Esame Finale - Anno 2021

To my Family.
To my Friends.

Abstract

The present research work focuses on the regionalisation of rainfall-runoff model parameters, fundamental for the implementation of hydrological models in ungauged or scarcely gauged basins and needed to reproduce the actual sequence of river discharge in time. Regionalisation of parameters is based on the transfer of information from hydrologically similar gauged basins (where parameters can be estimated) to ungauged basins. The literature reports on a great number of methods, which are characterised by different ways to capture basin similarity and to transfer parameters. This work provides further insights on parameter regionalisation and catchment similarity through the application of innovative methodologies to support the existing knowledge.

After a preliminary application of a set of consolidated parameter regionalisation techniques on a reference study region, the first experiment develops a methodology to test the robustness of regionalisation procedures to the availability of data in the study region, here named "informative content". In particular, the effect of the density of streamflow gauging stations and the topological relationships between their corresponding drainage basins on the different regionalisation techniques is investigated. Such work provides useful information for the choice of the most appropriate method for transferring parameters to ungauged locations, based on data availability in the study region. In particular, the use of approaches taking advantage of the information from more gauged basins, selected relying of their hydrological similarity to the target, result to be preferable for regionalisation purposes in both data-poor and data-rich regions.

The focus is then moved to the value of hydrological similarity at sub-basin scale. Driven by the fact that similarity is generally defined between entire catchments, neglecting some significant differences in the within-basin rainfall-runoff transformation processes, a methodology to differentiate hydrological processes at sub-basin scale and to transfer model parameters from similar sub-basins is proposed. The analysis is based on the diversification of the parameter values, and therefore of the corresponding hydrological dynamics, across elevation, one of the main factors influencing the runoff generation processes. Results show that the differentiation of runoff dynamics across elevation benefits the simulation at gauged sites. Even if with some limitations, the proposed methodol-

ogy is able to improve simulations also at ungauged sites in one of the considered study regions.

Finally, an innovative catchment signature is proposed for improving our knowledge about hydrological similarity, meant for the delineation of hydrologically similar regions. A new methodology for identifying the dominant rainfall-runoff transformation dynamics is presented: the interaction between the entire time series of runoff generation forcings (i.e. precipitation, actual evapotranspiration and snowmelt) and runoff itself is quantified taking advantage of the concepts of the Information Theory and used to characterise and classify catchments with promising results. In particular, the approach is able to identify similar hydrological dynamics relying exclusively on the information flow from forcing metereological signals to river streamflow, and should be seen as a complementary approach to consolidated classification techniques.

This Thesis contributes to the hydrological community, on one hand by providing further insights on the practical application of parameter regionalisation techniques, and on the other hand by proposing new challenging methodologies for understanding similarity in rainfall-runoff processes.

Contents

List of Figures

List of Tables

Premise

Quantifying river flows is vital to many practical water engineering applications, such as the forecast and mitigation of flood and drought risks, the design of water supply and drainage systems, but also water quality and ecosystem studies. Naturally, the large majority of river sections of interest is not provided with gauging stations, which are extremely heterogeneous in space and sparsely distributed across the globe. For this reason, the estimate of streamflow in ungauged basins has always been of paramount importance in Hydrology. 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).

Depending on the practical application, different hydrological streamflow characteristics can be considered to describe the water regime in a river section. The hydrological behaviour of a catchment can be summarised with a great (potentially infinite) number of indexes representing different flow conditions and aspects. Scientists use to refer to such indexes with the term *signatures*. By definition, the most complete and versatile runoff signature is the *runoff hydrograph* (i.e. the graph of the water level or rate of flow of a body of water as a function of time) which is the aggregated results of the many hydrological processes which lead to the transformation of precipitation into river discharge in the watershed. All the other indexes (e.g. mean annual runoff, seasonality, flow duration curve, high and low flows) are extracted from the hydrograph by averaging, estimating probabilities or taking extremes of the runoff time series and highlight information which would be difficult to see by observing the entire hydrograph itself. Other indexes can describe some of the characteristics of the runoff sequences, as for example repeating patterns (e.g. often estimated with autocorrelation functions). However, while synthetic indexes summarise the full range of runoff temporal variability, they miss one critical piece of information in the runoff hydrograph: the *actual* sequence of runoff in time (Parajka et al., 2013a); knowing the evolution of the river flow in time is essential to understand how different individual dynamics combine to produce catchment response.

One of the options to predict the entire runoff hydrograph in ungauged (or scarcely gauged) basins is the use of rainfall-runoff models. Hydrological models are *process-based methods* that estimate the runoff hydrograph from precipitation and other climatic variables: fed with a set of meteorological inputs they are able to simulate the hydrological processes which led to the production of catchment response in time. All rainfall-runoff models are governed by a set of parameters which characterise the basin. Whatever the model adopted for practical application, preliminary calibration of all, or at least part of, model parameters is always necessary: model parameters have to be optimised in order to minimise the difference between simulated and observed discharge. Such procedure is indeed possible exclusively at gauged locations, where the historical time series of streamflow is available. At ungauged locations a possible solution to this problem is to transpose model parameters from hydrologically similar gauged catchments in the region, where parameters can be calibrated. This operation is named *parameter regionalisation*, and it includes an extremely large variety of approaches which take advantage of different concepts depending also on model complexity.

This Thesis specifically addresses the issue of hydrological similarity and parameter regionalisation for rainfall-runoff models. In the last two decades, hydrologic scientists from all around the world have focused on the determination of the more accurate regionalisation techniques for different case studies and rainfall-runoff models (see e.g. the reviews of Merz et al., 2006; He et al., 2011; Peel and Blöschl, 2011; Parajka et al., 2013b; Hrachowitz et al., 2013; Razavi and Coulibaly, 2013; Guo et al., 2021). The purpose of this Dissertation is to provide innovative analyses for supporting rainfall-runoff modelling in ungauged basins, focusing on some peculiar aspects and issues of parameter regionalisation so far not explored in detail by the literature.

The first issue which will be addressed regards the influence of the quantity and the type of hydrological data used for validating regionalisation approaches: the synthesis of existing studies presented in Parajka et al. (2013b) has shown that different groups of regionalisation approaches have similar efficiency. Still, the regionalisation performance is related to the quality and quantity of data used for the analysis. So, a very important aspect for choosing the most adequate regionalisation technique is the informative content of the study region, i.e. which gauged stations are available for inferring the hydrological behaviour at the target, ungauged section. Taking advantage of a very densely gauged set of Austrian basins, we develop here a methodology for assessing how and in which measure the informative content of a region influences the accuracy of different regionalisation techniques applied through different rainfall-runoff models. In particular, the effects of the amount of information (gauging density) and the topological relationships between

catchments (i.e. presence of nested basins) are explored. The purpose is to provide further insights for assessing the performances and selecting the parameter regionalisation approaches most suitable to a specific study region, keeping into account the impact of data availability.

The choice of the most appropriate approach includes the choice of the type of similarity measure between gauged and ungauged basins considered for estimating parameters at ungauged locations. The question of what makes two catchments hydrologically similar is of fundamental importance to the understanding of catchment hydrology. In addressing it, we are addressing the underlying question of what controls various aspects of hydrological response (Reichl et al., 2009). Reviewing the existing studies applied in different parts of the world yields to a collection of different catchment attributes that have been used for representing similarity (e.g. Beven and Kirkby, 1979; Post and Jakeman, 1999; Peel et al., 2000; Chiew and Siriwardena, 2005; Hundecha and Bárdossy, 2004a; Laaha and Blöschl, 2006; Lowe and Nathan, 2006; Young, 2006). In addition, when applying rainfall-runoff models in ungauged basins, similarity is often defined with average catchment characteristics which are supposed to identify behaviour of the whole catchment, thus likely missing in representing some significant processes that have a strong spatial variability. Within-catchment variability of rainfall-runoff dynamics is influenced by multiple factors, and one of the strongest impact is due to the elevation. Elevation in fact, ruling the temperature gradient, is able to control many processes and, of course, to influence also on their similarity. A possible solution for taking into account of within-catchment variability in the rainfall-runoff simulation is the use of *semi-distributed* model structures, which divide the catchment into sub-basins differentiating the hydrological transformations. This is the second issue we address in this Dissertation: taking advantage of a well-know and simple model structure we implement a semi-distributed parameterisation framework based on an elevation zone structure, focusing on understanding if and how similarity of rainfall-runoff generation processes change with elevation. The purpose is to assess if we can gain useful information by optimising similarity measures at different altitudes and at sub-basin scale, and if such knowledge can improve simulation both in gauged and ungauged catchments.

Finally, the third analysis which will be presented in the Thesis is not directly related to parameter regionalisation, but it is framed within a topic which is complementary to the transfer of hydrological information: basin classification based on hydrological similarity, that is considered a necessary step for most of regionalisation approaches (Rosbjerg et al., 2013). Preliminary grouping of watersheds in clusters that are considered homogeneous in terms of dominant hydrological processes and rainfall-runoff dynamics may

be able to enhance the choice of rainfall runoff model structure, as well as the transfer of model parameters between catchments. In the last years, literature about catchment classification has explored the use of similarity between climatic indexes and signals, similarity of the geo-morphological basin characteristics and similarity in the runoff responses and signatures. No studies have so far implemented a similarity measure for catchment classification which quantifies the *interaction* between the time sequence of meteorological forcings and catchment response. In fact, such measure may capture the predominant role of some of the hydrological processes in rainfall-runoff transformation. During the years of my PhD, I had the opportunity to spend seven months at the Water Resources and Hydrologic Modeling Lab of McMaster University (Hamilton, Canada), led by Prof. Paulin Coulibaly; taking advantage of the expertise of the research team in the use of Information Theory (Shannon, 1948) in hydrology, I had the chance to explore the potential of Entropy measures (in particular of the quantity called Transfer Entropy, Schreiber, 2000) for classifying catchments based on the measure of the interaction between forcing data and catchment discharge time-series. Of course, this is a gauged-base signatures and cannot directly be applied to ungauged catchments, i.e. for parameter regionalisation based on signatures similarity. However, understanding similarity in rainfall-runoff processes and identifying dominant hydrological dynamics may be useful for enhancing parameter transfer and for identifying hydrologically similar regions.

The Thesis is structured as follows. Chapter 1 presents a state of the art of parameter regionalisation methods and the development of the most common techniques in the last decades, classifying them and describing their main features, finally deepening on the addressed research questions. Chapter 2 gives an overview of the study regions and the datasets used in the experiments of the Dissertation. Chapter 3 describes the methods: implemented rainfall-runoff models, calibration procedures and regionalisation methods applied in the next chapters are presented. In Chapter 4, the rainfall-runoff models are calibrated across all the basins of the main study region, and the proposed parameter regionalisation methods are tested and compared. Chapter 5 presents the research work performed for assessing the impact of the informative content in the region when regionalising model parameters. Chapter 6 reports the semi-distributed calibration and regionalisation framework developed with the purpose to understand how runoff generation dynamics change with elevation and to test the potential benefits of sub-basin similarity measures. Finally, Chapter 7 shifts the attention on catchment classification, in particular exploring the potential of an Entropy measure for enhancing the grouping of similar hydrological dynamics.

Chapter 1

Rainfall-runoff modelling in ungauged basins

1.1 Calibration-dependent nature of conceptual models

1.1.1 Rainfall-runoff models

Rainfall-runoff models are instruments able to simulate the hydrological dynamics occurring in a catchment and producing river discharge: they can provide streamflow at the basin outlet as function of a set of meteorological input time-series and catchment characteristics. Meteorological inputs are also called *model forcing* and are calculated from observed data, generally by interpolation and averaging procedures. Hydrological transformation and internal model state variables are governed by model equations, constrained by a set of model parameters which characterise the basin. The output of the simulation is the basin runoff, the result of the hydrological transformation, which normally has the same temporal resolution of model inputs. Rainfall-runoff models differ between each others by both structure and application purpose. A plethora of rainfall-runoff models is available in the literature but they can be generally classified in *physically-based* and *conceptual* models.

Physically-based models describe distributed mechanics of hydrological processes and their parameters are (or supposed to be) reflected in field measurements (Beven, 1989). Model equations respect the principles of conservation of mass, energy and momentum and internal model states simulate the actual storage and path of water and energy inside the catchment. Fully physically-based models are appropriate for studying the details of the hydrological processes but their implementation requires a large amount of information,

not available in most of basins. Such models are not considered in this Dissertation.

Conceptual models instead make assumptions about the flow processes and represent the watershed system with a number of storages connected by fluxes. Their spatial resolution can be either lumped (the catchment is a single entity) or spatially distributed (the catchment is divided into single elementary units). They are governed by empirical equations which do not comply with physical principles (except for mass balance) and their parameters are not necessarily linked to measurable basin physical characteristics. Conceptual models provide "simplified representations of key hydrological processes using a perceived system" (Dawson and Wilby, 2001) and are more frequently implemented for practical application. In fact, they require minimum amount of data, they are generally quite easy to implement and the computational cost of their simulations is limited. On the other hand and for the same reasons, the estimate of their model parameters plays a key role in the model set-up: given their empirical nature, the estimate of model parameter cannot generally be supported by field measurements. Various types of models have been adopted in different part of the world (e.g. Kokkonen et al., 2003; Littlewood et al., 2003; Post, 2009) and for a suite of some popular conceptual models, the reader can refer to Lee et al. (2005).

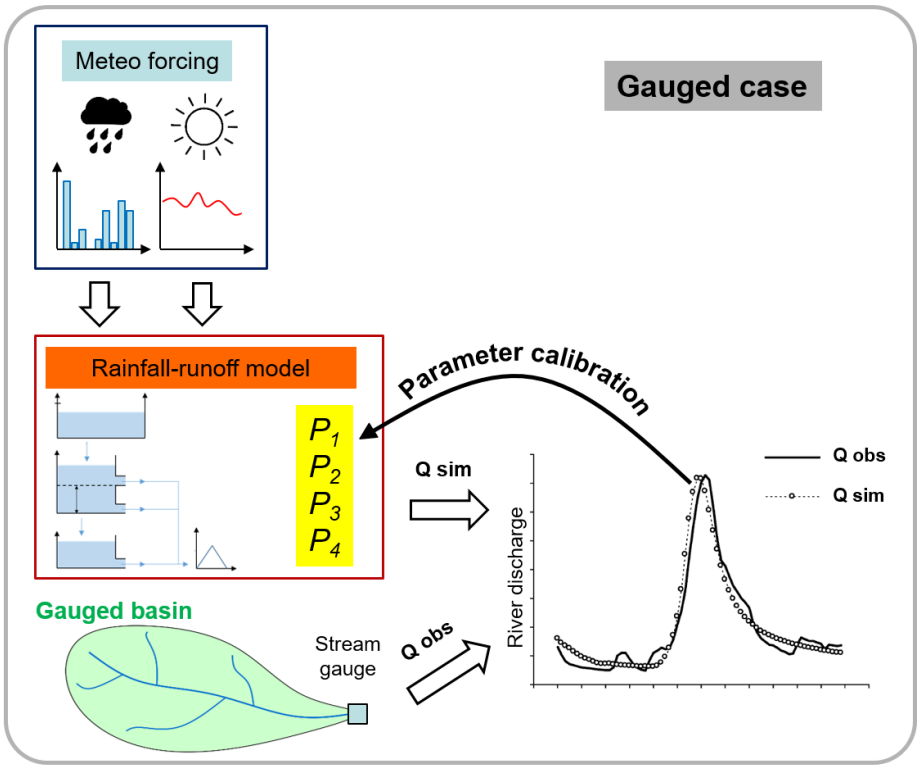

Figure 1.1: Set-up of a conceptual rainfall-runoff model in case of a gauged basin: model parameters ($P_1 \ldots P_4$) are calibrated minimising the difference between simulated (Q sim) and observed (Q obs) discharges.

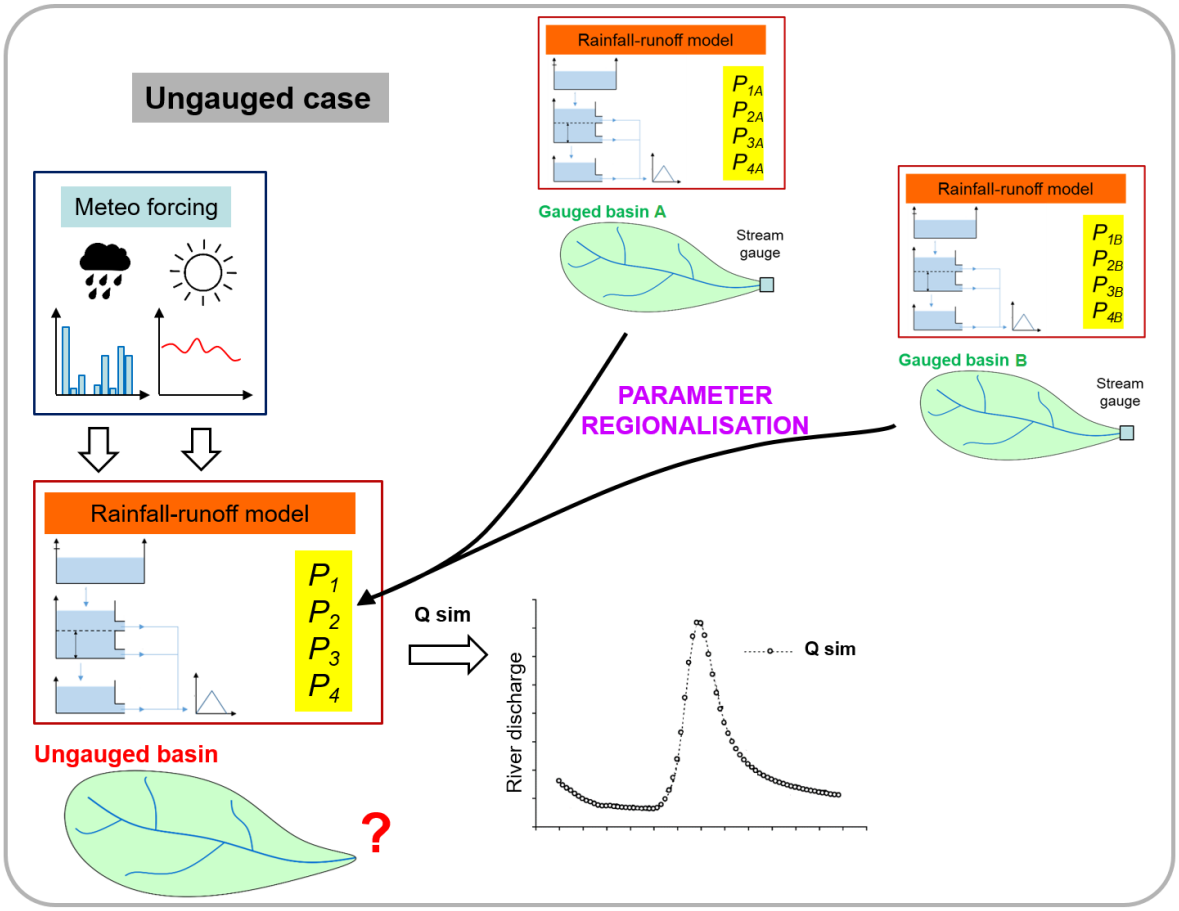

Figure 1.2: Set-up of a conceptual rainfall-runoff model in case of an ungauged basin: model parameters ($P_1 \ldots P_4$) are estimated from hydrologically similar gauged basins (A and B) through parameter regionalisation.

1.1.2 Model parametrisation in gauged case: calibration

If historical observations of river discharges are available at the basin outlet (i.e. in case of a gauged catchment), parameter values can be manually or automatically optimised by minimising the differences between model output and observations (Fig. 1.1). Such procedure is named *model calibration*.

1.1.3 Model parametrisation in ungauged case: regionalisation

The calibration-dependent nature of conceptual models makes the estimates of model parameters particularly trivial in ungauged basins. In this case (Fig. 1.2), observations are not available and parameters cannot be calibrated: a regionalisation of model parameters is needed, i.e. they are estimated based on the values of the parameters calibrated for hydrologically similar catchments.

1.2 Hydrological similarity

In a general way, catchments can be considered hydrologically similar if they filter climate variability in a similar fashion, as expressed by their (scaled) hydrological response (Wagener et al., 2013). Hydrological similarity in terms of direct similarity of rainfall-runoff processes is very difficult to capture, because most of the processes are difficult to observe. The most common approach to understand similarity is learning from similarities in climate and catchment characteristics, which are synthesised with indexes. Such approach makes the assumption that basins with similar characteristics show also hydrologically similar behaviour and similar processes (Burn and Boorman, 1992; Oudin et al., 2010).

1.2.1 Attributes to define similarity

Climate is one of the primary sources of similarity (e.g. Köppen, 1936; Budyko, 1974). For instance, a typical climatic index which let us understand the availability of water and energy in the catchment is the aridity index, defined as the ratio of annual potential evapotranspiration and annual precipitation. Moreover, climate influences most of other catchment characteristics as vegetation which also have a central role in rainfall interception and water propagation through the catchment, even at at finer temporal scales. Temperature have a central role both in soil moisture and snow processes, sometimes dominating the runoff events at various scales. Anyway, climate alone is often not enough to understand similarity of hydrological processes (Mcdonnell and Woods, 2004): the representation of catchment functioning in terms of partitioning, transmission, storage and release of water can be improved by catchment physical and morphological characteristics. For instance, infiltration dynamics can be related to soil properties, vegetation indexes can help understanding the role of evapotranspiration at large time scales; topography (slope, aspect and so on) and land cover may influence water propagation; geology and soil depth are linked, between others, to catchment storage; also drainage area is an indicator of catchment storage, in addition to runoff regulation capacity and concentration time. These are just some of the most common features which can be considered. More complex indexes can of course be taken into account: for example, stream network density (stream length per area) encloses the co-evolution of landscape, soil, vegetation, climate and geology in a particular region (Abrahams, 1984; Wang and Wu, 2013).

As already said, the above mentioned features are synthesised in numerical indexes (named attributes or descriptors), calculated by measuring and averaging spatial distributed characteristics.

1.2.2 Similarity measures

Once defined which attributes characterise the hydrological processes, similarity between different catchments has to be measured based on the obtained synthetic indexes. This section presents the most common approaches to express such measures.

Spatial proximity

We mentioned that similarity can be reflected in a number of catchment characteristics: if one assumes that such features vary smoothly in space, the simplest feature for similarity can be the geographical location of the catchments, which embraces all the other characteristics. In this case, the similarity between two basins B_1 and B_2 can be expressed by the Euclidean (geographical) distance between their coordinates X_{geo} and Y_{geo}:

$$d_{geo}(B_1, B_2) = \sqrt{[X_{geo}(B_1) - X_{geo}(B_2)]^2 + [Y_{geo}(B_1) - Y_{geo}(B_2)]^2} \tag{1.1}$$

In the following sections, it will be seen that spatial proximity often demonstrates to be effective for runoff prediction and it is implemented in a great number of studies, depending on the study region.

Geo-morphoclimatic distance

Spatial proximity alone is often not enough to properly represent similarity and a set of catchment attributes must be implemented. In this case, a measure of similarity Φ between two basins B_1 and B_2 can be thought as the sum of the "distances" $d_k(B_1, B_2)$ of their attributes X_k, standardised in general by their ranges or maximums:

$$\Phi(B_1, B_2) = \sum_{k=1}^{m} \frac{d_k(B_1, B_2)}{\max(d_k)} \tag{1.2}$$

where:

$$d_k(B_1, B_2) = |X_k(B_1) - X_k(B_2)| \tag{1.3}$$

Eq. 1.3 is valid for attributes described by a single value (as for instance mean catchment elevation or drainage area). Some descriptors instead, are identified by more than a single value: class-based attributes like land cover, geology and soil type typically indicate the portion of the catchment area associated to each of their classes. Therefore, each land cover/geology/soil type class is associated to a numerical value, i.e. the percentage of the catchment covered by each class: the attributes are thus identified by vectors $\overline{X}_k$, formed by all such percentage values instead of single values; in this case the difference between

the vectors can be calculated as an Euclidean distance (Eq. 1.4).

$$d_k(B_1, B_2) = \sqrt{\sum_c [X_{k,c}(B_1) - X_{k,c}(B_2)]^2} \tag{1.4}$$

The choice of the attributes X_k is strongly related to the study area and of course strongly depends on the hydrological process of interest. Generally, hydrologists refer to the "distance" Φ between two basins basins with the term *dissimilarity*.

Eq. 1.2, based on the distance in the attribute space, is one of the most straightforward way to express similarity. It gives the same relevance to all the attributes included in the index, i.e. assigning them equal weights; this is what done by a number of studies, in absence of previous knowledge about attributes relative importance (e.g. Parajka et al., 2005; Kay et al., 2007; Oudin et al., 2008; Zhang and Chiew, 2009). Alternative, a weighting component x_k can be included in the expression of dissimilarity (Eq. 1.5); such weights can be set based on expert judgement (e.g. McIntyre et al., 2005) or even optimised (e.g. Reichl et al., 2009; Wallner et al., 2013; Pagliero et al., 2019).

$$\Phi_x(B_1, B_2) = \sum_{k=1}^{m} x_k \frac{d_k(B_1, B_2)}{\max(d_k)} \tag{1.5}$$

Similarity measures in regionalisation

Similarity measures are used in the literature in two (complementary) fashions:

- the first is to use hydrological similarity to group catchments into large regions (sometimes not geographically contiguous) or to identify a "hydrologically consistent" region, then making assumptions for regionalisation purposes. This approach is sometimes used as a preliminary step in a class of methods which will be introduced in the next sections, called *regression-based* methods; and, if the dataset includes basins that are known to be extremely different, it may also the first step to identify large, more homogeneous regions inside which limiting the search for the smaller set of the closest donors (see next point);

- the second way to use similarity measures is for identifying, inside the study region, a few catchments "closest" to the target (ungauged) basin of interest and for transferring their hydrological information (in this case rainfall-runoff model parameters). Such catchments are called *donors* and this approach is the fundamental of the second class of techniques called *distance-based* methods. In order to select donors, a distance-threshold can be fixed and all the catchments within the limit are

considered as suitable donors, or alternatively a fixed number of donors, the closest one, can be set.

So far we referred to similarity as the distance in the attribute space between whole catchments considered as a single, undivided entity. However, nothing prevents one to consider similarity between basin portions as sub-basin, elevation zones or elementary Hydrological Response Units (HRUs).

All the above mentioned characteristics, being based on either climatic or geological, climatic, soil or morphological attributes, are measurable both at gauged and ungauged locations. For the purpose of identifying homogeneous regions inside an extremely diversified available dataset, similarity in runoff signatures, that is in metrics derived from streamflow measures, can be also used to improve classifications. Such classification is feasible only at gauged locations, but can be used to improve our knowledge about similarity in rainfall-runoff processes and eventually, in a regionalisation perspective, to use the obtained homogeneous group to train a classification framework based on climate and catchment characteristics, also available at ungauged locations, as done, for example, by Toth (2013) and Jehn et al. (2020).

1.3 Parameter regionalisation methods

Once introduced the concept of hydrological similarity, this section describes and classifies the major techniques for regionalising rainfall-runoff model parameters available in the literature, providing the theoretical concepts they are based on. Such state of the art will focus in particular on the methods used for the regionalisation of conceptual models. It was already introduced how the estimation of model parameters in ungauged locations is based on the transposition of information from gauged catchments. On the most general level, as already introduced in the previous section, regionalisation approaches for model parameterisation can be classified into two wide categories (He et al., 2011):

- *regression-based* methods;
- *distance-based* methods.

In the first case the type of information transferred is the relationship linking parameter values to catchment features, while in the second the parameters themselves are transferred.

1.3.1 Regression-based methods

Regression-based approaches are among the first methods developed for regionalisation of hydrological model parameters (first notable application dates back to Nash, 1960). They assume that each single model parameter can be described independently of each other by a set of basin characteristics (which can be climatic, geographical or physical basin attributes). In order to do that, they try to define empirical relationships which link parameters to catchment descriptors: commonly, multiple linear regressions are used and some or all of the catchments characteristics are transformed (e.g. logarithmically). A preliminary identification of homogeonous group of catchments for estimating such relationships, as highlighted in Sec. 1.2, can be implemented based on a similarity measure.

Regression-based methods, in turn, can be divided into two groups: *two-steps approaches* and *single-step approaches* (also called *regional calibration*).

Two-steps approaches

As the name suggests, this methods are based on two phases: (1) the calibration of model parameters at gauged locations maximising the goodness-of-fit statistics between simulated and observed discharges, and then (2) the implementation of multiple-regression models, solved typically by least square solutions, to relate each parameter to catchment characteristics. The obtained relationships are assumed to be valid regionally and used to estimate parameter values at ungauged locations. Regression models can be trained using all the available catchments in the dataset or a local group of catchments considered to be homogeneous to the target one.

A great variety of studies following this approach have been developed in the years; the reader can find some of the most notable examples in Servat and Dezetter (1993), Tung et al. (1997), Abdulla and Lettenmaier (1997), Kull and Feldman (1998), Sefton and Howarth (1998), Post and Jakeman (1996), Seibert (1999), Peel et al. (2000) and Kokkonen et al. (2003), Mwakalila (2003), Xu (2003), Merz and Blöschl (2004), Wagener and Wheater (2006), Heuvelmans et al. (2006), Parajka et al. (2005), Young (2006), Deckers et al. (2010), Samuel et al. (2011), Bao et al. (2012).

The method is apparently simple and the estimate of model parameter in ungauged catchments is straightforward. But, literature shows that resulting relationships between catchment properties and parameters are generally weak and the accuracy of the simulation in ungagued basin is not always good. Some studies (e.g. Hirsch et al., 1993; Blöschl, 2006) tried to reduce instability of parameter relationships using linear combinations of highly correlated catchment characteristics, but the low reliability of regression models is mainly due to the "problem" of *equifinality* (see e.g. Beven and Freer, 2001) of differ-

ent parameter sets: in conceptual models there is always more than one optimal set of parameters leading to the same simulation performances. In addition, regression-based methods regionalise parameter independently between each other despite their high interaction. On this themes, Bárdossy (2007) demonstrated that non-linear dependencies of parameters and non-uniqueness of parameter sets makes the fit of a regional transfer function unsuitable.

A strategy to enhance the parameter identifiability and to constrain regression is the so-called *sequential regression* (e.g. Hogue et al., 2000; Wagener and Wheater, 2006; Calver et al., 2005). Instead of calibrating all model parameters together, calibration is performed for one parameter at a time (remaining parameters have to be set arbitrarily), starting from the most identifiable and so on. Each parameter is calibrated against a different aspect of hydrological response, supposed to be more representative of the parameter itself. At each step, a regression model is fitted and the considered parameter is estimated at the ungauged location. In general, this type of regression brings moderate benefits to parameter identifiability and regionalisation performances (e.g. Lamb and Kay, 2004; Calver et al., 2005) at the expense of method complexity.

Single-step approaches: regional calibration

Scientist tried to overcome the above cited limitation of regression approaches by implementing a variant of the classical regression method in which calibration and regression are performed concurrently: instead of calibrating directly rainfall-runoff parameters, regression relationships are previously established across the study region and the regression coefficients are calibrated. In this case, the calibration is performed simultaneously for all the stream gauges in the region. The objective is to exploit the spatial information contained in catchment characteristics to find more reliable model parameters. One of the first studies on this line was performed by Fernandez et al. (2000), who tried to optimise both regression coefficients and model performances for a monthly water balance model in the south-east of United States, obtaining more solid relationships between parameters and catchments characteristics, but no performance improvement. Similar outcomes were found by, e.g. Szolgay et al. (2003), who calibrated uniform parameters on pre-set clusters. Slight better results were found when only model performance are optimised (and not regression relationships) as, for instance, by Parajka et al. (2007) who calibrated together 320 Austrian catchments, or by Hundecha and Bárdossy (2004b) who adopted such one-step method for almost 100 German catchments.

More recent applications have focused on the regional optimisation of the coefficients of transfer equations linking model parameters to landscape and climate predictors for

enhancing the knowledge of hydrological variability in space: they include Bastola et al. (2008) who calibrated TOPMODEL for 26 catchments in the UK, Rakovec et al. (2016) who calibrated the mHM model for 400 catchments across Europe, Mizukami et al. (2017) who calibrated the VIC model for 531 catchments in the conterminous United States and Beck et al. (2020) who calibrated transfer functions for a grid version of HBV at global scale. Merz et al. (2020) developed a new approach for the estimation of spatially consistent distributed parameters in Germany based on transfer equations, which does not require a priori assumption on the functional relationships between parameters and catchment characteristics. Interestingly, such novel methodology tries to enhance the regional consistence of model parameters by inferring regionalisation rules which consider several calibrated parameter sets (with equally good performance due to parameter equifinality) for each gauged catchment.

1.3.2 Distance-based methods

Distance-based methods rely on the identification of one or more hydrologically similar (i.e. hydrologically "close") gauged catchments, called *donors*, and on the transfer of their calibrated parameters to the target (ungauged) catchment. Methods belonging to this class differ for two aspects: (1) the definition of hydrological similarity and (2) the procedure for transferring parameters from donors to ungauged basin. The definition of similarity, already treated in Sec. 1.2, is used to select donors and is the foundation of all distance based methods.

Transfer methods

For what concerns the transfer of parameters, in the recent years a great variety of approaches have been developed. Here we will refer to different classes of the techniques discerning between:

- *single-donor* approaches which adopt the entire set of model parameters from the selected "best candidate" donor, maintaining the correlation strucuture of the parameter set;
- *multi-donor* approaches which detect a number of donor basins (either fixing a number of best donors or a maximum dissimilarity threshold). Within such class, methods are divided according to the way in which the information from more donors is exploited:
  - *parameter-averaging* approaches derive each target parameter independently,

as a function (generally a weighted average) of the calibrated parameters of the donors;

– *output-averaging* approaches transfer the entire set of model parameters from each one of the donor catchments, thus maintaining correlation among parameters. The model is run multiple times with the different parameter sets and the simulations are averaged.

Within multi-donor classes, parameter or outputs can be either averaged equally or donors can be associated to weights based on their similarity to the target. Next sections will separate the literature implementing techniques which use the only spatial proximity (i.e. geographical location) as similarity measure from the studies which represent similarity with climatic and geo-morphologic basin features.

State-of-the-art of methods based on spatial proximity

Assuming that hydrological characteristics and hydrological response vary smoothly in space, the most simple measure of similarity is the geographical distance with the candidates donor catchments. Given its simplicity it is adopted by several studies in the literature and deserves a separate section. If we use spatial proximity as measure of similarity, Euclidean (geographical) distances between ungauged basin U and donor basins D_i are the dissimilarity indexes; thus, the attributes are the coordinates of basin outlets or centroids X_geo and Y_geo (Eq. 1.1).

The most straightforward method for parameter regionalisation is probably the nearest neighbour approach in which spatial proximity is applied in a single-donor fashion: the entire set of model parameters of the nearest donor is assumed to hold also for the target catchment. It was applied with satisfactory performances by a number of studies: for instance, Parajka et al. (2005) successfully validated the method regionalising a semi-distributed version of the HBV model across a large set of Austrian basins, Li et al. (2009) applied it successfully for the regionalisation of the Xinanjiang model in south-east Australia, Vandewiele and Elias (1995) found that 44% of 75 basins in Belgium are well modelled by nearest available parameter set, and Randrianasolo et al. (2011) used the nearest donor catchment to enhance hydrological forecasting at ungauged locations in France.

When considering more donors, a simple approach, within the parameter-averaging class, is to assume that donors within a certain radius from the target are equally good and that each parameter can be estimated with their average calibrated values (e.g Parajka et al., 2005; Cislaghi et al., 2019); however, results are generally outperformed by all the other distance-based methods. On this line, more sophisticated approaches interpolate

parameter values based on the distance from the target, thus assigning a weight to each donor: simple inverse-distance average or more complex geostatistical techniques based on *kriging*: Vandewiele and Elias (1995), Merz and Blöschl (2004), Parajka et al. (2005), Viviroli et al. (2009), Swain and Patra (2017) obtained very good performances using ordinary kriging (i.e. interpolating each parameter based on their spatial correlation), but when compared to the accuracy of a single-donor approach using neighbouring catchment parameter sets, they are often similar (see e.g. Parajka et al., 2005). All parameter-averaging approaches suffer of part of the problems affecting regression based methods due to the fact that they do not maintain the correlation structure of the parameter set.

In order to overcome problems due to parameter inter-correlation, the combination of output-averaging and spatial proximity is one of the best solutions, adopted by a number of authors: for instance, McIntyre et al. (2005) were among the first to test this type of approach on a large set of British catchments and improved significantly the poor performances of the single-donor approach; Oudin et al. (2008) obtained the best regionalisation perfomances by averaging the simulation of the first four or seven nearest donor basins, respectively for the TOPMO and GR4J models, across a very dense set of more than a thousand French catchments; Li and Zhang (2017) implemented output-averaging both in lumped and grid-based fashions for the SIMHYD and Xinanjiang rainfall-runoff models across more than 600 Australian catchments; Arsenault et al. (2019) discovered that a donor-based spatial-proximity method was to be preferred for the regionalisation of the GR4J, HMETS and MOHYSE hydrological models in semi-arid and humid ungauged catchments in Mexico.

Depending on the study region, the assumption that spatial proximity is a good proxy for similarity of the hydrological response is not always valid. Climatic conditions of course influence similarity: among others, Patil and Stieglitz (2012) and Li and Zhang (2017) demonstrated that nearby gauged catchments are less useful for prediction in ungauged basins in arid regions.

State-of-the-art of methods based on geo-morphoclimatic distance

The main alternative to the simple spatial proximity is the use of climatic and physiographic catchment characteristics to define similarity. Sec. 1.2 introduced that the rational is that basin properties can predetermine hydrological behaviour (Burn and Boorman, 1992; Oudin et al., 2010) and that both climatic and geo-morphological catchment structural characteristics are essential for representing hydrological response (Wagener et al., 2007), which have to be somehow synthetised in indexes and used in the expression of similarity measure.

As already stressed, the hardest challenge in these approaches is the choice of the attributes to include in the definition of similarity (Eq. 1.2) or to set their weights (Eq. 1.5). Such choices strongly depend on the region and an accurate selection of the most suitable attributes is always needed. However, looking at studies conducted all around the globe, some attributes are demonstrated to be frequently needed, as elevation (see e.g. Kokkonen et al., 2003; Parajka et al., 2005; Zhang and Chiew, 2009), annual precipitation or aridity index (see e.g. McIntyre et al., 2005; Parajka et al., 2005) and catchment area (see e.g. McIntyre et al., 2005; Zhang and Chiew, 2009). Oudin et al. (2010) pointed out that such attributes should include the description of geological and lithological characteristics, which are often the cause of misrepresentation of catchment similarity.

Also with the geo-morphological and climatic similarity measures, several approaches to transfer parameters were tested. Similarly to what reported for spatial proximity, most straightforward approach can be to optimise similarity measures for identifying a single-donor from which transferring the entire parameter set (e.g. Bárdossy, 2007; Parajka et al., 2005; Zhang and Chiew, 2009; Viviroli et al., 2009);

Alternatively, parameter can be averaged individually from more than one single donor: some studies, as for example the ones of Kay et al. (2007), Samuel et al. (2011) and Cislaghi et al. (2019), identified pooling groups of donor catchments with attribute distances below a maximum threshold of dissimilarity and take the (eventually weighted) average of model parameters.

Finally, others tried to use the added value of multi-donor approaches, but without loosing parameter correlation, i.e. implementing output-averaging for a pool of selected "nearest" catchments in the space of the characteristics (e.g. McIntyre et al., 2005; Oudin et al., 2008; Reichl et al., 2009; Viney et al., 2009; Zelelew and Alfredsen, 2014; Beck et al., 2016; Li and Zhang, 2017). Again, in an output-averaging framework, donors can either contribute equally (e.g. Oudin et al., 2008) or be weighted based on their dissimilarity to the target (e.g. Reichl et al., 2009).

Suitability of donors

Distances in the geographical and/or attributes space between ungauged and gauged catchments are indeed the guiding ratio for the choice of the most suitable donors. At the same time, an additional aspect which must be taken into account is the suitability of a gauged catchment to serve as donor: keeping in mind that all the distance-based approaches are based on the unavoidable assumption that parameters reflect the hydrological behaviour of the catchment, the reliability of the calibrated model is a necessary condition for a candidate donor: if a model does not capture the correct processes of

rainfall-runoff generation over a catchment, there is no point of transferring the probably flawed model parameters for such catchment to the ungauged one (He et al., 2011). For this reason, inadequate model performance at-site (i.e. when calibrated against observations) of gauged catchments cannot be ignored: unsuitable basins can be excluded fixing a minimum performance threshold (e.g. Boldetti et al., 2010; Beck et al., 2016), or penalising unbehavioural modelling as done for instance by McIntyre et al. (2005), Reichl et al. (2009) or Zelelew and Alfredsen (2014): they applied an output-averaging approach where donor contributes are weighted based on their similarity to the target and, in order to give less importance (lower weight) to less suitable donors, they defined similarity measured as a the product of two indexes: "prior likelihood", function of at-site accuracy, and "posterior likelihood", based on the similarity between donor and target in term of attributes.

1.3.3 Overview of previous studies

The reliability of different parameter regionalisation approaches depends on a great number of factors linked to the characteristics of both study region and available data. Each technique has its advantages and its drawbacks and, in fact, the outcomes of the above cited studies, which compare different methods applied all around the world, not always match. However, certain tendencies in the results allow hydrologists to draw some general conclusions we summarise here (see e.g. the reviews of He et al., 2011; Parajka et al., 2013b; Razavi and Coulibaly, 2013; Guo et al., 2021).

The limitations of all the regression-based methods in predictive accuracy is due to the concurrent effect of (1) the equifinality of different parameter sets which leads to weak relationships between parameters and basin attributes (e.g. Beven and Freer, 2001) and (2) the strong inter-correlation between parameters of the same hydrological model that is neglected in such approaches (e.g Bárdossy, 2007). For such reason, regression-based methods have been demonstrated to be generally outperformed by distance-based methods by a number of recent studies conducted in different part of the world (e.g. Kokkonen et al., 2003; Merz and Blöschl, 2004; Parajka et al., 2005; Oudin et al., 2008; Reichl et al., 2009; Bao et al., 2012; Steinschneider et al., 2015; Yang et al., 2018; Arsenault et al., 2019; Cislaghi et al., 2019). Of course, such outcomes cannot be completely generalised, there are exceptions: for instance, Samuel et al. (2011) found that predictive performance of regressions were better than a simpler parameter-averaging method; Kay et al. (2006) noticed that the use of pooling groups for site-similarity failed the simulation for catchments with higher contributes of baseflow, and it was outperformed by a regression based approach. Even if regressions methods, due to the above limitations, are not applied in the

present work, they are still useful to investigate the relationship between the parameters and catchment characteristics.

Within distance-based group, the common hydrological sense would suggest to avoid to average single parameters independently, as it is done in parameter-averaging methods, in order to maintain the correlation structure among parameters, which have been demonstrated to be a necessary condition for reliable runoff predictions (Bárdossy, 2007). On the other hand, some inter-comparison studies reported very good accuracies for parameter-averaging techniques: in Austria, Merz and Blöschl (2004) and Parajka et al. (2005) regionalised the single parameters of the HBV model with an ordinary kriging approach (which interpolates parameter based on their spatial correlation) finding good performances, comparable to those of a single-donor approach also based on geographical distance; with the same model structure, Samuel et al. (2011) found the use of the single parameter set of the most similar or nearest donor to be unsuitable for a set of sparsely located basin in Ontario (Canada), in respect to an inverse-distance weighting approach based on catchment characteristics for averaging parameters. Such results are probably due to the fact that the use of a single-donor approach supposes (1) the presence of a "companion" (gauged) donor catchment in the region, enough hydrologically similar to the target to be suitable for the transfer of the entire parameter set, and (2) that the selected similarity measure is able to identify it.

Output-averaging approaches are able to overcome the limitations of both parameter-averaging and single-donor approaches: transferring the entire sets of model parameters obtained over more than one basin and averaging the simulations allows on one hand to maintain the correlation among parameters, and on the other hand to use the information of more than a single hydrologically similar watershed, eventually averaging out the effect of a misrepresenting donor. In fact, starting from the first application of McIntyre et al. (2005), output-averaging have been adopted successfully in an increasing number of studies (e.g Oudin et al., 2008; Viviroli et al., 2009; Reichl et al., 2009; Zelelew and Alfredsen, 2014; Beck et al., 2016; Arsenault et al., 2019).

1.4 Research questions

1.4.1 Choice of the donor catchments and transfer methods in the study region

It was already pointed out that evidences from the literature, with the exception of the few general guidelines provided in the studies listed in the previous section, do not provide

specific guidance on which and where different parameter regionalisation methods perform best. However, it was also highlighted that distance-based methods are demonstrated to generally overcome other types of approaches. Their application involve several choices, strictly related to the region and data of interest: the selection of the similarity measure, the number of donors to include in the regionalisation process and, of course, also the method to transfer parameters from gauged to ungagued basins. Such aspects are indeed strongly interconnected and have to be carefully evaluated in practical applications.

The first analysis, presented in Chapter 4, was conducted to set up and test different parameter regionalisation frameworks based either on spatial proximity or geomorphoclimatic distance, applied for two hydrological models. This involves the assessment of the type of similarity and the method features which better suit parameter regionalisation for each of the selected rainfall-runoff models. The experiment is conducted over the main study region of this Dissertation: a large set of 209 Austrian catchments. After selecting a group of techniques, based on the literature experience, we will answer to the following preliminary research questions:

Which similarity measure better suits the transferability of the parameters of different rainfall-runoff models? How many donors should be used? And finally: which transfer method performs best?

1.4.2 Impact of the informative content of the region

The reliability of the regionalisation techniques is linked to a number of features of the study region concerning climate, topography and so on; but, one of these aspects is known for playing a central role in method performances: gauged data availability. In very densely gauged areas, spatial proximity is expected to be a good similarity measure, as demonstrated by Merz and Blöschl (2004) and Parajka et al. (2005), who tested different regionalisation approaches on a dense dataset of more than 300 watersheds across Austria. Similar results are presented in Oudin et al. (2008), who examined spatial proximity on a set of 913 French catchments without snow impact. But different outcomes may be obtained when the gauged stations are less dense and less interconnected (that is, with less availability of stations along the same river). For example, Samuel et al. (2011) regionalised the parameters of the HBV model for a sparsely gauged dataset (135 watersheds in the wide area of Ontario, Canada) and found that the best approach for such a study area was an inverse-distance parameter averaging of a pre-selected set of physically similar catchments. Some efforts have been made by hydrologists in order to assess the impact of the number of stream gauges in the region of the target catchment (i.e. gauging station density): Oudin et al. (2008) tested also the effect of the density

of streamflow stations on the above mentioned set of France catchments, finding strong dependency of the performance of output-averaging techniques (based on physycal similarity and spatial proximity) on the number of neighbouring stations. In the same study region, the more recent study Lebecherel et al. (2016) confirmed the central role of the density of the hydrometric network on the use of spatial proximity as similarity measure. On the same line, Patil and Stieglitz (2012) observed that density of measures affects more favourably the regionalisation procedure in humid, rather than in dry, regions. In the review of Parajka et al. (2013b), the outcomes of the methods were compared to the number of stations used in the experimental datasets, but not to their spatial density.

A further characteristic of the available data which has to be taken into account is the topological relationship between the river sections in the dataset: let us assume that we have to set up a rainfall-runoff model for an ungauged river section, and there are available parameter sets calibrated on gauged sections located downstream or upstream on the same river, i.e. the ungauged catchment is *nested* in respect to one or more gauged catchments; in this case the down or upstream gauged basins share part of their drainage area with the target, and they are probably more hydrologically similar to it than any other basin in the dataset. Therefore, it is also likely that parameter regionalisation approaches will perform particularly well in this condition. In companion research fields, geostatistical methods have been developed ad hoc to take advantage of the topological relationship between basins for the regionalisation of hydrological variables not related to the use of rainfall-runoff models (Top-Kriging, Skøien et al., 2006). In model regionalisation, few applications have analysed the usefulness of nested donors: for instance, Merz and Blöschl (2004) showed that averaging parameters from down/upstream donors gives the best streamflow prediction; however, if the condition of "nestedness" applies to most of the basins in the experimental dataset, the significance of the results may be not transferable to a region with less strong topological relationships between the catchments, i.e. "poor" of nested basins. So far no studies have investigated the effect of nested basins on different parameter regionalisation techniques and different rainfall-runoff models.

The data availability, which includes density of gauging stations and topological relationship of the basins in the dataset, will be named here *informative content* of the study region. In order to give further instruments to hydrological modellers for choosing the most suitable technique, one of the analyses of the Dissertation, presented in Chapter 5, will try to answer the following research questions:

How much do i) the presence of nested catchments and ii) station density influence the performance of methods for regionalising rainfall-runoff models? Which techniques and which similarity measures are more affected by a deterioration of the informative content

of the region?

1.4.3 Value of sub-catchment similarity for parameter transfer

The notion of similarity is, and will always be, a critical issue in hydrology. As already pointed out, the use of basin properties to define a similarity measure is the first step of all distance-based techniques, as well as in many other PUB studies (see Blöschl et al., 2013).

Hydrological processes change between regions, between single catchments, but also inside the same basin. Similarity is often defined at catchment scale to transfer hydrological parameters from catchment to catchment, neglecting the pronounced within-catchment variability of the rainfall-runoff processes (Samaniego et al., 2017; Kling and Gupta, 2009; RouholahnejadFreund et al., 2019), that is instead often needed to better reproduce the spatially heterogeneous hydrological processes taking place in the basin. In fact, when applying spatially distributed models, the catchment is usually divided in elementary units (HRUs) or grid cells, by which meteorological inputs and rainfall-runoff generation processes are differentiated; therefore, it is possible to evaluate similarity between the single HRUs, but this has not been explored in detail so far, in particular in order to quantify the benefit of such approach (applied at higher spatial resolution), in respect to those of a lumped modelling framework. A notably exception is the study of Li and Zhang (2017), who compared the impact of the same similarity measures on two rainfall-runoff models (SIMHYD and Xinanjiang) when applied at catchment or at grid scale in Australia: in the case of the grid modelling approach, they transferred the model parameters between grid model elements rather than between entire catchments, finding only marginal improvement in regionalisation accuracy, in respect to the classical lumped approach (but they do not allow the similarity indexes to vary in space and between the two approaches).

In particular, literature so far has not focused specifically on the regionalisation of semi-distributed models, which represent a trade-off between parsimonious lumped and complex fully distributed conceptualisations: some studies have focused on the transfer of the parameters of semi-distributed models between nested basins of the same watershed (e.g. Pechlivanidis et al., 2010; Lerat et al., 2012; de Lavenne et al., 2016), but to the best of our knowledge no studies have so far developed a framework for the transfer of parameters between sub-catchments referring to different rivers.

Looking at the drivers of the spatial variability of the hydrological processes, one of the main factors which influences rainfall-runoff transformations inside the same catchment is with no doubts the elevation. In fact, elevation indirectly influences most of processes through the temperature gradient: precipitation state (liquid or solid), and evapotranspi-

ration, but also the vegetation covers are ruled, between other factors, by temperature. Differentiating the characterisation of hydrological processes, and with them the notion of similarity, across elevation zones, may improve our knowledge about the transformation taking place at different altitudes. In fact, it was already argued that elevation is often included in similarity measures (see e.g. Kokkonen et al., 2003; Parajka et al., 2005; Zhang and Chiew, 2009). Viviroli et al. (2009) favoured the transfer of parameters between catchments at the same elevation considering three homogeneous bands of altitude in Switzerland with marginal benefits, but they used the same index to identified donor at different elevations.

The above mentioned matters lead us to the following part of the Dissertation, exposed in Chapter 6, which will investigate the variability of similarity measures with elevation: using a semi-distributed version of the rainfall-runoff TUW model structure, an innovative semi-distributed calibration and regionalisation frameworks based on elevation zone are proposed. Special focus will be dedicated to the similarity of the runoff generation processes, highly impacted by elevation since involving snow and soil moisture dynamics. Applying the analyses to two very large study regions including overall more than 700 catchments (the already mentioned Austrian case study and an additional dataset of more than 500 US catchments), the research questions which are addressed are:

How is similarity changing across elevation? Is the proposed semi-distributed approach giving benefits to the rainfall-runoff simulations? Does a semi-distributed calibration and regionalisation approach improve simulation in ungauged catchments?

1.4.4 Similarity as interaction between forcing data and runoff

The last issue which will be addressed by this Dissertation represents a separated analysis which, even if not directly aimed at the regionalisation of rainfall-runoff model parameters, lies in the field of catchment classification, which, as already mentioned, is a companion analysis and often a necessary step of regionalisation approaches.

In Sec. 1.2 it was argued that the evaluation of similarity in rainfall-runoff transformations is often assessed also by taking advantage of similarity in catchment runoff response, even if available only at gauged locations.

In the case of rainfall-runoff model regionalisation, similarity should reflect the interaction between meteorological forcings and river streamflow time series, in particular at fine temporal scale, in order to reproduce similarity in the rainfall-runoff transformation processes. A quantification of such interaction may be used to better classify and identify similar watersheds with the purpose of model improvement or parameter regionalisation. Hydrologists have so far tried to capture similarity between basin meteorological forcings

(i.e. similarity of climate) or between streamflow time-series (i.e. similarity of runoff response), but no studies have so far implemented a similarity measure based on the interaction of model inputs and outputs.

Chapter 7, after reporting past research experiences in such direction, explores the potential of an innovative tool, based on Shannon's Information Theory, for measuring the interaction between hydrological signals, in particular between meteorological forcing and river runoff, and introduces a procedure for classifying catchments accordingly. The "gauged nature" of the analysis, which implies the knowledge of the streamflow time series, prevents the direct application of such methodology for regionalisation purposes, but it may help to improve our knowledge about catchment similarity. The analysis tries to answer the following research questions:

Are we able to quantify the interaction between meteorological forcing and river discharge by relying on the concepts of Information Theory? Is such measure representative of the hydrological behaviour of a watershed, improving our knowledge about catchment similarity?

Chapter 2

Study regions and data

In order to test the experiments conducted in this dissertation, the main study region considered is the Austrian country, used for all the analyses. An additional case study extending all across the Conterminous United States (named CONUS) is the employed for the experiment described in Chapter 6. For both regions, data for a large set of catchments, including meteorological inputs and streamflow measures, are available, but basin mutual relationship and climatic variability across the dataset are strongly different between the two cases study.

This chapter presents and describes the study regions and the available data, as well as the processes followed for the extraction of additional catchment characteristics needed for the analysis.

2.1 Main study region: Austria

In Austria, the available case study is composed by 209 catchments (see Fig. 2.1, panel a)) covering a large portion of the country. Their size varies considerably, mainly under 1000 km^2 (90% of the basins) and just 3 watersheds extend over more than 3000 km^2. The topography of the country varies significantly from the flat and hilly area in the north-east to the Alps in the centre and in the south-west, particularly steep in the extreme west. The annual precipitation ranges from about 600 mm in the east, where the evaporation plays an important role in the water balance, to more than 2000 mm in the west, mainly due to orographic lifting of north-westerly airflows at the rim of the Alps (Viglione et al., 2013). Even if the amount of precipitation and evapotranspiration ranges consistently, the aridity index assumes values from 0.2 to 1, meaning that the all the watersheds are mainly wet or weakly arid (annual evapotranspiration is never higher than precipitation). Land use is mainly agricultural in the lowlands and forest in the

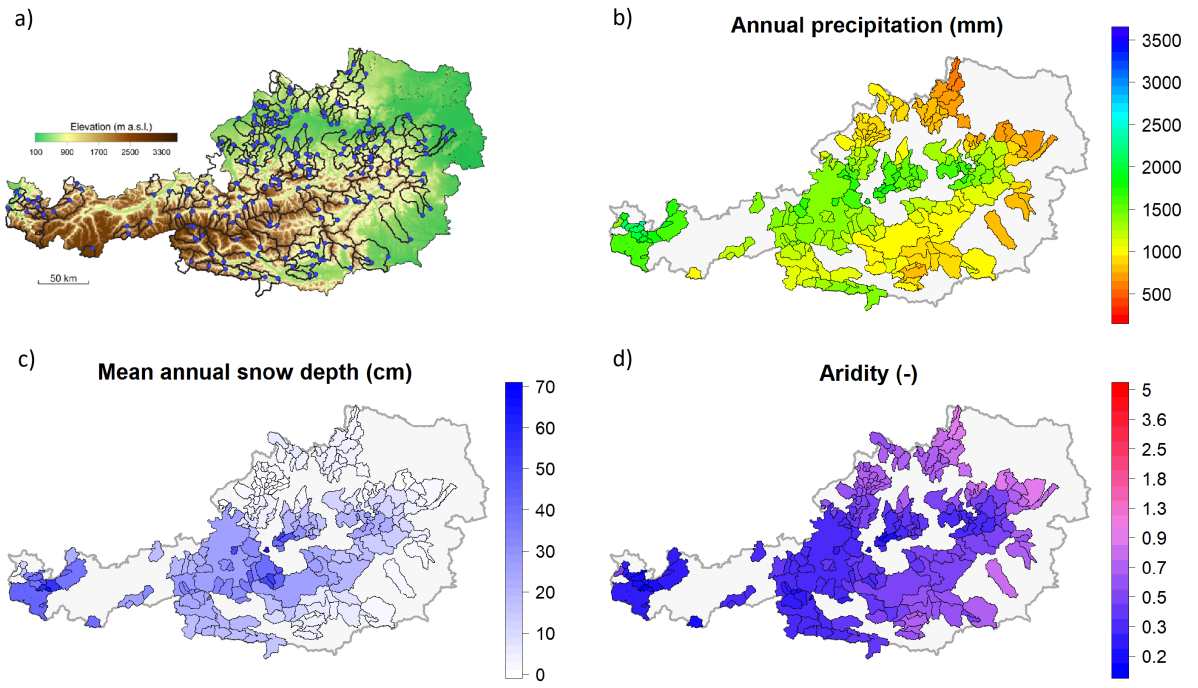

Figure 2.1: Panel a) Study area, blue points refer to stream gauges and black lines to catchment boundaries. Panels b), c) and d) Spatial pattern of some climatic attributes across the study area.

medium elevation ranges. Alpine vegetation and rocks prevail in the highest catchments (Parajka et al., 2005). Fig. 2.1 (panels b), c) and d)) shows the spatial pattern of mean annual precipitation, snow depth and aridity index across the study area.

Data have been provided by the Institute of Hydraulic Engineering and Water Resources Management (Vienna University of Technology), which previously screened the runoff data for errors and removed all stations with significant anthropogenic effects. Hydro-meteorological data include daily streamflow and daily inputs to the rainfall-runoff models for the 33 years period 1976-2008: daily average precipitation, temperature and potential evapotranspiration defined for 200 meters elevation zones for all the study catchments. Model inputs were prepared by the Austrian colleagues, who spatially interpolated daily precipitation and air temperature values from 1091 and 212 climatic stations respectively, using methods which include elevation as auxiliary information: in particular, external drift kriging was used for precipitation, and the least-squares trend prediction method was used for air temperatures (Pebesma, 2001). The potential evapotranspiration was estimated by a modified Blaney-Criddle method (Parajka et al., 2005) using interpolated daily air temperature and grid maps of potential sunshine duration calculated by the Solei-32 model (Mészáros et al., 2002). Spatially interpolated values were then averaged for each elevation zone.

In addition to hydro-meteorological data, a set of geo-morphoclimatic catchment at-

Table 2.1: Available catchment attributes for the Austrian case study.

Code	Unit	Min	Median	Max	Description
Elev	m a.s.l.	287	915	2964	Average elevation
Area	km^2	14	168	6214	Drainage area
Slope	m/m	0.9	12.4	28.5	Mean slope
Precip	mm	675	1230	2310	Mean annual total precipitation
maxP	mm	35	49	84	Mean annual maximum daily precipitation
PET	mm	281	608	715	Mean annual total evapotranspiration
SnowF	-	0.06	0.17	0.60	Fraction of precipitation fallen as snow (i.e. precipitation fallen in days below 0°)
SnowD	cm	1	14	68	Mean annual snow depth
Aridity	-	0.21	0.46	0.96	Aridity index (meanPET/meanP)
Irrad	$kWh/m^2/day$	1750	1899	2274	Mean annual solar irradiance
RiverD	m/km^2	0	830	1256	Stream network density
FARL	-	0.56	1	1	Flood attenuation index by reservoir and lakes
Land Cover	%	-	-	-	Portions of land use coverage
Geology	%	-	-	-	Portions of geological formations
Soils	%	-	-	-	Portions of regional soil types
Forest	-	0	0.47	0.93	Fraction of catchment covered in forest
AcqPort	-	0	0.01	0.83	Fraction of catchment with porous aquifers

tributes were also made available: topographic attributes such as mean catchment elevation and mean slope, which were previously derived from 1×1 km digital elevation model; climatic features such as mean and maximum annual precipitation, aridity index, and snow fraction of precipitation were instead derived from climatic input time series, while average snow depth was interpolated (analogously to what done for daily precipitation) from the vast snow gauge network covering the country; stream network density was calculated from the digital river network map at the 1:50000 scale for each catchment (Merz and Blöschl, 2004); FARL (i.e. flood attenuation index by reservoirs and lakes), boundaries of porous aquifers, and main geological formation were the same used and described in detail in (Parajka et al., 2005).

The set of available catchment attributes was completed with three additional features which were ad-hoc calculated for the each watershed, based on its drainage contour:

- land use coverage, derived from CORINE Land Cover maps updated to year 2012 (https://land.copernicus.eu/pan-european/corine-land-cover/clc-2012)

- mean annual solar irradiation, computed with the function *r.sun.daily* of GRASS GIS software (GRASS Development Team, 2017) as function of the latitude and the terrain topography (slope and aspect), derived from the Shuttle Radar Topography Mission (SRTM) Global Digital Elevation Model with a resolution of 3 Arc-Seconds ($\sim$ 90 meters), freely available at http://srtm.csi.cgiar.org (Reuter et al., 2007)

- soil type coverage, extracted from the FAO/UNESCO Soil Map of the World (http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/en/)

Tab. 2.1 reports a summary of the presented basin descriptors, along with their main statistics across the study region. For land cover classes, as well as for geology and soil type classes, the catchments are associated to more than one single attribute: each basin is described by the portions of the total catchment area corresponding to each class (and for this reason, Tab. 2.1 does not report the min/median/max values of such descriptors.

2.2 Additional study region: United States (CAMELS dataset)

The second selected case study belongs to the Contiguous United States of America (i.e. continental portion of the USA with the exception of Alaska, also called CONUS). The choice of such study region is due to the recent publication of a very large-sample (and open access) watershed-scale hydrometeorological dataset of 671 catchments covering great portion of the country and minimally impacted by human activities: the CAMELS dataset, provided by Newman et al. (2015), who developed and made available the hydrometeorological data, and Addor et al. (2017), who complemented the dataset with catchment attributes.

The 671 basins span the entire CONUS and cover a wide range of hydroclimatic conditions. They range from wet, warm basins in the southeastern (SE) US to hot and dry basins in the southwestern (SW) US, to wet, cool basins in the northwestern (NW) and dry, cold basins in the intermountain (Rocky Mountains) western US (Newman et al., 2015). Fig. 2.2 shows the variability of mean annual precipitation (mm) and aridity (-) across the watersheds; square dots refer to snowy catchments, where the fraction of solid precipitation is higher than 10%. The availability of water varies considerably in the country: there are many energy-limited basins with dryness ratios as small as 0.2 and many water-limited basins with dryness ratios as large as 5.

The dataset was specifically built for hydrological modelling: daily meteorological forcing data are available at different spatial scales, in order to fit a variety of rainfall-runoff model configurations: lumped (entire catchment), 100 m elevation bands and hydrologic response units (HRUs). They were calculated via areal average from the gridded meteorological dataset Daymet (Thornton and Running, 1999): it is a daily, gridded (1×1 km) data set over the CONUS and southern Canada and is available from 1980 to present. Forcing variables are available in the 30 years period 1980-2011 and include daily maximum and minimum temperature, precipitation, shortwave downward radiation, day

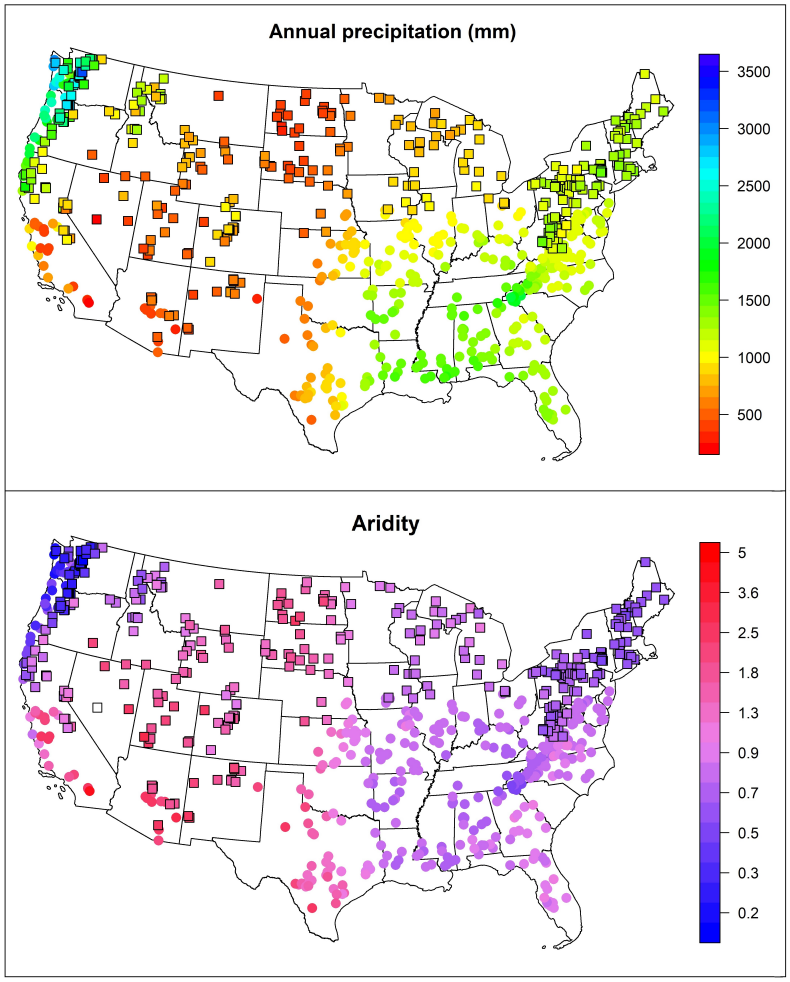

Figure 2.2: Pattern of annual precipitation (upper panel) and aridity index (lower panel) across CAMELS dataset. Circles denote basins with >90% of their precipitation falling as rain, squares with black outlines denote basins with >10% of their precipitation falling as snow as determined by using a 0 °C daily mean Daymet temperature threshold

length, humidity and snow water equivalent. Exclusively for the lumped configuration, additional meteorological daily inputs derived from Maurer et al. (2002) and National Land Data Assimilation System (NLDAS) (Xia et al., 2012) 12 km gridded datasets are available, but they were not used for this dissertation.

Daily streamflow data for the 671 gages were obtained from the USGS National Water Information System server (http://waterdata.usgs.gov/usa/nwis/sw) over the same forcing data time period, 1980–2011.

Additional details about hydrometeorological data collection and characteristics of CAMELS dataset can be found in Newman et al. (2015).

Addor et al. (2017) provides a rich set geo-morphoclimatic attributes aggregated at basin scale. Tab. 2.2 reports a summary of main catchment characteristics across CAMELS dataset. Along with hydrometeorolical data, shapefiles relating to the contours

Table 2.2: Summary of catchment characteristics for CAMELS dataset.

Code	Unit	Min	Median	Max	Description
Elev	m a.s.l.	10	463	3571	Average elevation
Area	km^2	4	330	25761	Drainage area
Slope	m/km	0.8	28.8	255.7	Mean slope
Precip	mm	235	1179	3262	Mean annual total precipitation
PET	mm	693	981	1732	Mean annual total evapotranspiration
SnowF	-	0	0.10	0.91	Fraction of precipitation fallen as snow (i.e. precipitation fallen in days below 0°)
SWE	mm	0	4	575	Mean annual snow water equivalent
Aridity	-	0.22	0.86	5.21	Aridity index (meanPET/meanP)
Irrad	$kWh/m^2/day$	2172	2926	4278	Mean annual solar irradiance
Forest	-	0	0.63	0.99	Fraction of catchment covered in forest

of basin drainage areas, elevation bands and HRUs are also made available in the dataset. Corresponding measures of drainage area are provided as well.

2.2.1 Dataset screening and basin selection

The available data were entirely screened in order to check for missing data and eventually for inconsistency. Such process brings to the exclusion of part of the watersheds of the dataset, which were considered to be unsuitable for the analysis. More in specific, the catchments presenting the following issues were excluded:

- one or more elevation zones with negative values of drainage area
- inconsistency between the value of the basin drainage area and the sum of the areas corresponding to each elevation zone (error $> 10\%$)
- portion of missing streamflow data greater than 10%
- missing forcing data

In addition, in order to be able to compare the results between the two cases study, 3 very large basin extending over more than 6000 km^2 were excluded. Finally, 527 catchments are selected for rainfall-runoff model simulations.

2.2.2 Preparation of rainfall-runoff model inputs: computation of potential evapotranspiration

The original dataset was re-organised and prepared for rainfall-runoff model simulation: in particular, meteorological data at 100 m elevation bands were gathered and

assembled to be given as input to the rainfall-runoff models.

It may be noticed that Daymet database does not include estimates of potential evapotranspiration (PET), a commonly needed input for conceptual hydrologic models (including those used for this dissertation). Therefore, PET was estimated here using the simplified method proposed by Blaney and Criddle (1962), which is based only on the measures of average daily temperature. The monthly potential evapotranspiration is estimated with the following formula:

$$PET_m = a + b\, N(i)\, W_{ta}(i)\, T_m(i) \tag{2.1}$$

where:

- T_m is the average temperature for month i
- $N(i)$ is the number of maximum hours of daylight for month i, function of the Latitude
- a and b are parameters to be estimated
- W_{ta} are monthly correction factors depending on long term average monthly temperature and on average elevation. They are calculated making use of tabulated values for specific altitudes and temperatures, used for estimating the coefficients A, B and C of the following formula:

$$W_{ta}(i) = A\, T_m^2(i) + B\, T_m(i) + C \tag{2.2}$$

For the estimate of a and b, a reference value of evapotranspiration from Thornthwaite formula (Thornthwaite, 1948) is used and parameters are estimated through a linear regression. Eq. 2.1 provides monthly values of potential evapotranspiration, thus it has to be divided by the number of time steps of the current month. In this case, the method was applied at elevation zone level to the available time series of air temperature for all the study catchments.

Chapter 3

Materials and methods

3.1 Rainfall-runoff models

This section presents in detail the rainfall-runoff models implemented in the analysis conducted for this dissertation. They are both continuous-simulation models operating at daily time steps: the TUW model (Section 3.1.1) and the CemaNeige-GR6J model (Section 3.1.1).

3.1.1 TUW model

The TUW model is a semi-distributed version of the HBV model (see e.g. Bergström, 1976; Lindström et al., 1997) developed by Viglione and Parajka (2018), and available through the R-package *TUWmodel*. It is widely applied in the literature, for instance testing parameter regionalisation techniques (see e.g. Parajka et al., 2005), hydrological projection (see e.g. Melsen et al., 2018) or the combined use of model calibration and data assimilation techniques (see e.g. Nijzink et al., 2018).

The model consists in a snow routine, a soil moisture routine and a flow response and routing routine (fully described below). The semi-distributed structure allows the user to conceptually divide the catchment in sub-basins, typically elevation zones (as done so far in the literature) but different catchment subdivisions could be considered as well (e.g. main land cover classes). The model processes the elevation zones as autonomous entities that contribute separately to the total outlet flow. It was designed for daily rainfall-runoff simulations (as used in the present work) but, if properly calibrated, can be applied also at different time scale. The model inputs are:

- T: air temperature (°C);
- P: precipitation (mm/day);

- PET: potential evapotranspiration (mm/day);
- *Sub-basin areas*: portions of catchment area corresponding to each sub-basin (-).

Each of the climatic input is a $n \times m$ matrix where n is the number of elevation zones and m the length of the time series: elements $T_{i,t}$, $P_{i,t}$ and $PET_{i,t}$ refer to the average value of the inputs at time step t for the elevation zone i. As anticipated, the model is run independently over the different elevation zones in the version of Fig. 3.1, described below. Finally, the outputs from the elevation zones are averaged taking into account sub-catchment areas. While model inputs and model states are defined over each of the sub-basins, model parameters can be either unique for the entire catchment (i.e. one single set of parameters for all the elevation zones) or differentiated across the sub-basins. All internal model state variables and all model outputs are given in mm, normalized in respect to catchment area.

Description of model routines

For the sake of simplicity, model routines are here described omitting the notation i which refers to the elevation zone. TUW model parameters (bold type in the text) are summarized in Tab. 3.1 which recalls their brief description as well. The snow routine is based on a simple degree-day concept and it is ruled by five parameters. Two threshold temperature, $\boldsymbol{T_R}$ and $\boldsymbol{T_S}$, are used to separate the total precipitation input P into rainfall P_R and snowfall P_S:

$$P_{R,t} = \begin{cases} P_t & \text{if} \quad T_t \geq \boldsymbol{T_R} \\ P_t \dfrac{T_t - \boldsymbol{T_S}}{\boldsymbol{T_R} - \boldsymbol{T_S}} & \text{if} \quad \boldsymbol{T_R} < T < \boldsymbol{T_R} \\ 0 & \text{if} \quad T_t \leq \boldsymbol{T_S} \end{cases} \tag{3.1}$$

$$P_{S,t} = P_t - P_{R,t} \tag{3.2}$$

where $\boldsymbol{T_R}$ and $\boldsymbol{T_S}$ are model parameters. The solid fraction of precipitation is multiplied by a snow correction factor $\boldsymbol{SCF}$ (model parameter) and feeds the snow water equivalent reservoir SWE (Eq. 3.3). The use of the coefficient $\boldsymbol{SCF}$ allows to better calibrate the contribution of precipitation to the snow storage. SWE produces a snowmelt M when air temperature exceeds the melting temperature threshold $\boldsymbol{T_M}$ (Eq. 3.4), which is then subtracted from the reservoir (Eq. 3.5) at the same timestep.

$$SWE_t = SWE_{t-1} + \boldsymbol{SCF}\, P_{S,t} \tag{3.3}$$

$$M_t = \begin{cases} \Big(T_t - \boldsymbol{T_M}\Big)\,\boldsymbol{DDF} & \text{if} \quad T_t > \boldsymbol{T_M} \;\;\&\;\; SWE_t > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.4}$$

$$SWE_t = SWE_t - M_t \tag{3.5}$$

where $\boldsymbol{T_M}$ and $\boldsymbol{DDF}$ (degree-day factor) are also model parameters. Both rainfall and snowmelt contribute to soil moisture. The soil moisture routine represents soil moisture state changes and runoff generation and involves three parameters: the maximum soil moisture storage $\boldsymbol{FC}$, a parameter representing the soil moisture state above which evapotranspiration is at its potential rate, $\boldsymbol{LP}$, and a parameter $\boldsymbol{\beta}$ ruling the non-linear function of runoff generation. The portion of snowmelt and rainfall ΔS_{UZ}, which generates runoff, is function of the soil moisture level SM and of the parameters $\boldsymbol{FC}$ and $\boldsymbol{\beta}$ (Eq. 3.6). The level of soil moisture storage at each time step is computed through the water balance (Eq. 3.7) between the recharge contributes of rainfall and snowmelt and ΔS_{UZ}. If SM exceeds $\boldsymbol{FC}$, surplus water is summed to ΔS_{UZ}. The actual evapotranspiration term AET is expressed as function of potential evapotranspiration PET (model input), $\boldsymbol{LP}$ and $\boldsymbol{FC}$ (Eq. 3.8), and it exits the system (Eq. 3.9).

$$\Delta S_{UZ,t} = \left(\frac{SM_{t-1}}{\boldsymbol{FC}}\right)^{\boldsymbol{\beta}} \Big(P_{R,t} + M_t\Big) \tag{3.6}$$

$$SM_t = SM_{t-1} + P_{R,t} + M_t - \Delta S_{UZ,t} \tag{3.7}$$

$$AET_t = \min\Big(PET_t\,\frac{SM_t}{\boldsymbol{LP}\,\boldsymbol{FC}},\;\; PET_t\Big) \tag{3.8}$$

$$SM_t = SM_t - AET_t \tag{3.9}$$

Finally, an upper and a lower soil reservoirs and a triangular transfer function compose the runoff response and routing routine, involving seven additional parameters $\boldsymbol{L_{UZ}}$, $\boldsymbol{k_0}$, $\boldsymbol{k_1}$, $\boldsymbol{k_2}$, $\boldsymbol{C_{PERC}}$, $\boldsymbol{B_{MAX}}$ and $\boldsymbol{C_{ROUTE}}$. Excess rainfall and snowmelt enter the upper zone reservoir of level S_{UZ} (Eq. 3.10) and leaves this reservoir through three paths: outflow from the reservoir based on a fast storage coefficient $\boldsymbol{k_1}$ (Eq. 3.12); percolation to the lower reservoir of level S_{LZ} with a constant percolation rate $\boldsymbol{C_{PERC}}$ (Eq. 3.14) and, if the threshold of the storage state $\boldsymbol{L_{UZ}}$ is exceeded, through an additional outlet based on a very fast storage coefficient $\boldsymbol{k_0}$ (Eq. 3.11). The three contributes are then deducted from S_{UZ} (Eq. 3.13). Water leaves the lower reservoir of level S_{LZ} based on a slow storage

coefficient $\boldsymbol{k_2}$ (Eq. 3.15 and Eq. 3.16).

$$S_{UZ,t} = S_{UZ,t-1} + \Delta S_{UZ,t} - q_{0,t} - q_{1,t} - \boldsymbol{C_{PERC}} \tag{3.10}$$

$$q_{0,t} = \frac{S_{UZ,t} - \boldsymbol{L_{UZ}}}{\boldsymbol{k_0}} e^{-\frac{1}{\boldsymbol{k_0}}} \tag{3.11}$$

$$q_{1,t,t} = \frac{S_{UZ,t}}{\boldsymbol{k_1}} e^{-\frac{1}{\boldsymbol{k_1}}} \tag{3.12}$$

$$S_{UZ,t} = S_{UZ,t} - q_{0,t} - q_{1,t} - \boldsymbol{C_{PERC}} \tag{3.13}$$

$$S_{LZ,t} = S_{LZ,t-1} + \boldsymbol{C_{PERC}} \tag{3.14}$$

$$q_{2,t} = \frac{S_{LZ,t}}{\boldsymbol{k_2}} e^{-\frac{1}{\boldsymbol{k_2}}} \tag{3.15}$$

$$S_{LZ,t} = S_{LZ,t} - q_{2,t} \tag{3.16}$$

Table 3.1: TUW model parameters.

Parameter	Description	Units
$\boldsymbol{SCF}$	Snow correction factor	-
$\boldsymbol{DDF}$	Degree day factor	mm/°C/day
$\boldsymbol{T_R}$	Threshold temperature above which precipitation is rain	°C
$\boldsymbol{T_S}$	Threshold temperature below which precipitation is snow	°C
$\boldsymbol{T_M}$	Threshold temperature above which melt starts	°C
$\boldsymbol{LP}$	Parameter related to the limit of evaporation	-
$\boldsymbol{FC}$	Field capacity (i.e., max soil moisture storage)	mm
$\boldsymbol{\beta}$	Non linear parameter for runoff production	-
$\boldsymbol{k_0}$	Storage coefficient for very fast response	days
$\boldsymbol{k_1}$	Storage coefficient for fast response	days
$\boldsymbol{k_0}$	Storage coefficient for slow response	days
$\boldsymbol{L_{UZ}}$	Threshold storage state above which very fast response start	mm
$\boldsymbol{C_{PERC}}$	Constant percolation rate	mm
$\boldsymbol{B_{MAX}}$	Maximum base at low flows	days
$\boldsymbol{C_{ROUTE}}$	Scaling parameter	days^2/mm

The outflows from both reservoirs is then routed by a triangular transfer function representing runoff routing in the streams, where the base of transfer function B_Q (Eq. 3.17) is

estimated with the scaling of the outflow by the $\boldsymbol{C_{ROUTE}}$ and $\boldsymbol{B_{MAX}}$ parameters (Parajka et al., 2005).

$$B_{Q,t} = \max\Big(\boldsymbol{B_{MAX}} - \boldsymbol{C_{ROUTE}}\,(q_{0,t} + q_{1,t} + q_{2,t}),\quad 1\Big) \tag{3.17}$$

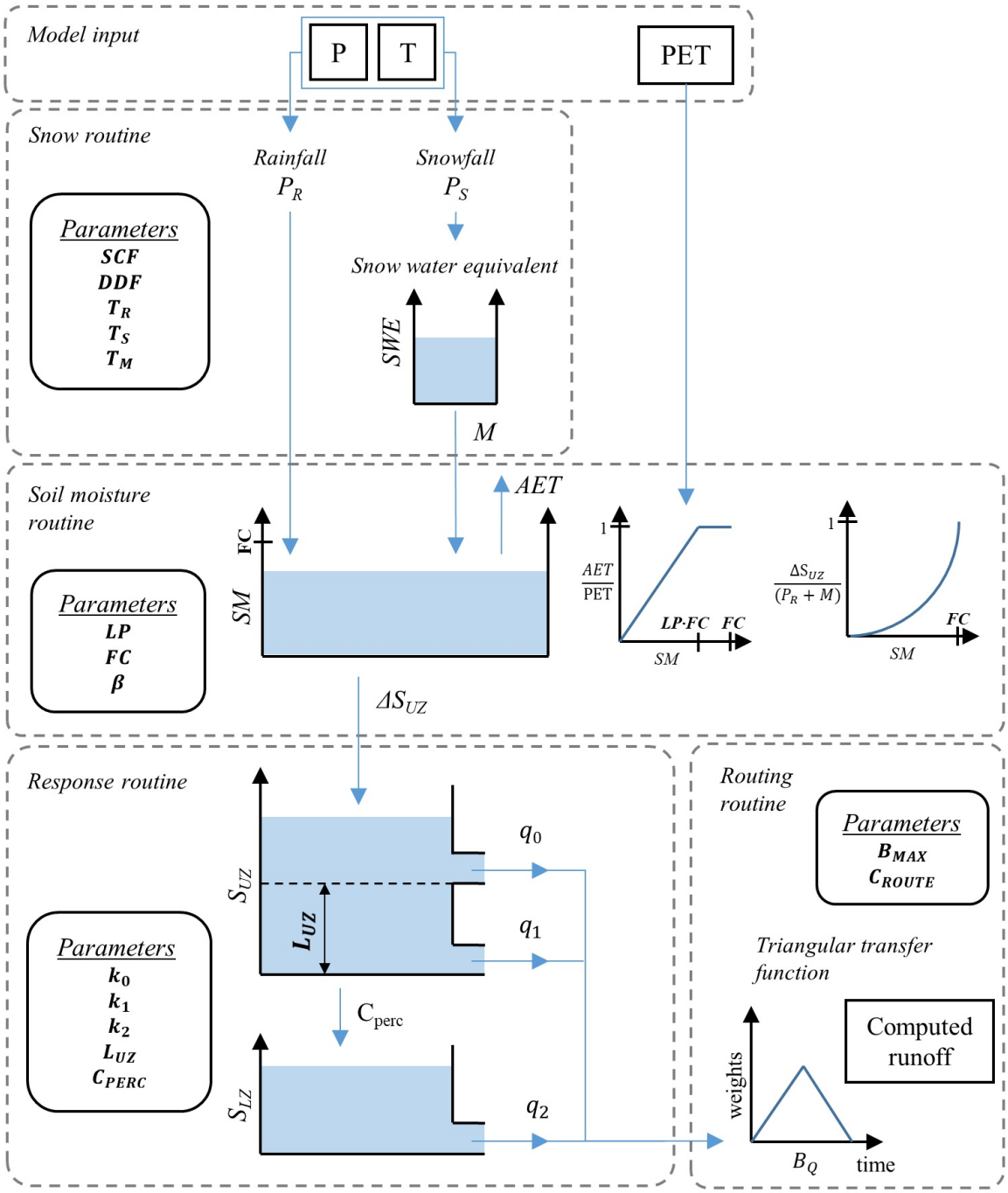

Figure 3.1: TUW model scheme - Lumped version (Neri et al., 2020).

3.1.2 CemaNeige-GR6J model

The second model is the French CemaNeige-GR6J (Coron et al., 2017b). It is the combination of the CemaNeige snow accounting routine (Valéry et al., 2014) with the GR6J model (Pushpalatha et al., 2011), a daily conceptual continuous rainfall-runoff model belonging to the suite of the *GR Hydrological Models*, developed at INRAE (Antony, France), by the Équipe Hydrologie des Bassins versants. The software is freely available in the *airGR* R-package (Coron et al., 2017a). Similar to the TUW, the literature shows how the GR models are implemented for different purposes: for instance they were used to regionalise rainfall-runoff parameters (see e.g. Oudin et al., 2008), to investigate the reasons behind pattern of model performance (see e.g. Poncelet et al., 2017) or to include satellite data in hydrological simulation (see e.g. Riboust et al., 2019).

As anticipated, the model is composed by a snow module called CemaNeige, whose output is processed by the GR6J model which simulates the total runoff at the basin outlet. Overall, the model is lumped, so it requires spatially-averaged catchment daily inputs: air temperature T (°C), precipitation P (mm/day) and potential evapotranspiration PE (mm/day).

The following sections report the description of the two model routines, where model parameters are again highlighted in bold type and then summarised in Tab. 3.2. Here again, all internal model state variables and all model outputs are normalized in respect to catchment area and expressed in mm.

Description of CemaNeige model routines

The CemaNeige snow accounting routine is based on an improved degree-day concept, which is based on the same principle of the snow module described for the TUW (Eq. from 3.18 to 3.20), but considering two additional features to enhance the accuracy of snow accumulation and melting:

- the thermal inertia of the snowpack is taken into account in order to estimate its cold-content, which can delay the beginning of the snow melt;
- a simplified approach to account for partial snow cover of the basin surface is considered.

Although the module requires daily lumped inputs (precipitation and air temperature), for better simulating snow accumulation and melting it allows to divide the catchment into more elevation zones of equal area. In this case, the hyspometric curve of the catchment is also required. Inputs for each elevation zone (T_e and P_e) are extracted through

interpolation of the mean catchment values using precipitation and temperature gradients (described in detail in Valéry et al., 2010), and not from "clipping" of the actual spatial fields like for the TUW elevation zones. The module functions are applied with a lumped set of parameters; but internal states are allowed to vary over each elevation layer according to the different extrapolated inputs. On each elevation layer, two outputs are computed: rain and snowmelt, which are summed in order to find the total water quantity feeding the hydrological model. At every time step, the total liquid output of CemaNeige at catchment scale is the average of every elevation zone outputs. Here, the description of the routines applied to each single elevation zone is reported.

The average air temperature value T_e is used first to compute the snow fraction SF which allow to divide the total precipitation P_e into rainfall P_R and snowfall P_S, following the rule proposed by the *US Army Corps of Engineers* (1956):

$$SF_t = \begin{cases} 0 & \text{if} \quad T_{e,t} > 3°\text{C} \\ 1 & \text{if} \quad T_{e,t} < \text{-}1°\text{C} \\ 1 - \dfrac{T_{e,t} + 1}{4} & \text{otherwise} \end{cases} \tag{3.18}$$

$$P_{S,t} = SF_t \, P_{e,t} \tag{3.19}$$

$$P_{R,t} = P_{S,t} - P_{e,t} \tag{3.20}$$

Snowfall is accumulated in an unique storage of level G (Eq. 3.21), that is the snow water equivalent. The cold-content (or heat capacity) of the snow pack eT_G is updated at every time step following Eq. 3.22 as function of the air temperature and of the model parameter $\boldsymbol{\theta_{G2}}$, which accounts for the thermal inertia of the system. If both air temperature and cold-content are above 0°C, the potential snowmelt PM is generated based on a degree-day approach (Eq. 3.23), which can never exceed G.

$$G_t = G_{t-1} + P_{S,t} \tag{3.21}$$

$$eT_{G,t} = \boldsymbol{\theta_{G2}} \, eT_{G,t-1} + (1 - \boldsymbol{\theta_{G2}}) \, T_{e,t} \tag{3.22}$$

$$PM_t = \begin{cases} \boldsymbol{\theta_{G1}} \, T_{e,t} & \text{if} \quad eT_{G,t} \geq 0°\text{C} \;\; \& \;\; T_{e,t} > 0°\text{C} \\ 0 & \text{otherwise} \end{cases} \tag{3.23}$$

where $\boldsymbol{\theta_{G1}}$ is the degree-day factor (model parameter). The model then simulates the

portion of basin covered by snow SC (Eq. 3.24) as the ratio between the actual snow water equivalent and an empirical fixed threshold $G_{threshold}$, which is automatically set by the model to the 90% of the annual solid precipitation. Such quantity allows to calculate the effective snowmelt M following Eq. 3.25, which is deducted form the snow storage (Eq. 3.26).

$$SC_t = \min\left(\frac{G_t}{G_{threshold}},\ 1\right) \tag{3.24}$$

$$M_t = (0.9\, SC_t + 0.1)\, PM_t \tag{3.25}$$

$$G_t = G_t - M_t \tag{3.26}$$

The total water quantity CN_{OUT}, the output of the CemaNeige model, is the sum of effective rainfall and snow melt:

$$CN_{OUT} = P_{R,t} + M_t \tag{3.27}$$

Description of GR6J model routines

The total liquid output of CemaNeige module averaged over all the elevation layers $\overline{CN}_{OUT}$ and potential evapotranspiration PE are the inputs of the GR6J rainfall-runoff model. In the model, the water balance is controlled by a soil moisture accounting reservoir and a conceptual "groundwater" exchange function, while the routing part of the structure consists in two flow components routed by two unit hydrographs, a non-linear store and an exponential-store, with a total of six parameters. The structure of the model is represented in Fig. 3.2: it can be noticed that such representation refers to the original structure of the model when no snow module is included (in fact, input are PE and P). In this case, precipitation is of course replaced by the total liquid contribute of CemaNeige:

$$P = \overline{CN}_{OUT} \tag{3.28}$$

Effective liquid contribute P_n and net evapotranspiration E_n are computed with the following equations:

$$P_{n,t} = \begin{cases} P_t - PE_t & \text{if} \quad P_t \geq PE_t \\ 0 & \text{otherwise} \end{cases} \tag{3.29}$$

$$E_{n,t} = \max\left(PE_t - P_t,\ 0\right) \tag{3.30}$$

The soil moisture module is composed by a non-linear production storage of maximum

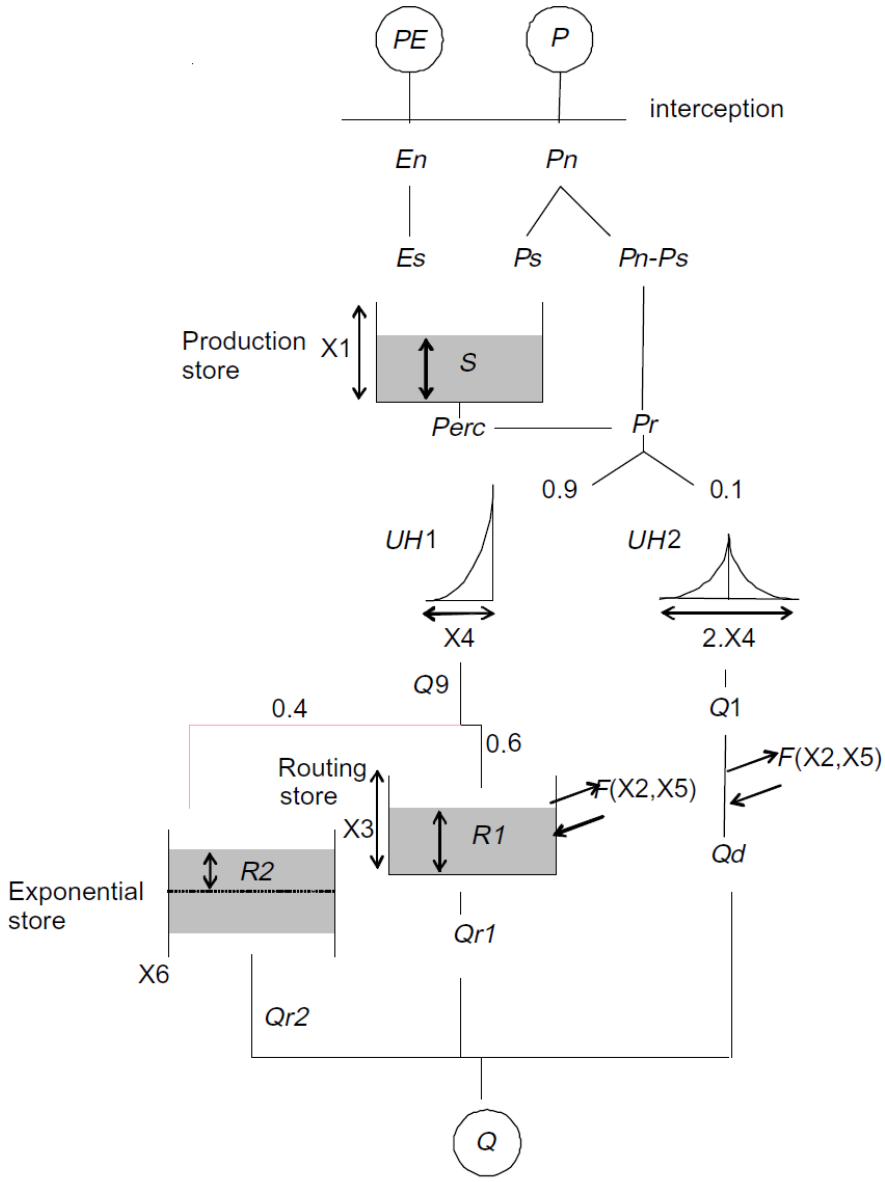

Figure 3.2: GR6J model scheme (Pushpalatha et al., 2011).

capacity $\boldsymbol{X1}$ (model parameter) and level S. At each time step, the level S is updated (Eq. 3.33) considering that a portion of the effective liquid contribute, termed P_s, enters in the storage, following the relation Eq. 3.31, while the actual evapotransipration E_s (Eq. 3.32) is removed from it.

$$P_{s,t} = \frac{\boldsymbol{X1}\left[1 - \left(\frac{S_{t-1}}{\boldsymbol{X1}}\right)^2\right]\tanh\left(\frac{P_{n,t}}{\boldsymbol{X1}}\right)}{1 + \frac{S_{t-1}}{\boldsymbol{X1}}\tanh\left(\frac{P_{n,t}}{\boldsymbol{X1}}\right)} \tag{3.31}$$

$$E_{s,t} = \frac{S_{t-1}\left(1 - \frac{S_{t-1}}{\boldsymbol{X1}}\right)\tanh\left(\frac{E_{n,t}}{\boldsymbol{X1}}\right)}{1 + \left(1 - \frac{S_{t-1}}{\boldsymbol{X1}}\right)\tanh\left(\frac{E_{n,t}}{\boldsymbol{X1}}\right)} \tag{3.32}$$

$$S_t = S_{t-1} - E_{s,t} + P_{s,t} \tag{3.33}$$

For better understanding the relationship between the effective rainfall/evapotranspiration and the level in the production store, Fig. 3.3 reports such relationship in a normalised graph.

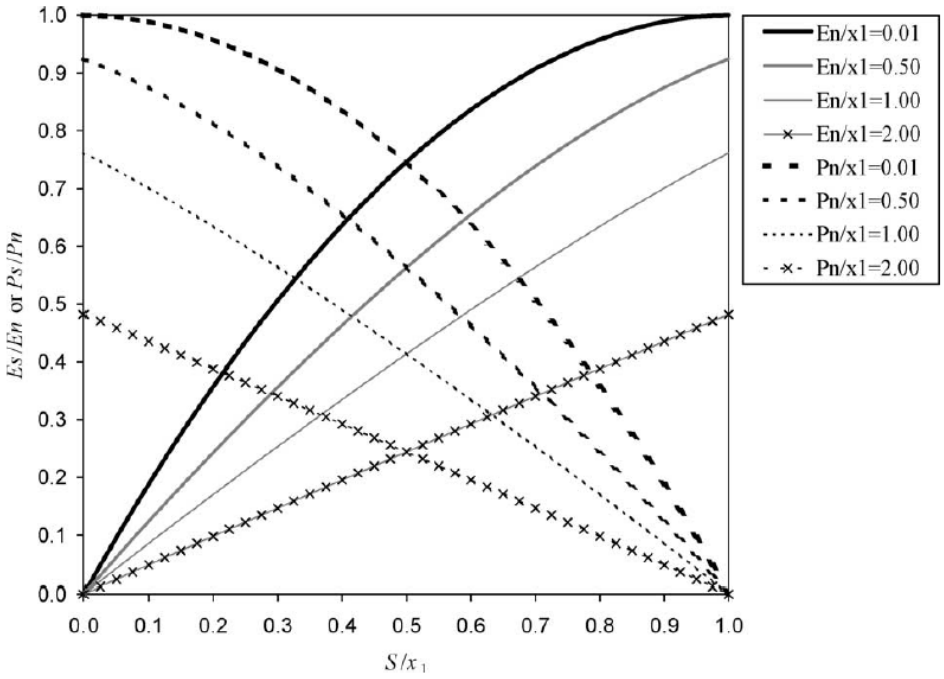

Figure 3.3: Illustration of the behaviour of the production functions (E_s/E_n: solid line; P_s/P_n: dashed line) as a function of storage rate $S/\boldsymbol{X1}$ for different values of $E_s/\boldsymbol{X1}$ or $P_s/\boldsymbol{X1}$ (Perrin et al., 2003).

A percolation rate then is released from the storage: it is simulated as function of the updated level S_t, which is again updated at the same time step:

$$Perc_t = S_t\left\{1 - \left[1 + \left(\frac{4}{9}\frac{S_t}{\boldsymbol{X1}}\right)^4\right]^{-\frac{1}{4}}\right\} \tag{3.34}$$

$$S_t = S_t - Perc_t \tag{3.35}$$

Observing Eq. 3.31 and Fig. 3.3, it can be noticed how S can never exceed $\boldsymbol{X1}$ by nature; $Perc$, in turn, can not be grater than S itself.

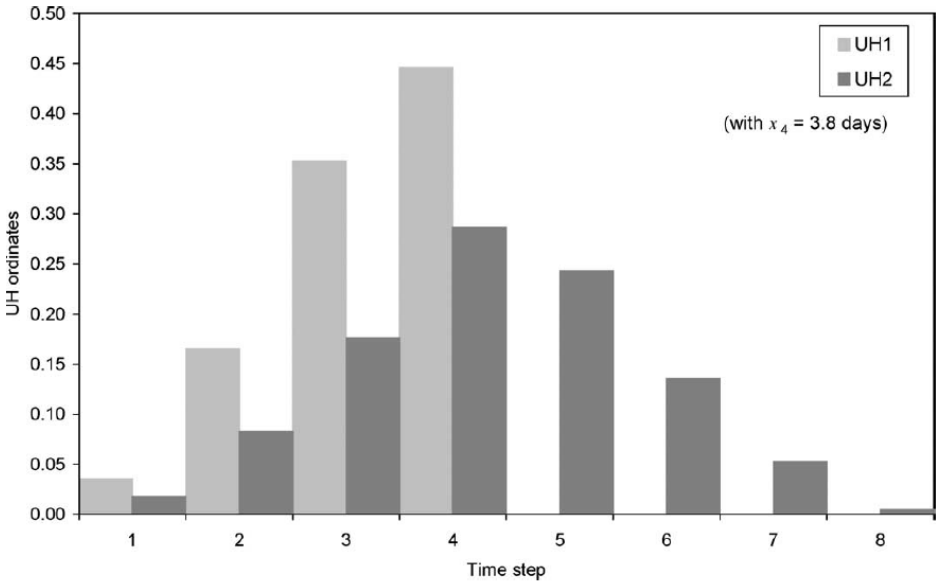

Figure 3.4: Example of the ordinates of $UH1$ and $UH2$ for parameter $\boldsymbol{X4} = 3.8$ days (Perrin et al., 2003).

The total runoff generated P_r from the soil moisture routine is computed as the sum of the percolation rate and the portion of effective rainfall, which bypass the production store:

$$P_{r,t} = Perc_t + (P_{n,t} - P_{s,t}) \tag{3.36}$$

This total amount of water is then divided into two components: the 90% of P_r is routed by the unit hydrograph $UH1$ and the remaining 10% is transformed by a single unit hydrograph $UH2$, whose output are termed respectively $Q9$ and $Q1$. With $UH1$ and $UH2$, one can simulate the time lag between the rainfall event and the resulting streamflow peak. Their ordinates are several successive time steps. Both unit hydrographs depend on the same time parameter $\boldsymbol{X4}$ expressed in days. However, $UH1$ has a time base of $\boldsymbol{X4}$ days whereas $UH2$ has a time base of $2\,\boldsymbol{X4}$ days. $\boldsymbol{X4}$ can take real values and is greater than 0.5 days. Fig. 3.4 shows an example of unit hydrograph ordinates for $\boldsymbol{X4} = 3.8$ days In their discrete form, unit hydrographs $UH1$ and $UH2$ have n and m ordinates, respectively, where n and m are the smallest integers exceeding $\boldsymbol{X4}$ and $2\,\boldsymbol{X4}$; respectively. This means that the water is staggered into n unit hydrograph inputs for $UH1$ and m inputs for $UH2$. The ordinates of both unit hydrographs are derived from the corresponding S-curves (cumulative proportion of the input with time) denoted by

$SH1$ and $SH2$, respectively (Perrin et al., 2003). $SH1$ curve is defined along time t by:

$$SH1_t = \begin{cases} 0 & \text{if} \quad t \leq 0 \\ \left(\frac{t}{\boldsymbol{X4}}\right)^{\frac{5}{2}} & \text{if} \quad 0 < t < \boldsymbol{X4} \\ 1 & \text{if} \quad t \geq \boldsymbol{X4} \end{cases} \tag{3.37}$$

while $SH2$ is similarly defined by:

$$SH2_t = \begin{cases} 0 & \text{if} \quad t \leq 0 \\ \frac{1}{2}\left(\frac{t}{\boldsymbol{X4}}\right)^{\frac{5}{2}} & \text{if} \quad 0 < t < \boldsymbol{X4} \\ 1 - \frac{1}{2}\left(2 - \frac{t}{\boldsymbol{X4}}\right)^{\frac{5}{2}} & \text{if} \quad \boldsymbol{X4} < t < 2\,\boldsymbol{X4} \\ 1 & \text{if} \quad t \geq 2\,\boldsymbol{X4} \end{cases} \tag{3.38}$$

and the unit hydrographs are finally computed following:

$$UH1_j = SH1_j - SH1_{j-1} \tag{3.39}$$

$$UH2_j = SH2_j - SH2_{j-1} \tag{3.40}$$

where j is an integer. Therefore, at each time step t, the output of the two hydrographs are respectively:

$$Q9_t = 0.9 \sum_{k=1}^{n} UH1_k\, P_{r,t-k+1} \tag{3.41}$$

$$Q1_t = 0.1 \sum_{k=1}^{m} UH2_k\, P_{r,t-k+1} \tag{3.42}$$

The 60% of the hydrograph output $Q9$ enters in a non-linear production store, while the remaining 40% is routed into an exponential store. The levels in the two reservoirs, termed respectively $R1$ and $R2$, are updated at every time step, through the water balance between the inflow and the water exchange function F, which represents the interaction of the reservoirs with the groundwater:

$$R1_t = \max\Big(0, \quad R1_{t-1} + 0.6\,Q9_t + F_t\Big) \tag{3.43}$$

$$R2_t = \max\Big(0, \quad R2_{t-1} + 0.4\,Q9_t + F_t\Big) \tag{3.44}$$

$$\text{with} \quad F_t = \boldsymbol{X2}\left(\frac{R1_{t-1}}{\boldsymbol{X3}} - \boldsymbol{X5}\right) \tag{3.45}$$

where $\boldsymbol{X3}$ is the maximum capacity of the non-linear production store (model parameter), while $\boldsymbol{X2}$ and $\boldsymbol{X5}$ are model parameters controlling the groundwater exchange function; in particular $\boldsymbol{X2}$ is a water exchange coefficient, it can be either positive of negative, and $\boldsymbol{X2}$ is a non-dimensional threshold parameter that allow the inversion of the water exchange with groundwater. The presence of the F function allows the GR6J model to break the water balance between model input and output; for these reason, the simulation is particularly sensitive to parameter $\boldsymbol{X2}$ and $\boldsymbol{X5}$.

The outflow from the reservoirs $Q_{r1,t}$ and $Q_{r2,t}$ are then computed following Eq. 3.46 and 3.48 and the reservoir levels are updated again at the same time step (Eq. 3.47 and 3.49).

$$Q_{r1,t} = R1_t\left\{1 - \left[1 + \left(\frac{R1_t}{\boldsymbol{X3}}\right)^4\right]^{-\frac{1}{4}}\right\} \tag{3.46}$$

$$R1_t = R1_t - Q_{r1,t} \tag{3.47}$$

$$Q_{r2,t} = \begin{cases} R2_t + \dfrac{\boldsymbol{X6}}{e^{\frac{R2_t}{\boldsymbol{X6}}}} & \text{if} \quad \dfrac{R2_t}{\boldsymbol{X6}} > 7 \\ \boldsymbol{X6}\, e^{\frac{R2_t}{\boldsymbol{X6}}} & \text{if} \quad \dfrac{R2_t}{\boldsymbol{X6}} < -7 \\ \boldsymbol{X6}\log\left(e^{\frac{R2_t}{\boldsymbol{X6}}} + 1\right) & \text{otherwise} \end{cases} \tag{3.48}$$

$$R2_t = R2_t - Q_{r2,t} \tag{3.49}$$

where $\boldsymbol{X6}$ is a model parameter related to the maximum capacity of the exponential store. It is important to underline that reservoir level cannot exceed its maximum value.

The output $Q1$ of $UH2$ contributes directly to the model output but, similarly to the volumes of the routing stores, it is first subjected to the same groundwater exchange function F (Eq. 3.50. Finally, the total streamflow Q is obtained by Eq. 3.51.

$$Q_{d,t} = \max\left(0, \;\; Q1_t + F_f\right) \tag{3.50}$$

$$Q_t = Q_{r1,t} - Q_{r2,t} + Q_{d,t} \tag{3.51}$$

Table 3.2: CemaNeige-GR6J model parameters.

Parameter	Description	Units
$\boldsymbol{\theta_{G1}}$	Snowmelt (degree-day) factor	mm/°C/day
$\boldsymbol{\theta_{G2}}$	Col content factor	-
$\boldsymbol{X1}$	Non-linear production storage capacity	mm
$\boldsymbol{X2}$	Groundwater exchange coefficient	mm/day
$\boldsymbol{X3}$	Non-linear routing store capacity	mm
$\boldsymbol{X4}$	Time parameter for unit hydrographs routing	mm
$\boldsymbol{X5}$	Threshold parameter for water exchange with groundwater	-
$\boldsymbol{X6}$	Exponential routing store capacity	-

3.2 Calibration of rainfall-runoff models

In gauged river sections, rainfall-runoff model parameters can be calibrated minimising the differences between simulated and observed discharges. Although the procedure could be performed even manually, the availability of high computational power in modern computers allows to easily perform automatic calibrations almost with any machine.

An automatic calibration procedure is generally characterised by two main features:

- the objective function;
- the optimisation algorithm;

3.2.1 The objective function: Kling-Gupta Efficiency

The objective function is the real-valued mathematical function whose value is to be either minimised or maximised over the set of feasible alternatives during the calibration process.

For what concern the calibration of rainfall-runoff models, the choice of the objective function should be weighted on the purpose of the application. For instance, the most common objective function in hydrological modelling is the Nash-Sutcliffe efficiency (NSE, Nash and Sutcliffe, 1970), a standardised version of the Root Mean Square Error (RMSE). Many previous studies have investigated variants of these metrics (e.g. Oudin et al., 2006; Kumar et al., 2010; Pushpalatha et al., 2012; Ding et al., 2016; Garcia et al., 2017). However, MSE-based metrics have been thought to be useful in model calibration to reduce simulation errors associated with high-flow values, because these metrics typically magnify the errors in higher flows more than in the lower flows (Mizukami et al., 2019), but Gupta et al. (2009) demonstrated that use of NSE for calibration underestimates the

response variability, with a consequently tendency to generally underestimate high flows while overestimating low flows.

If interested in reproducing a specific hydrologic process of a catchment, an option is to consider as objective function a specific stramflow signature (e.g. Olden and Poff, 2003; Yilmaz et al., 2008; Westerberg et al., 2011).

In general, if the purpose of the simulation is not focused on a specific aspect of the hydrograph, but rather on the reproduction of the entire variety of streamflow regimes, an alternative is to use an objective function embedding more than a single error (or efficiency). However, it is difficult (or even impossible) to optimise all the aspects of an hydrograph: for instance, it is impossible to improve the simulation of flow variability (to improve high-flow estimates) without simultaneously affecting the mean and correlation properties of the simulation (Mizukami et al., 2019). To provide a way to achieve balanced improvement of simulated mean flow, flow variability, and daily correlation, Gupta et al. (2009) proposed the KGE (Kling-Gupta Efficiency) as an improvement of the widely used Nash-Sutcliffe efficiency, which consider three types of model errors, namely the error in the mean, the variability, and the dynamics:

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \tag{3.52}$$

where r is the Pearson product-moment correlation coefficient, α is the ratio between the standard deviations of the simulated and observed values and β is the ratio between the means of the simulated and observed values. It can be noticed that in the form of Eq. 3.52, the three components have equal importance. However, Gupta et al. (2009) do not exclude the use of scaling factors (see e.g. Guse et al., 2017) in order to assign different weights to the components (not considered in this Dissertation).

Given its multi-objective nature, and thus its ability to consider different aspects of the runoff hydrograph, KGE is increasingly gaining dominance for hydrological model calibration in recent literature (e.g. Kling et al., 2012; Hirpa et al., 2018; Pool et al., 2018; Becker et al., 2019; Liu, 2020; Quintero et al., 2020). For this reason, it was selected for the calibration of the rainfall-runoff models in the analyses of this Dissertation.

3.2.2 Optimisation algorithm: Dynamically Dimensioned Search

An optimisation algorithm is a logical-mathematical procedure for the resolution of an optimisation problem. Such problems are decision problems whose final objective consists in the minimisation or maximisation of an objective function, which depends on one or more decision variables (i.e. parameters). The problem is usually formalised into

a mathematical model. Optimisation algorithms normally work iteratively and can be based on different approaches.

One possible discern is between local and global methods: the first are able to look for the optimal solution in a limited portion of the research space and consequently the solution may strongly depend on the starting point, while the former try to avoid such issue extending the searching process to the whole parameter space. For this reason, global method are preferred for the majority of optimisation problems, including most of the hydrological ones. Global optimisers, in turn, can be divided into *deterministic* and *stochastic* approaches. For the calibration of hydrological models, stochastic techniques are preferred.

Optimization algorithms can also be divided into *exact* (or *direct*) methods, which consider the shape and the gradient vector of the objective function, and *heuristic* methods. A heuristic technique is a problem solving approach, lying between Operational Research and Artificial Intelligence, which applies concepts from social sciences, medicine and biology to optimisation problems (Colorni et al., 1996). This kind of approaches were initially developed in order to solve problems including non-derivable or non-continuous objective functions. Their main advantage is the use of probabilistic rules to guide the research process towards regions of the parameter space more likely to improve the solution. At the same time, the main disadvantage is their non-deterministic and non-exact form: there are no criteria to evaluate the quality of the obtained *sub-optimal* solutions and the degree of exploration of the parameter space. In fact, all heuristic based optimization methods can yield good-quality or near-optimal solutions promptly to solve various optimization problems, but do not guarantee convergence to local or global optimal solutions (Rardin and Uzsoy, 2001). Heuristics are often problem-dependent, namely that they are built to solve a specific problem. Starting from the 70s, increasing importance was gained by the so-called *meta-heuristic* approaches, problem-independent techniques which can suite a broad range of optimisation problems. They can be divided into two large categories: those based on a "population of solutions" or on a "single solution" which differ for the number of solutions iterated at a time (multiple in the first case, a single one in the former).

In the present work, different optimisation problems are faced: the most evident is the calibration of the hydrological models, but also the optimisation of the similarity measures for RR-models regionalisation, treated in Chapter 6.

The algorithm selected for the resolution of the issues addressed in this thesis is the *Dynamically Dimensioned Search* (DDS) algorithm, introduced by Tolson and Shoemaker (2007) for the automatic calibration of watershed simulation models. It is a meta-heuristic,

stochastic, global optimisation algorithm based on a single solution. The algorithm was designed to converge as quickly as possible to the region of global minimum of the objective function, especially in those problems requiring high computational cost. The main feature of DDS is its ability to automatically scale the search on a user-specified simulation budget (maximum number of function evaluations). To achieve this, the value of the user-specified budget is integrated into the calculation of the probability of perturbing each search dimension (parameter). Therefore, at the beginning of the optimization, DDS perturbs a large number of model parameters (search dimensions) and, as the optimization progresses, it gradually reduces the number of dimensions to be searched to finally

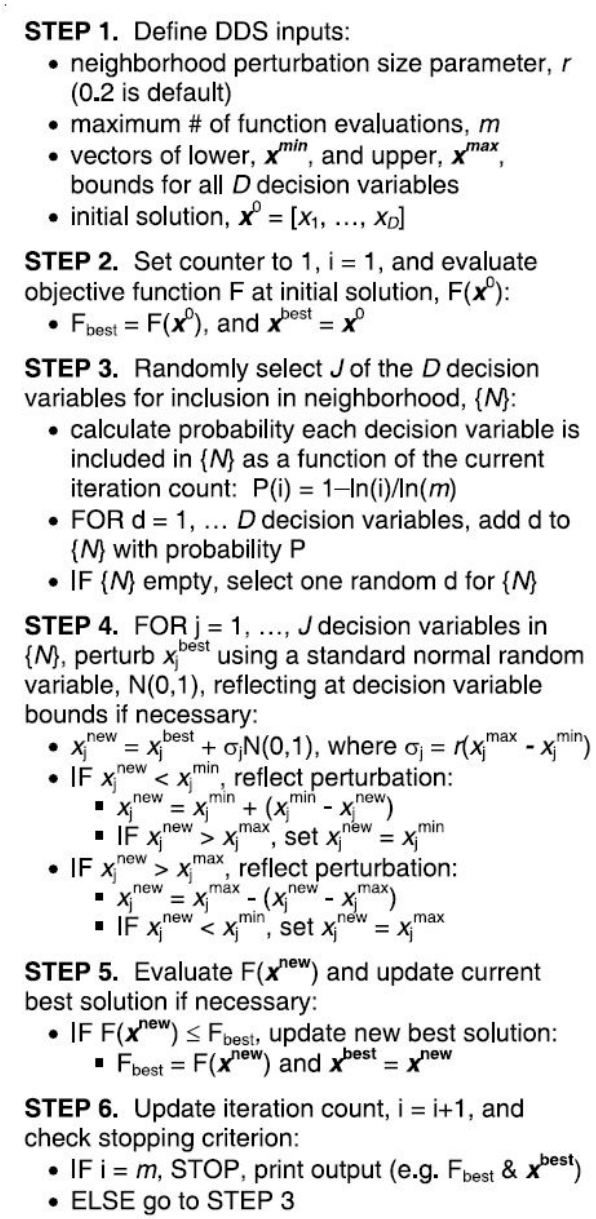

Figure 3.5: Framework of the Dynamically Dimensioned Search (DDS) algorithm (Tolson and Shoemaker, 2007).

perturb only one dimension at a time (Huot et al., 2019). Thus, the maximum number of function evaluations (m) represents the unique stopping condition and it is an input to the algorithm, as well as the initial parameter set from which the optimisation starts ($\boldsymbol{x^0}$). The only algorithm parameter is the scalar neighbourhood size perturbation parameter (r) which is recommended to be set to 0.2, as done for all the experiments of this Thesis. Fig. 3.5 reports a framework of the main algorithm steps, while for a detailed description of the DDS routines, readers are invited to refer to Tolson and Shoemaker (2007). The authors tested DDS for computationally expensive optimization problems and demonstrated how its performances outperformed those obtained with the noted and consolidated *shuffled complex evolution* (SCE) algorithm (Duan et al., 1993): DDS requires considerably less model evaluations (sometimes up to 15-20%) than SCE in order to find equally good values of the objective function, showing that DDS rapidly converges to good calibration solutions avoiding poor local optima. Similarly, following studies (Razavi et al., 2010; Arsenault et al., 2014, e.g.) confirmed its ability to efficiently calibrate hydrological models, quickly targeting good-quality solutions. For these reasons, after a few preliminary tests, the algorithm was chosen for the purposes of this work.

The algorithm was implemented through the R Programming Environment (R Core Team, 2019) using the *dds* function available in the *mcu* R package, developed by David Kneis and distributed through the GitHub repository (https://github.com/dkneis/mcu).

3.3 Parameter regionalisation methods

It has already been discussed as the set-up of any rainfall-runoff model in a river section where no discharge measures are available (ungauged catchment) always requires the estimate of model parameters based on the values calibrated in hydrologically similar (gauged) watersheds. Such gauged catchments are usually called *donor catchments* and the process of tranferring of model parameters from gauged to ungauged catchments is called *parameter regionalisation*. Chapter 1 introduced the main differences between such regionalisation approaches, dividing them into two big classes: regression-based and distance-based methods.

The following sections aims to focus on the techniques used in one or more of the experiments conducted for this dissertation, all belonging to the distance-based group of approaches, since recent studies have demonstrated that they are generally to be preferred to regression-based techniques (see e.g. Kokkonen et al., 2003; Merz and Blöschl, 2004; Oudin et al., 2008; Reichl et al., 2009; Bao et al., 2012; Steinschneider et al., 2015; Yang et al., 2018; Cislaghi et al., 2019).

3.3.1 Ordinary Kriging (KR)

Belonging to parameter-averaging sub-group, Ordinary Kriging (termed in the following KR) is one of the simplest geostatistical approaches and it is the most widely used kriging method. It based exclusively on the spatial location of the catchments, thus assuming as similarity measure the only spatial proximity. Each model parameter is regionalised independently of each other, based on their spatial correlation; therefore, the correlation of the parameter set is not maintained. Like all kriging approaches, KR produces predictions of hydrological variables of interest at ungauged sites with a linear interpolation of the empirical information (in this case of model parameters) collected at neighbouring gauging catchments. Through this method, the unknown value of a model parameter at prediction location U (i.e. the ungauged catchment), $p(U)$, can be estimated as a weighted average of the regionalised parameter values, calibrated for the donor catchments i within the neighbourhood:

$$p(U) = \sum_{i=1}^{n} \lambda_i \, p(i) \tag{3.53}$$

where λ_i is the kriging weight for the empirical value $p(D_i)$ at donor i, and n is the number of neighbouring donor basins used for interpolation. Kriging weights can be found by solving the typical ordinary kriging linear system (see Eq. 3.54a) with the constraint of unbiased estimation (see Eq. 3.54b):

$$\sum_{j=1}^{n} \gamma_{i,j} \lambda_j + \theta = \gamma_{U,i} \qquad i = 1, \ldots, n \tag{3.54a}$$

$$\sum_{j=1}^{n} \lambda_j = 1 \tag{3.54b}$$

where θ is the Lagrange parameter, $\gamma_{i,j}$ is the semi-variance between catchments i and j and $\gamma_{U,i}$ between catchments i and the ungauged location U (Isaaks and Srivastava, 1990). The semivariance, or variogram, represents the spatial variability of the regionalised variable p. The single realisations $\gamma_{i,j}$ that are produced from the sample observations of $p(i)$ cannot be considered to represent the system: the sample may produce a matrix that is singular or not positive definite, conditions required for solution of the system. Furthermore, the elements $\gamma_{U,i}$, by nature, are unobservable as the value of the dependent variable at the ungauged location, $p(U)$, is what is being estimated. However, with additional assumptions of stationarity, the semivariance can be modeled as a function of separation distance between catchments. Several classical models are available to ensure positive definiteness. These models are parameterised by calibration to the empirical variogram of

observed semivariance as a function of distance (Farmer, 2016). Once a variogram model is selected, the quantities $\gamma_{i,j}$ and $\gamma_{U,i}$ can be estimated and the system of Eq. 3.54 is solvable.

3.3.2 Nearest Neighbour (1 donor, NN-1)

The *Nearest Neighbour* method (NN-1) is a very simple approach in which the entire set of model parameters is transposed from the geographically nearest donor catchment.

3.3.3 Most Similar (1 donor, MS-1)

The third technique, termed *Most Similar* approach (MS-1), a single donor catchment is again identified, for transposing the entire parameter set. Instead of choosing the catchment that is geographically the closest, the "hydrologically most similar" donor is identified, based on a set of geomorphological and climatic descriptors implemented in the dissimilarity index introduced in Eq. 1.2. As stressed in Sec. 1.2, a set of optimal descriptors has to be identified through preliminary tests.

3.3.4 Output-averaging version of NN and MS techniques (NN-OA and MS-OA)

Evidences from the literature, reported in Sec. 1.3.2, highlighted the added value of the use of multiple donors from which transferring model parameters, exploiting the available information more than one single "sibling" catchment. It was also argue that this can be done both averaging single parameters independently or averaging the simulation obtained by simulating the model with each of the entire donor parameter set (first introduced by McIntyre et al., 2005).

Since the second option is demonstrated to outperform the former, in this Dissertation Nearest Neighbour (NN) and Most Similar (MS) techniques will be implemented also with such output-averaging approach and will be named respectively NN-OA and MS-OA. We would like to recall that output averaging (but, in general, all multi-donor apporaches) can be applied either identifying a set of best n donors or fixing a maximum distance-threshold and considering all the gauged catchments within the limit. In this case the former solution is adopted: n donor catchments are identified based on their spatial proximity (for the Nearest Neighbour method) or on their similarity (for the Most Similar method) to the target. The regionalised streamflow for the ungauged catchment is calculated from all the simulations $Q(t, P_i)$, obtained by running the model (fed by the meteorological input of the target catchment) with each one of the n parameter sets (P_i, with $i = 1, \ldots, n$)

corresponding to each of the donor catchments. Streamflow for time t, $Q_U(t)$, is computed as the weighted average of the simulated outputs (Fig 3.6):

$$Q_U(t) = \sum_{i=1}^{n} w_{i,U}\, Q(t, P_i) \tag{3.55}$$

where $w_{i,U}$ is the weight associated with each donor catchment D_i, computed as a function of a measure of dissimilarity between the donor and the target catchments. Such versions of the methods are here termed NN-OA and MS-OA. In the NN-OA case, the dissimilarity is defined by the spatial distance $d_{geo}(D_i, U)$ (Eq. 1.1) between the centroids of donor D_i and target catchment U (Eq. 3.56), while in the MS-OA method it corresponds to the dissimilarity index $\Phi(D_i, U)$ (see Eq. 1.2). The corresponding weights $w_{i,U}$ are computed accordingly to Eqs. 3.56 and 3.57, respectively.

$$w_{i,U} = \frac{\dfrac{1}{d_{geo}(D_i, U)}}{\displaystyle\sum_{i=1}^{n} \frac{1}{d_{geo}(D_i, U)}} \tag{3.56}$$

$$w_{i,U} = \frac{\dfrac{1}{\Phi(D_i, U)}}{\displaystyle\sum_{i=1}^{n} \frac{1}{\Phi(D_i, U)}} \tag{3.57}$$

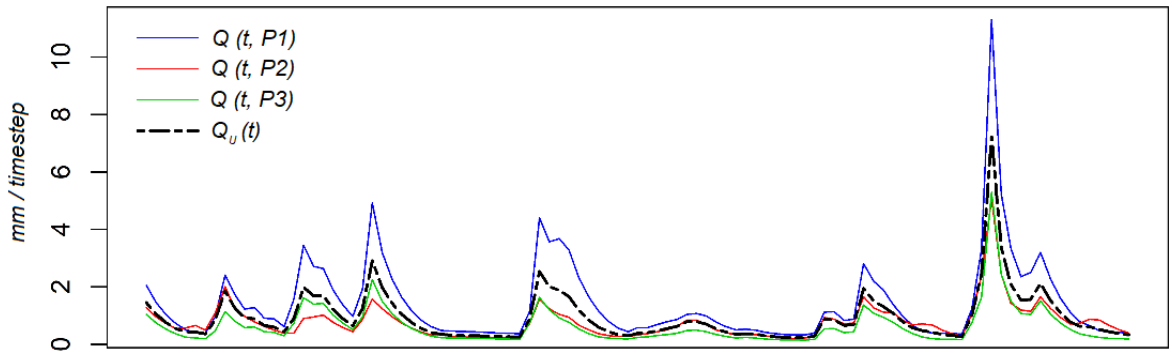

Figure 3.6: Example of output-averaging with 3 donors: the model is simulated with each of the entire donor parameter sets (coloured hydropraphs) and then the simulation are averaged (dotted hydrograph).

3.3.5 Testing a regionalisation method: leave-one-out cross-validation

In order to test the accuracy of any regionalisation approach, a validation procedure normally assumes a sub-set of the available gauged catchments (i.e. a validation set) to be ungauged and apply the regionalisation method using the hydrological information (i.e. the calibrated parameters) of the remaining portion of the dataset. In the field of parameter regionalisation the most common validation procedure, adopted here, is the so called *leave-one-out cross validation* which considers as "ungauged" sub-set each single catchment at a time: in turn, each basin is considered to be ungauged, and all the remaining catchments are available in the donors set for testing the regionalisation approach. This approach, even if computationally demanding, well simulates what would be done in a real world application; in addition, there is no randomness associated to the choice of the validation set.

Chapter 4

Comparison of regionalisation methods for different rainfall-runoff models

4.1 Introduction

This Chapter presents the analyses carried out for the set-up of a parameter regionalisation framework in the main study region of the Thesis, the Austrian country (Sec. 2.1). The two different continuous-simulation daily rainfall-runoff models presented in Sec. 3.1, the TUW model (already widely applied in the region) and the GR6J model implemented with the Cemaneige snow routine (never used so far for regionalisation in the Austrian region) are first calibrated at-site for all the study catchments. Then, the regionalisation approaches introduced in Sec. 3.3, consolidated for the study area, are applied. Their application involves a number of choices regarding regionalisation settings which will be explored in detail. Finally, the accuracy of the techniques are compared between each others and between the two models.

Research question

Which similarity measure better suits the transferability of the parameters of different rainfall-runoff models? How many donors should be used? And finally: which transfer method performs best?

4.2 Calibration of the rainfall-runoff models

The first phase of the experiment involves the parametrisation of the TUW and of the CemaNeige-GR6J models across all the 209 Austrian catchments. This section reports the detail of the calibration process.

TUW is run for all the study catchments with the semi-distributed model structure obtained by dividing them into 200-meters elevation zones. While model daily inputs (precipitation, temperature and potential evapotranspiration) and model states are defined over such zones, model parameters are assumed to be the same for the entire catchment. As introduced in Sec. 3.1.1, TUW parameters can be didved into runoff generation parameters and runoff propagation parameters: the former group could be defined differently across sub-basins, while the second does not (since model structure does not allow propagation between them, which are considered autonomous entities contributing separately to the total outlet flow. However, in this application it was decided to use a single parameter set for the entire catchment as done so far in the literature (e.g. Parajka et al., 2005; Melsen and Guse, 2019): as showed in Fig. 4.1, each basin is divided into more elevation zones over which inputs and model state variables (i.e. simulated hydrological processes) can vary (left), but all the sub-basin are ruled by the same parameter set (right).

Following the work by Parajka et al. (2005) on the same study area, 4 out of the 15 total parameters are pre-set, and 11 are calibrated: threshold temperatures $\boldsymbol{T_R}$ and $\boldsymbol{T_S}$

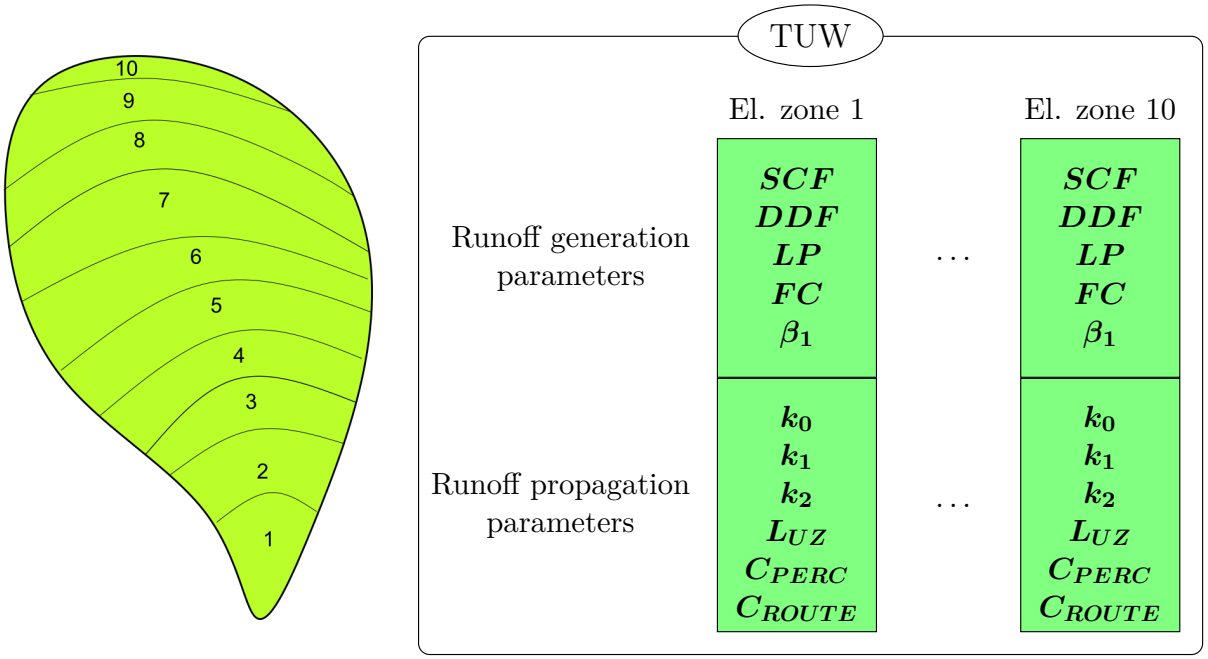

Figure 4.1: Example of the TUW parameter set for a catchment extending over ten elevation zones: parameters are the same for all the elevation zones and unique for the whole catchment.

Table 4.1: Calibrated TUW model parameters and their ranges.

Parameter	Units	Range
$\boldsymbol{SCF}$	-	0.9 - 1.5
$\boldsymbol{DDF}$	mm/°C/day	0 - 5
$\boldsymbol{LP}$	-	0 - 1
$\boldsymbol{FC}$	mm	0 - 600
$\boldsymbol{\beta}$	-	0 - 20
$\boldsymbol{k_0}$	days	0 - 2
$\boldsymbol{k_1}$	days	2 - 30
$\boldsymbol{k_0}$	days	30 - 250
$\boldsymbol{L_{UZ}}$	mm	0 - 100
$\boldsymbol{C_{PERC}}$	mm	0 - 8
$\boldsymbol{C_{ROUTE}}$	days^2/mm	0 - 50

Table 4.2: Cemaneige-GR6J model parameters and their transformed real value ranges.

Parameter	Units	Range
$\boldsymbol{\theta_{G1}}$	mm/°C/day	0 - 109
$\boldsymbol{\theta_{G2}}$	-	0 - 1
$\boldsymbol{X1}$	mm	0 - 21807
$\boldsymbol{X2}$	mm/day	-1903 - 1903
$\boldsymbol{X3}$	mm	0 - 21807
$\boldsymbol{X4}$	mm	0 - 22
$\boldsymbol{X5}$	-	0 - 1
$\boldsymbol{X6}$	-	0 21807

are fixed respectively to 2 and 0 °C, $\boldsymbol{T_M}$ to 0 °C and the maximum base of the transfer function at low flows $\boldsymbol{B_{MAX}}$ to 10 days. Tab. 4.1 presents the parameters to be calibrated and the corresponding ranges.

The CemaNeige-GR6J model is fed by mean catchment daily precipitation, air temperature and potential evapotranspiration. All the eight parameters of the combined model (2 for CemaNeige, 6 for GR6J) are calibrated. Lower and upper bounds of the parameters space are kept as default (note that the parameters are normalised in the calibration procedure). Tab. 4.2 reports the corresponding boundaries. For the sake of brevity, we will refer to this model just with the acronym GR6J, even if it will always include the CemaNeige snow module.

The sets of parameters for both rainfall-runoff models are estimated for all the study catchments with an automatic model calibration procedure, using the Dynamically Dimensioned Search algorithm (Sec. 3.2.2, Tolson and Shoemaker, 2007) and the Kling-

Gupta efficiency (Eq. 3.52, Gupta et al., 2009) between observed and simulated streamflow as objective function to be maximised. The DDS algorithm is implemented using 4000 function evaluations (set after several preliminary tests). In order to decrease the chance to run into a local minima, for each catchment the calibration procedure is repeated ten times and the set of calibrated parameters corresponding to the best KGE score is selected.

The 33 years of observation (1976-2008) are split into two sub-periods: the first one, from 1 November 1976 to 31 October 1992, is used for model calibration, and the second one, from 1 November 1991 to 31 October 2008, for model validation. Warm-up periods of one year are used in all cases.

In addition to KGE, model performances are evaluated through Nash-Sutcliffe efficiency (Eq. 4.1) as well. As mentioned in the Sec. 3.2.1, while KGE considers different types of model errors (the error in the mean, the variability and the dynamics of runoff), NSE is a standardised version of the mean square error.

$$NSE = 1 - \frac{\sum(Q_{sim} - Q_{obs})^2}{\sum(Q_{obs} - \overline{Q_{obs}})^2} \tag{4.1}$$

where Q_{sim} is the simulated runoff, Q_{obs} is the observed runoff and $\overline{Q_{obs}}$ is the average observed runoff.

4.2.1 Model performance at-site

Tab. 4.3 shows the model performances obtained by calibrating the models at-site, that is over the streamflow measured in each catchment during the calibration period (1977-1992) and validated over the years 1992-2008 (no regionalisation procedure is involved).

Both rainfall-runoff models behave well for the study area. While the median Kling-Gupta efficiencies are 0.85 for TUW and 0.88 for GR6J model in the calibration period,

Table 4.3: At-site performances: values of the 25% (1st quart.), 50% (med.) and 75% (3rd quart.) quantiles for Kling-Gupta (KGE) and Nash-Sutcliffe (NSE) efficiencies.

		KGE (-)			**NSE (-)**		
		1st quart.	**med.**	**3rd quart.**	**1st quart.**	**med.**	**3rd quart.**
TUW	**Calibration 1977-1992**	0.82	0.85	0.90	0.65	0.72	0.80
	Validation 1992-2008	0.72	0.76	0.82	0.59	0.66	0.72
GR6J	**Calibration 1977-1992**	0.86	0.88	0.91	0.72	0.77	0.81
	Validation 1992-2008	0.75	0.81	0.84	0.67	0.74	0.79

they deteriorate to 0.76 and 0.81 in the validation period, respectively. In the calibration period, KGE is always above 0.66 (TUW) and 0.76 (GRJ6). In contrast, the KGE is over 0.72 for both models for 75% of the basins (even if it drops below 0.3 for one and two basins, respectively for GR6J and TUW) in the validation period.

Looking at Nash-Sutcliffe efficiency, the difference between the two models is even more marked than for the KGE. It is interesting that despite the lower number of parameters GR6J model tends to perform better than TUW.

4.3 Regionalisation approaches

Three types of parameter regionalisation techniques are tested: KR, NN and MS (see Sec. 3.3). All belong to the distance-based group. NN and MS approaches are implemented both following a single-donor fashion (named NN-1 and MS-1) and an output-averaging framework (named NN-OA and MS-OA).

Similarly to the previous application of Merz and Blöschl (2004) and Parajka et al. (2005), Ordinary Kriging approach is based on an exponential variogram with a nugget of 10% of the observed variance, a sill equal to the variance, and a renge of 60 km both for TUW and GR6J model parameters. For both KR and NN approaches the position of the catchment is defined by the coordinates of the catchment centroid.

Next section will go deep into the choice of the attributes to estimate the similarity measure used in MS approaches to identify most suitable donors and into the choice of the number of donor catchments for NN-OA and MS-OA approaches.

For all the following analyses, the regionalisation approaches are tested through leave-one-out cross-validation (Sec. 3.3.5). The parameter sets of the donor catchments were obtained through a calibration procedure over the years 1977-1992. In contrast, for assessing the performances of the regionalisation methods, only the results obtained over the validation period (1992-2008) are reported. Spatiotemporal transfer of model parameters is, therefore, the most exacting task (as confirmed by the study of Patil and Stieglitz, 2015) since we are using parameters obtained over different catchments (in regionalisation) and over a different observation period. On the other hand, this is exactly what would happen in a real-world forecasting application or for assessing the impact of a climate change scenario, where you have to identify the parametrisation of a model to be used for independent hydro-climatic conditions and in any possible river section in the region.

4.3.1 Choice of best catchment descriptors for MS approaches

The implementation of the Most Similar approaches (MS-1 and MS-OA) requires the choice of the geo-morphologic and climatic attributes to be used for selecting the donor catchment(s), i.e. to calculate the dissimilarity indices of Eq. 1.2.

In order to individuate the best catchment descriptors (all reported in Table 1 with a brief description), the simpler Most Similar approach with one single donor catchment (MS-1) is applied sequentially to the entire dataset in leave-one-out cross-validation, using at each step an increasing number of attributes when defining the dissimilarity index Φ (Sec. 1.2); attributes weight are set equal. At each step, the method is tested multiple times, adding one by one each of the attributes and the one which gives the best regionalisation performances is selected. For greater clarity, Fig. 4.2 (panel a refers to TUW and panel b to GR6J) shows the boxplots of the consecutive best combinations of descriptors:

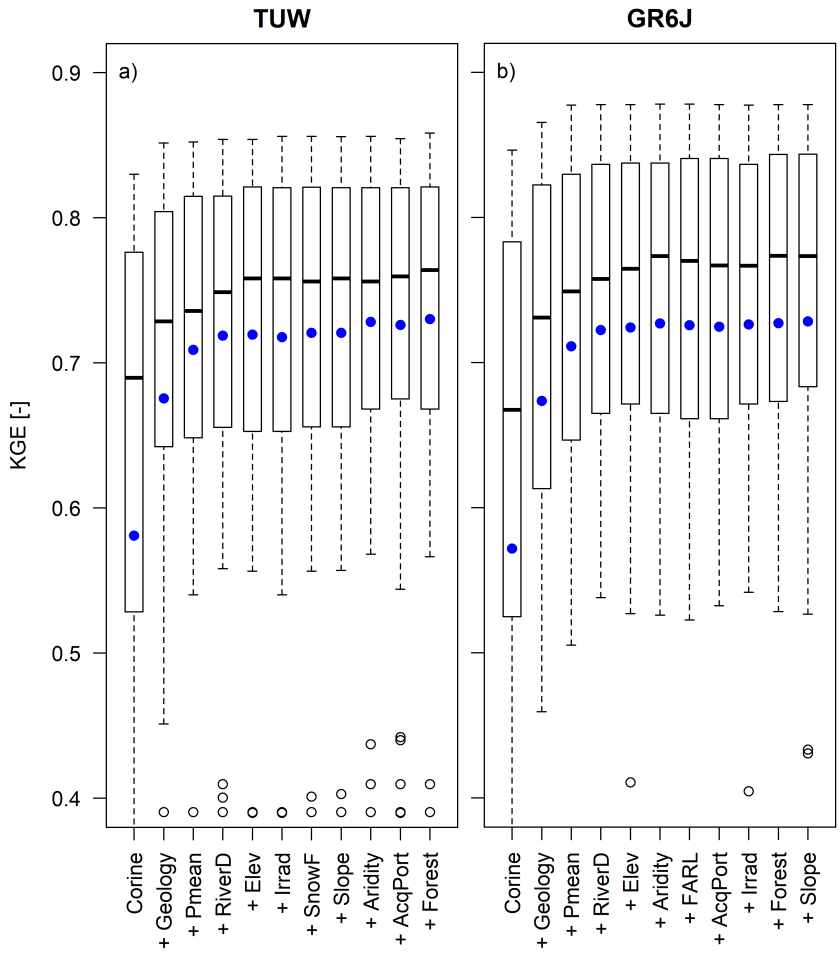

Figure 4.2: Kling-Gupta efficiencies for TUW (panel a) and GR6J (panel b) models for the consecutive steps of the similarity analysis. Boxes refer to 25% and 75% quantiles, whiskers refer to 10% and 90% quantiles and the blue points to the average.

at the first step, only one attribute is used, the Most Similar approach is tested for all the available catchment features, and the similarity in the land cover classes (Corine) gave the best efficiency. At the second step, the operation is repeated using land cover and each of the remaining attributes one at a time, finding the geology classes to be the best attribute to add, and so on. The analysis stops when the performances are decreasing or stop improving.

As can be inferred from Fig. 4.2, both rainfall-runoff models reach good regionalisation performances when using up to 5 attributes. Since the first best 5 attributes are the same for both models and from the sixth step the performances are not substantially improved, we decide to choose those five descriptors to characterise catchment similarity: land use classes, geological classes, mean annual precipitation, stream network density and mean elevation.

4.3.2 Choice of the number of donor catchments for NN-OA and MS-OA

Before comparing performances of regionalisation methods, it is necessary to choose the optimal settings for the output-averaging versions of Nearest Neighbour (NN-OA) and Most Similar (MS-OA) techniques.

The choice of the number of donor catchments represents a central issue in the methodology. Previous studies showed that the optimal number of donors is strongly related to the rainfall-runoff model and, of course, to the case study. McIntyre et al. (2005) were amongst the first to apply an ensemble (output-averaging) approach and to explore the use of different numbers of donors on the performance of the Probability Distribution Model (PDM, Moore, 1985) for a set of more than 100 UK catchments. They tested the impact of an increasing number of donors, either selecting the first n catchments with the smallest dissimilarity measure or including all the donors with a value of dissimilarity below a defined threshold (in the latter case, the number of donors may thus vary depending on the target-donors attributes). They found that a fixed number of ten donors resulted in the best regionalisation performances. Oudin et al. (2008) applied an output-averaging regionalisation for the TOPMO and GR4J models to a large French dataset of almost 1000 basins, but with no weights in flow averaging, since they used an arithmetic average (thus not taking into account magnitude of donor dissimilarities). They found that the two models performed optimally with a different number of donor catchments (seven and four respectively) and the efficiency of the regionalised model decreased almost linearly when increasing the number of donors above such values. The higher is the number of donor basins included in the regionalisation process, the more dissimilar will be the donors for

the target watershed, possibly leading to a deterioration of the results. The use of weights in flow averaging may indeed help to smooth this effect, giving less and less importance to the donors as their similarity decreases.

In the present work, the effect on regionalisation performances due to the number of donor basins is explored in detail, applying NN-OA and MS-OA for increasing number n of donor catchments: in particular, values between 1 and 50 are tested for both regionalisation techniques.

Regionalisation methods are repeated through leave-one-out cross-validation for each number of donors n and the median Kling-Gupta efficiency obtained for each value of n over all the 209 catchments is computed. Tests are performed for calibration and validation periods, but results are reported only for the validation period.

Fig. 4.3 shows the median Kling-Gupta efficiency when the changing number of donors for TUW (upper panel) and GR6J (lower panel). Looking at the figures, results show that in all the four cases, the index always deteriorates when more than 10 donors are chosen. On the other hand, there is not a unique optimal number of donors for the two models nor for the two regionalisation techniques. The optimal number of donors identified according to the median of the KGE varies between 3 and 7 depending both on the rainfall-runoff model (TUW or GRJ6) and on the regionalisation approach (NN-OA or MS-OA). Since the KGE differences between 3 and 7 donors are small (around 0.02), we decided to use 3 donors for both regionalisation methods and both models, which is also the most parsimonious option. The choice of a low number of donors is convenient also in view of the analysis to be done on decreasing informative content (Chapter 5),

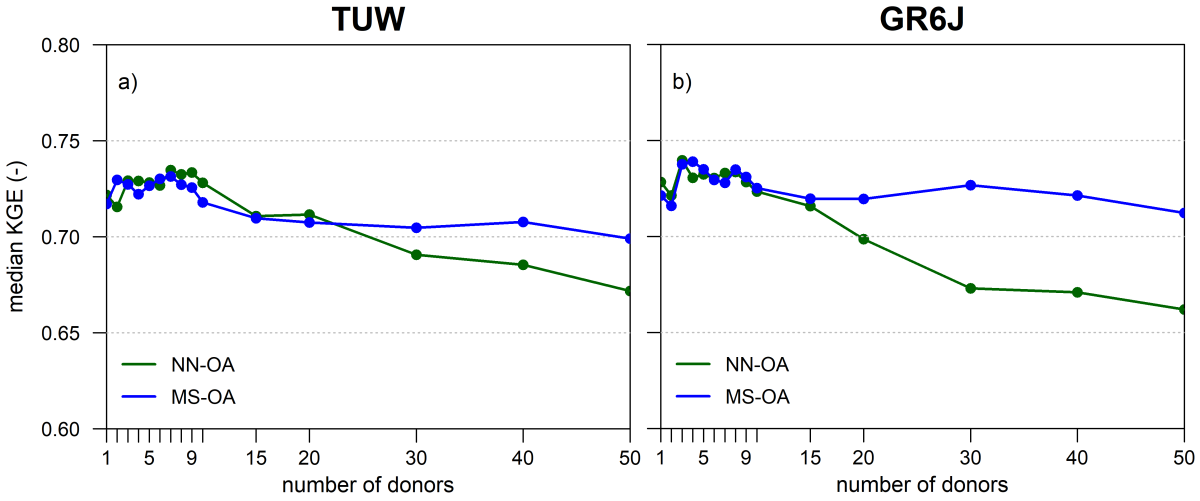

Figure 4.3: Impact of the number of donors on output-averaging Nearest Neighbour (NN-OA) and Most Similar (MS-OA) regionalisation methods for TUW (panel a) and GR6J (panel b) model.

where a large number of donors would imply the use of catchments that are less and less similar to the target one.

It may be noted that the results by Oudin et al. (2008) highlighted a clearer pattern of model performances when increasing the number of donors, with a stronger decrease in efficiency when using high numbers of donors. This result may be explained by the fact that they were using a simple not-weighted average of outputs. Here instead, the influence of the additional donors is gradually poorer, due to the weights implemented in the output-averaging procedure (Eq. 3.55). When adding further donors to the approaches, the corresponding weights in the average are gradually lower according to the increasing distance (for NN-OA) or dissimilarity index (for MS-OA) from the target. Thus, the impact of the less similar catchments is dampened, compared to what may be achieved using a not-weighted output average.

4.3.3 Comparison of the regionalisation methods

The above described regionalisation methods are tested over all the 209 study catchments through leave-one-out cross validation, for both models. Here all the basins in the dataset are used as potential donors. In turn, each basin is considered to be ungauged, and all the remaining (208) catchments are available in the donors set for testing the regionalisation approaches.

Fig. 4.4 reports Kling-Gupta and Nash-Sutcliffe efficiency boxplots for the two models when regionalising following each of the techniques. For better clarity, Tab. 4.4 shows corresponding inter-quartile values of model efficiencies.

For TUW (Fig. 4.4, upper panels), all regionalisation methods provided good simulations concerning the validation model performances obtained when the models have been calibrated on the target section (at-site simulations, white boxes). The loss in model ef-

Table 4.4: Inter-quartile values of Kling-Gupta and Nash-Sutcliffe efficiencies when regionalising TUW and GR6J models with the different techniques compared to at-site performances.

	Inter-quartile KGE (-)		**Inter-quartile NSE (-)**	
	TUW	**GR6J**	**TUW**	**GR6J**
At-site	0.72/0.82	0.75/0.84	0.59/0.72	0.67/0.79
NN-1	0.60/0.78	0.60/0.80	0.49/0.69	0.49/0.74
NN-OA	0.61/0.80	0.59/0.81	0.53/0.71	0.57/0.75
MS-1	0.61/0.78	0.60/0.81	0.48/0.69	0.50/0.74
MS-OA	0.63/0.80	0.64/0.81	0.56/0.72	0.59/0.75
KR	0.60/0.79	0.51/0.78	0.48/0.69	0.47/0.71

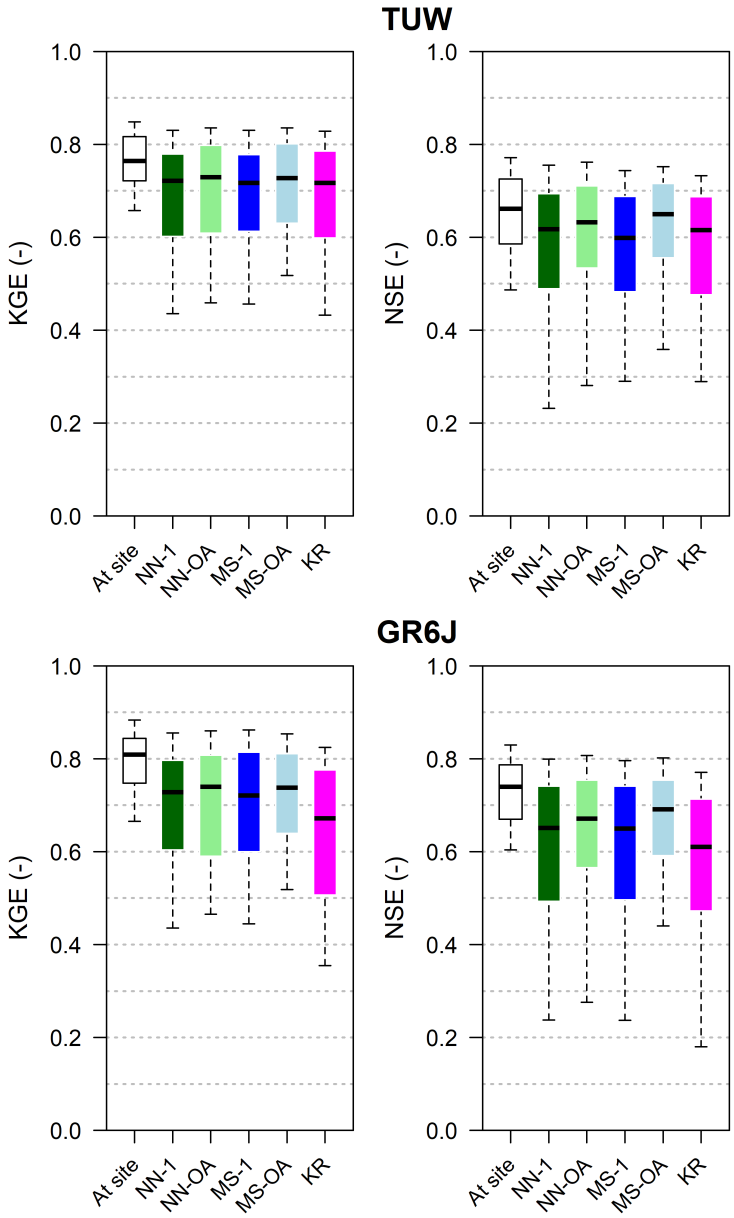

Figure 4.4: Original performances of the regionalisation methods for TUW (upper panels) and GR6J model (lower panels) for the 209 Austrian catchments in the validation period 1992-2008. Boxes extend to 25% and 75% quantiles while whiskers refer to 10% and 90% quantiles.

ficiency is, overall, small. The Nash-Sutcliffe efficiencies of KR, MS-1 and NN-1 methods are consistent with the findings of Parajka et al. (2005), who computed only the NS. Their results are very similar to the present ones, even if they worked on a greater number of Austrian catchments and calibrating the model against a different objective function.

For the GR6J model (Fig. 4.4, lower panels), the efficiencies of the Nearest Neighbour (NN-1 and NN-OA) and Most Similar (MS-1 and MS-OA) regionalisations are closer to those of the TUW in respect to what happened when the models are calibrated at-site.

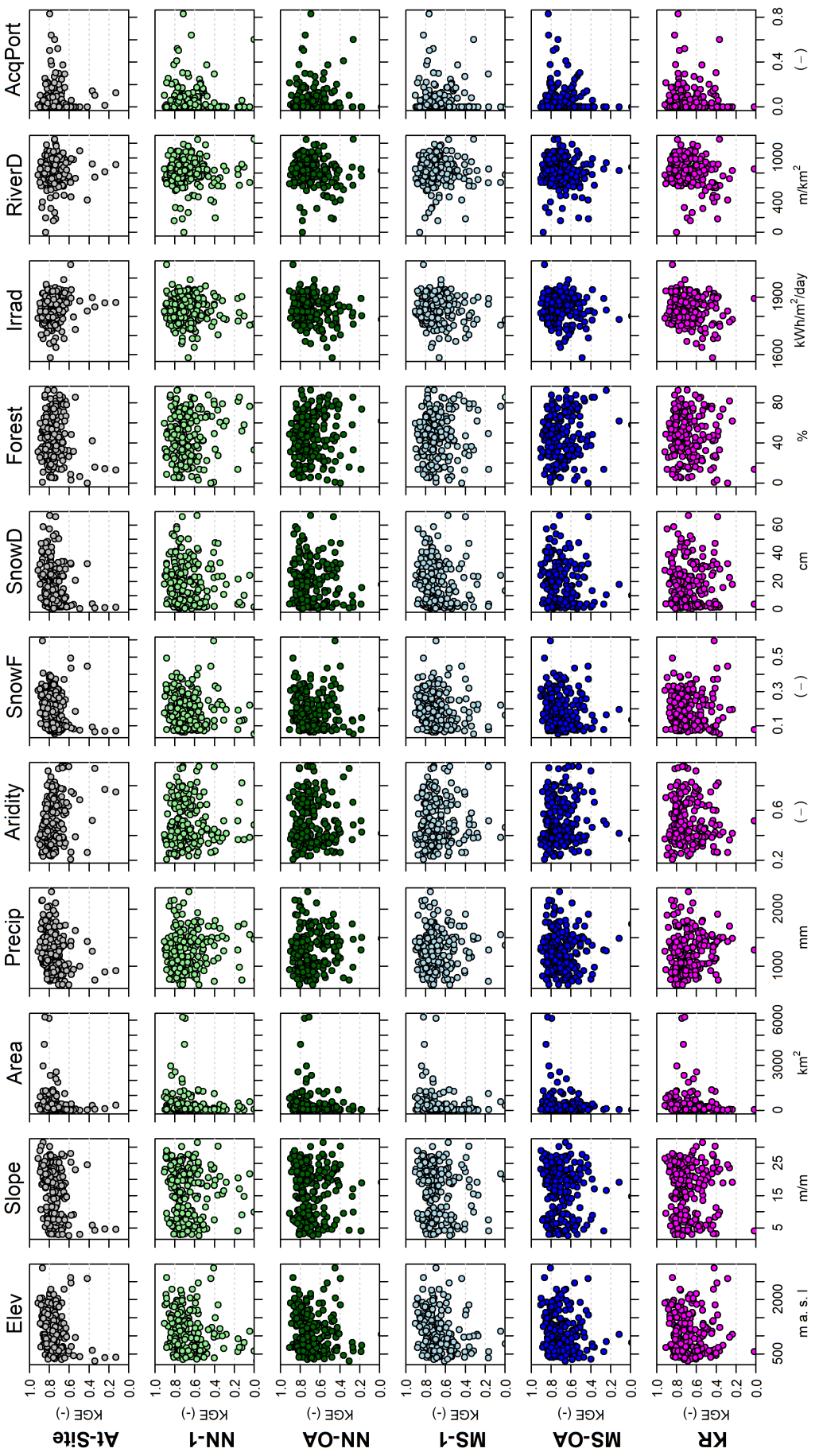

Figure 4.5: Model performance at-site (grey points, first line) and for the different regionalisation apporaches (coloured points, second to sixth line) for the TUW model against catchment attributes.

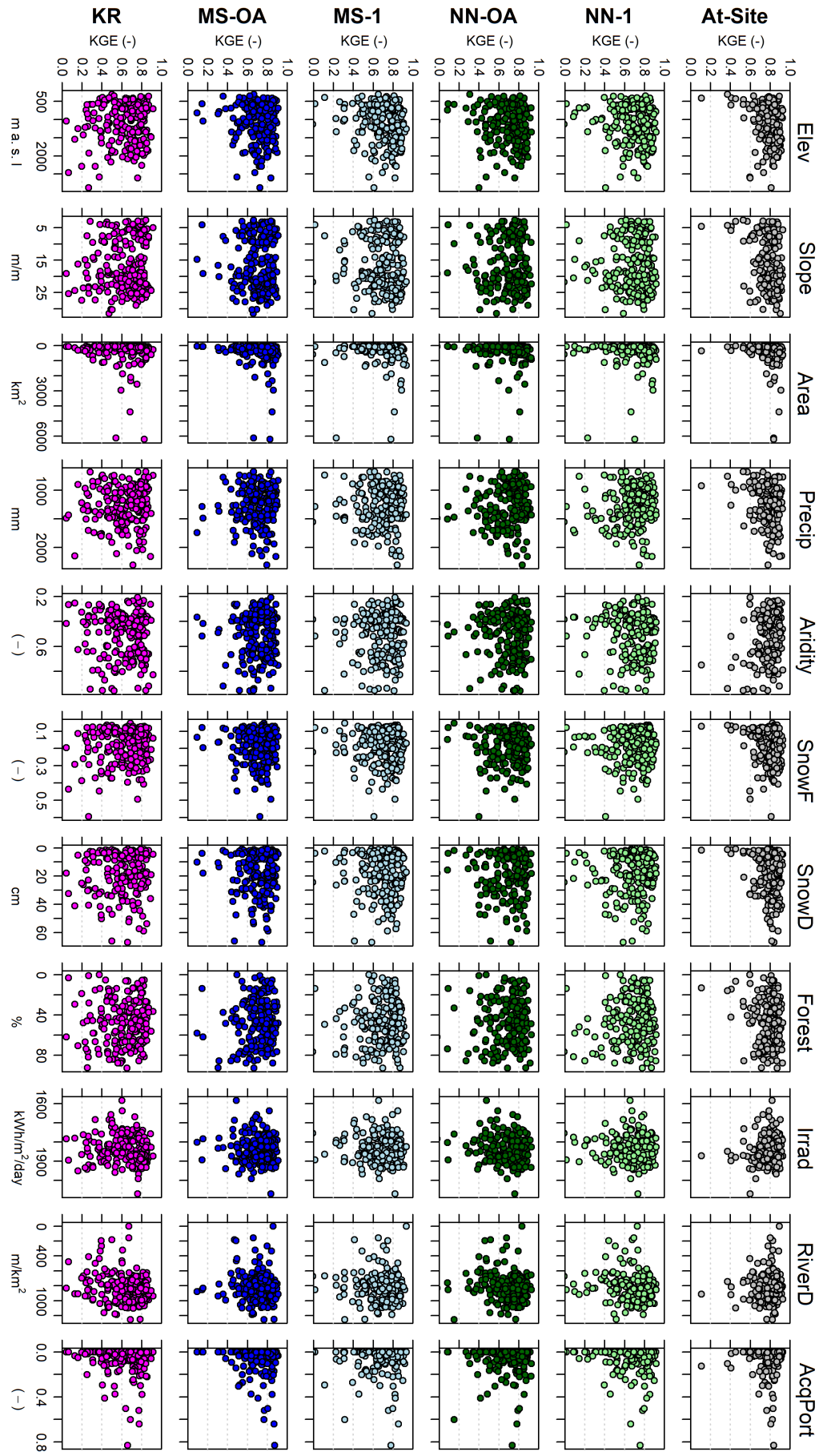

Figure 4.6: Model performance at-site (grey points, first line) and for the different regionalisation apporaches (coloured points, second to sixth line) for the GR6J model against catchment attributes.

In fact, with respect to the corresponding at-site calibration, the performances in the ungauged case (that is when parameters are regionalised) suffer a larger deterioration for GR6J than for TUW. In addition, we notice that, for GR6J model, the Ordinary Kriging has performances always poorer than all the other regionalisation methods.

For both rainfall-runoff models MS-OA tends to provide the best results and, in general, the two methods based on output average (NN-OA and MS-OA), that exploit the information from more than one donor, outperform NN-1 and MS-1, in particular in terms of Nash-Sutcliffe efficiency. It confirms the usefulness of regionalising based on more than one donor, as indicated by previous studies (e.g. McIntyre et al., 2005; Oudin et al., 2008; Viviroli et al., 2009; Zelelew and Alfredsen, 2014).

4.3.4 Model performances against catchment characteristics

In order to verify if catchment physiographic features have influence on simulation performances, the accuracy of the two rainfall-runoff models, both at-site and in regionalisation mode, were analysed against most of the catchment characteristics described in Tab. 2.1.

Figs. 4.5 and 4.6 show the scatterplots of model performances at-site (grey points in the first line) or in regionalisation mode (coloured points, second to sixth line) against basin attributes, respectively for TUW and GR6J models. No significant relationships between model performances and basin physical or climatic characteristics are observed for any of the two hydrological models. Interestingly, despite the different drainage areas of the catchments in the dataset, regionalisation accuracies do not show a clear relation with the size of the watershed, even if for some of the smaller catchments the performances were sub-optimal. This result is consistent with previous evidence from the literature (see, e.g. Parajka et al., 2013b). Moreover, probably due to the fact that the climate of Austria is basically wet or weakly arid, climate does not strongly influence model accuracy, in contrast to what observed in other studies which considered datasets covering a wider range of climatic conditions (e.g. Patil and Stieglitz, 2012; Li and Zhang, 2017). Possible links between the performance differences of the two models and basin characteristics were investigated (not showed here), but no significant relationships were found.

4.4 Discussion and summary of findings

This chapter presents a full application of a set of regionalisation techinques over the main study region of this Dissertation, setting the features of each approach and comparing their performances.

Two rainfall-runoff models for simulating daily streamflow have been first calibrated for the 209 study watersheds: a semi-distributed version of the HBV model (TUW model), and the lumped GR6J model coupled with the Cemaneige snow routine.

Both models perform very well when applied in the at-site mode, where the calibration and validation performances are very good for both rainfall-runoff models. In particular, TUW accuracy is very similar to what found by previous application of the same region (e.g. Parajka et al., 2005). Model efficiencies are somewhat larger for the GR6J model, which was applied for the first time with very good results in this Alpine dataset.

In order to assess the model performance when used in ungauged basins, the stream-gauge data for every section was, in turn, considered not to be available, and five regionalisation approaches were implemented for using the rainfall-runoff models in the validation period. This is indeed an exacting task because we are attempting to use the model over an ungauged catchment and for an observation period different from the one used for parameterising the gauged donor catchments. The first regionalisation approach is an Ordinary Kriging approach (KR), which separately interpolates each of the model parameter based on their spatial correlation in the study area. Two regionalisation approaches that select one single donor catchment and transpose its parameter set to the target basin have also been tested: in the first (NN-1) the geographically nearest catchment is selected, while in the second approach (MS-1) the single donor is the most similar one in terms of a set of physiographic and climatic attributes. The latter two approaches are implemented also in the output-averaging (OA) version, where the parameter sets of more than one donor are used for the simulation on the target section and the model outputs are then averaged accordingly to the distance/dissimilarity between donors and target.

The application of the MS based approaches involved the optimisation of the similarity measure used for identifying the donor basins, selecting the attributes to include in the dissimilarity index. The use of land cover, geology, annual precipitation, stream network density and elevation gave the best results for both models.

On the other hand, for the implementation of output-averaging based approaches (NN-OA and MS-OA), the impact of the number of donor basins for the transfer of the parameter sets was tested: results were similar when using a number of donors between three and seven, and in order to select the most parsimonious option the lower value was chosen. The pattern of model performance when increasing the number of donors is less clear than what observed by Oudin et al. (2008), but a different approaches to assign donor weights was used.

In regionalisation mode, the performances of the GR6J model deteriorates more than those of the TUW model, in comparison with the “gauged”, at-site parameterisation.

Reasons for this behaviour may lie in the different model structures and in the different transferability of model parameters (depending also on their meaning and their relation with the available catchment attributes). Such issue would deserve further attention and investigation but it would need a separate ad-hoc analysis, since the comparison of the structures and physical meaning of the parameters of the two models is not the specific objective of our work. The regionalisation accuracy of TUW across the region when using single donor methods (NN-1 and MS-1) and KR are similar, consistently to what observed in the comparative studies performed by previous studies which applied the same type of approaches to the same hydrological model structure (e.g. Merz and Blöschl, 2004; Parajka et al., 2005). For both rainfall-runoff models, the use of the output-averaging approach outperforms the use of a single donor (NN-OA and MS-OA performed better than NN-1 and MS-1), confirming the outcomes of other studies on the importance of exploiting the information available from more than only one donor (see e.g., McIntyre et al., 2005; Oudin et al., 2008; Viviroli et al., 2009; Zelelew and Alfredsen, 2014). The output-averaging methods also outperform the parameter-averaging Kriging method (especially for the GR6J model), showing that it is preferable transferring the entire parameter set of each donor, thus maintaining the correlation between the parameter values. The results of the MS-OA are close but tend to be better than those of the NN-OA, indicating that hydrological similarity is more important than geographical proximity for choosing the donors.

We expect that spatial proximity alone may be even less representative of hydrological similarity in a drier climate: Patil and Stieglitz (2012) and Li and Zhang (2017) have shown that in dry runoff-dominated regions, nearby catchments tend to exhibit less hydrological similarity than in more humid regions.

Chapter 5

Importance of the informative content of the study area: the role of nested catchments and gauging station density

5.1 Introduction

In the formulation of the research questions of Sec. 1.4, it was shown how evidence from the literature proved that a very important aspect for choosing the most adequate regionalisation technique is the informative content of the study region. This Chapter analyses how the performances of parameter regionalisation approaches are influenced by the "information richness" of the available regional data set. This first section resumes the concepts which led us to undertake this investigation.

The availability in the data set of gauged river stations representative of hydrological conditions similar to the ungauged ones plays an essential role in the assessment of the best regionalisation method. This availability can be, in some way, estimated with the station density (i.e. number of stations per km^2) and with the topological relationship between catchments.

Research question

How much do i) the presence of nested catchments and ii) station density influence the performance of methods for regionalising rainfall-runoff models? Which techniques and which similarity measures are more affected by a deterioration of the informative content of the region?

5.1.1 Influence of the nested donors

The presence of several nested catchments (i.e. gauged river sections on the same river) in the study region can strongly influence the performance of some regionalisation techniques. If for an ungauged basin model parameter sets are available for down/upstream gauged river sections, then donor and target watersheds share part of their drainage area, and thus they may also be hydrologically very similar. Such similarity may lead to very good regionalisation performances for a given approach, but may not represent the accuracy that would be obtained in different conditions. Therefore, regionalisation performances obtained for datasets with a high degree of "nestedness" may be not transferrable to study regions poor of nested basins.

So far, very few studies examined the impact of the presence of nested catchments on the performances of parameter regionalisation techniques. Merz and Blöschl (2004), Parajka et al. (2005) and Oudin et al. (2008) tested the effect of the removal of nested catchments from the available donor catchments, but only for one or two regionalisation techniques, without analysing in detail the differences between different types of approaches. Additionally, the contribution of the immediate downstream and/or upstream gauged stations has never been compared to that of the other nested catchments that share significant portions of drainage area with the ungauged one.

5.1.2 Influence of gauging station density

Also, the influence of gauging density on the regionalisation of rainfall-runoff model parameters has been little explored, with two notable exceptions. Oudin et al. (2008) applied the spatial proximity and physical similarity output-averaging techniques for decreasing values of station density in France and Lebecherel et al. (2016) tested the robustness of the spatial proximity output-averaging approach to an increasing sparse hydrometric network on the same study region. In Austria, the effect of station density has been investigated by Parajka et al. (2015), but in reference to the interpolation of streamflow time-series and not to the parameterisation of rainfall-runoff models.

5.1.3 Study aim

The purpose of the present chapter is to analyse the role of the informative content of the available regional data set, that is which and how many gauged catchments are available to be used as donors for the regionalisation in a target, ungauged section. This will be done comparing first the impact of the presence of nested donors and then the effect of the reduction of station density on the performances of the parameter regionalisation

techniques applied in the previous chapter for the very densely gauged Austrian dataset. The study is applied for the regionalisation of both TUW and GR6J rainfall-runoff models.

We believe that the present analysis may provide further insights for assessing the performances and selecting the parameter regionalisation approaches most suitable to a specific study region, keeping into account the impact of data availability, and in particular of gauging density and of the presence of nested catchments.

The analyses presented in this chapter have been recently published in Neri et al. (2020).

5.2 Impact of nested donors

As already introduced, the first purpose of the present analysis is to quantify the impact of the presence of several nested catchments on the regionalisation techniques. In particular, since nested catchments may have a strong hydrological similarity with the ungauged one, they are expected to play an essential role in the determination of method performances.

In the previous chapter, the performances of the five proposed regionalisation approaches (KR, NN-1, NN-OA, MS-1 and MS-OA) have been evaluated using all the study catchments as potential donors. Here the regionalisation procedures are repeated for each target basin (assumed to be ungauged) by excluding, from the donors set, the watersheds which are considered to be nested in relation to the target section.

5.2.1 Which catchments should be considered (to be) nested?

In general, two or more catchments are nested between each other if their closure sections are located on the same river, i.e. they share part of their drainage area. Since several gauged stations can be located on the same river, we propose to follow two different criteria to identify the nested basins:

- *Criterion 1*: the gauged sections that are immediately downstream and upstream of the target section (Fig. 5.1, panel a).
- *Criterion 2*: all the catchments sharing a given percentage of drainage area with the ungauged one (Fig. 5.1, panel b).

Catchments identified as nested by the two criteria

Following Criterion 1, 81% of the catchments in the dataset have at least one downstream or upstream nested donor (red dots in Fig. 5.2, panel a).

Instead, Criterion 2 requires the definition of a percentage threshold value of shared drainage area. A preliminary sensitivity analysis (not reported here) was performed, investigating the effect of different values between 5% and 20% for such percentage. Results show that differences in terms of regionalisation performance are not significant, and the threshold was fixed to 10%. The choice of the threshold influences the number of catchments which can be included in the study: in fact, the higher is the threshold, the lower is the number of basins classified as nested following Criterion 2. Using 10% as a threshold allows to include most of the watersheds in the analysis: 65% (137 catchments) of the basins have at least one nested donor catchment sharing at least the 10% of its area (red dots in Fig. 5.2, panel b).

All the watersheds having potential nested donors according to the second criterion have nested gauged catchments also according to the first criterion, but not vice versa. The impact of nested catchments on regionalisation performances is therefore evaluated

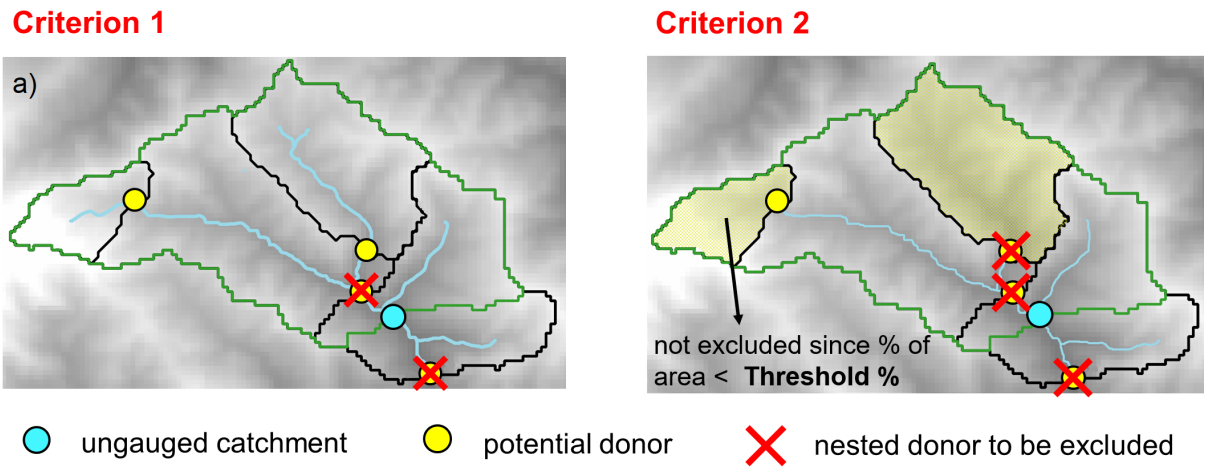

Figure 5.1: Criteria for excluding nested catchments when regionalising model parameters.

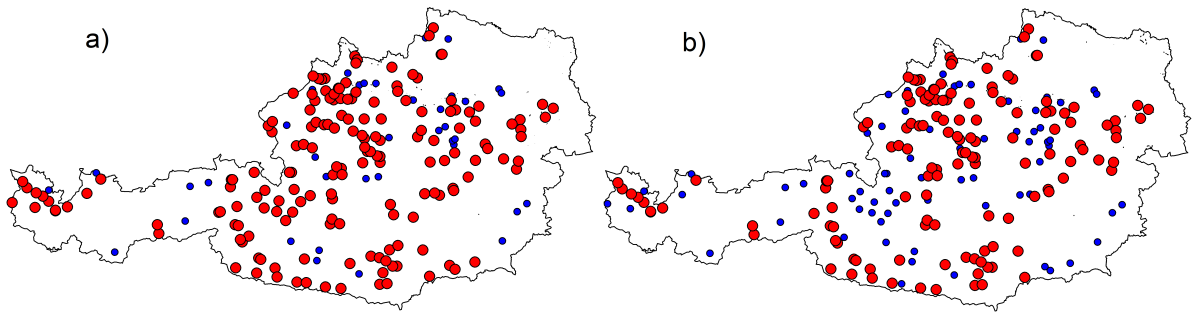

Figure 5.2: Panel a) Red dots (170) refer to catchments with at least one upstream or downstream nested gauged catchment (Criterion 1). Panel b) Red dots (137) refer to catchments with at least one nested gauged catchment sharing more than 10% of the drainage area (Criterion 2).

only for those 137 catchments that have at least one nested catchment according to both criteria. It is important to highlight that the remaining 35% of the basins are still used as potential donor catchments. The regionalisation approaches are not repeated using such basins as targets (since they have no nested donors, their performance would not change and they would distort the results).

Among the 137 catchments considered for the analysis of the nestedness, 43% have only downstream nested donor(s), 28% only upstream nested donor(s), and 29% at least one upstream and one downstream nested donors.

5.2.2 Performance losses in regionalisation when excluding nested donors

The regionalisation methods are applied again in leave-one-out cross-validation, but excluding from the available donors the catchments which are nested in relation to the target (ungauged) basin. This approach is done for both "nestedness criteria" (down/upstream or overlapping of drainage area) and the analysis applies exclusively to the 137 catchments classified as nested according to both of them (red dots in Fig. 5.2, panel b). The figures of this section (Figs. 5.3 and 5.4) therefore refer to such subset.

Fig. 5.3 compares the different performances (Kling-Gupta and Nash-Sutcliffe efficiencies in the upper and lower panels respectively) obtained in regionalisation (always over the validation period), when nested catchments are available or not as candidate donor basins for both TUW model (Fig. 5.3, upper panels) and GR6J (Fig. 5.3, lower panels). Each group of boxplots refers to a different regionalisation method: within such groups, the first box indicates the performance when no basins are excluded from the donor set, while the second and the third boxes report the performances due to the exclusion of the nested donors following Criterion 1 or 2 respectively.

The performance deterioration is highlighted by bar plots in Fig. 5.4, showing the mean loss in Kling-Gupta and Nash-Sutcliffe efficiencies when excluding nested donors following the two criteria.

Finally, Tab. 5.1 reports the interquartile variability of Kling-Gupta and Nash-Sutcliffe efficiencies for both models and all the regionalisation approaches when nested donors are excluded or not.

The less affected method is the Ordinary Kriging, especially for the TUW model. It is because the Ordinary Kriging is not based on the identification of one or more "sibling" donors which may have been excluded if nested. On the other hand, it should also be highlighted that such a method is the regionalisation approach that performs worst when nested basins are available.

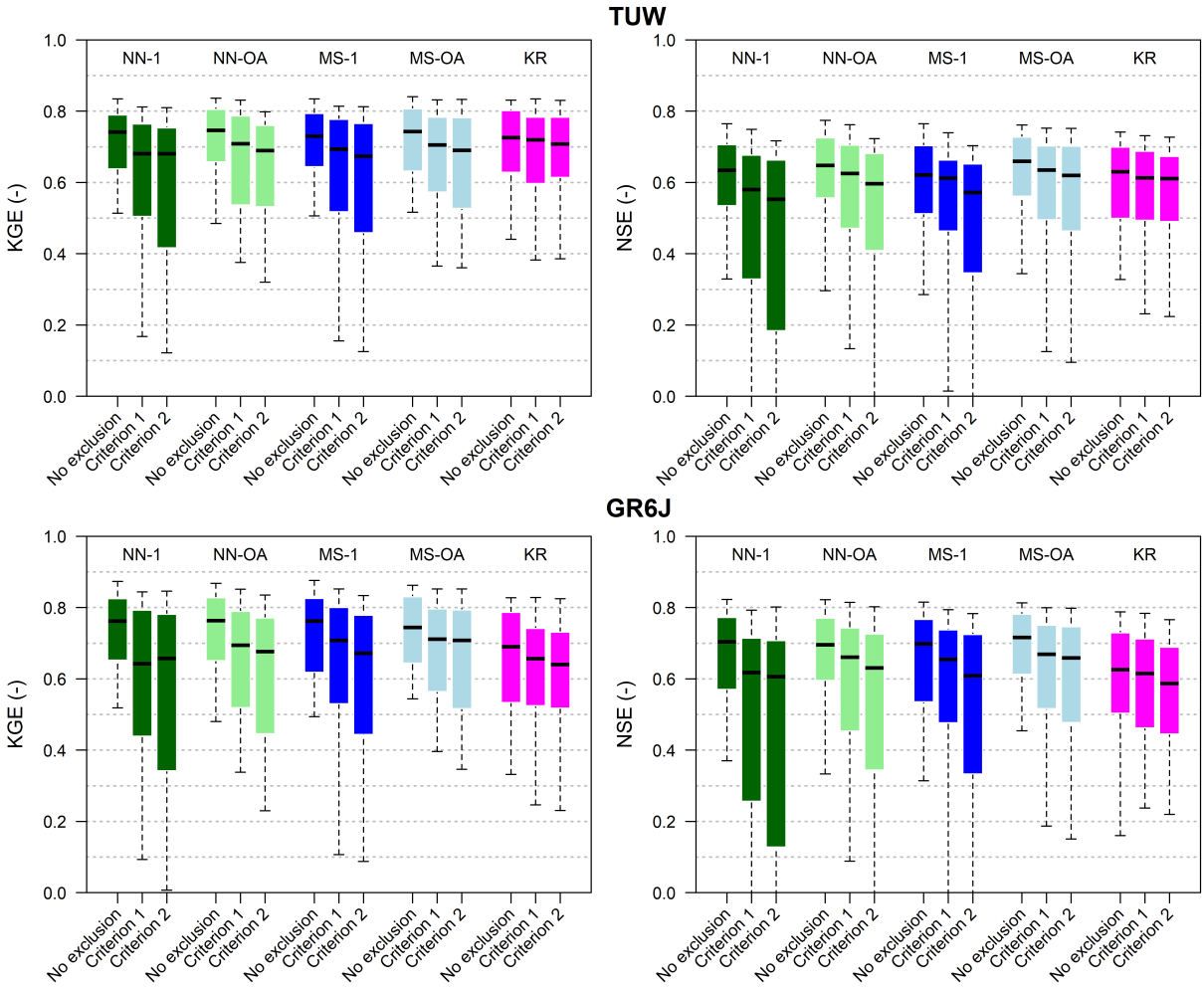

Figure 5.3: Effect of the exclusion of nested catchments for the subset of 137 watersheds classified as nested: Kling-Gupta (left panels) and Nash-Sutcliffe (right panels) efficiencies when regionalising the TUW (upper panels) and GR6J (lower panels) models. "No exclusion": all the donors are available. "Criterion 1" or "Criterion 2": nested catchments are excluded from donor set. Box colours refer to the different methods. Boxes extend to 25% and 75% quantiles while whiskers refer to 10% and 90% quantiles.

As expected, for both TUW and GR6J, NN-1 is always the most heavily affected method (dark green bars in bottom panels of Fig. 5.4). This is likely because the nearest donor is a nested one in more than 80% of the catchments for both criteria and its exclusion seriously compromise the performance.

Excluding the nested catchments also has a strong impact on MS-1 (dark blue bars in bottom panels of Fig. 5.4), even if to a lesser extent than for NN-1, since for more than 60% of the catchments the most similar donor is a nested one according to both criteria.

The degradation of performance moving from Criterion 1 (upstream/downstream) to Criterion 2 (overlapping drainage area) highlighted in Fig. 5.3 demonstrates that considering as donors not only the immediate downstream or upstream gauged river sections

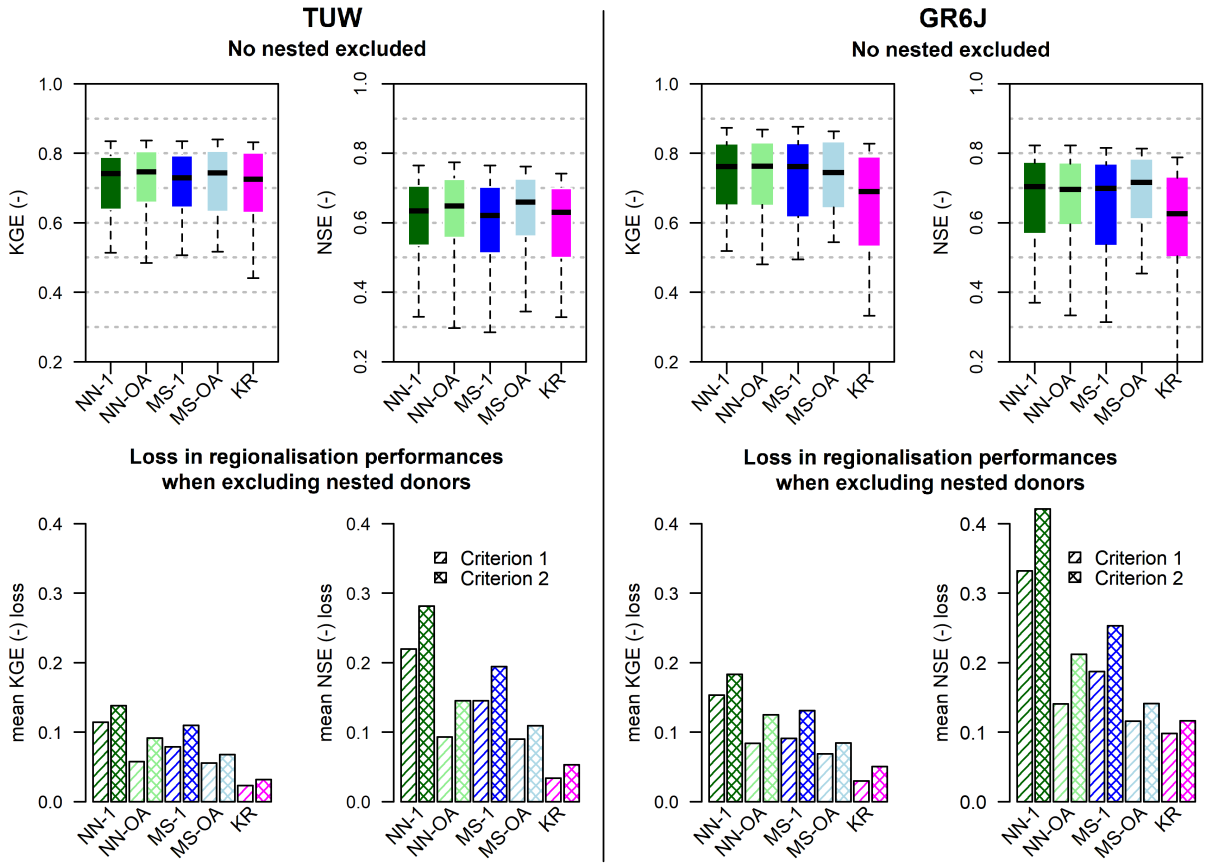

Figure 5.4: Kling-Gupta and Nash-Sutcliffe efficiencies and mean losses in the same methods resulting when excluding the nested donors with Criterion 1 and 2 (bottom panels) for TUW and GR6J models.

(Criterion 1), but also all the catchments partially sharing their drainage area with the target one (Criterion 2), has a strong positive influence on the regionalisation performance.

Furthermore, the use of output-averaging for both Nearest Neighbour and Most Similar approaches (NN-OA and MS-OA) not only outperforms the NN-1 and MS-1 when using all (nested and non-nested) donors (see also Sec. 4.3.3), but it also improves the robustness of the methods when the nested donors are excluded. The bottom panels of Fig. 5.4 show that the loss in the efficiencies of NN-OA and MS-OA are always smaller than those corresponding to the single donor approaches (NN-1 and MS-1), for both rainfall-runoff models and regionalisation methods. This confirms that the use of output-averaging and the use of more than one donor basin is preferable for regionalisation purposes also for regions that do not have so many nested catchments as the Austrian study area.

Finally, the values reported in Tab. 5.1 (as well as Figs. 5.4) show how, especially for NSE, the losses resulting when excluding nested donors from the regionalisation are higher for the GR6J model than for the TUW. The GR6J seems to be slightly more affected by the presence of nested basins, except for MS-1 and MS-OA whose performances remain

Table 5.1: Inter-quartile values of Kling-Gupta and Nash-Sutcliffe efficiencies when regionalising TUW and GR6J models excluding or not excluding nested donor catchments.

		Inter-quartile KGE (-)				
		NN-1	**NN-OA**	**MS-1**	**MS-OA**	**KR**
TUW	**No nested excluded**	0.64/0.79	0.66/0.81	0.64/0.79	0.63/0.81	0.63/0.80
	Criterion 1	0.50/0.76	0.54/0.79	0.52/0.78	0.57/0.78	0.60/0.78
	Criterion 2	0.42/0.75	0.53/0.76	0.46/0.77	0.53/0.78	0.61/0.78
GR6J	**No nested excluded**	0.65/0.82	0.65/0.83	0.62/0.83	0.64/0.83	0.53/0.79
	Criterion 1	0.44/0.79	0.52/0.79	0.53/0.80	0.56/0.80	0.52/0.74
	Criterion 2	0.34/0.78	0.45/0.77	0.44/0.78	0.52/0.79	0.52/0.73
		Inter-quartile NSE (-)				
		NN-1	**NN-OA**	**MS-1**	**MS-OA**	**KR**
TUW	**No nested excluded**	0.53/0.71	0.56/0.73	0.51/0.70	0.56/0.73	0.50/0.70
	Criterion 1	0.33/0.68	0.47/0.70	0.46/0.66	0.50/0.70	0.49/0.69
	Criterion 2	0.18/0.66	0.41/0.68	0.35/0.65	0.46/0.70	0.49/0.67
GR6J	**No nested excluded**	0.57/0.77	0.60/0.77	0.54/0.77	0.61/0.78	0.50/0.73
	Criterion 1	0.26/0.71	0.45/0.74	0.48/0.74	0.52/0.75	0.46/0.71
	Criterion 2	0.13/0.71	0.34/0.73	0.33/0.72	0.48/0.75	0.45/0.69

more similar to those of TUW. It may be due to the different structure and parameter transferability of the models, which would indeed deserve a dedicated study.

5.3 Impact of station density

Another way to evaluate the performances of regionalisation methods taking into account the richness in hydrometric information of the study area is to analyse the spatial density of the potential donors.

It is expected that the effect of the presence of several nested watersheds in a dataset is related to the effect due to station density. Because of that, the further purpose of the study is to analyse the impact of station density on regionalisation accuracy. Parajka et al. (2015) tested the impact of the station density for the direct weighted interpolation of daily runoff time-series with the topological-kriging (or Top-kriging) approach (see Skøien et al., 2006), and found that direct interpolation is superior to hydrological model regionalisation if station density exceeds 2 stations per 1000 km^2. Here, the same approach for analysing the density is applied to all the parameters regionalisation techniques.

The full station density in the dataset is about 2.4 gauges per 1000 km^2, estimated dividing the total number of stations by the area of Austrian territory, which is approximately 84000 km^2. The regionalisation approaches applied in the previous chapter are

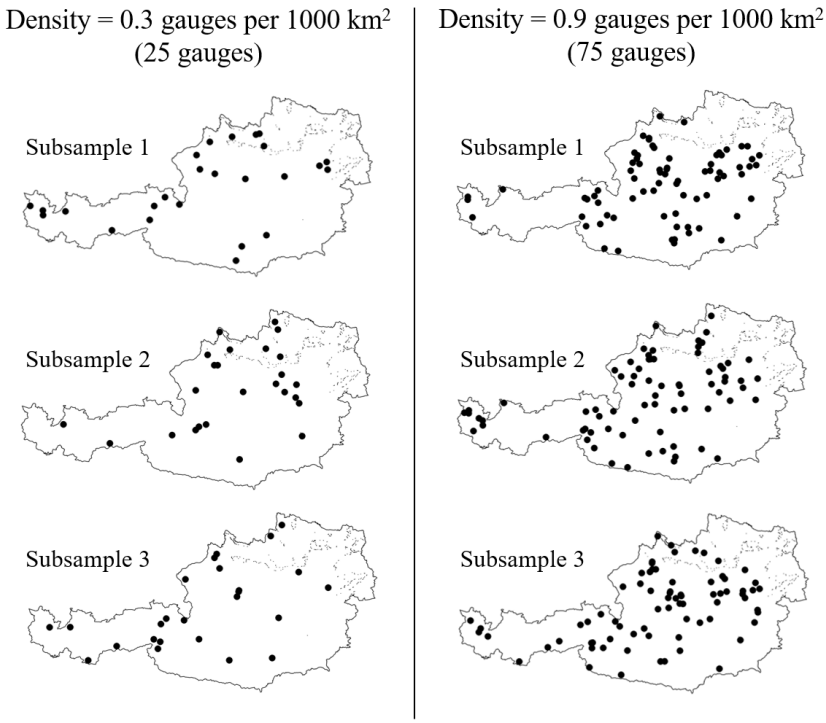

Figure 5.5: Example of three samples for two different station densities.

tested for decreasing station density in the catchments dataset. Seven different values of station density (ranging from 0.3 to 2.1 gauges per 1000 km^2) are tested, which correspond to a total number of stations between 25 and 175. For each value of station density, the corresponding number of gauged stations is randomly sampled (simple automatic non-supervised sampling) from the original set of 209 catchments, and the regionalisation approaches are applied on this subsample (catchments input dataset) in leave-one-out cross-validation. In turn, each of the catchment in the subsample is considered to be ungauged, and the remaining basins are used as potential donors. This operation is repeated 100 times to consider different samples of watersheds with the same density across the study area. Fig. 5.5 shows an example of three samples for two different station densities, corresponding to 25 and 100 stations in the input dataset.

5.3.1 Distribution of the sub-samples

For each of the seven assigned density values, the described procedure provides 100 different sets of regionalised target catchments. For a given density, each of 100 subsamples is formed by the same number of target catchments, resulting in the same number of efficiencies to be analysed.

First, it is important to verify that catchment samples are evenly distributed across the country: to do so we consider the distance of each catchment from its closer potential

donor as shown in panel a of Fig. 5.6. The average of the distances (d_1, d_2, d_3, d_4, d_5) of each catchment from the closest catchment (i.e. a potential donor) in a sample can be considered as a measure of the sample spatial distribution: the higher the distance, the less dense the sample. As above said, for each density, 100 different samples are generated, so that for each density, we have 100 different values for such averages. Panel b of Fig. 5.6 shows the average "distance within sample" of the closest available donor catchment across the 100 generated sub-sets for the different values of station density (each boxplot refers to the 100 values of average distance calculated for each sub-set). The average distance from the closest donor in the original, full density dataset (grey point in the figure) is around 8.5 km. As expected, the median target/donor distance (middle black solid line in each box) increases with decreasing density. It may be noticed that also the variability of the distance, as shown by box size and whiskers, gradually increases with the reduction of station density. Still, such increase is overall modest: even for the lowest density, it is limited to +/- 18% of the median for the 80% of the samples. The fact that, on average, the distance between a target catchment and the closest gauged catchment consistently increases with decreasing density proves that the samples with lower density do not tend to cluster/concentrate the catchments in a small region, but they are evenly distributed over the country.

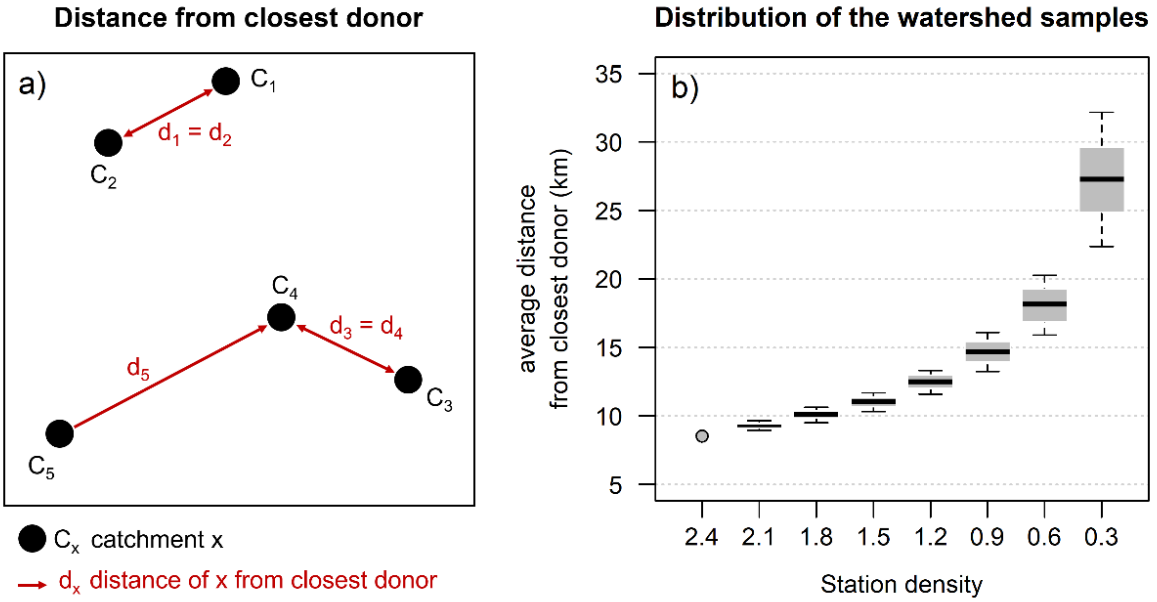

Figure 5.6: Panel a) Example of distance from the closest donor. Panel b) Boxplots of the average distance within a sample from the nearest available potential donor catchment across the 100 generated sub-sets, for different values of station density (gauges/1000km^2). Whiskers extend to 10th and 90th percentiles. The grey point indicates the average distance from the closest donor in the original dataset.

5.3.2 Performance losses in regionalisation when reducing station density

To analyse the results, the median regionalisation performances of each subsample are computed and presented here: thus, for each gauging density, the results consist of 100 values of median performances.

For the sake of brevity, only the median Kling-Gupta efficiencies over the validation periods are reported. They are shown in Fig. 5.7 for both TUW and GR6J models: each plot contains the boxplots of the median Kling-Gupta efficiencies for each station density (i.e. number of gauges per 1000 km^2), i.e. each boxplot presents the 100 values of median Kling-Gupta efficiencies obtained applying the regionalisation approaches to the 100 subsamples generated with an assigned density. The coloured point and the dotted line in the plots indicate the "original" (and maximum) median regionalisation efficiency of the approaches, that is the one obtained when using all available donors (i.e. full station density, corresponding to 2.4 gauges/1000 km^2).

The NN-1 method (Fig. 5.7, panels a and f) is the most affected by the decreasing density. In fact, when the density declines, there is a higher probability that the less dense subsamples do not include the catchment that is the nearest one to each target river section. And, as we have seen in the analyses on the nested donors, in the large majority of the cases, the nearest catchment is a nested one. In contrast, the second best may be substantially different from the target basin.

Also, the output-averaging version of the Nearest Neighbour methods (Fig. 5.7, panels b and g) strongly deteriorates for less dense networks. In general, Nearest Neighbour methods are highly sensitive to gauging density. Geographical distance results to be a good similarity measure only for densely gauged study areas (like Austria), since they firmly rely on the presence of gauged catchments in the immediate surroundings that are also hydrologically very similar. If the density decreases, the closest donor may be relatively far from the target, and it may therefore have little in common with it.

As far as the MS-1 (Fig. 5.7, panels c and h) is concerned, its performances degrade more gracefully (except for the GR6J model for the minimum density) than the NN-1 or the NN-OA. Also in this case (like for the NN-1), when the density decreases it becomes less probable that the most hydrologically similar catchment (identified by MS-1 in full density) is still part of the subsample. The results also indicate that there is more than one catchment in the original data set that is similar enough to the target in terms of catchment attributes.

This also holds true for the output–averaging MS (Fig. 5.7, panels d and i), which is even less affected by a reduction in donors' density and is the best-performing approach

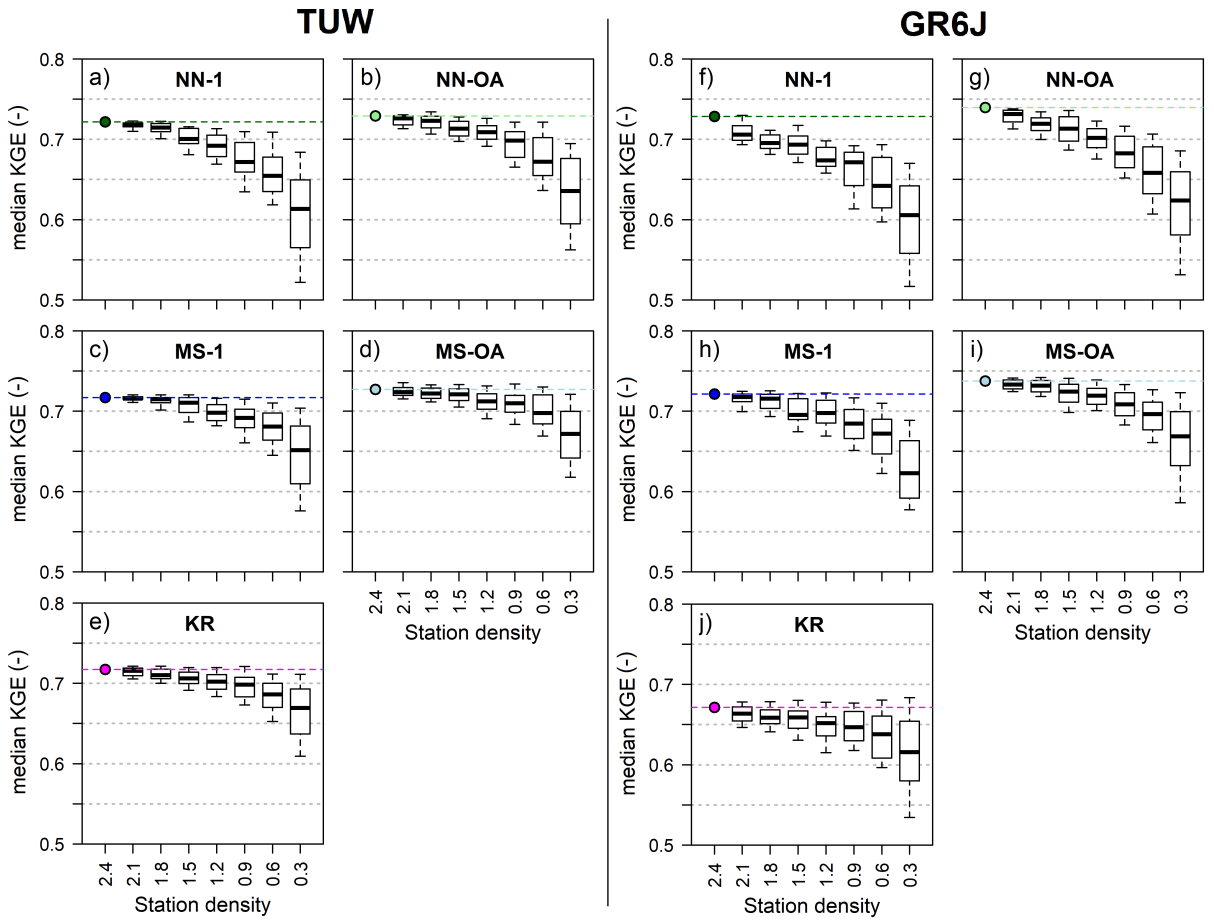

Figure 5.7: Median Kling-Gupta efficiency of the 100 sampled datasets for varying station density (number of gauges per 1000 km^2) for the TUW and GR6J models using NN-1 (panels a and f), NN-OA (panels b and g), MS-1 (panels c and h), MS-OA (panels d and i) and KR (panels e and j) regionalisation methods. The coloured point and dotted line in the plots indicate the original median regionalisation efficiency of the approaches when using all available donors (i.e. full station density, corresponding to 2.4 gauges/1000 km^2).

for any density (for both rainfall-runoff models). We may note that, also in this analysis, analogously to what resulted for the exclusion of nested catchments, for both approaches (NN and MS), the implementation of output-averaging allows to reduce the degradation in the performances in comparison to the corresponding 1-donor version.

The impact of station density is similar to that of excluding nested catchments also for the Ordinary Kriging approach (Fig. 5.7, panels e and j), which deteriorates less than the other methods for decreasing values of station density. For the TUW model, the Kriging regionalisation, starting from an already high KGE in full density, results in performances that are inferior only to those of MS-OA when the density goes below 0.9. For the GR6J model, even if the deterioration is limited since KR was poorly performing for the full density regionalisation (Fig. 4.4), the median KGE is always worse than those of all the other regionalisation approaches, for all the station densities.

Overall, all methods (excluding the poorly performing NN-1 and KR for the GR6J) result in relatively good performances provided that the station density is at least 0.9 gauges per 1000 km^2. On the other hand, leaving aside the Kriging method, the median KGE drops very steeply when the density reduces from 0.6 to 0.3 gauges per 1000 km^2.

5.4 Discussion and summary of findings

An assessment of the impacts of the presence of nested catchments and station density on the performance of parameter regionalisation techniques in a large Austrian dataset has been performed. The main motivation for this work lies in the lack of systematic studies in the literature about the effects of data-richness and informative content on the accuracy of various methods for transferring rainfall-runoff model parameters to ungauged catchments. Studies conducted on different study sets often do not lead to the same ranking of the tested approaches and the obtained results are not transferable to different study regions. This finding is indeed due to the diverse topological relationships between catchments (nestedness) in the datasets and the diverse density of the streamgauges.

The purpose of the work is to give support to the choice of the most appropriate parameter regionalisation approaches based on the available hydrometric information in the region. The study shows and quantifies how the informative content of the available gauged sections, here expressed by the presence of several nested catchments in a dataset or by the gauging density of the study region, can influence the predictive power of a certain technique.

The research has been conducted on the for a very densely gauged dataset covering a large portion of Austria, and for two rainfall-runoff models: a semi-distributed version of the HBV model (TUW model), and the lumped GR6J model coupled with the Cemaneige snow routine.

The impact of the richness of the data set (i.e. the informative content of the region) was analysed to assess the deterioration of the five regionalisation approaches, applied to the same study region in the previous chapter, for decreasing availability and "worth" of the available donors, starting from the influence of using nested basins as donors.

Two criteria have been proposed for identifying a basin that is nested with the target one. The first one, already used in the few analysis of nestedness in the literature, classifies as nested the first upstream and the first downstream gauges on the river network. The second, novel criterion, identifies as nested all the catchments that share more than a given percentage (here chosen as 10%) of the drainage area with the target one. It results that the first criterion identifies a larger number of nested catchments with at least one

potential donor. The first criterion considers as nested also a number of catchments that share less than 10% of area with the target one: this means that, in some cases, the first downstream or upstream gauge may be not representative of the same drainage area and their catchments may be governed by very different hydrological processes.

All the regionalisation approaches have been repeated by excluding from the donor set the catchments assumed to be nested with each target basin, according to each one of the two criteria.

For both rainfall-runoff models and all the regionalisation approaches, when excluding all the basins that share a significant portion of the same watershed (second criterion), the regionalisation procedure deteriorates more than when excluding the only first up-/downstream river sections: in fact, such first up/downstream catchment may, in some cases, not have much in common with the target one.

Looking at the two rainfall-models, when excluding the nested catchments, the regionalisation performances tend to deteriorates more for the GR6J than for the TUW: this seems to indicate that the TUW model may be more robust for regionalisation purposes, even when nested donors are not available.

Comparing the different regionalisation approaches, the parameter-averaging Kriging is the method that is less impacted by the exclusion of the nested donors, since it does not depend only on the choice of one or few "sibling" donors, that are very often the nested ones, but it takes into account some of the donors in a given radius. This is consistent to the outcomes of Merz and Blöschl (2004) and Parajka et al. (2005) who observed almost no deterioration of regionalisation performances when excluding the first down and upstream nested donors using the same Ordinary Kriging approach. When using, instead, a method transferring the entire parameter set from one or more donor catchments, the deterioration is more noticeable. The method that experiences the worst deterioration is the NN-1, since in 80% of the cases, the nearest basin is a nested one, and it is thus excluded from the potential donors. The second worst is the MS-1, that, when free to choose any single potential donor in the entire region, would choose a nested one in 60% of the cases. The output-averaging methods degrade less severely, showing that exploiting the information resulting from more than one donor increases the robustness of the approach also in regions that do not have so many nested catchments as in Austria (where the importance of nested donors in regionalising model parameters is highlighted also by Merz and Blöschl, 2004).

Finally, an assessment of the impact of station density on the regionalisation has also been implemented. The Nearest Neighbour approaches (both NN-1 and NN-OA) are the methods that suffer more from the decrease in gauging density. In contrast the Most

Similar methods (MS-1 and MS-OA), which use as similarity measure a set of catchment descriptors, are more capable of adapting to less dense datasets. In fact, in a more "sparse" monitoring network, the Most Similar methods are able to find other adequate donors, that may be anywhere in the region. On the other hand, the Nearest Neighbour techniques, when applied in low station density networks, risk to identify a "not so near" donor that may be very different from the target one. The impact of decreasing station density on the performance of the output-averaging approach based on spatial proximity (NN-OA) is in line with what observed by Lebecherel et al. (2016). The performances of both the output-averaging methods, in agreement with the results obtained for similar methods by Oudin et al. (2008), strongly deteriorate when the station density drops below 0.6 gauges per 1000 km^2.

The study confirms how the predictive accuracy of parameter regionalisation techniques strongly depends on the informative content of the dataset of available donor catchments, quantifying the contribution of nested catchments and station density for different approaches and rainfall-runoff models. The outcomes obtained for the Austrian data set indicate that the reliability and robustness of the regionalisation of rainfall-runoff model parameters can be improved by making use of output-averaging approaches, that use more than one donor basin but preserving the correlation structure of the parameter set. Such approaches result to be preferable for regionalisation purposes in both data-poor and data-rich regions, as demonstrated by the analyses on the degradation of the performances resulting from either removing the nested donor catchments or decreasing the gauging station density.

Chapter 6

Exploring elevation zone similarity for the semi-distributed regionalisation of model parameters

6.1 Introduction

6.1.1 Crucial effect of the altitude in runoff generation processes

In the previous chapter, it was highlighted how for transferring model parameters from gauged to ungauged catchments the combined use of i) a similarity measure based on catchment attributes rather than spatial distance and ii) an output-averaging approach is preferable for applications in both data-poor and data-rich regions.

In Sec. 1.4 we argued that similarity is usually defined at catchment scale when implementing most of lumped or simple semi-distributed models and descriptors refer to the entire basin, neglecting within-catchment heterogenity in hydrological processes. On the other hand, distributed models are able to account for hydrological processes at a finer scale and similarity may be defined between basic elements as grid cells or HRUs. The regionalisation of distributed model is addressed typically in a regression-based fashion linking basin elements to landscape characteristics (e.g. Wallner et al., 2013; Rakovec et al., 2016; Mizukami et al., 2017; Beck et al., 2020; Merz et al., 2020), optimising in a single-step approach the transfer functions; in such approaches, single parameters are regionalised independently based on calibrated relationships. However, literature has not explored in detail the use of similarity measures in a distance-based regionalisation framework for identifying or grouping donor HRUs of different basins and for the transfer of their model parameter set, especially evaluating the benefit of a higher spatial resolution

if compared to a classical lumped approach; an exception is the work of Li and Zhang (2017) who found marginal improvement in the use of grid based versions of two models in ungauged basins if compared to their lumped counterparts, but they did not differentiated similarity across space, i.e. the same similarity measure was used for all the grid elements (grid version) and for catchments (lumped version). We also highlighted that, to the best of our knowledge, no works in the literature have explored the use of semi-distributed regionalisation approaches for taking advantage of variability of hydrological processes and similarity at sub-catchment scale.

It is clear that understanding how dominant dynamics and similarity change in space is the key for improving rainfall-runoff simulations. Above cited studies have tried to describe parameters variability in space, across different regions and landscapes, but so far no studies have focused specifically on understanding how rainfall-runoff dynamics (and hydrological similarity) vary with elevation. In fact, the altitude at which precipitation occurs indirectly influences runoff generation processes through:

- the temperature gradient which is the main driver of snow dynamics and at the same time plays a central role for soil moisture dynamics, i.e. by ruling the evapotranspiration processes;
- the landscape features as vegetation and soil type which strongly change with elevation and may be essential to model the first interaction between rainfall and terrain, e.g. for estimating maximum soil moisture content and runoff release rate.

In this study we want to point the attention towards the effect of basin and sub-basin altitude on runoff generation processes, trying to understand firstly if we can gain useful information by learning how catchment similarity changes with elevation, and secondly if such information may help us to improve hydrological modelling.

The methodology tested in this chapter, involves the use of an elevation zone based semi-distributed model structure applied in a semi-distributed calibration and regionalisation framework. The model selected is the already described TUW model. Even if more sophisticated semi-distributed rainfall-runoff models were made available during the recent years by different authors, we are particularly interested in testing a well-known, and consolidated model structure as the HBV model, on which the TUW is based, in order to make the analyses as much replicable as possible. In addition, TUW (even in its semi-distributed version) is computationally not expensive, thus facilitating the several calibration procedures involved in the work to be performed for a large number of basins.

The study is applied to each of the two large case studies presented in Chapter 2: Austrian and US (CAMELS) datasets. The use of such deeply different regions with such high informative content helps to further validate the proposed analyses.

Next sections will present the ratio and the outline of the tested approach.

6.1.2 Semi-distributed rainfall-runoff modelling with TUW

As introduced in Sec. 3.1.1, TUW model is a semi-distributed version of the popular HBV model. It allows to divide the catchment into sub-basins over which the meteorological forcing are differentiated, along with rainfall-runoff transformations. However, the model does not take into account any propagation routines between catchment sub-entities: each of the sub-basin is considered as an autonomous entity which contributes separately to the total outlet flow, calculated as the average runoff from the different sub-basins according to their portion of total drainage area.

While meteorological inputs are differentiated over the elevation zones all the elevation zones belonging to the same basin are supposed to be characterised by the same dynamics and governed by the same model parameters that are therefore uniform over the entire basin. This is exactly what was done in the analyses of previous chapters, where a unique set of model parameters characterises all the elevation zones of each Austrian catchment (Fig. 4.1).

In the innovative framework proposed in this chapter, the intent is to differentiate the hydrological dynamics across elevation, i.e. allowing model parameters to assume different values at different altitudes. The purpose is to characterise changes in rainfall-runoff processes with elevation and use the obtained information to improve model simulation both in the gauged ("at-site" simulation) and ungauged (transfer of model parameters) cases. While the following sections will go into the details of the methodology, the essential aspect and steps of the semi-distributed modelling approach which will be proposed are introduced here for better clarity.

Spatial resolution of the model

The available meteorological inputs for the two selected case studies are defined over quite fine elevation zones: 200 and 100 meters bands for Austria and US respectively. This subdivision is good to account for variability in meteorological forcing but too fine to evaluate differences in hydrological similarity: such narrow zones may cover just a little portion of the catchment drainage area, not enough to evaluate macro differences in runoff generation processes. In addition, on the practical side, the number of donor "entities" (i.e. in the present case, elevation zones at the same altitude belonging to gauged catchments) must be large enough to guarantee a reasonably-sized sample for the evaluation of hydrological similarity: in fact, in order to transfer hydrological information (i.e. model parameters), a sufficient number of donors is required to determine the best

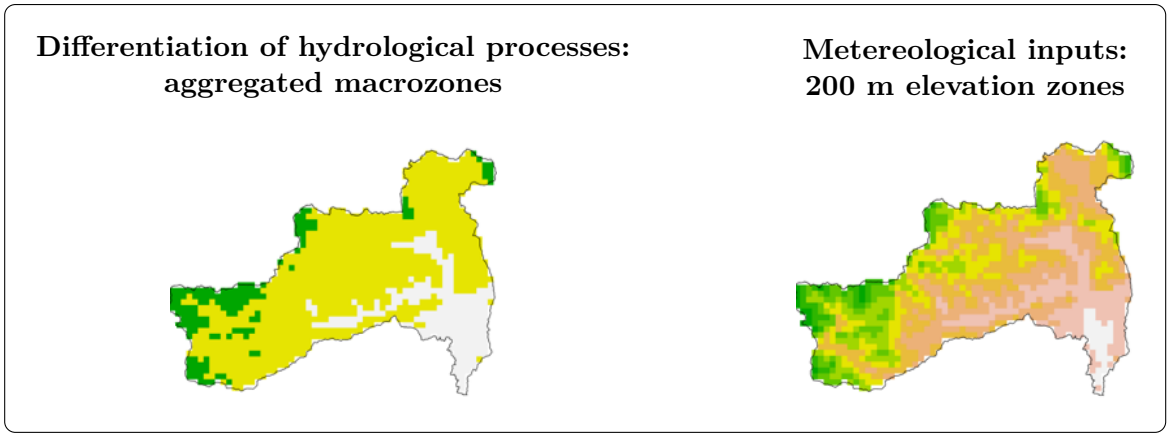

Figure 6.1: Example of the macrozones used for differentiating rainfall-runoff dynamics (left) and spatial resolution of meteorological forcings (right) for an Austrian basin.

similarity measure (and thus the most appropriate donors). For this reason, a coarser subdivision of the elevation zone ranges in the study regions is needed. As will be described in detail in Sec. 6.2, the 100 or 200 meters elevation zones are aggregated into wider bands named *macrozones*, over which the hydrological processes may be differentiated. The extension of such macrozones must be set uniformly within the same study region in order to allow the transfer of information between entities at the same elevation.

It must be clarified that the spatial resolution of the meteorological inputs is always kept as the original: in all the applications presented in this chapter (and more in general in this dissertation) rainfall-runoff forcings are always defined over the finer 100 or 200 meters elevation zones. For sake of clarity, Fig. 6.1 shows an example of the differences between the finer elevation zones of the meteorological inputs and the aggregated macrozones used for the differentiation of hydrological dynamics in the innovative framework which will be proposed.

New parametrisation approach

Sec. 3.1.1 described TUW routines and highlighted that the model is composed by a snow routine and a soil moisture routine which identify the runoff generation processes, and by a response and a routing routines, representing the propagation of the produced runoff across the basin.

It was already introduced how the proposed modelling approach involves the differentiation of model parameters across the aggregated elevation zones (i.e. macrozones) to account for diversity in rainfall-runoff dynamics. But during the model parametrisation (i.e. calibration procedure), not all the model routines will be differentiated across elevation: given the particular structure of the TUW model, which does not take into account runoff exchanges between sub-basins (macrozones), only the parameters governing the

generation of runoff will be allowed to vary across the catchments; the features of runoff propagation will be considered, again, uniform over the catchments.

In summary, the modelling approach we are proposing (which will be called *TUW-parD*) differentiates the runoff generation processes (snow and soil moisture routines) across elevation, while propagation routines will be related to the whole catchments.

Such modelling approach will be benchmarked against the standard approach, used in the previous chapters and in the literature, where hydrological dynamics are controlled by the same parameters, uniform across the elevation zones of each catchment (here named *TUW-parU*). This step of the analysis, treated in Sec. 6.3, will evaluate the potential of the proposed modelling approach for the at-site simulation.

Regionalisation and similarity: focusing on the runoff generation

As mentioned in Sec. 6.1.1, once the model is parametrised with the new approach, the analysis will be directed towards the characterisation of hydrological similarity at different elevations, focusing on the dynamics leading to the production of runoff, rather than its propagation through the catchments. This choice allows to better adapt the approach to the structure of the model and at the same time to concentrate the attention to the portion of rainfall-runoff processes that are most heavily influenced by the altitude at which precipitation occurs. A regionalisation based on the transfer of model parameter sets from gauged to ungauged macrozones based on sub-basin similarity is proposed and benchmarked against a classic approach were uniform parameters are transferred between entire basins. This analysis is presented in Sec. 6.4.

Research question

How is similarity changing across elevation? Is the proposed semi-distributed approach giving benefits to the rainfall-runoff simulations? Does a semi-distributed calibration and regionalisation approach improve simulation in ungauged catchments?

6.2 Defining the macrozones

In the previous section it was argued that the study will account for changes in hydrological processes and similarity at different altitudes considering elevation bands larger than the orginal resolution of meteorological inputs, here termed *macrozones*.

In fact, the number of donors covering each macrozone must be enough to guarantee a good sample for the evaluation of hydrological similarity (i.e. to have enough donors for transferring hydrological information).

We need, therefore, to decide how aggregating the original, fine 100 or 200 meters elevation zones in order to identify elevation entities characterised by specific hydrological conditions. A multitude of landscape characteristics is strongly related to the altitude, but the most noticeable is the vegetation. Models of vegetation zonation are complicated by multiple factors and thus the elevation ranges at which vegetation classes begin and end are generally not known a priori. However it is possible to split the altitudinal gradient into five main zones used by ecologists under varying names (Troll, 1973):

- Nival (glaciers): covered in snow throughout most of the year. Vegetation is extremely limited to only a few species.
- Alpine zone: the highest zone where vegetation typically exists. This area is shaped by the frequent frosts that restrict extensive plant colonisation.
- Sub-alpine: characterised by a closed carpet of vegetation that includes alpine meadows, shrubs and sporadic dwarfed trees.
- Montane: extends from the mid-elevation forests to the tree line. The exact level of the tree line varies with local climate, but typically the tree line is found where mean monthly soil temperatures never exceed 10.0 °C and the mean annual soil temperatures are around 6.7 °C (Nagy and Grabherr, 2009).
- Lowland (or foothill zone): this lowest section varies distinctly across climates.

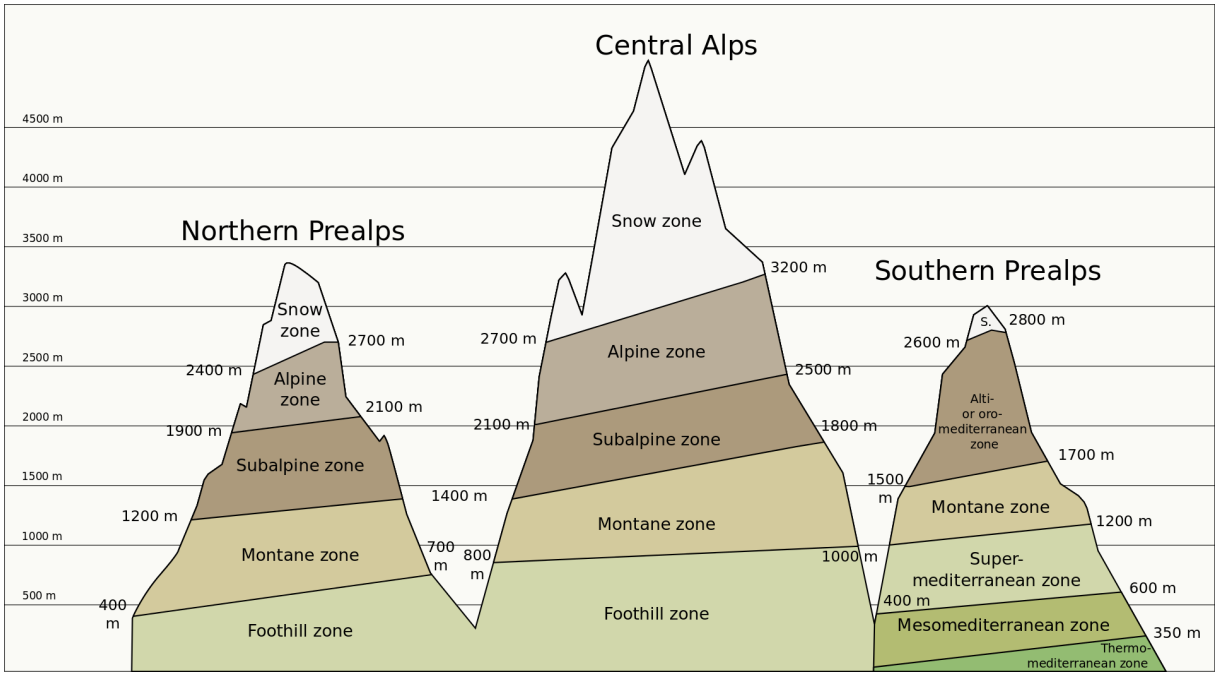

Figure 6.2: Vegetation zonation in the Alps (Altitudinal Zonation, *From Wikipedia, the free encyclopedia*)

Table 6.1: Elevation ranges of the macrozones for the two study regions. "no." refers to the no. of macrozones in each dataset, corresponding to a portion of the catchment area greater than 10%.

Country	Macrozone	Elevation range (m a.s.l.)	no.
Austria	1A	< 800	135
	2A	800 - 1400	148
	3A	1400 - 2000	99
	4A	2000 - 2400	46
	5A	> 2400	20
USA	1C	< 400	297
	2C	400 - 800	247
	3C	800 - 2000	152
	4C	> 2000	72

Tab. 6.1 shows the elevation ranges defined for the two case studies and the size of the sub-samples, i.e. the number of basins where a given elevation range covers at least 10% of the basin.

For what concerns the Austrian case study, five macrozones (named 1A,...,5A) are set accordingly to the zonation of central and northern Alps (Nagy and Grabherr, 2009, Fig. 6.2), which allows to maintain a sufficient number of sub-basin for each zone as well. Histogram in Fig. 6.3 (left) shows the number of sub-basins for each macrozone: it can be seen that macrozones are distributed quite uniformly across elevation till 2000 m a.s.l (upper range of macrozone 3A), while for the higher zones (4A and 5A) the sample is smaller, as expected. Fig. 6.3 (right) reports the number of catchments occupying one or more macrozones: just 50 basins extend over a single macrozone while more than one

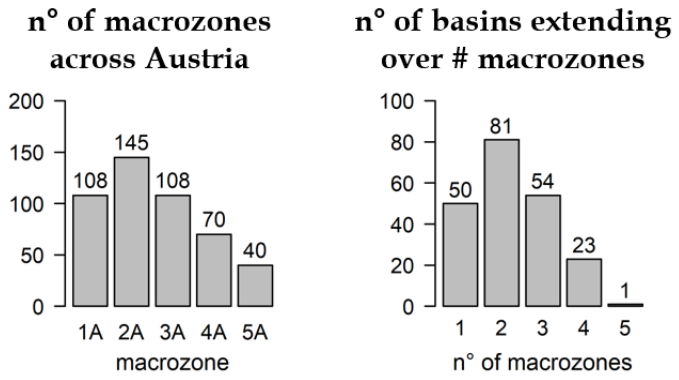

Figure 6.3: Left: number of sub-basins for each macrozone in Austria. Right: number of catchments occupying 1,2...5 macrozones. Only macrozones extending over at least the 10% of basin area are considered.

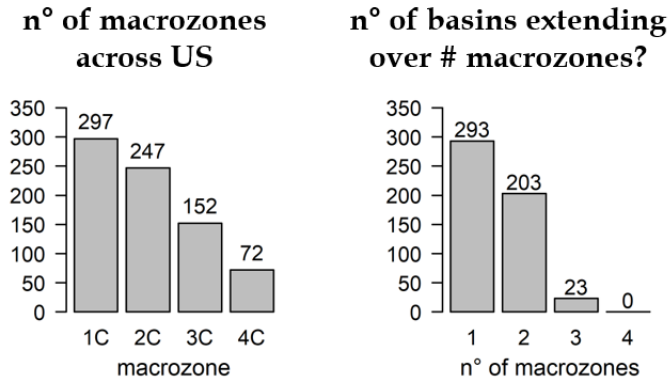

Figure 6.4: Left: number of sub-basins for each macrozone in Austria. Right: number of catchments occupying 1,2...4 macrozones. Only macrozones extending over at least the 10% of basin area are considered.

third of them (78 over 209) cover at least 3 macrozones.

For the US country instead, given the extension of the study region and the variability of the climatic conditions, it is more difficult to identify a general country-wide zonation and the literature does not provide additional information. However, a fixed subdivision is required for the analyses: four macrozones (named 1C,...,4C) are identified trying to have sufficiently large sub-basin samples. Similarly, Fig. 6.4 shows the numerosity of the macrozones across the country and the number of zones occupied by the catchments. As can be noticed, differently from Austria and even if lower macrozones are defined with narrower ranges than the Austrian ones (e.g. macrozone 1A includes both 1C and 2C), most of the catchments are flat and located at low elevation: macrozone 1C is the most

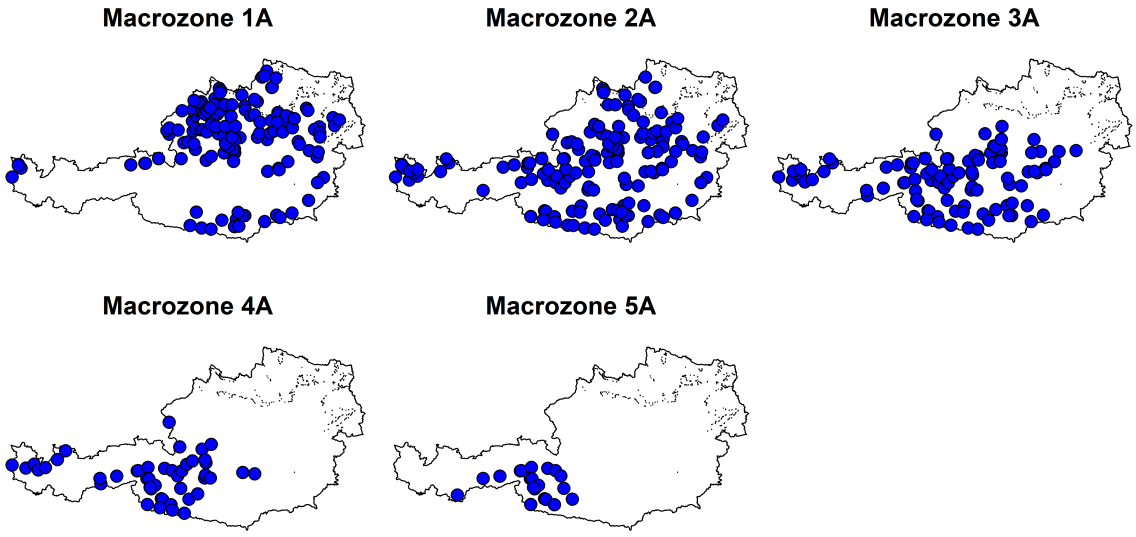

Figure 6.5: Location of the catchments including each of the five Austrian macrozones for a portion of drainage area > 10%. Points refer to basin outlets.

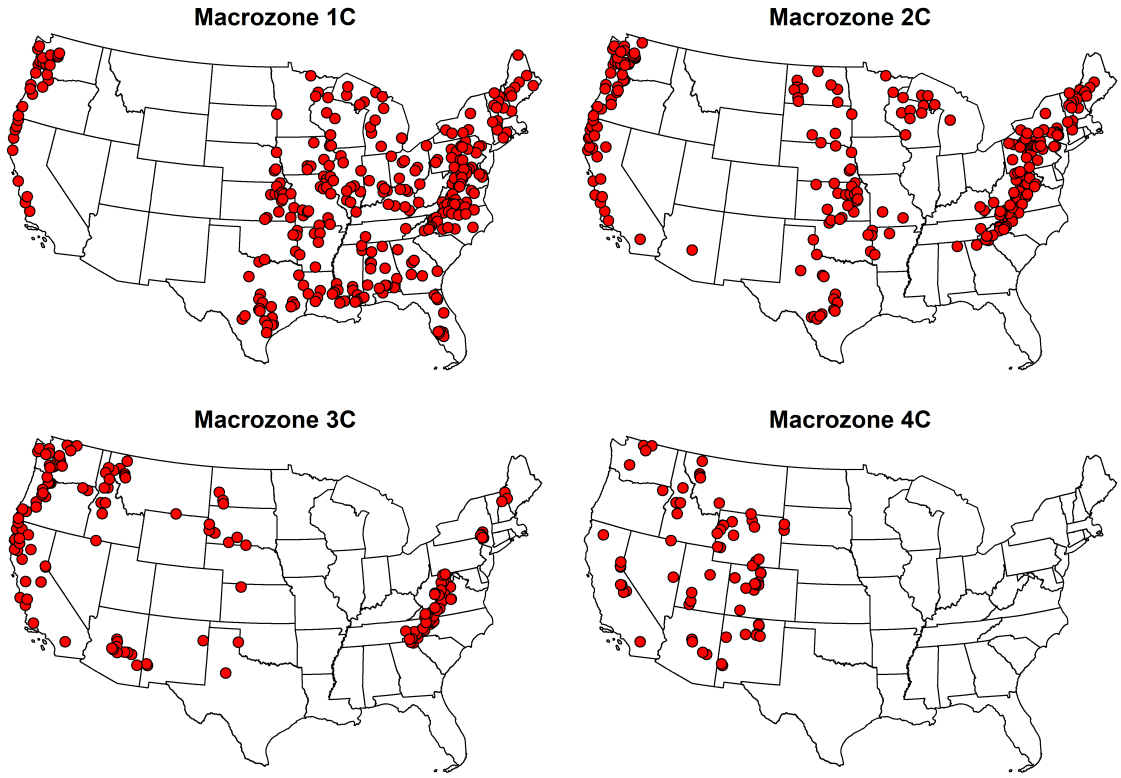

Figure 6.6: Location of the catchments including each of the four US macrozones for a portion of drainage area $> 10\%$. Points refer to basin outlets.

populated and more than half of the catchments occupy a single macrozone. Most of the remaining catchments extend over two macrozones.

Figs. 6.5 and 6.6 report, respectively for Austria and US regions, the location of the catchments covering each macrozone for a portion of drainage area greater than 10%. We see that, for each entity, the distance from the closer sub-basin(s) in Austria is always limited and similar across the different macrozones; for US instead, that distance increase substantially for higher macrozones, indeed depending on the location: midwest macrozones of the group 3C are quite distant between each others as well as arid south west for macrozone 4C.

Notice that catchments extending over just one macrozone are not excluded from the analysis: even if they could benefit less from semi-distributed modelling, their contribute to characterise rainfall-runoff generation processes at the corresponding elevation cannot be ignored.

6.2.1 Computation of attributes at macrozone-scale

As introduced above, the similarity between basins' macrozones (and not just between entire catchments) is taken into account: climatic and geomorphological attributes, which

Table 6.2: Catchment and macrozone descriptors used for the analysis.

Attribute	Description	Data source	
		Austria	USA
Elev	Average elevation	SRTM	
Area	Drainage area	R-R input data	
Slope	Mean slope	SRTM	
Precip	Mean annual total precipitation	R-R input data	
PET	Mean annual total evapotranspiration	R-R input data	
SnowF	Fraction of precipitation fallen as snow (i.e. precipitation fallen in days below 0°)	R-R input data	
SnowD/SWE	Mean annual snow depth / SWE	R-R input data	
Aridity	Aridity index	R-R input data	
Irrad	Mean annual solar irradiance	SRTM	R-R input data
Land Cover	Portions of land use coverage	NLCD	CORINE
Forest	Fraction of catchment covered in forest	NLCD	CORINE
Soils	Portions of regional soil types	FAO global soil maps	

are originally available at catchment-scale in the data-bases, have to be calculated again at macrozone level. The set of catchment features considered for the work is the same reported in Tab. 2.1. Exception is made for some of them, meaningful only from a whole catchment perspective, which are excluded for this study: geological formations, fraction of catchment with porous aquifer, FARL and stream network density.

Mean annual total precipitation and evapotranspiration, aridity index and snow fraction of precipiation were computed similarly to what done at catchment scale for both case studies consistently: since meteorological rainfall-runoff model inputs are available for 100 or 200 m elevation zones, respectively for USA and Austria, features can be aggregated at macrozone scale, instead of at catchment scale, and the corresponding climatic features can be computed. Similar procedure is followed for mean annual snow depth, but it is available only for Austria; for CAMELS, it is replaced with Daymet snow water equivalent (SWE).

The computation of all the remaining features involves the use of the contours of the elevation zones of each catchment: for Austria, they are derived from the SRTM Global Digital Elevation Model; for US, they are provided in the CAMELS dataset. Then, average macrozone elevation and slope are derived directly from the SRTM for both regions. Land cover classes and forest fraction are extracted from CORINE Land Cover maps (see Sec. 2.1) in Austria and from National Land Cover Database (NLCD) 2001 (http://www.mrlc.gov/nlcd2001.php) with 30 m resolution (whose classes are very similar to those of the European CORINE) in US dataset. Soils classes are extracted from FAO global soil maps consistently in both case studies. Finally, mean solar irradiation is computed differently between the two regions: in Austria, it is computed as function of the latitude and the terrain topography (slope and aspect), analogously to what done at

catchment scale (see Sec. 2.1); in US, it is replaced with the mean downward shortwave radiation (Daymet forcing variable). Tab. 6.2 recaps the descriptors used in this analysis and the corresponding data sources for each study region.

6.3 Parameterisation of the TUW model in a semi-distributed framework: can a macrozone-based approach improve model performance at gauged sites?

As introduced in Sec. 6.1.2, the first experiment of the work involves the development of an innovative semi-distributed calibration of the TUW on the gauged sites, which allows to differentiate the runoff generation dynamics across different altitudes. The outcomes will be benchmarked to those of the standard application of the TUW.

6.3.1 Benchmark calibration - TUW-parU

In the benchmark calibration approach, named *TUW-parU*, a single set of runoff generation parameters is calibrated for all the macrozones, which are therefore governed by the same parameters (Fig. 6.7). It is the same calibration approach followed in Sec. 4.2 for Austria. The generation of runoff can change across different altitude only due to differences in meteorological inputs but the hydrological dynamics are controlled by the same parameters. The number of parameters is the same for all the catchments.

6.3.2 Semi-distributed calibration - TUW-parD

In the innovative semi-distributed approach, here termed *TUW-parD*, the subset of the TUW parameters controlling runoff generation is allowed to vary across the different macrozones, enabling different hydrological behaviour in the catchments and different dynamics of the snow routine and the soil moisture routine. The parameters ruling the propagation of the runoff are instead kept constant within the basin: as already mentioned, the model structure does not include propagation of runoff between sub-basins, which are considered autonomous entities contributing separately to the total outlet flow. Fig. 6.8 shows an example of the parameter set in case of three macrozones.

It is important to recall that catchments may obviously extent over different elevation, thus covering different macrozones (Figs. 6.3 and 6.4). For this reason, the number of parameters to calibrate changes accordingly to the number of macrozones covered by each

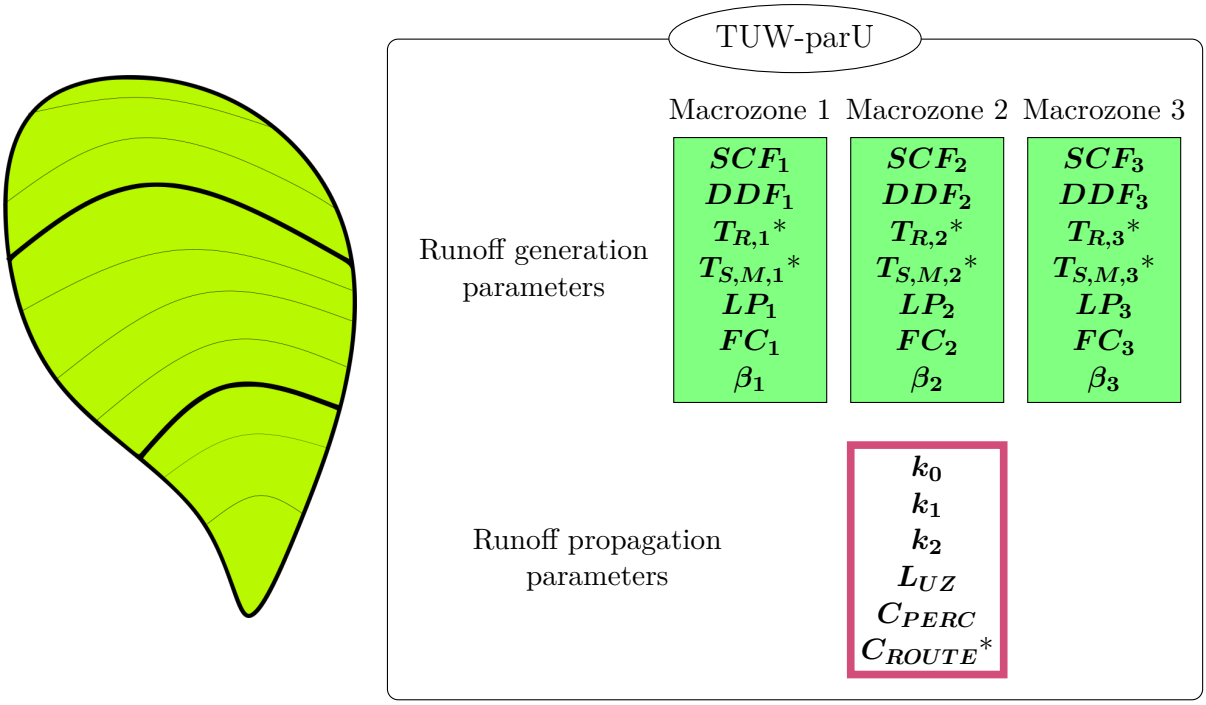

Figure 6.7: Example of the TUW-parU parameter set for a catchment extending over three macrozones: runoff generation parameters are the same for all the macrozones and runoff propagation parameters are unique for the whole catchment. * denotes parameters calibrated only for US region, and fixed for Austria.

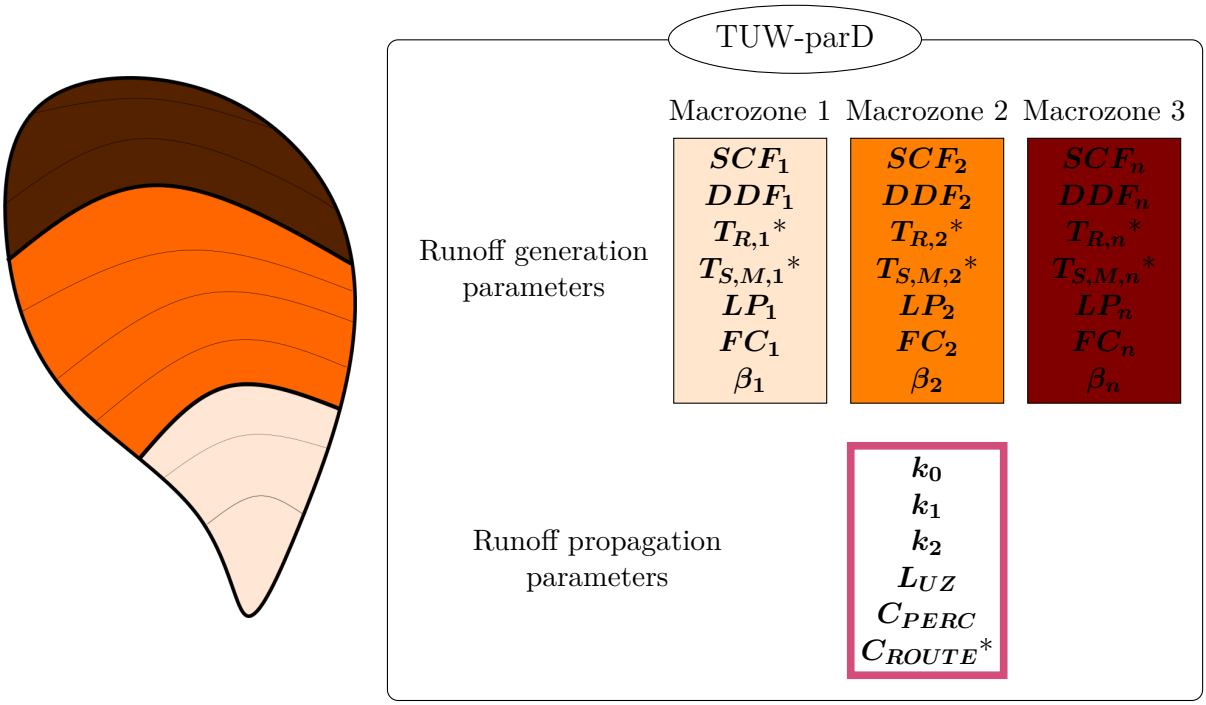

Figure 6.8: Example of the TUW-parD parameter set for a catchment extending over three macrozones: runoff generation parameters are differentiated across macrozones while runoff propagation parameters are unique for the whole catchment. * denotes parameters calibrated only for US region, and fixed for Austria.

Table 6.3: Calibrated TUW model parameters and their ranges for both calibration approaches.

Parameter	Units	Range	
		Austria	**USA**
$\boldsymbol{SCF}$	-	0.9 - 1.5	0.9 - 1.5
$\boldsymbol{DDF}$	mm/°C/day	0 - 5	0 - 5
$\boldsymbol{T_R}$	°C	fixed to 2	-3 - 10
$\boldsymbol{T_{S,M}}$	°C	fixed to 0	-3 - 3
$\boldsymbol{LP}$	-	0 - 1	0 - 1
$\boldsymbol{FC}$	mm	0 - 600	0 - 1000
$\boldsymbol{\beta}$	-	0 - 20	0 - 20
$\boldsymbol{k_0}$	days	0 - 2	0 - 2
$\boldsymbol{k_1}$	days	2 - 30	2 - 30
$\boldsymbol{k_0}$	days	30 - 250	30 - 250
$\boldsymbol{L_{UZ}}$	mm	0 - 100	0 - 100
$\boldsymbol{C_{PERC}}$	mm	0 - 8	0 - 20
$\boldsymbol{C_{ROUTE}}$	$days^2$/mm	fixed to 25	0 - 50

catchment. After the calibration process, still performed separately at each location, a set of runoff-generation parameters is associated to each macrozone of each catchment in the datasets, while a single set of runoff-propagation parameter is associated to the entire basins.

6.3.3 Calibration settings in common between the TUW-parU and TUW-parD approaches

In order to highlight the effect of the different parameter structures of the two techniques, all the remaining calibration features are the same for the two approaches:

- The *historical period* used for model calibration/validation: in Austria, like in Sec. 4.2, the model is calibrated between November 1st 1976 and October 31st 1992, while it is validated between November 1st 1991 and October 31st 2008, using a single year warm-up. For the USA, calibration period is from October 1st 1980 to September 30th 1998 and validation period is from October 1st 1993 to September 30th 2011; here the warm-up is raised to five years, based on previous applications of the HBV model on the same study region (Abebe et al., 2010; Melsen et al., 2018; Melsen and Guse, 2019) and on the presence of arid catchments across the datasets, which usually require a longer initialisation time (Kim et al., 2018).

- *Rainfall-runoff meteorological inputs* are always defined over the finer elevation zones

(200 m for Austria and 100 m for USA), as detailed in Sec. 6.1.2.

- *Parameter ranges*: in Austria, the same parameter ranges implemented in the analysis of Chapter 5 (Tab. 4.1) is used. Exception is made for $\boldsymbol{C_{ROUTE}}$ parameter, which is fixed for the semi-distributed appraoch to 25 days^2/mm (based on the expertise of model authors on the region and on preliminary attempts); in USA, the parameter ranges are chosen based on the works of Abebe et al. (2010), Melsen et al. (2018) and Melsen and Guse (2019) and on the advises of model authors: in particular, in the snow module the threshold temperatures $\boldsymbol{T_R}$ and $\boldsymbol{T_S}$ are calibrated, while melting temperature $\boldsymbol{T_M}$ is set equal to $\boldsymbol{T_S}$; in addition, $\boldsymbol{FC}$ and $\boldsymbol{C_{PERC}}$ boundaries are expanded in order to fulfil a larger variety of hydrological behaviours (including more arid catchments). Tab. 6.3 compares calibration boundaries between the two study regions. The number of parameter to estimate in the standard calibration is respectively 11 and 13 for Austria and USA, while in the semi-distributed calibration it varies from 10 (1 macrozone covered) and 30 (5 macrozones covered) in Austria and from 13 to 34 (4 macrozones covered) in USA.

- *Objective function*: Kling-Gupta efficiency (Eq. 3.52, Gupta et al., 2009) between simulated and observed daily discharges is used.

- Settings of the *optimisation algorithm*: DDS algorithm (Sec. 3.2.2, Tolson and Shoemaker, 2007) is applied maintaining the same number of function evaluations (set to 4000). Analogously to what done for the calibration of R-R models in the previous chapter, for each catchment and each of the approaches, the calibration procedure is repeated ten times for ten different starting points in the parameter space, and the set of obtained calibrated parameters corresponding to the best KGE score is selected. The reliability of the calibration processes is further monitored verifying the pattern of the objective function score for increasing number of function evaluations.

Computational time for the calibration process varies across catchment, mainly depending on the number of sub-catchments on which the model is run: single model run ranges between 0.1 and 2 seconds. In order to reduce the duration of the calibration processes, the algorithm is run taking advantage of parallel computing: when calibrating each single catchment, the DDS algorithm is run at the same time on more machine threads. This is done through the *parallel* R Package.

6.3.4 Calibration results: validation of the semi-distributed parameterisation TUW-parD at-site

The results of the two calibration procedures, i.e. where the runoff generation parameters are uniform for all the elevation zones (TUW-parU) and where they change with the elevation zone (TUW-parD), are evaluated considering both Kling-Gupta efficiency, the objective function for model calibration, and Nash-Sutcliffe efficiency, the standardised version of the root mean square error (Eq. 4.1). Next paragraphs compare the calibration approaches for each study region.

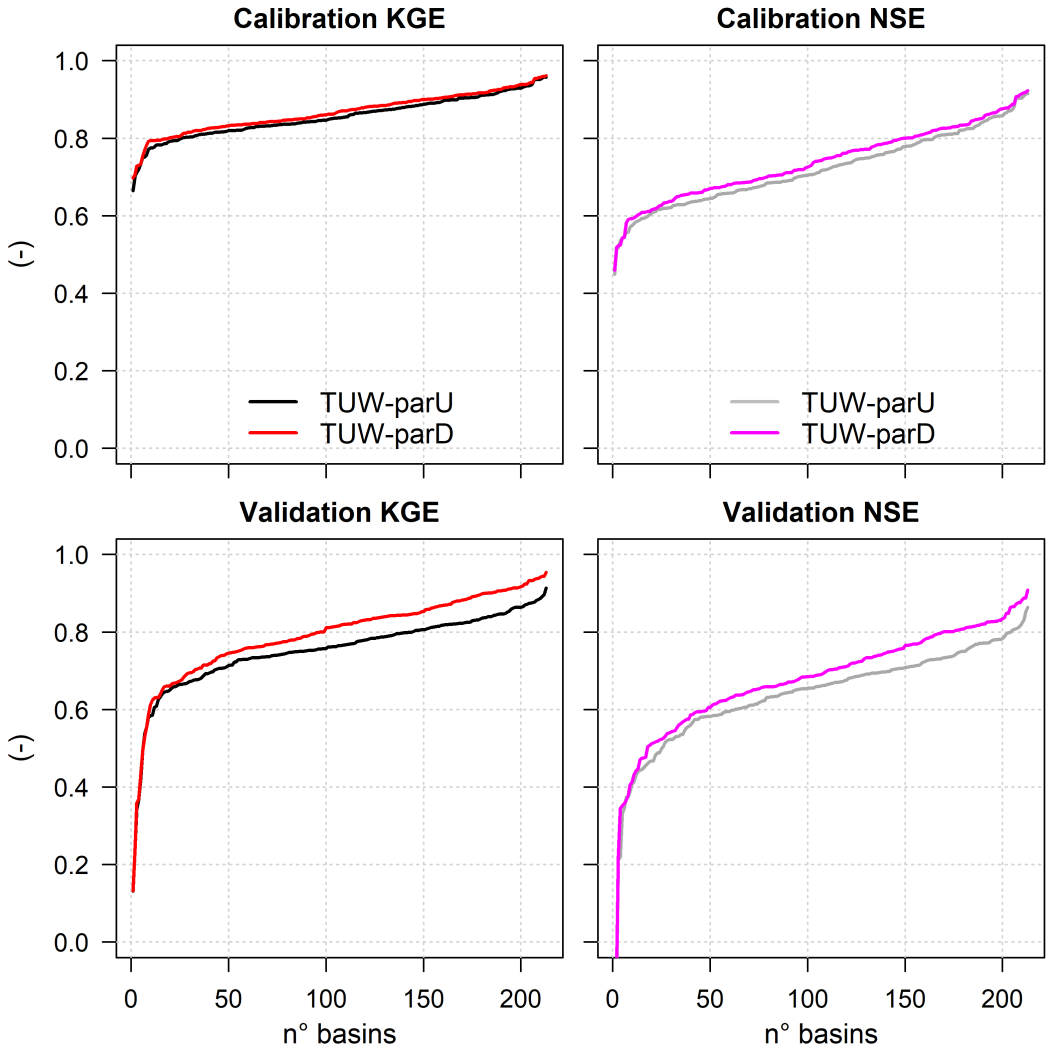

Figure 6.9: Comparison of the standard and semi-distributed at-site calibration approaches for Austria: cumulative distribution function of KGE and NSE.

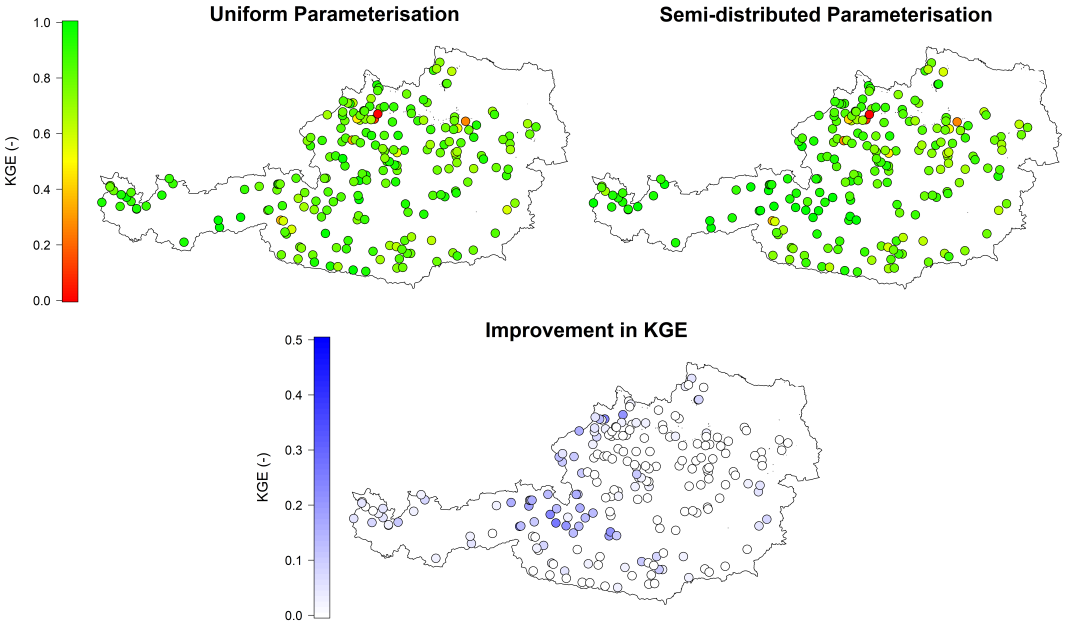

Figure 6.10: Pattern of the at-site standard and semi-distributed KGE (upper panels) across Austrian dataset and corresponding improvement brought by the semi-distributed approach (lower panel) in the validation period.

Austria

Fig. 6.9 reports the results of the calibration approaches for the Austrian dataset in terms of cumulative distribution function of model efficiencies (KGE and NSE) of daily runoff: the plot indicates the value of efficiency which is not exceed by the corresponding number of catchments. In the calibration period the two approaches perform similarly and the two lines are almost overlapping; in validation mode, i.e. when the model is run on an independent data period, the performance of the TUW-parD is slightly better than the TUW-parU. However, it has to be pointed out that the original performances of the TUW-parU are already optimal, and thus difficult to improve. Fig. 6.10 shows the spatial pattern of KGE efficiency for the two parameterisations (upper panels) and the corresponding improvement brought by the novel approach (lower panel): except for a few basins, the performances are uniformly good across the country for both parameterisations; in addition, the KGE score is improved by more than 0.1 for a group of basins, mainly located in the central mountainous area. The observed performance improvement are analysed against catchment characteristic (not showed here) but, unfortunately, no significant relationships are observed.

At this phase, an additional outcome is that the TUW-parD, despite the higher number of model parameters to calibrate, does not lead to over-parameterisation of the model, i.e. the situation in which, due to the large number of parameters, the model fits also

the noise of the calibration and it is not able to generalise to independent (Jakeman and Hornberger, 1993): this is guaranteed by the fact that the model works well also on validation periods. In fact, model testing, or model validation, is crucial to avoid running into over-parameterisation (Seibert et al., 2019).

Given the results obtained by the TUW-parD on the validation period, we can say that the differentiation of the runoff generation processes across the macrozones brings benefit, even if modest, to the at-site simulation.

USA

Similarly to what presented for Austria, Fig. 6.11 compares the model efficiencies obtained by the TUW-parD and TUW-parU across the 527 US catchments. During the calibration period the model performs similarly when following the standard or the semi-

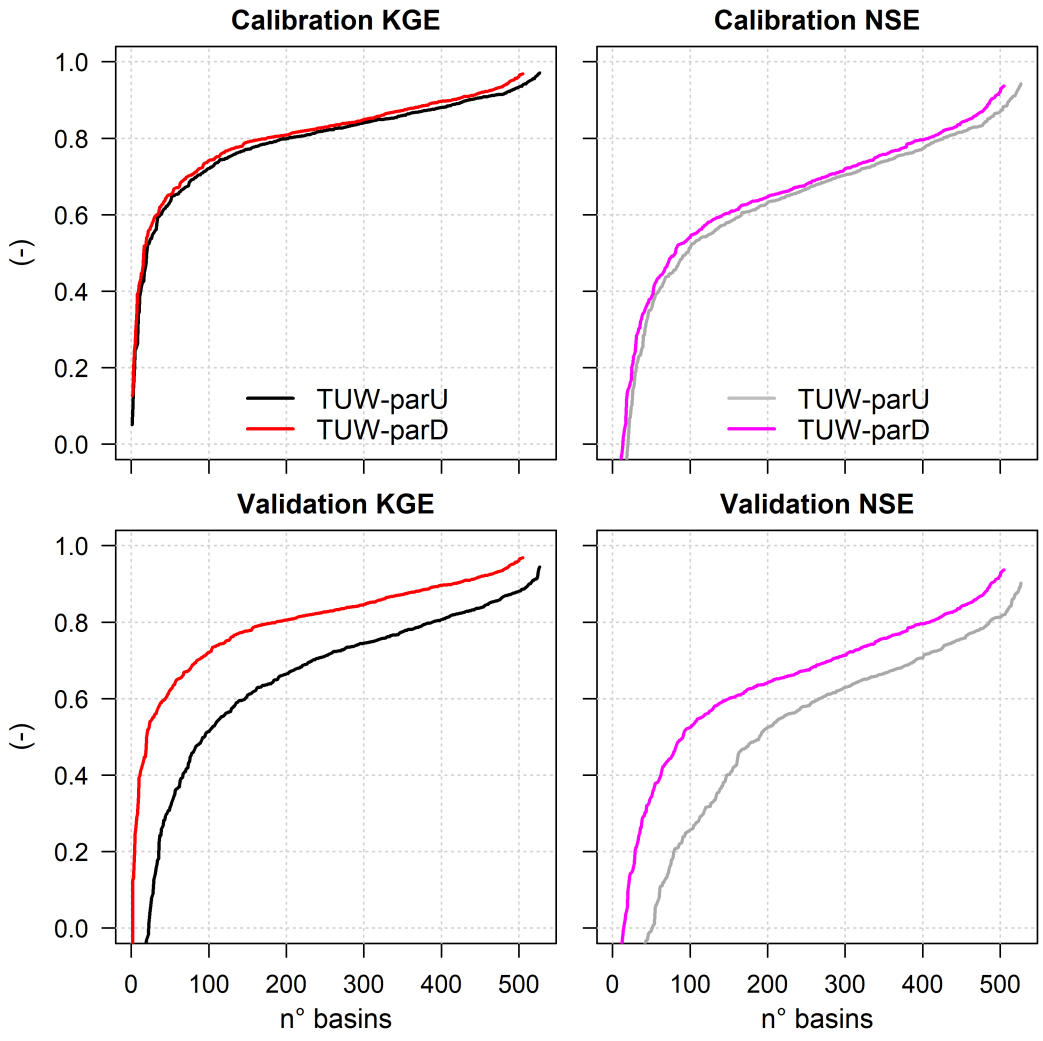

Figure 6.11: Comparison of the standard and semi-distributed at-site calibration approaches for USA: cumulative distribution function of KGE and NSE.

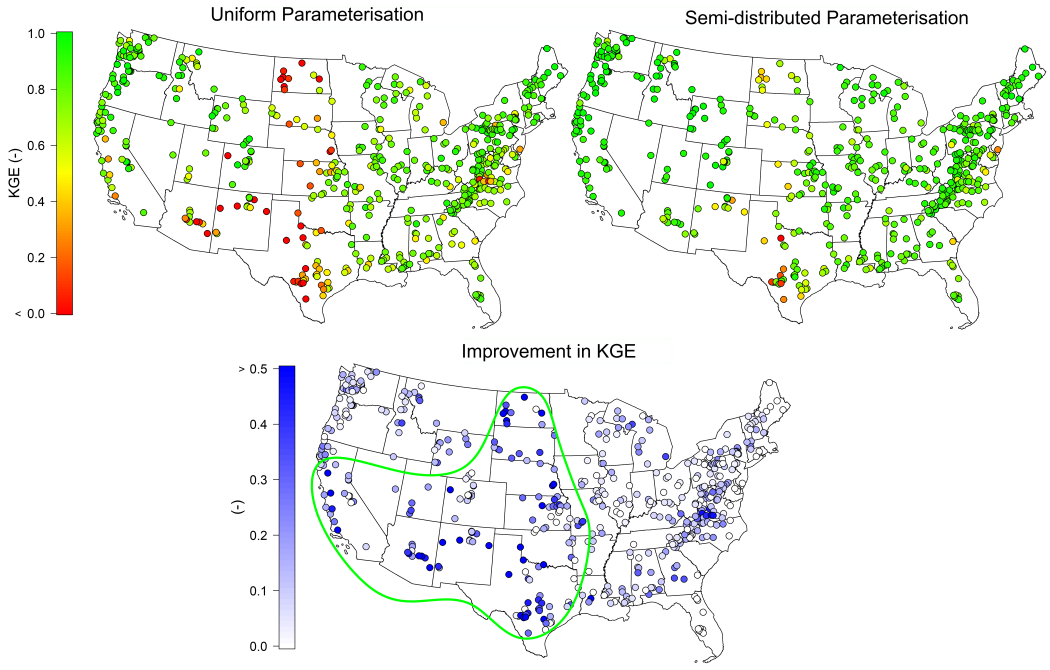

Figure 6.12: Pattern of the at-site standard and semi-distributed KGE (upper panels) across US dataset and corresponding improvement brought by the semi-distributed approach (lower panel) in the validation period (green contour highlights regions with higher KGE improvement.

distributed parameterisation approach. Instead, looking at the validation period it can be noticed that TUW-parD definitely outperforms TUW-parU: median KGE improves of approximately about 15%, but in particular, the number of catchments which fail the simulation decreases substantially: if we consider as reference values to detect model misbehaviour respectively KGE=0.3 and NSE=0.5 (Knoben et al., 2019), the TUW-parD reduces by 80% the number of catchments which fail the simulation in terms of KGE (from 43 to 8 catchments with KGE<0.3) and by the 50% in terms of NSE (from 187 to 91 catchments with KGE<0.5).

The eight catchments with the lowest Kling-Gupta efficiency in the validation period for both approaches are excluded from the next analysis, since evidently the model is not able to simulate properly the rainfall-runoff processes in such basins and their presence in the dataset could distort the following applications.

Fig. 6.12 shows the spatial pattern for KGE across the country and highlights the improvement allowed by the use of the proposed TUW-parD: most of the benefits are centred in the more arid regions (highlighted with the green contour), in the southern and mid-west states including the Great Plains (see Fig. 2.2) where the "standard" TUW-parU fails more frequently, as can be observed in previous applications (e.g. Melsen et al., 2018).

One could expect that model improvement would be higher in such basins covering

a larger number of macrozones, since they may have a greater within-catchment variability of hydrological processes; but instead, no clear dependency between the number of macrozones and model efficiency (or improvement) have been observed. However, the differentiation of model parameters across macrozones appear to be substantially helpful to model some of these catchments. It is important to highlight that the improvement is necessarily to be ascribed only to those catchments extending over more than one single macrozone, for which TUW-parD and TUW-parU parameter sets differ.

Pattern of semi-distributed model parameters across macrozones

A further aspect of interest when implementing a semi-distributed approach for model calibration could be seeking for a possible pattern of the parameters values across the macrozones. In this case in fact, the identification of a trend in parameter values related to macrozones would help to better understand the hydrological processes varying as a function of altitude. In particular, it may help to identify a specific "behavioural" range of each parameter value for each macrozone: this would also allow to further constrain calibration, facilitating the transfer of the parameter from gauged to ungauged entities when regionalising the model.

Figs. 6.13 and 6.14 show the distribution of TUW-parD model parameters values within the calibration ranges across the macrozones, respectively for Austria and USA (only the parameter governing the runoff generation are shown, since the remaining are kept constant). Unfortunately, no strong pattern can be detected for any of the param-

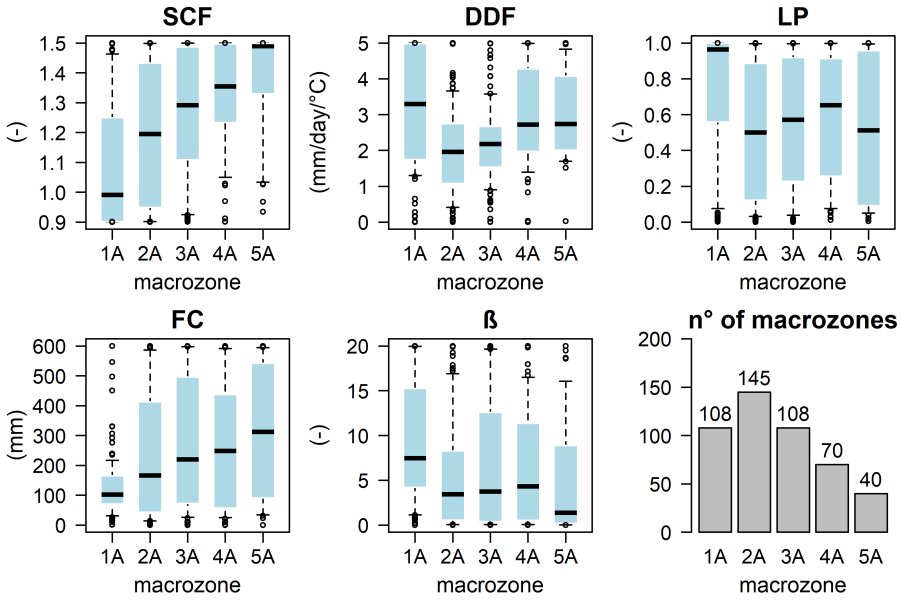

Figure 6.13: Distribution of the runoff generation parameters across macrozones in Austria. Boxplots whiskers refert to 10% and 90% quantiles.

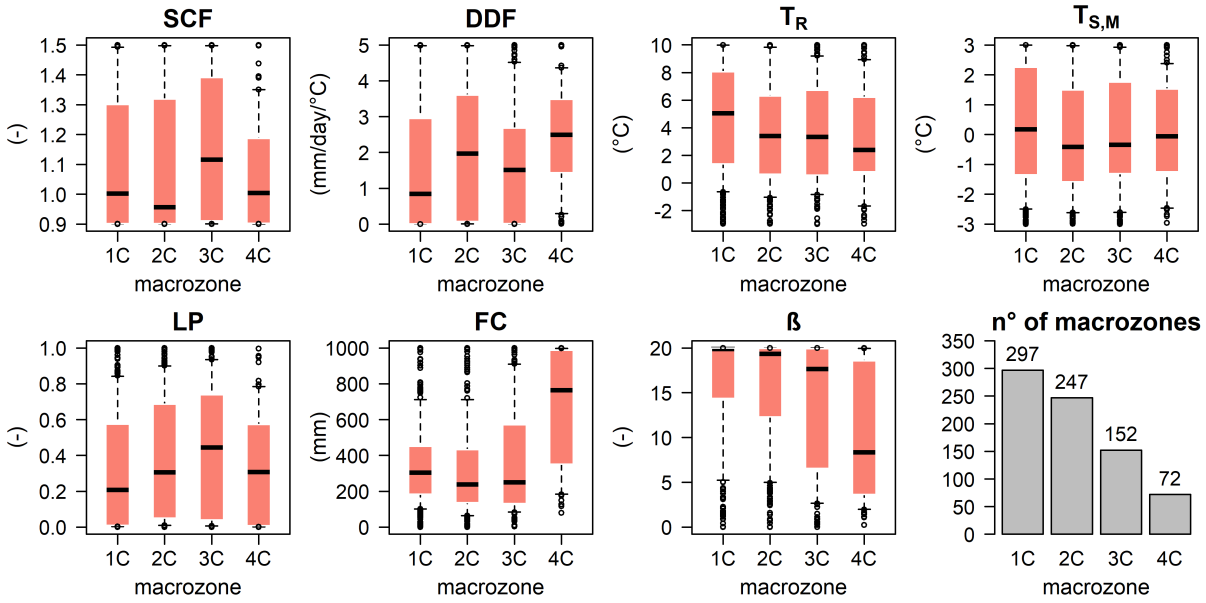

Figure 6.14: Distribution of the runoff generation parameters across macrozones in USA. Boxplots whiskers refert to 10% and 90% quantiles.

eters: in fact, the values are in general spread quite uniformly between the boundaries and/or the trend is not clear.

In Austria, median ***SCF***, which is a correction factor for the snowfall estimate (i.e. as higher the parameter as snowfall increases), exhibits an upward trend for increasing altitude: this is coherent with an increase in the snow pack, expected at higher altitude; however, in lower macrozones the values are widely distributed in the parameter range. Median ***FC***, which estimates the capacity of the soil moisture storage, has a weak upward trend with elevation, but values are quite spread across the range except for the lower macrozone. On the other hand, $\boldsymbol{\beta}$, which governs the release of runoff from soil moisture (i.e. as higher is the parameter, as higher is the release, under the same soil moisture content) show an opposite weak tendency.

Parameters ***FC*** $\boldsymbol{\beta}$ show similar tendencies also in USA, even if the patterns are not strong neither. Differently from what observed in Austria, ***SCF*** exhibits less clear pattern with altitude: this may be due to the fact that basins are more sparsely distributed across a so much larger region, dominated by much more different climate conditions and catchment physiographic characteristics.

Further calibration analyses using parameter research bounds differentiated across macrozones may be tested, but in the present work, it was decided to move through the regionalisation phase of the analysis with these calibrated parameter sets.

6.4 Semi-distributed regionalisation of the TUW_G-parD: can the macrozone structure improve simulation at ungauged sites? How is similarity changing with altitude?

Sec. 6.1.2 introduced that the second phase of the analysis focuses only on the regionalisation of the parameters governing the runoff generation. This is done to adapt the regionalisation to the structure of the model and to focus on the portion of the hydrological processes most influenced by landscape elevation.

However, since the contributes from the macrozones are then propagated together through the catchments, it is not possible to isolate the single performances of sub-catchments at the basin outlet. For this reason, the model routines governing the runoff generation process are separated from the original TUW model and used independently for evaluating the hydrological similarity and for applying the proposed regionalisation framework over the elevation macrozones. The propagation routines are therefore excluded from the regionalisation and similarity evaluation applications.

In order to exclude propagation processes, the model outcome to be considered is not the streamflow but the amount of runoff generated over each sub-catchment, i.e. the portion of the water "leaving" the soil moisture storage, to be successively propagated through the catchments/sub-catchments. Such quantity in the model is called ΔS_{UZ} (Eq. 3.6, Fig. 6.15) and it is given in mm/day (it does not depend on sub-catchment area, which is used exclusively to aggregate the contributions from different entities). Since the original version of the *TUWmodel* R package does not provide automatically ΔS_{UZ} as a model output, the model is decomposed, and the runoff generation routines are extracted. A new model, called TUW_G, is re-coded in R environment: it provides the runoff producted and leaving the soil moisture storage, which can be available either at macrozone level or basin level. Model inputs and routines are the same as for TUW but only runoff generation parameters are required, of course. Fig. 6.15 shows the portion of the model (in red) corresponding to TUW_G: it has to be pointed out that such scheme is related to a single model entity (i.e. elevation zone), and the outputs from all the zones are then aggregated to find the total catchment contribution.

Of course, a limitation of this methodology is that the use of the only runoff generation module of the TUW (TUW_G) prevents the validation of the method against observed values, since the only available observations are the streamflows resulting from the propagation of all the macrozones contributions to the basin closure section. In fact, the

real runoff production in each macrozone is unknown also when considering the basin as gauged. In absence of observed runoff production values, the performances of the proposed regionalisation method will be assessed against the best available at-site simulation of runoff production: that is, for each elevation zone, we assume as "true" the runoff production value that is simulated by the best performing model calibrated at-site. Previous section shows how TUW-parD outperforms TUW-parU for at-site simulations. Thus, assuming that the accuracy of TUW-parD in reproducing streamflow at the basin outlet is reflected also in its estimation of the runoff generation, the at-site ΔS_{UZ} values simulated with TUW$_G$-parD are used as reference for validating the regionalisation approaches: the

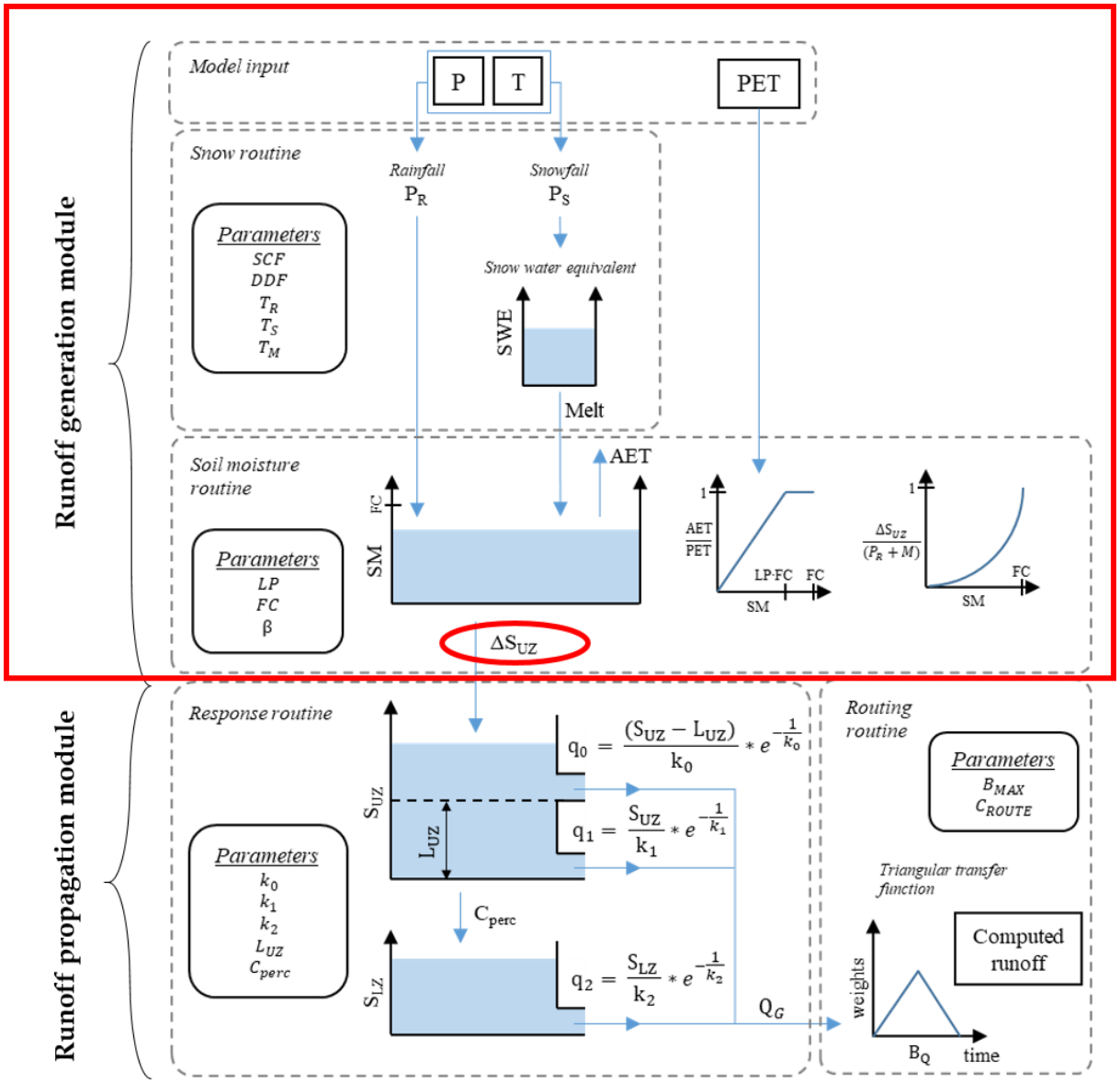

Figure 6.15: TUW$_G$: portion of module routines of TUW model related to runoff generation, highlighted in red. The scheme is related o a single model entity (i.e. elevation zone).

$\Delta S_{UZ,site}$ is considered as the "true" amount of runoff production.

Once the TUW-parD and the TUW-parU have been calibrated (see Sec. 6.3), the corresponding TUW_G-parD and TUW_G-parU can be used for the regionalisation analysis. It is important to recall that TUW_G-parD and TUW_G-parU can be run independently for each macrozone.

This section aims to develop a regionalisation framework especially adapted to the TUW_G-parD model in order to evaluate the benefit in terms of regionalisation efficiency of such semi-distributed parameterisation in comparison to the benchmark method, where all the macrozones are controlled by the same parameters (TUW_G-parU).

Regionalisation technique: MS-OA

In order to evaluate the importance of different physical and climatic descriptors and test the possible benefits due to the implementation of similarity and parameter transfer at sub-basin scale, an appropriate regionalisation method must be identified.

It was already argued that similarity-based approaches taking advantage of output-averaging concept, like the NN-OA or MS-OA methods (Sec. 3.3.4), are preferred by most authors for the regionalisation of conceptual model parameters (e.g. McIntyre et al., 2005; Viviroli et al., 2009; Lebecherel et al., 2016). Additionally, the experiments of the previous Chapter (Secs. 5.2 and 5.3) proved the MS-OA method to be accurate and robust when applied to regions with varying informative content.

In view of such findings, here the MS-OA approach is also used. Similarly to what done in Chapter 4 (Secs. 4.3.1 and 4.3.2), this regionalisation procedure involves the choice of two settings: the number of donors n and the attributes to be included in the dissimilarity index of Eq. 1.2. Even if both aspects are crucial, the latter is of particular interest because it may allow us to understand how similarity changes at different altitude (different macrozones) or at different scales (catchment/macrozone). Details about the methodology, in respect to its application to the TUW_G, are given in Appendix A (Sec. 6.6).

6.4.1 Benchmark: regionalisation of TUW_G-parU with the MS-OA approach

In the more standard regionalisation framework for the TUW_G-parU model, the runoff generation parameters, which are uniform for all the macrozones, are regionalised at catchment scale and considering the similarity between the entire catchments and not

between the sub-basins. Therefore, a single set of attributes has to be selected for each basin and controls the inter-catchment dissimilarity (expressed by Eq. 1.2).

Choice of the set of attributes for the dissimilarity index Φ

A widely used methodology to select the set of attributes is the one used in Sec. 4.3.1, in which MS-OA approach is applied sequentially to the entire dataset in leave-one-out cross-validation, using at each step an increasing number of attributes when defining the dissimilarity index Φ until the best accuracy is reached: at each step, the method is tested multiple times, adding one by one each of the attributes and the one which gives the best regionalisation performances is selected. Alternatively, one can test all the possible combinations of a pre-set number of attributes, as will be done in this case.

Preliminary tests (not showed here) highlighted that including three attributes in Φ (i.e. fixing $m = 3$ in Eq. 1.2) was a good trade-off which allows to maximise regionalisation performances while keeping acceptable computational times (i.e. number of attribute combinations to test). Thus, it was decided to proceed accordingly, testing all the possible combinations of three of the 12 available catchment descriptors (Tab. 6.2), for a total of 220 combinations. For each combination the regionalisation procedure is applied to all the N catchments across the dataset in leave-one-out cross validation, i.e. each catchment is considered to be ungauged, and all the remaining basins are available in the donors set: after computing the dissimilarities with the donors, the MS-OA is applied obtaining the regionalised runoff production $\Delta S_{UZ,U,basin}$. Then, its accuracy is evaluated against the "true" at-site generated runoff $\Delta S_{UZ,site,basin}$ at the target location. Each combination provides values of model efficiencies KGE^G (i.e. KGE between the regionalised and the "true" at-site runoff production) for all the N catchments. In order to select the best attribute combination, the overall regionalisation performance across the entire dataset has to be taken into account (i.e. considering the simulations in all catchments): for this reason, the interquartile KGE^G is considered to identify the combinations leading to the best regionalisation performances:

$$KGE^G_{IQ} = \frac{q_{25}(\overline{KGE}^G) + q_{75}(\overline{KGE}^G)}{2} \tag{6.1}$$

where $\overline{KGE}^G$ is the vector of obtained KGE^G efficiencies for the entire set of catchments and q_{25} and q_{75} are the 25% and 75% quantiles.

6.4.2 Regionalisation of TUW_G-parD with the MS-OA approach

When applying the MS-OA approach for the regionalisation of the TUW_G-parD, the parameters are transferred at macrozone scale: each macrozone is considered as an autonomous entity and it is regionalised independently; donor parameter sets are transferred from donor macrozones at the same elevation. Similarly, the choice of the attributes to include in the dissimilarity index (Eq. 1.2) is performed separately for each of the elevation ranges Z, thus differentiating the dissimilarity index:

$$\Phi_Z = \sum_{k=1}^{3} \frac{d_{Z,k}(U_Z, D_{i,Z})}{\max(d_{j,k})} \tag{6.2}$$

and a set of attributes for each elevation range Z has to be selected (where Z can be 1A,...,5A for Austria and 1C,...,4C for US). The attributes used for the distance in the attributes space $d_{Z,k}$ refer to the characteristics of the macrozones.

Choice of the attributes for the dissimilarity indexes Φ_Z

As already stressed, different steps are performed to select the attributes independently for each elevation range Z. Similarly to what done for the benchmark regionalisation of TUW_G-parU, all the possible combinations of three attributes are tested. For each combination, the procedure is the same exposed in the previous section but it is applied at macrozone scale (instead of a basin scale): given a set of three descriptors, the dissimilarities between macrozones at the elevation Z are defined and the regionalisation procedure is applied to all the macrozones at elevation range Z across the dataset in leave-one-out cross validation (i.e. each sub-basin is considered to be ungauged, and all the remaining macrozones at the same elevation are available in the donors set). Then, the accuracy of the regionalised runoff production $\Delta S_{UZ,U,Z}$ is evaluated against the at-site generated runoff $\Delta S_{UZ,site,Z}$ at elevation range Z at the target location (macrozones covering a percentage of total basin area lower than 10% are excluded from the donor set). Similarly to what described for the benchmark approach, each combination provides model efficiencies for all the macrozones at elevation Z, and the overall regionalisation performance across the dataset is taken into account with the interquartile KGE^G between the regionalised and the "true" at-site runoff production ($\Delta S_{UZ,site,Z}$).

At the end of the process, for each elevation range Z, a combination of three attributes which maximises the KGE^G can be detected.

6.4.3 Common settings for the choices of the best set of attributes to quantify the similarity in the MS-OA approaches

For both the regionalisation approaches, TUW_G-parU and TUW_G-parD, the following settings are chosen:

- *Number of donors*: in Austria, the number of donors is set equal to three, based on the outcomes obtained in the work of previous chapter (Sec. 4.3.2). In USA, a preliminary test based on the application of the MS-OA approach for the regionalisation of the entire parameter set of the benchmark TUW-parU model was performed, leading to fix the number of donors to three as well; the test is described in Appendix B (Sec. 6.7).

- *Excluded donors*: after initial tests, catchments (and the corresponding macrozones) with an at-site KGE efficiency (Figs. 6.9 and 6.11) in validation period lower than 0.5 are excluded from the donor sets.

- *Simulation period*: for both datasets only the validation period (see Sec. 6.3.3) is considered for regionalisation and all the results presented in the next sections will refer to that data periods. This was done for the same reasons exposed in Sec. 4.3.

- *Computational shortcut to reduce simulation time*: testing each three-attribute combination requires significant computational time since the model have to be run for all the study catchments/macrozones (and three times, due to output averaging): even taking advantage of parallel computing, as briefly described in Sec. 4.2, this would have led to quite long simulation time. In order to substantially reduce the computational cost of the processes, each catchment/macrozone was previously run with all the parameter sets obtained for each potential donor in order to create ready-made databases of potential simulations: in such a way, instead of re-running the model at each iteration (for each ungauged location), the algorithm reads directly the simulation (outcome of the regionalisation process) from the ready-made database, saving computational time. A ready-made database is assembled for each regionalisation process: one at catchment-scale and one for each elevation range of macrozones.

6.4.4 Regionalisation results (MS-OA)

A drawback of the MS-OA approach is that, by definition, it does not allow to understand the relative importance of each of the attributes included in the dissimilarity index.

In fact, Eq. 1.2 gives equal weights to each of the m (three in this case) descriptors considered. Thus, the outcomes of the optimisation processes (i.e. the steps for choosing the attributes at catchment/macrozone scale) described in the previous sections (Secs. 6.4.1 and 6.4.2), allow to select a single set of three attributes (having the same importance in determining the dissimilarity Φ) leading to the best regionalisation performances (i.e. KGE^G_{IQ} between regionalised and at-site runoff production).

On the other hand, results also highlight that there are many sub-optimal combinations with similar performances. This was expected, for the principle of equifinality: a given end-state of the process can be reached by many potential combinations of descriptors. A possible way to indirectly evaluate attribute importance may therefore be to consider more than one best combination and see which attributes appear more frequently in Φ. In this case, we tested such solution but it did not lead to meaningful results, since they depend strongly on the number of sub-optimal combinations considered, that is on how many combinations are included in the above list.

Since such approach did not provide the desired information on the importance of the attributes, we will consider only the single combination of attributes leading to the best KGE^G_{IQ} scores at catchment scale (TUW_G-parU) or at macrozone scale (TUW_G-parD): the results reported in this section concern exclusively the best obtained regionalisation performances.

Austria: best combinations of attributes

Tab. 6.4 reports the optimised combinations of attributes, leading to the best performance (KGE^G_{IQ}) in Austria: of course, a single set of attributes is obtained for the regionalisation of TUW_G-parU, while TUW_G-parD requires a three-attributes set for each elevation range Z.

Uniform parameterisation. It can be observed that the optimal regionalisation of the runoff production at catchment scale (TUW_G-parU) requires three climatic attributes describing i) mean solar irradiation , ii) the presence of snow (mean annual snow depth) and iii) the annual potential evapotranspiration (PET) to be included in the dissimilarity index. Such result shows that climate characteristics, even if not extremely variable across the country, have strong effect on the first phase of the rainfall runoff dynamics. Interestingly, land cover, which resulted to be the most important attribute to represent similarity for the regionalisation of the complete parameter set of TUW model in the work presented in the previous Chapter (Fig. 4.2), seems to lose its significance when focusing only on the processes governing the generation on runoff (TUW_G-parU). Such results indicate that land cover in Austria mainly influences runoff propagation, while climate

characteristics, even if not extremely variable across the country, have strong effect on the first phase of the rainfall runoff dynamics.

Semi-distributed parameterisation. This is confirmed also when regionalising the TUW_G-parD at different elevation ranges: climatic attributes (irradiation, annual precipitation, aridity, PET or snow depth) are good to catch macrozone similarity at all the elevations, where at least one between such characteristics appears in all the best combinations. The role of the geo-morphological attributes is difficult to interpret: for instance, land cover and area appear only at elevation 5A, soil types only at 2A, average catchment elevation is used only at elevation range 1A and 5A, and the percentage of basin covered by forests seems useful only at elevation 4A. This difficulty in highlighting the role of the attributes may be due to the fact that MS-OA approach does not allow to identify the importance of the single attributes and to fully understand changes of similarity with altitude, but only to select a set of best descriptors. Moreover, the issue of equifinality discussed above prevent us from taking such results as unique, with respect to what concerns the choice of the best combination of attributes. However, the maximisation of the regionalisation performances is on the other hand guaranteed.

Table 6.4: Best combination of attributes for the regionalisation of TUW_G-parU and TUW_G-parD with the MS-OA approach in Austria.

		Attributes for MS-OA
TUW_G-parU	**Entire catchment**	Irradiation PET Snow Depth
TUW_G-parD	**Elev. range 1A**	Elevation Irradiation Aridity
	Elev. range 2A	Soils Precipitation Snow Depth
	Elev. range 3A	PET Aridity Snow Depth
	Elev. range 4A	Elevation Irradiation Forest
	Elev. range 5A	Area Land Use Snow Depth

Austria: regionalisation accuracy

When analysing the model performances, the proposed distributed regionalisation allows a slight improvement in the simulations of the runoff generation at least as far as the "true" modelled values are concerned: Fig. 6.16 compares the accuracy (KGE^G) of the regionalisation methods for estimating the runoff generated over the entire catchment when applied with the benchmark approach TUW_G-parU with uniform parameters (orange boxplots) or with the semi-distributed approach transferring TUW_G-parD parameters between macrozones (violet boxplots).

The first frame on the left refers to the accuracy of the model in reproducing the at-site total basin runoff generation $\Delta S_{UZ,site,basin}$: in the case of TUW_G-parD, it is obtained averaging the contributes optimised for the different macrozones (weighted on their corresponding drainage areas).

The frames on the right report the accuracy in reproducing the at-site runoff production for each of the macrozones ($\Delta S_{UZ,site,1A}, \ldots, \Delta S_{UZ,site,5A}$): in this case, the regionalised production for the TUW_G-parU (orange boxplots) is estimated from the simulation run at catchment scale but considering the runoff generated on the separate macrozone.

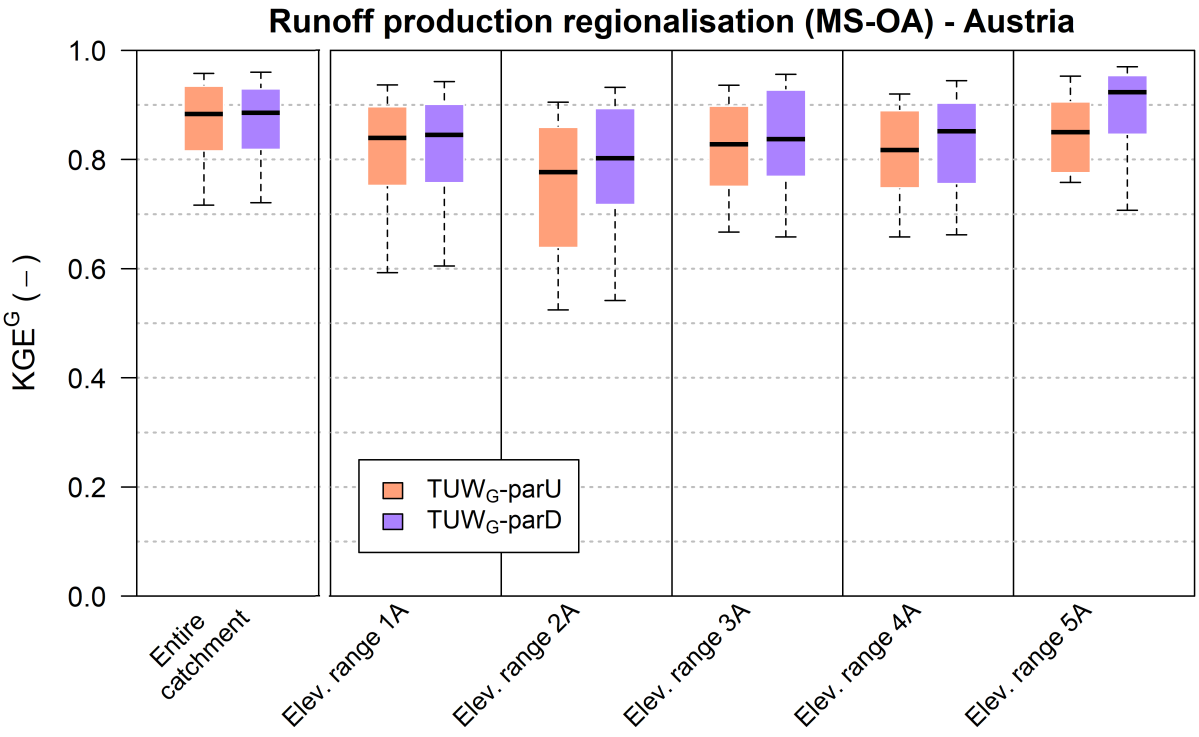

Figure 6.16: Regionalisation accuracy in Austria when regionalising runoff production with the MS-OA method for the standard and semi-disitributed approaches (i.e. tranferring runoff generation parameters at catchment and macrozone scales) against at-site simulations. Boxplot whiskers refer to 10% and 90% quantiles.

We should here remember that even if the parameters are uniform over the zones, the model runs the simulation at the spatial resolution of meteorological inputs (i.e. 100 or 200 meters elevation zones) and the corresponding runoff generation outputs can be averaged for the different macrozones, as stressed in Sec. 6.1.2.

When regionalising the runoff production of the entire catchment (left panel), the semi-distributed approach does not bring benefits. On the other hand, regionalisation efficiencies are slightly improved for single macrozones: this indicates that transferring parameters between macrozones may enhance the simulation of runoff production at macrozone scale.

USA: regionalisation accuracy and best combinations of attributes

Analogously to what reported for the Austrian case study, Fig. 6.17 and Tab. 6.5 show respectively regionalisation efficiencies and the best three-attributes combination for the US region.

Looking first at the regionalisation efficiencies (Fig. 6.17), it can be easily noticed that the semi-distributed based regionalisation of TUW_G-parD is not able in this case

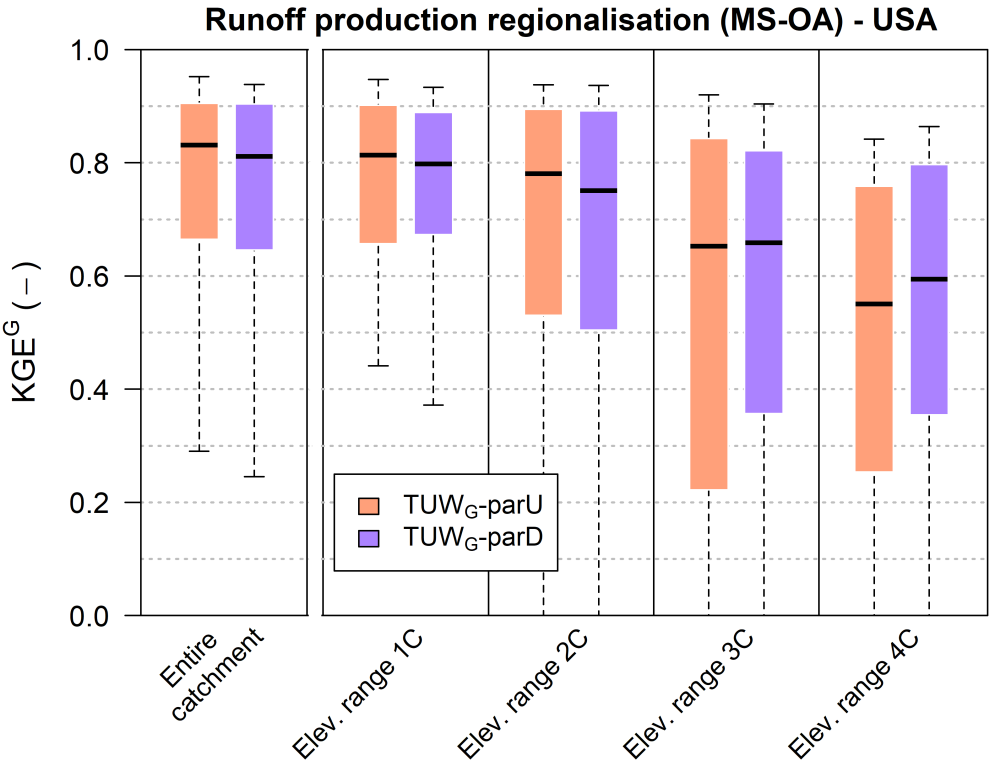

Figure 6.17: Regionalisation accuracy in USA when regionalising runoff production with the MS-OA method for the standard and semi-disitributed approaches (i.e. tranferring runoff generation parameters at catchment and macrozone scales) against at-site simulations. Boxplot whiskers refer to 10% and 90% quantiles.

to improve the accuracy of runoff production estimates in respect to the benchmark approach: the two approaches behave similarly both at catchment and macrozone scale. In addition, while when regionalising the runoff production for the whole catchments the efficiency are satisfactory, they substantially fail for the reproduction of macrozone contributes; in fact, all macrozones except 1C show quite poor KGE^G (in comparison to the efficiency at catchment scale): even with the semi-distributed approach, for several catchments the MS-OA is not able to properly reproduce the runoff generation. Both methodologies are not able to find an appropriate set of donors to faithfully reproduce the at-site simulations.

One possible reason is that macrozone 2C, 3C and 4C are distributed very heterogeneously across the country (see Fig. 6.6) and this affects the accuracy of the methodology.

In addition, the method is based on the assumption that the accuracy of the model in reproducing the discharge at-site is reflected in its ability to reliably reproduce the generated runoff, which may be not valid in several catchments.

Looking at the first row of Tab. 6.5, we see that also in US the best combination of attributes for the runoff production regionalisation at catchment scale includes climatic features: the presence of snow (even if quantified with a different snow attribute,

Table 6.5: Best combination of attributes for the regionalisation of TUW_G-parU and TUW_G-parD with the MS-OA approach in USA.

		Attributes for MS-OA
TUW_G-parU	**Entire catchment**	Slope Arid SnowF
TUW_G-parD	**Elev. range 1C**	Irradiation PET SnowF
	Elev. range 2C	Elevation PET SnowF
	Elev. range 3C	Elevation Irradiation PET
	Elev. range 4C	Slope Soils Forest

referring to the portion of precipitation fallen as snow) and basin aridity, but also a geo-morphological descriptor, average catchment slope, is included.

Similarly to what happens for Austria and for the same reasons, the best combinations for regionalisation at macrozone scale are very difficult to interpret. But this may be probably due to the fact that, for the US, such results are only partially reliable, since as shown by Fig. 6.17 the regionalisation of TUW_G-parD fails substantially for the higher macrozones

In order to better explore changes in similarity across different elevations, the relative importance of the attributes may be investigated with different approaches. An additional experiment (not reported here) is also conducted modifying the regionalisation method, including all the attributes in the dissimilarity measure and assigning weights to them, which have to be calibrated against the "true" runoff generation. Results are again not satisfactory and difficult to interpret, probably due to same reasons mentioned before: the issue of equifinality and the use of simulated runoff generation to optimise the dissimilarity index. For the sake of simplicity, we prefer not to show them here.

6.5 Summary of findings

This Chapter presented a semi-distributed framework for the implementation of an innovative parameterisation of the TUW model, which differentiates the dynamics of runoff generation at different altitudes. The approach was applied both in the gauged and ungauged case. Such framework was compared to a benchmark approach in which the conceptual model is applied with parameters that are the same for all the elevation zones of each catchment. The analysis are implemented on two very large case studies: a densely gauged set of more than 200 Austrian catchments and a database for rainfall-runoff simulation of more than 500 catchments across the United States of America.

The purpose of the study, besides evaluating the benefits in terms of at-site model efficiency given by the semi-distributed parameterisation, was to explore the changes in hydrological similarity across different altitudes applying an ad-hoc regionalisation method both at macrozone and catchment scale.

In the gauged case, the study involved a new calibration of the model, allowing the parameters controlling runoff generation routines to vary across previously defined macro elevation zones (called macrozones). Compared to the uniform parameterisation, the innovative parameter structure allowed an improvement of the at-site model performances especially in the validation period. In Austria the benefit is limited, mainly due to the

fact that the performances on such region are already very good with the uniform parameters, while in USA there is a significant improvement. This demonstrates that i) even if the number of parameters is larger, the model is not overparameterised and that ii) a differentiation of rainfall-runoff generation dynamics across basin elevation zones helps to closely reproduce the overall rainfall-runoff transformations.

No significant patterns of the calibrated model parameters across elevation zones have been observed, but further investigation could deepen the issue, testing the benefit of calibration ranges differentiated across different altitudes to further constrain calibration, facilitating the transfer of model parameters from gauged to ungauged entities.

Successively, a semi-distributed regionalisation framework applied only to the runoff generation module was developed. The TUW model was decomposed, extracting the routines responsible of the runoff production, before its propagation across the basin. Since no actual measure of runoff production is available, the "at-site" generated runoff modelled with the semi-distributed parameterisation was considered to be the reference, "true" runoff generation value, against which the regionalisation approach is evaluated. The two benchmark and semi-distributed runoff generation parameterisation approaches were regionalised accordingly to the similarity between entire catchments or single macrozones respectively. In the benchmark approach, the uniform parameter sets are transferred to the target from the most similar donor catchments; in the semi-distributed approach, the regionalisation is performed independently at the different elevations: the parameter set for generating runoff on each macrozone is transferred from the most similar donor macrozones having the same elevation. The regionalisation method used is the MS-OA, which requires the choice of the attributes to be used for transferring parameters; in such analysis, all the possible combinations of three attributes were tested to maximise the regionalisation accuracies. Such approach, by definition, does not allow to account for relative attribute importance, but just to identify the three attributes leading to the best performance at catchment scale (benchmark approach) or for the macrozones at different elevation (semi-distributed approach).

A limitation of this analysis is the use of at-site simulations of runoff production for optimising similarity measures in the regionalisation process, since actual measurements of runoff production are of course not possible. Such choice was made in order to be consistent with the particular structure of the TUW model which does not consider the interaction between sub-catchments: in fact, since the runoff generated over each elevation zone is propagated independently towards the stream outlet, without considering the internal hydrological relationships of the entities, it would have been meaningless to

differentiate the propagation routine (i.e. different contributions) as well.

Unfortunately the resulting best attribute sets across different altitudes were not clear enough to let us better understand the role of the descriptors in the rainfall-runoff generation dynamics at different elevation. Reasons behind this results may lie either in the incapability of the available attributes to characterise sufficiently well the hydrological behaviour of the single macrozones, but also especially in the equifinality of the process which leads to have more solutions to the same problem and thus not univocal optimal attribute combinations.

In Austria, the indexes assessing the regionalisation efficiency demonstrated that semi-distributed parameterisation is able to improve the simulation of the runoff production estimates at single macrozone level if compared to the uniform parameterisation. In USA, both approaches substantially fail in reproducing the at-site runoff generation at high altitudes. This can be due to a number of factors: first of all and as already stressed, the use of "at-site" simulated time series of runoff production for validating the method is based on the assumption that the accuracy of the model in reproducing the discharge at-site is reflected in its ability to reliably reproduce the generated runoff, which may be not valid in several catchments; in addition, regionalisation procedures in US are in general very challenging (see e.g. Pool et al., 2019), given the very high variability of the climatic and hydrological conditions across the country; in order to find more homogeneous group of catchments, the dataset may be divided into sub-regions, but the choice is not straightforward, as it strongly depends on the hydrological variable of interest (see e.g. the classification of Yaeger et al., 2012; Berghuijs et al., 2014b; Mcmanamay et al., 2014; Brunner et al., 2020; Jehn et al., 2020).

Moreover, being the US a very large region, with an extremely variable topography, the macrozones defined to divide the whole elevation range led to even more heterogeneously distributed groups of macrozones. A different subdivision of the elevation range, differentiated across the country, may solve some of the present issues.

Future studies on this line will also focus on the application of similar frameworks to other semi-distributed or fully-distributed hydrological models, able to explicitly simulate interaction between sub-basins: in fact, this would allow to detect the direct contribution of each macrozone to the total discharge, rather than to the runoff generation only.

Furthermore, the approach will be reapplied dividing the US dataset in subregions and considering also to differentiate the elevation ranges over such regions.

6.6 Appendix A: Applying MS-OA to $\mathrm{TUW_G}$

This section recalls the main steps of the MS-OA technique when adapted to the regionalisation of the runoff generation model $\mathrm{TUW_G}$ for each macrozone of an ungauged catchment:

1. n donors (i.e. elevation macrozones, where the runoff generation parameters are assumed to be known and are denoted $P_{G,i}$) are selected based on a dissimilarity index Φ (Eq. 1.2), which is defined through the "distance" in the normalised attributes space;

2. the $\mathrm{TUW_G}$ model is simulated for the ungauged macrozone, using the entire parameter set $P_{G,i}$ of each donor i ($i = 1, \ldots, n$);

3. the simulated runoff generation values are weighted-averaged, according to the dissimilarity between the donor and the ungauged macrozone:

$$\Delta S_{UZ,U}(t) = \sum_{i=1}^{n} w_{i,U}\, \Delta S_{UZ}(t, P_{G,i}) \tag{6.3}$$

   where $\Delta S_{UZ,U}$ is the regionalised runoff production at the ungauged macrozone U and $w_{i,U}$ is the weight associated to donor i, calculated as function of the dissimilarity index (see Eq. 3.57).

It is clear that such method requires the choice of two settings: i) the number of donors n and ii) the the set of attributes used in the dissimilarity measure Φ to quantify the distance in the attribute space between target and donors.

6.7 Appendix B: Choice of the number of donors for the US region

The implementation of the MS-OA technique is based on the concept of output-averaging. The number of donors used in the methods is fixed previously to the optimisation of the best attribute set to define Φ.

In Austria the donor number is fixed to three, based on previous analysis. In USA, a preliminary test is performed to test the effect of the number of donors. It consists in the application of the MS-OA approach for the regionalisation of the entire set of parameters (and not only those governing runoff generation) optimised in the benchmark approach

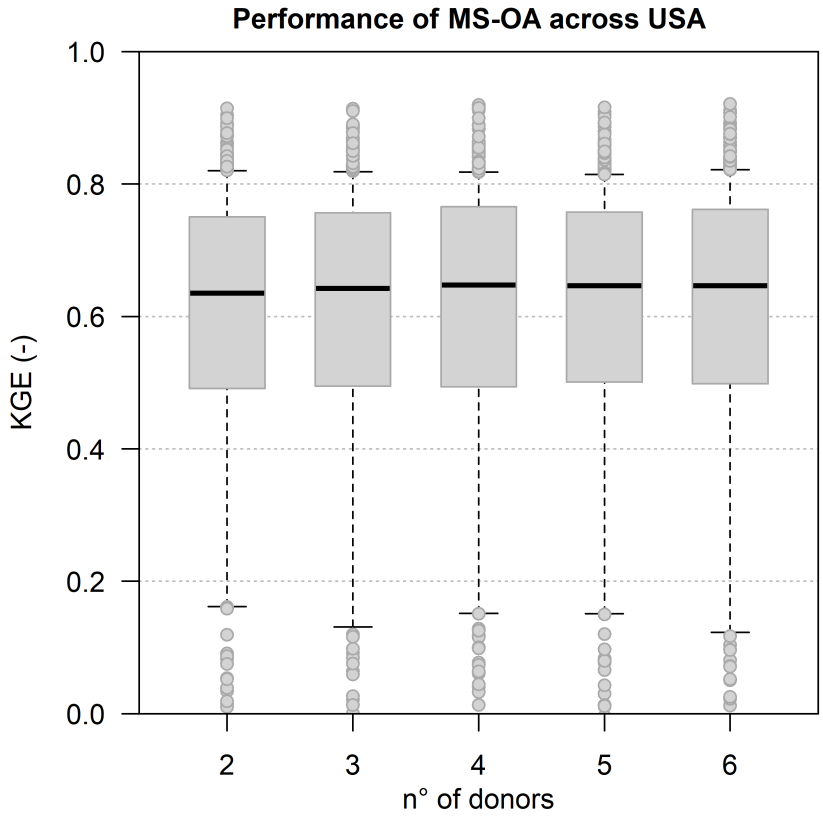

Figure 6.18: Optimised accuracy of the modified MS-OA technique applied to the entire TUW model parameter set for different number of donor catchments. Boxplot whiskers refer to 10% and 90% quantiles.

(TUW-parU), thus for the *classic* regionalisation of the whole TUW model (as operated in Austria for the analysis of the previous chapter).

The choice of the best attribute set of Sec. 6.4.1 is repeated for different numbers of donor catchments using the same settings, except for the objective function which is the interquartile KGE between simulated and observed discharges at basin outlets (and not related to runoff production). Number of donors between two and six are considered. Fig. 6.18 shows the optimised performance of the modified MS-OA approach across the entire US region: the optimisation process reaches very similar results in all cases. So, the number of donors is set to three, to be consistent with the Austrian dataset.

Chapter 7

Exploring the potential of transfer entropy for catchment dynamics characterisation and classification

7.1 Introduction

As introduced in the Premise of the Dissertation, this Chapter presents a stand-alone analysis started during the research period at the Water Resource and Hydrologic Modeling Lab of McMaster University (Ontario, Canada). The work is the results of the several tests made to explore the possible use of Information Theory for enhancing our knowledge about hydrological similarity.

Catchment classification has always been addressed as one of the essential steps for transferring information between similar watersheds, through the identification of the dominant hydrological processes and their main characteristics. The delineation of similar groups of basins is required for several regionalisation applications (Rosbjerg et al., 2013).

Whichever is the clustering method used (for a concise summary of the most common techniques, see e.g. Razavi and Coulibaly, 2013), most of the studies implement streamflow indices and signatures as metrics for classifying the watersheds, looking at the similarity between some of the hydrograph features (e.g. Sawicz et al., 2011; Yaeger et al., 2012; Archfield et al., 2014). In the case of ungauged basins, climatic and physical characteristics are typically used instead (e.g. Knoben et al., 2018; Swain and Patra, 2019). Some studies use the clusters obtained through streamflow signatures (gauged case) to train the classification algorithm in the ungauged case (see, e.g. Toth, 2013; McManamay and Derolph, 2019; Jehn et al., 2020).

Clustering algorithms based on streamflow signatures have been implemented, for in-

stance, for improving regional flood frequency analysis (e.g. Castellarin et al., 2001; Merz and Blöschl, 2005; Rao and Srinivas, 2006), for assessing water availability at annual or seasonal scale (e.g. Holmes et al., 2002; Viglione et al., 2007; Berghuijs et al., 2014a) or for low flow statistics estimation (e.g. Laaha and Blöschl, 2006; Vezza et al., 2010). However, less attention has been given to the delineation of homogenous groups of catchments for the transfer of information regarding the rainfall-runoff processes at fine temporal scale (needed for the regionalisation of hydrological model parameters), where it is essential to take into account the dynamic of streamflow generation and propagation, considering the features of the entire runoff hydrograph and its forcings. To address this issue, a few recent studies have focused on different classification metrics that could synthesise the temporal correlation structure of streamflow processes: some of them proposed to analyse similarity of runoff temporal dynamics through the parameters of linear models estimating the global autocorrelation function (ACF) of the streamflow time series (De Thomasis and Grimaldi, 2001; Chiang et al., 2002; Grimaldi, 2004; Corduas, 2011); others used single autocorrelation coefficients for regionalisation purposes (e.g. Montanari and Toth, 2007; Castiglioni et al., 2010; Lombardi et al., 2012). More recently, notable contributes were given by Singh et al. (2016), who used a "data depth" function to explore similarity of the whole dynamics of streamflow time series for the transfer of rainfall-runoff model parameters, or by Pérez Ciria and Chiogna (2020) who applied a classification framework based on the analysis of daily streamflow at multiple temporal scales with Discrete Wavelet Transform. Alternatively, functional analysis can be used to represents different hydrological regimes as functions which enable the exploitation of the full information stored in the time series or annual hydrograph when clustering catchments (e.g. Brunner et al., 2020). Toth (2013) provided the first ever watershed clustering including measures of the fine time-scale variability and correlation structure of both rainfall and streamflow series, including a correlation scaling exponent to classify the shape of the hydrograph ACFs, but such indexes are obtained independently for each one of the two time series.

To the best of our knowledge, no studies have so far considered the interaction between the entire time series of forcing data (e.g. precipitation) and streamflow, quantifying it through measures to be used as clustering metrics. Such measures may be effective for assessing the similarity of the main hydrological processes taking places in different watersheds, especially useful if one aims to transfer information regarding rainfall-runoff dynamics as, for instance, when regionalising hydrological model parameters. Linear cross-correlation analysis, which measures the correlation of two time series as function of the displacement (in time) of one relative to the other, may measure the similarity of different forcings and streamflow signals in order to identify the dominant factors

guiding the runoff generation. However, given the high nonlinearity of the rainfall-runoff relationship, the use of a metric able to look beyond linear correlation analysis between meteorological inputs and streamflow would be required. One of the potential approach for this purpose is the use of the concepts of the *information theory*: they are based on the notion of *entropy*, i.e. the content of information of a signal (as a time series), or, in the multivariate case, the content of information shared between more variables (i.e. *mutual information*). An entropy-based measure of particular interest is the so-called *transfer entropy* (Schreiber, 2000), a time asymmetric quantity which analyses the interaction between different signals.

Information theory-based approaches have been widely used in hydrology and water resources during the last decades. Singh (1997) reports a detailed overview of the development of these methods, describing typical application in hydrological sciences. Recent studies made use of entropy terms derived from Shannon's information theory (Shannon, 1948) for identifying hydrological similarity. For example, Rajsekhar et al. (2013) used *direction information transfer* (standardised version of mutual information) to regionalise drought characteristics. Loritz et al. (2018) made use of normalised mutual information between the simulated streamflow time series of sub-catchment nested within the same downstream catchment for clustering hillslope models into functional groups of similar runoff generation and implemented Shannon entropy as measure of diversity in simulations. Such studies applied the entropy concepts for grouping and characterising discharge time series across the considered study regions.

More interestingly for the present study, Bennett et al. (2019) demonstrated how the above cited quantity called *transfer entropy* can be used to quantify the active transfer of information between hydrologic processes at various timescale and help understanding the dynamic of rainfall-runoff system, in the perspective of evaluation of model behaviour. They measured the connectivity of simulated hydrological processes in single catchments for different rainfall-runoff model structures.

Based on the findings of Bennett et al. (2019), in this study the concept of transfer entropy is applied for identifying the dominant hydrological processes occurring in a catchment, measuring the transfer of information from different forcings to streamflow time series. Then, similarity between dominant processes is used for grouping similar catchment dynamics. In a first step, the different amounts of information transferred from the three main meteorological forcing variables (i.e. precipitation, snow melt and actual evapotranspiration) to observed runoff are computed through transfer entropy. For all the study watersheds the estimated transfer entropy values are then used as signatures to characterise catchment dynamics. Then, a classification of the basins is obtained as-

suming that similar values of transfer entropy for the three forcings identify similar basins, through a simple hierarchical clustering algorithm. The methodology is tested on the large and dense Austrian dataset (Sec. 2.1). The classification results are compared against the values of a set of typical streamflow signatures across the clusters. The "traditional" streamflow signatures are also used as attributes in a benchmark classification method, which is also compared to the clusters obtained through the transfer entropy values.

The purpose of the present study is to show the potential of transfer entropy for characterising and classifying catchment dynamics as alternative to typical runoff signatures and flow indices.

The Chapter is organised as follows: Sec. 7.2 gives the basic principles of information theory based on Shannon entropy, defines its terms, and finally introduces the concept of transfer entropy. In Sec. 7.3, the basins are characterised based on the information flow from meteorological forcings to runoff, estimated through transfer entropy, and the results are analysed against catchment characteristics and against a set typical streamflow signatures. In Sec. 7.4, transfer entropy values are used as similarity features to classify different hydrological dynamics. In Sec. 7.5, a benchmark classification method based on streamflow signatures is implemented and compared to the resulted transfer entropy classification. Finally, Sec. 7.6 reports the concluding remarks.

Research question

Are we able to quantify the interaction between meteorological forcing and river discharge by relying on the concepts of Information Theory? Is such measure representative of the hydrological behaviour of a watershed, improving our knowledge about catchment similarity?

7.2 Entropy theory and transfer entropy

Information theory was initially developed in the field of communication engineering. In information theory, entropy quantifies the uncertainty associated to a variable. By extension, entropy is a measure of the amount of information content in a generic signal. Shannon entropy, $H(X)$, provides a mathematical formula for explaining the information content from a variable X (e.g. the streamflow time series at a given gauge), which has a set of discrete probabilities, $p_1, \ldots, p_n$ (Shannon, 1948; Singh, 1997). The average information content associated to X is called *marginal entropy* and it is given as:

$$H(X) = -\sum_{i=1}^{n} p(x_i)\, log_2\, p(x_i) \tag{7.1}$$

where n is the total number of class intervals (also called bins), and $p(x_i)$ is the occurrence probability of X in the ith class interval. Strictly speaking, assuming that one bit is the information content of a binary random variable that is 0 or 1 with equal probability, Shannon's entropy measures the number of bits needed to optimally encode a sequence of realisations of X. If, for instance, X is a constant signal (i.e. it has a unique known value) the probability of that event (i.e. the probability that X assumes the constant value) will be one, while all the other probability will be zero; in this case there will be no uncertainty associated to X, and the information content $H(X)$ will be zero. On the other hand, if X has a uniform distribution (i.e. the probability of each event is equal to $1/N$) the marginal entropy assumes its maximum at $log_2 N$, since:

$$H(X) = -\sum_{i=1}^{n} \frac{1}{N} log_2 \frac{1}{N} = N \frac{1}{N} log_2 N \tag{7.2}$$

In the multivariate case, the *joint entropy* represents the whole amount of information embedded in more variables. In case of two signals X and Y, and it is defined as follows:

$$H(X,Y) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) log_2 p(x_i, y_j) \tag{7.3}$$

The marginal entropies of a number of signals often contain duplicated information, so that the joint entropy should be less than the sum of marginal entropy unless every variable is independent of each other (Keum et al., 2017). Consequently, it is possible to measure the portion of the information content shared between the two variables X and Y. It is named *mutual information* (Cover and Thomas, 2005):

$$I(X,Y) = H(X) + H(Y) - H(X,Y) \tag{7.4}$$

Mutual information quantifies the knowledge we gain about Y by measuring the variable X, or vice versa. Conditional forms of the above presented measures of information can be also defined. The *conditional entropy*, i.e. the additional information provided by the entire time series X if we already know the information content of Y, is given by:

$$H(X|Y) = H(X) - H(X,Y) \tag{7.5}$$

Similarly, *conditional mutual information* (given a third variable Z) can be also defined as the information content shared between X and Y, provided we already know the information content of Z:

$$I(X,Y|Z) = H(X|Z) + H(Y|Z) - H(X,Y|Z) \tag{7.6}$$

All the above mentioned quantities are symmetric, except for the choice of the conditioning variables. They are not describing the flow of information between variables; in other words, we do not know if a variable is somehow influencing the others and vice versa.

Schreiber (2000) developed a method for accounting for information transfer, which considers the flow of information in time between signals. In order to do so, variables are artificially "shifted" in time: given two original variables X_t and Y_t, let's consider X_{t-lx} and Y_{t-ly} as new time series, built by shifting X_t and Y_t respectively by lx and ly time lags. The quantity, called *transfer entropy*, can be written as a particular case of conditional mutual information:

$$T_{X\to Y}(lx, ly) = I(Y_t, X_{t-1}, \ldots, X_{t-lx}|Y_{t-1}, \ldots, Y_{t-ly}) \tag{7.7}$$

where $T_{X\to Y}$ is the transfer entropy from X to Y associated to time lags lx and ly: it represents the mutual information between the target variable Y at time t and the dependent variable X at all the previous lx time lags, conditioned by the knowledge of the history of variable Y itself at the previous ly time lags (in case of zero lag, it would lose its "directional" nature, corresponding to the simple mutual information between X and Y). More intuitively, transfer entropy can be thought as the additional knowledge we gain about the variable Y at time t by measuring X at the previous lx time steps, with respect to the information already given by knowing the previous ly states of Y. It quantifies the effective information flow from X to Y, given fixed time lags.

For a more rigorous treatment and definition of transfer entropy, see Schreiber (2000). The reader can find other applications of transfer entropy as tool for causal effects estimation, for instance, in Ruddell and Kumar (2009), who used it for quantifying information flows within ecohydrological systems, or in Hlinka et al. (2013), who employed transfer entropy to estimate causality of climate networks.

It has been underlined in the literature that, when choosing the values of the two time lag parameters lx and ly, there is a trade-off between computational complexity and estimation accuracy and stability (Schreiber, 2000). Hlinka et al. (2013) showed how the reliability of transfer entropy estimate decreases for increasing values of time lag, due to the difficulties in the estimation of high-dimensionality probability distributions.

7.3 Catchment characterisation with transfer entropy

The aim of the study is to show the use of transfer entropy (TE) as a catchment attribute for identifying dominant hydrological processes and for the characterisation of

different types of catchment dynamics.

In this experiment, we decided to consider the information flow transferred to the observed daily streamflow time series, from the observed daily precipitation and from the other two main components responsible for the runoff generation: daily actual evapotranspiration and snow melt, chosen due to their dominant impact on the water balance, similarly to what done also by Bennett et al. (2019).

The approach we propose in this work is based on the following hypotheses: i) the flow of information between catchment forcing data (including the main contributes to runoff generation, i.e. rain, snow and evapotranspiration) and the observed streamflow can be quantified with its transfer entropy, ii) if two or more catchments have similar values of information flow (quantity of transferred information), they are similar and the rainfall-runoff transformation is dominated by similar processes.

This section describes and reports the results of the first phase of the methodology which consists in the characterisation of each catchment with transfer entropy values. First, actual evapotranspiration and snow melt time series are estimated through the application of a rainfall-runoff model. Then, transfer entropy values from precipitation, actual evapotranspiration and snow melt are computed for all the study catchments. A first interpretation of the results is conducted analysing TE values against catchment characteristics and a set of typical streamflow signatures.

In the second and last phase of the approach, which will be presented in the following Sec. 7.4, a simple hierarchical clustering algorithm is applied to the standardised values of transfer entropy and results for different number of clusters are analysed.

7.3.1 Estimation of snow melt and actual evapotranspiration through rainfall-runoff model application

As introduced previously, two of the dependent variables considered as sources of information to be transferred to catchment streamflow are actual evapotranspiration and snow melt. As in most cases, these quantities need to be estimated, since they are very difficult to measure. For this purpose, it was decided to use the TUW model, which was observed to behave very well in the study region based not only on the analysis of this Dissertation, but also on a number of previous applications (see, e.g. Merz and Blöschl, 2004; Parajka et al., 2005). The model is calibrated against the observed runoff and it is used exclusively for extracting the estimated actual evapotranspiration and snow melt at daily scale.

The model is calibrated and run at daily time steps for all the study catchments using the same structure and settings implemented for the analyses presented in Chapter 4

with the difference that, for this application, the 33 years of observation (1976-2008) were entirely used for model calibration, with a warm-up period of one year. We decided not to use a split-sample procedure, since we know from the previously presented analyses that the model behaves well, and we preferred using the entire record both for estimating the model parameters and for simulating the snow melt and actual evapotranspiration state variables in order to have the longest possible time series for the TE-HC analysis. The resulting simulation efficiencies are very good: Kling-Gupta Efficiency ranges between 0.56 and 0.94 over the catchments, with a median of 0.85, while Nash Sutcliffe Efficiencies are between 0.47 and 89, with a median of 0.71.

7.3.2 Computation of transfer entropy

Having simulated through the model the actual evapotranspiration and snow melt time series, the flow of information from catchment forcing data to the observed streamflow is quantified, computing the transfer entropy from each of the three daily time series of:

- precipitation (P)
- actual evapotranspiration (AET)
- snow melt ($melt$)

to the observed streamflow (Q).

The use of the transfer entropy introduced in Sec. 7.2 requires the preliminary choice of the values of the time lags lx and ly to be used in the computation. As state above, high time lags negatively influence the reliability of the estimates, due to problems in the estimate of high-dimensionality probability distributions. In order to minimise these effects, but also to keep the methodology as simple as possible for evaluating the potential of transfer entropy in rainfall-runoff dynamic classification, in this study it was decided to consider only the previous time step ($lx = ly = 1$), thus assessing the information transfer bewteen the forcing variable at day $t - 1$ and the streamflow at day t. It is of course clear that the target variable (i.e. daily streamflow time series Q) is influenced by the dependent variables (P, AET and $melt$) also at time lags greater than one day; however, choosing to consider just one single lag allows us to have a first order estimate of the impact of the different variables on the streamflow, and at the same time to simplify the TE calculation, for a better understanding of the information flow. Thus, for this parametrisation, the definition of transfer entropy (from Eq. 7.7) becomes:

$$T_{X \to Y}(1,1) = I(Y_t, X_{t-1} | Y_{t-1}) \tag{7.8}$$

where Y is the daily streamflow Q and X can be either P, AET or $melt$. The computation of transfer entropy has been carried out within the R Programming Environment (R Core Team, 2019), by using the package *RTransferEnropy* (Behrendt et al., 2019).

Transfer entropy, as every information theoretic measure, requires the estimation of probabilities of occurrence. Such estimate, which is a potential source of uncertainty, can be performed following various techniques. One option is the use of continuous distribution functions (e.g. Krstanovic and Singh, 1992), and in case of multivariate data the choice of the distribution is limited to normal or lognormal distribution, for which joint entropy can be calculated as a function of the covariance matrix (Ozkul et al., 2000), otherwise no equations are available to estimate multivariate entropy quantities. Alternatives are, for instance, the use of non-parametric density estimators (e.g. Mishra and Coulibaly, 2009) or the use of discrete probability distributions which define a number of class intervals to approximate the probabilities by the corresponding relative frequencies.

Entropy calculations using discrete distributions have been preferred in the recent specialised literature, due to the unavoidable assumptions at the basis of the chosen distribution functions and to the above mentioned difficulties in formularising the joint entropy in many distributions (Fahle et al., 2015). However, the discrete entropy term calculations also require an assumption of data quantisation and there is no consensus on which method should be used (Keum and Coulibaly, 2017). Among the simplest methods, for example, the *histogram method* divides the range of the variables in class intervals with equal width, while the *quantile method* uses fixed quantiles as bin boundaries and it is based on the notion that you want to have the tail events (i.e., extreme events) in separate bins. In general, the choice of the intervals/quantiles should be guided both by the distribution of the data and by the number of observations.

In this case, the authors of the R-package recommend that the number of bins is limited in order to avoid too many zero observations when calculating relative frequencies as estimators of the joint probabilities in the (effective) transfer entropy equations. The default version of the tool set the bin limits to the 5% and 95% quantiles isolating extreme tail events. On the other hand, the literature on the use of transfer entropy for purposes similar to our analysis is extremely limited, and, to the best of our knowledge, no reference on the quantisation method is provided. For this experiment, the effect of different bin widths and types was investigated: as expected, the results (not reported here) vary between different quantisation methods.

Finally, it was decided to divide the data interval into five bins based on the quantile method defining a large central interval between the quartiles, and dividing the tails of the distribution into two intermediate intervals (between 5% and 25% and between 75%

and 95% quantiles respectively) and two "extreme" intervals (outside the 5% and 95% quantiles).

Another complication is given by the fact that transfer entropy estimates are known to be biased due to small sample effects. To address this issue, the comparison of the results with those obtained with a shuffling procedure may be used: shuffled versions of the independent variable $X_{shuffled}$ can be generated, randomly drawing values from the time series of X and realigning them to generate new time series. Then, the corresponding "shuffled" transfer entropies can be computed $T_{X_{shuffled}\to Y}$: this procedure "destroys" the time series dependencies of X, as well as the statistical dependencies between X and Y, and it thus represents a "white noise" against which the calculated transfer entropy can be benchmarked. The implemented R tool allows the automatic computation of the *effective transfer entropy* (Marschinski and Kantz, 2002): in order to derive a consistent estimator, shuffling is repeated many times and the average of the resulting shuffled transfer entropy estimates across all replications $\overline{T}_{X_{shuffled}\to Y}$ is subtracted from the Shannon transfer entropy to obtain a bias corrected estimate:

$$ET_{X\to Y} = T_{X\to Y} - \overline{T}_{X_{shuffled}\to Y} \tag{7.9}$$

The number of shuffles is here set to 100. More details, may be found in the *RTransfer-Enropy* manual.

In this paper, transfer entropy is always computed in the effective form of Eq. 7.9. However, for the sake of simplicity, we will still refer to it as $T_{X\to Y}$ (the lag specification is also omitted and it is always 1).

The computation is carried out for all the 209 basins and for the three independent forcing variables. At the end of the process each watershed is characterised by three values of transfer entropy which estimate the information transferred from each of the independent variables to the observed runoff.

We are estimating the impact (i.e. the net information flow) of the dependent variables at the previous day: we expect that high values of transfer entropy from precipitation will be associated to watersheds with the fastest response times; at the same time, snow dominated catchments will probably show the highest impact of snow melt on the generation of runoff.

Fig. 7.1 shows the values of the standardised transfer entropies across the basins of the study region ($T^*_{P\to Q}$ in panel a, $T^*_{AET\to Q}$ in panel b and $T^*_{melt\to Q}$ in panel c). The spatial pattern of such results is analysed in detail in Sec. 7.4.1, together with the catchment classes which will be obtained at the end of the TE-HC methodology. At this point, we will just verify the presence of any potential relationship between the computed TE values

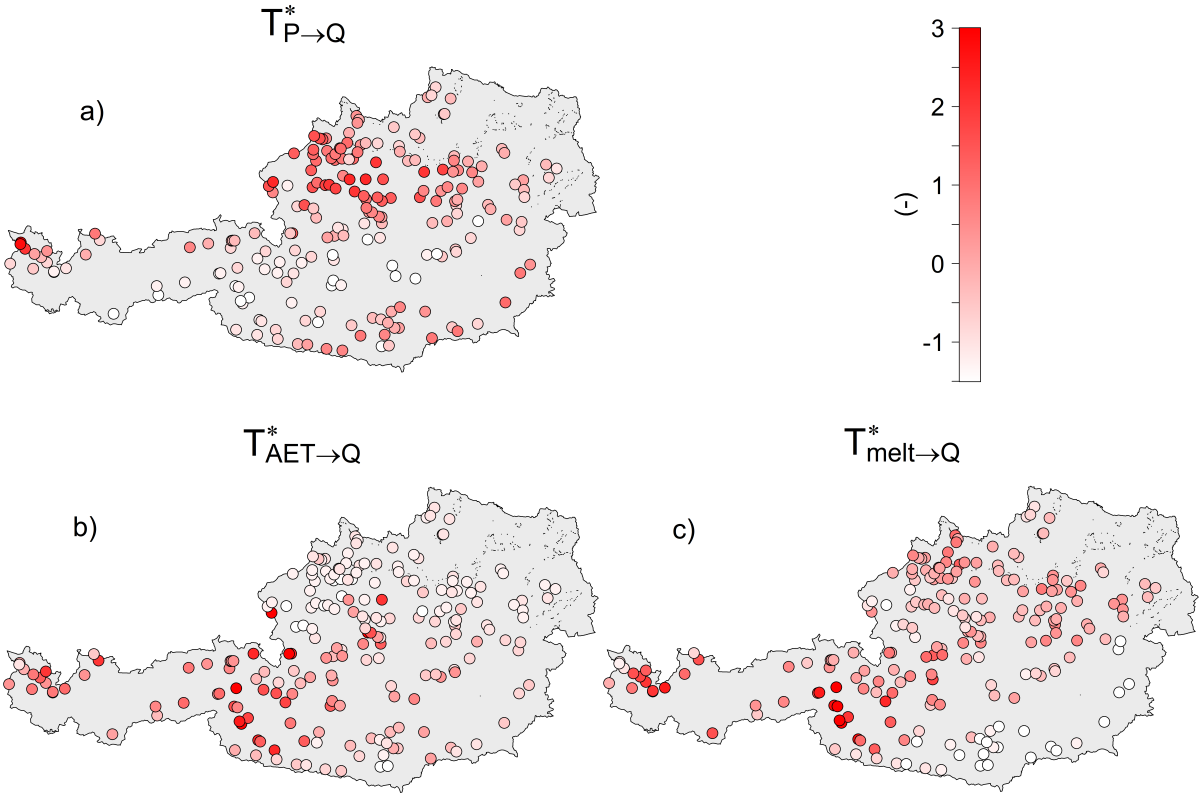

Figure 7.1: Standardised transfer entropy values across Austria: estimated information flow from a) precipitation, b) actual evapotranspiration and c) snow melt to the streamflow.

and a set of catchment features and streamflow signatures.

7.3.3 Analysis of TE values against catchment features

The interpretation of the resulting values of transfer entropy is not straightforward. In order to better understand how they are are linked to catchment features, Fig. 7.2 reports the scatterplots of each of the three obtained T^* against most of the basin attributes described in Tab. 2.1. It can be noticed that drainage area (*Area*), river network density (*RiverD*), forest fraction (*Forest*), solar irradiation (*Irrad*) and fraction of catchment with porous aquifers (*AcqPort*) are not clearly linked to any of the TE values. On the other hand, mean elevation (*Elev*), snow fraction (*SnowF*) and snow depth (*SnowD*), which are strongly correlated between each other, show the a consistent relation with the three entropy components: as expected, basins at high elevation, characterised by significant presence of snow, correspond to low values of transfer entropy from 1-day lag precipitation, while for decreasing altitudes (and decreasing *SnowF* and *SnowD*) the range of variability of $T^*_{P\to Q}$ increases and the relationship is less clear; $T^*_{AET\to Q}$ and $T^*_{melt\to Q}$ are instead positively correlated with elevation and presence of snow: looking at the remaining features, basin slope (*Slope*) is negatively (but not significantly, i.e. low p-value) correlated

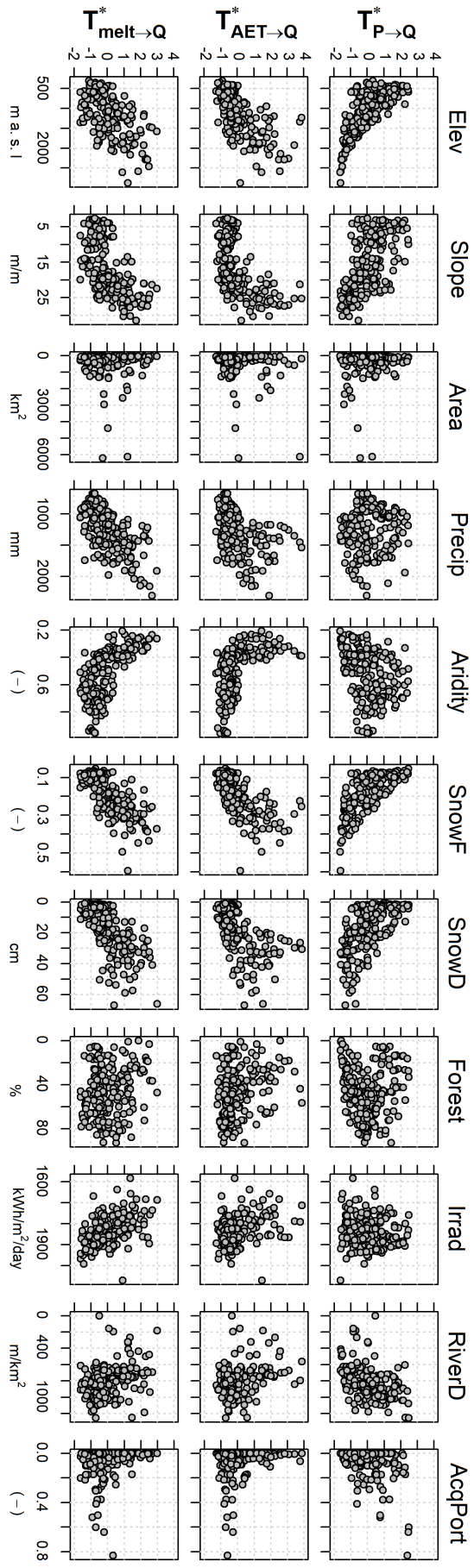

Figure 7.2: Standardised values of transfer entropy against basin characteristics.

with the impact of precipitation P and we can see that flatter catchments corresponds to low impact of AET and *melt*: this can be explained by the fact that slope is also related to elevation (i.e. higher catchments are generally steeper). The impact of P is not clearly related to other climatic attributes, but annual precipitation (*Precip*), higher in mountainous basins due to the orographic component, is positively correlated with the information flow from *melt*. More arid catchments are characterised by low information from actual evapotranspiration and snow melt. In general, from this analysis we can observe that in Austria the differences in transfer entropy tend to be explained mainly by those catchments attributes linked to elevation. On the other hand, for instance, no clear relationships between fast response time (high impact of 1 day-lag P) and the analysed catchment characteristics can be detected.

7.3.4 Analysis of TE values against streamflow signatures

Before implementing the obtained transfer entropies to classify Austrian catchments, the single TE values are compared also to the flow characteristics of the basins. Viglione et al. (2013) used six typical runoff signatures to characterise streamflow regimes at different time scales across the same Austrian dataset. As stressed in the Premise of this Dissertation, hydrological runoff signatures synthesise in indexes different flow conditions and aspects, i.e. portions of the rainfall-runoff processes. Here the purpose is to look for any potential relationships between such measures and the transfer entropies which, on the other hand, represent the interaction between forcing variables and runoff.

The signatures chosen for the study are:

- mean annual specific runoff (mm/yr):

$$\overline{Q}_y = 365\,\overline{Q}_d = \frac{365}{T}\sum_{t=1}^{T} Q_d(t) \tag{7.10}$$

  where $\overline{Q}_d$ is the mean daily specific runoff (mm/day) and T (days) is the record length; this index estimates the long term water availability of a catchment. More humid basins correspond to higher values of mean annual runoff, and viceversa.

- the range of the Pardé's coefficients Par_i (defined as the mean monthly runoff $\overline{Q}_i$ for month i divided by the mean annual runoff $\overline{Q}_y$):

$$Par_i = \frac{\overline{Q}_i}{\overline{Q}_y} \tag{7.11a}$$

$$\Delta Par = \max(Par_i) - \min(Par_i) \tag{7.11b}$$

This signature is an indicator of runoff seasonality: since each Parde's coefficient refers to the deviation of monthly runoff in respect to the annual average, their variability estimates how the runoff is distributed along the year. For instance, snow dominated catchments have a high values of such range, since the runoff is so much higher in the periods of the year when snow melt occurs;

- the slope of the flow duration curve (%/%) defined as:

$$mFDC = 100\,\frac{Q_{30\%} - Q_{70\%}}{40\,\overline{Q}_d} \tag{7.12}$$

where $Q_{30\%}$ (mm d^{-1}) is the value of daily runoff which is exceeded 30% of the time and $Q_{70\%}$ 70% of the time. Thus, $mFDC$ is a measure of slope of the central part of the flow duration curve and indicates the percentage of increase of runoff, with respect to the annual mean, for 1% decrease of exceedance probability; in other words, it estimates the variability of the central values of the distribution, excluding the extremes: seasonal watersheds have steep FDC, while catchments governed by groundwater contributions or more flashy catchments have flat FDC (in the central part);

- normalised low flow statistic q_{95} (-), calculated as the value of daily runoff $Q_{95\%}$ (mm d^{-1}) which is exceeded the 95% of the time divided by the mean daily runoff $\overline{Q}_d$:

$$q_{95} = \frac{Q_{95\%}}{\overline{Q}_d} \tag{7.13}$$

Catchments with higher q_{95} are typically characterised by pervious soils and are able to self-regulate the runoff, while more flashy catchments and highly seasonal watersheds have low q_{95};

- normalised high flow statistic q_{05} (-) calculated as the value of daily runoff $Q_{95\%}$ (mm d^{-1}) which is exceeded the 5% of the time divided by the mean daily runoff $\overline{Q}_d$:

$$q_{05} = \frac{Q_{05\%}}{\overline{Q}_d} \tag{7.14}$$

Such index is high in non self-regulating catchments and where the rainfall-runoff response is particularly fast, but also in watersheds with high seasonality;

- the integral scale $\tau_{1/e}$ (days) calculated as the time lag at which the autocorrelation function drops below $1/e \approx 0.368$. The integral scale is a raw measure of the runoff hydrograph memory (see e.g. Blöschl and Sivapalan, 1995, p. 255 and reference

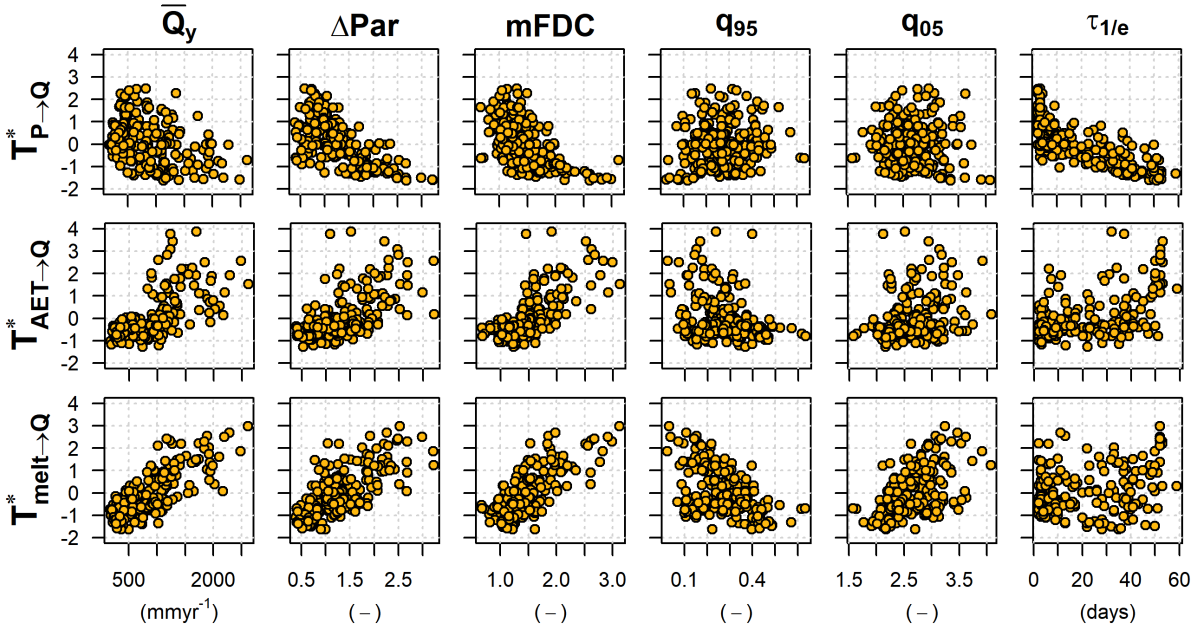

Figure 7.3: Standardised values of transfer entropy against runoff signatures.

therein). Thus, the more seasonal is a catchment, the higher is its hydrograph memory; in contrast, the response of more flashy catchments is highly depending on single rainfall events, and their memory is low.

The six streamflow signatures presented above are computed for the entire dataset using all the available streamflow records. Their spatial pattern will be reported in Sec. 7.4.1 (Fig. 7.5) when compared to the final output of the TE-HC approach. Here, Fig. 7.3 shows the scatterplots between the six signatures and the three transfer entropy values. As expected, the relationships with $\overline{Q}_d$ are similar to those observed against mean annual precipitation in Fig. 7.2; ΔPar and $mFDC$, which are measures of catchment seasonality, show the same relation as observed for elevation and presence of snow (which strongly influence seasonality in Austria). No significant relations between P entropy values and low/high flow statistics (q_{95} and q_{05}) are observed; on the other hand, lower impact of 1 day-lag AET and $melt$ corresponds to the highest low flow statistics and to the lowest high flows. The hydrograph memory $\tau_{1/e}$ is weakly negatively correlated to the impact of previous day precipitation P, even if for basins with lower memory the values of $T^*_{P\to Q}$ vary between almost the entire range.

The results presented here and in the previous section show that the transfer entropy measures provide additional information, not fully explained by catchment climatic, physical and hydrological characteristics; at the same time, they are still difficult to interpret. In order to evaluate their potential in identifying dominant hydrological dynamics across the region, the TE values are then used for the classification of the catchments.

7.4 Catchment classification based on transfer entropy (TE-HC)

The second step of the methodology is the classification of the catchments based on their transfer entropy values. These values (that will be the input vectors of the classification algorithm) are previously standardised to zero mean and unit variance, so to remove the scale effect and to give them equal importance when used as distance metrics. This classification is here named TE-HC (Transfer Entropy - Hierarchical Clustering).

Distance between catchments are calculated as the Euclidean distance in the three-dimensional standardised transfer entropy space and their classification is performed using a hierarchical cluster analysis based on Ward's minimum variance method (Ward, 1963). Even if recent literature proposes more refined and sophisticated classification algorithms and procedures, several studies still use it for grouping catchments based on signatures with satisfactory results (e.g. Kuentz et al., 2017; Jehn et al., 2020). Ward's clustering is a simple algorithm which minimises the total within-cluster variance. At the first step, each element defines a single cluster. Then, at each step, the algorithm finds the pair of clusters that leads to minimum increase in total within-cluster variance after merging. This is recursively repeated until all the elements belong to one single cluster. Hierarchical cluster analysis has the advantage that it does not require the a priori definition of the number of clusters. The algorithm produces a classification tree and the number of clusters is chosen only in the second phase by cutting the tree. In this study the results for different numbers of clusters are presented.

Fig. 7.4 shows the results of the TE-HC classifications. They are reported for different number of clusters: from the top to the bottom panel they refer respectively to 3, 4, 5 and 6 clusters. Each panel presents on the left the position of the Austrian catchment outlets, whose color indicates the cluster they belong to. For each of them, a bar plot on the right (identified by the same color) reports the average value of standardised transfer entropy from each of the dependent variables. The clusters are generated by cutting the classification tree produced by the hierarchical algorithm. Thus, when adding a further cluster, one of the groups (according to the minimum variance criteria) is split into two new clusters. In order to underline this process, the split clusters are linked in the figure.

Starting from the classification with 3 clusters, it can be easily noticed how the methodology is able to capture and isolate the catchments with the greatest amount of information flow from precipitation (Cluster 2, blue dots): these basins have the fastest response and, consistently, low impact of AET (and thus low seasonality); in fact, as pointed out previously, besides their direct effect on the water balance, transfer entropy from AET

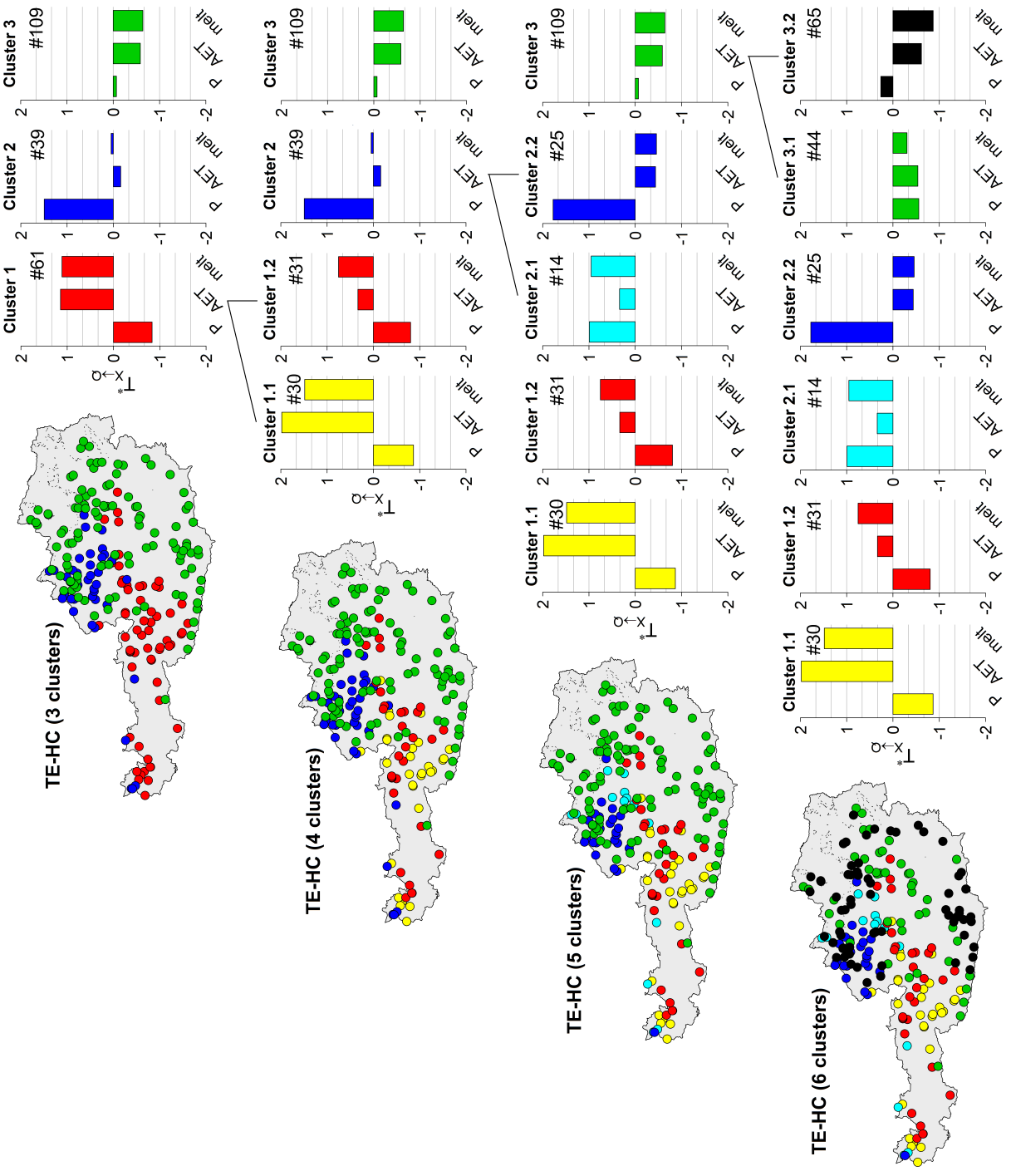

Figure 7.4: TE-HC classification results. Each line refers to a different number of clusters. The points on the map correspond to the watershed closing sections and their color to the cluster they belong to. The bar plots show the average (standardised) transfer entropy values for the cluster and for each independent variables. The size of the sample is also reported.

(highly depending on temperature) may be linked also to the runoff seasonality: catchments with high seasonality have AET and Q in phase during the year, while this is not valid in catchments with low seasonal behaviour. This is confirmed also looking at Cluster 1 (red): it identifies the snow-dominated catchments with the highest impact of both snow and AET components, and the lowest information flow from P. Consistently, most of them correspond to mountainous Alpine catchments (see Fig. 2.1), where seasonality is particularly pronounced. Finally, Cluster 3 (green) groups the basins with lowland behaviour, which have an in-between response time and weak impact of seasonality or snow, located at lower altitudes and in the eastern lowlands.

If a further cluster is considered (4 clusters classification, second line in Fig. 7.4), two new groups are generated from former Cluster 1, basically based on the magnitude of AET impact (seasonality) and on the predominant role of snow melt as runoff generation process. Cluster 1.1 (yellow) includes watershed with the highest amount of information transferred from both AET and $melt$, and with the lowest impact of 1 day-lag precipitation accordingly, completely dominated by snow. By contrast, catchments with less pronounced effect of snow related processes belong to Cluster 1.2 (red).

Moving to the classification into 5 clusters (third line in Fig. 7.4), the second cluster is split in Cluster 2.1 and 2.2 (light and dark blue respectively). This subdivision allows separating the catchments (dark blue) with the fastest response time, very strong impact of precipitation and lower seasonality and snow impact, from those (light blue) where information from precipitation is still high but the contribution of snow and AET is also very important, with T^* values similar to Cluster 1.2. In fact, Cluster 2.1 can be seen as an intermediate group between clusters 1.2 and 2.2.

Finally, considering 6 clusters (bottom line in Fig. 7.4), the still large group of catchments (109) belonging to Cluster 3 is divided: the resulting two groups (Cluster 3.1 and 3.2) differ between each other for the different contribution of snow and precipitation. In fact, in cluster Cluster 3.1 the effect of $melt$ is still close to the average and that of P is low, while in Cluster 3.2 the transfer entropy from snow contribution is minimum and the impact of precipitation is higher. To confirm this, let's look at the group of basins in the southern part of the country: from west to east the catchments change cluster as the effect of snow decreases with altitude.

7.4.1 Analysis of the TE-HC classes against streamflow signatures

In this section, the values of the six streamflow signatures presented in Sec. 7.3.4 across the study area and inside the clusters obtained through the TE-HC methodology

are analysed. Here, the results of the novel proposed classification are analysed against the pattern of such signatures. In the following section, a benchmark clustering algorithm is also applied using the six signatures as classification criteria, and the obtained clusters are also compared to the TE-HC.

Fig. 7.5 displays the spatial pattern of the signatures across the Austrian country (each point refers to the catchment closing section). In order to qualitatively analyse the streamflow behaviour inside the clusters of the classification approach introduced in this study, the figure also reports again the TE-HC resulting map (panel g). In addition, Fig. 7.6 shows the variability of the six streamflow signatures inside the six TE-HC clusters.

Looking at the mean annual runoff $\overline{Q}_y$ (Fig. 7.5a and Fig. 7.6a), which provides information about the all-year-round water availability, the majority of the most humid (and mountainous) catchments in the west and in the centre are characterised by high transfer entropy between snow melt and discharge (Fig. 7.1c) and belong to Clusters 1.1 and 1.2 (red and yellow points in Fig. 7.5g) which are those at higher elevation and, consistently, dominated by snow processes. It should be pointed out that the high runoff volume is actually more related to the orographic lifting of northwesterly airflows at the rim of the Alps rather than directly to the presence of snow itself, which is also due to catchment altitude. On the other hand, drier catchments present minimum information flow from AET (Fig. 7.1b), since high values of evapotranspiration would correspond to a reduction of basin discharge.

A marked seasonality is more directly related to snow: in fact, high values of the Pardè's range ΔPar (Fig. 7.5b and Fig. 7.6b) corresponds again to Clusters 1.1 and 1.2. In particular, the strongest differences in Pardé's coefficients, representing high seasonality, match with Cluster 1.2 (yellow points in Fig. 7.5g), whose watersheds have the largest values of transferred information from $melt$, but also from AET (see Fig. 7.1b-c and Fig. 7.4): in fact, the effect of AET is also linked to snow watersheds; nevertheless, we expected it to be even more pronounced in catchments with the most marked seasonality, given the fact that AET is highly dependent on air temperature: in general, besides the effect of the 1-day lag, the value of the AET transfer entropy will be as much higher as AET and Q are in phase during the year. In contrast, the lowest seasonality characterises Clusters 2.2 and 3.2 (blue and black boxes in Fig. 7.6b and Fig. 7.6c), which have the lowest impact of $melt$ and AET. The slope of the flow duration curve $mFDC$ (Fig. 7.5c and Fig. 7.6c) shows very similar pattern to what observed for ΔPar.

Low flow statistics q_{95} (Fig. 7.5d and Fig. 7.6d) are smaller for snow catchments where flow seasonality is more pronounced (red and yellow basins in Fig. 7.5g) and larger for low

lands catchments, in particular for Cluster 3.1 and 3.2 (green and black dots in Fig. 7.5g) where the effect of precipitation at previous lag is limited and catchments have more storage capacity (due to pervious soil). It can be appreciated that q_{95} has a positive trend

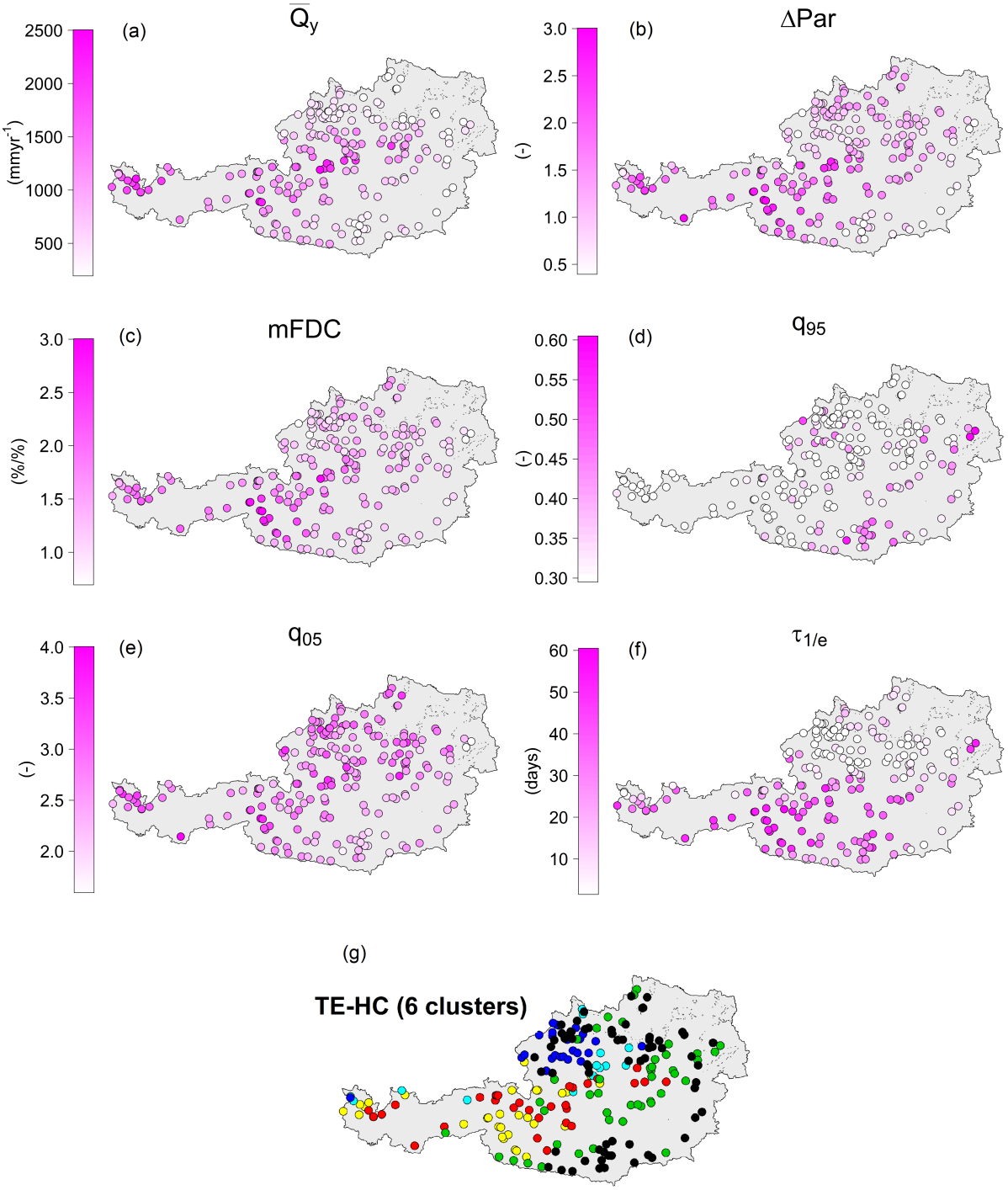

Figure 7.5: Pattern of streamflow signatures: mean annual runoff (a), range of Pardé's coefficients (b), slope of the flow duration curve (c), low flow statistic (d), high flow statistic (e) and integral scale (f), compared to TE-HC classification with 6 clusters (g). Points refer to catchment outlets.

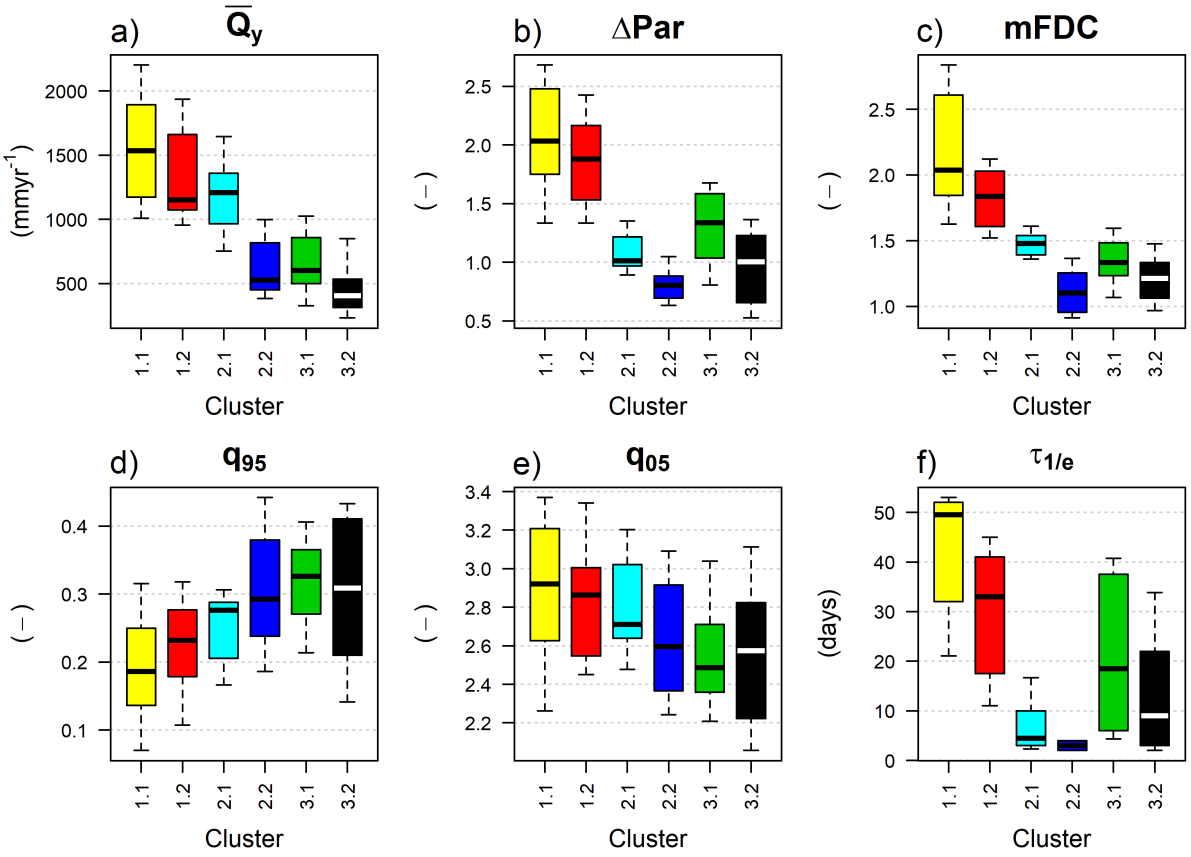

Figure 7.6: Variability of the six streamflow signatures inside the different clusters obtained through the TE-HC classification.

moving from Cluster 1.1 to 3.2 (Fig. 7.6d). For the same reasons, the high flow statistic q_{05} (Fig. 7.5e and Fig. 7.6e) presents opposite behaviour.

Finally, catchments with high values of information flow from precipitation (Fig. 7.1a) match the regions labelled as more "flashy" by previous studies (e.g. Viglione et al., 2013), showing low integral scale $\tau_{1/e}$ (Fig. 7.5f and Fig. 7.6f). In fact, $\tau_{1/e}$ estimates the hydrograph memory and is indeed minimum for the catchments of Clusters 2.1 and 2.2 (light and dark blue dots in Fig. 7.5g) which have the highest values of transfer entropy for the precipitation at previous day. The highest integral scales $\tau_{1/e}$ can be find instead in the snow catchments of Clusters 1.1 and 1.2, where we observe the minimum impact of 1 day-lag precipitation, and in the basins in southern Austria with slower dynamics and highly pervious geology, belonging to Cluster 3.1 and 3.2. However, slow response is not the primary feature of Clusters 3.2, which includes also catchments with low hydrograph memory (black points in the north-east, Fig. 7.5g), but more in general this cluster groups lowlands catchments also in the south-east with the lowest impact of snow, where the impact of 1-day lag precipitation is not uniform (the contribution of precipitation signal is nor strong nor weak when averaged).

Such analysis demonstrates how the proposed TE-HC classification can be able to capture similarity between more features of the runoff hydrograph, here represented by the selected streamflow signatures through the analysis of the role of the drivers. On the other hand, given the different nature of the similarity measure used in the classification process, we do not expect TE-HC to be able to detect the same hydrological differences between catchments that are highlighted by the considered hydrograph characteristics (as for instance for low and high flow statistics q_{95} and q_{05}). In fact, transfer entropy may be an additional instrument to characterise catchment dynamics that are not fully captured by streamflow signatures alone, by analysing the interaction between forcing variables and runoff.

7.5 Comparison of the TE-HC classes to a benchmark classification based on signatures

This last section of the Chapter compares the results obtained by the TE-HC classification to the application of a benchmark clustering approach using as classification metrics the six catchment signatures proposed in the previous section rather than the values of the transfer entropy values between climatic forcing and streamflow time series. The comparison is performed observing how watersheds are differently grouped: we do not expect identical clustering, considering that the algorithms work on conceptually different inputs.

7.5.1 Streamflow signatures classification (QS-HC)

In order to consistently compare the results, the Ward's hierarchical clustering is used again as classifier, which allows to easily evaluate the effect of different classification metrics.

Similarly to what done for transfer entropy values, signatures are standardised to zero mean and unit variance. But in this case, given the significant number of signatures that we are considering and their strong mutual correlation, a Principal Component Analysis (PCA) is performed prior to the classification process. PCA (see e.g., Krzanowski, 1988) is a multivariate analysis statistical method that enables one to obtain smaller number of uncorrelated variables from a larger number of possibly correlated variables by constructing an orthogonal basis for the original variables themselves. The derived uncorrelated variables are called principal components (PC). The full set of PC's has the same dimensionality of the original set of variables. They are ordered in such a way that the first

Table 7.1: Coefficients of the linear transformation for the first three PC's of the streamflow signatures and relative proportion of the cumulative variance explained.

Signature	**PC1**	**PC2**	**PC3**
$\overline{Q}_y$	-0.381	0.183	-0.881
ΔPar	-0.480	0.133	0.191
$mFCD$	-0.484	0.167	0.078
q_{95}	0.419	0.444	-0.187
q_{05}	-0.392	-0.504	0.094
$\tau_{1/e}$	-0.247	0.685	0.370
Cumulative variance	0.629	0.866	0.953

component accounts for as much of the variability in the original dataset as possible, and each following PC accounts for as much of the remaining variability as possible (Di Prinzio et al., 2011). Tab. 7.1 shows a summary of the analysis with the loadings of the linear transformation for the three first PC's, which allow to account for the 95% of the variance. It can be noticed that the first component (PC1) is positively associated with low flows q_{95} and negatively associated with high flows q_{05}, slope of the FDC $mFDC$, seasonality ΔPar and annual runoff $\overline{Q}_y$ (with similar loadings, around 0.4), but also to the hydrograph memory $\tau_{1/e}$ with lower loading: therefore, high positive values of PC1 indicate flashy catchments, while negative values refer to a seasonal behaviour. On the other hand, PC2 has strong positive loading with $\tau_{1/e}$ and q_{95} and strong negative contribution of q_{05} and may help to capture self-regulating catchments with slow response. Finally, most of the load on PC3 is associated to $\overline{Q}_y$ and may help to further detect very humid and mountainous catchments, characterised by higher rainfall (due to orographic effect). Such three PC's are used as classification metrics for Ward's algorithm, also maintaining the same input dimension of used for TE-HC.

Fig. 7.7 reports the results of the classification, here named QS-HC: six clusters are considered in order to be consistent with the finest TE-HC clusterisation. On the top, histograms reports the average values of the PCs (Fig. 7.7a) and of the standardised streamflow signatures (Fig. 7.7b) among the different clusters. Below the histograms, a map of the country shows the location of the catchments (referring to their outlets) where squares are colored accordingly to the resulting clusters (Fig. 7.7c). Finally, TE-HC classification into 6 clusters is reported again in the bottom for comparison (Fig. 7.7d).

Runoff volume ($\overline{Q}_y$) is greater in Clusters A and B and it decreases moving to the remaining three clusters, mainly located in the east and at lower altitudes. Similar behaviour is observed for seasonality (ΔPar and $mFDC$). In particular, Cluster A represents the

more extreme seasonal/snow behaviour and includes just a few catchments distributed along the Alpine chain. The hydrograph memory (integral scale $\tau_{1/e}$) is greater in the

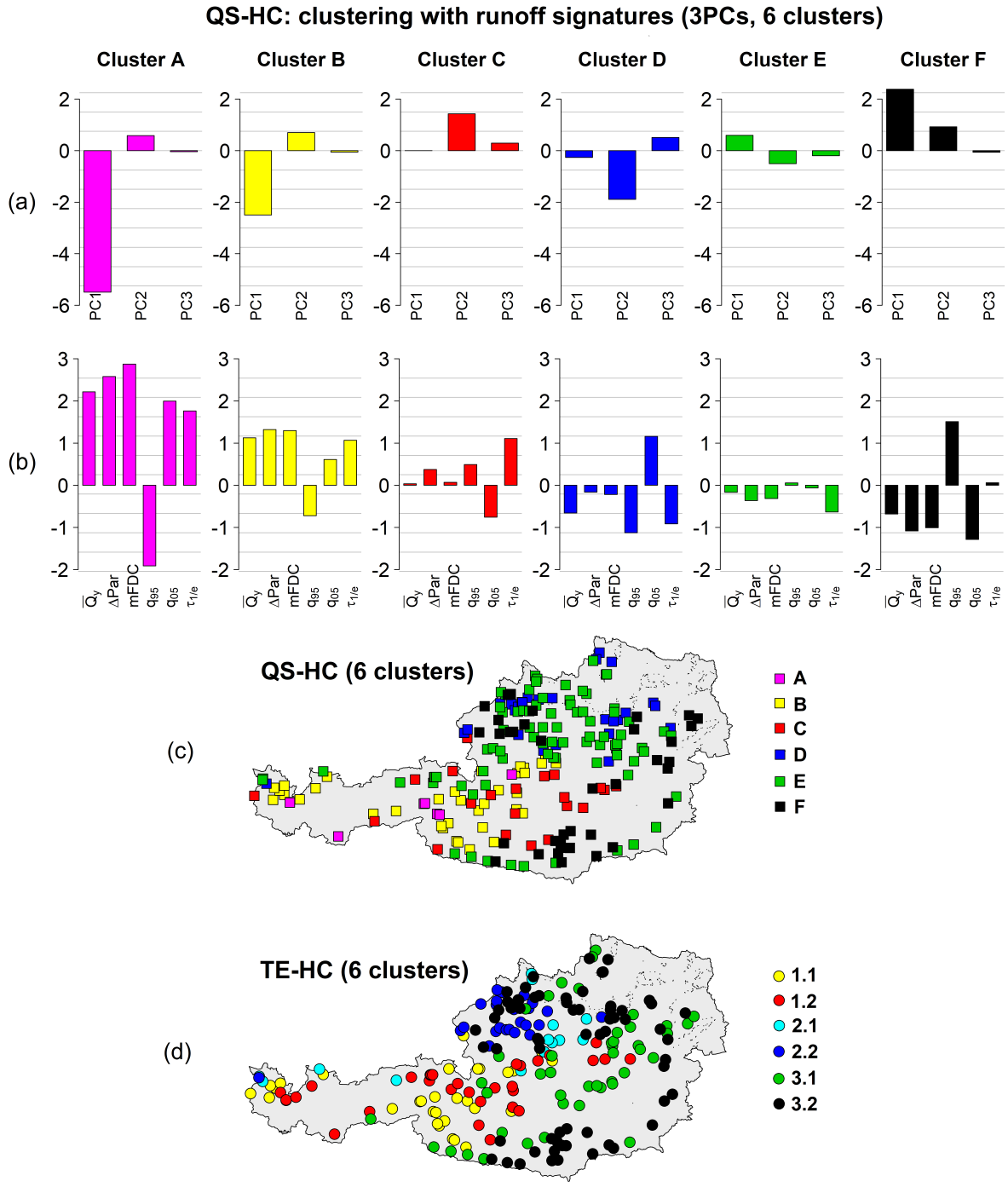

Figure 7.7: QS-HC benchmark classification with Ward's algorithm and runoff signatures' PCs. The top bar plots (a) show the average values of the PCs for all the clusters, while bar plots (b) show the corresponding values of standardised streamflow signatures. Panel (c) reports the location of the catchment outlets across the country according to QS-HC, square colors refer to clusters. Bottom panel (d) reports TE-HC classification into 6 clusters for the comparison.

Clusters A, B and C, it is minimum for Clusters D and E and it assumes values around the country average (i.e. zero, according to standardisation) in Cluster F. It appears that the variability of runoff is moderate (higher low flows and weaker high flows statistics) for the catchments in Clusters C and F, while extremes are pronounced (small q_{95} and large q_{05}) in Clusters A, B and D.

Looking more specifically at the single groups, Clusters A, B and C (magenta, red and yellow dots) include the catchments dominated by snow, given the strongest seasonality and the long hydrograph memory: Cluster A and B in particular show also the highest runoff, due to the large precipitation volume at high elevations, while C includes less extreme snow behaviours with decreasing seasonality, lower humidity (runoff volume) and less extreme low and high flows. These three groups cover most of the mountainous area of the country. Watersheds of Cluster E (green) have runoff volume similar to Cluster C but the lower hydrograph memory and the weaker seasonality suggest faster response times; moreover runoff variability increases. Cluster D and F include the drier catchments, but they differ between each others substantially for the runoff autocorrelation and for the magnitude of high and low flows: Cluster D (dark blue) has the more extreme behaviour with very low autocorrelation $\tau_{1/e}$ and very pronounced extremes (small q_{95} and large q_{05}), typical behaviour of flashy catchments; Cluster F (black) has the weakest seasonality, one of the lowest runoff volumes and a typical behavior of pervious soil watersheds, which are able to self-regulate the runoff, leading to the weaker extremes (high minimum flows and low maximum flows).

7.5.2 Comparison between TE-HC and QS-HC classifications

As expected, the classification based on streamflow signatures (QS-HC) differs from the one obtained through TE-HC, especially in low land regions of the study area. Both techniques are able to identify snow dominated catchments: TE-HC takes advantage of the information flow from *melt* and *AET* towards runoff (Clusters 1.1 and 1.2 in Fig. 7.4 or in bottom panel of Fig. 7.7), while QS-HC catches typical snow behaviour from seasonality (ΔPar) and high autocorrelation of streamflow $\tau_{1/e}$ (Clusters A, B and C). Catchments with faster response time are highlighted by the large information flow from P to runoff in TE-HC (Cluster 2.1 and 2.2 in Fig. 7.4) and by the low autocorrelation in QS-HC, see Clusters D and E. Nevertheless, it can be clearly noticed that the flashy behaviour is identified by both classifiers (i.e. dark blue Clusters 2.2 and D), even if they do not exactly match. In addition, the effect of pervious soils in lowlands catchments is more easily detected with signatures describing both extremes and hydrograph memory (Cluster F). However, it can be also noticed that Cluster F is similar to Cluster 3.2, which includes

those catchments with minimum impact of snow.

In terms of goodness of the clustering processes, the internal cluster validation could be performed with a number of different metrics. One of the simplest and more classical indexes for evaluating cluster consistency is the *Average Silhouette Width* (*ASW*). *ASW* is the average of the *silhouette* values (introduced by Rousseeuw, 1987): the silhouette value s_i is a measure defined for each catchment i of how similar it is to its own cluster compared to other clusters, ranging from -1 to 1 (for more details please see Appendix C). Here, both classifications score around 0.27, showing similar internal consistency. The obtained scores are not optimal, but good scores on an internal criterion do not necessarily translate into good effectiveness in an application.

On the other hand, the affinity of the two clustering approaches may be quantified with the *Rand Index RI* (Rand, 1971, see details in Appendix C). It varies between 1 (perfect agreement between the two partitions) and 0 (no agreement). *RI* between TE-HC and the classification based on signatures is 0.71. Hubert and Arabie (1985) proposed an adjustment to the Rand Index, which reduces its dependency on the number of clusters, called *ARI* (*Adjusted Rand Index*): its expected value is equal to zero for random partitions having the same number of objects in each class. The index is still equal to 1 for perfect agreement but it can also take negative values (and have higher discriminatory power than *RI*). In this case, *ARI* is 0.14. Such score confirms that the two finest classifications only partially match. The goodness of the two approaches could be actually verified in future by their implementation in practical applications, e.g. using the obtained clusters to support a PUB problem, as can be the regionalisation of rainfall-runoff model parameters.

In general, both procedures bring some interesting and meaningful results, which differ from each other given the different similarity measures they are based on. TE-HC is based exclusively on the interaction between the components forcing runoff and runoff itself, not considering the direct similarity in runoff characteristics as, for instance, runoff volume or flow statistics.

7.6 Concluding remarks on the use of TE for catchment dynamics characterisation and classification

A novel approach for the characterisation and classification of catchment dynamics through the quantification of the information flows between meteorological runoff forcings and runoff itself has been proposed. The central purpose of the analysis is to demonstrate the potential of transfer entropy measures for assessing the similarity in the main hydro-

logical processes taking place in watersheds. Transfer entropy considers the interaction between each one of the entire meteorological time series and it is able to quantify the amount of information transferred from them to the streamflow time series. The literature regarding catchment classification normally involves similarity measures based on different catchment signatures or geo-morphological and climatic features independently from each other. This work aimed to show how transfer entropy can be used as a complementary tool for identifying and highlighting similar catchment dynamics. This may be particularly interesting when the interaction between forcing variables and runoff needs to be understood, as for instance in the case of rainfall-runoff model calibration and regionalisation.

The research has been conducted for a very densely gauged set of watersheds covering a large portion of the Austrian country. A consolidated semi-distributed rainfall-runoff model was firstly applied at daily scale for all the study catchments and the time-series of the not measurable runoff forcings were simulated: actual evapotranspiration and snow melt.

Then, the information flow from each i) measured daily precipitation ii) simulated actual evapotranspiration and iii) simulated snow melt time series to daily runoff was estimated through the computation of transfer entropy values. In order to avoid incurring into further uncertainties due to the estimate of high-dimensional probability distribution estimate and at the same time to simplify the TE calculation, allowing a better understanding of the information flow, transfer entropy was computed in this work considering only one previous time step (1 day-lag transfer entropy). Even though runoff is certainly affected by forcing variables at greater time lags, the analysis can still provide a first order estimate of the transfer entropy and of its ability in identifying useful catchment dynamics. The obtained standardised transfer entropies were first analysed, comparing them to the values of available catchment attributes and to a set of six streamflow signatures, trying to better understand the results of the TE calculation. Then, the TE values were used as catchments features when applying a hierarchical clustering algorithm (Ward's minimum variance method). The rational of the method is that similar values of information flow between meteorological forcing and streamflow time series mean similar catchment dynamics.

Results for different number of clusters were considered. Within the limitation of the study, the results of the classification are promising. The method is able to distinguish the predominant or partial role of snow melt and evapotranspiration across the dataset. At the same time, the amount of information flow transferred from the precipitation to the runoff hydrograph, measuring the impact of rainfall events occurring in the previous

day, can help to assess differences in catchment response time.

Also, the TE-HC clustering was analysed against the pattern of the identified set of runoff signatures: the methodology demonstrates its ability to capture similarity between more features of the runoff hydrograph described by the indexes, and therefore also the potential of transfer entropy to measure interaction between forcing meteorological data and runoff.

Additionally, the TE-HC was compared to a benchmark classification method, applying the Ward's algorithm using the set of six signatures as basin features. A Principal Component Analysis was previously applied to reduce the dimension of the input variables and to account for their correlation. The obtained clusters match only in part with those of TE-HC: both algorithms are able to isolate snow dominated catchments, but, especially in the lowlands, the classes partially differ. This was expected since the algorithm are based on different similarity measures. Moreover, the study has two main limitation: the first is that transfer entropy values are calculated based on a single time lag; secondly, since the purpose was to assess the potential of transfer entropy for catchment classification in the most parsimonious way, TE-HC was implemented considering information flow just from the three main meteorological forcing components and it can not take into account all the governing hydrological phenomena.

Despite the limitations, we believe that the present analysis demonstrates the potential of TE-HC, and in particular the potential of transfer entropy, as further instrument for assessing hydrological similarity and for quantifying the connection between different processes. It should be seen as a complementary approach to classification techniques based on consolidated streamflow signatures. In particular, given that TE-HC is based on the concept of information flow, and keeping into account the role of the main forcing mechanisms, it appears promising especially when the classification aims to detect similarity in dominant hydrological dynamics, like for the transfer of rainfall-runoff model parameters, for instance. In fact, gauged classifications like TE-HC may be used to train clustering algorithm based on basin attributes (as done e.g. by Toth, 2013; McManamay and Derolph, 2019; Jehn et al., 2020) in the ungauged case.

Forthcoming experiments will focus on coupling the use of transfer entropy in catchment classification to the application of rainfall-runoff model regionalisation techniques, testing the improvement due to the use of such novel clusters as founding point of the regionalisation framework.

Moreover, future analysis will indeed require the inclusion of higher time lags in transfer entropy computation, as well as the inclusion of further runoff forcing components as independent variables.

7.7 Appendix C: Details on clustering evaluation metrics

Silhouette Width

Given a dataset $X = \{x_1, \ldots, x_n\}$ and it clustering $C = \{c_1, \ldots, c_n\}$, the silhouette width for an observation $i \in X$ is defined as follows:

$$s_i(C) = \frac{c(i) - a(i)}{max\{a(i), b(i)\}} \tag{7.15}$$

where

$$a(i) = \frac{1}{|c_i| - 1} \sum_{j \in c_i, i \neq j} d(i,j) \quad \text{and} \quad b(i) = min_{k \neq i} \Big(\frac{1}{|c_k|} \sum_{j \in c_k} d(i,j) \Big) \tag{7.16}$$

We can interpret $a(i)$ as a measure of how well i is assigned to its cluster (the smaller the value, the better the assignment): it is mean distance between i and all other data points in the same cluster c_i, where $d(i,j)$ the distance between data points i and j. On the other hand, $b(i)$ is the minimum distance of i from all points in any other cluster, of which i is not a member. The Average Silhouette Width $ASW(C)$ of the clustering C is the average the single silhouettes $s_i(C)$ across the dataset.

Rand Index

Comparing two partitions ($P1$ and $P2$) of the same data set, a couple of objects (i.e., catchments) can belong to the same class or different classes in $P1$ and $P2$. Let us define N_{00} as the number of catchments that belong to the same class both in $P1$ and $P2$; N_{10} as the number of catchments that belong to the same class in $P2$ but not in $P1$; N_{01} as the number of catchments that belong to the same class in $P1$ but not in $P2$; N_{11} as the number of catchments that belong to different classes both in $P1$ and $P2$. Under these assumptions, RI is defined as:

$$RI = \frac{N_{11} + N_{00}}{N_{11} + N_{00} + N_{10} + N_{01}} \tag{7.17}$$

Conclusions

This Thesis focuses on the development of innovative analyses to enhance the reliability of rainfall-runoff modelling in ungauged basins and the understanding of hydrological similarity. On one hand, relevant issues concerning the regionalisation of model parameters are addressed: in particular, the value of the regional informative content of the available dataset when choosing the regionalisation method and the benefit of similarity measures differentiated across elevation in a semi-distributed modelling framework. On the other hand, the similarity of rainfall-runoff transformation processes is explored through an innovative approach proposing the quantification of the interaction between the meteorological forcings and the corresponding streamflow time-series.

First, a preliminary experiment was conducted for the main case study: Austria. Two rainfall-runoff models were calibrated over a very data-rich region covering most of the country. A set of consolidated parameter regionalisation techniques was implemented in leave-one-out cross-validation. This required analysing the different options for their optimal setting, such as the choice of the similarity measures and of the number of donor catchments from which transferring the model parameters, that were optimised. Finally, the simulation perfomances of the different approaches were compared.

Second, an analysis was set up for assessing the performances and selecting the parameter regionalisation approaches most suitable to a specific study region, keeping into account the impact of gauging density and of the presence of nested catchments. The work aims at giving support to the choice of the techniques based specifically on the availability of hydrometric information, i.e. informative content, in the region. With such aim, we tested the robustness of the regionalisation methods, previously applied to the very densely gauged Austrian region, to the reduction of the informative content in the region. On one hand, the regionalisation approaches were re-applied by excluding from the donor set the basins considered to be nested to the target, measuring the degradation of the simulation efficiency, in respect to the original experiment, in which nested basins could serve as donors. On the other hand, the impact of decreasing the density of the gauged stations (i.e. overall number of stations across the study region) was analysed, testing different subsets of the original monitoring network and computing, for each of

them, the entity of median performance deterioration.

Results were in general consistent between the two analyses, confirming the strong dependency of all methods on the informative content of the dataset and quantifying its impact on each of the different approaches. However, they also provided additional guiding principles to the choice of the techniques. In fact, while for a high information richness (original condition of the experimental set of Austrian basins), the performances of all the techniques are all good and very close, when the data availability decreases, the results show that i) the use of "output-averaging" approaches, exploiting the information of more than one donor basins but preserving the correlation structure of the parameter set, and ii) the choice of a similarity measure based on catchment descriptors, rather than spatial proximity, improve the robustness of regionalisation.

Third, we tested the benefit of a semi-distributed based modelling framework for the calibration and regionalisation of rainfall-runoff model parameters, exploring the use of a similarity measure differentiated with elevation. A procedure to differentiate the parameters and thus the rainfall-runoff generation dynamics across elevation was especially adapted to the semi-distributed structure of the TUW model. The twofold purpose of the study was to evaluate the potential improvement of the simulation both in gauged (at-site) and ungauged cases, and to acquire additional understanding about changes in similarity (defined as function of landscape attributes) due to differences in rainfall-runoff generation dynamics at different altitudes. The main motivation for this work lies in the lack of studies in the literature testing the transfer of model parameters at sub-catchment scale, especially with a model structure based on elevation zones. The proposed methodology was applied to two large study regions: Austria and USA. Such framework was applied only to the runoff generation module of the hydrological model.

The proposed calibration procedure allows the set of parameters ruling the runoff generation to vary across elevation. Results showed substantial improvement of at-site efficiency, in respect to those obtained with a standard calibration approach, and no over-parameterisation, especially across US catchments. In Austria the improvement was more limited since the at-site performances were already very good with the "traditional" calibration (where such parameters are the same for all the elevation zones of the catchment).

Then, a regionalisation framework was developed for the transfer of the runoff generation parameters differentiated across elevation, optimising different similarity measures for different altitudes.

Unfortunately, no significant patterns of the attributes with altitude were detected, probably due to equifinality of different similarity measures. Compared to a standard regionalisation framework (transferring the runoff generation parameters uniformly between

catchments and not between elevation zones) the proposed approach slightly improved the regionalisation accuracy in reproducing the at-site runoff generation in Austria. In USA, the performances of the simulation at ungauged sites obtained with both the regionalisation approaches were not satisfactorily; this was probably due, besides all the limitation of the study, to the high climatic and hydological diversity of the study region across which the considered macrozones were not homogeneously distributed: a subdivision of the country into more homogeneous sub-regions is needed, in order also to differentiate the elevation ranges across the extremely various topography of the country.

The semi-distributed structure of the tested rainfall-runoff model, where the elevation zones do not propagate the runoff generated between each other but are considered as autonomous entities, may be one of the reasons why the proposed approach failed to improve the simulation at ungauged basins. In fact, the strongest limitation of the study is probably that the regionalisation was validated against the runoff production simulated at-site (that is from the model calibrated as gauged), since the actual runoff production cannot be observed. Given the structure of the semi-distributed model, the combination of the runoff generation and routing to the closure section probably does not allow to distinguish the actual contributions of the elevation zones in the at-site calibration and the runoff production simulated by the model is therefore affected by relevant uncertainty.

Further work will apply similar frameworks, to make the most of the available information on the hydrological similarity at sub-basin scale, to different rainfall-runoff semi-distributed or fully-distributed model structures in the future. Moreover, a proper subdivision of the US dataset into more homogeneous sub-regions will be implemented.

Finally, we focused on a complementary research field, not directly involving parameter regionalisation but devoted to improve our knowledge about similarity of rainfall-runoff processes. We tested the potential of approaches recently proposed in Information Theory for classifying the rainfall-runoff catchment dynamics. In particular, we focused on measuring the interaction between the entire time series of runoff forcings (i.e. precipitation, actual evapotranspiration and snowmelt) and runoff itself, in order to use such quantities to define a similarity in the hydrological behaviour of the catchments. In fact, no studies in the literature have so far tried to use such fine-scale interaction to characterise and classify basins. A methodology based on the concept of Transfer Entropy (i.e. transfer of information) from meteorological forcing to catchment streamflow, called TE-HC, was set up and applied to the Austrian dataset. Analysing the obtained classification against a set of well-established streamflow signatures, the results are interesting and highlight the potential of such new signature for classifying catchments and to identifying common dominant drivers in rainfall-runoff transformation. Further research is of course needed to

deepen in the capabilities of Transfer Entropy, testing additional variants of the proposed methodology.

In conclusion, we believe that this Thesis provides important contributes for supporting hydrological modelling in ungauged basins.

On one hand, the role of the quantity and quality of the gauged data available for transferring rainfall-runoff model parameters was analysed, offering an additional instrument to the hydrological community for the choice of the regionalisation technique more appropriate for each study region. At the same time, the efforts spent to set up a semi-distributed calibration and regionalisation framework, in view of the differences in the runoff generation processes occurring at different altitudes, highlight the potential but also the limitations of the proposed approach, mainly due to the lack of measurable output for the calibration of the parameters characterising a specific elevation zone.

On the other hand, an innovative approach to quantify and compare dominant hydrological processes was proposed with interesting results. The potential of transfer entropy to identify similar catchment behaviours appears promising especially for enhancing the transfer of rainfall-runoff model parameters.

Bibliography

Abdulla, F. A. and Lettenmaier, D. P. (1997). Development of regional parameter estimation equations for a macroscale hydrologic model. *Journal of Hydrology*, 197(1):230 – 257.

Abebe, N. A., Ogden, F. L., and Pradhan, N. R. (2010). Sensitivity and uncertainty analysis of the conceptual HBV rainfall-runoff model: Implications for parameter estimation. *Journal of Hydrology*, 389(3-4):301–310.

Abrahams, A. D. (1984). Channel networks: A geomorphological perspective. *Water Resources Research*, 20(2):161–188.

Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P. (2017). The CAMELS data set: Catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, 21(10):5293–5313.

Archfield, S. A., Kennen, J. G., Carlisle, D. M., and Wolock, D. M. (2014). An objective and parsimonious approach for classifying natural flow regimes at a continental scale. *River Research and Applications*, 30(9):1166–1183.

Arsenault, R., Breton-Dufour, M., Poulin, A., Dallaire, G., and Romero-Lopez, R. (2019). Streamflow prediction in ungauged basins: analysis of regionalization methods in a hydrologically heterogeneous region of mexico. *Hydrological Sciences Journal*, 64(11):1297–1311.

Arsenault, R., Poulin, A., Côté, P., and Brissette, F. (2014). Comparison of stochastic optimization algorithms in hydrological model calibration. *Journal of Hydrologic Engineering*, 19(7):1374–1384.

Bao, Z., Zhang, J., Liu, J., Fu, G., Wang, G., He, R., Yan, X., Jin, J., and Liu, H. (2012). Comparison of regionalization approaches based on regression and similarity for predictions in ungauged catchments under multiple hydro-climatic conditions. *Journal of Hydrology*, 466-467:37–46.

Bárdossy, A. (2007). Calibration of hydrological model parameters for ungauged catchments. *Hydrology and Earth System Sciences*, 11(2):703–710.

Bastola, S., Ishidaira, H., and Takeuchi, K. (2008). Regionalisation of hydrological model parameters under parameter uncertainty: A case study involving TOPMODEL and basins across the globe. *Journal of Hydrology*, 357(3-4):188–206.

Beck, H. E., Pan, M., Lin, P., Seibert, J., van Dijk, A. I. J. M., and Wood, E. F. (2020). Global fully distributed parameter regionalization based on observed streamflow from 4,229 headwater catchments. *Journal of Geophysical Research: Atmospheres*, 125(17):e2019JD031485.

Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J., and Bruijnzeel, L. A. (2016). Global-scale regionalization of hydrologic model parameters. *Water Resources Research*, 52(5):3599–3622.

Becker, R., Koppa, A., Schulz, S., Usman, M., aus der Beek, T., and Schüth, C. (2019). Spatially distributed model calibration of a highly managed hydrological system using remote sensing-derived et data. *Journal of Hydrology*, 577:123944.

Behrendt, S., Zimmermann, D., Dimpfl, T., and Peter, F. (2019). *RTransferEntropy: Measuring Information Flow Between Time Series with Shannon and Renyi Transfer Entropy*. R package version 0.2.8.

Bennett, A., Nijssen, B., Ou, G., Clark, M., and Nearing, G. (2019). Quantifying Process Connectivity With Transfer Entropy in Hydrologic Models. *Water Resources Research*, 55(6):4613–4629.

Berghuijs, W. R., Sivapalan, M., Woods, R. A., and Savenije, H. H. (2014a). Patterns of similarity of seasonal water balances: A window into streamflow variability over a range of time scales. *Water Resources Research*, 50(7):5638–5661.

Berghuijs, W. R., Woods, R. A., and Hrachowitz, M. (2014b). A precipitation shift from snow towards rain leads to a decrease in streamflow. *Nature Climate Change*, 4(7):583–586.

Bergström, S. (1976). *Development and Application of a Conceptual Runoff Model for Scandinavian Catchments*. A: Bulletin series. Department of Water Resources Engineering, Lund Institute of Technology, University of Lund.

Beven, K. (1989). Changing ideas in hydrology - the case of physically-based models. *Journal of Hydrology*, 105(1):157 – 172.

Beven, K. and Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the glue methodology. *Journal of Hydrology*, 249(1):11 – 29.

Beven, K. J. and Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, 24(1):43–69.

Blaney, H. and Criddle, W. (1962). *Determining Consumptive Use and Irrigation Water Requirements*. USDA Technical Bulletin 1275, US Department of Agriculture, Beltsville.

Blöschl, G. (2006). *Rainfall-Runoff Modeling of Ungauged Catchments*.

Blöschl, G. and Sivapalan, M. (1995). Scale issues in hydrological modelling: A review. *Hydrological Processes*, 9(3-4):251–290.

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.

Boldetti, G., Riffard, M., Andréassian, V., and Oudin, L. (2010). Data-set cleansing practices and hydrological regionalization: is there any valuable information among outliers? *Hydrological Sciences Journal*, 55(6):941–951.

Brunner, M., Melsen, L., Newman, A., Wood, A., and Clark, M. (2020). Future streamflow regime changes in the United States: assessment using functional classification. *Hydrology and Earth System Sciences Discussions*, pages 1–23.

Budyko, M. I. (1974). Climate and life. *Academic Press, N. Y.*

Burn, D. H. and Boorman, D. B. (1992). Catchment classification applied to the estimation of hydrological parameters at ungauged catchments. Technical report, Wallingford. scanned legacy working document.

Calver, A., Crooks, S., Jones, D., Kay, A. Kjeldsen, T., and Reynard, N. (2005). National river catchment flood frequency method using continuous simulation. In *Joint Defra/EA Flood and Coastal Erosion Risk Management R&D Programme, R&D Technical Report FD2106/TR*.

Castellarin, A., Burn, D., and Brath, A. (2001). Assessing the effectiveness of hydrological similarity measures for flood frequency analysis. *Journal of Hydrology*, 241(3):270 – 285.

Castiglioni, S., Lombardi, L., Toth, E., Castellarin, A., and Montanari, A. (2010). Calibration of rainfall-runoff models in ungauged basins: A regional maximum likelihood approach. *Advances in Water Resources*, 33(10):1235 – 1242. Special Issue on Novel Insights in Hydrological Modelling.

Chiang, S. M., Tsay, T. K., and Nix, S. J. (2002). Hydrologic regionalization of watersheds. I: Methodology development. *Journal of Water Resources Planning and Management*, 128(1):3–11.

Chiew, F. H. S. and Siriwardena, L. (2005). Prediction of runoff hydrographs in ungauged basins. In Zerger, A. and Argent, R. M., editors, *MODSIM 2005 International Congress on Modeling and Applications*, pages 2883–2889. Modell. and Simul. Soc. of Aust. and N. Z., Canberra.

Cislaghi, A., Masseroni, D., Massari, C., Camici, S., and Brocca, L. (2019). Combining a rainfall-runoff model and a regionalization approach for flood and water resource assessment in the western po valley, italy. *Hydrological Sciences Journal*, 65(3):348–370.

Colorni, A., Dorigo, M., Maffioli, F., Maniezzo, V., Righini, G., and Trubian, M. (1996). Heuristics from nature for hard combinatorial optimization problems. *International Transactions in Operational Research*, 3(1):1–21.

Corduas, M. (2011). Clustering streamflow time series for regional classification. *Journal of Hydrology*, 407:73–80.

Coron, L., Perrin, C., and Michel, C. (2017a). *airGR: Suite of GR Hydrological Models for Precipitation-Runoff Modelling*. R package version 1.0.9.64.

Coron, L., Thirel, G., Delaigue, O., Perrin, C., and Andréassian, V. (2017b). The suite of lumped GR hydrological models in an R package. *Environmental Modelling & Software*, 94:166–171.

Cover, T. M. and Thomas, J. A. (2005). Differential entropy. In *Elements of Information Theory*, chapter 8, pages 243–259. John Wiley & Sons, Ltd.

Dawson, C. and Wilby, R. (2001). Hydrological modelling using artificial neural networks. *Progress in Physical Geography - PROG PHYS GEOG*, 25:80–108.

de Lavenne, A., Thirel, G., Andréassian, V., Perrin, C., and Ramos, M.-H. (2016). Spatial variability of the parameters of a semi-distributed hydrological model. *Proceedings of the International Association of Hydrological Sciences*, 373:87–94.

De Thomasis, E. and Grimaldi, S. (2001). Introduzione di una metrica tra modelli parametrici lineari nelle applicazioni di tipo idrologico. In *Giornata di Studio: Metodi Statistici and Matematici per l'Analisi delle Serie Idrologiche, Roma*.

Deckers, D. L. E. H., Booij, M. J., Rientjes, T. H. M., and Krol, M. S. (2010). Catchment Variability and Parameter Estimation in Multi-Objective Regionalisation of a Rainfall–Runoff Model. *Water Resources Management*, 24(14):3961–3985.

Di Prinzio, M., Castellarin, A., and Toth, E. (2011). Data-driven catchment classification: application to the pub problem. *Hydrology and Earth System Sciences*, 15(6):1921–1935.

Ding, J., Wallner, M., Müller, H., and Haberlandt, U. (2016). Estimation of instantaneous peak flows from maximum mean daily flows using the HBV hydrological model. *Hydrological Processes*, 30(9):1431–1448.

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

Fahle, M., Hohenbrink, T. L., Dietrich, O., and Lischeid, G. (2015). Temporal variability of the optimal monitoring setup assessed using information theory. *Water Resources Research*, 51(9):7723–7743.

Farmer, W. (2016). Ordinary kriging as a tool to estimate historical daily streamflow records. *Hydrology and Earth System Sciences*, 20:2721–2735.

Fernandez, W., Vogel, R. M., and Sankarasubramanian, A. (2000). Regional calibration of a watershed model. *Hydrological Sciences Journal*, 45(5):689–707.

Garcia, F., Folton, N., and Oudin, L. (2017). Which objective function to calibrate rainfall-runoff models for low-flow index simulations? *Hydrological Sciences Journal*, 62(7):1149–1166.

GRASS Development Team (2017). *Geographic Resources Analysis Support System (GRASS GIS) Software, Version 7.2*. Open Source Geospatial Foundation.

Grimaldi, S. (2004). Linear parametric models applied to daily hydrological series. *Journal of Hydrologic Engineering*, 9(5):383–391.

Guo, Y., Zhang, Y., Zhang, L., and Wang, Z. (2021). Regionalization of hydrological modeling for predicting streamflow in ungauged catchments: A comprehensive review. *WIREs Water*, 8(1):e1487.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1-2):80–91.

Guse, B., Pfannerstill, M., Gafurov, A., Kiesel, J., Lehr, C., and Fohrer, N. (2017). Identifying the connective strength between model parameters and performance criteria. *Hydrology and Earth System Sciences*, 21(11):5663–5679.

He, Y., Bárdossy, A., and Zehe, E. (2011). A review of regionalisation for continuous streamflow simulation. *Hydrology and Earth System Sciences*, 15(11):3539–3553.

Heuvelmans, G., Muys, B., and Feyen, J. (2006). Regionalisation of the parameters of a hydrological model: Comparison of linear regression models with artificial neural nets. *Journal of Hydrology*, 319(1):245 – 265.

Hirpa, F. A., Salamon, P., Beck, H. E., Lorini, V., Alfieri, L., Zsoter, E., and Dadson, S. J. (2018). Calibration of the global flood awareness system (glofas) using daily streamflow data. *Journal of Hydrology*, 566:595 – 606.

Hirsch, R. M., Helsel, D. R., Cohn, T. A., and Gilroy, E. J. (1993). Statistical treatment of hydrologic data, chapter 17. In Maidment, D. R., editor, *The McGraw-Hill Handbook of Hydrology*. McGraw-Hill, New York.

Hlinka, J., Hartman, D., Vejmelka, M., Runge, J., Marwan, N., Kurths, J., and Paluš, M. (2013). Reliability of inference of directed climate networks using conditional mutual information. *Entropy*, 15(6):2023–2045.

Hogue, T. S., Sorooshian, S., Gupta, H., Holz, A., and Braatz, D. (2000). A Multistep Automatic Calibration Scheme for River Forecasting Models. *Journal of Hydrometeorology*, 1(6):524–542.

Holmes, M. G. R., Young, A. R., Gustard, A., and Grew, R. (2002). A new approach to estimating mean flow in the uk. *Hydrology and Earth System Sciences*, 6(4):709–720.

Hrachowitz, M., Savenije, H., Blöschl, G., McDonnell, J., Sivapalan, M., Pomeroy, J., Arheimer, B., Blume, T., Clark, M., Ehret, U., Fenicia, F., Freer, J., Gelfan, A., Gupta, H., Hughes, D., Hut, R., Montanari, A., Pande, S., Tetzlaff, D., Troch, P., Uhlenbrook, S., Wagener, T., Winsemius, H., Woods, R., Zehe, E., and Cudennec, C. (2013). A decade of predictions in ungauged basins (pub) - a review. *Hydrological Sciences Journal*, 58(6):1198–1255.

Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2:193–218.

Hundecha, Y. and Bárdossy, A. (2004a). Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model. *Journal of Hydrology*, 292(1-4):281–295.

Hundecha, Y. and Bárdossy, A. (2004b). Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model. *Journal of Hydrology*, 292(1):281 – 295.

Huot, P. L., Poulin, A., Audet, C., and Alarie, S. (2019). A hybrid optimization approach for efficient calibration of computationally intensive hydrological models. *Hydrological Sciences Journal*, 64(10):1204–1222.

Isaaks, E. H. and Srivastava, R. M. (1990). *Applied Geostatistics*. OUP USA.

Jakeman, A. J. and Hornberger, G. M. (1993). How much complexity is warranted in a rainfall-runoff model? *Water Resources Research*, 29(8):2637–2649.

Jehn, F. U., Bestian, K., Breuer, L., Kraft, P., and Houska, T. (2020). Using hydrological and climatic catchment clusters to explore drivers of catchment behavior. *Hydrology and Earth System Sciences*, 24(3):1081–1100.

Kay, A. L., Jones, D. A., Crooks, S. M., Calver, A., and Reynard, N. S. (2006). A comparison of three approaches to spatial generalization of rainfall-runoff models. *Hydrological Processes*, 20(18):3953–3973.

Kay, A. L., Jones, D. A., Crooks, S. M., Kjeldsen, T. R., and Fung, C. F. (2007). An investigation of site-similarity approaches to generalisation of a rainfall-runoff model. *Hydrology and Earth System Sciences*, 11(1):500–515.

Keum, J. and Coulibaly, P. (2017). Information theory-based decision support system for integrated design of multivariable hydrometric networks. *Water Resources Research*, 53(7):6239–6259.

Keum, J., Kornelsen, K. C., Leach, J. M., and Coulibaly, P. (2017). Entropy applications to water monitoring network design: A review. *Entropy*, 19(11):1–21.

Kim, K. B., Kwon, H.-H., and Han, D. (2018). Exploration of warm-up period in conceptual hydrological modelling. *Journal of Hydrology*, 556:194 – 210.

Kling, H., Fuchs, M., and Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, 424-425:264–277.

Kling, H. and Gupta, H. (2009). On the development of regionalization relationships for lumped watershed models: The impact of ignoring sub-basin scale variability. *Journal of Hydrology*, 373(3):337 – 351.

Knoben, W. J. M., Freer, J. E., and Woods, R. A. (2019). Technical note: Inherent benchmark or not? comparing nash–sutcliffe and kling–gupta efficiency scores. *Hydrology and Earth System Sciences*, 23(10):4323–4331.

Knoben, W. J. M., Woods, R. A., and Freer, J. E. (2018). A quantitative hydrological climate classification evaluated with independent streamflow data. *Water Resources Research*, 54(7):5088–5109.

Kokkonen, T. S., Jakeman, A. J., Young, P. C., and Koivusalo, H. J. (2003). Predicting daily flows in ungauged catchments: Model regionalization from catchment descriptors at the Coweeta Hydrologic Laboratory, North Carolina. *Hydrological Processes*, 17(11):2219–2238.

Köppen, W. (1936). Das geographische system del klimate. In Köppen, W. and Geiger, R., editors, *Handbuch der Klimatologie*, volume 1, pages 1–44. Gebrüder Bornträger, Berlin.

Krstanovic, P. and Singh, V. (1992). Evaluation of rainfall networks using entropy: Ii. application. *Water Resources Management*, 6:295–314.

Krzanowski, W. J. (1988). *Principles of Multivariate Analysis: A User's Perspective*. Oxford University Press, Inc., USA.

Kuentz, A., Arheimer, B., Hundecha, Y., and Wagener, T. (2017). Understanding hydrologic variability across europe through catchment classification. *Hydrology and Earth System Sciences*, 21(6):2863–2879.

Kull, D. W. and Feldman, A. D. (1998). Evolution of clark's unit graph method to spatially distributed runoff. *Journal of Hydrologic Engineering*, 3(1):9–19.

Kumar, R., Samaniego, L., and Attinger, S. (2010). The effects of spatial discretization and model parameterization on the prediction of extreme runoff characteristics. *Journal of Hydrology*, 392(1):54 – 69.

Laaha, G. and Blöschl, G. (2006). A comparison of low flow regionalisation methods-catchment grouping. *Journal of Hydrology*, 323(1):193 – 214.

Lamb, R. and Kay, A. L. (2004). Confidence intervals for a spatially generalized, continuous simulation flood frequency model for great britain. *Water Resources Research*, 40(7).

Lebecherel, L., Andréassian, V., and Perrin, C. (2016). On evaluating the robustness of spatial-proximity-based regionalization methods. *Journal of Hydrology*, 539.

Lee, H., McIntyre, N., Wheater, H., and Young, A. (2005). Selection of conceptual models for regionalisation of the rainfall-runoff relationship. *Journal of Hydrology*, 312(1-4):125–147.

Lerat, J., Andréassian, V., Perrin, C., Vaze, J., Perraud, J. M., Ribstein, P., and Loumagne, C. (2012). Do internal flow measurements improve the calibration of rainfall-runoff models? *Water Resources Research*, 48(2).

Li, H. and Zhang, Y. (2017). Regionalising rainfall-runoff modelling for predicting daily runoff: Comparing gridded spatial proximity and gridded integrated similarity approaches against their lumped counterparts. *Journal of Hydrology*, 550:279–293.

Li, H., Zhang, Y., Chiew, F. H., and Xu, S. (2009). Predicting runoff in ungauged catchments by using Xinanjiang model with MODIS leaf area index. *Journal of Hydrology*, 370(1-4):155–162.

Lindström, G., Johansson, B., Persson, M., Gardelin, M., and Bergström, S. (1997). Development and test of the distributed HBV-96 hydrological model. *Journal of Hydrology*, 201(1-4):272–288.

Littlewood, I., Croke, B., Jakeman, A., and Sivapalan, M. (2003). The role of "top-down" modelling for Prediction in Ungauged Basins (PUB). *Hydrological Processes*, 17:1673 – 1679.

Liu, D. (2020). A rational performance criterion for hydrological model. *Journal of Hydrology*, 590:125488.

Lombardi, L., Toth, E., Castellarin, A., Montanari, A., and Brath, A. (2012). Calibration of a rainfall-runoff model at regional scale by optimising river discharge statistics: Performance analysis for the average/low flow regime. *Physics and Chemistry of the Earth, Parts A/B/C*, 42-44:77 – 84. Estimating and representing uncertainty in applied hydrology, hydraulics and water quality studies.

Loritz, R., Gupta, H., Jackisch, C., Westhoff, M., Kleidon, A., Ehret, U., and Zehe, E. (2018). On the dynamic nature of hydrological similarity. *Hydrology and Earth System Sciences*, 22(7):3663–3684.

Lowe, L. and Nathan, R. (2006). Use of similarity criteria for transposing gauged streamflows to ungauged locations. *Australasian Journal of Water Resources*, 10(2):161–170.

Marschinski, R. and Kantz, H. (2002). Analysing the information flow between financial time series. *The European Physical Journal B*, 30:275–281.

Maurer, E. P., Wood, A. W., Adam, J. C., Lettenmaier, D. P., and Nijssen, B. (2002). A long-term hydrologically based dataset of land surface fluxes and states for the conterminous united states*. *Journal of Climate*, 15(22):3237–3251.

Mcdonnell, J. and Woods, R. (2004). On the need for catchment classification. *Journal of Hydrology*, 299:2–3.

McIntyre, N., Lee, H., Wheater, H., Young, A., and Wagener, T. (2005). Ensemble predictions of runoff in ungauged catchments. *Water Resources Research*, 41(12):1–14.

Mcmanamay, R. A., Bevelhimer, M. S., and Kao, S. C. (2014). Updating the US hydrologic classification: An approach to clustering and stratifying ecohydrologic data. *Ecohydrology*, 7(3):903–926.

McManamay, R. A. and Derolph, C. R. (2019). Data descriptor: A stream classification system for the conterminous United States. *Scientific Data*, 6:1–18.

Melsen, L. A., Addor, N., Mizukami, N., Newman, A. J., Torfs, P. J., Clark, M. P., Uijlenhoet, R., and Teuling, A. J. (2018). Mapping (dis)agreement in hydrologic projections. *Hydrology and Earth System Sciences*, 22(3):1775–1791.

Melsen, L. A. and Guse, B. (2019). Hydrological Drought Simulations: How Climate and Model Structure Control Parameter Sensitivity. *Water Resources Research*, 55(12):10527–10547.

Merz, R. and Blöschl, G. (2004). Regionalisation of catchment model parameters. *Journal of Hydrology*, 287(1-4):95–123.

Merz, R. and Blöschl, G. (2005). Flood frequency regionalisation: spatial proximity vs. catchment attributes. *Journal of Hydrology*, 302(1):283 – 306.

Merz, R., Blöschl, G., and Parajka, J. (2006). Regionalization methods in rainfall-runoff modelling using large catchment samples. *IAHS-AISH Publication*, 307:117–125.

Merz, R., Tarasova, L., and Basso, S. (2020). Parameter's controls of distributed catchment models: How much information is in conventional catchment descriptors? *Water Resources Research*, 56(2):e2019WR026008. e2019WR026008 2019WR026008.

Mészáros, I., Miklánek, P., and Parajka, J. (2002). Solar energy income modelling in mountainous areas. *Interdisciplinary Approaches in Small Catchment Hydrology: Monitoring and Research - Proceedings of the 9th Conference of the European Network of Experimental and Representative Basins*, pages 127–135.

Mishra, A. K. and Coulibaly, P. (2009). Developments in hydrometric network design: A review. *Reviews of Geophysics*, 47(2).

Mizukami, N., Clark, M. P., Newman, A. J., Wood, A. W., Gutmann, E. D., Nijssen, B., Rakovec, O., and Samaniego, L. (2017). Towards seamless large-domain parameter estimation for hydrologic models. *Water Resources Research*, 53(9):8020–8040.

Mizukami, N., Rakovec, O., Newman, A. J., Clark, M. P., Wood, A. W., Gupta, H. V., and Kumar, R. (2019). On the choice of calibration metrics for 'high-flow' estimation using hydrologic models. *Hydrology and Earth System Sciences*, 23(6):2601–2614.

Montanari, A. and Toth, E. (2007). Calibration of hydrological models in the spectral domain: An opportunity for scarcely gauged basins? *Water Resources Research*, 43(5).

Moore, R. J. (1985). The probability-distributed principle and runoff production at point and basin scales. *Hydrological Sciences Journal*, 30(2):273–297.

Mwakalila, S. (2003). Estimation of stream flows of ungauged catchments for river basin management. *Physics and Chemistry of the Earth, Parts A/B/C*, 28(20):935 – 942.

Nagy, L. and Grabherr, G. (2009). *The Biology of Alpine Habitats*. Biology of Habitats. OUP Oxford.

Nash, J. and Sutcliffe, J. (1970). River flow forecasting through conceptual models part i - a discussion of principles. *Journal of Hydrology*, 10(3):282 – 290.

Nash, J. E. (1960). A unit hydrograph study, with particular reference to british catchments. *Proceedings of the Institution of Civil Engineers*, 17(3):249–282.

Neri, M., Parajka, J., and Toth, E. (2020). Importance of the informative content in the study area when regionalising rainfall-runoff model parameters: the role of nested catchments and gauging station density. *Hydrology and Earth System Sciences*, 24(11):5149–5171.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T., and Duan, Q. (2015). Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: Data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrology and Earth System Sciences*, 19(1):209–223.

Nijzink, R. C., Almeida, S., Pechlivanidis, I. G., Capell, R., Gustafssons, D., Arheimer, B., Parajka, J., Freer, J., Han, D., Wagener, T., van Nooijen, R. R., Savenije, H. H., and Hrachowitz, M. (2018). Constraining Conceptual Hydrological Models With Multiple Information Sources. *Water Resources Research*, 54(10):8332–8362.

Olden, J. D. and Poff, N. L. (2003). Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. *River Research and Applications*, 19(2):101–121.

Oudin, L., Andréassian, V., Mathevet, T., Perrin, C., and Michel, C. (2006). Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations. *Water Resources Research*, 42(7).

Oudin, L., Andréassian, V., Perrin, C., Michel, C., and Le Moine, N. (2008). Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments. *Water Resources Research*, 44(3):48–54.

Oudin, L., Kay, A., Andréassian, V., and Perrin, C. (2010). Are seemingly physically similar catchments truly hydrologically similar? *Water Resources Research*, 46(11).

Ozkul, S., Harmancioglu, N. B., and Singh, V. P. (2000). Entropy-based assessment of water quality monitoring networks. *Journal of Hydrologic Engineering*, 5(1):90–100.

Pagliero, L., Bouraoui, F., Diels, J., Willems, P., and McIntyre, N. (2019). Investigating regionalization techniques for large-scale hydrological modelling. *Journal of Hydrology*, 570:220 – 235.

Parajka, J., Andréassian, V., Archfield, S., Bárdossy, A., Blöschl, G., Chiew, F., Duan, Q., A., G., Hlavčová, K., Merz, R., McIntyre, N., Oudin, L., Perrin, C., Rogger, M., Salinas, J. L., Savenije, H. G., Skøien, J. O., Wagener, T., Zehe, E., and Zhang, Y. (2013a). Prediction of runoff hydrographs in ungauged basins. In Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H., editors, *Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales*, pages 227–269. Cambridge University Press, Cambridge.

Parajka, J., Blöschl, G., and Merz, R. (2007). Regional calibration of catchment models: Potential for ungauged catchments. *Water Resources Research*, 43(6):1–16.

Parajka, J., Merz, R., and Blöschl, G. (2005). A comparison of regionalisation methods for catchment model parameters. *Hydrology and Earth System Sciences*, 9(3):157–171.

Parajka, J., Merz, R., Skøien, J. O., and Viglione, A. (2015). The role of station density forredicting daily runoff by top-kriging interpolation in Austria. *Journal of Hydrology and Hydromechanics*, 63(3):228–234.

Parajka, J., Viglione, A., Rogger, M., Salinas, J. L., Sivapalan, M., and Blöschl, G. (2013b). Comparative assessment of predictions in ungauged basins-Part 1: Runoff-hydrograph studies. *Hydrology and Earth System Sciences*, 17(5):1783–1795.

Patil, S. and Stieglitz, M. (2012). Controls on hydrologic similarity: role of nearby gauged catchments for prediction at an ungauged catchment. *Hydrology and Earth System Sciences*, 16(2):551–562.

Patil, S. D. and Stieglitz, M. (2015). Comparing spatial and temporal transferability of hydrological model parameters. *Journal of Hydrology*, 525:409–417.

Pebesma, E. J. (2001). *Gstat User ' S Manual*. Dept. of Physical Geography, Utrecht Universit.

Pechlivanidis, I., McIntyre, N., and Wheater, H. (2010). Calibration of the semi-distributed pdm rainfall-runoff model in the upper lee catchment, uk. *Journal of Hydrology*, 386(1):198 – 209.

Peel, M., Chiew, F., Western, A., and Mcmahon, T. (2000). *Extension of Unimpaired Monthly Streamflow Data andRegionalisation of Parameter Values to Estimate Streamflow in Ungauged Catchments*. Centre for Environmental Applied Hydrology, The University of Melbourne, Melbourne.

Peel, M. C. and Blöschl, G. (2011). Hydrological modelling in a changing world. *Progress in Physical Geography*, 35(2):249–261.

Pérez Ciria, T. and Chiogna, G. (2020). Intra-catchment comparison and classification of long-term streamflow variability in the alps using wavelet analysis. *Journal of Hydrology*, 587:124927.

Perrin, C., Michel, C., and Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, 279(1-4):275–289.

Poncelet, C., Merz, R., Merz, B., Parajka, J., Oudin, L., Andréassian, V., and Perrin, C. (2017). Process-based interpretation of conceptual hydrological model performance using a multinational catchment set. *Water Resources Research*, 53(8):7247–7268.

Pool, S., Vis, M., and Seibert, J. (2018). Evaluating model performance: towards a non-parametric variant of the Kling-Gupta efficiency. *Hydrological Sciences Journal*, 63(13-14):1941–1953.

Pool, S., Viviroli, D., and Seibert, J. (2019). Value of a limited number of discharge observations for improving regionalization: A large-sample study across the united states. *Water Resources Research*, 55(1):363–377.

Post, D. (2009). Regionalizing rainfall-runoff model parameters to predict the daily streamflow of ungauged catchments in the dry tropics. *Hydrology Research*, 40.

Post, D. A. and Jakeman, A. J. (1996). Relationships between catchment attributes and hydrological response characteristics in small Australian mountain ash catchments. *Hydrological Processes*, 10(6):877–892.

Post, D. A. and Jakeman, A. J. (1999). Predicting the daily streamflow of ungauged catchments in S.E. Australia by regionalising the parameters of a lumped conceptual rainfall-runoff model. *Ecological Modelling*, 123(2-3):91–104.

Pushpalatha, R., Perrin, C., Le Moine, N., Mathevet, T., and Andréassian, V. (2011). A downward structural sensitivity analysis of hydrological models to improve low-flow simulation. *Journal of Hydrology*, 411(1-2):66–76.

Pushpalatha, R., Perrin, C., Moine, N. L., and Andréassian, V. (2012). A review of efficiency criteria suitable for evaluating low-flow simulations. *Journal of Hydrology*, 420-421:171 – 182.

Quintero, F., Krajewski, W. F., Seo, B.-C., and Mantilla, R. (2020). Improvement and evaluation of the iowa flood center hillslope link model (hlm) by calibration-free approach. *Journal of Hydrology*, 584:124686.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rajsekhar, D., Mishra, A. K., and Singh, V. P. (2013). Regionalization of drought characteristics using an entropy approach. *Journal of Hydrologic Engineering*, 18(7):870–887.

Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., Attinger, S., Schäfer, D., Schrön, M., and Samaniego, L. (2016). Multiscale and multivariate evaluation of water fluxes and states over european river basins. *Journal of Hydrometeorology*, 17(1):287–307.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336):846–850.

Randrianasolo, A., Ramos, M. H., and Andréassian, V. (2011). Hydrological ensemble forecasting at ungauged basins: Using neighbour catchments for model setup and updating. *Advances in Geosciences*, 29:1–11.

Rao, A. R. and Srinivas, V. (2006). Regionalization of watersheds by fuzzy cluster analysis. *Journal of Hydrology*, 318(1):57 – 79.

Rardin, R. L. and Uzsoy, R. (2001). Experimental evaluation of heuristic optimization algorithms: A tutorial. *Journal of Heuristics*, 7(3):261–304.

Razavi, S., Tolson, B. A., Matott, L. S., Thomson, N. R., MacLean, A., and Seglenieks, F. R. (2010). Reducing the computational cost of automatic calibration through model preemption. *Water Resources Research*, 46(11):1–17.

Razavi, T. and Coulibaly, P. (2013). Classification of Ontario watersheds based on physical attributes and streamflow series. *Journal of Hydrology*, 493:81–94.

Reichl, J. P., Western, A. W., McIntyre, N. R., and Chiew, F. H. (2009). Optimization of a similarity measure for estimating ungauged streamflow. *Water Resources Research*, 45(10):1–15.

Reuter, H. I., Nelson, A., and Jarvis, A. (2007). An evaluation of void-filling interpolation methods for srtm data. *International Journal of Geographical Information Science*, 21(9):983–1008.

Riboust, P., Thirel, G., Moine, N. L., and Ribstein, P. (2019). Revisiting a simple degree-day model for integrating satellite data: implementation of swe-sca hystereses. *Journal of Hydrology and Hydromechanics*, 67(1):70–81.

Rosbjerg, D., Blöschl, G., Burn, D., Castellarin, A., Croke, B., Di Baldassarre, G., Iacobellis, V., Kjeldsen, T. R., Kuczera, G., Merz, R., Montanari, A., Morris, D., Ouarda, T. B. M. J., Ren, L., Rogger, M., Salinas, J. L., Toth, E., and Viglione, A. (2013). Prediction of floods in ungauged basins. In Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H., editors, *Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales*, pages 189–226. Cambridge University Press, Cambridge.

RouholahnejadFreund, E., Fan, Y., and Kirchner, J. (2019). Global assessment of how averaging over land-surface heterogeneity affects modeled evapotranspiration rates. *Hydrology and Earth System Sciences Discussions*.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.

Ruddell, B. L. and Kumar, P. (2009). Ecohydrologic process networks: 2. analysis and characterization. *Water Resources Research*, 45(3).

Samaniego, L., Kumar, R., Thober, S., Rakovec, O., Zink, M., Wanders, N., Eisner, S., Müller Schmied, H., Sutanudjaja, E. H., Warrach-Sagi, K., and Attinger, S. (2017). Toward seamless hydrologic predictions across spatial scales. *Hydrology and Earth System Sciences*, 21(9):4323–4346.

Samuel, J., Coulibaly, P., and Metcalfe, R. A. (2011). Estimation of continuous streamflow in ontario ungauged basins: Comparison of regionalization methods. *Journal of Hydrologic Engineering*, 16(5):447–459.

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G. (2011). Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern usa. *Hydrology and Earth System Sciences*, 15(9):2895–2911.

Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters*, 85:461–464.

Sefton, C. and Howarth, S. (1998). Relationships between dynamic response characteristics and physical descriptors of catchments in england and wales. *Journal of Hydrology*, 211(1):1 – 16.

Seibert, J. (1999). Regionalisation of parameters for a conceptual rainfall-runoff model. *Agricultural and Forest Meteorology*, 98-99:279–293.

Seibert, J., Staudinger, M., and van Meerveld, H. J. I. (2019). Validation and over-parameterization—experiences from hydrological modeling. In Beisbart, C. and Saam, N. J., editors, *Computer Simulation Validation: Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives*, pages 811–834. Springer International Publishing, Cham.

Servat, E. and Dezetter, A. (1993). Rainfall-runoff modelling and water resources assessment in northwestern ivory coast. tentative extension to ungauged catchments. *Journal of Hydrology*, 148(1):231 – 248.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423.

Singh, S. K., McMillan, H., Bárdossy, A., and Fateh, C. (2016). Nonparametric catchment clustering using the data depth function. *Hydrological Sciences Journal*, 61(15):2649–2667.

Singh, V. P. (1997). The use of entropy in hydrology and water resources. *Hydrological Processes*, 11(6):587–626.

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., Merz, R., and Blöschl, G. (2006). Top-kriging - geostatistics on stream networks. *Hydrology and Earth System Sciences*, 10(2):277–287.

Steinschneider, S., Yang, Y. C. E., and Brown, C. (2015). Combinaison de la régression et de la proximité spatiale pour la régionalisation des modèles de bassin versant: étude comparative. *Hydrological Sciences Journal*, 60(6):1026–1043.

Swain, J. B. and Patra, K. C. (2017). Streamflow estimation in ungauged catchments using regionalization techniques. *Journal of Hydrology*, 554:420 – 433.

Swain, J. B. and Patra, K. C. (2019). Impact of catchment classification on streamflow regionalization in ungauged catchments. *SN Applied Sciences*, 1(5):456.

Szolgay, J., Hlavčové, K., Kohnová, S., and Danihlík, R. (2003). Estimating regional parameters of a monthly balance model. *Journal of Hydrology and Hydromechanics*, 51:256–273.

Thornthwaite, C. W. (1948). An approach toward a rational classification of climate. *Geographical Review*, 38(1):55–94.

Thornton, P. E. and Running, S. W. (1999). An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. *Agricultural and Forest Meteorology*, 93(4):211–228.

Tolson, B. A. and Shoemaker, C. A. (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, 43(1).

Toth, E. (2013). Catchment classification based on characterisation of streamflow and precipitation time series. *Hydrology and Earth System Sciences*, 17(3):1149–1159.

Troll, C. (1973). High mountain belts between the polar caps and the equator: Their definition and lower limit. *Arctic and Alpine Research*, 5(3):A19–A27.

Tung, Y. K., Yeh, K. C., and Yang, J. C. (1997). Regionalization of unit hydrograph parameters: 1. comparison of regression analysis techniques. *Stochastic Hydrol Hydraul*, 11:145 – 171.

Valéry, A., Andréassian, V., and Perrin, C. (2010). Regionalization of precipitation and air temperature over high-altitude catchments - learning from outliers. *Hydrological Sciences Journal*, 55(6):928–940.

Valéry, A., Andréassian, V., and Perrin, C. (2014). 'As simple as possible but not simpler': What is useful in a temperature-based snow-accounting routine? Part 2 - Sensitivity analysis of the Cemaneige snow accounting routine on 380 catchments. *Journal of Hydrology*, 517:1176–1187.

Vandewiele, G. L. and Elias, A. (1995). Monthly water balance of ungauged catchments obtained by geographical regionalization. *Journal of Hydrology*, 170(1-4):277–291.

Vezza, P., Comoglio, C., and Rosso, M. (2010). Low flows regionalization in north-western italy. *Water Resources Management*, 24:4049–4074.

Viglione, A., Claps, P., and Laio, F. (2007). Mean annual runoff estimation in north-western italy. In La Loggia, G., Aronica, G. T., and Ciraolo, G., editors, *Water resources assessment and management under water scarcity scenarios*, pages 97–122. Centro Studi Idraulica Urbana, Torino.

Viglione, A. and Parajka, J. (2018). *TUWmodel: Lumped Hydrological Model for Education Purposes*. R package version 1.0-1.

Viglione, A., Parajka, J., Rogger, M., Salinas, J. L., Laaha, G., Sivapalan, M., and Blöschl, G. (2013). Comparative assessment of predictions in ungauged basins - Part 3: Runoff signatures in Austria. *Hydrology and Earth System Sciences*, 17(6):2263–2279.

Viney, N. R., Vaze, J., Chiew, F. H., Perraud, J., Post, D. A., and Teng, J. (2009). Comparison of multi-model and multi-donor ensembles for regionalisation of runoff generation using five lumped rainfall-runoff models. *18th World IMACS Congress and MODSIM 2009 - International Congress on Modelling and Simulation: Interfacing Modelling and Simulation with Mathematical and Computational Sciences, Proceedings*, (July):3428–3434.

Viviroli, D., Mittelbach, H., Gurtz, J., and Weingartner, R. (2009). Continuous simulation for flood estimation in ungauged mesoscale catchments of Switzerland - Part II: Parameter regionalisation and flood estimation results. *Journal of Hydrology*, 377(1-2):208–225.

Wagener, T., Blöschl, G., Goodrich, D. C., Gupta, H. V., Sivapalan, M., Tachicawa, Y., Troch, P. A., and Weiler, M. (2013). A synthesis framework for runoff prediction. In Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H., editors, *Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales*, pages 11–28. Cambridge University Press, Cambridge.

Wagener, T., Sivapalan, M., Troch, P., and Woods, R. (2007). Catchment classification and hydrologic similarity. *Geography Compass*, 1:901 – 931.

Wagener, T. and Wheater, H. S. (2006). Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty. *Journal of Hydrology*, 320(1-2):132–154.

Wallner, M., Haberlandt, U., and Dietrich, J. (2013). A one-step similarity approach for the regionalization of hydrological model parameters based on self-organizing maps. *Journal of Hydrology*, 494:59 – 71.

Wang, D. and Wu, L. (2013). Similarity of climate control on base flow and perennial stream density in the budyko framework. *Hydrology and Earth System Sciences*, 17(1):315–324.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244.

Westerberg, I. K., Guerrero, J.-L., Younger, P. M., Beven, K. J., Seibert, J., Halldin, S., Freer, J. E., and Xu, C.-Y. (2011). Calibration of hydrological models using flow-duration curves. *Hydrology and Earth System Sciences*, 15(7):2205–2227.

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Lettenmaier, D., Koren, V., Duan, Q., Mo, K., Fan, Y., and Mocko, D. (2012). Continental-scale water and energy flux analysis and validation for the north american land data assimilation system project phase 2 (nldas-2): 1. intercomparison and application of model products. *Journal of Geophysical Research: Atmospheres*, 117(D3).

Xu, C.-Y. (2003). Testing the transferability of regression equations derived from small sub-catchments to a large area in central sweden. *Hydrology and Earth System Sciences*, 7:317–324.

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.

Yang, X., Magnusson, J., Rizzi, J., and Xu, C.-Y. (2018). Runoff prediction in ungauged catchments in norway: comparison of regionalization approaches. *Hydrology Research*, 49(2):487–505.

Yilmaz, K. K., Gupta, H. V., and Wagener, T. (2008). A process-based diagnostic approach to model evaluation: Application to the nws distributed hydrologic model. *Water Resources Research*, 44(9).

Young, A. R. (2006). Stream flow simulation within uk ungauged catchments using a daily rainfall-runoff model. *Journal of Hydrology*, 320(1):155 – 172. The model parameter estimation experiment.

Zelelew, M. B. and Alfredsen, K. (2014). Transferability of hydrological model parameter spaces in the estimation of runoff in ungauged catchments. *Hydrological Sciences Journal*, 59(8):1470–1490.

Zhang, Y. and Chiew, F. H. (2009). Relative merits of different methods for runoff predictions in ungauged catchments. *Water Resources Research*, 45(7).

Acknowledgements

First of all, I would like to express my sincere gratitude to my Supervisor, Prof. Elena Toth, 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, Elena really encouraged me throughout the past four years, resulting in a friendly and wise guidance. Special thanks to the DICAM fellows: Simone, Aleppio, Aleddio, Lorenz, Giada, Cristiana, Marco, Serena, Sara, Francesca, Irene, Federica, Iuliia, Attilio, Peyman, Andrea, Antonio and Prof. Brath. Each one of them, in his own way, played a part in making the last three years positive and formative, in a very friendly and collaborative environment.

Thanks are due to all the co-authors of the scientific contributes I published during my PhD. Particular thanks to Prof. Juraj Parajka and Prof. Alberto Viglione, for the wise suggestions regarding the *TUWmodel* R-package and its implementation and for providing us the main dataset used for all the experiments of my Dissertation.

Many thanks are owed to Prof. Paulin Coulibaly and his research team, for the warm and friendly hospitality and valuable and stimulating supervision during my stay at McMaster University, Hamilton (Canada).

Last but not least, I warmly thank my Girlfriend Carla, my Family and my Friends, for being a peerless support during these years and all my life in general.

E vaisempre!