



1 **Importance of the information content in the study area when regionalising rainfall-runoff**
2 **model parameters: the role of nested catchments and gauging station density**

3 Mattia Neri¹, Juraj Parajka², Elena Toth¹

4 ¹ DICAM, University of Bologna, Bologna, Italy

5 ² Institute for Hydraulic and Water Resources Engineering, Vienna University of Technology, Austria

6 *Correspondence to:* Mattia Neri (mattia.neri5@unibo.it)

7 **Abstract.**

8 The set up of a rainfall-runoff model in a river section where no streamflow measurements are available for its
9 calibration is one of the key research activity for the Prediction in Ungauged Basins (PUB): in order to do so it is
10 possible to regionalise the model parameters based on the information available in gauged sections in the study region.

11 The information content in the data set of gauged river stations plays an essential role in the assessment of the best
12 regionalisation method: this study analyses how the performances of different model regionalisation approaches are
13 influenced by the “information richness” of the available regional data set, and in particular by its gauging density and
14 by the presence of nested catchments, that are expected to be hydrologically very similar.

15 The research is carried out over a densely gauged dataset covering the Austrian country, applying two different rainfall-
16 runoff models: a semi-distributed version of the HBV model (TUW model), and the Cemaneige-GR6J model. The
17 regionalisation approaches include both methods which transfer the entire set of model parameters from donor
18 catchments, thus maintaining correlation among parameters (“output averaging” techniques), and methods which derive
19 each target parameter independently, as a function of the calibrated donors’ ones (“parameter averaging” techniques).

20 The regionalisation techniques are first implemented using all the basins in the dataset as potential donors, showing that
21 the output-averaging methods outperform the parameter-averaging kriging method, highlighting the importance of
22 maintaining the correlation between the parameter values.

23 The regionalisation is then repeated decreasing the information content of the data set, by excluding the nested basins,
24 identified taking into account either the position of the closing section along the river or the percentage of shared
25 drainage area. The parameter-averaging kriging is the method that is less impacted by the exclusion of the nested
26 donors, whereas the methods transferring the entire parameter set from only one donor suffer the highest deterioration,
27 since the single most similar or closest donor is often a nested one. On the other hand, the output-averaging methods
28 degrade more gracefully, showing that exploiting the information resulting from more than one donor increases the
29 robustness of the approach also in regions that do not have so many nested catchments as the Austrian one.

30 Finally, the deterioration resulting from decreasing the station density on the regionalisation was analysed, showing that
31 the output averaging methods using as similarity measure a set of catchment descriptors, rather than the geographical
32 distance, are more capable to adapt to less dense datasets.

33 The study confirms how the predictive accuracy of parameter regionalisation techniques strongly depends on the
34 information content of the dataset of available donor catchments and indicates that the output-averaging approaches,
35 using more than one donor basin but preserving the correlation structure of the parameter set, seem to be preferable for
36 regionalisation purposes in both data-poor and data-rich regions.



37 1 Introduction

38 In the hydrological practice, it is often needed to gain information on ungauged river sections and one of the most
39 informative way to do so is implementing a rainfall-runoff model, when, as it is often the case, the meteorological input
40 variables are retrievable in reference to its drainage area. Since in such cases the model parameters may not be obtained
41 through a calibration procedure, it is necessary to regionalise them, exploiting the information of the hydrometric
42 measurements collected in hydrologically similar catchments in the study area.

43 Regionalisation approaches for model parameterisation can be classified into two wide categories: “regression-based”
44 methods and “distance-based” methods (He et al., 2011). The former techniques try to define relationships between
45 each model parameter and geomorpho-climatic catchment attributes (see e.g., Seibert 1999). The latter, instead, identify
46 a set of donor watersheds (with similar attributes) and transfer their calibrated parameters to the ungauged (“target”)
47 catchment. This last type of approaches includes both methods which transfer the entire set of model parameters from
48 donor catchments, thus maintaining correlation among parameters (also named “output averaging” techniques, which
49 run the model multiple times and average the simulations), and methods which derive each target parameter
50 independently, as a function (generally a weighted average) of the calibrated donors’ ones (“parameter averaging”
51 techniques). To the latter class (“distance-based” group of the “parameter averaging” type) belong also the kriging
52 methods, where the parameters are regionalised based on their spatial correlation and independently from each other
53 (Merz and Blöschl, 2004; Parajka et al., 2005).

54

55 In the last two decades, hydrologic scientists from all around the world have focused on the determination of the more
56 accurate regionalisation techniques for different case studies and rainfall-runoff models (see e.g., the reviews of Merz et
57 al. 2006, He et al. 2011, Peel & Blöschl 2011, Parajka et al. 2013, Hrachowitz et al. 2013, Razavi and Coulibaly 2013).

58

59 A very important aspect for choosing the most adequate regionalisation technique, and that is worthy of further
60 analyses, is the information content of the study region. In particular, in very densely gauged areas, spatial proximity is
61 expected to be a good similarity measure, as demonstrated by the studies by Merz and Blöschl (2004) and Parajka et al.
62 (2005), who tested different regionalisation approaches on a dense dataset of more than 300 watersheds across Austria,
63 and by Oudin et al. (2008), on a set of 913 catchments in France, finding that the techniques based on spatial proximity
64 alone provided excellent performances. But different outcomes may result for less densely and less interconnected (that
65 is with less availability of stations along the same river), as shown for instance, by Samuel et al. (2011): they
66 regionalised the parameters of HBV model for a strongly less densely gauged dataset (135 watersheds on the wide area
67 of Ontario, in Canada) and found that the best the best approach for such study area was an inverse-distance parameter
68 averaging for a pre-selected set of physically similar catchments.

69

70 The availability in the data set of gauged river stations representative of hydrological conditions similar to the ungauged
71 ones plays an essential role in the assessment of the best regionalisation method. This availability can be, in some way,
72 estimated with the station density (i.e. number of station per km²) and with the topological relationship between
73 catchments. In particular, the presence of several nested catchments (i.e. gauged river sections on the same river) in the
74 study region can strongly influence the performance of certain techniques: if for an ungauged basin, model parameter
75 sets are available for down/upstream gauged river sections, donor and target watersheds share indeed part of their
76 drainage area, and thus they may be also hydrologically very similar. This may actually lead to very good
77 regionalisation performances for a given approach, but such accuracy may not represent what would be obtained in



78 different conditions. Therefore, regionalisation performances obtained for datasets with high degree of “nestedness”
79 may be not transferrable to study regions poor of nested basins.

80

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

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

93

94 The purpose of the present paper is to compare the impact of the presence of nested donors on the performances of
95 different parameter regionalisation techniques for a dataset of 209 catchments across Austria. The effect of nested
96 donors is here tested for a set of consolidated techniques, applied to two different continuous simulating daily rainfall-
97 runoff models, for generalisation purposes: the first is the TUW model (semi-distributed version of HBV, used by
98 Parajka et al. 2005), and the second model, never used so far for regionalisation in the Austrian region, is the GR6J
99 model (Pushpalatha et al. 2011) implemented with the Cemaneige snow routine (Valery et al. 2014).

100 For the exclusion of nested basins, we propose two different criteria, taking into account either the position of the
101 closing section along the river or the percentage of shared drainage area. The results are also compared to the effect of
102 the reduction of station density, following a procedure similar to what was done, for different purposes, by Parajka et al.
103 (2015).

104 We believe that the present analysis may provide further insights for assessing the performances and selecting the
105 parameter regionalisation approaches most suitable to a specific study region, keeping into account the impact of the
106 topological information “richness” of the available regional dataset.

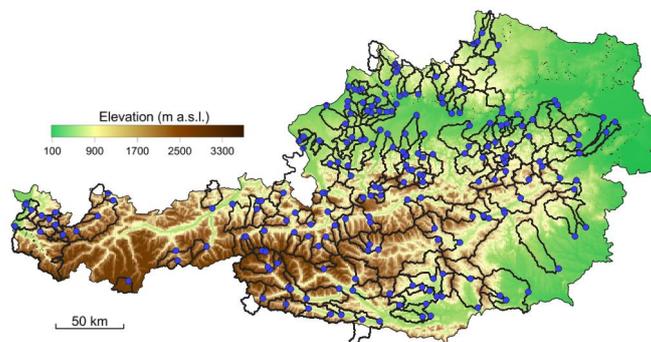
107 The paper is organized as follows: Section 2 introduces the case study and data. Section 3 first describes the rainfall-
108 runoff models and the tested regionalisation schemes, then the methodology for assessing the impact of nested
109 catchments and of station density is presented, while the results are presented in Section 4. Finally, Section 5 reports the
110 discussion and the conclusions.

111 **2 Study region and data**

112 The case study is composed by 209 catchments (see Figure 1) covering a large portion of Austria. Their size varies
113 considerably, from 13 to over 6000 km². The topography of the country varies significantly from the flat and hilly area
114 in the north-east to the Alps in the centre and in the south-west, particularly steep in the extreme west. The annual
115 precipitation ranges from about 600 mm in the east, where the evaporation plays an important role in the water balance,
116 to the more than 2000 mm in the west, mainly due to orographic lifting of north-westerly airflows at the rim of the Alps



117 (Viglione et al., 2013). Land use is mainly agricultural in the lowlands and forest in the medium elevation ranges.
118 Alpine vegetation and rocks prevail in the highest catchment (Parajka et al., 2005). The aridity index assumes values
119 from 0.2 to 1, meaning that the watersheds are mainly wet or weakly arid (annual evapotranspiration is never higher
120 than precipitation).
121 Data have been provided by the Institute of Hydraulic Engineering and Water Resources Management (Vienna
122 University of Technology), which previously screened the runoff data for errors and removed all stations with
123 significant anthropogenic effects. Hydro-meteorological data include daily streamflow and daily inputs to the rainfall-
124 runoff models for the 33 years period 1976-2008: daily average precipitation, temperature and potential
125 evapotranspiration defined for 200 meters elevation zones for all the study catchments. The potential evapotranspiration
126 is estimated by a modified Blaney-Criddle method (Parajka et al., 2005) using interpolated daily air temperature and
127 grid maps of potential sunshine duration (Mészáros et al., 2002).
128 In order to implement some of the parameter regionalisation approaches, we make use of several geo-morphoclimatic
129 catchment attributes, reported and briefly described in Table 1. Topographic attributes such as mean catchment
130 elevation and mean slope are derived from 1 x 1 km digital elevation model while climatic features such as mean annual
131 precipitation, and aridity index are derived from climatic input time series. Figure 2 shows the spatial pattern of mean
132 annual precipitation, snow depth and aridity index across the study area. Mean annual solar irradiance is computed
133 through GRASS GIS software (<http://grass.osgeo.org>). Stream network density, FARL (flood attenuation by reservoir
134 and lakes), boundaries of porous aquifers, areal portions of regional soil types and main geological formation were the
135 same used and described in detail in Parajka et al. (2005). Finally, Land use coverage is derived from CORINE Land
136 Cover maps updated to year 2012 (<https://land.copernicus.eu/pan-european/corine-land-cover/clc-2012>). For land cover
137 classes, as well as for geology and soil type classes, the catchments are associated to more than one single attribute:
138 each basin is described by the portions of the total catchment area corresponding to each class (and for this reason,
139 Table 1 does not report the min/median/max values of such descriptors.
140



141
142

Figure 1. Study area, blue points refer to stream gauges and black lines to catchment boundaries.

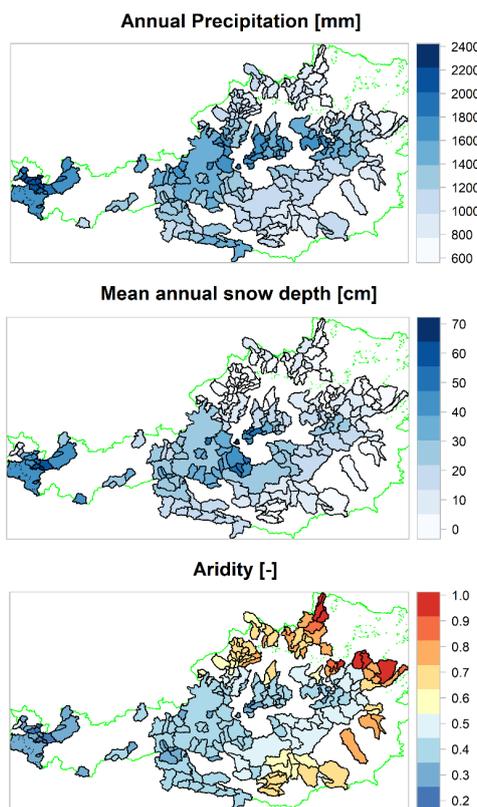


Figure 2. Spatial pattern of some climatic catchment attributes across the study area.

143
 144
 145
 146

Table 1. Available catchment attributes.

Code	Unit	Min	Median	Max	Description
Elev	m a.s.l.	287	915	2964	Mean elevation
Area	km ²	14	168	6214	Drainage area
Slope	m/m	0.9	12.4	28.5	Mean slope
meanP	mm	675	1230	2310	Mean annual total precipitation
maxP	mm	35	49	84	Mean annual maximum daily precipitation
meanPET	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	mm	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
Corine	%	-	-	-	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



147 **3 Materials and methods**

148 **3.1 Rainfall-runoff models structure and calibration**

149 Two models for simulating daily streamflow were applied in this study. This choice is made in order to analyse the
150 effect of nested catchments and station density on the performance of parameter regionalisation methods for different
151 model structures.

152 **3.1.1 TUW model**

153 The first is the TUW model, a semi-distributed version of the HBV model (Bergström 1976, Lindström et al., 1997)
154 developed by Parajka and Viglione (2019). It consists in a snow routine, a soil moisture routine and a flow response and
155 routing routine. The model processes the elevation zones as autonomous entities that contribute separately to the total
156 outlet flow. The inputs are daily air temperature, precipitation and potential evapotranspiration over the different
157 elevation zones, on which the model is run in the version schematized in Figure 3. Finally, the different outputs from
158 the elevation zones are averaged taking into account the sub-catchment areas.

159 The snow routine is based on a simple degree-day concept and it is ruled by five parameters: two threshold temperature
160 parameters distinguishing rain and snow, Tr and Ts , a melting temperature Tm , a snow correction factor SCF and the
161 degree-day factor DDF . The soil moisture routine represents soil moisture state changes and runoff generation and
162 involves three parameters: the maximum soil moisture storage FC , a parameter representing the soil moisture state
163 above which evapotranspiration is at its potential rate, LP , and a parameter β ruling the non-linear function of runoff
164 generation. Finally, an upper and a lower soil reservoirs and a triangular transfer function compose the runoff response
165 and routing routine, involving seven additional parameters. The sum of excess rainfall and snowmelt enters the upper
166 zone reservoir and leaves this reservoir through three paths: i) outflow from the reservoir based on a fast storage
167 coefficient k_1 ; ii) percolation to the lower zone with a constant percolation rate C_{perc} , iii) if a threshold of the upper
168 storage state L_{UZ} is exceeded, through an additional outlet based on a very fast storage coefficient k_0 . Water leaves the
169 lower zone based on a slow storage coefficient k_2 . The outflows from both reservoirs are then routed by a triangular
170 transfer function representing runoff routing in the streams, where the base of transfer function, B_Q , is estimated with
171 the scaling of the outflow by the C_{ROUTE} and B_{MAX} parameters. More details about the model structure and application in
172 R can be found in Parajka et al. (2007) and Ceola et al (2015), respectively.

173 The model is run for all the study catchments with the semi-distributed model structure obtained dividing them into
174 200-meters elevation zones: model daily inputs (precipitation, temperature and potential evapotranspiration) and model
175 states are defined over such zones, while model parameters are assumed to be the same for the entire catchment.

176 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
177 are calibrated: threshold temperatures Tr and Ts are fixed respectively to 2 and 0 °C, Tm to 0 °C and the maximum base
178 of the transfer function at low flows B_{MAX} to 10 days. Table 2 briefly reports and describes the calibrated parameters,
179 defining also their lower and upper bounds.

180
181
182
183
184
185

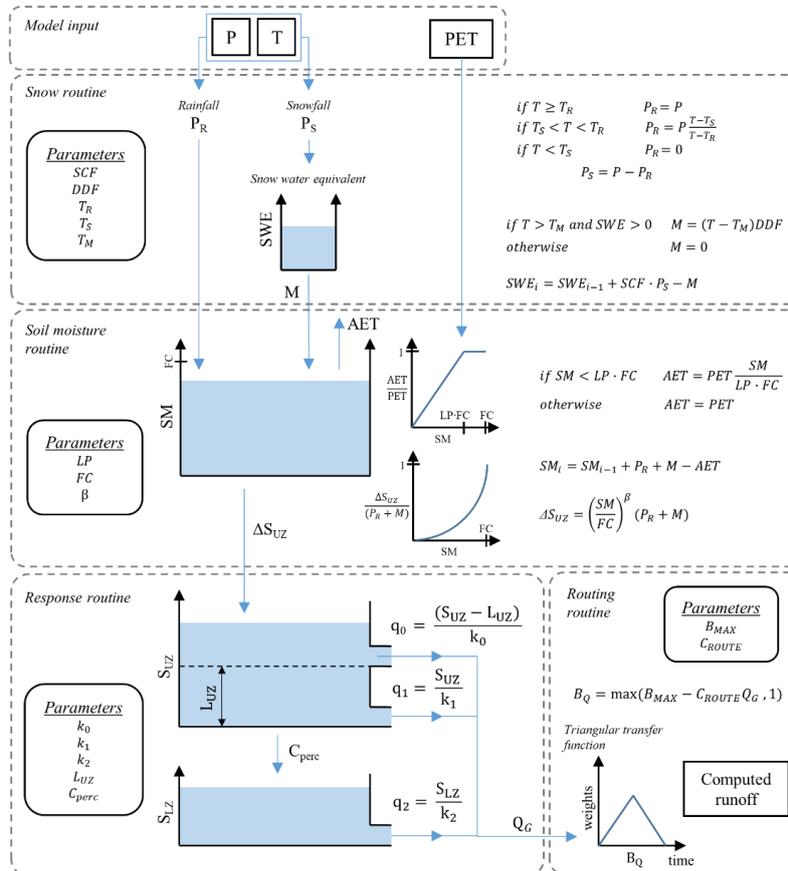


186

Table 2. TUW model parameters and their ranges.

Parameter	Units	Range	Description
SCF	-	0.9 - 1.5	Snow correction factor
DDF	mm/(°C*day)	0 - 5	Degree day factor
LP	-	0 - 1	Parameter related to the limit of evaporation
FC	mm	0 - 600	Field capacity, i.e., max soil moisture storage
β	-	0 - 20	Non linear parameter for runoff production
k_0	days	0 - 2	Storage coefficient for very fast response
k_1	days	2 - 30	Storage coefficient for fast response
k_2	days	30 - 250	Storage coefficient for slow response
L_{UZ}	mm	0 - 100	Threshold storage state, very fast response starts if exceeded
C_{perc}	mm/day	0 - 8	Constant percolation rate
C_{ROUTE}	days ² /mm	0 - 50	Scaling parameter

187



188

189

Figure 3. TUW model scheme – Lumped version.



190 **3.1.2 CemaNeige-GR6J model**

191 The second model is the French CemaNeige-GR6J. It is the combination of the CemaNeige snow accounting routine
 192 (Valéry et al., 2014) with the GR6J model (Pushpalatha et al., 2011), a daily lumped continuous rainfall-runoff model,
 193 developed at IRSTEA (Anthony, France), by the Équipe Hydrologie des Bassins versants.

194 The inputs of the model are spatially-averaged catchment daily air temperature, precipitation and potential
 195 evapotranspiration. Catchment hypsometric curve is also required.

196 The CemaNeige snow accounting routine is based on a degree-day concept, where the thermal inertia of the snowpack
 197 is also taken into account. It involves two parameters, a snowmelt factor, θ_{G1} , and a cold-content factor, θ_{G2} . Although
 198 the module requires daily lumped inputs, for better simulating snow accumulation and melting it allows to divide the
 199 catchment into more elevation zones of equal area, through the use of the hypsometric curve. Inputs for each elevation
 200 zone are extracted through interpolation of the mean catchment values using precipitation and temperature gradients
 201 (Valéry et al, 2010), and not from “clipping” of the actual spatial fields like for the TUW elevation zones. The module
 202 functions are applied with a lumped set of calibrated parameters; but internal states are allowed to vary over each
 203 elevation layer according to the different extrapolated inputs. On each elevation layer, two outputs are computed: rain
 204 and snowmelt, which are summed in order to find the total water quantity feeding the hydrological model. At every time
 205 step, the total liquid output of CemaNeige at catchment scale is the average of every elevation zone outputs. Here we
 206 decide to maintain, as default, the number of elevation layers equal to five. For a detailed description of CemaNeige
 207 routines, the readers may refer to Valéry et al. (2014).

208 The total liquid output of CemaNeige module and potential evapotranspiration are the inputs of the GR6J rainfall-runoff
 209 model. In the model, the water balance is controlled by a soil moisture accounting reservoir and a conceptual
 210 “groundwater” exchange function, while the routing part of the structure consists in two flow components routed by two
 211 unit hydrographs, a non-linear store and an exponential-store, with a total of six parameters. The structure of the model
 212 is represented in Figure 4 and a detailed description of the model routines is given in Pushpalatha et al. (2011).

213 The CemaNeige-GR6J model is fed with mean catchment daily precipitation, air temperature and potential
 214 evapotranspiration. All the 8 parameters of the combined model (2 for CemaNeige, 6 for GR6J) are calibrated. Lower
 215 and upper bounds of the parameters space are kept as default: all the parameters are allowed to vary between the
 216 normalized interval [-9.99 9.99] and then specific parameter transformations are applied before the model is run. Table
 217 3 reports brief parameters description and transformed boundaries. For the sake of simplicity, we will refer to this
 218 model just with the acronym GR6J, even if it will always include the CemaNeige snow module.

219
 220

Table 3. CemaNeige-GR6J model parameters and their transformed real ranges.

Parameter	Units	Range	Description
θ_{G1}	mm/(°C*day)	0 - 109	Snowmelt (degree-day) factor
θ_{G2}	-	0 - 1	Cold content factor
X1	mm	0 - 21807	Non-linear production storage capacity
X2	mm	-1903 - 1903	Groundwater exchange coefficient
X3	mm	0 - 21807	Non-linear routing store capacity
X4	days	0 - 22	Time parameter for unit hydrographs routing
X5	-	0 - 1	Threshold parameter for water exchange with groundwater
X6	mm	0 - 21807	Exponential routing store capacity

221



243 preferred to regression-based techniques (see e.g. Kokkonen et al. 2003, Merz and Blöschl 2004, Oudin et al. 2008,
244 Reichl et al. 2009, Bao et al. 2012, Steinschneider et al. 2015, Yang et al. 2018, Cislighi et al. 2019).

245 3.2.1 Ordinary kriging (KR)

246 The first is a parameter-averaging technique, based on an ordinary kriging approach (termed in the following KR),
247 where each model parameter is regionalised independently from each other, based on their spatial correlation.
248 Catchment position is defined by the coordinates of the catchment centroid and the ordinary kriging is based on an
249 exponential variogram with a nugget of 10% of the observed variance, a sill equal to the variance, and a range of 60 km
250 both for TUV and Cemaneige-GR6J model parameters.

251 3.2.2 Nearest Neighbour (1 donor, NN-1)

252 The second approach is a nearest neighbour method (NN-1), where the complete set of model parameters is transposed
253 from the geographically nearest donor catchment.

254 3.2.3 Most Similar (1 donor, MS-1)

255 In the third technique, termed “most similar” approach (MS-1), a single donor catchment is again identified, for
256 transposing the entire parameter set but, instead of choosing the catchment that is geographically the closest, the
257 “hydrologically most similar” donor is identified, based on a set of geo-morphological and climatic descriptors. Five
258 descriptors are used for assessing such similarity: mean catchment elevation, long-term mean annual precipitation,
259 stream network density, land cover classes, geology classes. Such set of descriptors was selected by preliminary tests:
260 since it is not the focus of the work, the analysis for the assessment of the best catchment descriptors is reported in
261 Appendix A. The donor catchment is identified as the catchment with the smallest dissimilarity index ϕ (e.g. Burn and
262 Boorman, 1993):

263

$$264 \quad \phi = \sum_{j=1}^5 \frac{d_j(D,U)}{\max(d_j)} \quad \text{Eq. 2}$$

265

266 which represents the sum of the differences d_j of the 5 descriptors of the donor catchment D and of the ungauged
267 catchment U of interest, normalised by their maximum. For the attributes described by a single value (mean catchment
268 elevation, long-term mean annual precipitation and stream network density), d_j is expressed by the absolute difference
269 between the descriptors X_j^D and X_j^U of the donor and target catchments respectively (Eq. 3). For land cover and geology,
270 whose attributes X_j are the vectors containing the portions of the total catchment area $X_{j,c}$ corresponding to each class c ,
271 the difference d_j is calculated as the Euclidean distance between such vectors (Eq. 4).

272

$$273 \quad d_j(D,U) = |X_j^D - X_j^U| \quad \text{Eq. 3}$$

274

$$275 \quad d_j(D,U) = \sqrt{\sum_c (X_{j,c}^D - X_{j,c}^U)^2} \quad \text{Eq. 4}$$

276



277 3.2.4 Output averaging version of NN and MS techniques (NN-OA and MS-OA)

278 Nearest Neighbour and Most Similar approaches allow to maintain correlation among model parameters, and
279 overcomes the well-known limitation of the regression approach due to interaction between them. In the “regression-
280 based” methods in fact, as well as in the parameter-averaging approaches (e.g. KR technique), parameters are
281 regionalised independently from each other, possibly affecting simulation performances. On the other hand, one single
282 donor catchment (as in NN-1 and MS-1 approaches) is often not fully representative of the hydrological behavior of the
283 target watershed. Recent studies have been demonstrating how averaging the outputs of the simulations (rather than
284 model parameters) obtained with different donor parameter sets may be preferred (see e.g., Oudin et al. 2008, Viviroli et
285 al. 2009). For this reason, NN and MS techniques are also tested identifying more than one donor (here termed NN-OA
286 and MS-OA respectively), with an ‘output-averaging’ approach (introduced by McIntyre et al., 2005): n donor basins
287 (the geographically closest ones for the nearest neighbour method, or those with the smallest similarity indexes for the
288 “most similar” method) are identified. The regionalised streamflow for the ungauged catchment is calculated from all
289 the simulations $Q(d, P_i)$, obtained running the model (fed by the meteorological input of the target catchment) with each
290 one of the n parameter sets (P_i , with i in $[1 ; n]$) corresponding to each of the donor catchments. Streamflow for day
291 d , $Q(d)$, is computed as the weighted average of the simulated outputs:

292

$$293 \quad Q(d) = \sum_{i=1}^n w_i Q(d, P_i) \quad \text{Eq. 5}$$

294

295 where w_i is the weight associated to each donor catchment i , computed as function of a measure of dissimilarity
296 between the donor and the target catchments. In the NN-OA case, the dissimilarity is defined by the spatial distance D_i
297 between the centroids of donor i and target catchments (Eq. 6), while in the MS-OA method it corresponds to the
298 dissimilarity index ϕ_i (Eq. 7).

299

$$300 \quad w_i = \frac{\frac{1}{D_i}}{\sum_{i=1}^n \frac{1}{D_i}} \quad \text{Eq. 6}$$

301

$$302 \quad w_i = \frac{\frac{1}{\phi_i}}{\sum_{i=1}^n \frac{1}{\phi_i}} \quad \text{Eq. 7}$$

303

304 3.2.5 Choice of the number of donor catchments for NN-OA and MS-OA

305 The choice of the number of donor catchments for output averaging represents a central issue in the methodology.
306 Previous studies showed that the optimal number of donors is strongly related to the rainfall-runoff model and, of
307 course, to the case study. McIntyre et al. (2005) were amongst the first to apply an ensemble (“output averaging”)
308 approach and to explore the use of different numbers of donors on the performance of the Probability Distribution
309 Model (PDM, Moore, 1985) for a set of more than 100 UK catchments. They tested the impact of an increasing number
310 of donors, either selecting the first n catchments with the smallest dissimilarity measure, or including all the donors with
311 a value of dissimilarity below a defined threshold (in the latter case, the number of donors may thus vary depending on
312 the target-donors attributes). They found that a fixed number of ten donors resulted in the best regionalisation
313 performances. Oudin et al. (2008) applied an output-averaging regionalisation for the TOPMO and GR4J models to a



314 large French dataset of almost 1000 basins, but with no weights in flow averaging, since they used an arithmetic
315 average (thus not taking into account magnitude of donor dissimilarities). They found that the two models performed
316 optimally with a different number of donor catchments (seven and four respectively) and the efficiency of the
317 regionalised model decreased almost linearly when increasing the number of donors above such values. In fact, the
318 higher is the number of donor basins included in the regionalisation process, the more dissimilar will be the donors with
319 respect to the target watershed, possibly leading to a deterioration of the results. The use of weights in flow averaging
320 may indeed help to smooth this effect, giving less and less importance to the donors as their similarity decreases.
321 In the present work, the effect on regionalisation performances due to the number of donor basins is explored in detail,
322 applying NN-OA and MS-OA for increasing number n of donor catchments, as discussed in Section 4.2.
323

324 **3.3 Impact of nested catchments: which catchments should be considered (to be) nested?**

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

329 Once the performances have been evaluated using all the study catchments as potential donors, the regionalisation
330 procedures are repeated for each target basin (assumed to be ungauged) by excluding, from the donors set, the
331 watersheds which are considered to be nested in relation to the target section.

332 In general, two or more catchments are nested between each other if their closure sections are located on the same river,
333 i.e. they share part of their drainage area. Since it may happen that several gauged stations are located on the same river,
334 we propose to follow two different criteria in order to identify the nested basins:

- 335 - *Criterion 1*: the gauged sections that are immediately downstream and upstream of the target section
- 336 - *Criterion 2*: all the catchments sharing a given percentage of drainage area with the ungauged one.

337

338 **3.4 Impact of station density**

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

341 It is expected that the effect of the presence of several nested watersheds in a dataset is related to the effect due to
342 station density. Because of that, further purpose of the study is to compare the results obtained from the above described
343 nested catchments analysis to the impact of station density on regionalisation accuracy. Parajka et al. (2015) tested the
344 impact of the station density not for rainfall-runoff modelling but for the direct weighted interpolation of daily runoff
345 time-series with the topological-kriging (or Top-kriging) approach (see Skøien et al., 2006). Here, the same approach
346 for analysing the density is applied to all the parameters regionalisation techniques.

347 The full station density in the dataset is about 2.5 gauges per 1000 km², estimated dividing the total number of stations
348 by the area of Austrian territory, which is approximately 84000 km². All the applied regionalisation approaches are
349 tested for decreasing station density in the catchments dataset. Given a certain value of station density, the
350 corresponding number of gauged stations is randomly sampled from the original set of 209 catchments and the
351 regionalisation approaches are applied on this subsample (catchments input dataset) in leave-one-out cross validation: in

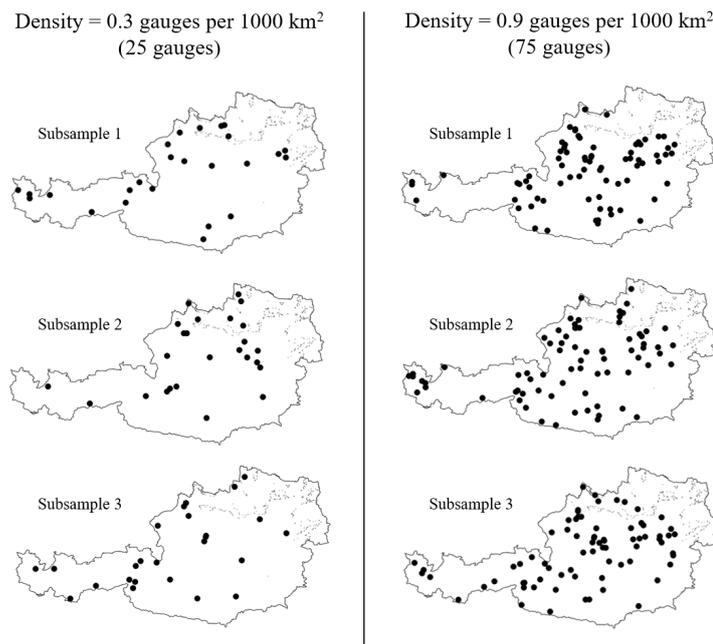


352 turn, each of the catchment in the subsample is considered to be ungauged and the remaining basins are used as
353 potential donors. Figure 5 shows an example of three samples for two different station densities, corresponding to 25
354 and 100 stations in the input dataset.

355 For each given value of station density, the following procedure is carried out:

- 356 - 100 different random samples (i.e. 100 different subsamples) with the same number of catchments are
357 generated.
- 358 - for each subsample, the regionalisation approaches are applied, through leave-one-out cross validation and the
359 deterioration of the performances with decreasing density is analysed.

360



361

362

Figure 5. Example of three samples for two different station densities.

363

364 4 Results and discussion

365 4.1 Model performances “at site”

366 As anticipated, the rainfall-runoff models are calibrated against Kling-Gupta Efficiency (Eq. 1). In addition to KGE,
367 model performances are evaluated through Nash-Sutcliffe Efficiency (Eq. 8) as well. While KGE considers different
368 types of model errors (the error in the mean, the variability and the dynamics of runoff), NSE is a standardize version of
369 the mean square error.

370

371

$$NSE = 1 - \frac{\sum(Q_{sim} - Q_{obs})^2}{\sum(Q_{obs} - Q_{obs})^2} \quad Eq. 8$$



372
 373 where Q_{sim} is the simulated runoff, Q_{obs} is the observed runoff and $\overline{Q_{obs}}$ is the average observed runoff.
 374 Table 4 shows the model performances obtained calibrating the models “at site”, that is over the streamflow measured
 375 in each catchment during the calibration period (1977-1992) and validated over the years 1992-2008 (no regionalisation
 376 procedure is involved).
 377 Both rainfall-runoff models behave well for the study area: in calibration, the median Kling-Gupta efficiencies are 0.85
 378 for TUW and 0.88 for GR6J model, while in validation they deteriorate to 0.76 and 0.81 respectively. In the calibration
 379 period, KGE is always above 0.66 and 0.76, respectively for TUW and GRJ6, whereas in validation, the KGE is over
 380 0.72 for both models for 75% of the basins (even if it drops below 0.3 for two and one basins, respectively for TUW
 381 and GR6J).
 382 Looking at Nash-Sutcliffe efficiency the difference between the two models is even more marked than for the KGE:
 383 GR6J model tends to perform better than TUW, despite the lower number of parameters.

384
 385
 386

Table 4. “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

387

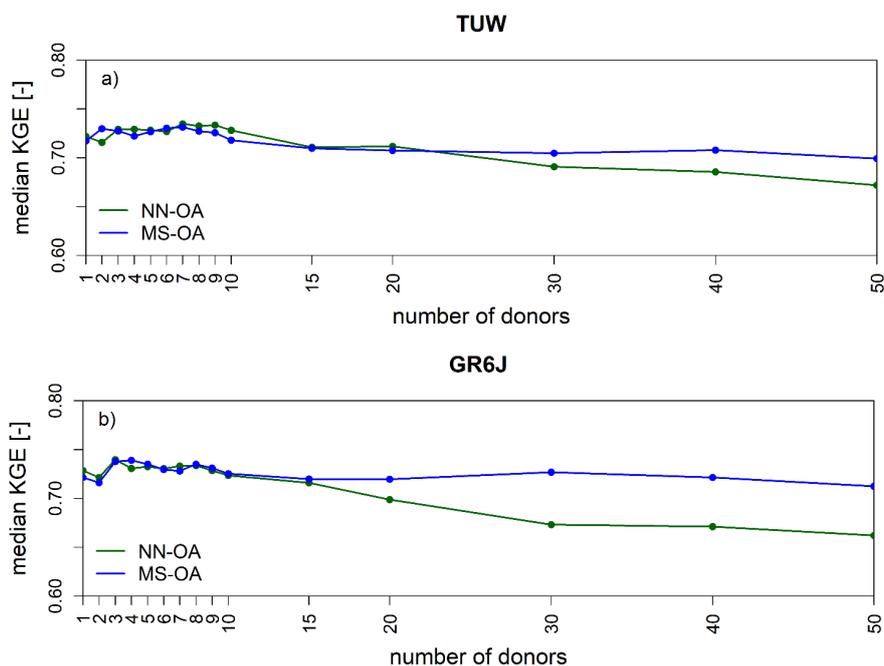
388 4.2 Regionalisation performances using all catchments as potential donors

389 4.2.1 Choice of the donors for the “output averaging” regionalisation methods

390 Before comparing performances of regionalisation methods, it is necessary to choose the optimal settings for the output-
 391 averaging versions of nearest neighbour (NN-OA) and “most similar” (MS-OA) techniques.
 392 As anticipated in the methodology Section 3.2.5, we first investigate the effect of using different numbers of donors: in
 393 particular, values between 1 and 50 are tested for both regionalisation techniques.
 394 Regionalisation methods are repeated through leave-one-out cross-validation for each number of donors n and the
 395 median Kling-Gupta efficiency obtained for each value of n over all the 209 catchments is computed. Tests are
 396 performed for calibration and validation periods, but results are reported only for the validation period.
 397 Figure 6 shows the median Kling-Gupta efficiency when the changing number of donors for TUW (upper panel) and
 398 GR6J (lower panel). Looking at the figures, we may see that in all the four cases, the index always deteriorates when
 399 more than 10 donors are chosen. On the other hand, there is not a unique optimal number of donors for the two models
 400 nor for the two regionalisation techniques. The optimal number of donors identified according to the median of the
 401 KGE varies between 3 and 7 depending both on the rainfall-runoff model (TUW or GRJ6) and on the regionalisation
 402 approach (NN-OA or MS-OA). Since the KGE differences between 3 and 7 donors are not so relevant (around 0.02),



403 we decided to use 3 donors for both regionalisation methods and both models, which is also the most parsimonious
404 option. In addition, the choice of a low number of donors is convenient also in view of the analysis to be done on
405 decreasing density, where a large number of donors would imply the use of catchments that are less and less similar to
406 the target one.
407 It may be noted that the results by Oudin et al. (2008) highlighted a clearer pattern of model performances when
408 increasing the number of donors, with a stronger decrease in efficiency when using high numbers of donors. This may
409 be explained by the fact that they were using a simple not-weighted average of outputs. Here instead, the influence of
410 the additional donors is gradually poorer, due to the weights implemented in the output averaging procedure (Eq. 5):
411 when adding further donors to the approaches, the corresponding weights in the average are gradually lower according
412 to the increasing distance (for NN-OA) or dissimilarity index (for MS-OA) from the target. Thus, the impact of the less
413 similar catchments is smoothed, compared to what may be achieved using a not-weighted output average.
414



415
416 Figure 6. Impact of the number of donors on output-averaging nearest neighbour (NN-OA) and 'most similar' (MS-OA)
417 regionalisation methods for TUV (panel a) and GR6J (panel b) model.
418

419 4.2.2 Performances of the regionalisation methods

420 This section shows the performances of the regionalisation methods without excluding any candidate donor: the above
421 described regionalisation methods will be tested over all the 209 study catchments through leave-one-out cross
422 validation, for both models. Here all the basins in the dataset are used as potential donors: in turn, each basin is
423 considered to be ungauged and all the remaining (208) catchments are available in the donors set for testing the
424 regionalisation approaches. The parameter sets of the donor catchments used in the regionalisation are obtained through



425 a calibration procedure over the years 1977-1992, whereas for assessing the performances of the regionalisation
426 methods, only the results obtained over the validation period (1992-2008) are reported. Spatiotemporal transfer of
427 model parameters is therefore the most exacting task (as confirmed by the study of Patil et al. 2015), since we are using
428 parameters obtained over different catchment (in regionalisation) and over a different observation period. On the other
429 hand, this is exactly what would happen in a real-world forecasting application or for assessing the impact of a climate
430 change scenario, where you have to identify the parametrization of a model to be used for independent hydro-climatic
431 conditions and in any possible river section in the region.

432

433 Figure 7 reports Kling-Gupta and Nash-Sutcliffe efficiency boxplots for the two models when regionalising following
434 each of the techniques.

435 For TUW (Figure 7, panels a and b), all regionalisation methods provided good simulations: with respect to the
436 performances (always on the validation period) obtained when the models have been calibrated on the target section (“at
437 site” simulations, white boxes): the loss in efficiency indexes is, overall, limited. The Nash-Sutcliffe efficiencies of KR,
438 MS-1 and NN-1 methods are consistent with the findings of Parajka et al. (2005), who computed only the NS: their
439 results are very similar to the present ones, even if they worked on a greater number of Austrian catchments and
440 calibrating the model against a different objective function.

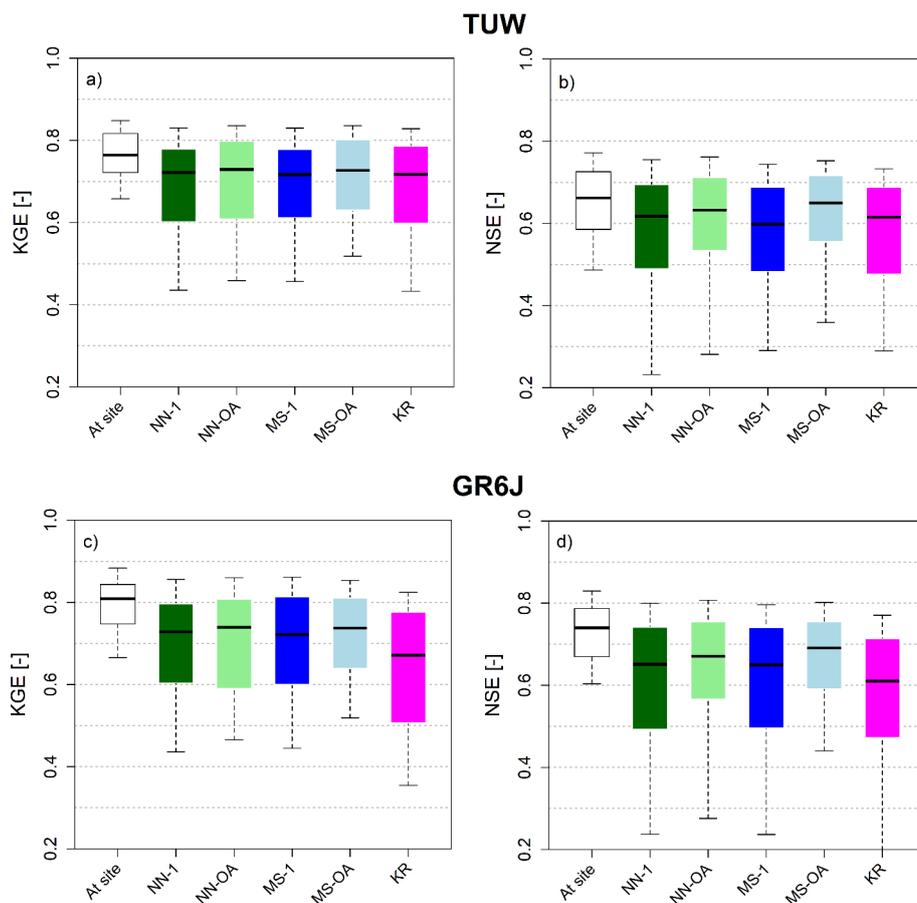
441 For the GR6J model (Figure 7, panels c and d), the efficiencies of the nearest neighbours (NN-1 and NN-OA) and
442 “most similar” (MS-1 and MS-OA) regionalisations are closer to those of the TUW in respect to what happened when
443 the models are calibrated “at site”. In fact, the GR6J model in regionalisation mode deteriorates more than HBV in
444 respect to the parametrization obtained considering the target as gauged.

445 In addition, we notice that, for this model, the ordinary kriging has performances always poorer than all the other
446 regionalisation methods.

447

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

453



454
 455 Figure 7. Original performances of the regionalisation methods for TUW model (panels a and b) and for GR6J model
 456 (panels c and d) for the 209 Austrian catchments in the validation period 1992-2008. Boxes extend to 25% and 75%
 457 quantiles while whiskers refer to 10% and 90% quantiles.

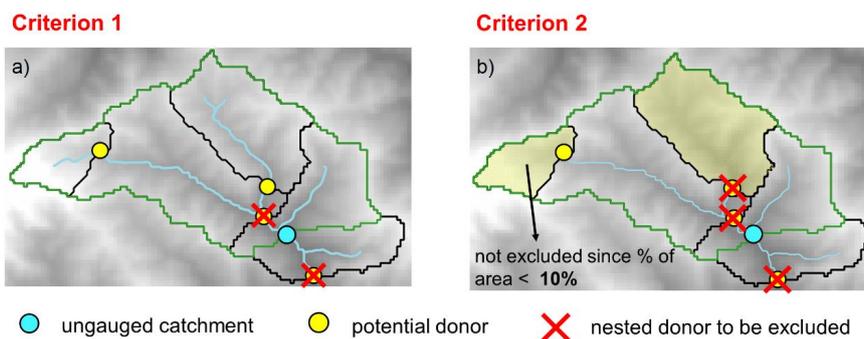
458 4.3 Impact of nested donors: performance losses in regionalisation

459 4.3.1 Catchments identified as nested by the two criteria

460 As introduced in Section 3.3, two different Criteria are implemented for identifying which donor catchments are
 461 considered to be nested in relation to an ungauged catchment: *Criterion 1* (Figure 8, panel a) assumes that the only
 462 nested donors are the first downstream and the first upstream gauged sections. Following this approach, 81% of the
 463 catchments in the dataset have at least one downstream or upstream nested donor (red dots in Figure 9, panel a).
 464 Instead, *Criterion 2* (Figure 8, panel b) excludes all the potential donors sharing a given percentage of drainage area
 465 with the target catchment. It requires the definition of a percentage threshold value of shared drainage area. A
 466 preliminary sensitivity analysis (not reported here) was performed, investigating the effect of different values between
 467 5% and 20% for such percentage. Results show that differences in terms of regionalisation performance are not
 468 significant and it is fixed to 10%. The choice of the threshold influences the number of catchments which can be
 469 included in the study: in fact, the higher is the threshold, the lower is the number of basins classified as nested following

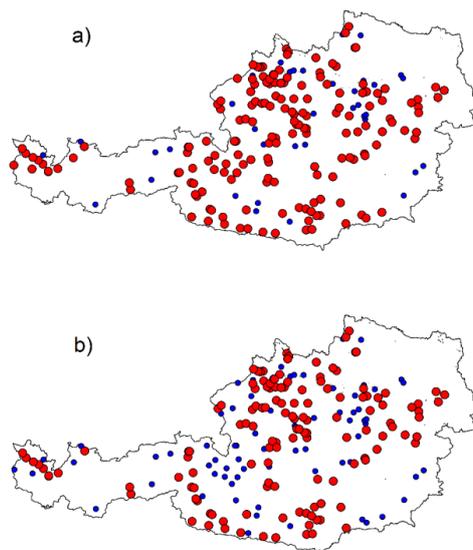


470 *Criterion 2.* Using 10% as a threshold allows to include most of the watersheds in the analysis: 65% (137 catchments)
471 of the basins have at least one nested donor catchment sharing at least the 10% of its area (red dots in Figure 9, panel b).
472 All the watersheds having potential nested donors according to the second criterion have nested gauged catchments also
473 according to the first criterion, but not vice versa: the impact of nested catchments on regionalisation performances is
474 therefore evaluated only for those 137 catchments which are considered to have nested gauged catchments following
475 both criteria.
476 It is important to highlight that the remaining 35% of the basins are still used as potential donor catchments, but the
477 regionalisation approaches are not repeated using such basins as targets (since they have no nested donors, their
478 performance would not change and they would distort the results).
479 Among the 137 catchments considered for the analysis of the “nestedness”, 43% result to have only downstream nested
480 donor(s), 28% only upstream nested donor(s), and 29% at least one upstream and one downstream nested donors.
481



482
483
484

Figure 8. Criteria for excluding nested catchments when regionalising model parameters.



485

486 Figure 9. Panel a: red dots (170) refer to catchments with at least one upstream or downstream nested gauged
487 catchments (Criterion 1). Panel b: red dots (137) refer to catchments with at least one nested gauged catchment
488 more than 10% of drainage area (Criterion 2).

489

490 4.3.2 Performance losses in regionalisation when excluding nested donors

491 The regionalisation methods are applied again in leave-one-out cross validation, this time excluding from the available
492 donors the catchments which are nested in relation to the target (ungauged) basin. This is done for both “nestedness”
493 criteria (down/upstream or overlapping of drainage area) and the analysis applies exclusively to the 137 catchments
494 classified as nested according to both criteria (red dots in Figure 9, panel b). The figures of this section (Figures 10 to
495 13) therefore refer to such subset.

496

497 Figures 10 and 11 compare the different performances (Kling-Gupta and Nash-Sutcliffe efficiencies in the upper and
498 lower panels respectively) obtained in regionalisation (always over the validation period), when nested catchments are
499 available or not as candidate donor basins for both TUW model (Figure 10) and GR6J (Figure 11). Each group of
500 boxplots refers to a different regionalisation method: within such groups, the first box indicates the performance when
501 no basins are excluded from the donor set, while the second and the third boxes report the performances due to the
502 exclusion of the nested following Criterion 1 or 2 respectively.

503

504 The performance deterioration is highlighted by bar plots in Figures 12 and 13, showing the mean loss in Kling-Gupta
505 and Nash-Sutcliffe efficiencies when excluding nested following the two criteria.

506

507 Finally, Table 5 reports the interquartile variability of Kling-Gupta e Nash-Sutcliffe efficiencies for both models and all
508 the regionalisation approaches when nested donors are excluded or not.



509

510 The method that is less affected is the ordinary kriging, especially for the HBV model, due to the fact that such method
511 is not based on the identification of one or more ‘sibling’ donors which may have been excluded if nested. On the other
512 hand, it should also be highlighted that such method is the regionalisation approach that performs worst, when nested
513 basins are available.

514

515 As expected, for both TUW and GR6J, NN-1 is always the most heavily affected method (dark green bars in bottom
516 panels of Figure 12 and 13): this is due to the fact that the nearest donor is a nested one in more than 80% of the
517 catchments, for both criteria and its exclusion seriously compromise the performance.

518

519 Excluding the nested catchments has also a strong impact on MS-1 (dark blue bars in bottom panels of Figures 12 and
520 13), even if to a lesser extent than for NN-1, since for more than 60% of the catchments the most similar donor is a
521 nested one according to both criteria.

522

523 The degradation of performance moving from Criterion 1 (upstream/downstream) to Criterion 2 (overlapping drainage
524 area) highlighted in Figure 10 and 11 demonstrates that using as donors not only the immediate downstream or
525 upstream gauged river sections, but also all the catchments partially sharing their drainage area with the target one, have
526 a strong positive influence on the regionalisation performance.

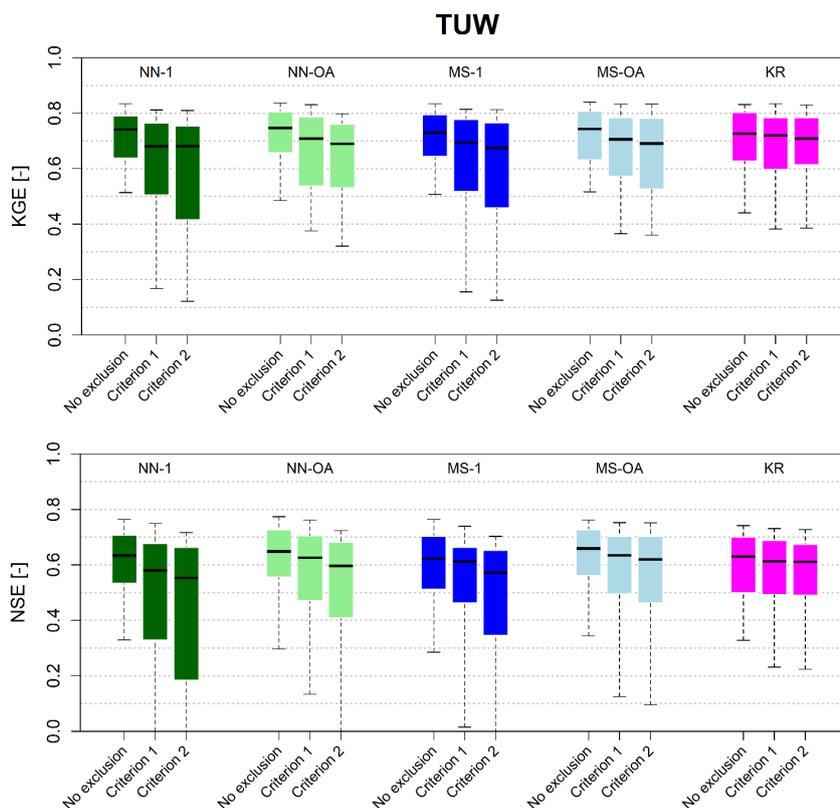
527

528 Furthermore, it is clear how the use of output-averaging for both nearest neighbour and “most similar” approaches (NN-
529 OA and MS-OA), in addition to perform better than the NN-1 and MS-1 when using all (nested and non-nested) donors
530 (see also Section 4.3) can also improve the robustness of the methods to the exclusion of the nested donors: the bottom
531 panels of Figures 12 and 13 in fact show that the loss in the efficiencies of NN-OA and MS-OA are always smaller
532 than those corresponding to the single donor approaches (NN-1 and MS-1), for both rainfall-runoff models and for both
533 regionalisation methods. This confirms that the use of output-averaging (or more in general the use of more than one
534 donor basin) is preferable for regionalisation purposes also for regions that do not have so many nested catchments as
535 the Austria study area.

536

537 Finally, the values reported in Table 5 (as well as Figure 12 and 13) shows how, especially for NSE, the losses resulting
538 when excluding nested donors from the regionalisation are higher for the GR6J model than for the HBV: the GR6J
539 seems to be slightly more affected by the presence of nested basins, except for MS-1 and MS-OA whose performances
540 remain more similar to those of TUW.

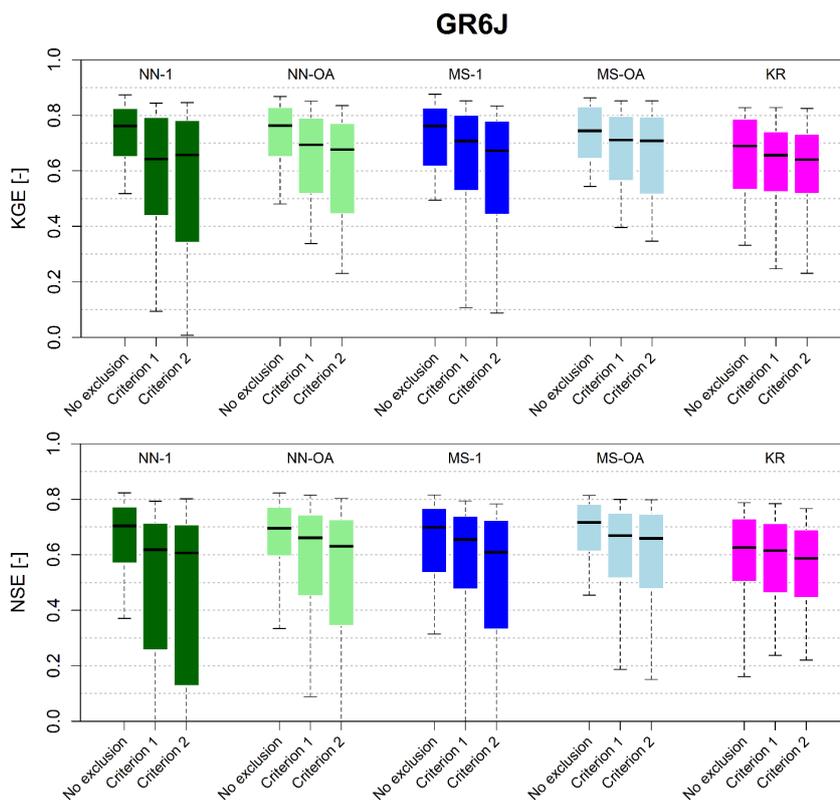
541



542
 543 Figure 10. Effect of the exclusion of nested catchments for the subset of 137 watersheds classified as nested: Kling-
 544 Gupta (upper panel) and Nash-Sutcliffe (lower panel) efficiencies when regionalising the TUW model. “No exclusion”:
 545 all the donors are available. “Criterion 1” or “Criterion 2”: nested catchments are excluded from donor set. Box colours
 546 refer to the different methods: green is nearest neighbour (1 donor is dark green and 3 is light green), blue is most
 547 similar (1 donor is dark blue and 3 is light blue) and magenta is ordinary kriging. Boxes extend to 25% and 75%
 548 quantiles while whiskers refer to 10% and 90% quantiles.
 549

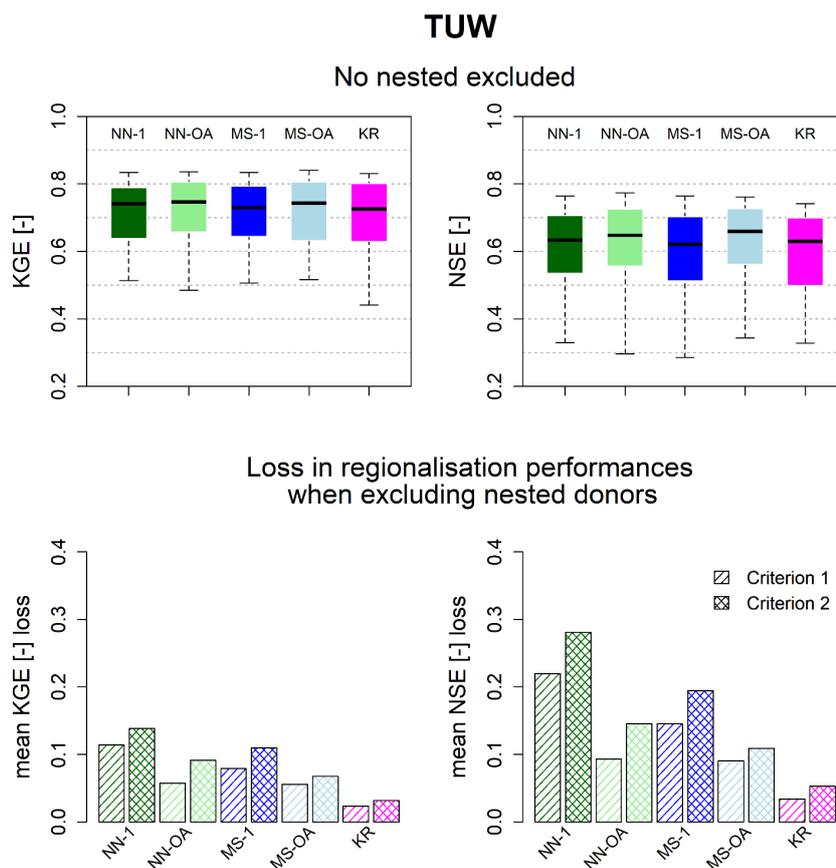


550



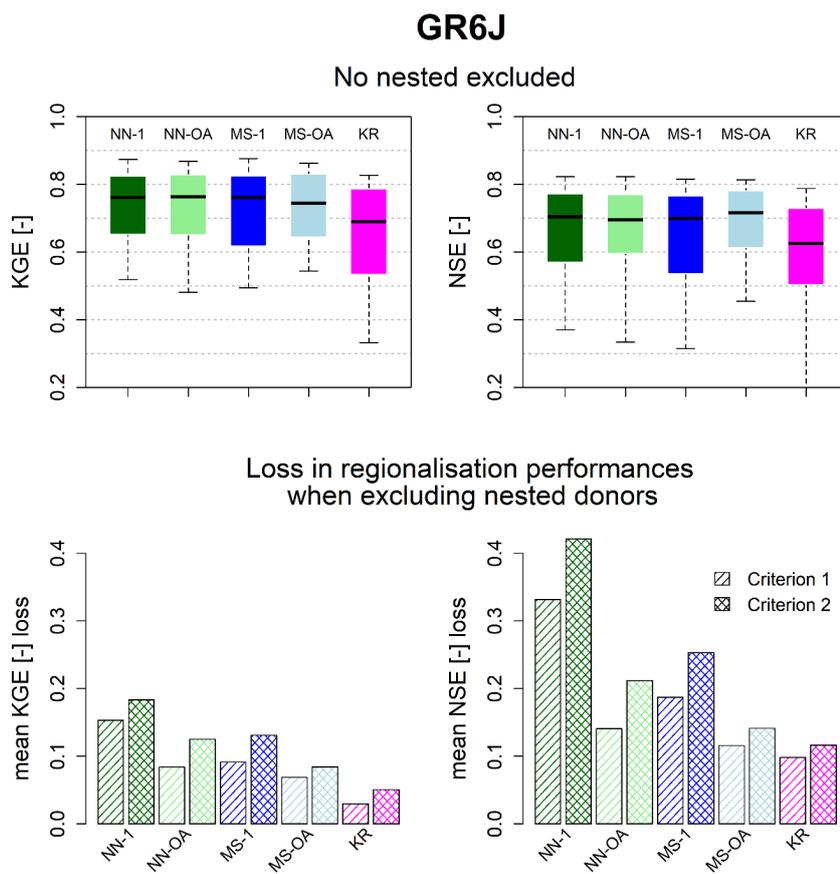
551
 552
 553
 554
 555
 556
 557
 558

Figure 11. Effect of the exclusion of nested catchments for the subset of 137 watersheds classified as nested: Kling-Gupta (upper panel) and Nash-Sutcliffe (lower panel) efficiencies when regionalising the GR6J model. “No exclusion”: all the donors are available. “Criterion 1” or “Criterion 2”: nested catchments are excluded from the donor set. Box colours refer to the different methods: green is nearest neighbour (1 donor is dark green and 3 is light green), blue is most similar (1 donor is dark blue and 3 is light blue) and magenta is ordinary kriging. Boxes extend to 25% and 75% quantiles while whiskers refer to 10% and 90% quantiles.



559
 560
 561
 562

Figure 12. Kling-Gupta and Nash-Sutcliffe efficiencies and mean losses in the same criteria resulting when excluding the nested donors with Criterion 1 and 2 (bottom panels) for TUW model.



563
 564
 565
 566
 567
 568

Figure 13. Kling-Gupta and Nash-Sutcliffe efficiencies and mean losses in the same criteria resulting when excluding the nested donors with Criterion 1 and 2 (bottom panels) for GR6J model.

Table 5. Inter-quartile values of Kling-Gupta and Nash-Sutcliffe efficiencies when regionalising TUV and GR6J models excluding or not excluding nested donor catchments.

		Inter-quartile KGE [-]				
		NN-1	NN-OA	MS-1	MS-OA	KR
TUV	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
TUV	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



569 4.4 Impact of station density: performance losses in regionalisation

570 The last results concern the analysis on the impact of station density on regionalisation performances. As introduced in
571 Section 3.4, for decreasing values of station density across Austria, 100 random samples of stream gauges are generated
572 from the 209 catchments data set and the regionalisation methods are repeated in leave-one-out cross validation over
573 each one of such samples. Seven different values of station density from 0.3 to 2.1 gauges per 1000 km² are tested,
574 which correspond to a total number of stations across Austria from 25 to 175.

575 For each assigned density value, the described procedure provides 100 different sets of regionalised target catchments.
576 For a given density, each one of these 100 subsamples is formed by the same number of target catchments, resulting
577 therefore in the same number of efficiencies to be analysed.

578 In order to analyse the results, the median regionalisation performances of each subsample are computed and presented
579 here: thus, for each gauging density, the results consist in 100 values of median performances.

580 For sake of brevity, only the median Kling-Gupta efficiencies over the validation periods are reported. They are shown
581 in Figure 14 for TUW model and in Figure 15 for GR6J model: each plot contains the boxplots of the median Kling-
582 Gupta efficiencies for each station density (i.e. number of gauges per 1000 km²), that is, each boxplot presents the 100
583 values of median Kling-Gupta efficiencies obtained applying the regionalisation approaches to the 100 subsamples
584 generated with an assigned density. The colored point and the dotted line in the plots indicate the “original” (and
585 maximum) median regionalisation efficiency of the approaches, that is the one obtained when using all available donors
586 (i.e. actual station density, corresponding to 2.4 gauges/1000 km²).

587

588 The NN-1 method (Figures 14 and 15, panels a) is the most affected by the decreasing density. In fact, when the density
589 declines, there is a higher probability that the less dense subsamples do not include the catchment that is the nearest
590 one to each target river section. And, as we have seen in the analyses on the nested donors, in the large majority of the
591 cases, the nearest catchment is a nested one, whereas the second best may be substantially different from the target
592 basin.

593 Also the output-averaging version of the nearest neighbours method (Figures 14 and 15, panels b) strongly deteriorates
594 for less dense networks. In general, nearest neighbour methods are highly sensitive to gauging density: geographical
595 distance results to be a good similarity measure only for densely gauged study area (like Austria), since they firmly rely
596 on the presence of gauged catchments in the immediate surroundings that are also hydrologically very similar. If the
597 density decreases, the closest donor may be relatively far from the target, and it may therefore have little in common
598 with it.

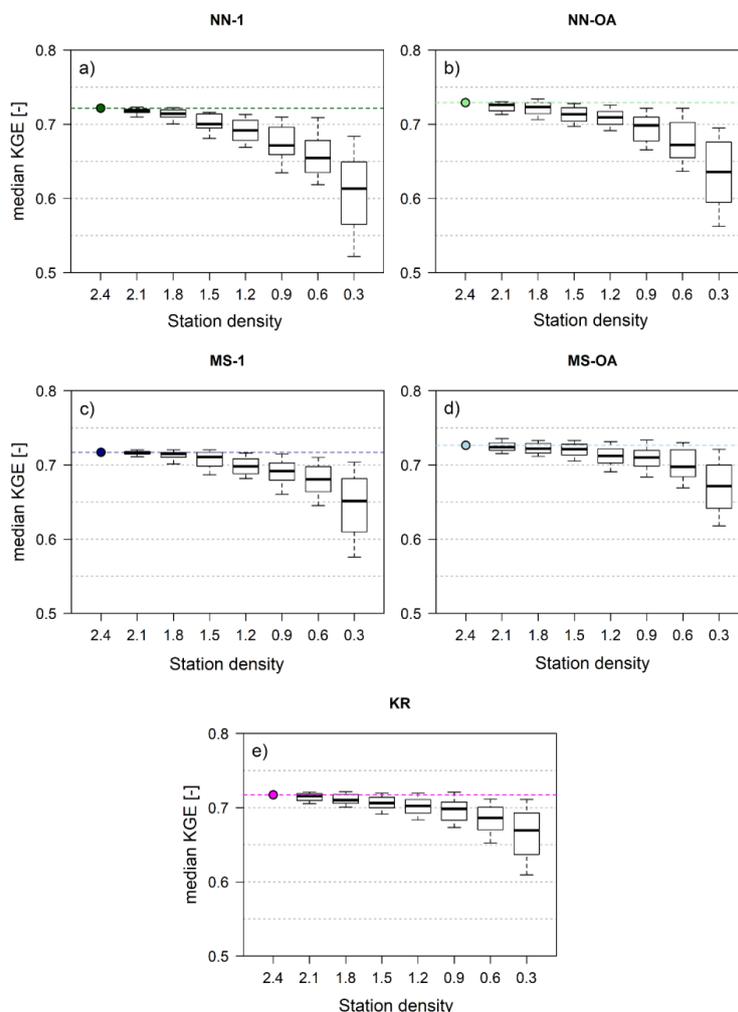
599 As far as the MS-1 (Figures 14 and 15, panels c) is concerned, its performances degrades more gracefully (with the
600 exception of the GR6J model for the minimum density) than the NN-1 or the NN-OA. Also in this case (like for the
601 NN-1), when the density decreases it becomes less probable that the most hydrologically similar catchment (identified
602 by MS-1 in full density) is still part of the subsample; but it is also true there is more than one catchment in the original
603 data set that is similar enough to the target in terms of catchment attributes.

604 This holds also for the output-averaging MS (Figures 14 and 15, panels d), which is even less affected by a reduction in
605 donors' density and is the best-performing approach for any density (for both rainfall-runoff models).

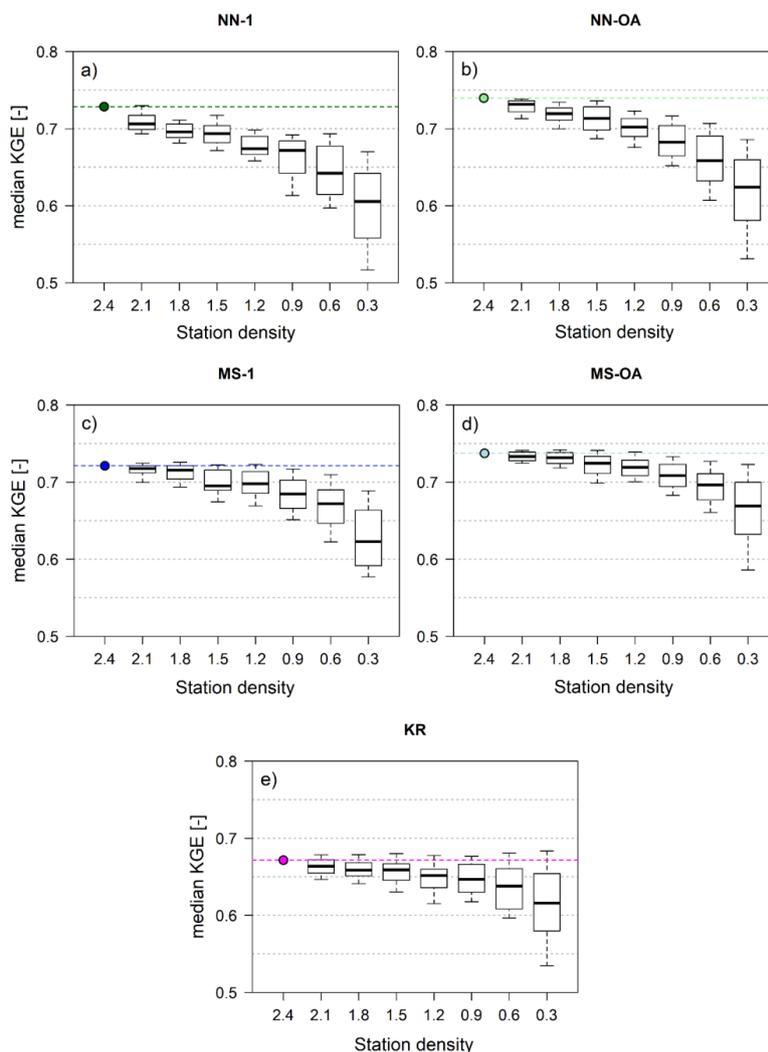
606 We may note that, also in this analysis, analogously to what resulted for the exclusion of nested catchments, for both
607 approaches (NN and MS), the implementation of output-averaging allows to reduce the degradation in the performances
608 in comparison to the corresponding 1-donor version.



609 The impact of station density is similar to that of excluding nested catchments also for the ordinary kriging approach
 610 (Figures 14 and 15, panels e), which deteriorates less than the other methods for decreasing values of station density.
 611 For the TUW model, the kriging regionalisation, starting from an already high KGE in full density, results in
 612 performances that are inferior only to those of MS-OA when the density goes below 0.9. For the GR6J model, even if
 613 the deterioration is limited, since the kriging was poorly performing for the full density regionalisation (Figure 7), the
 614 median KGE is always worse than those of all the other regionalisation approaches, for all the station densities.
 615 Overall, all methods (excluding the poorly performing NN-1 and the kriging for the GR6J) result in relatively good
 616 performances provided that the station density is at least 0.9 gauges per 1000 km². On the other hand, leaving aside the
 617 kriging method, the median KGE drops very steeply when the density passes from 0.6 to 0.3 gauges per 1000 km².



618
 619 Figure 14. Median Kling-Gupta efficiency of the 100 sampled datasets for varying station density (number of gauges
 620 per 1000 km²) for the TUW model using NN-1 (panel a), NN-OA (panel b), MS-1 (panel c), MS-OA (panel d) and KR
 621 (panel e) regionalisation methods. The colored point and dotted line in the plots indicate the original median
 622 regionalisation efficiency of the approaches when using all available donors (i.e. actual station density, corresponding to
 623 2.4 gauges/1000 km²).



624
625 Figure 15. Median Kling-Gupta efficiency of the 100 sampled datasets for varying station density (number of gauges
626 per 1000 km²) for the GR6J model using NN-1 (panel a), NN-OA (panel b), MS-1 (panel c), MS-OA (panel d) and KR
627 (panel e) regionalisation methods. The colored point and dotted line in the plots indicate the original median
628 regionalisation efficiency of the approaches when using all available donors (i.e. actual station density, corresponding to
629 2.4 gauges/1000 km²).

630 5 Discussion and conclusions

631 An assessment of the impact of the presence of nested catchments and of station density on the performance of
632 parameter regionalisation techniques in a large Austrian dataset has been performed. The main motivation for this work
633 lies in the lack of systematic studies in the literature about the effect of data-richness and information content when
634 evaluating the accuracy of various methods for transferring rainfall-runoff model parameters to ungauged catchments.
635 In fact, studies conducted on different study sets often do not lead to the same ranking of the tested approaches and the



636 obtained results are not extendable to different study regions. This is indeed due also to the diverse topological
637 relationships between catchments (“netstedness”) in the datasets and to the diverse density of the streamgauges.

638

639 The purpose of the work is to give support to the choice of the most appropriate parameter regionalisation approaches,
640 showing and quantifying if and how the presence of several nested catchments in a dataset or the gauging density can
641 distort the predictive power of a certain technique.

642

643 The research has been conducted for a very densely gauged dataset covering a large portion of the Austrian country.
644 Two rainfall-runoff models for simulating daily streamflow have been calibrated for the 209 study watersheds: a semi-
645 distributed version of the HBV model (TUW model), and the lumped GR6J model coupled with the Cemaneige snow
646 routine.

647 Both models perform very well when applied in “at-site” mode, that is when parameterised in the traditional, (not
648 regionalised) way, and for each target section the historical gauged streamflow data are used for fitting the parameter
649 set. The calibration and validation performances are very good for both rainfall-runoff models, with better values of the
650 chosen goodness-of-fit indexes for the GR6J model, which demonstrates to perform very well also in this Alpine
651 dataset.

652

653 In order to assess the capability of the models when used on ungauged basins, the streamgauge data for every section
654 was, in turn, considered not to be available, and five regionalisation approaches were implemented for using the
655 rainfall-runoff models in such ‘ungauged’ sections over the validation period. This is indeed an exacting task because
656 we are attempting to use the model over an ungauged catchment and for an observation period different from the one
657 used for parameterising the gauged donor catchments. The first regionalisation approach is an ordinary kriging
658 approach (KR), which separately interpolates each of the model parameter based on their spatial correlation in the study
659 area. Two approaches selecting one single donor catchment and transposing its parameter set to the target basin are also
660 tested: in the first (NN-1) the geographically nearest catchment is selected, while in the second approach (MS-1) the
661 single donor that “lends” all its parameters to the target one is the most similar one in terms of a set of physiographic
662 and climatic attributes. The latter two approaches are implemented also in the output-averaging (OA) version, where the
663 entire parameter set of more than one donor is used for the simulation on the target section and the model outputs are
664 then averaged accordingly to the distance/dissimilarity between donors and target.

665 In regionalisation mode, the performances of the GR6J model deteriorates more than those of the TUW model, in
666 comparison with the ‘gauged’, at-site parameterisation. For both rainfall-runoff models, the use of the output averaging
667 approach outperform the use of a single donor (NN-OA and MS-OA performed better than NN-1 and MS-1),
668 confirming the outcomes of other studies on the importance of exploiting the information available from more than only
669 one donor (see e.g., McIntyre et al. 2005, Oudin et al. 2008, Viviroli et al. 2009, Zelelew and Alfredsen 2014). The
670 output-averaging methods also outperform the parameter-averaging kriging method (especially for the GR6J model),
671 showing that it is preferable transferring the entire parameter set of each donor, thus maintaining the correlation
672 between the parameter values. The results of the MS-OA are close but tend to be better than those of the NN-OA,
673 indicating that hydrological similarity is more important than geographical closeness for choosing the donors.

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



677

678 The impact of the “richness” of the data set was then analysed, in order to assess the deterioration of the regionalisation
679 approaches for decreasing availability and ‘worth’ of the available donors, starting from the influence of using nested
680 basins as donors.

681

682 Two criteria have been proposed for identifying a basin that is nested with the target one: the first one, already used in
683 the few analysis of “nestedness” in the literature, classifies as nested the first upstream and the first downstream gauges
684 on the river network. The second, novel criterion, identifies as nested all the catchments that share more than a given
685 percentage (here chosen as 10%) of the drainage area with the target one. It results that the first criteria includes in the
686 list of basins being nested with at least one potential donor many more sections than those identified by the second
687 criteria. In fact, the first criterion considers as nested also a number of catchments that share less than 10% of area with
688 the target one: this means that, in some cases, the first downstream or upstream gauge may be not representative of the
689 same drainage area and their catchments may be governed by very different hydrological processes.

690

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

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

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

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

713

714 Finally, an assessment of the impact of station density on the regionalisation has been also implemented. The nearest
715 neighbour approaches (both NN-1 and NN-OA) are the methods that suffer more from the decrease in gauging density,
716 whereas the “most similar” methods (MS-1 and MS-OA), which use as similarity measure a set of catchment
717 descriptors, are more capable to adapt to less dense datasets: in fact the “most similar” methods are able to find other



718 adequate donors, that may be anywhere in the region, whereas the nearest neighbours techniques, in a more ‘sparse’
719 monitoring network risk to identify a “not so near” donor that may be very different from the target one.
720 The impact of decreasing station density on the performance of the output-averaging approach based on spatial
721 proximity (NN-OA) is in line to what observed by Lebecherel et al. (2016).
722 The performances of both the output-averaging methods, in agreement with the results obtained for similar methods by
723 Oudin et al. (2008), strongly deteriorate when the station density drops below 0.6 gauges per 1000 km².
724
725 The study confirms how the predictive accuracy of parameter regionalisation techniques strongly depends on the
726 information content of the dataset of available donor catchments, quantifying the contribution of nested catchments and
727 station density for different approaches and rainfall-runoff models. The outcomes obtained in reference to the Austrian
728 data set indicate that the reliability and robustness of the regionalisation of rainfall-runoff model parameters can be
729 improved by making use of output-averaging approaches, that use more than one donor basin but preserving the
730 correlation structure of the parameter set. Such approaches result to be preferable for regionalisation purposes in both
731 data-poor and data-rich regions, as demonstrated by the analyses on the degradation of the performances resulting from
732 either removing the nested donor catchments or decreasing the gauging station density.

733 References

- 734 Bao, Z., Zhang, J., Liu, J., Fu, G., Wang, G., He, R., Yan, R., Jin, J., and Liu, H.: Comparison of regionalization
735 approaches based on regression and similarity for predictions in ungauged catchments under multiple hydro-
736 climatic conditions, *J. Hydrol.*, 466-467, 37-46, <https://doi.org/10.1016/j.jhydrol.2012.07.048>, 2012.
- 737 Bergström, S.: Development and application of a conceptual runoff model for Scandinavian catchments, Dept. of Water
738 Resour. Engineering, Lund Inst of Technol./Univ. of Lund, Bull. Ser. A, No. 52, 1976.
- 739 Burn, D.H., and Boorman, D. B.: Catchment classification applied to the estimation of hydrological parameters at
740 ungauged catchments, Wallingford, Institute of Hydrology, vol. 143, pp. 429–454, 1992.
- 741 Ceola, S., Arheimer, B., Baratti, E., Blöschl, G., Capell, R., Castellarin, A., Freer, J., Han, D., Hrachowitz, M.,
742 Hundecha, Y., Hutton, C., Lindström, G., Montanari, A., Nijzink, R., Parajka, J., Toth, E., Viglione, A., and
743 Wagener, T.: Virtual laboratories: new opportunities for collaborative water science, *Hydrol. Earth Syst. Sci.*,
744 19, 2101–2117, <https://doi.org/10.5194/hess-19-2101-2015>, 2015.
- 745 Cislaghi, A., Masseroni, D., Massari, C., Camici, S., and Brocca, L.: Combining a rainfall–runoff model and a
746 regionalization approach for flood and water resource assessment in the western Po Valley, Italy, *Hydrol. Sci.*
747 *J.*, <https://doi.org/10.1080/02626667.2019.1690656>, 2019.
- 748 Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G.F.: Decomposition of the mean squared error and NSE
749 performance criteria: Implications for improving hydrological modelling, *J. Hydrol.*, 377, 1–2, 80–91,
750 <https://doi.org/10.1016/j.jhydrol.2009.08.003>, 2009.
- 751 Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B.,
752 Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W.,
753 Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R.
754 A., Zehe, E., and Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB)—a review. *Hydrol. Sci. J.*,
755 58, 6, 1198–1255, <https://doi.org/10.1080/02626667.2013.803183>, 2013.
- 756 He, Y., Bárdossy, A., and Zehe, E.: A review of regionalisation for continuous streamflow simulation, *Hydrol. Earth*
757 *Syst. Sci.*, 15, 3539–3553, <https://doi.org/10.5194/hess-15-3539-2011>, 2011.
- 758 Kokkonen, T.S., Jakeman, A. J., Young, P. C., and Koivusalo, H. J.: Predicting daily flows in ungauged catchments:
759 Model regionalization from catchment descriptors at the Coweeta Hydrologic Laboratory, North Carolina,
760 *Hydrol. Process.*, 17, 11, 2219–2238, <https://doi.org/10.1002/hyp.1329>, 2003.
- 761 Lebecherel, L., Andréassian, V., and Perrin, C.: On evaluating the robustness of spatial-proximity-based regionalisation



- 762 methods, *J. Hydrol.*, 539, 196–203, <https://doi.org/10.1016/j.jhydrol.2016.05.031>, 2016
- 763 Li, H., and Zhang, Y.: Regionalising rainfall-runoff modelling for predicting daily runoff: Comparing gridded spatial
764 proximity and gridded integrated similarity approaches against their lumped counterparts, *J. Hydrol.*, 550, 279–
765 293, <https://doi.org/10.1016/j.jhydrol.2017.05.015>, 2017.
- 766 Lindström, G., Johansson, B., Persson, M., Gardelin, M., and Bergström, S.: Development and test of the distributed
767 HBV-96 hydrological model, *J. Hydrol.*, 201, 1–4, 272–288, [https://doi.org/10.1016/S0022-1694\(97\)00041-3](https://doi.org/10.1016/S0022-1694(97)00041-3),
768 1997.
- 769 McIntyre, N. R., Lee, H., Wheeler, H., Young, A., and Wagener, T.: Ensemble predictions of runoff in ungauged
770 catchments, *Water Resour. Res.*, 41, 12, 1–14, <https://doi.org/10.1029/2005WR004289>, 2005.
- 771 Merz, R., and Blöschl, G.: Regionalisation of catchment model parameters, *J. Hydrol.*, 287, 1–4, 95–123,
772 <https://doi.org/10.1016/j.jhydrol.2003.09.028>, 2004.
- 773 Merz, R., Blöschl, G., and Parajka, J.: Regionalisation methods in rainfall-runoff modelling using large samples, *Large
774 Sample Basin Exp. Hydrol. Model Parameterization Results Model Param. Exp. IAHS Publ.*, 307, 307, pp.
775 117–125, 2006.
- 776 Mészáros, I., Mikl'ánek, P., and Parajka, J.: Solar energy income modelling in mountainous areas, in: *RB and
777 NEFRIEND Proj.5 Conf. Interdisciplinary Approaches in Small Catchment Hydrology: Monitoring and Research*,
778 edited by: Holko, L., Mikl'ánek, P., Parajka, J., and Kostka, Z., Slovak NC IHP UNESCO/UH SAV, Bratislava,
779 Slovakia, 127–135, 2002.
- 780 Moore, R. J.: The probability-distributed principle and runoff production at point and basin scales, *Hydrol. Sci. J.*, 30, 2,
781 273–297, <https://doi.org/10.1080/02626668509490989>, 1985.
- 782 Oudin, L., Andréassian, V., Perrin, C., Michel, C., and Le Moine, N.: Spatial proximity, physical similarity, regression
783 and ungauged catchments: A comparison of regionalization approaches based on 913 French catchments, *Water
784 Resour. Res.*, 44, 3, 1–15, <https://doi.org/10.1029/2007WR006240>, 2008.
- 785 Parajka, J., Merz, R., and Blöschl, G.: A comparison of regionalisation methods for catchment model parameters,
786 *Hydrol. Earth Syst. Sci.*, 9, 157–171, <https://doi.org/10.5194/hess-9-157-2005>, 2005.
- 787 Parajka, J., Merz, R., Blöschl, G.: Uncertainty and multiple objective calibration in regional water balance modelling:
788 case study in 320 Austrian catchments, *Hydrol. Process.*, 21, 435–446, <https://doi.org/10.1002/hyp.6253>, 2007.
- 789 Parajka, J., Viglione, A., Rogger, M., Salinas, J. L., Sivapalan, M., and Blöschl, G.: Comparative assessment of
790 predictions in ungauged basins – Part 1: Runoff-hydrograph studies, *Hydrol. Earth Syst. Sci.*, 17, 1783–1795,
791 <https://doi.org/10.5194/hess-17-1783-2013>, 2013.
- 792 Parajka, J., Merz, R., Skøien, J. O., and Viglione, A.: The role of station density for predicting daily runoff by top-
793 kriging interpolation in Austria, *J. Hydrol. Hydromechanics*, 63, 3, 228–234, [https://doi.org/10.1515/johh-2015-
794 0024](https://doi.org/10.1515/johh-2015-0024), 2015.
- 795 Patil, S. and Stieglitz, M.: Controls on hydrologic similarity: role of nearby gauged catchments for prediction at an
796 ungauged catchment, *Hydrol. Earth Syst. Sci.*, 16, 551–562, <https://doi.org/10.5194/hess-16-551-2012>, 2012.
- 797 Patil, S. and Stieglitz, M.: Comparing spatial and temporal transferability of hydrological model parameters, *J. Hydrol.*,
798 525, 409–417, <https://doi.org/10.1016/j.jhydrol.2015.04.003>, 2015.
- 799 Peel M. C., and Blöschl, G.: Hydrological modelling in a changing world, *Prog. Phys. Geogr.*, 35, 2, 249–261,
800 <https://doi.org/10.1177/0309133311402550>, 2011.
- 801 Pushpalatha, R., Perrin, C., Le Moine, N., Mathevet, T., and Andréassian, V.: A downward structural sensitivity
802 analysis of hydrological models to improve low-flow simulation, *J. Hydrol.*, 411, 1–2, 66–76,
803 <https://doi.org/10.1016/j.jhydrol.2011.09.034>, 2011.
- 804 Razavi T., and Coulibaly, P.: Streamflow Prediction in Ungauged Basins: Review of Regionalization Methods, *J.
805 Hydrol. Eng.*, 18, 8, 958–975, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000690](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000690), 2013.
- 806 Reichl, J. P. C., Western, A. W., McIntyre, N. R., and Chiew, F. H. S.: Optimization of a similarity measure for
807 estimating ungauged streamflow, *Water Resour. Res.*, 45, W10423, <https://doi.org/10.1029/2008WR007248>,
808 2009.
- 809 Samuel, J., Coulibaly, P., and Metcalfe, A.: Estimation of continuous streamflows in Ontario ungauged basins:



- 810 comparison of regionalization methods, *J. Hydrol. Eng.*, 16, 5, 447-459,
811 [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000338](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000338), 2011.
- 812 Seibert, J.: Regionalisation of parameters for a conceptual rainfall-runoff model, *Agr. For. Met.*, 98–99, 279–293, 1999.
- 813 Skøien, J. O., Merz, R., and Blöschl, G.: Top-kriging - geostatistics on stream networks, *Hydrol. Earth Syst. Sci.*, 10,
814 277–287, <https://doi.org/10.5194/hess-10-277-2006>, 2006.
- 815 Steinschneider, C., Yang, Y. E., and Brown, C.: Combining regression and spatial proximity for catchment model
816 regionalization: a comparative study, *Hydrol. Sci. J.*, 60, 6, 1026-1043,
817 <https://doi.org/10.1080/02626667.2014.899701>, 2015.
- 818 Tolson B. A., and Shoemaker, C. A.: Dynamically dimensioned search algorithm for computationally efficient
819 watershed model calibration, *Water Resour. Res.*, 43, 1, 1–16, <https://doi.org/10.1029/2005WR004723>, 2007.
- 820 Valéry, A., Andréassian, V., and Perrin, C.: Regionalization of precipitation and air temperature over high-altitude
821 catchments – learning from outliers, *Hydrol. Sci. J.*, 55, 6, 928–940,
822 <https://doi.org/10.1080/02626667.2010.504676>, 2010.
- 823 Valéry, A., Andréassian, V., and Perrin, C.: ‘As simple as possible but not simpler’: What is useful in a temperature-
824 based snow-accounting routine? Part 2 - Sensitivity analysis of the Cemaneige snow accounting routine on 380
825 catchments, *J. Hydrol.*, vol. 517, pp. 1176–1187, <https://doi.org/10.1016/j.jhydrol.2014.04.058>, 2014.
- 826 Viglione, A., Parajka, J., Rogger, M., Salinas, J. L., Laaha, G., Sivapalan, M., and Blöschl, G.: Comparative assessment
827 of predictions in ungauged basins – Part 3: Runoff signatures in Austria, *Hydrol. Earth Syst. Sci.*, 17, 2263–
828 2279, <https://doi.org/10.5194/hess-17-2263-2013>, 2013.
- 829 Viglione A., and Parajka J.: TUWmodel: Lumped/Semi-Distributed Hydrological Model for Education Purposes. R
830 package version 1.1-0, <https://CRAN.R-project.org/package=TUWmodel>, 2019.
- 831 Viviroli, D., Mittelbach, H., Gurtz, J., and Weingartner, R.: Continuous simulation for flood estimation in ungauged
832 mesoscale catchments of Switzerland – Part II: parameter regionalisation and flood estimation results, *J.*
833 *Hydrol.*, 377, 1-2, 208-225, <https://doi.org/10.1016/j.jhydrol.2009.08.022>, 2009.
- 834 Yang, X., Magnusson, J., Rizzi, J., and Xu, C.: Runoff prediction in ungauged catchments in Norway: comparison of
835 regionalization approaches, *Hydrol. Res.*, 49, 2, 487-505, <https://doi.org/10.2166/nh.2017.071>, 2018.
- 836 Zelelew M. B., and Alfredsen, K.: Transferability of hydrological model parameter spaces in the estimation of runoff in
837 ungauged catchments, *Hydrol. Sci. J.*, 59, 8, 1470-1490, <https://doi.org/10.1080/02626667.2013.838003>, 2014.

838 **Appendix A: Choice of best catchment descriptors**

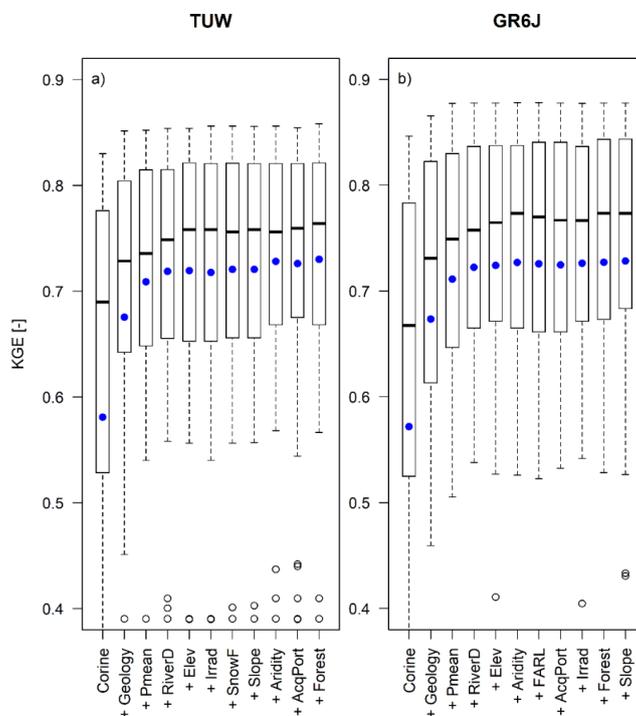
839 The implementation of the “most similar” approach requires the choice of the geo-morphologic and climatic attributes
840 to be used for selecting the donor catchment(s), i.e. to calculate the dissimilarity indices of equation 7.

841 This similarity study is part of a preliminary analysis carried out using the whole period of available daily data (from
842 1976 to 2008, again with 1 year of warm-up) for calibrating the rainfall-runoff models.

843 In order to individuate the best catchment descriptors (all reported in Table 1 with a brief description), the most similar
844 approach with one single donor catchment (MS-1) is applied sequentially to the entire dataset in leave-one-out cross-
845 validation, using at each step an increasing number of attributes when defining the dissimilarity index ϕ . At each step,
846 the method is tested multiple times, adding one by one each of the attributes and the one which gives the best
847 regionalisation performances is selected. For greater clarity, Figure A1 (panel a) refers to TUW and panel b) to GR6J)
848 shows the boxplots of the consecutive best combinations of descriptors: at the first step, only one attribute is used, the
849 most similar approach is tested for all the available catchment features, and the similarity in the land cover classes
850 (Corine) gave the best efficiency. At the second step, the operation is repeated using land cover and each of the
851 remaining attributes one at a time, finding the geology classes to be the best attribute to add, and so on. The analysis
852 stops when the performances are decreasing or stop improving.



853 As can be inferred from Figure A1, both rainfall-runoff models reach good regionalisation performances when using up
854 to 5 attributes. Since the first best 5 attributes are the same for both models and from the sixth step the performances are
855 not substantially improved, we decide to choose those five descriptors to characterize catchment similarity: land use
856 classes, geological classes, mean annual precipitation, stream network density and mean elevation.
857



858

859 Figure A1. Kling-Gupta efficiencies for TUW (a) and GR6J (b) models for the consecutive steps of the similarity
860 analysis. Boxes refer to 25% and 75% quantiles, whiskers refer to 10% and 90% quantiles and the blue points to the
861 average.