Diploma Thesis

Dealing with artificial trans-catchment diversions during runoff interpolation with TopKriging

submitted in satisfaction of the requirements for the degree of
Diplom-Ingenieur
of the TU Wien, Faculty of Civil Engineering

Diplomarbeit

Methodenentwicklung zur Berücksichtigung von künstlichen Überleitungen zwischen Pegeleinzugsgebieten bei der Interpolation von Jahresabflusswerten mit TopKriging

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines
Diplom-Ingenieurs
eingereicht an der Technischen Universität Wien, Fakultät für Bauingenieurwesen

von

Nikolaus Weber, BSc

Matr.Nr.: 01226339

unter der Anleitung von

Univ. Prof. Dipl.-Ing. Dr. **Matthias Zessner**

Dipl.-Geoökol. **Steffen Kittlaus**

Institut für Wassergüte und Ressourcenmanagement
Forschungsbereich Wassergütewirtschaft
Technische Universität Wien
Karlsplatz 13/226, 1040 Wien, Österreich

Wien, im Dezember 2020

Acknowledgements

At this point I would like to thank my supervisor, Dipl.-Geoökol. Steffen Kittlaus, who supported me during the whole work in a very committed and professional way. His useful advice and constructive criticism have helped me a lot.
A special thanks goes to Univ. Prof. Dipl.-Ing. Dr. Matthias Zessner who made it possible to write my diploma thesis in the Research Center for Water Quality Management at the Institute for Water Quality and Resource Management (IWAG).
To work as student assistant during the past two years was very instructive and important for me and I would also like to thank all employees of the Research Center for Water Quality Management for the motivating and pleasant working atmosphere.
Many thanks to all companies and authorities who kindly supported this work with data.
Sincere thanks to my fellow students for their mutual support, help and motivation during my studies of civil engineering at TU Wien. Especially to Michael, Doğan, Martin, Kristýna, Philipp, Alan and Sophia.
I also want to thank my parents, siblings and friends for their support, advice and open ears while I was going my own path. My friend Sunita deserves a big thank-you for her motivation and support during high school graduation. My wife and partner Nona deserves the most thanks for her love, patience and encouragement.
Finally, I would like to thank Preston A. Long, Ph.D and DI Maresi Grabner for the grammatical, formal and content-related proofreading.

Abstract

Diversions, artificial interventions on rivers, are abundant all over the alpine regions due to the high number of hydropower stations, but also in the mid and low lands which cause a disturbance of the natural flow regime. Specially, those between catchments, hence trans-catchment diversions, lead to uncertainties in the observed datasets that are used as input to current catchment scale emissions models like MoRE. Those uncertainties influence the calculation and further the results, which consequently reduce the accuracy of those models. It is a bother to those who work with these models or depend on their results. Although these models rely on accurate and reliable data as input, surprisingly very little research addresses this topic.

This issue is discussed in the submitted master thesis. A method is evaluated to take trans-catchment diversions in account, to improve input data and to reduce data uncertainty. Diversions were related to a diversion area, which were in many cases already available, and with that the catchment area of a runoff gauging station was corrected.

This approach is applied to an input dataset with annual MQ values for the STOBIMO project, a substance transport study using the MoRE model, where problems with diversions have been reported. Around 17% of the runoff gauging stations in the study area are affected by diversions. Therefore, the relevant diversions were identified, quantified, and used to correct the runoff gauge data. With TopKriging an interpolation was carried out and then transformed to an appropriated input format for the STOBIMO project. To validate this approach, all companies of the related diversions were asked for validation data and most were keen to provide observed diversion data for further use.

Cross-validation indicated that TopKriging prediction efficiency, measured in Nash-Sutcliffe efficiency (NSE), for diversion affected gauges can be increased by 83% from 0.40 to 0.73 and for the whole study area (17% diversion affected gauges) by 11%, hence from 0.63 to 0.70. Prediction comparison showed that MoRE runoff prediction efficiency (NSE) can be significantly improved for the diversion affected gauges by 51% from 0.63 to 0.95 and for the whole study area by 3%, hence from 0.92 to 0.95 in general.

Kurzfassung

Überleitungen sind künstliche Eingriffe an Flüssen und sind, bedingt durch den hohen Grad an Wasserkraftnutzung, in den Alpen reichlich vorhanden. Aber auch in den flacheren Gebieten sind sie vorzufinden und beeinflussen dadurch das natürliche Abflussgeschehen an Flüssen. Besonders Überleitungen zwischen Pegeleinzugsgebieten führen zu einer erhöhten Unsicherheit der and den Pegeln gemessenen Zeitreihen. Diese wiederum finden als Eingangsdaten für aktuelle Emissionsmodellierungen auf Einzugsgebietsebene, wie etwas MoRE, Verwendung.

Diese Unsicherheiten beeinflussen sowohl die Berechnungen als auch die davon gewonnenen Erkenntnisse. Daraus wird die Genauigkeit der Modelle negativ beeinflusst, was ein Ärgernis für jene darstellt, die mit diesen Modellen arbeiten oder von diesen Ergebnissen abhängen. Erstaunlich wenig Forschung wurde diesem Thema gewidmet, obwohl solche Modelle zu einem hohen Grad von genauen und zuverlässigen Eingangsdaten abhängen.

Die vorliegende Diplomarbeit diskutiert dieses Thema und zeigt eine Methode, mit der durch Berücksichtigung von Überleitungen zwischen Pegeleinzugsgebieten die Eingangsdaten solcher Modelle sowie deren Unsicherheiten verbessert werden. Als Bezugsgröße für Überleitungen wurde die Überleitungsfläche verwendet, welche in vielen Fällen von offiziellen Quellen bezogen wurde. Damit wurden dann die Einzugsflächen der betroffenen Pegel korrigiert.

Das STOBIMO Projekt führt eine Stoffbilanzmodellierung, basierend auf dem MoRE Modell, für Spurenstoffe auf Einzugsgebietsebene aus. An einem Eingangsdatensatz dieseses Projektes welcher aus jährlichen mittleren Abflüssen (MQ) besteht, wird diese Methode angewendet.
Im STOBIMO Projektbericht [3] wurde über Unsicherheiten durch Überleitungen bereits berichtet. Denn ungefähr 17% der verwendeten Pegel im Projektgebiet sind beeinflusst durch Überleitungen. Daher wurden die relevanten Überleitungen identifiziert, quantifiziert und verwendet, um die gemessenen Datenreihen der Pegel zu verbessern. Mit TopKriging wurde eine Interpolation der gemessenen Abflussspenden durchgeführt und anschließend in ein geeignetes Format für das MoRE Modell des STOBIMO Projektes überführt. Um die angewandte Methode zu überprüfen, wurden bei den relevanten Unternehmen Überleitungsdaten erhoben. Die meisten Unternehmen waren interessiert und lieferten Daten zu Überleitungen, welche dann für die Validierung verwendet wurden.

Eine Kreuzvalidierung zeigte, dass die Güte der Interpolationeffizienz mit TopKriging für überleitungsbeeinflusste Pegel um 83%, von 0,40 zu 0,73 Nash-Sutcliffe-Modelleffizienz (NSE), verbessert wurde. Für das ganze Projektgebiet betrugen die allgemeinen Verbesserungen 11%, von 0,63 zu 0,70 NSE. Ein Vergleich der simulierten Abflüsse im MoRE Modell zu den beobachteten Abflüssen der Pegel, zeigte für überleitungsbeeinflusste Pegel eine Verbesserung der Prognoseeffizienz um 51%, von 0,63 zu 0,95 NSE. Verbesserungen gab es auch für das ganze Projektgebiet, diese betrugen 3%, von 0,92 zu 0,95 NSE.

Contents

Chapter 1

Introduction

Water is an unique resource on our planet. It is a heritage, not a commercial product and therefore must be protected, defended and treated as such [1].

Due to the continuous growth in demand for sufficient quantities of clean water by our society, the worlds waterbodies (rivers and lakes, groundwater and bathing waters) became under increased pressure [1]. Protection of the water bodies is therefore a priority of society and policymakers to protect them from pollution, and in many cases to clean polluted water bodies. The increasing public awareness in the past decade put a pressure on the policy makers to tackle these threats and to have a more holistic view on the resource water.

In the year 2000, the European Parliament established a framework, called EU Water Framework Directive (WFD), to meet that concern. Its main goals are to expand the scope of policies from administrative or political boundaries to river basins, the natural geographical and hydrological units [1], to expand the scope of water protection to all waterbodies. To ensure high level protection, as "good status" by chemical, ecological and hydrological standards achieved through the "combined approach", hence by considering both the source and receiving environment equally [1]. Further goals are to set appropriate prices to the resource water and to involve public participation.

The co-ordination of the WFD goals requires a "river management plan" for each river basin. The plan accounts for analysis of the current and target status, the requirements to achieve these objectives, and tangible actions for implementation [1]. Once this document is established, it is updated every 6 years, to summarize the achievements, highlight the problematic fields, and to set new or adjust existing programs as well as projects of measurements for progressive achievement of the WFD goals. The transfer to the member states is done by the National water management plan (NWP).

1.1 Motivation

In the second half of the last century great improvements could be achieved in terms of water quality, by cleaning our wastewater emissions from nutrients like carbon, nitrogen and phosphorus. Now the research is focusing more and more on the micro pollutant, hence hazardous substances, in our rivers. Their visibility and effects are less direct in nature but nevertheless harmful to human and environment. Gender-change of fish caused by hormones, micro plastic or painkillers in rivers to name a few topics of concern, which even received attention by mainstream media.

Hazardous substances are defined in the WFD as "...‘substances or groups of substances that are toxic, persistent and liable to bio-accumulate, and other substances or groups of substances which give rise to an equivalent level of concern’" [2] and are listed as defined in Article 2(30) of the WFD [1].

Article 16 of the WFD aims at the progressive reduction of those priority hazardous substances [1] and therefore the strategy against water pollution through chemical substances has to be considered in the "river management plan" and NWP.

One of those projects to support the Austrian National water management plan (NWP) is the STOBIMO trace substance project [3] by the Austrian Federal Ministry of Agriculture, Regions and Tourism (BMLRT). Its focus is on the identification of the pollution pathways in the study area to derive the best actions to tackle the pathways of relevant priority hazardous substances. Those are selected trace substances, inorganic and organic, which are all listed in the WFD.

The STOBIMO project [3] uses the Modeling of Regionalized Emissions (MoRE) model [4], a pathway-specific substance emission model, as model and software, to calculate the emissions of each pathway and subcatchment unit, called analytical unit (AU). The study area covers nearly all of the Austrian surface and additionally parts of Germany and Switzerland, namely the Inn river regions. A map with the MoRE AU can be seen in figure 1.2. The emission pathways are distinguished between point pathways like wastewater treatment plants and diffuse pathways like agriculture, roads or tile drainage. An overview can be seen in figure 1.1.

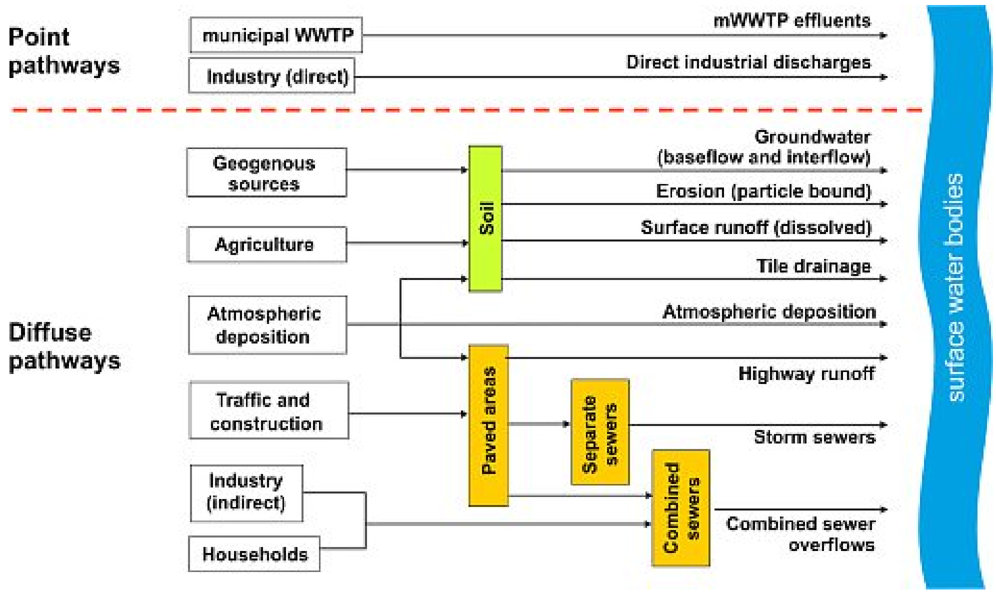


Fig. 1.1: Emission pathways considered in the MoRE model. (Taken from [3])

The MoRE model [4] calculates the trace substance emissions into river bodies by each emission pathway in annual steps for each hydrological subcatchment based on the input parameter provided by the STOBIMO project. One of those input parameters is the net runoff generated in each analytical unit (AU), hence a key parameter in the MoRE model. And a model is only as good as their input parameters, the goal is therefore to find the best possible input data for the MoRE model.

Due to the fact that there are only a limiting number of runoff gauging stations available, those measurements have to be interpolated and extrapolated to net runoff for all analytical units (AUs).

Until now, those net runoffs were calculated as MQ runoffs out of a dataset, containing interpolated daily runoff measurements using measurements from runoff gauging stations [3]. Through comparison between modelled and measured runoff gauge station runoff a twofold problem could be identified. First, due to artificial interventions on riverbodies in the alpine region and second due to a region with very high infiltration rates (Wiener Becken) [3]. Specially, the first problem could not be tackled due to insufficient available data and caused high model uncertainties in those regions. Furthermore, artificial interventions on riverbodies are abundant all over the study area of the STOBIMO project. For example, hydro power diversions in the

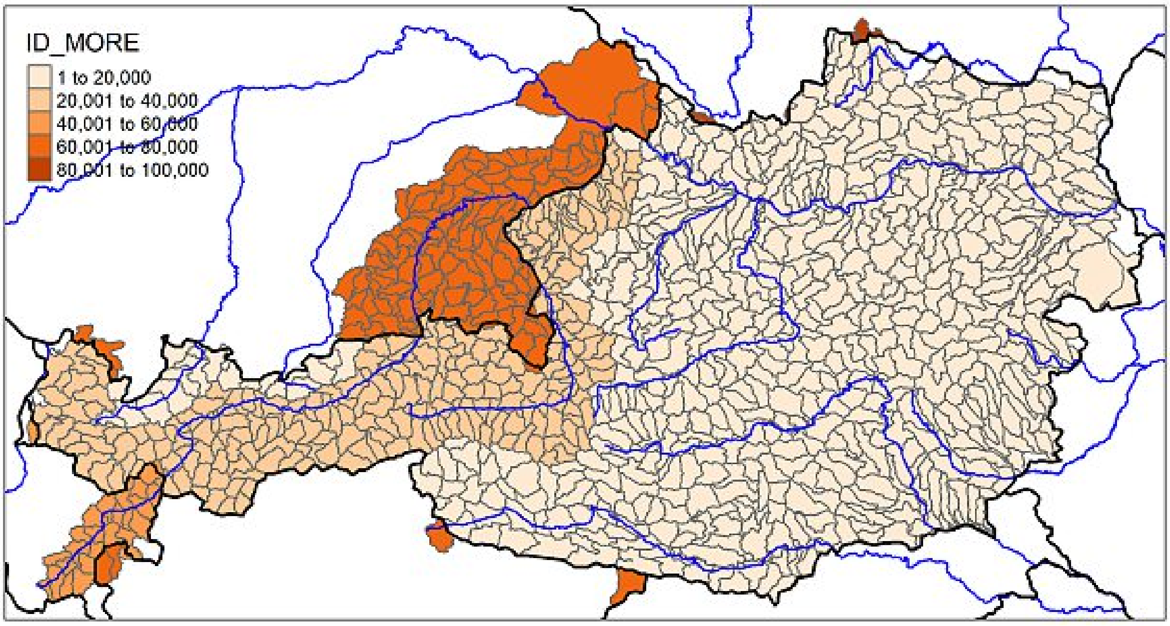


Fig. 1.2: STOBIMO project area with MoRE AU distinguished by *MoRE ID* with main rivers and country borders. (© data.umweltbundesamt.at © EuroGeographics for the administrative boundaries)

high lands (west) or artificial channels in the low lands (north-east) for irrigation, industry or other purposes. All of those affect the runoff regime of rivers and their corresponding watersheds. Amann et al. [3] emphasizes that closing data gaps is necessary to improve model quality to use the full range of the MoRE model.

1.2 Research question

All kinds of those artificial interventions will be defined as trans-catchment-diversions (diversions), meaning an artificial transfer from one river to another, causing increasing runoff in the receiving river and decreasing the runoff downstream of the dividing river. As Wesemann et al. state, many Austrian runoff gauging stations are influenced by reservoir management or artificial diversions [5, p. 2]. Because the measurements from runoff gauging stations are used for the interpolations, the idea is to improve the hydrological data used in the MoRE model of the STOBIMO project by considering diversion. This leads to the following research question:

Which improvement can be achieved in terms of prediction efficiency of net runoff through the consideration of diversion?

Therefore, the purpose of this diploma thesis is to identify diversion within the study area and to develop a method to consider diversion in the interpolation process and the MoRE model. This should result in an improved runoff dataset used as model input for the STOBIMO project [3] causing a more accurate MoRE model [4].

Chapter 2

State of the Art

2.1 River Diversions

Diversions are artificial interventions on riverbodies to redirect water from one place on the river to another. Usually a diversion dam is placed in the river to raise the waterlevel and to redirect the water to its purpose like irrigation, hydropower, or industrial or municipal drinking water use. According to Egré and Milewski [6], river diversions consist of two cases: In-stream diversions dam a river and divert the service water through pipes or tunnels to a downstream position of the same river, causing decreasing runoff downstream of the diversions. Secondly, there are trans-watershed diversions where a river is also dammed and then divided to a further river in another watershed. This causes increasing runoff at the receiving river and decreasing runoff downstream of the diversion. Specially, the second case influences not only parts of the downstream like case one does, but all downstream gauges. In particular, if it examines if the diversion is over basins borders, e.g. between Danube and Rhine river basins.

Diversions decrease not only the runoff immediately after diversion but also influence shore erosion, water temperature, and water quality, Egré and Milewski [6] further assert. The most effective mitigation methode, according to Egré and Milewski [6] is to ensure a minimum ecological flow downstream of a diversion, called residual water. This is in the most cases legally required and as minimum flow requirements prescribed by the authorities.

As Wesemann et al. [5] stated, many Austrian runoff gauging stations are influenced by reservoir management or artificial diversions and therefore do not fully represent the natural (undisturbed) conditions [5, p. 495]. For example, diversions for hydropower stations or drinking water extraction in the high lands or artificial channels in the low lands for irrigation, industry, or other purposes. All those affect the runoff regime of rivers and their corresponding watersheds. Due to this bias the time series of those gauges are affected and are potentially limited suitable for geostatistics and hydrological modelling.

There is a paucity of research to date on the topic of diversion consideration. Wesemann et al. [5] analysed a highly disturbed subcatchment in the Stubach valley in Salzburg to calibrate a rainfall-runoff forecasting model. Beside the usual input data (runoff gauge stations, rainfall, landuse, soil moisture, ect.) secondary informations (timeseries of water intakes, diversion tunnels, pressure lines and artificial reservoir waterlevels) were used to derive local discharges. All in-and-out flows and reservoir changes were balanced to obtain a timeseries resolution of 12 and 24 h. The model achieved a NSE of 0.79 for both simulation and validation. The paper shows that specially for the headwaters in the mountains diversions have huge effects on the predictions. However, this requires sufficient data and time to process it, as the volume of work can be very high.

Besides this paper, no other relevant research could be found related to the topic of diversion consideration.

2.2 Interpolation Methods

One of the main issues in hydrological regionalisation is the prediction of unknown values with a limited number of known values, according to Blöschl [7]. This thesis is of that case, as a limited number of runoff gauge stations provide measurements which are used to predict the runoff at every single point along the river network (flow tree). Geostatistical methods provide a solution to solve that issue by the use of spatial correlation or as Rizo-Decelis et al. [8, p. 277] state, the purpose is to improve predictions for unknown places and to make estimations close to real data measured in situ.

2.2.1 Ordinary Kriging

Ordinary Kriging, first presented by Deutsch und Journe in 1997, is currently widely used in geostatistics to predict the value of a random variable over a spatial region [7]. The predicted value is assumed to be an estimator of the real value $E[z]$ and the method is being described as the best linear unbiased estimator (BLUE) [7]. This means that the expected bias is zero, hence there is no systematic error and the expected kriging error is minimised, and consequently the mean quadratic error is minimized (Eq 2.1):

$$E[(z - \hat{z})] = 0 \qquad \& \qquad E[(z - \hat{z})^2] \rightarrow min \tag{2.1}$$

Equation 2.2 shows that the predicted value is calculated as the weighted average (linear combination) of the n-nearest neighbours. With the variables unknown value $\hat{z}(x_0)$ at position x_0, the interpolated weight λ_i of the measurement $z(x_i)$ at position x_i and the number of neighbours n used for the interpolation:

$$\hat{z}(x_0) = \sum_{i=1}^{n} \lambda_i z(x_i) \tag{2.2}$$

By solving the kriging system in equation 2.3, the unknown weights can be found:

$$\begin{gathered} \sum_{j=1}^{n} \lambda_j \gamma_{ij} - \lambda_i \sigma_i^2 + \mu = \gamma_{0i} \qquad i = 1, ..., n \\ \sum_{j=1}^{n} \lambda_j = 1 \end{gathered} \tag{2.3}$$

The γ_{ij} is called the gamma value of the theoretical semivariogram model, hence the expected semi-variance between the measurements i and j. σ_i^2 is the uncertainty of the measurements, or the measurement error. μ is called the Lagrange parameter. When using the measurement error σ_i^2 in the kriging equation (equation 2.3) the literature speaks about kriging with uncertain data [9].

One of the big advantages of Ordinary Kriging is that it also provides an estimate of the prediction uncertainty of the prediction itself, hence the estimation-error.

2.2.2 Variogram

The spatial variance γ is used in geostatistics to describe the spatial variability of random variables. According to equation 2.4, it's calculated as the halved expected value of the squared difference between two random variables with distance h. High variance suggest that the two variables have very different values, whereas low variance suggest that they have very similar values. x is refereed as the position vector of variable 1 and $x + h$ of variable 2, which is in distance h apart from variable 1.

$$\gamma(h) = \frac{1}{2}E[z(x+h) - z(x)]^2 \tag{2.4}$$

A Variogram is a diagram describing the spatial dependency throughout the region. It shows the correlation of spatial variance γ over the distance between random variables. In figure 2.1 a typical variogram is shown. The nugget is the variance at distance zero and can be interpreted as a measurement error or as Blöschl [7] said the natural variability to small distances smaller than the smallest distance between measurements. The sill represents the level of the variogram, hence the variance of all variables. If the sill is high the measurements vary widely. The distance over which the variables are correlated is called correlation distance or lag. The higher the lag the smoother the spatial prediction. The variable in figure 2.1 is stationary, meaning that the mean is constant over the field. In some circumstances, this is not the case and a trend in data can be seen. For example, for groundwater level, where there is an up- and a downstream direction and the variables vary not randomly but systematically. Blöschl [7] points out that in these cases the variogram would look different. This is because for large distances no horizontal plateau is reached but it continues to rise continuously.

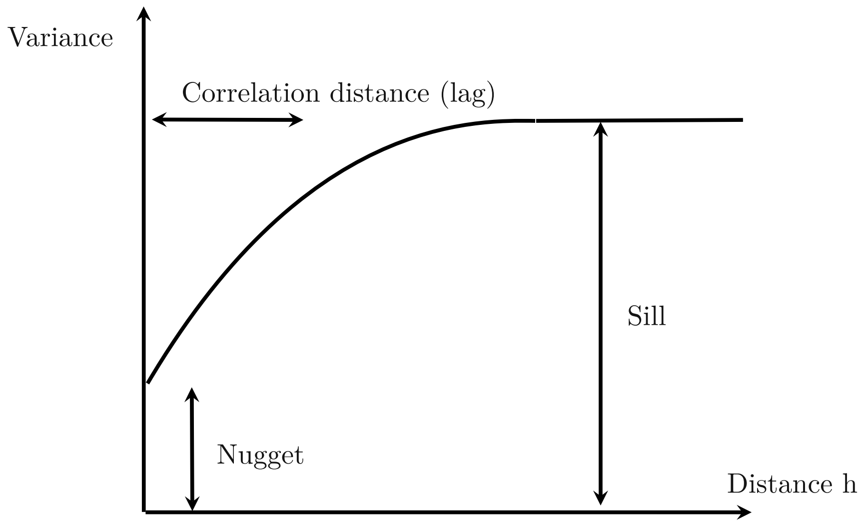


Fig. 2.1: Typical theoretical variogram (adapted from Blöschl [7])

For real applications the variogram is at first unknown and has to be calculated out of the available data. In a fist step the variance of all variable pairs is plotted over the distance, this is called the empirical variogram. Then a compensation curve is fitted to the data, the so called variogram model or theoretical variogram. In this step, the whole properties of the population are estimated out of the statistical properties of the samples, Blöschl [7] highlights. Different models are available like exponential, Gaussian or spherical, and the parameters are chosen to fit the data best.

For the estimation of the variogram it is sufficient to use the half mean squared euclidean distance for each lag incement, also called semivariance. Therefore, the mean is used in equation 2.5 instead of the estimator in equation 2.4:

$$\gamma(h) = \frac{1}{2\ n(h)} \sum_{i=1}^{n} (h)(z(x_i) - z(x_i + h))^2 \tag{2.5}$$

If the semivariance is plotted over the distance, the diagram is called semivariogram. The theoretical semivariogram is then used for the Ordinary Kriging interpolation.

2.2.3 TopKriging

For variables continuous in space like rainfall or snow cover, Ordinary Kriging and similar methods generate good and reliable results. But for variables related to the river network (flow tree) like mean annual runoff (MQ) or stream temperature, these methods fall short as they are conceptually improper as their correlation do not consider the inherent nature of water networks.

In 2006, Skøien et al. [10] introduced a method called topological kriging, short TopKriging, to solve that issue. It splits the stream flow process into two processes. The first process is runoff generation which is continuous in space (e.g. rainfall). The second process is runoff aggregation which subject to routing along the river network, hence the nested tree structure of water networks [11]. The result is that measurements along the same river course are considered more correlated than from other river courses even when they have the same euclidean distance.

Instead of Ordinary Kriging, which uses point values, in TopKriging the runoff generation is assumed to be a spatially continuous process over the whole landscape and the measurements are the aggregation (integral) of a point runoff over a catchment [11]. Hence the spatial variable is defined over a non-zero catchement area A, called support:

$$z(A) = \frac{1}{|A|} \int_A z(\vec{x}) d\vec{x} \tag{2.6}$$

$z(A)$ is the spatial variable, $|A|$ the size of the support and $z(\vec{x})$ the value at location $\vec{x}$. A constant mean is assumed, hence the variables are assumed to be stationary.

The spatial prediction is made by the so called block-kriging predictor given in equation 2.7, where the prediction $\hat{z}(A_0)$ for river location x_0, with catchment area A_0, is a linear aggregation from non-point samples $z(A_1)$, $z(A_2)$, ..., $z(A_n)$.

$$\hat{z}(A_0) = \sum_{i=1}^{n} \lambda_i z(A_i) \tag{2.7}$$

For TopKriging, the kriging system (equation 2.3) stays the same but the semivariances between the measurements must be integrated over the support, Skøien et al. [12] emphasise. For this a point variogram γ_p is assumed, describing the relation between areas and their spatial support. For the regularization of the variogram, which is the calculation of aggregated variograms for the catchments out of the point variogram [11], the semivariance between two observations is given as:

$$
\begin{aligned}
\gamma_{ij} &= 0.5 \times Var(Z(A_i) - Z(A_j)) \\
&= \frac{1}{|A_i||A_j|} \int_{A_i} \int_{A_j} \gamma_p(|\vec{x}_i - \vec{x}_j|) d\vec{x}_i d\vec{x}_j \\
&\quad - 0.5 \times \left[\frac{1}{|A_i|^2} \int_{A_i} \int_{A_i} \gamma_p(|\vec{x}_i - \vec{x}_j|) d\vec{x}_i d\vec{x}_j \right. \\
&\qquad \left. + \frac{1}{|A_j|^2} \int_{A_j} \int_{A_j} \gamma_p(|\vec{x}_i - \vec{x}_j|) d\vec{x}_i d\vec{x}_j \right]
\end{aligned} \tag{2.8}
$$

The position vectors $\vec{x}_i$ and $\vec{x}_j$ are used for integration within the area. The first part gives the semivariance of the variogram whereas the second part subtracts the semivariance within the catchments. Therefore, the semivariance γ_{ij} will be the smallest for close-by areas on the same river, because their support overlap.

To simplify this computationally expensive step, the support is discretised to regular grid points and the integrals are replaced by sums. Another simplification is the use of mean distances between areas instead of calculating the distance between each discretization point, this step is called by Skøien et al. [12] as the use of *Gosh(1951)*-distance.

The semivariance values are inserted into the kriging matrix in equation 2.3 to calculate the kriging weights λ_i.

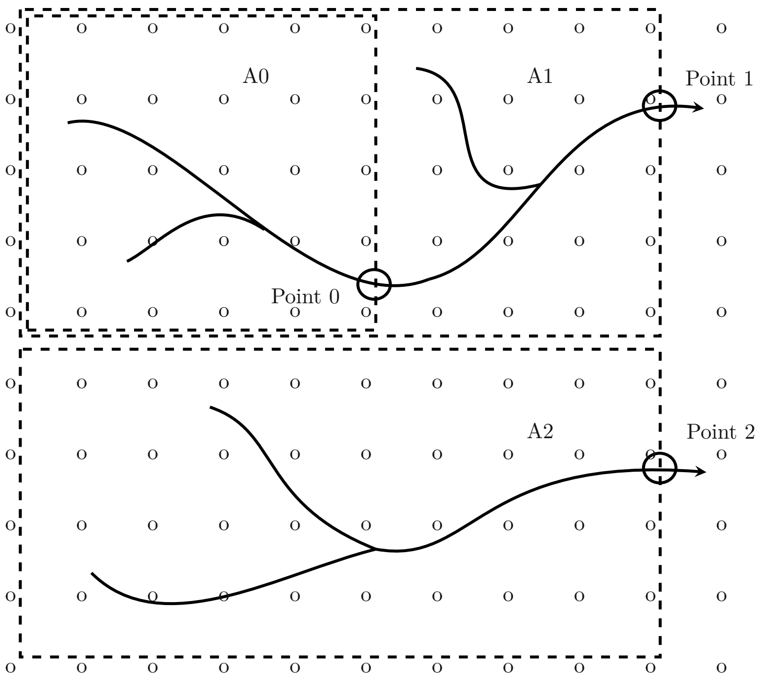


Fig. 2.2: Schematic river network and catchment boundaries with point pairs shown, redrawn from [13]

The schematic figure 2.2 consists of three watersheds and the regular grid points for discretization. To predict the runoff at point 0, Ordinary Kriging would use equal weights for point 1 and 2 as they have equal sized support and similar euclidean (direct) distances to point 0. But top kriging will use a higher weight for point 1 as it belongs to the same river system, hence it shares common discretization points with point 0. If point 1 would be far away (approx 10 times) TopKriging would give more weight to point 2 of the adjacent river system because it is closer to the target point 0 [11]. Hence, Kriging weights depend in TopKriging on both river network topology and distance, or as Laaha et al. [11] described as a more natural way.

In 2014, a package called *rtop* for the statistical software R was introduced by Skøien et al. [12] to make TopKriging interpolation simple, efficient and available to a broad audience.

2.2.4 Application of TopKriging

TopKriging is well suited for prediction of stream flow and stream flow-related variables [11]. Parajka et al. [14] investigated the role of station density and advised not to use TopKriging for station densities below 1.0 stations per 1000 km^2. For a station density above 2.0 the TopKriging predictions are better than one from hydrological model regionalisation. A mean model efficiency above 0.7 (NSE) can be expected for station densities of 2.4. With a station density above 6.0 stations per 1000 km^2 in the STOBIMO project area, proper and precise predictions can be expected. The distribution of kriging weights between observations was intensively discussed by Laaha et al. [11] and they concluded a very realistic distribution of kriging wheights which follows the logical, physical based, understanding of watershed correlation and steam flow behaviour.

For prediction of non-stationary variables like stream temperature, which is highly related to catchment altitude, Laaha et al. [11] suggested to use TopKriging with external drift (TKED), where the deterministic pattern of the variable is modelled with an external drift function (e.g. exponential regression).

TopKriging was also used to predict water quality variables along a main river channel in Mexico [8]. Several interpolation methodes were compared to predict 28 water quality variables. The number of overlapping watersheds was in total 10, which is very low compare to the dataset used in Skøien et al. [10] or Skøien et al. [12]. The prediction efficiency was compared among the interpolation methods with cross-validation. The TopKriging method (TopKriging (TK), TopKriging with external drift (TKED) and Regression TopKriging (RTK)) was overall the best prediction methodology, by being the best predictor for the vast majority (total 79%) of the assessed variables, specially for variables related to wastewater discharge [8]. Within the TopKriging methods TK was the best predictor as it predicted 39% of the variables best, followed by RTK with 36% and TKED with 4%.

Chapter 3

Material & Methods

3.1 Study area

The study area is the same as used in the STOBIMO project [3]. It covers almost the entirety of Austria and additionally the Inn river in Germany and Switzerland and some bordering watersheds of Italy, Liechtenstein and Czech Republic, which results in a total area of 94,596 km^2. The landscape is very diverse and covers a wide range of different landscapes from alpine regions with altitudes over 3500 m above sea level, to continental arid climate in the east.

Figure 3.1 shows the runoff gauging stations in the study area distinguished by diversion affection. Around 17% of them are diversion affected. Apparently, in the east the diversions are due to drinking water extraction, artificial canals and infiltration, and in the west they are due to to hydropower plant diversions. In evidence, main rivers like the Danube or Inn are negligibly affected by diversions in terms of watershed area. Yet, for the headwaters in the mountains, it may have huge effects on the predictions.

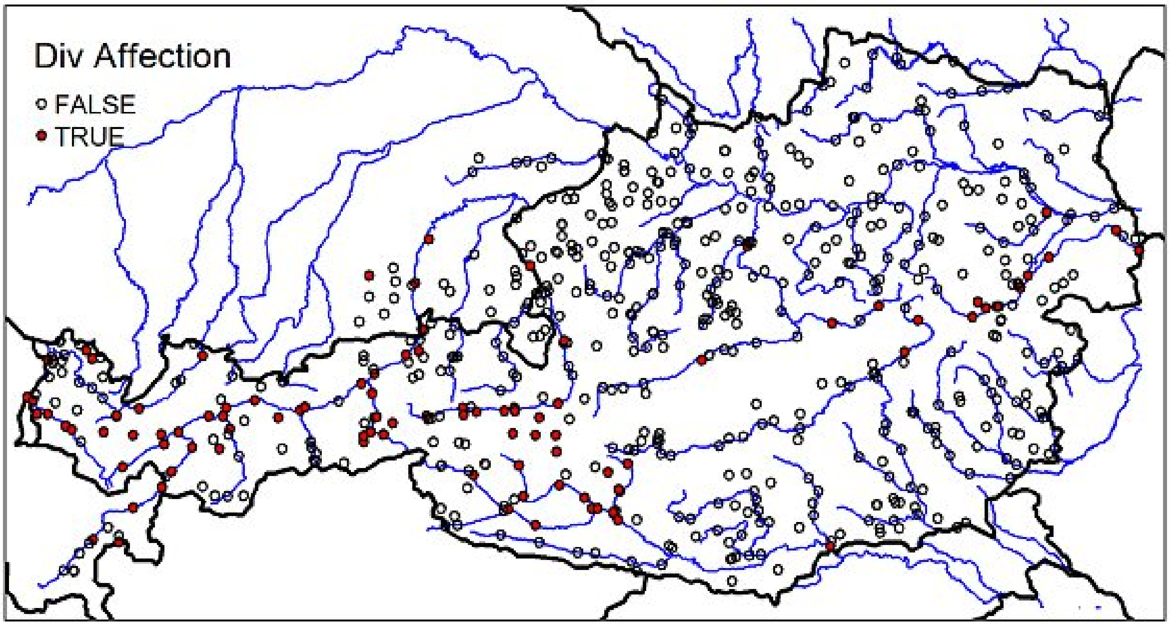


Fig. 3.1: Runoff gauging stations in the study area with main rivers and country borders. Colour referees to diversion affection. (© data.umweltbundesamt.at © EuroGeographics for the administrative boundaries)

3.2 General method

This thesis, after concluding section 2.1, will focus only on diversion areas as correction-measures for diversions, obtained with analysing the diversions to find a simple and practical solution and workflow to consider them in the interpolation. Therefore, the approach of this thesis is to take trans-catchment diversions based on watershed areas in account.

Each diversion has an related watershed area (A_{Div}). This area is either subtracted from or added to the orographic watershed area (A_{oro}) of the downstream river gauge, to calculate the specific runoff, or is subtracted from or added to the MoRE AU to simulate the diversion process along the flowtree (See 3.4.1.1 for additional explanations).

The classic water catchment station is a structural element to dam water. In a routine mode, one part of the water is diverted for hydropower or other purpose, the other part called residual water is left in the river for environmental reasons. In case of an high flow event due to the limited intake of the diversion, the exceeding amount of water is discharged via the spillway into the river as shown in figure 3.2. Due to the simplicity of the approach of this thesis, the following effects were not taken in account:

- Minimum flow requirements: Each water catchment station has to leave legally required quantity of residual water in the river
- Overflow: In case of highflow events only parts of the collected runoff will be diverted. In case the intake water is too dirty or full with sediments, the energy supply companies stop the intake until the highflow event is over.
- Revision, unplanned downtime of the hydropower plants: For about 5 days a year the discharge capacity of the hydropower plant is not reached due to planned or unplanned events.

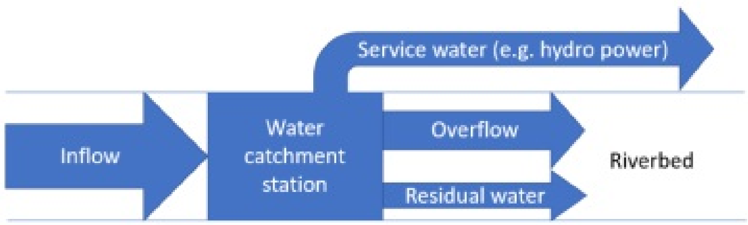


Fig. 3.2: Water catchment station in- & outflow

It's assumed that for annual modelling those effects cancel each other out and the approach with diversion areas will represent the actual diversion quantities. In subsubsection 3.4.4.3, a validation method will be discussed to verify this approach.

Also to make the results comparable and to show which improvements can be made by considering trans-catchment-diversions (diversions), two cases are considered:

- A case without diversion consideration (Div=FALSE), using the orographic watershed area (A_{oro}) for specific runoff calculation, resulting in the natural specific runoff q_{nat} in mm/a.
- A case with diversion consideration (Div=TRUE), using the effective watershed area (A_{eff}) for specific runoff calculation, resulting in the effective specific runoff (q_{eff}) in mm/a.

3.3 Material (Data mining)

For spatial estimation there are raw data from the fields of hydrology and topography needed. They are introduced in the following section.

3.3.1 Existing data

3.3.1.1 Watersheds overlapping

The *HORA Watersheds* is a SpatialPolygon dataset which consists of 7774 watersheds in Austria. The structure is that the watersheds are overlapping, therefore each watershed includes the upstream watersheds areas. The area sizes of the watershed polygons are asa result very diverse and range from 0.33 km^2 of small headwaters to 39,800 km^2 of the downstream Danube watersheds. This nesting structure is necessary for the TopKriging interpolation [10]. It covers the whole surface within the Austrian border and some neighbouring watersheds, where the biggest one is the upper Inn in Switzerland. The lower Inn in Germany is not mapped and has to be implemented in a further step (See 3.4.1.2). Each watershed has an unique ID (*EZGID*) to identify each watershed and to match them with the river network (*HORA edges*). In addition, each watershed has two corresponding areas, *AREASQKM* is the subcatchment size within the study area, hence the area sizes of the watershed polygons and *AREA_KOR* the whole upstream watershed area. For example, the Danube river at location Passau where Danube and Inn river confluence, *AREASQKM* would be the whole Inn river watershed, whereas *AREA_KOR* would be the Inn and Danube river watershed area.

The total watershed area (*AREA_KOR*) of a watershed at March river (*EZGID = 7688*) had to be corrected because it was identical with the watershed area within the study area (*AREASQKM*). Which could not be the case because a huge part of March river is located outside the study area. Also, watersheds which are nearly identical to their neighbouring watersheds (less than 0.02 km^2 area difference), caused problems in the TopKriging interpolation and were therefore removed. Furthermore, watersheds of the Rhine river itself were removed, because they are beyond the study area (STOBIMO) and are not representative to the related upstream watersheds inside the study area, hence they share nearly the same areas in the model but represent different watershed areas in order of magnitudes. All those changes can be seen in the attached scripts in appendix C.2.2 (Line-No.: 156 - 162, 283 - 287).

3.3.1.2 River network

The *HORA edges* is a SpatialLine dataset which consists of the 5775 river-sections in Austria and parts of neighbouring countries like partly the Inn river region in Switzerland. The length varies from less than 1 km to 28 km. With the columns *EZGA* and *EZGE* the corresponding up- and downstream watershed from the *HORA Watershed* dataset can be identified. With columns *HZBNRA* and *HZBNRE* the corresponding runoff gauging stations could be additionally identified.

Minor changes had to be made as there were mistakes in the dataset, like missing or wrong runoff gauging stations or wrong upstream AU IDs. Those can be seen in the attached scripts in appendix C.2.2 (Line-No.: 60 - 84).

3.3.1.3 STOBIMO watersheds

The *STOBIMO watersheds* is a SpatialPolygon dataset which consists of 894 watersheds in Austria. In contrast to the *HORA Watershed* (see 3.3.1.1) the *STOBIMO watersheds* are not nested. The area size is therefore less diverse and ranges from 9 km^2 to 1946 km^2. It covers the whole surface within the Austrian border and some neighbouring watersheds like the upper and lower Inn in Switzerland and Germany, which is the whole study area in this diploma thesis. The watersheds of this dataset are also the MoRE analytical units (AUs) of the STOBIMO project [3] and therefore the input data to the MoRE Model (see 1.1) and the target dataset in this diploma thesis. Throughout this thesis, "STOBIMO AU" refers to the MoRE AU in the STOBIMO project area.

Minor changes had to be made to the dataset, because some AUs had missing runoff gauging stations, which were added. Others refereed to runoff gauging stations which did not represent the whole AU, causing a non-compliance with the validation method and were therefore removed. Also, the precluded gauges (See 3.3.1.4) were removed from the MoRE AU in the *STOBIMO watersheds* spatial dataset. This can be seen in the attached scripts in appendix C.2.2 (Line-No.: 44 - 55).

3.3.1.4 Runoff gauging stations

Gauges in Austria

The *eHyd Pegel 2011* is SpatialPoint dataset containing 771 river gauges in Austria. 587 are capable to measure the discharge and for all of them are discharge data available. They can also be found at the Internet portal for hydrographic data of Austria (eHyd) provided by the BMLRT [15].

Gauges in Switzerland

The *Pegel_CH* is a SpatialPoint dataset containing 15 river gauges in Switzerland along the Inn river. For 12 gauges discharge data are available [16]. Missing watershed area was added to some gauge stations in the dataset, those can be seen in the attached scripts in appendix C.2.2 (Line-No.: 144 - 148).

Gauges in Germany

The *Pegel_BY* is a SpatialPoint dataset containing 88 river gauges in Bavaria along the lower Inn basin. For 36 gauges discharge data are available [17].

Discharge data for gauges

The *Jahresabfluesse_alle_Pegel_IWAG* is a dataset containing annual mean discharge (MQ in m^3/s) of 645 gauges in Austria [15], Switzerland [16] and Bavaria [17]. The records range from 1 to 119 years, however the median is 44 years. Only data from 2009 to 2017 were used in the analysis as this is the analysis timeline in this thesis. Some runoff gauging stations do not have measurements for each day. Those with less than 356 observation days were excluded due to the increased chance an extreme event could be missed and the calculated MQ does not represent the characteristic of this year. The number of gauges available per year consequently range from 580 to 620 stations and can be seen in table 3.1. This leads to a station density of 6 stations per 1000 km^2, which are very high and should give a high prediction efficiency [14]. Therefore it is better to exclude a few gauges with less representative observations, because the station density remains regardless high. The MQ is the arithmetic mean of all daily runoff values in a considered time span (discharge year) [17].

Tab. 3.1: Number of runoff gauging stations (gauge) with useable data per year.

year	2009	2010	2011	2012	2013	2014	2015	2016	2017
number of gauges	618	617	620	613	612	596	588	585	580

Preclusion of gauges

The following runoff gauging stations are excluded due to their undefinable watershed area which is affected by known diversions with unknown diversion areas:

- Gauge Singerin (Steg) [208694]
- Gauge Singerin (Höllental) [208702]
- Gauge Kraiburg [18004007]
- Gauge Mühldorf [18004506]

The following gauges are excluded because they measure only parts of the stream and therefore aren't representative:

- Gauge Wiener Neustadt (Flußbauhof) [208975]
- Gauge Katzelsdorf (EVN) [208967]
- Gauge Rosenheim Q / Hammerbach [18312010]
- Gauge Trostberg / Alzkanal [18409009]
- Gauge Gufflham / Alzkanal [18409508]
- Gauge Deutsch-Jahrndorf (Neurießäcker) [210435]

3.3.1.5 Identification of diversions

As part of the *Nationaler Gewässerbewirtschaftungsplan 2015* [18] an online map of the Austrian river network is provided [19]. This map provides after a certain zoom level the artificial diversion pathways which were used to identify and allocate diversions. Also, a list with the diversion-affected gauges in Austria (See 3.3.2.1) was provided as appendix of the *Hydrographisches Jahrbuch von Österreich 2017: Hydrographischer Dienst in Österreich* [15]. Those datasets are not complete, hence not all diversion are mapped or listed. Therefore additional research had been carried out. Further information from the websites of energy supply companies and mapped diversions from OpenStreetMap (OMS) were used to identify diversions. Wikipedia articles are also very informative to get information about hydropower plants.

3.3.2 Collected data

Additional data had to be collected to reach an adequate data situation for further processing.

3.3.2.1 Diversion areas

Austria & Liechtenstein

The BMLRT provided with the *Hydrographisches Jahrbuch von Österreich 2017: Hydrographischer Dienst in Österreich* [15] a list with the diversion-affected gauges in Austria, with the inlet and outlet diversion watershed area in km^2.

Switzerland & Italy

The Engadiner Kraftwerke AG (EKW) provided watershed areas for their own water catchments stations and for those from A2A S.p.A. (Italian public utility company) (A2A) in Italy. To get the diversion area, those areas where adjusted eighter by the gauge data provided by Swiss Federal Office for the Environment (FOEN) or by data of the *Restwasserkarte Schweiz* [20].

Germany

No diversion area could be collected for Inn region in Bavaria. Therefore the diversion watershed area (A_{Div}) was measured out of Geographic information system (GIS) maps.

3.3.2.2 Energy supply companies

The research on relevant diversions was a significant part in this thesis. Getting an overview of all hydropower plants and understanding their complex structures demanded considerable effort, especially in contacting the companies and asking for diversion data.

To validate the model results, the annual discharge per diversions was needed. Therefore, all relevant companies and authorities had to be asked to provide data of their trans-catchment diversions. Fortunately, most (except 2) companies were keen to support scientific research and provided data as far as available. Table 3.2 shows all relevant hydropower plants which were considered in this thesis.

In some cases, no data were collected or only partially available and couldn't be used. In most cases, data were available and provided as tables in text, *.csv* or in Excel *.xlsx* format. Data were either monthly or annual total sums in m^3 or available as mean discharge in m^3/s. In a further step the data were cumulated to represent each AU and transformed into mean discharge in m^3/s for further processing. Due to company regulations, this data is classified as confidential and therefore not published in this master thesis. Consequently the data is only used for validation purposes.

The following company names are abbreviated: Österreichische Bundesbahnen (ÖBB), Tiroler Wasserkraft AG (TIWAG), illwerke vkw AG (illwerke vkw), Elektrizitätswerke Reutte AG (EWR), Stadtwerke München GmbH (SWM), Kärntner Elektrizitäts-Aktiengesellschaft (KELAG), Liechtensteinische Kraftwerke (LKW), Engadiner Kraftwerke AG (EKW) & A2A S.p.A. (Italian public utility company) (A2A).

3.3.2.3 Special diversion cases

Wiener Wasser

The City of Vienna - MA 31 - Wiener Wasser (Wiener Wasser) operates as public water supply company two major pipelines with several water catchment stations (springs) in Lower Austria and Styria. Since the water transferred is drinking quality, they are not relevant for trace substance transport and therefore for the MoRE model. Nevertheless, because of the large quantities withdrawn, the diversions are relevant for the runoff gauging stations. Therefore, diversion data were collected and provided by Wiener Wasser. For diversion relevant are the *First Vienna Spring Water Main (I. Wiener Hochquellenleitung)* and *Second Vienna Spring Water Main (II. Wiener Hochquellenleitung)*.

Marchfeld Kanal

The *Betriebsgesellschaft Marchfeldkanal* operates an irrigation channel in Lower Austria and Vienna. It transports water from the Danube river to the Marchfeld plane. Diversion data were provided, but not used as they are not included in the MoRE model of the STOBIMO project.

Special case: Infiltrations of Leitha river

Schwarza river and Leitha river in the border region of Lower Austria and Burgenland are subject to infiltrations and several artificial channels. In 2009, a report was published to this issue [21]. The infiltrations of Leitha river are calculated with the formulas given in chaper 4.1.3.2.2 of the report LEITHA - Referenzzustand und Zielzustand WRRL [21]. Because eHyd [15] provided data for *Wiener Neustädter* canal and *Katzelsdorf* canal, their available runoff data were used instead of the given values in the paper. The infiltrations of Leitha river are included as diversions, despite that they are of natural origin and that the research question addresses artificial diversions in particular. This is due to the fact that the diverted amounts are not negligible.

Tab. 3.2: Hydropower plants relevant for trans catchment diversions by energy supply company.

company	power plant group	hydropower plant	data provided
illwerke vkw	Obere Ill/Lünersee	Kopswerk I & II	TRUE
illwerke vkw	Obere Ill/Lünersee	Obervermuntwerk I & II	TRUE
illwerke vkw	Obere Ill/Lünersee	Rodundwerk I & II	TRUE
illwerke vkw	Obere Ill/Lünersee	Lünerseewerk	TRUE
illwerke vkw	Obere Ill/Lünersee	Latschauwerk	TRUE
illwerke vkw	Obere Ill/Lünersee	Vermuntwerk	TRUE
illwerke vkw	Obere Ill/Lünersee	Rellswerk	TRUE
illwerke vkw	-	Walgauwerk	TRUE
illwerke vkw	-	KW Langenegg	TRUE
illwerke vkw	-	KW Klösterle	TRUE
ÖBB	Stubachtal	KW Tauernmoos	TRUE
ÖBB	-	KW Spullersee	TRUE
TIWAG	-	KW Achensee	TRUE
TIWAG	-	KW Imst	TRUE
TIWAG	-	KW Kaunertal	TRUE
TIWAG	Sellrain-Silz	KW Kühtai	TRUE
TIWAG	Sellrain-Silz	KW Silz	TRUE
EWR	-	KW Heiterwang	FALSE
KELAG	-	KW Koralpe	TRUE
KELAG	Fragant	KW Fragant	FALSE
Salzburg AG	Remsach, Bockstein, Nassfeld	KW Naßfeld	TRUE
VERBUND	-	KW Schwarzach	TRUE
VERBUND	-	KW Sölk	TRUE
VERBUND	Reißeck-Kreuzeck	KW Galgenbichl	TRUE
VERBUND	Reißeck-Kreuzeck	KW Reißeck II	TRUE
VERBUND	Reißeck-Kreuzeck	KW Rottau	TRUE
VERBUND	Reißeck-Kreuzeck	KW Kolbnitz	TRUE
VERBUND	Kaprun	KW Kaprun	TRUE
VERBUND	Zillertal	KW Funsingau	TRUE
VERBUND	Zillertal	KW Häusling	TRUE
VERBUND	Zillertal	KW Roßhag	TRUE
VERBUND	Zillertal	KW Mayrhofen	TRUE
LKW	-	KW Samina	TRUE
SWM	-	Leitzachwerke	TRUE
EKW	-	KW Ova Spin	TRUE
EKW	-	KW Pradelle	TRUE
EKW	-	KW Martina	TRUE
A2A	-	Plant at Premadio	FALSE

3.4 Methods (Data munging)

To obtain results. the raw data had to be processed. This is discussed in this section which is divided into subsections pre-processing, processing and post-processing. Figure 3.3 shows a simplified representation of the processing steps.

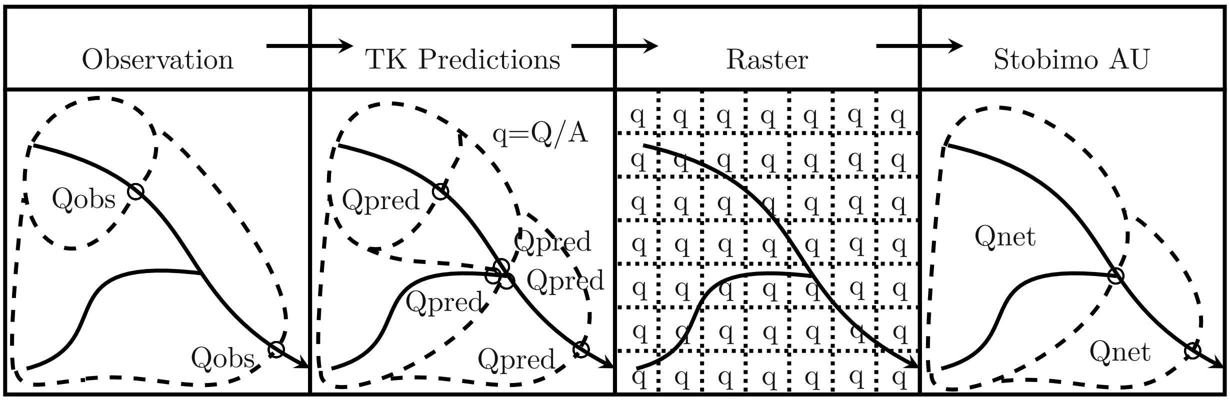


Fig. 3.3: Simplified diagram of process steps from observed runoff data to TK interpolated values, which are converted to specific runoff for rasterization to the transfer to STOBIMO AUs and the final conversion to net specific runoff per STOBIMO AU.

3.4.1 Pre-processing

Before the TopKriging interpolation [12] in 3.4.2, collected data had to be pre-processed. The different datasets with their corresponding tables, hence informations had to be merged to consistent datasets to meet the input requirements of the TopKriging interpolation.

3.4.1.1 Gauge area correction

The collected data from 3.3.2.1 were used to correct the watershed area of the disturbed gauge station. The inlet ($A_{Div.Inlet}$) and outlet diversion areas ($A_{Div.Outlet}$) were added/subtracted from the given orographic watershed area A_{oro} (equation 3.2). A definition of the used watershed areas can be seen in figure 3.4.

Some gauges (e.g. gauge station *Lorüns-Äule*) resulted in implausible effective watershed A_{eff} area and were corrected manually by diversion data or GIS analysis. Strictly speaking lowland rivers (e.g. Danube river) are also disturbed by diversions in their headwaters. But due to this minimal influence their stream gauge station were assumed to be undisturbed. The limit was set to 0.05%, hence a ratio $A_{Div}/A_{oro} < 0.0005$.

$$A_{eff} = A_{oro} + (A_{Div.Inlet} - A_{Div.Outlet}) \tag{3.1}$$

A_{eff} ... effective watershed area of the gauge in km^2
A_{oro} ... orographic watershed area of the gauge in km^2
$A_{Div.Inlet}$ ... diversion watershed area added to the orographic watershed area in km^2
$A_{Div.Outlet}$ ... diversion watershed area subtracted from the orographic watershed area in km^2

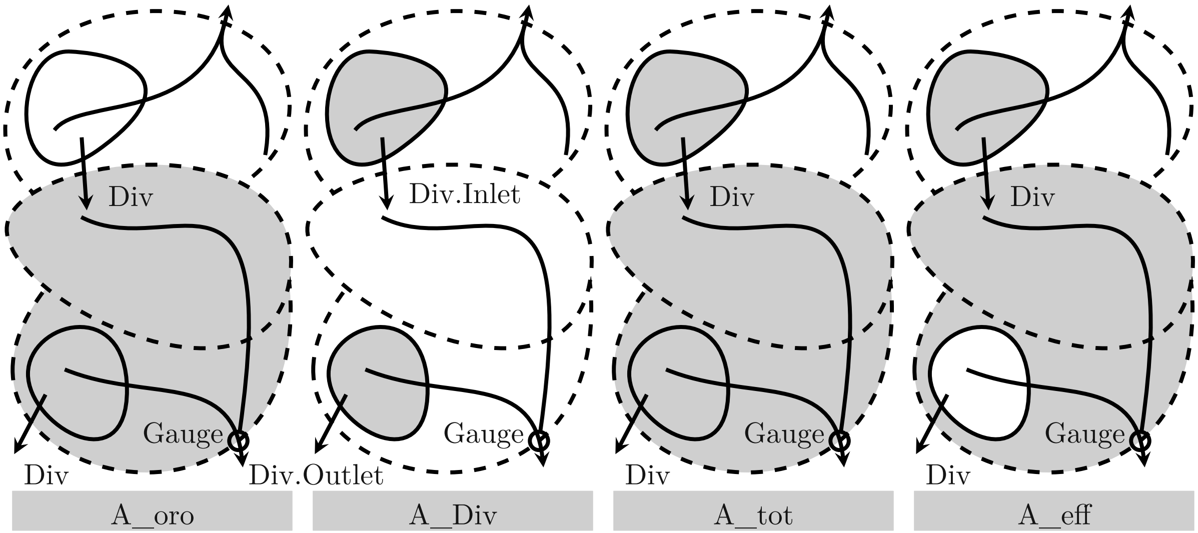


Fig. 3.4: Definitions of watershed areas used in this thesis. Referred to the runoff gauging station (gauge) are: orographic (A_{oro}), effective (A_{eff}), net diversion (A_{Div}), added diversion ($A_{Div.Inlet}$), subtracted diversion ($A_{Div.Outlet}$) watershed area of the runoff gauging station (gauge).

Special case: Artificial channels in the flatlands

Artificial channels and river infiltrations in the flatlands are a special case. This is because instead of a diversion watershed area a diversion amount in m^3/s is provided by eHyd [15] or calculated from literature (See 3.3.2.3).

In order to be able to apply the chosen approach in this thesis here as well, the diversion watershed areas must be derived from the diversion amounts (Q_{Div}). Therefore, the diversion watershed area (A_{Div}) will be calculated with equation 3.2 for each year and further the mean of it will be then used to define the diversion watershed area for the gauge in the TopKriging interpolation and the AU in the MoRE-Model.

$$A_{Div} = \frac{1}{n} \cdot \sum_{i=1}^{n} \frac{Q_{Div.i}}{Q_{gauge.i} + Q_{Div.i}} \cdot A_{oro} \tag{3.2}$$

A_{Div} ... diversion watershed area in km^2
A_{oro} ... orographic watershed area of the gauge in km^2
Q_{Div} ... sum of all diversions and infiltrations in m^3/s
Q_{gauge} ... runoff measured at the gauge in m^3/s
i ... year of observation (2009-17)
n ... total number of years with observations ($n_{max} = 8$)

3.4.1.2 Preparation of spatial data

The available spatial datasets had to be prepared to meet the interpolation input requirements. The following paragraphs explain of what the individual input datasets consist.

Prediction Locations

The dataset *PredictionLocations* is an input (as *SpatialPolygonDataFrame*) for the rtop-package [12]. It has to consist of overlapping watershed areas along the river flowtree, hence the hierarchical structure of rivers from source to mouth, where predictions are made.

For the Austrian Danube watershed, the *HORA Watershed* dataset was combined with the *HORA edges* dataset to find the corresponding runoff gauging stations and upstream watersheds for each watershed. All watersheds without matches were filtered out. This was required mainly at the upper Inn region. The lower Inn river watershed was not covered by this dataset.

For the upper (Switzerland) and lower Inn river (Bavaria) watershed, overlapping watersheds had to be created manually with a GIS programm out of the *STOBIMO watersheds* dataset. That leads to a more coarse structure and therefore less precise interpolation in this regions. The upstream watershed and stream gauge station for each watershed were also taken from the *STOBIMO watersheds* dataset.

To create the *PredictionLocations* dataset for the whole study area, the overlapping watersheds of Austrian Danube watershed were merged with the upper and lower Inn region watersheds.

Each watershed has following informations:

- *EZGE* ... watershed ID
- *EZGA* ... upstream watershed ID
- *EZGTO* ... downstream watershed ID
- *EZGE_AREA* ... watershed area in km^2
- *EZGA_AREA* ... upstream watershed area in km^2
- *ID_GAUGE* ... stream gauge station at the outlet of the watershed

Runoff gauging stations

For the dataset *rnet_gauges* the datasets from 3.3.1.4 were merged to a single dataset (as *SpatialPointDataFrame*) containing all available runoff gauging stations. Dataset *"eHyd Pegel 2011"* contains the runoff gauging station informations and in particularly the corresponding watershed of the *PredictionLocations* dataset.

Each runoff gauging station has following information:

- *EZGE* ... watershed ID
- *EZGA* ... upstream watershed ID
- *EZGTO* ... downstream watershed ID
- *EZGE_AREA* ... watershed area in km^2
- *EZGA_AREA* ... upstream watershed area in km^2
- *ID_GAUGE* ... stream gauge at the outlet of the watershed

Observations

The dataset *Observations* is an input (as *SpatialPolygonDataFrame*) for the rtop-package in data processing (See 3.4.2) and is a subset of the *PredictionLocations* dataset. It only consists of watershed areas with a runoff gauge station with available runoff data (MQ) (See 3.4.1.3). Some runoff gauging stations were excluded due to aforementioned reasons and the numbers of stations used for the input can be taken from table 3.3.

Each watershed area has following informations:

- *EZGE* ... ID of the watershed itself
- *EZGE_AREA* ... watershed area of the watershed itself in km^2
- A_{eff} ... effective watershed area of the gauge in km^2
- A_{oro} ... orographic watershed area of the gauge in km^2
- Q_{gauge} ... runoff measured at the gauge in m^3/s

Tab. 3.3: Number of runoff gauging stations (gauges) used per year as input (observations) for interpolation.

year	2009	2010	2011	2012	2013	2014	2015	2016	2017
number of gauges	575	574	576	572	571	557	550	547	543

STOBIMO_EZG

The dataset *STOBIMO_EZG* is a spatial data table and the target dataset of this diploma thesis. The predictions of the TopKriging Interpolation have to be transferred to the analytical units (AUs) of the *STOBIMO_EZG*.

Each analytical unit (AU) includes following informations:

- *ID_MORE* ... ID of the analytical unit (AU) itself
- *TO_ID_MORE* ... ID of the downstream AU (natural flow)
- *TO_ID_2_MORE* ... ID of the diversion AU (diversion flow)
- *HZB_PEGEL1* ... ID of the downstream AU (natural flow)
- *AREAKM2* ... watershed area of the AU (only watershed inside the study area) in km^2
- *AREAKM2_korr* ... watershed area of the AU (whole upstream watersheds) in km^2

Diversions between STOBIMO analytical unit (AU)

The collected data from 3.3.2.1 were used to set the diversions and diversion areas between the individual MoRE AUs. During the analysis it went hand in hand with the gauge corrections (See 3.4.1.1). Apart from data collection, this step was the most complex and costly, as it required the highest allotment of time.

In the MoRE-Model the river flowtree is simulated and each analytical unit (AU) can has two downstream AU, one natural AU (*TO_ID_MORE*) and one diversion AU (*TO_ID_2_MORE*). So each diversion had to be assigned to a giving AU. The receiving AU ID was recorded in the diversion column *TO_2_ID_MORE* of the giving AU.
During this process a few problems occurred:

- If an AU has two assigned diversions, which is not possible inside the MoRE model and the reconstructed flowtree calculation, the most important diversion was selected and the less important had to be assigned to the next possible downstream AU.

- Diversions which direct to the same downstream AU as their natural flow, hence *TO_2_ID_MORE* is identical to *TO_ID_MORE*, can not be calculated by the MoRE model. Despite this, the reconstructed flowtree calculation was able to calculated them in this thesis. Therefore these diversions were retained and only removed when exporting to the MoRE model.

- If a diversion flows into the upstream AU, then there is a hydraulic short circuit. An example can be seen in figure 3.5. This cannot be represented in the MoRE-Model nor the reconstructed flowtree calculation. Since some of these diversions affect runoff gauging stations, they are special considered in the post-processing process (See 3.4.3.3).

3.4.1.3 MQ data

First, the gauges from 3.3.1.4 were precluded from the runoff table *Jahresabfluesse alle Pegel IWAG* (3.3.1.4). Then, they are merged with the diversion table from 3.4.1.1 to add the orographic

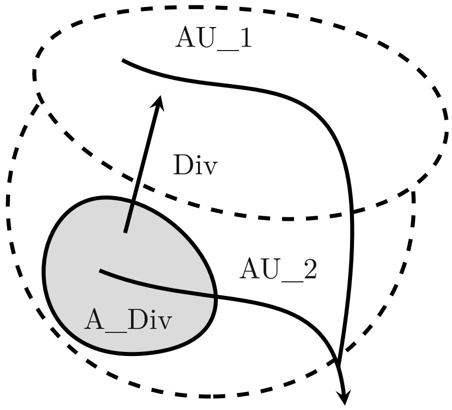


Fig. 3.5: Example of a hydraulic short circuit. A part of downstream analytical unit (AU) (AU_2) is diverted (A_Div) to the upstream AU (AU_1) causing a hydraulic short circuit.

watershed area (A_{oro}) and effective watershed area (A_{eff}) for each disturbed gauge station. Afterwards, the orographic watershed area for the undisturbed gauge station was added from the *rnet_gauges* dataset. The effective watershed area was set equal to the orographic watershed area for the undisturbed gauge station as they are unaffected by diversions.

For further processing (See 3.4.2) the rtop-package [12] requires as input the average runoff per unit area in e.g $m^3/(s\,km^3)$, namely specific runoff. Therefore, this is calculated with the orographic watershed area for undisturbed (natural) case as reference situation and with the effective watershed area for disturbed (with diversion consideration) case (equation 3.3) and added to the table. For better comparability the unit was changed to mm/a, see equation 3.4. To identify leap years (days per year) for the calculation the *R* [22] package*hydro TSM* [23] was used.

The table contains the following information:

- ID ... stream gauge ID
- $YEAR$ ... year of runoff data
- MQ ... mean annual runoff (MQ) measured at the gauge in m^3/s
- A_{oro} ... orographic watershed area of the gauge in km^2
- A_{eff} ... effective watershed area of the gauge in km^2
- q_{nat} ... natural (undisturbed) specific runoff in $m^3/(s\,km^3)$
- q_{eff} ... effective (disturbed) specific runoff in $m^3/(s\,km^3)$
- $q_{nat_{mm}}$ ... natural (undisturbed) specific runoff in mm/a
- $q_{eff_{mm}}$ ... effective (disturbed) specific runoff in mm/a

$$q_{nat} = \frac{MQ\_m3\_s}{A_{oro}} \quad \& \quad q_{eff} = \frac{MQ\_m3\_s}{A_{eff}} \tag{3.3}$$

$$\begin{aligned} q_{nat_{mm}} &= (\tfrac{3600}{1000} \cdot 24 \cdot 365^{*}) \cdot q_{nat} \\ q_{eff_{mm}} &= (\tfrac{3600}{1000} \cdot 24 \cdot 365^{*}) \cdot q_{eff} \\ &\qquad\qquad {}^{*}366 \; for\ leap\ years \end{aligned} \tag{3.4}$$

Looking at the *MQ table* of the *runoff gauging stations* the observed specific runoff over all analysis years (2009-17) can vary between 10 mm/a in the north eastern flatlands and 8490 mm/a in small alpine regions as shown in table 3.4.

Tab. 3.4: Statistics of all observed specific runoff from runoff gauging stations in the project area between 2009 - 2017

variable	Min.	1% Qu.	1^{st} Qu.	Median	Mean	3^{rd} Qu.	99% Qu.	Max.
q_{nat_mm}	10	50	342	674	811	1130	2784	8490
q_{eff_mm}	19	52	360	726	831	1170	2555	8490

3.4.2 Processing

3.4.2.1 Input

The spatial data from 3.4.1.2 are the input for the rtop-package [12]. As the package requires the specific runoff (runoff per unit area) [12] they are added to the *Observations* from the runoff data (*MQ data*) in 3.4.1.3. Two variants are calculated to compare the effects of diversion area consideration. For without diversion consideration (Div = FALSE) the natural specific runoffs (q_{nat}) in mm/a are used. For with diversion consideration (Div = TRUE) the effective specific runoffs (q_{eff}) in mm/a are used. As example the observed effective specific runoff (q_{eff}) of 2009 is shown in figure 3.6.

- Observations ... dataset of locations with the observed values (runoff measurements)
- PredictionLocations ... dataset of locations with absent runoff measurements (including the locations with observed values)
- Parameters ... list with changed default parameters:
  gDist = TRUE ... use Ghosh-distance to reduce computation time
  cloud = FALSE ... binned variograms instead of a variogram cloud
  rresol = 25 ... min. number of discretization points in each area (as suggested in [12])
  singularSolve = TRUE ... because kriging matrices are singular (when two or more areas being (almost) identical)

3.4.2.2 TopKriging interpolation

The rtop-package [12] in the *R* environment [22] makes the TopKriging interpolation simple and efficient. First a *rtop-object* is created by calling *createRtopObject* and adding the *prediction-Location* spatial dataset as well the *Observations* spatial dataset with the observed values. All results will be stored subsequently in this *rtop-object*.

Then *rtopFitVariogram* will create a sample variogram and will fit a variogram model to it, some diagnostic plots produced by call *checkVario* can be seen in figure 3.7 on page 34. Skøien et al. [12] stated that, "the first two explore the data before variogram fitting and interpolation, whereas the last two show the correspondence between the sample variogram and the fitted variogram." [12, p. 185].

The actual interpolation is done by calling *rtopKrige*, this solves the kriging system based on the regularized semivariances seen in figure 3.7d on page 34.

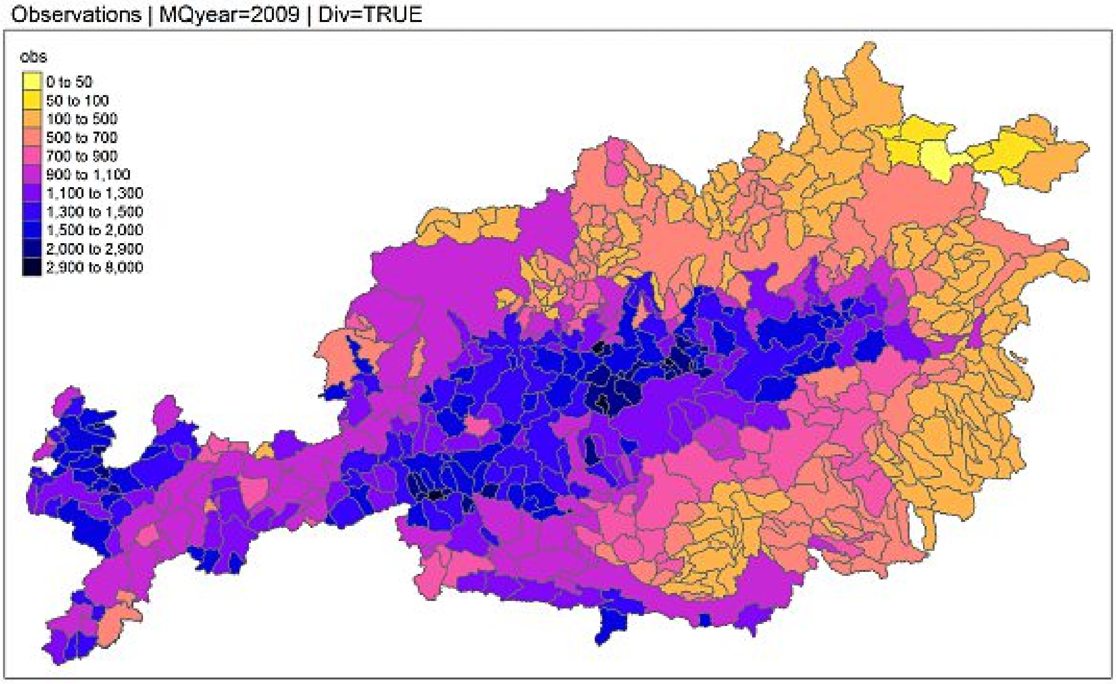


Fig. 3.6: Overlapping watersheds with observed natural (undisturbed) specific runoff in mm from year 2009 with diversion consideration (MQyear=2009, Div=TRUE).

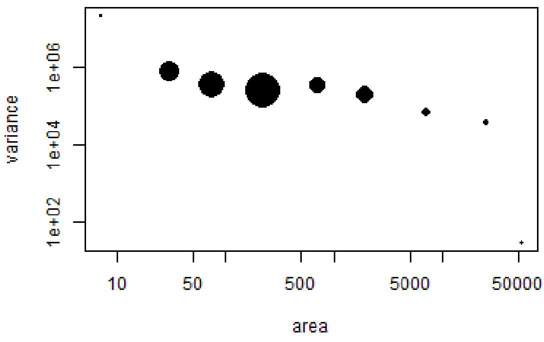
3.4.2.3 Diagnostic plots

Figure 3.7a on the following page shows that the variance is negative correlated to the catchment size. This is a condition for using TopKriging as Skøien et al. [12] state, a decreasing dispersion variance with increasing area is on of the assumptions of top-kriging [12, p. 185].

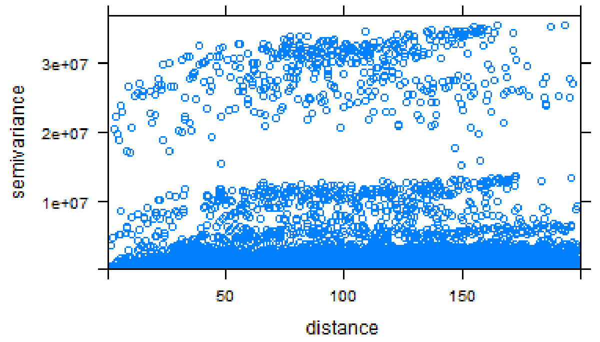
The semi-variance over distance of catchment pairs can be seen in figure 3.7b on the next page. High semi-variance values result when the catchment pair has very different values and low semi-variance values indicate similar values. Increasing semi-variance with distance and a spatial autocorrelation up to a range of 40 km lag distance can be seen. The two accumulations at semi-variance value of 1×10^7 and 3×10^7 are caused by pairs with catchment 1034 (Area=18.4 km^2) and 1045 (Area=9.4 km^2). Those catchments have very high specific MQ runoff and belong to the runoff gauging stations *Roßleithen* (HZBNR: 206482) and *Polsterluck* (HZBNR: 205773) respectively. As there is no other information that give reasons to exclude those watersheds, they are continued to be used for further analysis.

In figure 3.7c on the following page the observed and regularized semi-variance values (gamma) are compared. The circle size is relative to the number of observations per bin and the diagonal 1:1 line represents a perfect fit. Apparently, most of the big bins are centered around the 1:1 line which indicates a good fit of the model. The accumulation on the right is caused by the same outliers as discussed in the previous figure.

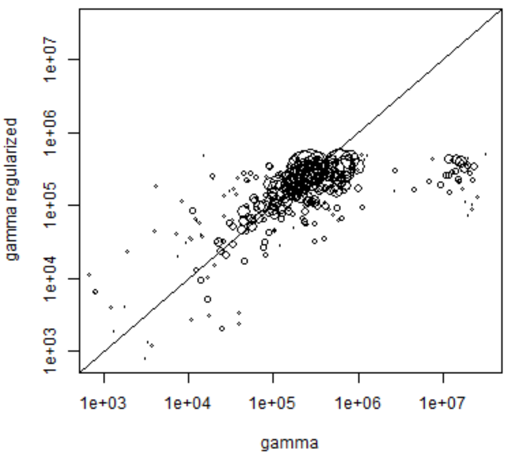
Figure 3.7d on the next page compares the sample variogram and the regularized gamma (semi-variance) for some selected bins. For example, the *"300 vs. 30"* line shows the regularized semi-variogram between a catchment of size 300 and 30 km^2. This approach is based on quadratic catchments, while natural catchments have a more rectangular shape. The regularization can not fully reproduce the reduction of the variance as function of area, especially for middle distances. The black solid line represents the point variogram used for the calculation of the semi-variogram.



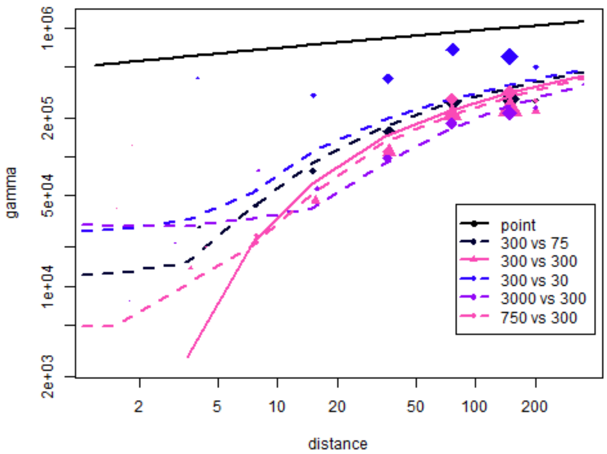
(a) Ratio between area size and spatial variance of observations cumulated in bins. Circle size is relative to the number of observations per bin.



(b) Observations sample Variogram cloud. Circles represent catchment pairs.



(c) Semivariogram values ratio between regularization and the observation. Circle size is relative to the number of observations per bin.



(d) Regularized semivariograms and sample variogram by distance and area. Dotted lines show regularized semivariograms for combinations of catchment sizes whereas solid lines equal sized catchments represent. Circle size is relative to the number of observations per bin.

Fig. 3.7: TopKriging diagnostic plots (Div=TRUE, MQyear=2009) (adapted from Skøien et al. [12]).

3.4.2.4 Output

The kriging results of the interpolation appear as 3 added column to the *predictionLocation* spatial dataset:

- var1.pred ... predicted specific runoff for each watershed in mm/a
- var1.var ... variance (prediction error) of predicted specific runoff for each watershed in mm^2/a^2
- sumWeights ... sum of the Kriging weights used

As an example, the *predictionLocation* spatial dataset concerning the year 2009 with diversion consideration can be seen in figure 3.8. The two figures show the predicted specific runoff for each watershed (*var1.pred*) in mm/a, hence the interpolation result and the prediction error of the predicted specific runoff for each watershed in mm^2/a^2 respectively.

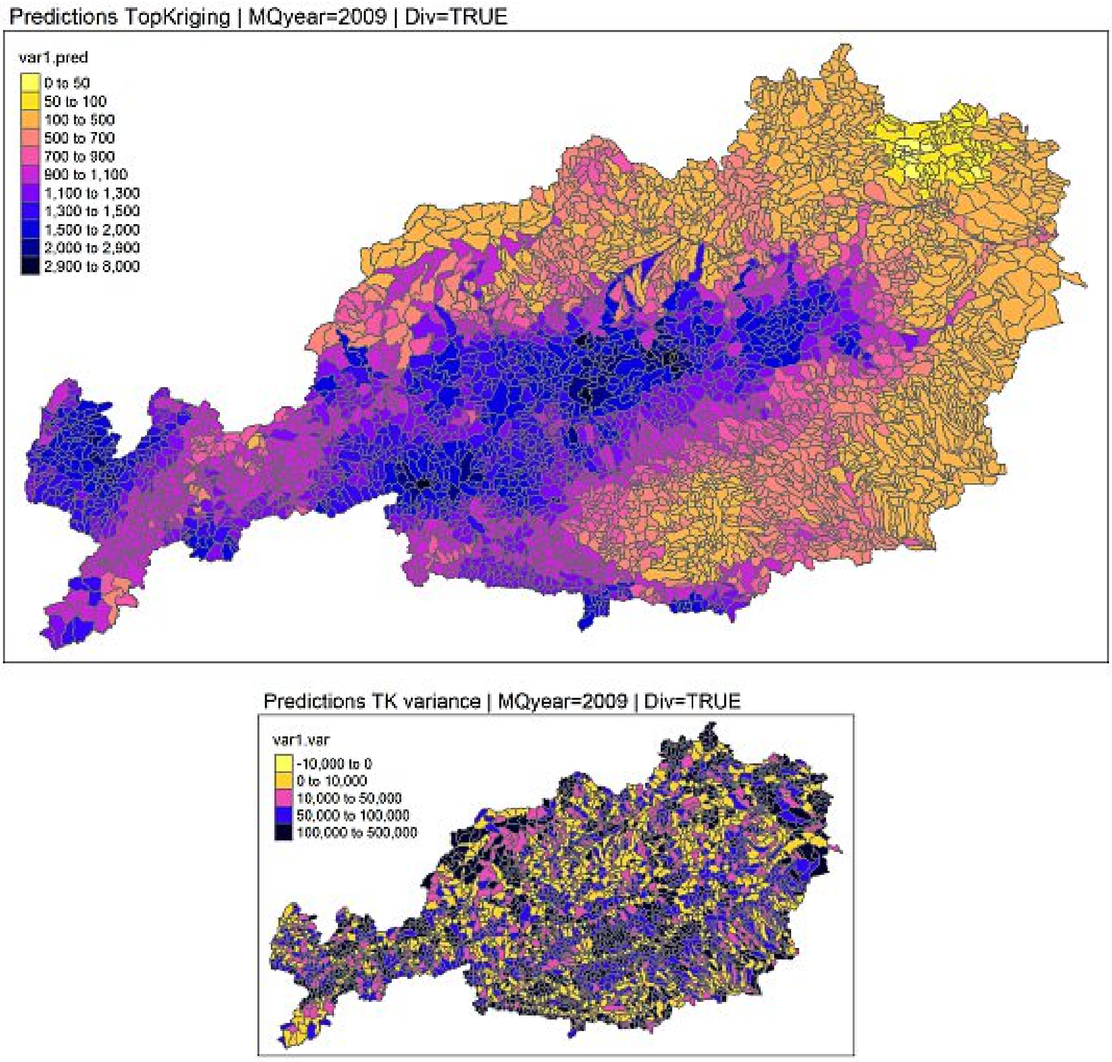


Fig. 3.8: Overlapping watersheds with top-kriging predicted natural (undisturbed) specific runoff in mm/a and the prediction error in mm^2/a^2 (MQyear 2009, Div=TRUE)

3.4.3 Post-processing

The reason for post-processing is to convert the output (TopKriging interpolation results) from the last section to the *STOBIMO watershed* dataset. This is done by converting the rtop output, a spacial object with predictions to a raster object and then back to a spacial object. The outcome is a specific runoff for each MoRE AU.

3.4.3.1 Transformation to raster

To transfer the specific runoff from the *predictionLocation* spatial dataset to the *STOBIMO* spatial dataset, a simple and comprehensible method to do is to take an intermediate step by converting (or discretizing) the spatial object to a raster object. Beforehand, the *predictionLocation* spatial dataset is sorted by area size, therefore the smallest, hence most upstream, watershed is the most top one. This ensures that in the next step the value (specific runoff) of the most upstream watershed is used. A raster object with an 2 by 2 km resolution is created and the specific runoff value of the most top layer is taken from the *predictionLocation* spatial dataset and projected to the raster pixel. This is done by calling *rasterize* function of the *raster* package [24]. An example can be seen in figure 3.9.

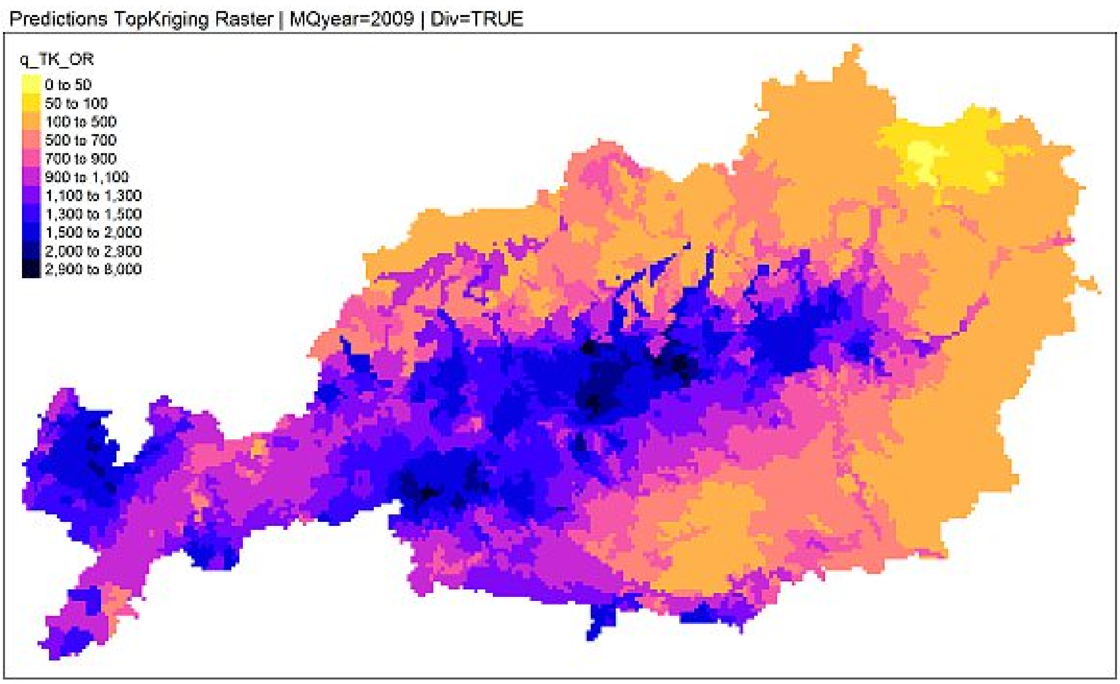


Fig. 3.9: Raster object with predicted specific runoff in mm/a (MQyear=2009, Div=TRUE). Transformed out of 3.8 on the previous page.

3.4.3.2 Transformation to STOBIMO watersheds

The transformation back to the *STOBIMO* spatial dataset is done by calling *extract* function [24]. This function calculates the mean value out of all overlapping raster pixel, specific runoff (*q_mm_sim*) in mm/a, for each MoRE AU. An example can be seen in figure 3.10.

Because the MoRE model [4] requires the runoff per AU in m^3/s the specific runoff in mm/a is converted to runoff in m^3/s by using equation 3.5.

$$Q_{AU} = \frac{q_{AU\_mm}}{(3.6 \cdot 24 \cdot 365^{*})} \cdot A_{AU} \qquad {}^{*}366\ for\ leap\ years \tag{3.5}$$

q_{AU_mm} ... specific runoff for each AU in mm/a
A_{AU} ... watershed area of AU in km^2
Q_{AU} ... runoff of the AU in m^3/s

Some Diversions go to the upstream AU, causing a hydraulic short circuit which cannot be simulated in a flow tree calculation, therefore the diversion runoff and diversion watershed area of the giving AU will be decreased by the same amount as the receiving AU will be increased.

3.4.3.3 Splitting factor and flow tree calculations

If an analytical unit (AU) has two downstream AUs the MoRE model [4] uses a splitting factor (SF), called *RM_FCT_Q_SPLIT* here named $SF_{Q.Split}$, to split the loads (e.g. runoff, nutrients, ect.) among the two downstream AU.

To calculate this splitting factor (See equation 3.8), the total area along the flow tree, hence the hierarchical structure of rivers from source to mouth by considering diversions has to be calculated and compared with the diversion area (3.3.2.1).

The aggregation of runoff (and watershed area) along the flow tree is done by reproducing the workflow of the MoRE model [4] in a *R* [22] script. This SF ranges from 0.002 to 0.998, as

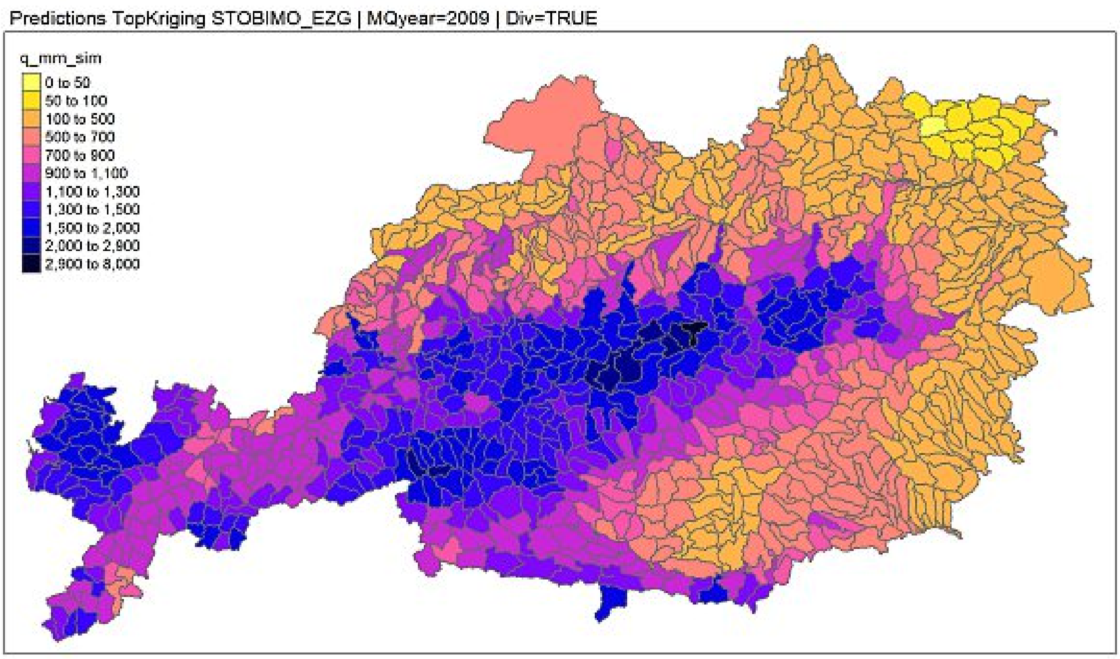


Fig. 3.10: STOBIMO spatial dataset with predicted specific runoff in mm/a for each MoRE AU (MQyear=2009, Div=TRUE). Transformed by calculation of the mean value per AU from 3.9 on the preceding page.

shown in table 3.5 and does not change by year, hence stays constant as the diversion area does not change by year.

The script of splitting factor calculation can be seen in appendix C.4.1 and the aggregation of runoff along the flow tree in appendix C.4.2.

Tab. 3.5: Splitting factor (SF) for MoRE model (Without zero values).

Variable	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Splitting factor	0.002	0.099	0.360	0.413	0.740	0.998

In a first step, the total effective watershed area is aggregated for each AU by considering all upstream diversions including the outlet diversion area ($A_{Div.Outlet}$) in the current AU (see equation 3.6). A definition of the used watershed areas can be seen in figure 3.4.

$$A_{tot} = A_{eff} + A_{Div.Outlet} \tag{3.6}$$

In the second step, the total area for each AU is used to calculate the SF for each AU (see equation 3.7). This SF is used in further calculations and also exported to the MoRE model for use in the *STOBIMO* project.

$$SF_{Q.Split} = \frac{A_{Div.Outlet}}{A_{tot}} \tag{3.7}$$

The third and final step calculates the effective and diversion runoff for each AU (see equation 3.8). This modelled runoff data can be compared for validation with the observed runoff data (See 3.4.4.4).

$$\begin{aligned} Q_{eff} &= (1 - SF_{Q.Split}) \cdot Q_{tot} \\ Q_{Div.Outlet} &= SF_{Q.Split} \cdot Q_{tot} \end{aligned} \tag{3.8}$$

$SF_{Q.Split}$ ... SF for diversion split
A_{tot} ... total area at the end of the AU incl. all diversion areas in km^2
A_{eff} ... effective area at the end of the AU (area of river runoff) in km^2
$A_{Div.Outlet}$ ... diversion area leaving the AU in km^2
Q_{tot} ... total runoff at the end of the AU incl. all diversion areas in m^3/s
Q_{eff} ... effective runoff at the end of the AU (river runoff) in m^3/s
$Q_{Div.Outlet}$ ... diversion runoff leaving the AU in m^3/s

Finally, the *STOBIMO* spatial dataset contains the information needed for further processing in the MoRE model [4].

3.4.4 Validation

3.4.4.1 Cross Validation

To validate the prediction improvements which can be obtained by diversion consideration in TopKriging interpolation, a leave-one-out-cross-validation is performed on the *Observation* spatial dataset. Cross-validation was first introduced by Cressie (1991) and is a common method to measure the prediction efficiency of an interpolation.

At each increment, one observed catchment is removed and then predicted by the TopKriging interpolation. This increment is repeated until all observation catchments are interpolated. The deviation of prediction to the actual observed value is measured using the Nash-Sutcliffe efficiency (NSE) and Modified Nash-Sutcliffe efficiency (mNSE) coefficients (See 3.4.4.4). They show the goodness-of-fit of the predictions.

The actual cross-validation is done before the TopKriging interpolation (See 3.4.2). As a build-in-function of the rtop-package [12] it can be easily started by calling function *rtopKrige(rtopObj,cv=TRUE))*. This combines leave-one-out-cross-validation with TopKriging for the *Observation* dataset. An example can be seen in figure 3.11.

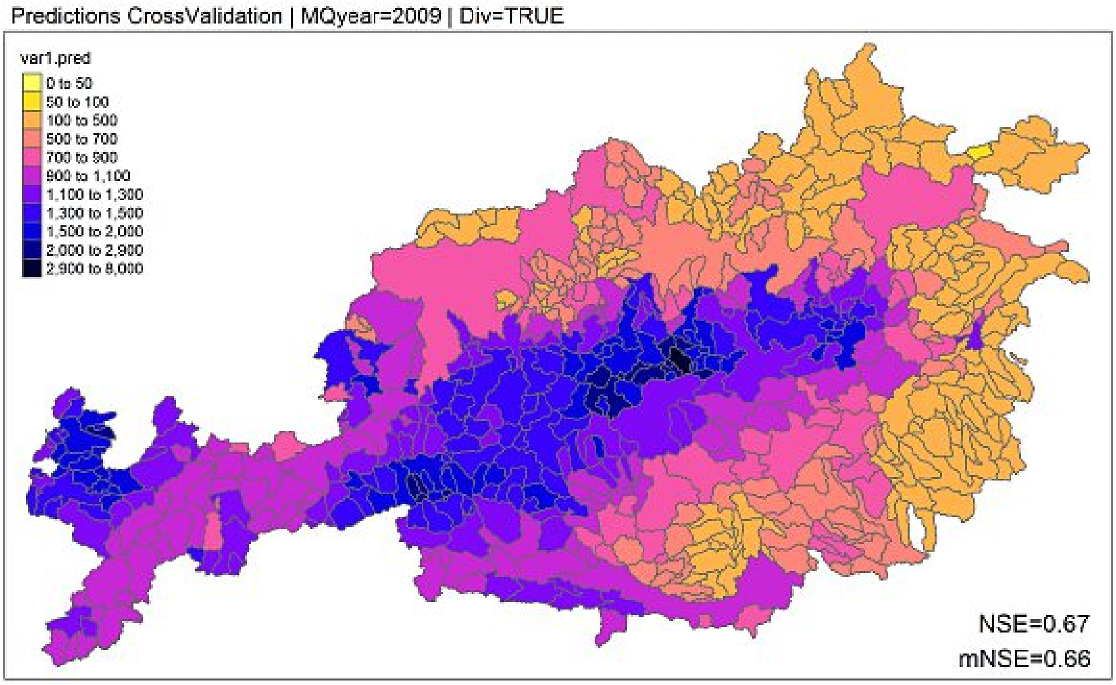


Fig. 3.11: Cross-validation prediction as specific runoff in mm/a of year 2009 (Div=TRUE) of the observations in figure 3.6 including the model efficiency coefficients.

3.4.4.2 Comparison simulated with observed runoff values

The modelled stream gauging station runoff (3.4.3.3) is compared with the observed runoff (3.3.1.4) and the modelled diversion runoff (3.4.3.3) is compared with the observed diversion runoff (3.3.2.2). Model efficiency is measured by using the NSE and mNSE model efficiency coefficient (3.4.4.4). For some groups of diversion only the sum of observed runoff data was available. In that case, the simulated diversion runoffs were cumulated to a single runoff to make it comparable.

Runoff gauging stations MQ runoff difference
For further examination of the performance of each runoff gauging station, a list with the MQ runoff difference between predictions to observations for each runoff gauging station was created, visible in appendix B. The MQ runoff difference is calculated with the equation 3.9.

$$MQ_{Diff} = \frac{MQ_{obs} - MQ_{sim}}{MQ_{obs}} \cdot 100 \tag{3.9}$$

MQ_{Diff} ... MQ runoff difference in %
MQ_{obs} ... observed MQ runoff in m^3/s
MQ_{sim} ... simulated (predicted) MQ runoff in m^3/s

3.4.4.3 Validation of the assumption

The theory from subsection 3.2 is that effects of overflow, minimum flow requirements, and revisions, cancel each other out. And the assumption is that watershed area represents the diversion runoff. To validate this assumption, the relationship between model efficiency improvement (NSE and mNSE) of validation in each year and the total runoff were compared. If no correlation exists, the theory of cancelling-out-effects is valid and the approach with watershed area is suitable to represent the diversion runoff in both wet and dry years. On the other hand a correlation between those variables would indicate that some effects are under- or overestimated and the approach with diversion areas is too static to reflect those effects.

The total runoff per year is calculated as the sum of all annual runoff over all runoff gauging stations using the data from MQ table (subsection 3.4.1.3). The model efficiency improvement are calculated as the value difference of NSE and mNSE between with and without diversion consideration (Div=TRUE/FALSE) for both cross-validation (3.4.4.1) and effective runoff validation (3.4.4.2).

For all statistical tests the level of significance (α) is set to 5%.

3.4.4.4 Nash–Sutcliffe model efficiency coefficient

Nash-Sutcliffe efficiency (NSE), an indicator of model performance, was first introduced by Nash and Sutcliffe in 1970. It indicates how well the simulated data fit the observed data. The range is between minus infinite (-Inf) and 1. According to equations 3.10 & 3.11, the definition is 1 minus the ratio between the sum of squared deviations of simulated and observed runoff and the variance of the observed runoff.

$$NSE = 1 - \frac{\sum_{i=1}^{n}(Q_{sim.i} - Q_{obs.i})^2}{\sum_{i=1}^{n}(Q_{obs.i} - \bar{Q}_{obs})^2} \tag{3.10}$$

$$\bar{Q}_{obs} = \frac{1}{n} \cdot \sum_{i=1}^{n} Q_{obs.i} \tag{3.11}$$

A NSE-value of 1 would represent a perfect fit of the model, hence predicted values are equal to observed values. In real-world applications, due to uncertainties and measurement errors, the aim is to get as close as possible to a NSE-value of 1. If NSE-value is below 0 the performance is not acceptable as it indicates that the mean observed value is a better predictor than the simulation itself [25].

The mNSE, with default value j=1, is a modification of the NSE, introduced by Legates and McCabe (1999) to make the coefficient less sensitive to high extreme values compared to the NSE.

$$mNSE = 1 - \frac{\sum_{i=1}^{n}|Q_{sim.i} - Q_{obs.i}|^j}{\sum_{i=1}^{n}|Q_{obs.i} - \bar{Q}_{obs}|^j} \tag{3.12}$$

NSE ... Nash-Sutcliffe efficiency
$mNSE$ ... Modified Nash-Sutcliffe efficiency
$Q_{sim.i}$ ... simulated runoff in m^3/s
$Q_{obs.i}$ ... observed runoff in m^3/s
$\bar{Q}_{obs}$ ... mean of all observed runoffs in m^3/s
j ... exponent to be used in the computation of the mNSE (Default = 1)

Analysis was conducted in R [22] and NSE & mNSE-value were calculated using the package *hydroGOF* [26].

Model efficiency

Moriasi et al. [25] proposed a range of NSE values, shown in table 3.6, to assess model efficiency on a monthly time intervals basis. They also suggested that increasing time intervals could bear generally stricter performance rating, which is the case in this thesis were MQ-values are based on annual time intervals.

Tab. 3.6: Typical NSE value range for yearly time steps to assess model efficiency (after Moriasi et al. [25]).

Value range NSE	Model efficiency
0.75 < NSE < 1.00	very good model efficiency
0.65 < NSE < 0.75	good model efficiency
0.50 < NSE < 0.65	sufficient model efficiency
NSE < 0.50	insufficient model efficiency

3.4.5 Software

To create this document LaTeX was used with *TeXstudio*. For literature collection *Citavi* was used. Further was *ArcMap* a great tool for analysis and modification of spatial data and *Dia* a useful tool to create the diagram in this thesis.

Analysis was conducted in *R* (version 4.0.2) [22] a powerful open-source software for statistics and data science. Its functionality can be extended by packages, which are extensions of the base software. For this master thesis the following *R* packages were used:

- sp: Classes and Methods for Spatial Data [27] & [28] (version: 1.4-4)
- sf: Simple Features for R: Standardized Support for Spatial Vector Data [29] (version: 0.9-6)
- rtop: Interpolation of Data with Variable Spatial Support [12] & [30] (version: 0.5-14)
- data.table: Extension of 'data.frame' [31] (version: 1.13.2)
- raster: Geographic Data Analysis and Modeling [24] (version: 3.4-5)
- hydroGOF: Goodness-of-Fit Functions for Comparison of Simulated and Observed Hydrological Time Series [26] (version: 0.4-0)
- tmap: Thematic Maps in R. [32] (version: 3.2)
- ggplot2: Elegant graphics for data analysis [33] (version: 3.3.2)

- hydroTSM: Time Series Management, Analysis and Interpolation for Hydrological Modelling [23] (version: 0.6-0)

The script to analyse the data in this master thesis can bee seen in the appendix C. The parent script (C.1) connects all other child scripts (C.2 to C.4.4).

Chapter 4

Results

In this chapter the results of interpolation are shown. To compare the results and to conclude the improvements which can be achieved through diversion consideration, the results are shown for two cases, namely without and with diversion consideration. Those cases are distinguished by area data, which are used as input data for the interpolation:

- Without diversion consideration (Div=FALSE), using the orographic watershed area (A_{oro}) for specific runoff calculation, resulting in the natural specific runoff (q_{nat}) in mm/a.
- With diversion consideration (Div=TRUE), using the effective watershed area (A_{eff}) for specific runoff calculation, resulting in the effective specific runoff (q_{eff}) in mm/a.

For easier understanding in the following chapter it will be referred as without diversion consideration (Div=FALSE) and with diversion consideration (Div=TRUE).

Analysis was conducted in R [22] and figures were produced using the package *ggplot2* [33] or package *tmap* [32].

4.1 TopKriging Interpolation

4.1.1 Observations

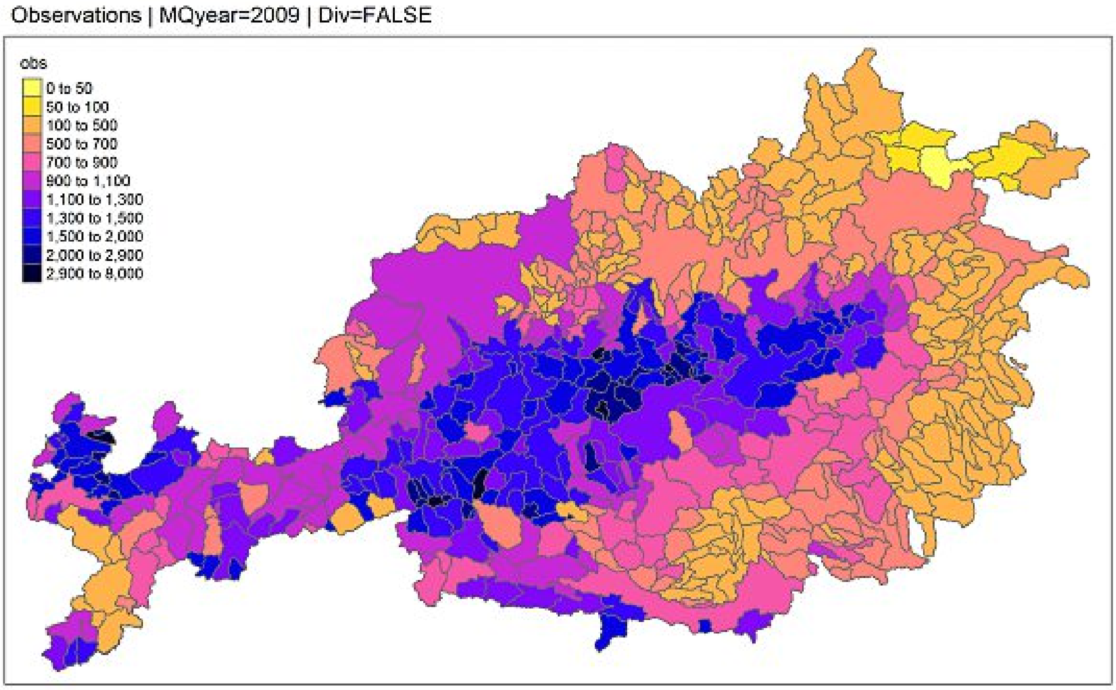
The input for the interpolation is the *Observation* spatial dataset with the observations attached, which can be seen in figure 4.1. The observations, the observed specific runoffs in mm/a, are plotted onto the overlapping watersheds which are sorted by area size. That means that only headwaters without upstream watersheds are shown entirely and all other watersheds are overlapped by their upstream watersheds. In the upper part of the map some bigger and streched areas can be seen, those are the big streams like Inn and Danube river with a lower density of stream gauge stations.

The specific runoff without diversion consideration can be viewed in figure 4.1a and the specific runoff with diversion consideration can be viewed in figure 4.1b.

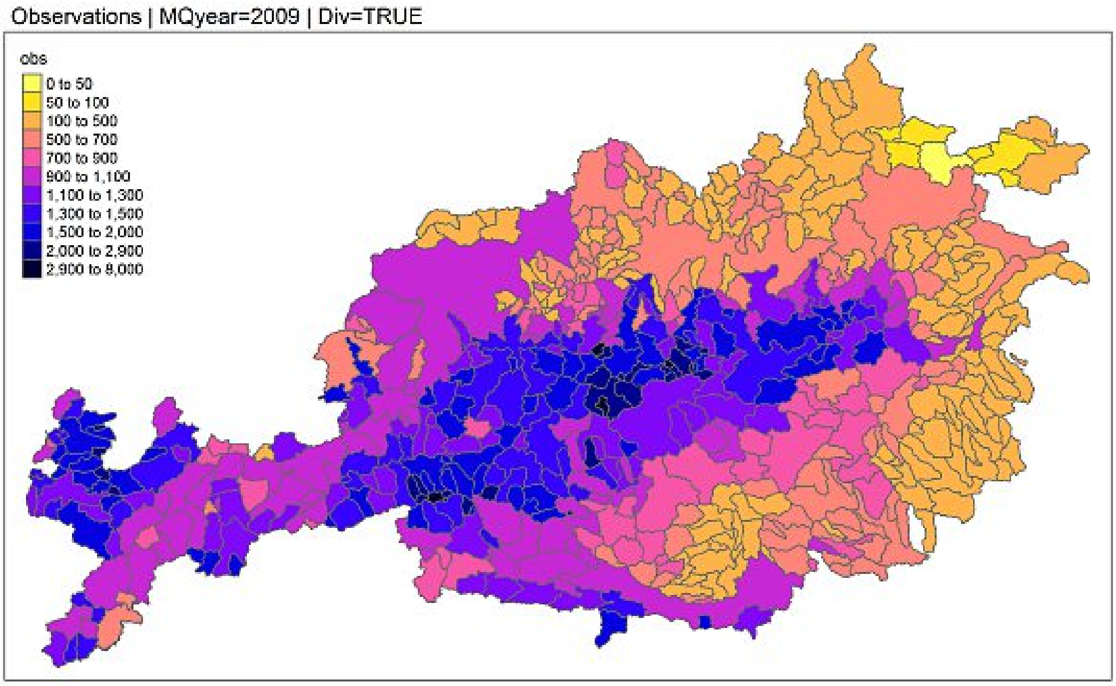
4.1.2 Predictions

The output of the TopKriging interpolation is the *Prediction Locations* spatial dataset with the predictions attached, which can be seen in figure 4.2. The predictions, the predicted specific runoffs in mm/a, are plotted onto the overlapping watersheds which are sorted by area size. That means that only headwaters without upstream watersheds are shown entirely and all other watersheds are overlapped by their upstream watersheds.

The prediction error in mm^2/a^2 is shown in figure 4.3.

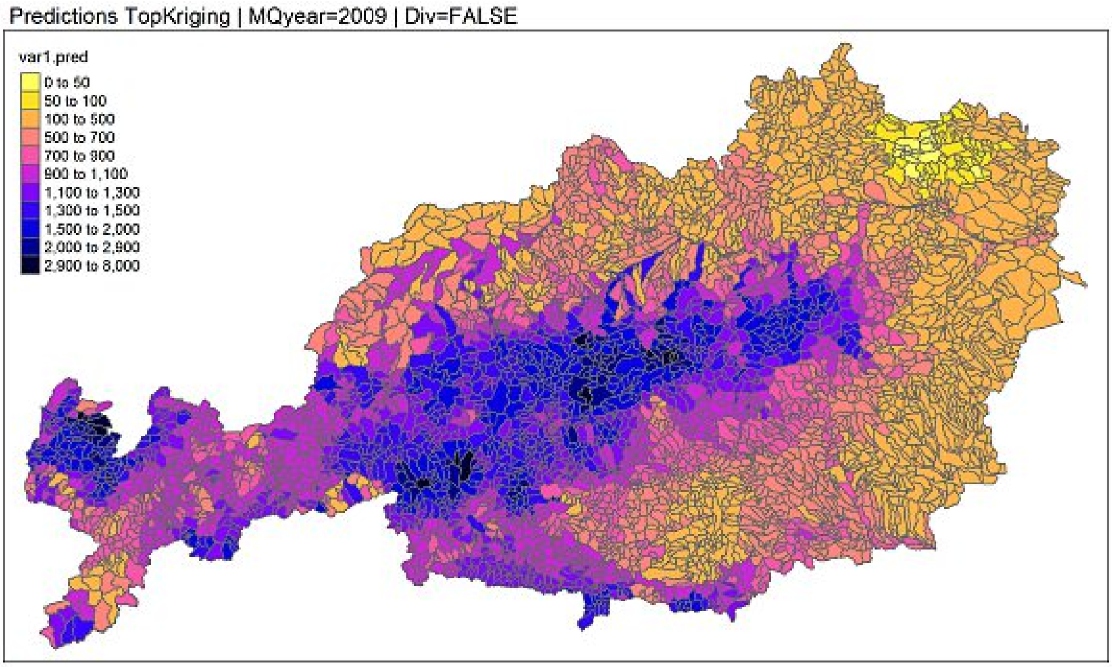


(a) Without diversion consideration (Div=FALSE)

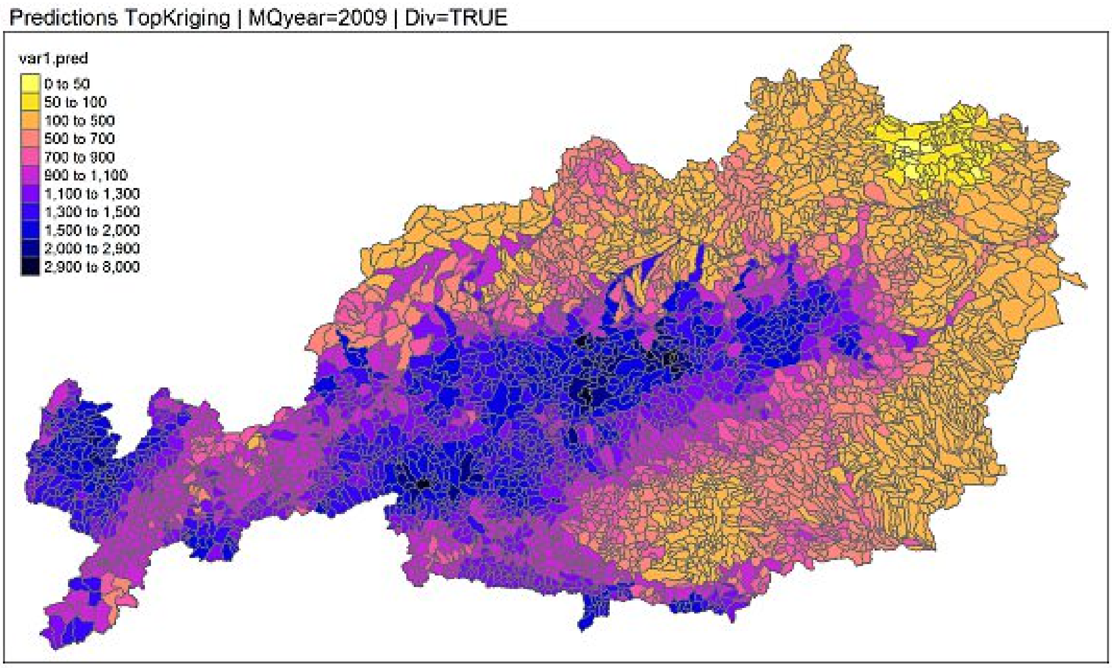


(b) With diversion consideration (Div=TRUE)

Fig. 4.1: Observed specific runoff (*obs*) in mm/a for the year 2009.

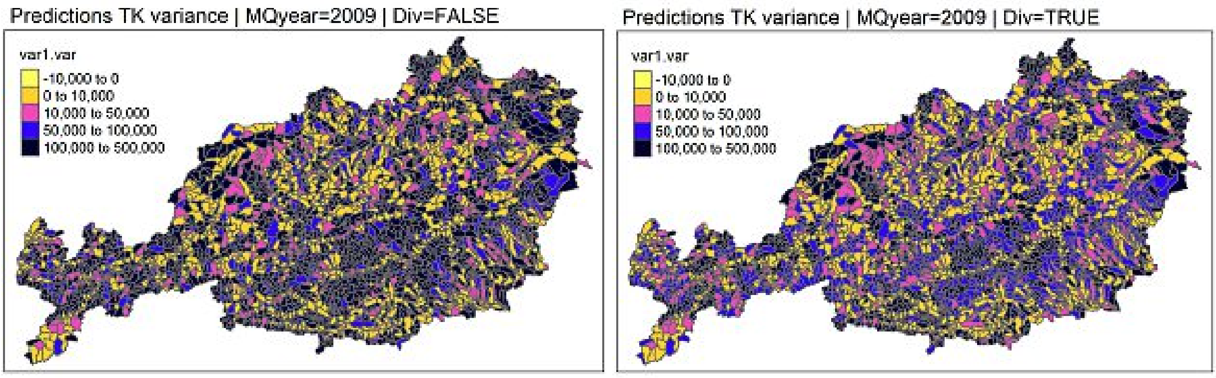


(a) Without diversion consideration (Div=FALSE)



(b) With diversion consideration (Div=TRUE)

Fig. 4.2: Predicted specific runoff ($var1.pred$) in mm/a interpolated with TopKriging out of fig. 4.1 for the year 2009.

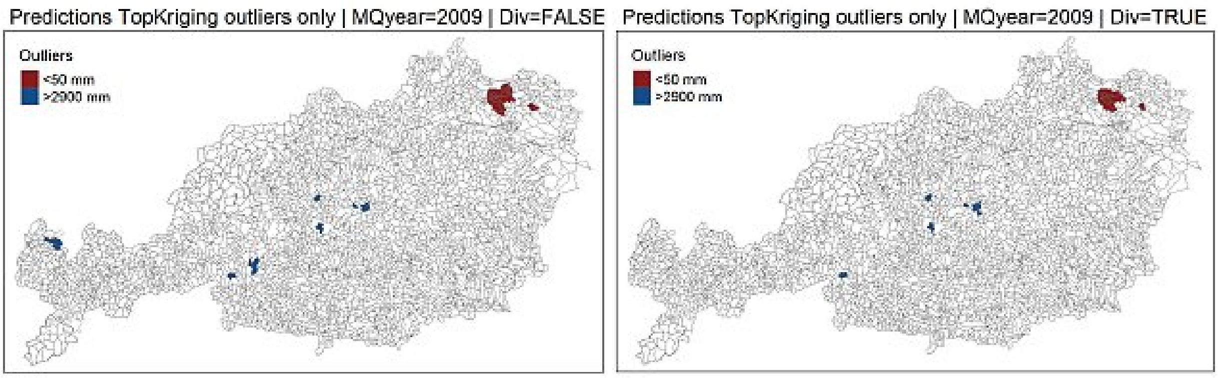


(a) Without diversion consideration (Div=FALSE) **(b)** With diversion consideration (Div=TRUE)

Fig. 4.3: Prediction error or estimated kriging Variance ($var1.var$) in mm²/a² of the interpolation with TopKriging for the year 2009.

4.1.3 Outliers

Outliers are defined as the watersheds exceeding the 1% and 99% percentile, hence lower and upper limit respectively. In figure 4.4, the outlier watersheds for year 2009 are highlighted and distinguished by upper and lower limit, hence specific runoff higher 2900 mm/a and lower 50 mm/a respectively.



(a) Without diversion consideration (Div=FALSE) **(b)** With diversion consideration (Div=TRUE)

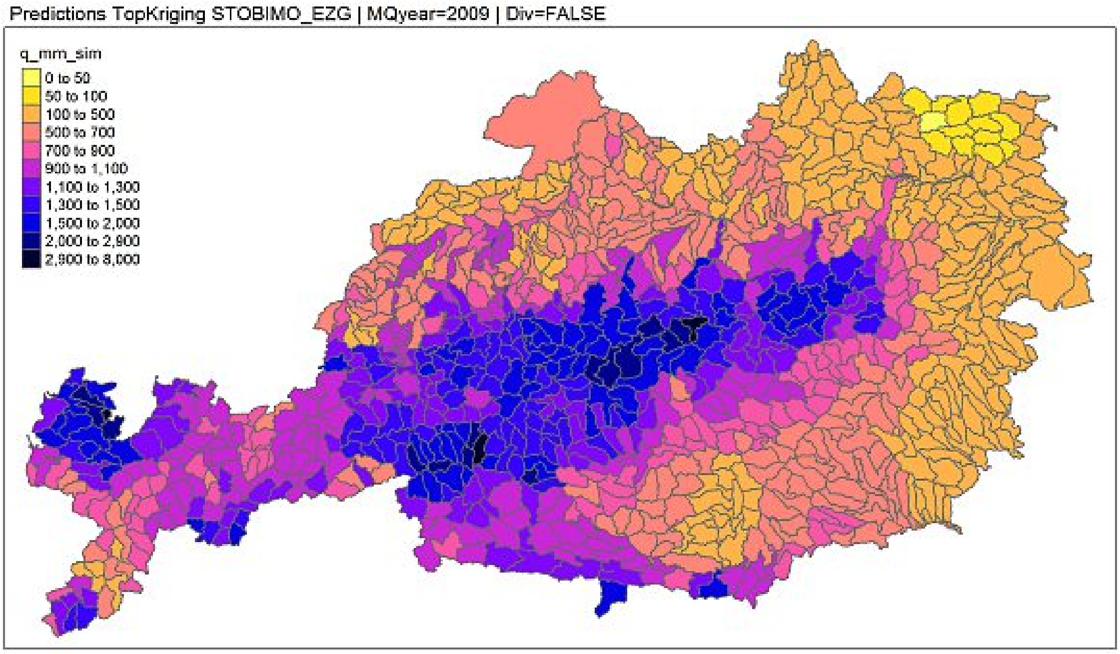
Fig. 4.4: Outliers of the TopKriging predictions for the year 2009 distinguished by upper and lower limit.

4.1.4 Predictions STOBIMO

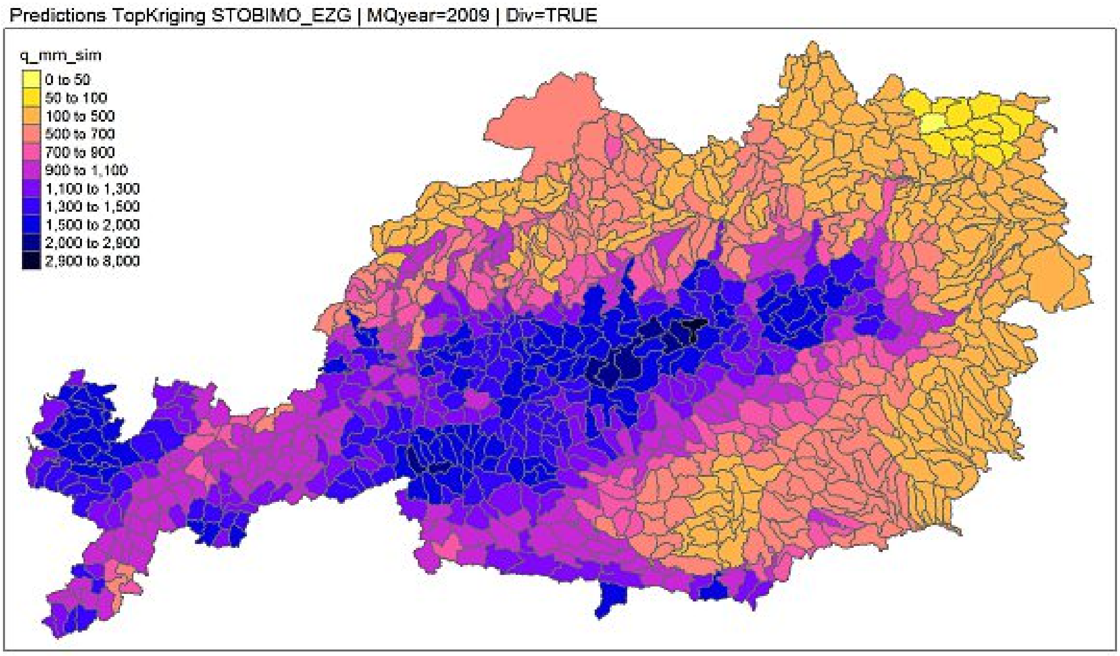
The specific runoff per MoRE AU for the *STOBIMO* spatial dataset can be seen in figure 4.5.

4.1.5 Process steps comparison

A comparison of all runoff values distinguished by process step can be seen in figure 4.6. The figure illustrates the distribution of data and percentile statistic for Observation (obs), cross-validation (pred_CV), TopKriging predictions (pred_TK) and STOBIMO watersheds (STOBIMO) each for with and without diversion consideration. Due to the different resolution of watersheds (Step:



(a) Without diversion consideration (Div=FALSE)



(b) With diversion consideration (Div=TRUE)

Fig. 4.5: Predicted specific runoff (*q_mm_sim*) for the STOBIMO watersheds in mm/a for the year 2009.

pred_TK) the number of watersheds vary more than an order of magnitude within the process steps.

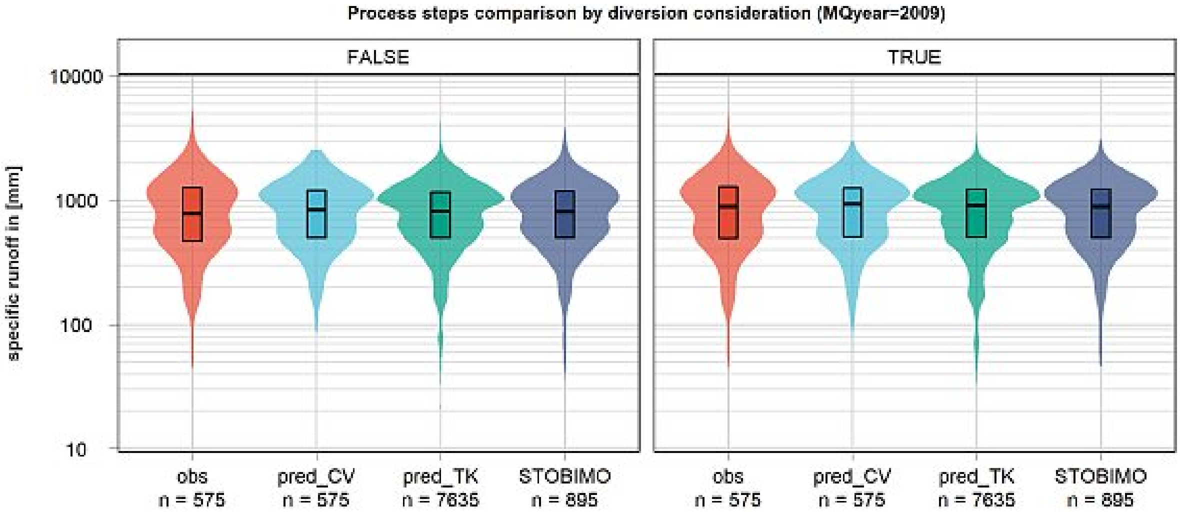


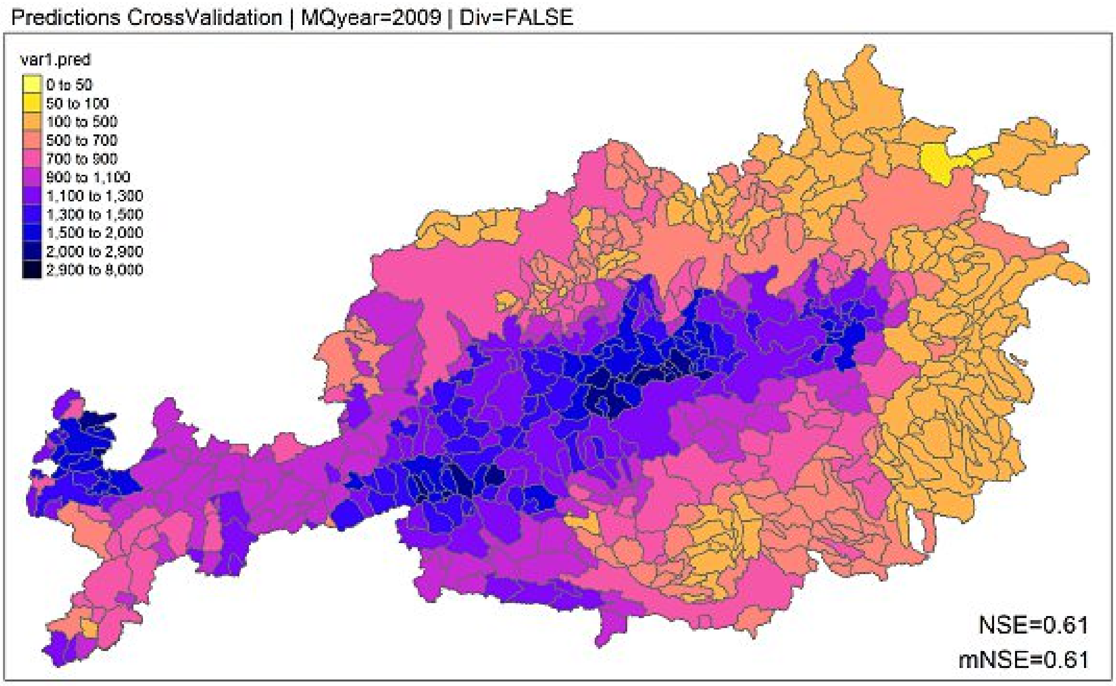
Fig. 4.6: One year comparison (2009) of the specific runoff values in mm/a for each process step by diversion consideration (Div). Kernel density shows distribution of data, black box shows statistics (25%, 50% (median) & 75% percentile) for the data.

4.2 Validation

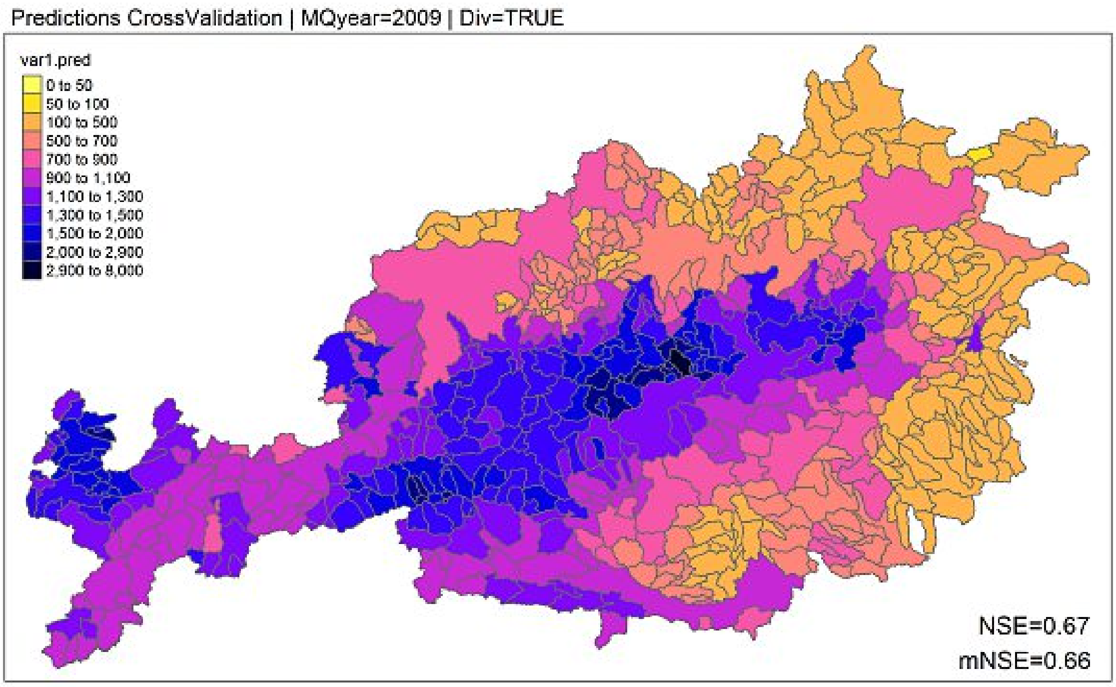
4.2.1 Observations validated with cross-validation

The result for the year 2009 of the leave-one-out-cross-validation, which is performed on the *Observation* spatial dataset with the specific runoff in mm/a, can be seen in figure 4.7. In the right bottom corner the model efficiency coefficients Nash-Sutcliffe efficiency (NSE) and Modified Nash-Sutcliffe efficiency (mNSE) are displayed.

The coefficient values (NSE and mNSE) of Cross-Validation for all simulated years (2009-17) can be viewed in table 4.1. For the year 2009 the NSE increases by 0.06 and the mNSE by 0.05 in general over all runoff gauging stations. Visually these improvements can be seen as a smoother change over the AU (figure 4.7) and the reduced prediction variance in figure 4.8. Throughout all years improvements can be achieved in terms of both coefficients NSE (mean 0.07) and mNSE (mean 0.06), this can be seen in figure 4.9 and table 4.1. Taking only diversion affected gauges in account (figure 4.9b & right side of table 4.1), the cross-validation model efficiency coefficient values increase significantly through diversion consideration in average by 0.33 for NSE and 0.21 for mNSE.

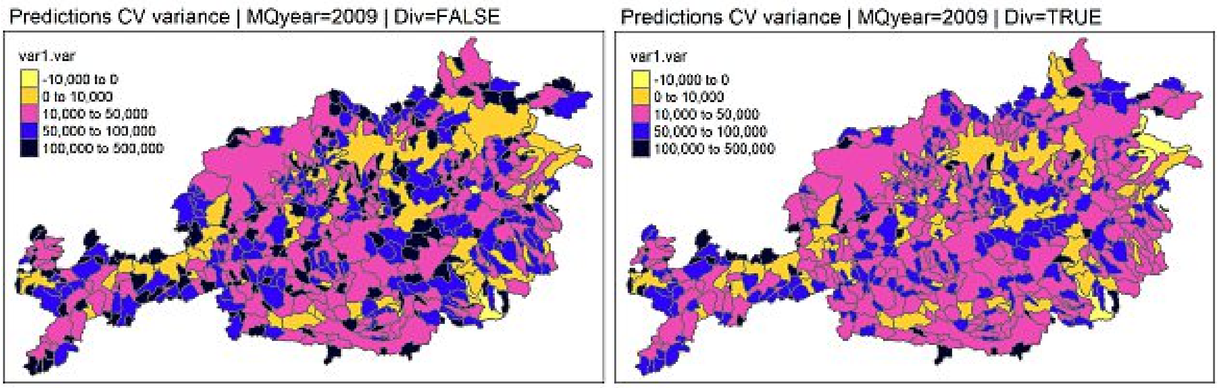


(a) Without diversion consideration (Div=FALSE)



(b) With diversion consideration (Div=TRUE)

Fig. 4.7: With Cross-Validation (CV) predicted specific runoff (*var*1.*pred*) in mm/a with model efficiency coefficient (NSE & mNSE) for year 2009.

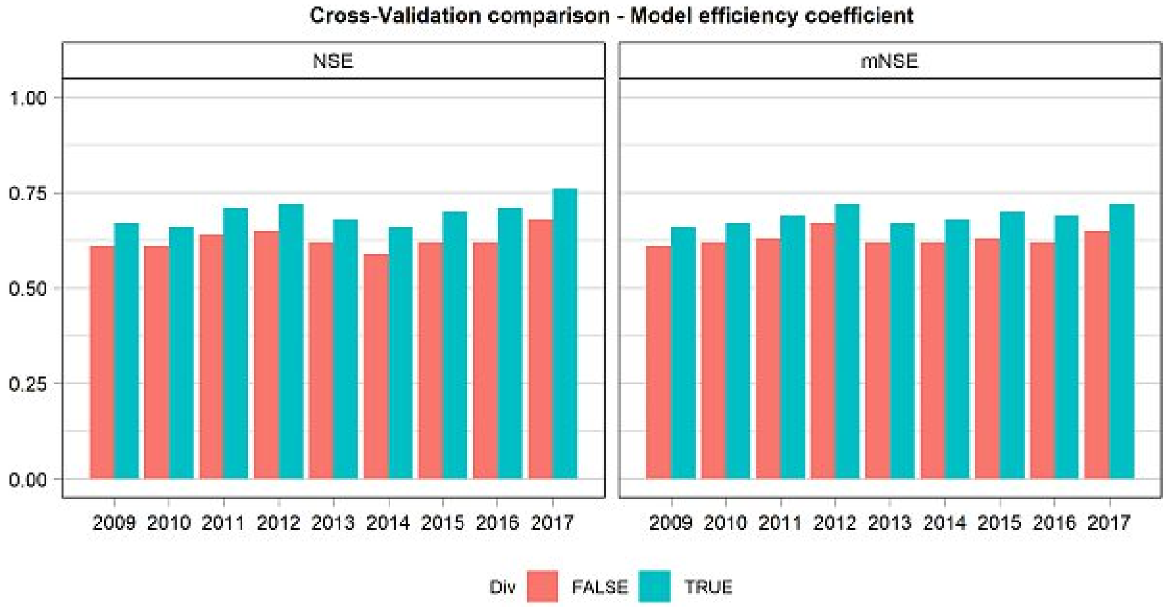


(a) Without diversion consideration (Div=FALSE) **(b)** With diversion consideration (Div=TRUE)

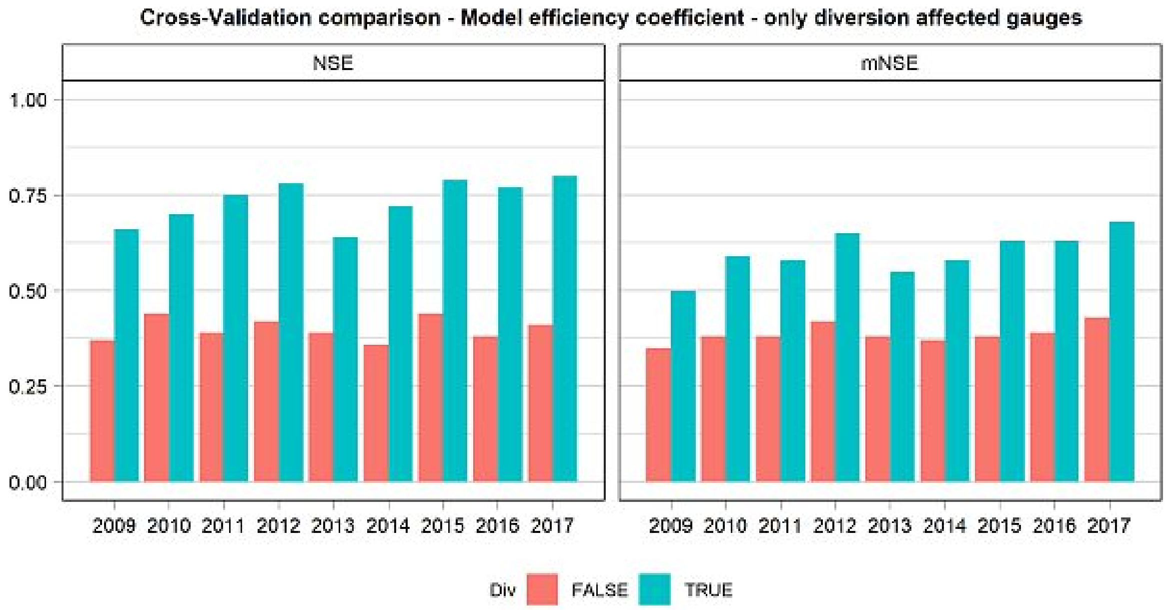
Fig. 4.8: Prediction error (estimated kriging Variance) ($var1.var$) in mm^2/a^2 of the cross-validation (CV) for year 2009.

Tab. 4.1: Cross-Validation model efficiency coefficient for years 2009-17. TRUE/FALSE refers to diversion consideration. Distinguished between all and only diversion affected runoff gauging stations. Gauge count located in column N.

	All runoff gauging stations					Only diversion affected stations				
	NSE		mNSE			NSE		mNSE		
year	FALSE	TRUE	FALSE	TRUE	N	FALSE	TRUE	FALSE	TRUE	N
2009	0.61	0.67	0.61	0.66	575	0.37	0.66	0.35	0.50	96
2010	0.61	0.66	0.62	0.67	574	0.44	0.70	0.38	0.59	96
2011	0.64	0.71	0.63	0.69	576	0.39	0.75	0.38	0.58	97
2012	0.65	0.72	0.67	0.72	572	0.42	0.78	0.42	0.65	97
2013	0.62	0.68	0.62	0.67	571	0.39	0.64	0.38	0.55	97
2014	0.59	0.66	0.62	0.68	557	0.36	0.72	0.37	0.58	93
2015	0.62	0.70	0.63	0.70	550	0.44	0.79	0.38	0.63	92
2016	0.62	0.71	0.62	0.69	547	0.38	0.77	0.39	0.63	92
2017	0.68	0.76	0.65	0.72	543	0.41	0.80	0.43	0.68	92
mean	0.63	0.70	0.63	0.69	563	0.40	0.73	0.39	0.60	95



(a) All runoff gauging stations.



(b) Only diversion affected runoff gauging stations (DivBias=TRUE)

Fig. 4.9: Comparison of model efficiency coefficient (NSE & mNSE) per year for cross-validation distinguished by diversion consideration (Div=TRUE/FALSE) and between all and only diversion affected runoff gauging stations.

4.2.2 Predictions validated with runoff comparison

4.2.2.1 Diversion runoff comparison

Obviously, diversion runoff comparison can only be made for the case "With diversion consideration (Div=TRUE)", because in case "without diversion consideration (Div=FALSE)" no diversions were considered at all.

Figure 4.10 compares the simulated diversion MQ runoff, with the observed diversion MQ runoff, hence the collectedobserved diversion MQ runoff (See 3.3.2). The closer the values are to the 1:1 line the better the model efficiency and hence the prediction. For further information the individual diversions are distinguished between the source of diversion area. For *DivAREA* the diversion area was supplied as area, and for *MQ_to_AREA* calculated out of given MQ values (See 3.4.1.1). The outliers of category *Storage hydropower* belong to 4 diversions from hydropower stations.

The outlier of category *Pumped storage hydropower* belongs to one diversion of a hydropower station with pumpstorage scheme which even has negative runoffs for some years, hence the total amount pumped is greater than the amount used for energy production. Those negative runoffs are out of plotting range due to the logarithmic scale.

For this reason, and for increased comparability, both MQ runoff simulated and observed, were divided by their diversion area (A_{Div}) to obtain the specific diversion runoff in $\mathrm{m}^3/(\mathrm{s\,km}^3)$. Those specific runoffs were compared in figure 4.11 for each diversion category including the 1:1 line of perfect prediction. For further information the individual diversions are distinguished between the source of diversion area (as in figure 4.10). In category *Pumped storage hydropower* the same outlier diversion with negative runoff values as in figure 4.10 can be clearly identified. The Outliers of category *Storage hydropower* belong to 4 diversions where 3 of them are the same as in figure 4.10.

A comparison of the model efficiency coefficients of MQ runoff for all analysing years can be seen in figure 4.12 with the corresponding left side of table 4.2. Prediction efficiency is in average 0.84 for NSE and 0.67 for mNSE. Right side of table 4.2 shows prediction efficiency by diversion category over all years (2009-17).

Tab. 4.2: Runoff comparison model efficiency coefficient distinguished by years (left) and by diversion category (right). Gauge count located in column N.

	MQ_{Div}		
year	NSE	mNSE	N
2009	0.87	0.72	47
2010	0.90	0.72	52
2011	0.88	0.71	52
2012	0.84	0.67	54
2013	0.83	0.66	53
2014	0.74	0.57	53
2015	0.82	0.64	55
2016	0.83	0.66	55
2017	0.81	0.64	52
mean	0.84	0.67	53

	MQ_{Div}		
Diversion category	NSE	mNSE	N
Storage hydropower	0.73	0.52	342
Canal & Others	0.39	0.39	33
Pumped storage hydropower	0.81	0.6	53
Run-of-river hydropower	0.73	0.63	45

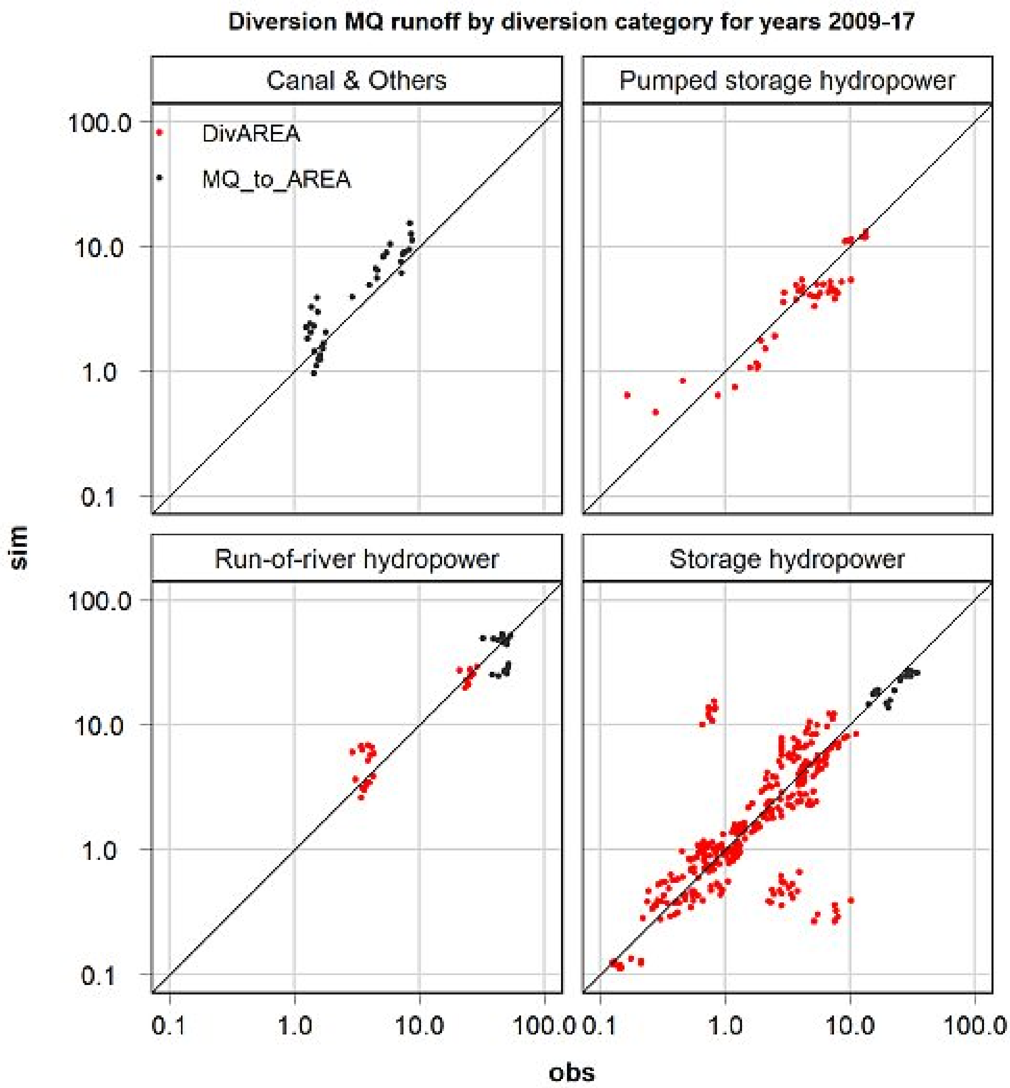


Fig. 4.10: Comparison of simulated to observed diversion runoff in m^3/s distinguished by diversion area transformation for each diversion category for all analysis years (2009-17).

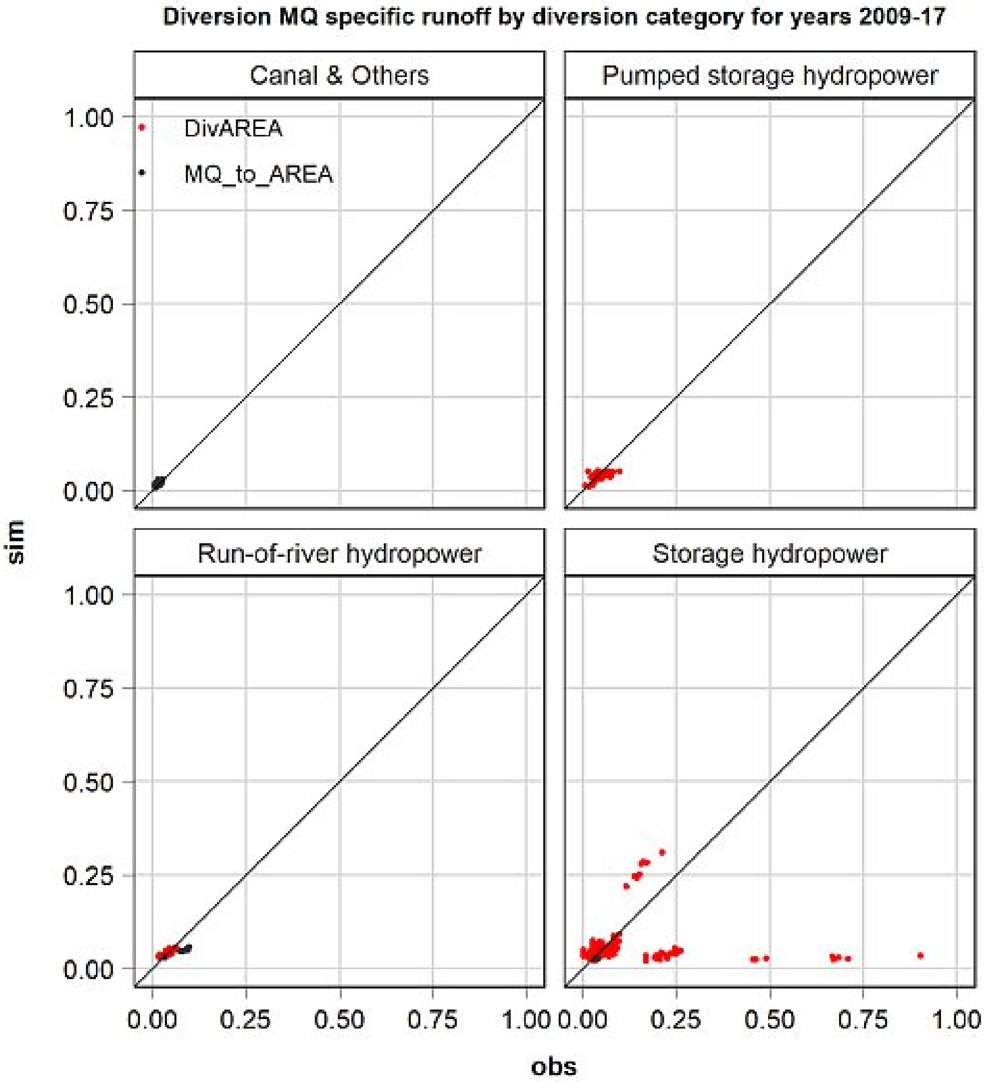


Fig. 4.11: Comparison of simulated to observed diversion specific runoff in $m^3/(s\,km^3)$ distinguished by diversion area transformation for each diversion category for all analysis years (2009-17).

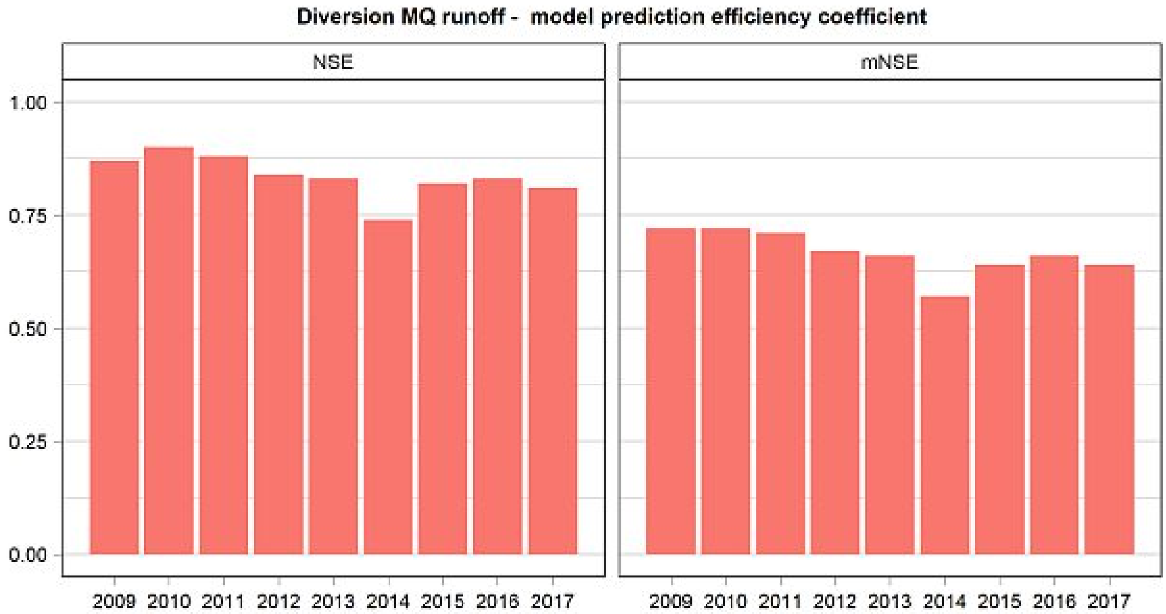


Fig. 4.12: Comparison of model efficiency coefficient (NSE & mNSE) for diversion runoff prediction distinguished by years (Div=TRUE).

4.2.2.2 Gauge runoff comparison

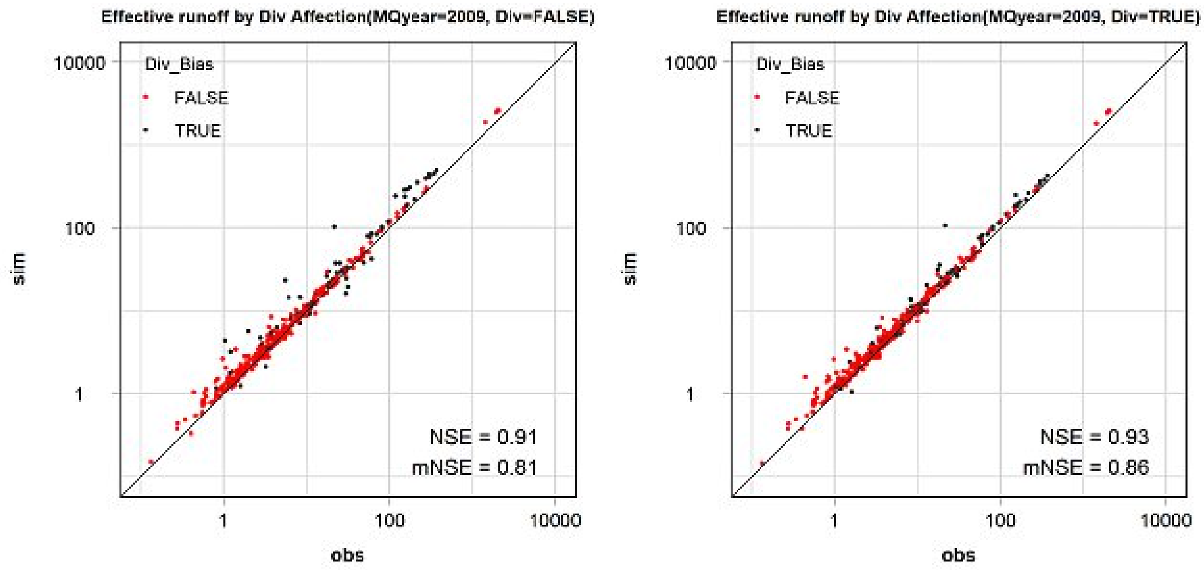
Gauge runoff comparison compares the simulated effective MQ runoff, hence the aggregated predicted runoff per AU with the observed effective MQ runoff, hence the MQ runoff measured at the stream gauge station. A comparison between with and without consideration for the year 2009 can be seen in figure 4.13. For easier comparison the model efficiency coefficients NSE and mNSE are displayed in the right bottom corner. The closer the values are to the 1:1 line the better the model efficiency. The highest values (MQ$\geq$1000 m^3/s) are the runoff gauges at Danube river. The second highest values (MQ=100$-$1000 m^3/s) belong to the following rivers sorted by size: lower Inn, Salzach, Enns, Drau, intermediate Inn, Mur and Traun. Runoff gauging station *Flattach* (HZBR-Nb.:213124) can be seen as outlier in both figures because the prediction efficiency is very low. The comparisons for each year (2009-17) for only division affected gauges (Div_Bias=TRUE) are illustrated in figures 4.15 and 4.16. The figures show similar results over all years, both with and without consideration.

Around 17%, or 80 out of 408 runoff gauging stations, are diversion affected (Div_Bias=TRUE) in year 2009. The comparison of only those diversion affected runoff gauging stations for year 2009 can be seen in figure 4.14. The gauge runoff prediction efficiency in 2009 increases for diversion affected gauges by 0.23 from 0.71 to 0.94 NSE whereas for all gauges it increases from a higher level by 0.02 from 0.91 to 0.93 NSE. The left side of table 4.3 shows the model efficiency coefficients comparison for all runoff gauging stations in the study area. There are slightly improvements thought diversion consideration, in mean a difference of 0.03 for NSE and 0.06 for mNSE. The right side of table 4.3 shows the model efficiency coefficients comparison only for diversion affected runoff gauging stations in the study area. The improvements through diversion consideration are here more significant, in mean a difference of 0.32 for NSE and 0.24 for mNSE. Figure 4.17 show a graphic illustration of table 4.3. It is evident that in figure 4.17a the coefficients do not fluctuate over the years. The situation is different in second figure 4.17b where

only diversion affected gauges are considered. The coefficients are subject to large fluctuations if diversion are not considered. With diversion consideration a constant improvement over all years can be seen and the fluctuations of coefficients are significantly diminished.
For further examination of the performance of each runoff gauging station, a list with the MQ runoff difference between predictions and observations for each runoff gauging station was created, which can be seen in appendix B.

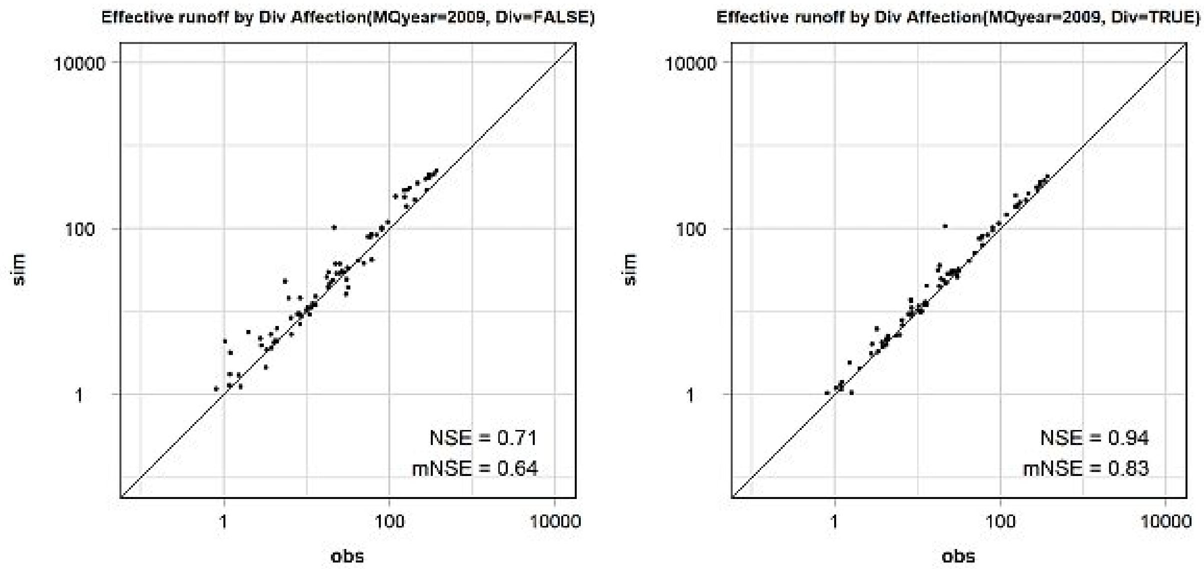
Tab. 4.3: Gauge runoff comparison model efficiency coefficient for years 2009-2017. TRUE/-FALSE refers to with/without diversion consideration. Distinguished between all and only diversion affected runoff gauging stations. Gauge count located in column N.

	All runoff gauging stations MQ_{eff}					Only diversion affected stations MQ_{eff}				
	NSE		mNSE			NSE		mNSE		
year	FALSE	TRUE	FALSE	TRUE	N	FALSE	TRUE	FALSE	TRUE	N
2009	0.91	0.93	0.81	0.86	408	0.71	0.94	0.64	0.83	80
2010	0.93	0.96	0.83	0.88	408	0.69	0.94	0.63	0.83	80
2011	0.93	0.97	0.82	0.88	410	0.63	0.93	0.59	0.83	81
2012	0.91	0.94	0.81	0.87	405	0.78	0.96	0.68	0.87	81
2013	0.95	0.97	0.84	0.89	404	0.76	0.96	0.67	0.86	81
2014	0.90	0.95	0.79	0.87	392	0.52	0.95	0.57	0.85	77
2015	0.90	0.96	0.78	0.88	385	0.48	0.94	0.54	0.83	76
2016	0.91	0.95	0.80	0.87	383	0.55	0.95	0.56	0.84	76
2017	0.91	0.95	0.80	0.87	382	0.54	0.94	0.56	0.83	76
mean	0.92	0.95	0.81	0.87	397	0.63	0.95	0.60	0.84	79



(a) Without diversion consideration (Div=FALSE) **(b)** With diversion consideration (Div=TRUE)

Fig. 4.13: Comparison of simulated to observed gauge runoff in m^3/s with model efficiency coefficient (NSE & mNSE) distinguished of all gauges for the year 2009. Runoff gauging stations are distinguished by diversion affection (Div_Bias) and diversion consideration (Div).



(a) Without diversion consideration (Div=FALSE) **(b)** With diversion consideration (Div=TRUE)

Fig. 4.14: Same figure as above but only diversion affected gauges (Div_Bias=TRUE) are displayed with the corresponding model efficiency coefficient (NSE & mNSE).

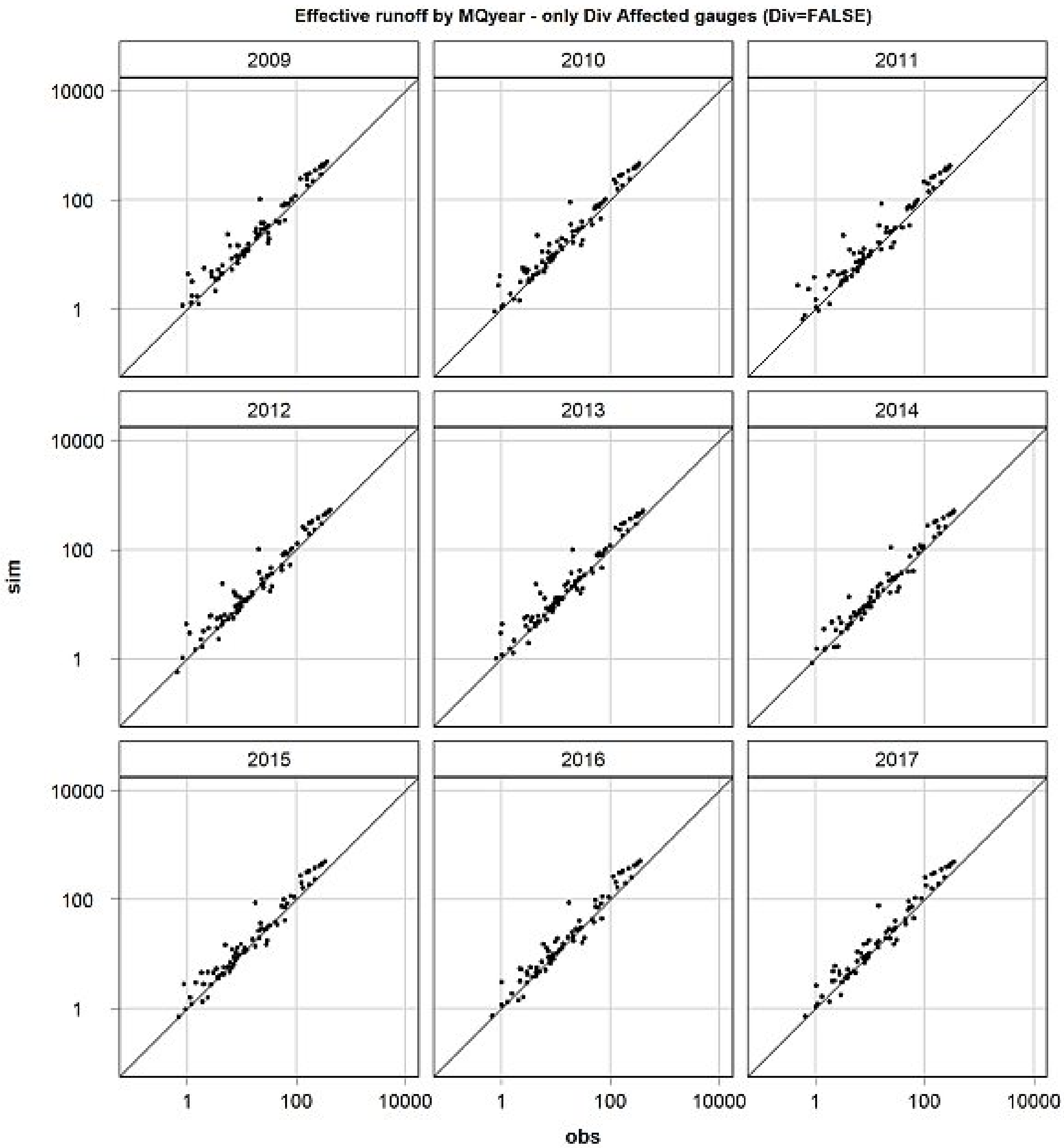


Fig. 4.15: Comparison of simulated to observed gauge runoff in m^3/s only for diversion affected gauges (Div_Bias=TRUE) and for all analysis years (2009-17) and Div=FALSE.

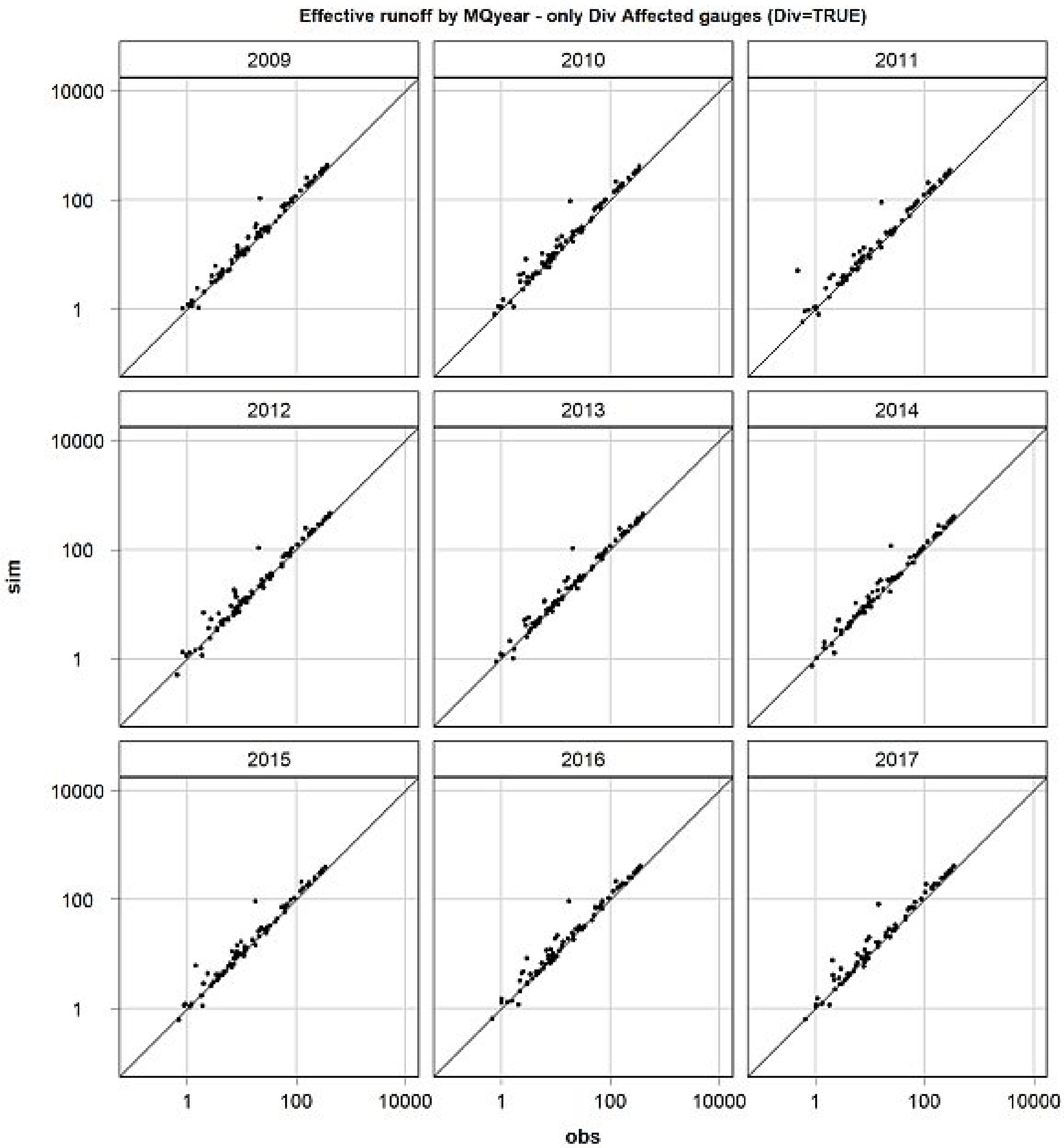
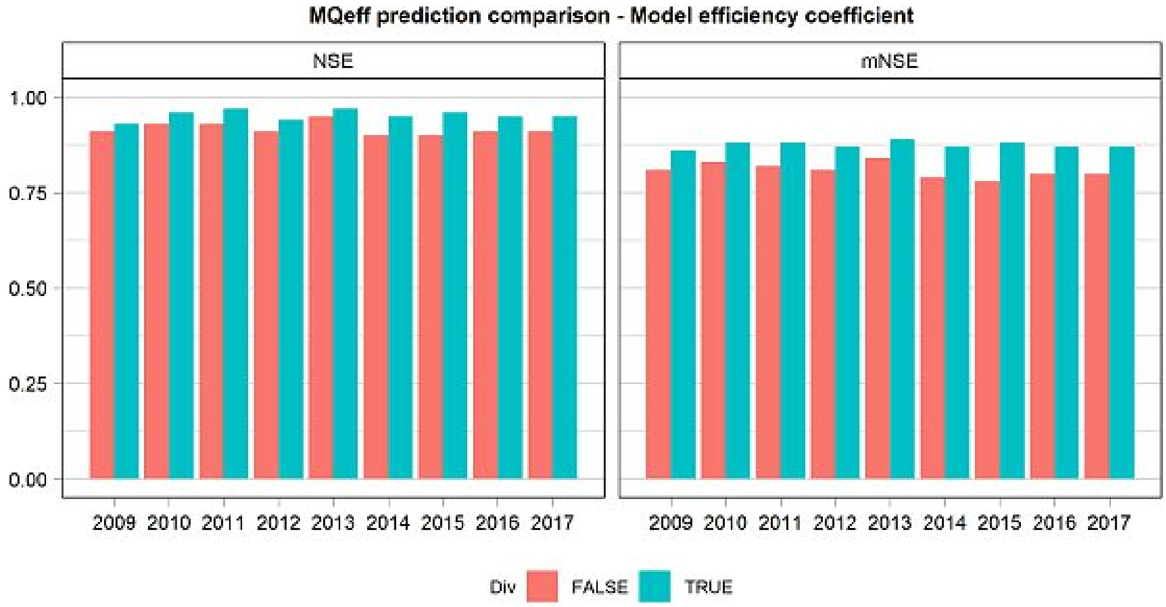
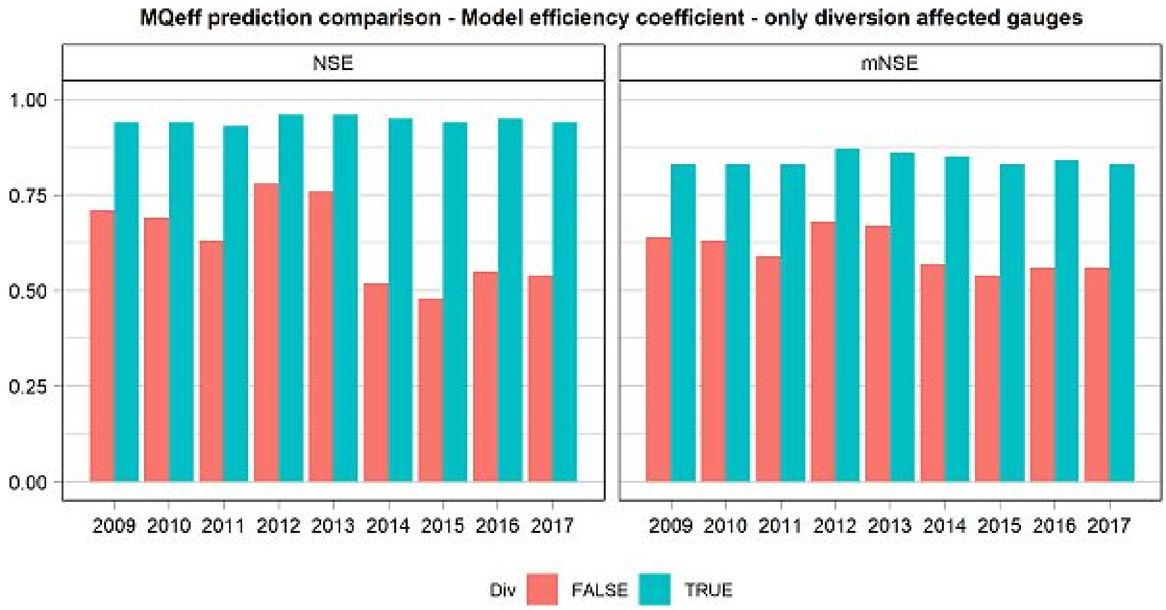


Fig. 4.16: Comparison of simulated to observed gauge runoff in m^3/s only for diversion affected gauges (Div_Bias=TRUE) and for all analysis years (2009-17) and Div=TRUE.



(a) All runoff gauging stations.



(b) Only diversion affected runoff gauging stations (DivBias=TRUE)

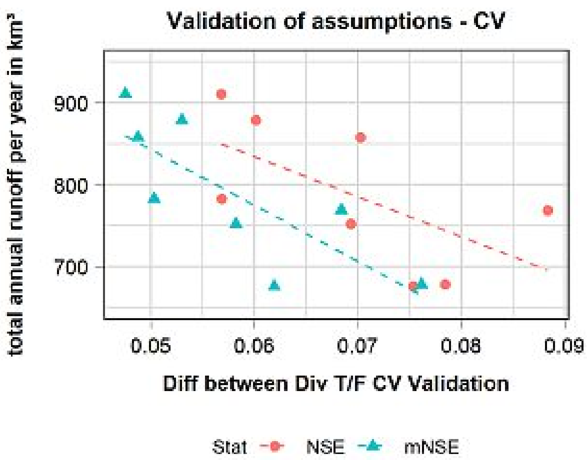
Fig. 4.17: Comparison of model efficiency coefficient (NSE & mNSE) per year for gauge runoff prediction distinguished by diversion consideration (Div=TRUE/FALSE) and between all and only diversion affected runoff gauging stations.

4.2.3 Validation of assumptions

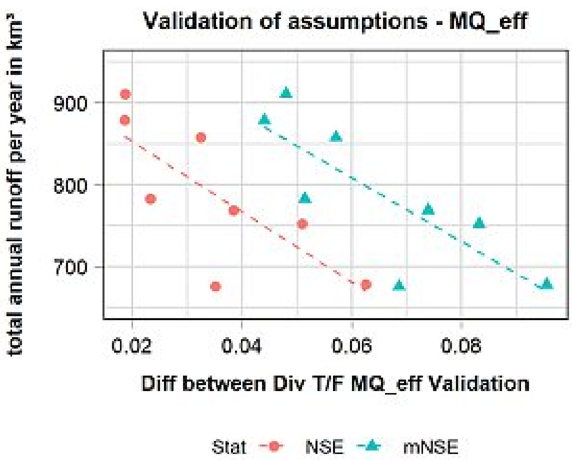
The total runoff per year varies between 911 km^3 in 2013 and 646 km^3 in 2017 as shown in figure 4.19. To validate the assumption, that watershed area represents the diversion runoff, model efficiency improvements (NSE and mNSE) of validation in each year and the total runoff were compared. The scatterplots (figures 4.18) show the total runoff per year in cubic kilometres over the model efficiency improvements of cross-validation (figure 4.18a) and of gauge runoff validation (figure 4.18b) for both NSE and mNSE. The model prediction efficiency improvements are calculated as the difference in model efficiency coefficient of validation between with (Div=TRUE) and without diversion consideration(Div=FALSE). The regression line shows evidence of a moderate, negative correlation between total annual runoff and model efficiency improvements. Due to the low number of samples (n=9) normality could not be assumed. Therefore, a Spearman's rank correlation has been performed on the data and results are shown in table 4.4. There is evidence that a negative correlation between total annual runoff and model efficiency improvements exists, as all p-values are below the level of significance ($\alpha = 0.05$).

Tab. 4.4: Spearman's rank correlation results of total annual runoff and model efficiency improvements distinguished by validation data and model efficiency coefficients.

	Observation cross-validation		Gauge runoff validation	
Coefficients	NSE	mNSE	NSE	mNSE
ρ	-0.717	-0.767	-0.783	-0.717
p	0.037	0.021	0.017	0.037



(a) For observation cross-validation.



(b) For gauge runoff (MQeff) validation.

Fig. 4.18: Relation between model efficiency improvements to the total annual runoff. *Stat* refers to the used indicator of model performance, hence NSE and mNSE.

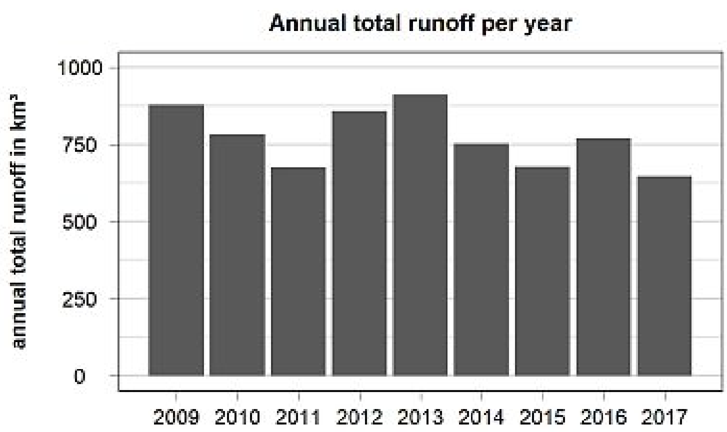


Fig. 4.19: Total annual runoff of all runoff gauging stations over the years.

Chapter 5

Discussion

Figure 3.1 shows the runoff gauging stations in the study area distinguished by diversion affection. Around 17% of them are diversion affected. This figure emphasizes the statement of Wesemann et al. [5] that a large part of the runoff gauging stations in the alpine region are diversion affected.

5.1 TopKriging Interpolation

The general improvements of diversion consideration for observed specific runoff in figure 4.1 can be seen by comparing 4.1a with 4.1b. Without diversion consideration the map in figure 4.1a shows a irregular picture with partially high changes in the specific runoff within neighbouring watersheds. For example the specific runoff in the Engadine valley in Switzerland, located at the left bottom corner, ranges between 111 mm/a to 1360 mm/a. With diversion consideration the specific runoff in figure 4.1b ranges from 624 mm/a to 1360 mm/a and the map is more regular.

The TopKriging predictions in figure 4.2 show the same picture. Through diversion consideration the pattern gets smoother and specially in the alpine headwaters. In the left bottom and middle of the map, the specific runoff values look more plausible comparing to the neighbouring watersheds. The east-west difference due to the diverse Austrian landscape is distinguishable. Diversion consideration also decreases the prediction error as seen in figure 4.3. The variance gets reduced by half a magnitude. Looking closely at the plot obviously the variance plot reflects the river network, hence the variance is lower for higher rank rivers and higher for headwaters. In evidence, Mur-Mürz valley in the south east, Drau river in the east, and Inn river in the mid west are clearly visible. This is due to the number of correlated observations influencing the prediction error, hence headwaters have less correlated observations and therefore the variance increases.

The predictions for the STOBIMO watersheds (MORE AUs) in figure 4.5 show also a smoother picture specially in the alpine regions. It should be mentioned that the upper Inn river in Switzerland and Ill river in Vorarlberg (far west) are strongly influenced by hydropower plants. Therefore without diversion consideration this strong alpine regions in figure 4.5a show a specific runoff typical for lowlands. With diversion consideration in figure 4.5a this gets corrected, hence the watersheds match their neighbours.

The process comparison in figure 4.6 shows that throughout the process steps the distribution of data does not significantly vary. Median, 1^{st} and 3^{rd} quartiles doesn't change much. Only the density changes for the process step *pred_TK*, which is due to the ten times higher number of values, hence watersheds. Interesting is that due to the very high number of watersheds and their high variability in step *pred_TK*, TopKriging prediction produces a higher number of outlies, whereas in step *pred_CV* and *STOBIMO* the number of outliers gets reduced, concluding a smoothing effect through the lower number of watershed and their greater size in terms of area.

5.2 Validation

5.2.1 Observations validated with cross-validation

The cross-validation shows that there is a general improvement in prediction efficiency with diversion consideration over all years (figure 4.9), according to Moriasi et al. [25] (table 3.6) the general performance of the CV predictions is slightly raised from satisfactory to good model efficiency. Looking at the individual predictions (figure 4.7) it can be seen that some watersheds get overestimated, for example, in the far east or at the lower Inn river in the north west. Through diversion consideration the variance (4.8) gets significantly reduced in some regions but also increases in others. However, overall reduction prevails. If only diversion affected gauges (DivBias=TRUE) are considered, the prediction efficiency can be raised from insufficient throughout all years from 0.40 to 0.73 NSE by considering diversions. This improvement is significant and therefore diversion affection influences the prediction efficiency and should not be neglected when their number is high in the study area.

5.2.2 Predictions validated with runoff comparison

5.2.2.1 Diversion runoff comparison

Diversion runoff comparison can obviously only be made for the case with diversion consideration (Div=TRUE). Comparing the prediction efficiency in table 4.2 with table 3.6 the coefficients indicate for NSE a very good and for mNSE a satisfactory prediction efficiency. Looking at the comparison of TopKriging predictions over all years in figure 4.10 it is noticeable that the predictions are also quite good for diversions whose diversion area (A_{Div}) is calculated with MQ values (labeled in the figure as *MQ_to_AREA*), specially for higher values. For some diversions there is very poor prediction for all years, some overestimated and some underestimated. Those outliers have to be further evaluated and maybe their diversion area has to be derived from their MQ values (See 3.4.1.1). Because their diversion area is taken from official statistic, they were not corrected in this thesis.

Converting the MQ runoff to specific runoff the comparison reveals a similar but still different view at the prediction comparison (figure 4.11). Three outliers from the previous discussed (figure 4.10) can also be seen in figure 4.11 but some diversions can only be seen in one of those two figures. The breakdown by diversion category (left side of table 4.2) shows that, beside of the discussed outliers, the predictions are reliable across the diversion categories except for category *Canal & Others*. Despite all diversion areas of this category were calculated out of observed MQ runoffs, their prediction efficiency was insufficient. This can be due to low sample number (33 observations of 4 diversions). Or the variability between years can not be represent by the approach with diversion areas. The predictions of pumpstorage diversions were consistently reliable, probably due to the fact that within the study area none known multi-annual-reservoir exists and annual time steps smoothen the effects of pumpstorage operation.

5.2.2.2 Gauge runoff comparison

The prediction efficiency of the general model (left side of table 4.3) is already very good [25] but can be improved throughout the years by diversion consideration, hence 0.03 for NSE and 0.06 for mNSE. Concerning this study area with around 17% diversion affected gauges, an improvement exists but is not significant. When considering only diversion affected gauges (right side of table 4.3), the improvements are significant. Without diversion consideration (Div=FALSE) the prediction efficiency ranges from unsatisfactory to good according to Moriasi et al. [25], whereas with diversion consideration (Div=TRUE) the prediction efficiency is constant on a very high level, nearly as good as the total model.

In the comparison of without (Div=FALSE) to with diversion consideration (Div=TRUE) in figure 4.14 the gauges with MQ between 100 – 500 m^3/s are obviously overestimated in case of without diversion consideration but with diversion consideration they perfectly align with the 1:1 line. This can be seen for all analyzed years, hence the approach worked as desired for those gauges and caused a significant improvement in prediction efficiency for those runoff gauging stations. The observable visual noise reduction through diversion consideration in both figures 4.13 & 4.14 can be seen in the higher difference of mNSE compared to difference of NSE, in average 0.05 and 0.02 respectively.

In figure 4.16 the runoff gauging station *Flattach* (HZBR-Nb.:213124) can be seen as outlier, hence the diversion consideration brought no improvements. This gauge is located on the Möll river, which is strongly influenced by large storage hydropower plants in its upper reaches. Their diversion area need further investigation.

5.2.3 Validation of assumptions

Statistical tests confirm and both plots of figure 4.19 show a negative linear correlation between model improvements by diversion consideration and total annual runoff. Therefore, validation of the assumptions indicate that the influence of diversions is higher in dry years than in wet ones. Furthermore, the approach with diversion areas may underestimate the overflow into the river bed via the spillway of river catchment stations. However, due to the limited sample size (n=9), further research must be conducted to validate this claim.

5.2.4 Cost of data collection

The author emphasizes that the workload to identify, locate and analyse the diversions is very high. Therefore, it has to be discussed if the approach in this thesis is worth the effort. This can be evaluated based on the share of diversion affected gauges in the study area or the intended results. For detailed local results or high share of diversion affected gauges, the effort can be valuable, as the thesis showed significant increased model accuracy. On the other hand, if the number of diversion affected gauges is low or only the general or total results of a model are of interest, this workload may be not worth the effort due to its small degree of general improvements. Maybe with other model parameters improvements can be achieved more labour efficient.

A suggestion would be to include diversion pathways and diversion areas for diversion affected gauges at the national level into the available GIS applications (e.g eHyd [15] in Austria). A barrier-free digital access in a central place with up-to-date data would tremendously decrease the expenditure on data collection, which may convince researchers to implement diversion consideration in their models.

Chapter 6

Conclusion & Outlook

6.1 Conclusion

This thesis shows that diversion consideration in the MoRE model yields a clear improvement in prediction performance of annual runoffs. Leave-one-out-cross validation analysis reveals that diversion consideration improves the model performance and that the approach with diversion areas is a valid and well suited method to consider diversions on an annual time scale. In conclusion:

- Prediction efficiency of TopKriging can be improved by diversion consideration. The NSE increases in average by 83% from 0.40 to 0.73 (NSE) for diversion affected gauges. Concerning the whole study area, with 17% diversion affected gauges, the general improvements are in average 11%, hence from 0.63 to 0.70 (table 4.1).
- Prediction efficiency of gauge runoff in the MoRE model can be improved by diversion consideration. The NSE increases in average by 51% from 0.63 to 0.95 (NSE) for diversion affected gauges. For all runoff gauging stations in the study area, the general improvements are in average 3%, hence from 0.92 to 0.95 (table 4.3).
- Diversion consideration can be obtained by a single variable, namely diversion area (A_{Div}). Which can be derived from official statistics, surface measurements or by calculations from observed diversion runoffs.
- Good results can be achieved on all diversion categories except for category *Canal & Others* (see figures 4.10 & 4.11). For this category further research is suggested as the sample size was relative small (n=33).
- The method of diversion area tends to underestimates the overflow into the river bed via the spillway of river catchment stations. However, due to the limited sample size (n=9), further research must be conducted to validate this claim.
- The cost of data collection was very high, which has to be considered before diversion consideration in projects.

6.2 Outlook

Improvements on the TopKriging input datasets can be made in the upper and lower Inn region, as the overlapping watersheds are not as fine as in the remaining study area. This could lead to better TopKringing predictions in those areas.

Some runoff gauging stations show very high specific MQ runoff which necessitates further examination. For some diversions, the observed runoffs do not meet the simulated runoffs by more than a magnitude, hence the provided official diversion area should be re-evaluated.

Unfortunately, no statement can be made about the application of the thesis approach with diversion area to models with quarterly, monthly or weekly time steps. This requires a separate investigation.

In general, the author wants to emphasise that if the diversion area is available with barrier-free digital access as up-to-date data, the expenditure on data collection would tremendously decrease. An easily accessible data, as it is currently available for runoff data should be the target.

Bibliography

[1] European Commission. *Directive 2000/60/EC of the European Parliament and of the Council establishing a framework for community action in the field of water policy*. URL: `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32000L0060` (last access 11/22/2020).

[2] European Commission. *Decision 2455/2001/EC of the European Parliament and of the Council establishing the list of priority substances in the field of water policy and amending Directive 2000/60/EC*. URL: `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32001D2455&qid=1606044092349` (last access 11/22/2020).

[3] A. Amann, M. Clara, O. Oliver Gabriel, G. Hochedlinger, M. Humer, F. Humer, S. Kittlaus, S. Kulcsar, C. Scheffknecht, H. Trautvetter, M. Zessner, and O. Zoboli. *STOBIMO Spurenstoffe: Stoffbilanzmodellierung für Spurenstoffe auf Einzugsgebietsebene*. 2019. URL: `https://doi.org/10.34726/s80j-4g54` (last access 10/07/2020).

[4] S. Fuchs, M. Kaiser, L. Kiemle, S. Kittlaus, S. Rothvoß, S. Toshovski, A. Wagner, R. Wander, T. Weber, and S. Ziegler. "Modeling of Regionalized Emissions (MoRE) into Water Bodies: An Open-Source River Basin Management System". In: *Water* 9.4 (2017), p. 239. DOI: `10.3390/w9040239`.

[5] J. Wesemann, H. Holzmann, K. Schulz, and M. Herrnegger. "Behandlung künstlicher Speicher und Überleitungen in der alpinen Niederschlags-Abfluss-Vorhersage". In: *Österreichische Wasser- und Abfallwirtschaft* 70.9-10 (2018), pp. 485–496. ISSN: 0945-358X. DOI: `10.1007/s00506-018-0501-9`.

[6] D. Egré and J. C. Milewski. "The diversity of hydropower projects". In: *Energy Policy* 30.14 (2002), pp. 1225–1230. ISSN: 03014215. DOI: `10.1016/S0301-4215(02)00083-6`.

[7] G. Blöschl. "Geostatistische Methoden bei der hydrologischen Regionalisierung: Methoden der hydrologischen Regionalisierung". In: *Wiener Mitteilungen* Band 197 (2006).

[8] L. D. Rizo-Decelis, E. Pardo-Igúzquiza, and B. Andreo. "Spatial prediction of water quality variables along a main river channel, in presence of pollution hotspots". In: *The Science of the total environment* 605-606 (2017), pp. 276–290. DOI: `10.1016/j.scitotenv.2017.06.145`.

[9] G. de Marsily. *Quantitative hydrogeology*. Paris School of Mines, Fontainebleau. London: Academic., 1986.

[10] J. O. Skøien, R. Merz, and G. Blöschl. "Top-kriging - geostatistics on stream networks". In: *Hydrol. Earth Syst. Sci.* 10 (2006), pp. 277–287.

[11] G. Laaha, J. O. Skøien, F. Nobilis, and G. Blöschl. "Spatial Prediction of Stream Temperatures Using Top-Kriging with an External Drift". In: *Environmental Modeling & Assessment* 18.6 (2013), pp. 671–683. ISSN: 1573-2967.

[12] J. O. Skøien, G. Blöschl, G. Laaha, E. Pebesma, J. Parajka, and A. Viglione. "Rtop: An R package for interpolation of data with a variable spatial support, with an example from river networks". In: *Computers & Geosciences* 67. (2014), pp. 180–190. ISSN: 00983004. DOI: `10.1016/j.cageo.2014.02.009`. URL: `https://CRAN.R-project.org/package=rtop`.

[13] Fontanie F. "Ermittlung des Wasserkraftpotenzials in Österreich". Diplomarbeit. 2010.

[14] J. Parajka, R. Merz, J. O. Skøien, and A. Viglione. "The role of station density for predicting daily runoff by top-kriging interpolation in Austria". In: *Journal of Hydrology and Hydromechanics* 63.3 (2015), pp. 228–234. DOI: 10.1515/johh-2015-0024.

[15] A. I. W. BMLRT. *Hydrographisches Jahrbuch von Österreich 2017: Hydrographischer Dienst in Österreich*. 2020. URL: https://www.bmlrt.gv.at/wasser/wasser-oesterreich/wasserkreislauf/hydrographische_daten/jahrbuecher/jahrbuch2017.html (last access 09/11/2020).

[16] BAFU. *Hydrologisches Jahrbuch der Schweiz: Abfluss, Wasserstand und Wasserqualität der Schweizer Gewässer*. 2020. URL: https://www.bafu.admin.ch/bafu/de/home/themen/wasser/publikationen-studien/publikationen-wasser/hydrologisches-jahrbuch-der-schweiz.html (last access 11/24/2020).

[17] G. D. B. LfU. *Gewässerkundlicher Jahresbericht 2019*. 2020. URL: https://www.lfu.bayern.de/wasser/gewaesserkundlicher_jahresbericht_2019/index.htm (last access 11/24/2020).

[18] S. W. BMLFUW IV. *Nationaler Gewässerbewirtschaftungsplan 2015*. 2017. URL: https://www.bmlrt.gv.at/wasser/wasser-oesterreich/wasserrecht_national/planung/NGP-2015.html (last access 11/24/2020).

[19] BMLRT. *Wasserinformationssystem Austria - WISA: Wasserkörper 2015*. 2020. URL: https://maps.wisa.bmlrt.gv.at/gewaesserbewirtschaftungsplan-2015# (last access 11/24/2020).

[20] Swiss Federal Office for the Environment. *Restwasserkarte Schweiz*. 2007. URL: https://map.geo.admin.ch (last access 08/15/2020).

[21] *LEITHA - Referenzzustand und Zielzustand WRRL*. Amt d. NÖ Landesregierung, Gr. Wasser – Abt. Wasserwirtschaft und Amt d. Bgld. Landesregierung, Abt. 9 – Wasser- und Abfallwirtschaft, Hauptreferat Gewässeraufsicht und Gewässerentwicklung, 2009.

[22] R Core Team. *R: A Language and Environment for Statistical Computing*. 2020. URL: https://www.R-project.org/ (last access 11/21/2020).

[23] M. Zambrano-Bigiarini. *hydroTSM: Time Series Management, Analysis and Interpolation for Hydrological Modelling*. 2020. DOI: 10.5281/zenodo.83964. URL: https://github.com/hzambran/hydroTSM.

[24] R. J. Hijmans. *raster: Geographic Data Analysis and Modeling*. 2020. URL: https://CRAN.R-project.org/package=raster (last access 09/30/2020).

[25] D. N. Moriasi, J. G. Arnold, M. W. van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith. "Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations". In: *Transactions of the ASABE* 50.3 (2007), pp. 885–900. DOI: 10.13031/2013.23153.

[26] M. Zambrano-Bigiarini. *hydroGOF: Goodness-of-Fit Functions for Comparison of Simulated and Observed Hydrological Time Series*. 2020. DOI: 10.5281/zenodo.839854. URL: https://github.com/hzambran/hydroGOF (last access 10/01/2020).

[27] E. J. Pebesma and R. S. Bivand. "Classes and methods for spatial data in R". In: *R News* 5.2 (2005), pp. 9–13. URL: https://CRAN.R-project.org/doc/Rnews/.

[28] R. S. Bivand, E. Pebesma, and Virgilio. *Applied spatial data analysis with R*. Second edition. Springer, NY, 2013. URL: https://asdar-book.org/.

[29] E. Pebesma. “Simple Features for R: Standardized Support for Spatial Vector Data”. In: *The R Journal* 10 (2018), pp. 439–446. DOI: 10.32614/RJ-2018-009. URL: https://doi.org/10.32614/RJ-2018-009.

[30] J. O. Skøien. *Package ‘rtop’: Interpolation of Data with Variable Spatial Support*. 2018. URL: https://cran.r-project.org/web/packages/rtop/rtop.pdf (last access 11/03/2020).

[31] M. Dowle and A. Srinivasan. *data.table: Extension of ’data.frame’*. 2020. URL: https://CRAN.R-project.org/package=data.table (last access 09/30/2020).

[32] M. Tennekes. “tmap : Thematic Maps in R: Journal of Statistical Software, 84(6)”. In: (2018). DOI: 10.18637/JSS.V084.I06.

[33] H. Wickham. *ggplot2: Elegant graphics for data analysis*. Use R! Cham: Springer, 2016. ISBN: 9783319242774. DOI: 10.1007/978-3-319-24277-4.

List of Figures

List of Tables

Acronyms

A2A A2A S.p.A. (Italian public utility company)

AU analytical unit

BMLRT Austrian Federal Ministry of Agriculture, Regions and Tourism

diversion trans-catchment-diversion

eHyd Internet portal for hydrographic data of Austria

EKW Engadiner Kraftwerke AG

EWR Elektrizitätswerke Reutte AG

FOEN Swiss Federal Office for the Environment

GIS Geographic information system

illwerke vkw illwerke vkw AG

IWAG Institute for Water Quality and Resource Management

KELAG Kärntner Elektrizitäts-Aktiengesellschaft

LKW Liechtensteinische Kraftwerke

mNSE Modified Nash-Sutcliffe efficiency

MoRE Modeling of Regionalized Emissions

NSE Nash-Sutcliffe efficiency

NWP National water management plan

OMS OpenStreetMap

RTK Regression TopKriging

SF splitting factor

SWM Stadtwerke München GmbH

TIWAG Tiroler Wasserkraft AG

TK TopKriging

TKED TopKriging with external drift

WFD EU Water Framework Directive

Wiener Wasser City of Vienna - MA 31 - Wiener Wasser

ÖBB Österreichische Bundesbahnen

Translations

hydropower plants - Wasserkraftwerk

study area - Untersuchungsgebiet

watershed & catchment - Einzugsgebiet (EZG)

headwaters - Oberläufe, Quellgebiete

orographic watershed area (A_{oro}) - Fläche des orographischen Einzugsgebietes

effective watershed area in (A_{eff}) - Fläche des wirksamen Einzugsgebietes

river gauge runoff (Q_{gauge}) - Abfluss gemessen bei einer Messstelle (Pegel)

trans-catchment-diversion (short: diversion) - Überleitung in ein benachbartes Eizugsgebiet

diversion watershed area (short: diversion area) (A_{Div}) - Einzugsgebietfläche der Überleitung

Inlet diversion watershed area ($A_{Div.Inlet}$) - Überleitungsfläche in das EZG (Zuleitung)

Outlet diversion watershed area ($A_{Div.Outlet}$) - Überleitungsfläche aus dem EZG (Ableitung)

diversion amount (Q_{Div}) - Überleitungsmenge

hierarchical structure of rivers from source to mouth (the so-called *flow tree*) - hierarchische Gliederung der Teileinzugsgebiete (sog. *Abflussbaum*)

specific runoff - Abflussspende

net runoff (runoff generated in a singular AU) - Nettoabfluss (Abluss eines einzelnen AU)

natural specific runoff (q_{nat}) - natürliche Abflussspende

effective specific runoff (q_{eff}) - effektive Abflussspende

legally required quantity of residual water - behördlich vorgeschriebe Restwassermenge

water catchment stations - Wasserfassungen von Kraftwerken

minimum flow requirements - Restwassermengen

annual mean discharge (MQ) - Mittlerer Abfluss

gauge - Messstelle an einem Fluss

runoff gauging station (only Q)- Abflussmessstelle

undisturbed runoff gauging station - unbeeinflusste Abflussmessstelle

disturbed runoff gauging station - durch Überleitungen beeinflusste Abflussmessstelle

EU Water Framework Directive - EU Wasserrahmenrichtlinie (WRRL)

national water management plan - Nationaler Gewässerbewirtschaftungsplan (NGP)

Austrian Federal Ministry of Acriculture, Regions and Tourism - Österreichisches Bundesministerium für Landwirtschaft, Regionen und Tourismus (BMLRT)

Appendix A

Diversion Areas

A.1 Diversion areas

Here the used Diversion data can be seen:

Diversion areas of runoff gauge stations

Tab. A.1: Diversion area table for diversion affected runoff gauge stations.

$HZBR_{NR}$	Country/State	$A_{oro}[km^2]$	$A_{Div.Inlet}[km^2]$	$A_{Div.Outlet}[km^2]$	$A_{eff}[km^2]$
200014	Vorarlberg	4647.9	25.5	0.0	4673.4
200022	Vorarlberg	144.9	184.7	316.3	13.3
200030	Vorarlberg	39.3	0.0	20.9	18.4
200055	Vorarlberg	511.5	176.3	673.2	14.6
231662	Vorarlberg	535.2	176.3	587.0	124.5
200097	Vorarlberg	72.2	0.0	12.5	59.7
231670	Vorarlberg	854.6	170.0	673.2	351.4
200592	Vorarlberg	66.6	2.7	0.0	69.3
231688	Vorarlberg	1118.6	170.0	655.2	633.4
200501	Vorarlberg	70.0	0.0	25.5	44.5
200121	Vorarlberg	1268.7	170.0	25.5	1413.2
200147	Vorarlberg	1281.0	170.0	25.5	1425.5
200196	Vorarlberg	6301.1	170.0	0.0	6471.1
200295	Vorarlberg	111.6	0.0	97.6	14.0
200451	Vorarlberg	81.5	97.6	0.0	179.1
200303	Vorarlberg	199.3	97.6	187.0	109.9
200337	Vorarlberg	10907.0	170.0	0.0	11077.0
200360	Vorarlberg	25.2	0.0	1.8	23.4
200378	Vorarlberg	84.3	0.0	6.2	78.1
201012	Tyrol	247.9	0.0	6.2	241.7
201038	Tyrol	459.8	0.0	6.2	453.6
201087	Tyrol	1012.2	0.0	81.2	931.0
202184	Tyrol	115.5	75.0	0.0	190.5
201145	Tyrol	105.5	112.7	218.0	0.2
201178	Tyrol	2162.0	0.0	105.0	2057.0
201194	Tyrol	2461.5	237.4	105.0	2593.9
201277	Tyrol	271.3	0.0	33.8	237.5
201251	Tyrol	130.6	0.0	33.8	96.8
201210	Tyrol	97.6	0.0	66.2	31.4
201236	Tyrol	385.4	0.0	130.0	255.4
202036	Tyrol	727.0	0.0	163.8	563.2
201319	Tyrol	3842.0	274.5	268.8	3847.7
201335	Tyrol	165.4	0.0	87.7	77.7
201434	Tyrol	785.5	0.0	25.8	759.7
230342	Tyrol	890.0	0.0	53.3	836.7
201459	Tyrol	5118.8	0.0	322.1	4796.7
230078	Tyrol	5289.7	137.3	268.8	5158.2
201475	Tyrol	5420.2	137.3	268.8	5288.7
202069	Tyrol	52.0	0.0	29.3	22.7

Continued on next page

Tab. A.1: Diversion area table for diversion affected runoff gauge stations (continuation).

$HZBR_{NR}$	Country/State	$A_{oro}[km^2]$	$A_{Div.Inlet}[km^2]$	$A_{Div.Outlet}[km^2]$	$A_{eff}[km^2]$
202077	Tyrol	64.4	0.0	31.0	33.4
230706	Tyrol	204.7	0.0	60.3	144.4
201525	Tyrol	5771.6	23.7	268.8	5526.5
230714	Tyrol	280.9	0.0	23.7	257.2
201624	Tyrol	854.4	0.0	23.7	830.7
201673	Tyrol	7129.8	0.0	163.8	6966.0
201681	Tyrol	7230.7	218.0	268.8	7179.9
201699	Tyrol	135.3	0.0	113.8	21.5
201715	Tyrol	225.0	0.0	194.5	30.5
201723	Tyrol	129.2	0.0	13.3	115.9
201749	Tyrol	610.9	10.0	0.0	620.9
201756	Tyrol	696.3	10.0	0.0	706.3
201970	Tyrol	141.0	20.4	0.0	161.4
201772	Tyrol	196.8	30.7	10.0	217.5
201780	Tyrol	1094.7	30.7	0.0	1125.4
201806	Tyrol	8503.6	248.7	268.8	8483.5
201814	Tyrol	8508.7	248.7	268.8	8488.6
201889	Tyrol	9310.0	248.7	268.8	9289.9
201897	Tyrol	94.0	0.0	2.3	91.7
201905	Tyrol	9502.7	248.7	268.8	9482.6
18246006	Germany	202.7	0.0	170.3	32.4
18203003	Germany	381.7	0.0	282.4	99.3
18204006	Germany	754.0	0.0	452.7	301.3
18000403	Germany	9713.2	248.7	268.8	9693.1
18001508	Germany	10153.5	249.7	268.8	10134.4
18003004	Germany	11960.4	249.7	268.8	11941.3
18004007	Germany	12253.6	249.7	268.8	12234.5
18004506	Germany	12385.7	249.7	268.8	12366.6
18005000	Germany	13320.5	249.7	268.8	13301.4
18005019	Germany	13320.5	249.7	268.8	13301.4
18005702	Germany	22571.0	250.7	268.8	22552.9
18007209	Germany	25520.0	251.7	268.8	25502.9
18007800	Germany	26037.9	252.7	268.8	26021.8
18008008	Germany	26040.0	252.7	268.8	26023.9
203026	Salzburg	206.8	0.0	30.7	176.1
203075	Salzburg	582.6	0.0	30.7	551.9
203083	Salzburg	74.5	0.0	2.3	72.2
203554	Salzburg	127.9	14.4	0.0	142.3
203109	Salzburg	88.6	92.5	0.0	181.1
203125	Salzburg	1168.7	92.5	30.7	1230.5
203133	Salzburg	60.7	0.0	8.3	52.4
203141	Salzburg	161.0	0.0	16.4	144.6
203158	Salzburg	96.1	0.0	21.0	75.1
203166	Salzburg	242.2	0.0	21.0	221.2

Continued on next page

Tab. A.1: Diversion area table for diversion affected runoff gauge stations (continuation).

$HZBR_{NR}$	Country/State	$A_{oro}[km^2]$	$A_{Div.Inlet}[km^2]$	$A_{Div.Outlet}[km^2]$	$A_{eff}[km^2]$
203851	Salzburg	57.4	21.0	0.0	78.4
203208	Salzburg	220.7	21.0	0.0	241.7
203968	Salzburg	2143.0	76.1	30.7	2188.4
203323	Salzburg	3555.7	76.1	30.7	3601.1
204297	Salzburg	4425.7	76.1	30.7	4471.1
203539	Salzburg	6120.0	76.1	30.7	6165.4
206847	Salzburg	6690.5	76.1	30.7	6735.9
210732	Styria	194.3	0.0	107.9	86.4
210898	Styria	594.8	0.0	58.8	536.0
210864	Styria	280.3	0.0	6.4	273.9
211227	Styria	230.1	0.0	3.9	226.2
211243	Styria	726.8	0.0	3.9	723.0
205757	Upper Austria	5010.3	0.0	58.8	4951.6
208710	Lower Austria	469.2	0.0	104.2	365.1
208736	Lower Austria	89.5	0.0	12.4	77.1
208744	Lower Austria	113.5	0.0	12.4	101.1
208785	Lower Austria	704.7	0.0	505.3	199.4
208884	Lower Austria	1201.6	0.0	956.9	244.6
209288	Lower Austria	1242.2	0.0	1009.2	233.0
214031	Lower Austria	1598.9	0.0	292.6	1306.3
209007	Lower Austria	1982.0	0.0	292.6	1689.4
210013	Lower Austria	2131.3	0.0	292.6	1838.7
208157	Lower Austria	1028.9	180.3	0.0	1209.2
208199	Lower Austria	112.8	92.3	0.0	205.1
212092	Tyrol	518.4	0.0	12.1	506.3
212316	Tyrol	1876.8	0.0	12.1	1864.7
212324	Carinthia	2112.0	0.0	12.1	2099.9
212357	Carinthia	2561.4	0.0	107.5	2453.9
213926	Carinthia	142.3	0.0	76.8	65.5
212373	Carinthia	412.1	0.0	130.3	281.8
213124	Carinthia	705.3	23.2	64.0	664.5
212399	Carinthia	1043.8	23.2	139.3	927.7
213199	Carinthia	3674.4	130.3	76.1	3728.6
213207	Carinthia	56.5	0.0	11.3	45.2
212431	Carinthia	360.1	0.0	11.3	348.8
212472	Carinthia	131.3	0.0	93.3	38.0
212498	Carinthia	266.0	0.0	117.3	148.7
212530	Carinthia	1035.5	0.0	130.3	905.2
213215	Carinthia	4789.6	0.0	76.1	4713.5
213173	Carinthia	11043.9	0.0	76.1	10967.8
2403	Switzerland	733.0	0.0	604.2	128.9
2239	Switzerland	295.0	0.0	246.6	48.5
2265	Switzerland	1581.0	0.0	840.3	740.7
2067	Switzerland	1941.0	0.0	105.0	1836.0

Appendix B

Validation results

B.1 Runoff gauging stations MQ runoff difference

Here a list MQ runoff difference between predicted and observed MQ runoff per analysis year (2009-17) for each runoff gauging station from sub-subsection *Validation: Comparison simulated with observed runoff values* in 3.4.4.2 is shown.

Runoff gauging stations MQ runoff difference table

Tab. B.1: Runoff gauging stations with their MQ runoff difference between predicted and observed per analysis year.

	MQ runoff difference (obs-sim) in %								
$HZBR_{NR}$	2009	2010	2011	2012	2013	2014	2015	2016	2017
2067	-36	-33	-29	-33	-28				
2105	-1	1	4	1	1				
2239	-15	-18	-12	-21	-16				
2256	2	2	3	3	2				
2262	0	0	1	1	0				
2263	0	-2	-7	-5	-4				
2265	2	-7	-15	-1	-8				
2304	-8	-7	-8	-8	-8				
2403	8	2	-23	-6	-6				
2462	-9	-10	-11	-10	-10				
200048	-6	-4	-4	-3	1	-5	-8	-10	-3
200147	-3	-4	5	1	4	7	2	4	-12
200204	10	13	15	6	9	8	4	9	10
200253	-5	3	-8	-3	-5	2	-4	-7	0
200311	22	22	22	22	22	22	22	23	22
200329	2	2	2	-1	3	2	2	2	5
200378	-3	1	3	1	1	2	3	2	2
200410	-36	-35	-39	-34	-31	-34	-36	-36	-34
200501	34	34	31	38	39	43	41	41	35
200592	1	0	0	-1	3	4	1	3	2
201012	-3	1	-4	2	-2	-2	-4	-5	1
201087	4	6	-3	11	4	5	6	9	3
201095	1	-3	-7	-2	-1	-2	-7	-4	1
201111	-6	-3	-6	-3	-3	-6	-7	0	0
201160	-7	-9	-18	-15	-16	-11	-6	-12	-12
201178	-33	-30	-28	-32	-26	-16	-27	-38	-36
201194	-27	-25	-27	-27	-23	-17	-24	-33	-33
201210	5	9	5	12	14	2	2	9	6
201236	-18	-15	-19	-10	-15	-13	-10	-15	-12
201251	4	6	6	7	6	8	7	6	3
201277	-11	-13	-16	-12	-14	-21			
201319	-24	-21	-25	-22	-22	-22	-21	-27	-30
201335	-12	-8	-10	-6	-3	-7	-1	1	-3
201350	9	8	5	10	7	6	7	5	7
201376	2	2	2	3	4	3	3	4	0

Continued on next page

Tab. B.1: Runoff gauging stations with their MQ runoff difference (continuation).

	MQ runoff difference (obs-sim) in %								
$HZBR_{NR}$	2009	2010	2011	2012	2013	2014	2015	2016	2017
201392	-5	-9	-10	-11	-8	-6	-4	-8	1
201434	-7	-11	-8	-13	-10	-5	-7	-12	-2
201459	-20	-20	-19	-15	-17	-17	-16	-21	-23
201525	-18	-21	-17	-16	-17	-19	-19	-19	-24
201558	-8	-4	-5	-4	-5	-9	-10	-8	-10
201566	0	4	2	0	3	4	0	2	3
201574	-13	-6	-4	-9	-2	-3	-9	-7	-2
201582	-8	-12	-11	-7	-10	-10	-12	-10	-8
201624	-11	-11	-7	-10	-6	-8	-8	-7	-1
201640	-30	-10	-12	-18	-7	-4	-18	-21	3
201665	-5	-2	-4	-3	-4	-6	-6	-3	-6
201681	-19	-19	-16	-14	-15	-16	-17	-16	-18
201723	-9	-12	-11	-12	-11	-6	-10	-9	-11
201749	10	13	8	9	5	2	14	8	1
201756	4	7	4	6	2	-1	7	5	2
201772	-11	-6	-11	-4	-5	-7	-13	-8	-7
201780	-2	-1	-5	-2	-5	-9	0	-2	-7
201806	-15	-15	-14	-11	-11	-14	-14	-12	-16
201822	4	3	3	8	8	6	4	6	-1
201848	-2	-5	-6	-5	-5	-3	-8	-4	-3
201863	-10	-16	-8	-6	-7	-9	-6	-1	-6
201889	-9	-10	-10	-5	-5	-8	-10	-7	-10
201897	-41	-44	-54	-48	-44	-44	-45	-48	-52
201913	-5	-6	-6	-7	-4	-6	-8	-3	-7
201921	-11	-16	-17	-12	-8	-15	-13	-10	-5
201947	5	3	5	4	2	2	6	6	5
201970	-4	0	-16	-3	1	-1	-2	0	-10
202036	-11	-7	-10	-5	-4	-7	-2	-8	-10
202044	-11	-1	-5	-2	1	-1			
202101	-12	-10	-15	-14	-17	-19	-6	-14	-9
202127	-4	-5	-6	-5	-5	-5	-5	-5	-4
202218	0	1	-2	1	2	1	1	0	0
202283	7	9	7	8	8	7	11	10	10
202382	1	-2	-1	-2	-4	1	-3	-1	1
202523	-7	-27	-9	-26	-14	-13	-24	-6	-26
202549	10	9	11	8	9	8	9	9	10
202622	4	5	2	-9	-8	-8	-8	0	1
203026	-10	-9	-5	-8	-6	-9	-16	-13	-13
203034	-10	-8	-2	-5	-4	2	-14	-10	-15
203075	-16	-15	-12	-9	-14	-18	-30	-24	-28
203109	12	7	13	13	11	7	15	13	10
203125	-27	-29	-36	-27	-25	-32	-34	-30	-28
203133	3	5	5	5	5	2	2	4	6

Continued on next page

Tab. B.1: Runoff gauging stations with their MQ runoff difference (continuation).

	MQ runoff difference (obs-sim) in %								
$HZBR_{NR}$	2009	2010	2011	2012	2013	2014	2015	2016	2017
203141	-20	-25	-21	-12	-17	-14	-11	-23	-20
203158	6	3	3	8	6	3	4	4	4
203166	-13	-16	-20	-10	-10	-16	-15	-16	-16
203208	-2	0	2	4	5	4	3	1	-7
203224	1	-1	-2	4	0	1	4	6	2
203232	4	8	2	3	4				
203257	-22	-32	-38	-24	-21	-24			
203265	-10	-17	-21	-9	-16	-29	-23	-33	-27
203307	6	6	2	7	8	6	2	4	5
203315	7	8	7	7	8	7	8	7	8
203323	-13	-13	-17	-10	-9	-13	-22	-21	-16
203349	1	-5	-4	-2	2				
203364	5	6	7	7	4	4	5	11	5
203455	9	14	13	13	5	7	1	7	1
203463	-4	-1	-5	-3	-2	-3	-4	-5	-1
203471	4	-1	-6	-1	2	-3	-5	-2	-5
203489	-23	-28	-23	-24	-23	-31	-34	-23	-23
203497	-1	6	3	-1	1	1	1		
203505	-8	-8	-8	-8	-7	-7	-7	-9	-8
203521	0	3	3	-1	2	-3	-5	-3	-2
203539	-2	-4	-3	-2	1	-4	-8	-4	-8
203554	-12	-6	-17	-5	-4	-11	2	-4	-10
203596	-6	-5	-4	-3	1	-1	-1	1	1
203737	6	8	7	6	8	4	7	6	4
203745	14	16	19	12	16	13	17	16	14
203760	3	2	4	1	-1	3	1	3	3
203778	-20	-22	-19	-20	-22	-18	-20	-20	-13
203786	-13	-11	-10	-15	-9	-9	-14	-7	-10
203794	-13	-15	-22	-14	-12	-12	-10	-16	-12
203810	-1	-3	-2	2	-1	-2	0	-3	-7
203828	-15	-19	-25	-18	-13	-13	-16	-18	-15
203844	-1	-2	-2	0	1	2	1	2	0
203919	-22	-29	-34	-23	-22	-29	-27	-30	-28
203968	-20	-22	-28	-21	-17	-19	-19	-16	-17
203976	-15	-15	-12	-11	-9	-10	-14	-11	-14
204008	-13	-13	-14	-12	-13	-13	-12	-13	-13
204032	-19	-19	-22	-14	-7	-16			
204057	10	11	13	14	11	11	10	12	11
204297	-9	-9	-12	-5	-6	-5	-9	-4	-4
204545	-24	-28	-32	-22	-27	-32	-42	-16	-22
204586	3	5	4	4	4	4	-1	5	3
204677	-47	-38	-48	-52	-40	-45	-38	-43	-51
204701	-69	-86	-82	-86	-88	-85	-84	-50	-71

Continued on next page

Tab. B.1: Runoff gauging stations with their MQ runoff difference (continuation).

	MQ runoff difference (obs-sim) in %								
$HZBR_{NR}$	2009	2010	2011	2012	2013	2014	2015	2016	2017
204719	-5	4	1	-5	-9	-3	-9	1	-5
204735	-5	-3	-7	-4	-5	-3	-3	0	0
204768	-5	-8	-8	-7	-4	-7	-10	-7	-11
204776			-13						
204784	-28	-31	-31	-31	-45	-39	-41	-37	-32
204834	0	0	1	1	0	0	0	1	-2
204867	-17	-16	-15	-19	-11	-14	-19	-15	-20
204883	2	-1	-3	-2	-2	-11	-3	-10	0
204925	-3	-2	0	-2	-2	-3	-1	-2	-2
204933	-1	-1	1	0	-4	-11	-4	2	-2
204974	-5	-11	-9	-2	-8	-3	0	-7	-1
205021	-11	-22	-20	-27	-27	-23	-22	-27	-20
205054	-12	-4	-7	-6	-13	-15	-5	-10	-4
205104	5	6	7	6	8	6	7	3	4
205146	-3	0	-2	2	1	2	-5	1	-5
205153	0	1	2	2	3	-2	-8	-4	-6
205179	-1	-2	-2	-2	-2	1	0	-1	0
205187	1	-7	-4	0	0	5	0	1	2
205278	1	2	2	2	1	0	2	2	3
205369	-14	-9	-19	-13	-12	-11	-10	-7	-22
205377	18	17	24	16	18	18	21	16	15
205419	-10	-15	-15	-14	-14	-7	-10	-14	-4
205435	-7	-5	-8	-6	-4	-6	-8	-6	-7
205450	-6	-7	-8	-15	-15	-12	-13	-15	-14
205468	-7	-6	-10	-8	-6	-5	-8	-8	-6
205500	-22	-18	-25	-15	-18	-19	-30	-24	-18
205534	-26	-23	-23	-27	-22	-24	-21	-32	-28
205633	-17	-15	-16	-19	-14	-18	-11	-20	-23
205658	-16	-23	-30	-36	-19	-22	-25	-27	-30
205732	-15	-18	-19	-16	-13	-17	-16	-16	-13
205740	4	5	-7	0	1	-1	-2	-2	3
205781	-12	-12	-7	-11	-12	-11	-7	-12	-8
205799	-27	-40	-51	-42	-43	-44	-47	-45	-32
205823	-24	-33	-34	-28	-20	-32	-34	-27	-26
205831	-35	-45	-46	-41	-33	-36	-47	-35	-29
205864	-12	-21	-24	-19	-20	-26	-26	-20	-14
205898	-16	-18	-22	-16	-15	-20	-19	-19	-19
205914	-20	-26	-30	-24	-24	-27	-30	-34	-28
205922	-6	-5	-13	-5	-5	-7	-7	-7	-7
205948	-5	-5	-8	-5	1	-1	-7	-4	-5
205971	-16	-14	-9	-15	-10	-30	-20	-2	-11
205997	4	5	8	9	4	9	8	9	6
206037	-21	-21	-32	-18	-14	-27	-30	-22	-29

Continued on next page

Tab. B.1: Runoff gauging stations with their MQ runoff difference (continuation).

	MQ runoff difference (obs-sim) in %								
$HZBR_{NR}$	2009	2010	2011	2012	2013	2014	2015	2016	2017
206078	-145	-141	-148	-159	-152	-117	-119	-131	-135
206169	-6	-4	3	-2	-1	-3	-1	0	-1
206185	10	7	13	12	12	17	9	8	11
206326	-17	-9	-5	-8	-4	-4	-4	-6	-7
206383	-47	-56	-76	-53	-54	-60	-64	-58	-41
206409	-7	-6	-11	-12	-10	-11	-12	-12	-8
206573	-17	-10	-19	-11	-16	-29	-16	-20	-17
206581	-11	-9	2	4	-14	-2	7	-6	-9
206656	-73	-75	-70	-70	-65	-71	-67	-68	-64
206680	14	12	17	15	13	13	12	12	14
206730	-71	-82	-71	-86	-79	-63	-83	-77	-67
206763	-4	1	2	3	2	4	0	4	3
206771	-6	-3	-3	-3	-3	-6	-3	-2	-3
206805	-167	-143	-151	-169	-145	-157	-158	-142	-147
206813	-26	-30	-38	-41	-32	-38	-33	-30	-34
206839	-43	-34	-39	-49	-46	-22	-26	-39	-49
207035	-28	-22	-19	-26	-19	-26	-22	-24	-23
207134	-25	-20	-18	-22	-17	-21	-19	-21	-21
207274	-26	-19	-18	-25	-18	-21	-19	-22	-23
207357	-27	-20	-17	-25	-19	-21	-19	-21	-22
207373	-25	-20	-18	-25	-18	-20	-20	-22	-22
207613	-1	-2	1	0	1	-2	-6	-3	1
207654	4	4	2	3	6	4	2	7	2
207688	-4	-4	-7	-5	-1	-5	-12	-7	-12
207696	-8	-10	-8	-7	-5	-9	-9	-9	-9
207795	-12	-15	-22	-25	-12	-20	-17	-17	-19
207803	-9	-9	-13	-20	-10	-10	-13	-6	-15
207811	-24	-28	-28	-31	-25	-35	-30	-25	-33
207837	-38	-33	-47	-50	-44	-48	-52	-37	-52
207852	-21	-24	-33	-29	-19	-17	-21	-23	-37
207860	-28	-26	-45	-43	-25	-35	-39	-29	-43
207894	-2	0	-5	-12	-4	4	5	4	1
207902	6	8	5	-2	5	7	3	5	2
207910	-8	-3	-14	-23	-6	-12	-10	-3	-14
207936	-12	-10	-21	-20	-9	-14	-21	-11	-13
207944	-7	-7	-17	-15	-8	-12	-10	0	-15
207951	1	0	-1	-2	-1	-2	-2	-4	-2
207985	-13	-12	-32	-49	-27	-25	-11	-12	-36
207993	-21	-13	-25	-48	-29	-31	-12	-17	-37
208009	-29	-29	-39	-55	-41	-47	-46	-40	-52
208017	-6	1	-12	-30	-11	-19	-21	-14	-23
208041	-26	-9	-9	-29	-36	-15	-25	-20	-24
208058	-11	-8	-3	7	0	2	-3	0	-1

Continued on next page

Tab. B.1: Runoff gauging stations with their MQ runoff difference (continuation).

	MQ runoff difference (obs-sim) in %								
$HZBR_{NR}$	2009	2010	2011	2012	2013	2014	2015	2016	2017
208082	-19	-24	-14	-15	-10	-17	-8	-9	-16
208090	-6	-7	-7	-6	-2	-5	-6	-13	-6
208116	-3	-12	-7	-10	-6	-5	-4	-13	-4
208124	-11	-17	-11	-13	-17	-17	-13	-21	-18
208157	-57	-34	-21	-48	-55	-47	-39	-49	-49
208199	-26	-7	1	23	-12	14	10	5	4
208272	-2	1	9	4	-1	-3	3	2	5
208439	25	37	41	32	25	31	35	32	25
208462	-117	-130	-131	-121	-111	-103	-140	-146	-122
208520	-26	-18	-22	-25	-31	-23	-27	-33	-30
208579	-16	-5	3	6	1	5	5	-10	4
208611	-3	1	-9	-1	-1	-13	-12	-24	-13
208637	-44	-29	-35	-43	-40	-3	-17		
208678	-7	-4	-3	1	-3	-3	-2	-1	-3
208686	3	0	4	4	1	-1	1	3	1
208710	7	7	6	17	11	13	8	13	8
208744	-58	-37	-44	-64	-46	-37	-31	-48	-39
208769	7	20	10	17	29	15	-1	0	-11
208819	1	2	0	-1	2	1	0	4	1
208827	-16	-13	-8						
208835	-17	2	2	-6	-3	-2			
208843	-25	-17	-5	-18	-16	-12	-15	-11	-23
208884	-57	-189	-952	-260	-83	-91	-336	-187	-269
208918	-259	-186	-128	-255	-256	-166	-106	-249	-172
209007	-93	-70	-71	-154	-83	-79	-75	-104	-109
209130	-16	-16	-29	-21	-14	-13	-21	-23	-29
209155	-33	-3	-11	-27	-46	-41	-30	-60	-83
209189	-64	-61	-40	-114	-74	-53	-57	-63	-72
209197	2	0	-3	-4	1	-2	-4	-2	-5
209338	-30	-21	-16	-18	-25	-17	-23	-33	-33
209361	0	1	-3	-1	1	-2	1	-4	-2
209478	-15	-15	-16	-26	-18	-20	-18	-15	-19
209510	-35	-27	-22	-33	-42	-41	-36	-37	-42
209536	-40	-39	-37	-38	-42	-41	-42	-44	-41
209742	-2	2	5	-12	3	-4	6	-20	-8
209817	4	-3	-13	-12	-4	-1	-13	-8	-7
209882	1	6	3	-2	6	8	1	8	-4
210039	-15	-2	-12	-24	-17	-7	-6	-14	-9
210054	-37	-29	-29	-31	-16	-29	-20	-36	-34
210062	-6	-2	-7	-3	-3	-3	-4	-11	-4
210070	-40	-28	-32	-117	-62	-56	-71	-62	-101
210088	-92	-51	-43	-137	-89	-80	-99	-115	-208
210096	-68	-23	-20	-83	-73	-59	-52	-81	-104

Continued on next page

Tab. B.1: Runoff gauging stations with their MQ runoff difference (continuation).

	MQ runoff difference (obs-sim) in %								
$HZBR_{NR}$	2009	2010	2011	2012	2013	2014	2015	2016	2017
210203	-54	-45	-34	-31	-51	-39	-32	-33	-42
210211	-15	-18	-16	-20	-15	-15	-21	-18	-29
210237	-18	-19	-13	-25	-18	-5	-20	-18	-31
210245	-18	-21	-8	-38	-23	-14	2	-32	-29
210252	-25	-31	-39	-61	-25	-15	-30	-42	-52
210260	-29	-25	-14	-40	-21	-14	-20	-21	-39
210302	-8	7	-4	-6	-4	-2	-8	-6	-5
210310	-19	-17	-18	-20	-5	-10	3	-21	-21
210401	-15	-14	-15	-24	-13	-19	-17	-21	-21
210419	-11	-6	-8	-10	-12	-9	-3	4	5
210468	-18	-13	-9	-29	-6	-9	-8	-11	-20
210500	-38	-29							
210526	7	6	4	5	5	5	6	6	7
210625	-3	-2	-1	-3	-2	-2	-3	-2	-3
210641	-15	-17	-24	-18	-22	-12	-17	-10	-11
210732	-90	-96	-98	-77	-79	-93	-90	-89	-85
210773	2	3	4	4	3	2	5	6	3
210799	-21	-24	-26	-19	-27	-21	-20	-18	-25
210815	0	0	-5	12	5	2	6	6	4
210823	-19	-22	-25	-13	-20	-16	-15	-14	-19
210856	24	26	23	29	25	21	23	30	30
210864	-4	-6	-7	1	-10	-9	-6	4	1
210898	14	18	16	18	21	28	17	13	12
210963	-4	1	9	4	2	1	11	4	0
210971	1	-5	3	2	3	7	0	2	-1
210989	-11	-9	-9	-11	-2	0	-4	-4	-4
210997	-2	8	8	7	-2	1	8	0	1
211003	-16	-6	-8	-8	-16	-10	-12	-7	-21
211029	-10	-16	-1	2	-12	-14	-11	-8	-2
211037	-24	-23	-41	-49	-31	-9	-32	-63	-57
211086	-13	-21	-19	-14	-18	-18	-21	-18	-17
211102	-15	-19	-20	-17	-21	-18	-22	-19	-14
211110	-3	0	7						
211128	-9	-13	-16	-12	-18	-11	-12	-16	-14
211136	-24	-29	-30	-27	-33	-27	-28	-29	-23
211169	12	2	3	4	-5	3	3	-2	1
211185	-19	-23	-19	-19	-22				
211193	-20	-17	-18	-13	-19	-18	-16	-25	-19
211227	11	10	10	12	12	13	13	12	10
211243	-27	-20	-17	-14	-22	-31	-19	-22	-25
211250	-35	-41	-36	-31	-27	-38	-30	-40	-35
211268	-32	-33	-32	-23	-20	-31	-30	-35	-35
211276	-12	-10	-13	-8	-9	-17		-11	-12

Continued on next page

Tab. B.1: Runoff gauging stations with their MQ runoff difference (continuation).

	MQ runoff difference (obs-sim) in %								
$HZBR_{NR}$	2009	2010	2011	2012	2013	2014	2015	2016	2017
211292	-18	-17	-20	-16	-16	-20	-22	-21	-20
211334	-3	-7	-14	-26	-12	-12	-24	-26	-65
211342	3	-7	-4	-7	-7	3	-4	2	-11
211383	6	7	8	3	5	8	9	8	6
211391	-20	-15	-16	-13	-17	-13	-15	-12	-10
211441	-9	-8	-14	-14	-10	-12	-11	-12	-18
211458	-3	-6	-13	-11	-7	-7	-15	-6	-13
211474	0	3	-13	0	-2	1	-1	-1	-5
211508	-4	-5	-5	-3	-6	-6	-4	-3	0
211573	-17	-17	-17	-16	-16	-20	-21	-19	-21
211599	-7	-9	-16	-14	-15	-13	-28	-23	-35
211631	-43	-37	-43	-39	-41	-38	-70	-51	-55
211649	-4	11	-7	-10	0	0	-25	-13	-31
211656	-23	-21	-37	-36	-28	-31	-42	-20	-29
211664	2	7	6	8	7	3	8	5	7
211730	-2	0	3	-1	4	5	3	2	0
211763	-17	-16	-18	-15	-17	-20	-21	-21	-20
211797	6	2	1	7	10	4	3	-1	1
211854	-28	-35	-35	-37	-37	-31	-43	-49	-47
211870	-19	-18	-18	-17	-18	-21	-23	-20	-22
211888	-44	-55	-48	-43	-45	-51	-45	-65	-37
211896	9	-4	-7	-1	-1	1	1	-1	-6
211904	-3	-6	-9	-5	-7	-6	-7	-3	
211961	-30	-28	-31	-26	-23	-22	-24	-24	-26
211995	-28	-37	-44	-60	-37	-40	-59	-46	-36
212043	1	5	1	3	-1	1	2	3	0
212076	-14	-16	-14	-15	-13	-12	-15	-12	-11
212092	-24	-25	-26	-21	-14	-18	-29	-25	-24
212100	-9	-3	-3	-2	-2	-5	-7	-5	-7
212118	8	7	7	7	6	7	7	6	8
212167	-22	-21	-23	-19	-15	-20	-26	-25	-21
212217	0	-1	0	1	1	0	-1	1	1
212324	-19	-21	-22	-23	-15	-13	-20	-22	-31
212357	-17	-22	-21	-20	-18	-12	-20	-21	-23
212373	-30	-30	-23	-23	-16	-17	-23	-29	-28
212381	1	3	1	1	4	1	-2	6	2
212431	-68	-85	-92	-70	-91	-49	-73	-75	-71
212472	-16	-27	-26	-19	-27	-11	-32	-27	-16
212498	-17	-26	-18	-16	-20	-13	-22	-27	-34
212522	-10	-12	-5	2	-1	-12	-4	-12	-11
212530	-12	-14	-14	0	-4	-6	-17	-18	-14
212597	-80	-62	-69	-81	-64	-82	-77	-77	-70
212613	7	10	7	-5	0	11	0	-6	-12

Continued on next page

Tab. B.1: Runoff gauging stations with their MQ runoff difference (continuation).

	MQ runoff difference (obs-sim) in %								
$HZBR_{NR}$	2009	2010	2011	2012	2013	2014	2015	2016	2017
212647	6	2	2	-2	-3	1	-5	4	-17
212670	8	10	8	4	2	17	-3	5	0
212704	0	-4	-3	0	-1	-1	-7	-6	-2
212753	10	16	7	16	11	17	6	16	12
212787	8	12	6	10	11	9	4	9	8
212852	-12	-11	-12	-8	-7	-11	-12	-10	-10
212878	10	5	6	3	-8	3	10	8	7
212886	-5	-7	-8	-12	-22	-8	-5	-8	-12
212894	-21	-19	-8	-20	-16	-6	-15	-10	-9
212928	-1	2	5	8	1	3	4	3	5
212936	0	-6	-4	-7	-2	-4	-5	0	-7
212951	-18	-21	-14						
213025	-35	-17	-18	-27	-27	-20	-39	-19	-21
213033	-9	-3	0	-2	5	-3	-5	-1	-2
213041	-12	-12	-5	-10	-6	-11	-5	-9	-9
213082	2	-14	-4	-2	0	-1	-2	1	-1
213090	-10	-17	-14	-10	-9	-3	-12	-13	-1
213116	-18	-22	-21	-22	-17	-11	-13	-5	-10
213124	-399	-419	-465	-432	-420	-382	-425	-435	-470
213157	-13	-12	-11	-13	-13	-6	-7	-17	-17
213181	-1	-2	-5	-4	-3	1	-8	-5	-7
213207	-9	-5	-3	-2	-11	-4	-2	0	-6
213215	-63	-72	-77	-68	-64	-52	-72	-70	-78
213231	10	9	11	7	10	7	12	9	9
213249	-9	-8	-6						
213256	-27	-28	-27	-49	-38	-38	-36	-54	-44
213306	-2	0	5	-7	-3	-3	-6	3	-3
213322	-24	-15	-36	-21	-16	-13	-43	-28	-18
213355	-45	-65	-68	-74	-34	-46	-57	-63	-42
213371	-6	-14	-9	-7	-5	-2	-8	-7	-3
213397	-19	-17	-13	-22	-23	-11	-15	-19	-18
213405	-34	-36	-32	-38	-33	-29	-31	-28	-35
213926		-80	-100	-95	-93	-85	-86	-89	-98
214031	-78	-75	-74	-114	-81	-75	-77	-99	-106
214304			-27	-48	-32	-30	-31	-49	-53
214536	6	8	0	15	11	11	10	13	9
214668	-13	-18	-8	-29	-14	-19	-28		
215012	-39	-48	-50	-45	-49	-27	-41	-48	-73
215038	-42	-35	-24	-49	-36	-19	-28	-34	-47
230078	-18	-20	-17	-14	-16	-17	-16	-19	-23
230300	10	9	12	9	8	8	11	13	13
230342	-4	-2	-3	-1	1	-2	-2	-4	-1
230706	-15	-5	-9	-7	-6	-5	-1	-2	-10

Continued on next page

Tab. B.1: Runoff gauging stations with their MQ runoff difference (continuation).

	MQ runoff difference (obs-sim) in %								
$HZBR_{NR}$	2009	2010	2011	2012	2013	2014	2015	2016	2017
231092	-18	-19	-15	-21	-17	-7	-15	-11	-18
231100	-6	-6	-6	-5	-5	-6	-5		
231662	14	21	0	11	-8	-10	1	7	24
231688	-20	-7	-20	-6	-16	-30	-30	-21	-15
18000403	-20		-21	-13	-18	-15	-15	-13	-15
18001508	-10	-10	-12	-4	-5	-8	-17	-15	-15
18003004	-17	-21	-20	-13	-14	-19	-17	-18	-18
18209000	-55	-44	-59	-43	-46	-79	-70	-44	-44
18212004	-2	-2	-3	-2	-2	-3	-2	-2	-3
18242005	0	-1	0	-1	-2	0	0	0	0
18246006	-1	9	7	9	14	7	4	5	1
18262002	5	4	6	5	3	3	3	1	3
18263005	-6	-9	-8	-10	-6	-7	-4	-1	-6
18463004	-30	-26	-31	-26	-23	-30	-28	-25	-25
18483500	-14	-1	-4	-17	-8	7	-6	-21	-7
18484503	2	-1	2	2	2	-4	1	3	1
18487501	-2	-2	-6	-2	-2	-1	-3	-2	-3
18606000	-8	-7	-12	-11	-8	-12	-12	-3	-8
18620500	4	7	14	7	8	14	14	17	10
18622006	16	10	11	3	6	22	-57	-28	-22
18625004	4	-8	13	11	0	3	8	12	13
18642003	10	4	7	3	-1	-1	1	5	2
18643006	9	1	0	-8	-11	4		-20	-26
18646809	-34	-105	-85	-64	-34	-156	-54	-68	-51
18662000	-26	-11	-2	-28	-1	-10	-18	-11	-14
18683000	-16	-16	-5	-36	-18	-59	-42	-19	-46
18801005	-1	-2	-4	2	1	-2	-2	0	-9
18803805	-39	-13	-10	-36	-50	-46	-34	-34	-37
18804706	-15	-7	-9	-41	-49	-19	-17	-19	-24
18806406	-27	-15	-17	-41	-38	-34	-27	-27	-25
18825003	-85	-57	-138	-80		-91	-76	-107	-97

Appendix C

R script

C.1 Parent script

```r
##%##########################################################%##
#                                                              #
#   Diploma Thesis                                             #
#   TopKriging prediction with                                 #
#   diversion consideration                                    #
#                                                              #
#   Parent file                                                #
#   Creator:                                                   #
#   nikolaus.weber@tuwien.ac.at                                #
#   Last edit:                                                 #
#   02.12.2020 by Nikolaus Weber                               #
#                                                              #
##%##########################################################%##


## Libs ######################################################################

    if (!require("vctrs"))      install.packages("vctrs", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("ggplot2"))    install.packages("ggplot2", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("rgdal"))      install.packages("rgdal", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("sp"))         install.packages("sp", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("rtop"))       install.packages("rtop", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("tmap"))       install.packages("tmap", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("sf"))         install.packages("sf", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("igraph"))     install.packages("igraph", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("data.table")) install.packages("data.table", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("gridExtra"))  install.packages("gridExtra", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("tidyverse"))  install.packages("tidyverse", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("ggsci"))      install.packages("ggsci", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("digest"))     install.packages("digest", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("raster"))     install.packages("raster", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("hydroGOF"))   install.packages("hydroGOF", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("rgeos"))      install.packages("rgeos", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("readxl"))      install.packages("readxl", dependencies = TRUE, repos="https://cloud.r-project.org/")

    library(vctrs)
    library(ggplot2)
    library(rgdal)
    library(sp)
    library(rtop)
    library(tmap)
    library(sf)
    library(igraph)
    library(data.table)
    library(gridExtra)
    library(tidyverse)
    library(ggsci) #theme plotting
    library(digest)
    library(raster)
    library(gridExtra) # for table plot
    library(hydroGOF) # for statistic like NSE, ...
    library(rgeos) # for GIS work
    library(readxl) # for reading Excel-files


## Set WD & Create New Directory ##############################################
```

```r

wd <- "C:/Users/nikol/ownCloud/Diplomarbeit_BauIng" # Cloud @ HOME
setwd(wd)
getwd()


## Settings for Plotting #####################################################

theme_plot <- function() {
  theme_bw() %+replace%
  theme(
  axis.text.x =       element_text(size = 10 , lineheight = 1,  colour = "black", margin = margin(3,5,3,5,"pt")),
  axis.text.y =       element_text(size = 10 , lineheight = 1,  colour = "black", margin = margin(5,3,5,1,"pt")),
  axis.ticks =        element_line(colour = "black", size = 0.2),
  axis.title.x =      element_text(size = 10, angle = 0, vjust = -0.8, colour = "black", face = "bold", margin = margin(3,5,3,5,"pt")),
  axis.title.y =      element_text(size = 10, angle = 90, vjust = 6, colour = "black", face = "bold", margin = margin(5,3,5,1,"pt")),
  axis.ticks.length = unit(0.3, "lines"),

  legend.background = element_rect(colour=NA),
  legend.key =        element_blank(),
  legend.key.size =   unit(1.2, "lines"),
  legend.text =       element_text(size = 10, colour = "black", margin = margin(2,0,2,2,"pt")),
  legend.title =      element_blank(),
  legend.position =   "right",

  panel.background =  element_rect(fill = "white", colour = NA),
  panel.border =      element_rect(fill = NA, colour="black"),
  panel.grid.major =  element_line(colour = "grey85"),
  panel.grid.minor =  element_line(colour = "grey90"),
  panel.spacing =     unit(0.5, "lines"),

  strip.background =  element_rect(fill = "white", colour = "black"),
  strip.text.x =      element_text(size = 10),
  strip.text.y =      element_text(size = 10, angle = -90),

  plot.background =   element_rect(colour = NA),
  plot.title =        element_text(size = 11, colour = "black", vjust = 0, face = "bold", margin = margin(0,5,10,5,"pt")),
  plot.margin =       unit(c(1, 1, 1, 1), "lines")
  )
}

theme_set(theme_plot())

col_grg <-  c(rgb(33, 89, 46, maxColorValue = 255), rgb(230, 75, 52, maxColorValue = 255), rgb(50, 50, 50, maxColorValue = 255))
col_gr <-   c(rgb(33, 89, 46, maxColorValue = 255), rgb(230, 75, 52, maxColorValue = 255))

breaks <-       10^(-10:10)
minor_breaks <- rep(0:9, 21)*(10^rep(-10:10, each=10))


# plotting - range of runoff

# "l/s/km2"
q_at =       c(0,5,10,20,30,50,100,200) # typische MQ werte (https://de.wikipedia.org/wiki/Abflussspende)
q_col =      rev(bpy.colors(length(q_at))) # invert the Legend color

# "m3/s/km2"
m3_at =      c(seq(0,0.07, 0.005)) # predictions
m3_col =     rev(bpy.colors(length(m3_at))) # predictions
m3_at_var =  seq(0,0.00015,0.00001) # variance
m3_col_var = rev(bpy.colors(length(m3_at_var))) # variance

# "mm"
mm_at =      c(0,200,seq(500,1500,200),2000,2900) # predictions
```

```
    mm_col =     rev(bpy.colors(length(mm_at))) # predictions
    mm_at_var =  c(seq(0,10,0.5),20) # variance
    mm_col_var = rev(bpy.colors(length(mm_at_var))) # variance

    mm_at_OL <-  c(-20000,seq(600,1800,100),2000,20000) # for the outliers
    mm_col_OL <- rev(bpy.colors(length(mm_at_OL))) # for the outliers


## pre-processing data ########################################################

## load spacial data
source("analysis/1_load_spacial_data.R")
# OR
load("./data/prepareSPACIAL/01_saveVAR_loadSPACIAL.RData")

## prepare spacial data
source("analysis/1_prepare_spacial_data.R")
# OR
load("./data/prepareSPACIAL/02_saveVAR_prepareSPACIAL_processed.RData")

## create MQ table
source("analysis/1_MQ_table_for_rtop.R")
# or
load("./data/prepareSPACIAL/01_saveVAR_MQ_table.RData")


## processing & post-processing data #########################################

# setting for Output
version <- "V46" # version
saveDir <- paste(Sys.Date(),"STOBIMO_all", version, sep = "_") # save directory
dir.create(file.path(paste0("data/",saveDir,"/")), showWarnings = F)
writeLines(paste(date()," \nVersion: ", version, "\n",
                 "run final calculation of all STOBIMO AUs
                 V1.1 reduce raster size to 2km^2 \n") , # add description here
           paste0("data/",saveDir,"/00_READ_ME.txt"))

# processing parameter
## MQyear       # set year of MQ
## Div          # set TRUE if effective watershed area (A_eff) of the gauge should be used
## OL_Limit_Q   # set TRUE if outliers should be limited
## OL_Limit     # set outliers Limit(min, max)
set.seed    (1)  # to produce reproduceable results

# run rtop interpolation and post-processing
source("analysis/2_rtop_interpolation_final.R")
# rtop_interp(MQyear = 2009, Div = F, OL_Limit_Q = T, OL_Limit = c(50, 2900))
lapply(2009:2017, FUN = rtop_interp, Div = T, OL_Limit_Q = FALSE, OL_Limit = c(50, 2900))
lapply(2009:2017, FUN = rtop_interp, Div = F, OL_Limit_Q = FALSE, OL_Limit = c(50, 2900))


## Validation, Comparison & Export data ######################################

# Calc splitting factor
source("analysis/3_calc_splitting_factor.R")

# Validation
source("analysis/3_validation.R")

# result comparison
source("analysis/3_result_comparison.R")

# prepare data for export to MORE
source("analysis/3_Export2MORE.R")


## End parent script ##########################################################
```

C.2 Child script: Pre-Processing

C.2.1 Load spatial data

```r
##%#################################################%##
#                                                    #
#   Diploma Thesis                                   #
#   TopKriging prediction with                       #
#   diversion consideration                          #
#                                                    #
#   load spacial data                                #
#   Creator:                                         #
#   nikolaus.weber@tuwien.ac.at                      #
#   Editor:                                          #
#   nikolaus.weber@tuwien.ac.at                      #
#   Last edit:                                       #
#   23.09.2020 by Nikolaus Weber                     #
#                                                    #
##%#################################################%##


## Libs #######################################################################

    if (!require("rgdal")) install.packages("rgdal", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("sp")) install.packages("sp", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("sf")) install.packages("sf", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("data.table")) install.packages("data.table", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("raster")) install.packages("raster", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("rgeos")) install.packages("rgeos", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("readxl")) install.packages("readxl", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("dplyr")) install.packages("dplyr", dependencies = TRUE, repos="https://cloud.r-project.org/")

    library(rgdal)
    library(sp)
    library(sf)
    library(data.table)
    library(gridExtra)
    library(raster)
    library(rgeos) # for GIS work
    library(readxl)
    library(dplyr)


## load spacial data ##########################################################

# Use dir() to find directory name (Use dir() to take a look in your working directory.)
dir()

# Location of the shapefiles
dsn_test_region <- "raw_data/rohdaten_gis_raw.gdb"

# Call dir() with directory name
dir(dsn_test_region)

# Check layers in dir()
ogrListLayers(dsn_test_region)

# Read in shapefile with readOGR(): (insert layer without extension)

## river_network
rnet <- readOGR(dsn = dsn_test_region, layer = "HORA_edges")  # Austria 5775 obs.

## predictionLocations
predictionLocations <- readOGR(dsn = dsn_test_region, layer = "HORA_Watersheds_gesamt")          # Austria     7774 obs.
predictionLocations_CH <- readOGR(dsn = "./data/prepareSPACIAL/data_input/STOBIMO_EZG_CH.shp")  # Switzerland 19 obs.
predictionLocations_BY <- readOGR(dsn =
```

```
"./data/prepareSPACIAL/data_input/STOBIMO_EZG_BY.shp")  # Bavaraia    78 obs.

## rnet_gauges
rnet_gauges <- readOGR(dsn = dsn_test_region, layer = "ehyd_pegel_2011")  # gauges Austria (eHyd) 771 obs.
rnet_gauges_CH <- readOGR(dsn = dsn_test_region, layer = "Pegel_CH")      # gauges Switzerland     15 obs.
rnet_gauges_BY <- readOGR(dsn = dsn_test_region, layer = "Pegel_BY")      # gauges Bavaraia        88 obs.

## STOBIMO_EZG
STOBIMO_EZG <- st_as_sf(readOGR(dsn = dsn_test_region, layer = "STOBIMO_SPURENSTOFFE_EZG_V2")) # STOBIMO_EZG (MoRE AU)
STOBIMO_EZG$HZB_PEGEL1 <- as.character(STOBIMO_EZG$HZB_PEGEL1)
STOBIMO_EZG_Samina <- st_as_sf(readOGR(dsn = "./data/prepareSPACIAL/data_input/STOBIMO_EZG_Samina.shp")) # Samia river, Liechtenstein
STOBIMO_EZG_Samina <- st_transform(STOBIMO_EZG_Samina, st_crs(STOBIMO_EZG)) # change CRS
colnames(STOBIMO_EZG_Samina) <- colnames(STOBIMO_EZG) # make same colnames as STOBIMO_EZG
STOBIMO_EZG <- rbind(STOBIMO_EZG,STOBIMO_EZG_Samina) # merge
STOBIMO_EZG <- as(STOBIMO_EZG, Class = "Spatial") # convert back to sp
rm(STOBIMO_EZG_Samina)

## borders Austria
bord_AUT <- rgeos::gSimplify(readOGR(dsn = dsn_test_region, layer = "Oesterreich_ges"),tol = 100)   # borders Austria
bord_AUT_BL <- rgeos::gSimplify(readOGR(dsn = dsn_test_region, layer = "Oesterreich_BL"),tol = 100) # borders federal states Austria
bord_AUT_PB <- rgeos::gSimplify(readOGR(dsn = dsn_test_region, layer = "Oesterreich_PB"),tol = 100) # borders municipalities Austria

rm(dsn_test_region)


## save image ################################################################

save.image("./data/prepareSPACIAL/01_saveVAR_loadSPACIAL.RData")


## End load spacial data #####################################################
```

C.2.2 Prepare spatial data

```r
##%######################################################%##
#                                                          #
#   Diploma Thesis                                         #
#   TopKriging prediction with                             #
#   diversion consideration                                #
#                                                          #
#   prepare spacial data                                   #
#   Creator:                                               #
#   nikolaus.weber@tuwien.ac.at                            #
#   Editor:                                                #
#   nikolaus.weber@tuwien.ac.at                            #
#   Last edit:                                             #
#   18.11.2020                                             #
#                                                          #
##%######################################################%##



## Libs ######################################################################

    if (!require("rgdal")) install.packages("rgdal", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("sp")) install.packages("sp", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("sf")) install.packages("sf", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("data.table")) install.packages("data.table", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("raster")) install.packages("raster", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("rgeos")) install.packages("rgeos", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("readxl")) install.packages("readxl", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("dplyr")) install.packages("dplyr", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("nngeo")) install.packages("nngeo", dependencies = TRUE, repos="https://cloud.r-project.org/")

    library(rgdal) # for GIS work
    library(sp) # for GIS work
    library(sf) # for GIS work
    library(data.table) # for fast table processing
    library(gridExtra) # for plotting tables
    library(raster) # for GIS work
    library(rgeos) # for GIS work
    library(readxl) # for reding Excel files
    library(dplyr) # for easy table processing
    library(nngeo) # to remove holes in sf-objects


## prepare STOBIMO_EZG #######################################################

    # corrections ("STOBIMO_SPURENSTOFFE_EZG_V2")
    STOBIMO_EZG$HZB_PEGEL1[STOBIMO_EZG$ID_MORE == 70230] <- 18005000 #correction after GIS analysis (gauge station: Eschelbach / Inn)
    STOBIMO_EZG$HZB_PEGEL1[STOBIMO_EZG$ID_MORE == 70075] <- 18246006 #correction after GIS analysis (gauge station: Erb / Leitzach)
    STOBIMO_EZG$HZB_PEGEL1[STOBIMO_EZG$ID_MORE == 10980] <- 208157 #correction after GIS analysis (gauge station: Schwechat (Hallenbad))
    # removed gauges due to  non-compliance with the validation method
    STOBIMO_EZG$HZB_PEGEL1[STOBIMO_EZG$ID_MORE == 12225] <- NA #correction after GIS analysis (gauge station: 205229    Ebensee (Unterlangbath))
    STOBIMO_EZG$HZB_PEGEL1[STOBIMO_EZG$ID_MORE == 40065] <- NA #correction after GIS analysis (gauge station: 2319   Ova da Cluozza - Zernez)
    STOBIMO_EZG$HZB_PEGEL1[STOBIMO_EZG$ID_MORE == 70055] <- NA #correction after GIS analysis (gauge station: 18226009  Miesbach / Schlierach)

    # remove unplausible gauges
    gauges_dismiss <- setDT(read_excel("./data/Diversion_data/Diversion_data_table.xlsx", sheet = "DIV_obs_gauges_dismiss"))
    STOBIMO_EZG$HZB_PEGEL1[STOBIMO_EZG$HZB_PEGEL1 %in% gauges_dismiss$HZBR_NR] <- NA
```

```
## prepare rnet ##############################################################

    # corrections ("HORA_edges")
    rnet$EZGA[rnet$EZGE == 7422] <- rnet[rnet$EZGE == 7422,]$EZGE #correction after GIS analysis
    rnet$EZGA[rnet$EZGE == 429] <- rnet[rnet$EZGE == 429,]$EZGE #correction after GIS analysis
    rnet$EZGA[rnet$EZGE == 5011] <- 4759 #correction after GIS analysis

    rnet$HZBNRE[rnet$EZGE == 6664] <- 2265     #correction after GIS analysis (gauge station: Tarasp)
    rnet$HZBNRE[rnet$EZGE == 6799] <- 2067     #correction after GIS analysis (gauge station: Martina)
    rnet$HZBNRE[rnet$EZGE == 7274] <- 213173   #correction after GIS analysis (gauge station: Lavamünd/Drau)
    rnet$HZBNRE[rnet$EZGE == 6344] <- 207332   #correction after GIS analysis (gauge station: Marchegg (Fluss-km 14,98))
    rnet$HZBNRE[rnet$EZGE == 3019] <- 203851   #correction after GIS analysis (gauge station: Böckstein (Summenpegel))
    rnet$HZBNRE[rnet$EZGE == 6069] <- 214031   #correction after GIS analysis (gauge station: Deutsch Brodersdorf (Messseilbahn))
    rnet$HZBNRE[rnet$EZGE == 4350] <- 213926   #correction after GIS analysis (gauge station: Heiligenblut-OWF)
    rnet$HZBNRE[rnet$EZGE == 6961] <- 204297   #correction after GIS analysis (gauge station: Salzburg (Summenpegel))
    rnet$HZBNRE[rnet$EZGE == 7508] <- 201905   #correction after GIS analysis (gauge station: Kufstein (Bahnhofsbrücke))
    rnet$HZBNRE[rnet$EZGE == 7278] <- 206201   #correction after GIS analysis (gauge station: Schärding (Schreibpegel))
    rnet$HZBNRE[rnet$EZGE == 1262] <- 214445   #correction after GIS analysis (gauge station: Lunz am See (Wassercluster-Summe))
    rnet$HZBNRE[rnet$EZGE == 3798] <- 214304   #correction after GIS analysis (gauge station: Furth (Feuerwehrhaus))
    rnet$HZBNRE[rnet$EZGE == 7283] <- 0        #correction after GIS analysis (no gauge)
    rnet$HZBNRE[rnet$EZGE == 2368] <- 0        #correction after GIS analysis (no gauge)
    rnet$HZBNRE[rnet$EZGE == 2464] <- 0        #correction after GIS analysis (no gauge)
    rnet$HZBNRE[rnet$EZGE == 3024] <- 0        #correction after GIS analysis (no gauge)
    rnet$HZBNRE[rnet$EZGE == 7294] <- 18007800 #correction after GIS analysis (gauge station: Passau Ingling KW / Inn)
    rnet$HZBNRE[rnet$EZGE == 7137] <- 18005702 #correction after GIS analysis (gauge station: Braunau-Simbach KW / Inn)
    rnet$HZBNRE[rnet$EZGE == 7538] <- 18000403 #correction after GIS analysis (gauge station: Oberaudorf / Inn)
    rnet$HZBNRE[rnet$EZGE == 7506] <- 211490   #correction after GIS analysis (gauge station: Mureck (Schreibpegel))


## prepare rnet_gauges #######################################################

    ## get list of all gauges
    gauges <- unique(c( rnet_gauges$HZBNR01, rnet_gauges_BY$Messtellen, rnet_gauges_CH$ID))
    ## get list of all Austrian Q gauges
    gauges_Q_AT <- as.data.table(rnet_gauges@data)
    gauges_Q_AT <- gauges_Q_AT[Messstelle %like% "Durchfluss", HZBNR01]

    ## subset rnet_gauges to IWAG list (list of gauges with runoff data)
    IWAG_eHyd_Q <- as.data.table(read.csv2(file = "./raw_data/Jahresabfluesse/20200912_Jahresabfluesse_alle_Pegel_IWAG_updated.csv", header = T)) ## mean annual q [m3/s]
    gauges_dismiss <- setDT(read_excel("./data/Diversion_data/Diversion_data_table.xlsx", sheet = "DIV_obs_gauges_dismiss"))
    IWAG_eHyd_Q <- IWAG_eHyd_Q[!pegel_nr %in% gauges_dismiss$HZBR_NR, pegel_nr, by = "pegel_nr"] # 643 obs.
    IWAG_eHyd_Q <- unique(IWAG_eHyd_Q$pegel_nr)
    rm(gauges_dismiss)
```

```
    # subset to available data
    rnet_gauges <- st_as_sf(rnet_gauges[rnet_gauges$HZBNR01 %in% IWAG_eHyd_Q,]) # 586 obs.
    rnet_gauges_BY <- st_as_sf(rnet_gauges_BY[rnet_gauges_BY$Messtellen %in% IWAG_eHyd_Q,]) # 36 obs.
    rnet_gauges_CH <- st_as_sf(rnet_gauges_CH[rnet_gauges_CH$ID %in% IWAG_eHyd_Q,]) # 12 obs.

    # add gauges ID
    rnet_gauges_BY$EZGE <- as.integer(as.character(predictionLocations_BY$ID_MORE[match(rnet_gauges_BY$Messtellen,predictionLocations_BY$HZB_PEGEL1)]))
    rnet_gauges_CH$EZGE <- as.integer(as.character(predictionLocations_CH$ID_MORE[match(rnet_gauges_CH$ID,predictionLocations_CH$HZB_PEGEL1)]))
    rnet_gauges_CH$EZG <- as.numeric(predictionLocations_CH$AREAKM2[match(rnet_gauges_CH$EZGE,predictionLocations_CH$ID_MORE)])

    # prepare for merge
    rnet_gauges              <- rnet_gauges %>% dplyr::select("HZBNR01","MSTNAM02","GEW03","EGAREA05","EZGE","geometry")
    colnames(rnet_gauges)    <- c("ID","name","river","Area","PredL_EZGE","geometry")
    # rnet_gauges_BY
    rnet_gauges_BY           <- rnet_gauges_BY %>% dplyr::select(4,2,3,8,15,16)
                           # <- rnet_gauges_BY[,c("Messtellen","Pegelname","GewÃ.sser","GrÃ.ÃYe_Einz","EZGE","geometry")]
    colnames(rnet_gauges_BY) <- c("ID","name","river","Area","PredL_EZGE","geometry")
    rnet_gauges_CH           <- rnet_gauges_CH %>% dplyr::select("ID", "Ort", "GewÃ.sser","EZG","EZGE","geometry")
    colnames(rnet_gauges_CH) <- c("ID","name","river","Area","PredL_EZGE","geometry")

    # change class
    rnet_gauges_BY$ID         <- as.integer(rnet_gauges_BY$ID)

    rnet_gauges$PredL_EZGE    <- as.integer(as.character(rnet_gauges$PredL_EZGE))
    rnet_gauges_BY$PredL_EZGE <- as.integer(as.character(rnet_gauges_BY$PredL_EZGE))
    rnet_gauges_CH$PredL_EZGE <- as.integer(as.character(rnet_gauges_CH$PredL_EZGE))

    rnet_gauges$name          <- as.character(rnet_gauges$name)
    rnet_gauges_BY$name       <- as.character(rnet_gauges_BY$name)
    rnet_gauges_CH$name       <- as.character(rnet_gauges_CH$name)

    rnet_gauges$river         <- as.character(rnet_gauges$river)
    rnet_gauges_BY$river      <- as.character(rnet_gauges_BY$river)
    rnet_gauges_CH$river      <- as.character(rnet_gauges_CH$river)

    # merge
    rnet_gauges_all <- rbind(rnet_gauges_CH, rnet_gauges_BY, rnet_gauges)

    rnet_gauges <- as(rnet_gauges_all, Class = "Spatial")

    rm(IWAG_eHyd_Q, rnet_gauges_all, rnet_gauges_BY, rnet_gauges_CH, gauges)

    # add missing AREA
    rnet_gauges$Area[rnet_gauges$ID == 2067] <- 1941 #correction after BAFU analysis (gauge station: Inn - Martina)
    rnet_gauges$Area[rnet_gauges$ID == 2265] <- 1581 #correction after BAFU analysis (gauge station: Inn - Tarasp)
    rnet_gauges$Area[rnet_gauges$ID == 2319] <- 27   #correction after BAFU analysis (gauge station: Ova da Cluozza - Zernez)
    rnet_gauges$Area[rnet_gauges$ID == 2304] <- 55.3 #correction after BAFU analysis (gauge station: Ova dal Fuorn - Zernez, Punt la Drossa)


## prepare predictionLocations ###############################################

  ##### prepare predictionLocations AT

    # corrections ("HORA_Watersheds_gesamt")
    predictionLocations$AREA_KOR[predictionLocations$EZGID == 7688] <- 26847.1803 +
```

```r
    2441.506203 - 2438.180333 #correction after GIS analysis
    predictionLocations <- predictionLocations[predictionLocations$EZGID != 5974,] # removed because it causing problems
    predictionLocations <- predictionLocations[predictionLocations$EZGID != 1336,] # removed because it causing problems
    predictionLocations <- predictionLocations[predictionLocations$EZGID != 4883,] # removed because it causing problems
    predictionLocations <- predictionLocations[predictionLocations$EZGID != 3112,] # removed because it causing problems
    predictionLocations <- predictionLocations[predictionLocations$EZGID != 3985,] # removed because it causing problems
    predictionLocations <- predictionLocations[predictionLocations$EZGID != 5047,] # removed because it causing problems


    # rnet for identification
    rnet_EZG <- rnet[rnet$EZGA !=  rnet$EZGE, c("EZGA","EZGE","HZBNRE")]@data

    ### add variables to predictionLocations
    predictionLocations$EZGE <- as.integer(as.character(predictionLocations$EZGID))
    predictionLocations$EZGA <- as.integer(as.character(rnet_EZG$EZGA[match(predictionLocations$EZGID, rnet_EZG$EZGE)]))
    predictionLocations$EZGTO <- as.integer(as.character(NA))
    predictionLocations$EZGE_AREA <- round(predictionLocations$AREA_KOR[match(predictionLocations$EZGE, predictionLocations$EZGID)],digits = 2)
    predictionLocations$EZGA_AREA <- round(predictionLocations$AREA_KOR[match(predictionLocations$EZGA, predictionLocations$EZGID)],digits = 2)
    predictionLocations$ID_GAUGE <- as.integer(as.character(rnet$HZBNRE[match(predictionLocations$EZGID, rnet$EZGE)]))
    rm(rnet_EZG)
    #View(predictionLocations@data) # 7774 obs.

    # subset
    predictionLocations <- predictionLocations[predictionLocations$EZGE %in% unique(c(rnet$EZGA, rnet$EZGE)) |
                                               predictionLocations$EZGA %in% unique(c(rnet$EZGA, rnet$EZGE)),] # 7555 obs.

    # analyse for outliers
    DT_predL <- as.data.table(predictionLocations)
    DT_predL[EZGE_AREA < EZGA_AREA,] # non
    DT_predL[, AREA_DIFF := EZGE_AREA - EZGA_AREA][AREA_DIFF == 0 & AREA_KOR > 1,] # 6 obs. are ok because their A_DIFF are very tidy
    DT_predL[, AREA_DIFF := EZGE_AREA - EZGA_AREA][AREA_DIFF < 0.5 & AREA_KOR > 1000,] # a lot but not to change
    rm(DT_predL)

    # prepare for output
    predictionLocations <- predictionLocations[,c("EZGE","EZGA","EZGTO","EZGE_AREA","EZGA_AREA","ID_GAUGE")]
    #View(predictionLocations@data)


  ##### prepare predictionLocations CH

    ### add variables to predictionLocations
    predictionLocations_CH$AREASQKM <- round(predictionLocations_CH$AREAKM2,digits = 2)
    predictionLocations_CH$AREA_KOR <- round(predictionLocations_CH$AREAKM2,digits = 2)
    predictionLocations_CH$EZGE <- as.integer(as.character(predictionLocations_CH$ID_MORE))
    predictionLocations_CH$EZGA <- as.integer(as.character(predictionLocations_CH$ID_MORE[match(predictionLocations_CH$ID_MORE, predictionLocations_CH$TO_ID_MORE)]))
    predictionLocations_CH$EZGTO <- as.integer(as.character(predictionLocations_CH$TO_ID_MORE))
    predictionLocations_CH$EZGE_AREA <- predictionLocations_CH$AREA_KOR
    for (i in predictionLocations_CH$ID_MORE) {
```

```
      #print(c(i))
      predictionLocations_CH$EZGA_AREA[predictionLocations_CH$ID_MORE == i] <- sum(predictionLocations_CH$AREA_KOR[predictionLocations_CH$TO_ID_MORE == i])
      #print(predictionLocations_CH$EZGA_AREA[predictionLocations_CH$ID_MORE == i])
      if (abs(predictionLocations_CH$EZGA_AREA[predictionLocations_CH$ID_MORE == i] - predictionLocations_CH$EZGE_AREA[predictionLocations_CH$ID_MORE == i]) < 0.5) {
        predictionLocations_CH$EZGA_AREA[predictionLocations_CH$ID_MORE == i] <- NA
      }
    }
    predictionLocations_CH$ID_GAUGE <- as.integer(as.character(STOBIMO_EZG$HZB_PEGEL1[match(predictionLocations_CH$ID_MORE, STOBIMO_EZG$ID_MORE)]))

    # analyse for outliers
    DT_predL <- as.data.table(predictionLocations_CH)
    DT_predL[,list(ID_MORE,TO_ID_MORE,EZGE,EZGA,EZGTO,EZGE_AREA,EZGA_AREA)]
    DT_predL[EZGE == EZGA,] # non
    DT_predL[EZGE_AREA < EZGA_AREA,] # non
    DT_predL[, AREA_DIFF := EZGE_AREA - EZGA_AREA][AREA_DIFF == 0 & AREA_KOR > 1,] # non
    DT_predL[, AREA_DIFF := EZGE_AREA - EZGA_AREA][AREA_DIFF < 0.5 & AREA_KOR > 1000,] # non
    rm(DT_predL)

    # prepare for output
    predictionLocations_CH <- predictionLocations_CH[,c("EZGE","EZGA","EZGTO","EZGE_AREA","EZGA_AREA","ID_GAUGE")]
    #predictionLocations_CH@data


  ##### prepare predictionLocations BY

    ### add variables to predictionLocations
    predictionLocations_BY$AREASQKM <- round(predictionLocations_BY$AREAKM2,digits = 2)
    predictionLocations_BY$AREA_KOR <- round(predictionLocations_BY$AREAKM2,digits = 2)
    predictionLocations_BY$EZGE <- as.integer(as.character(predictionLocations_BY$ID_MORE))
    predictionLocations_BY$EZGA <- as.integer(as.character(predictionLocations_BY$ID_MORE[match(predictionLocations_BY$ID_MORE, predictionLocations_BY$TO_ID_MORE)]))
    predictionLocations_BY$EZGTO <- as.integer(as.integer(predictionLocations_BY$TO_ID_MORE))
    predictionLocations_BY$EZGE_AREA <- predictionLocations_BY$AREA_KOR
    for (i in predictionLocations_BY$ID_MORE) {
      #print(c(i))
      predictionLocations_BY$EZGA_AREA[predictionLocations_BY$ID_MORE == i] <- sum(predictionLocations_BY$AREA_KOR[predictionLocations_BY$TO_ID_MORE == i])
      #print(predictionLocations_BY$EZGA_AREA[predictionLocations_BY$ID_MORE == i])
    }
    predictionLocations_BY$ID_GAUGE <- as.integer(as.character(STOBIMO_EZG$HZB_PEGEL1[match(predictionLocations_BY$ID_MORE, STOBIMO_EZG$ID_MORE)]))

    # analyse for outliers
    DT_predL <- as.data.table(predictionLocations_BY)
    DT_predL[EZGE_AREA < EZGA_AREA,] # non
    DT_predL[, AREA_DIFF := EZGE_AREA - EZGA_AREA][AREA_DIFF == 0 & AREA_KOR > 1,] # non
    DT_predL[, AREA_DIFF := EZGE_AREA - EZGA_AREA][AREA_DIFF < 10 ,] # a lot but not to change
    rm(DT_predL)

    # prepare for output
    predictionLocations_BY <- predictionLocations_BY[,c("EZGE","EZGA","EZGTO","EZGE_AREA","EZGA_AREA","ID_GAUGE")]
    #predictionLocations_BY@data
```

```
  ##### merge predictionLocations
    # change into sf-objects
    predictionLocations <- st_as_sf(predictionLocations)
    predictionLocations_BY <- st_as_sf(predictionLocations_BY)
    predictionLocations_CH <- st_as_sf(predictionLocations_CH)
    # merge
    predictionLocations_all <- rbind(predictionLocations_CH, predictionLocations_BY, predictionLocations)

    # get overlapping watersheds
    Watersheds_without_Inn <- read_sf(dsn="./data/prepareSPACIAL/data_input/Watersheds_without_Inn.shp")
    Watersheds_without_Inn_and_Rott <- read_sf(dsn="./data/prepareSPACIAL/data_input/Watersheds_without_Inn_and_Rott.shp")

    predictionLocations_all[predictionLocations_all$EZGE %in% Watersheds_without_Inn$EZGID,"geometry"] <-
    st_union(predictionLocations_all[predictionLocations_all$EZGE %in% Watersheds_without_Inn$EZGID,"geometry"],
    predictionLocations_all[predictionLocations_all$EZGE==70320,"geometry"],
    by_feature=T)

    predictionLocations_all[predictionLocations_all$EZGE %in% Watersheds_without_Inn_and_Rott$EZGID,"geometry"] <-
    st_union(predictionLocations_all[predictionLocations_all$EZGE %in% Watersheds_without_Inn_and_Rott$EZGID,"geometry"],
    predictionLocations_all[predictionLocations_all$EZGE==70495,"geometry"],
    by_feature=T)

    # remove holes
    predictionLocations_all <-st_remove_holes(predictionLocations_all)

    # order
    predictionLocations_all <- predictionLocations_all[order(predictionLocations_all$EZGE_AREA),]

    # remove unnecessary watershed (artifacts from the merging process)
    predictionLocations_all <- predictionLocations_all[!predictionLocations_all$EZGE %in% c(3, 7267),]

    # remove watershed from rhine river causing problems in rtop
    predictionLocations_all <- predictionLocations_all[!predictionLocations_all$EZGE %in% c(7709, 6520, 6505, 6489, 6470, 6450, 6428, 1253, 9),]

    # write
    write_sf(predictionLocations_all, "./data/prepareSPACIAL/data_output/predictionLocations_overlapping.shp")


  ##### prep for export
    predictionLocations <- as(predictionLocations_all, Class = "Spatial")

    rm(predictionLocations_all, predictionLocations_BY, predictionLocations_CH, i, Watersheds_without_Inn, Watersheds_without_Inn_and_Rott)


## prepare observations ######################################################

    ## subset predictionLocations to IWAG list (list of gauges with runoff data)
    IWAG_eHyd_Q <- as.data.table(read.csv2(file = "./raw_data/Jahresabfluesse/20200912_Jahresabfluesse_alle_Pegel_IWAG_updated.csv", header = T)) ## mean annual q [m3/s]
    gauges_dismiss <- setDT(read_excel("./data/Diversion_data/Diversion_data_table.xlsx", sheet = "DIV_obs_gauges_dismiss"))
    IWAG_eHyd_Q <- IWAG_eHyd_Q[!pegel_nr %in% gauges_dismiss$HZBR_NR, pegel_nr, by = "pegel_nr"] # 643 obs.
    IWAG_eHyd_Q <- unique(IWAG_eHyd_Q$pegel_nr)
    rm(gauges_dismiss)

    # subset to available data
    observations <- st_as_sf(predictionLocations[predictionLocations$ID_GAUGE %in%
```

```
    IWAG_eHyd_Q,]) # 489 obs.

    # add variables to observations
    DIV_obs <- setDT(read_excel("./data/Diversion_data/Diversion_data_table.xlsx",
    sheet = "DIV_obs"))
    DIV_obs$HZBR_NR <- as.integer(DIV_obs$HZBR_NR)
    observations$A_oro <- DIV_obs$A_oro[match(observations$ID_GAUGE, DIV_obs$HZBR_NR)]
    observations$A_eff <- DIV_obs$A_oro[match(observations$ID_GAUGE, DIV_obs$HZBR_NR)]

    # analyse for outliers
    DT_Obs <- as.data.table(observations)
    DT_Obs[abs((EZGE_AREA-A_oro)/EZGE_AREA*100) >5] # 4 obs. -> all are OK
    rm(DT_Obs)

    # prep for export
    observations <- as(observations, Class = "Spatial")
    rm(IWAG_eHyd_Q, DIV_obs)


## check spacial data CRS #####################################################

    if ((proj4string(predictionLocations) != proj4string(STOBIMO_EZG) |
        proj4string(predictionLocations) != proj4string(bord_AUT) |
        proj4string(predictionLocations) != proj4string(bord_AUT_BL) |
        proj4string(predictionLocations) != proj4string(bord_AUT_PB) |
        proj4string(predictionLocations) != proj4string(rnet_gauges) |
        proj4string(predictionLocations) != proj4string(rnet) ) == T) {
      stop("CRS have to be the same")
    }


## save image #################################################################

    save.image("./data/prepareSPACIAL/02_saveVAR_prepareSPACIAL_processed.RData")


## write shape Files ##########################################################

    writeOGR(obj = predictionLocations,
             dsn = "./data/prepareSPACIAL/data_output/predictionLocations.shp",
             layer = "predictionLocations",
             driver = "ESRI Shapefile",
             check_exists=TRUE, overwrite_layer= TRUE)

    writeOGR(obj = observations,
             dsn = "./data/prepareSPACIAL/data_output/observations.shp",
             layer = "observations",
             driver = "ESRI Shapefile",
             check_exists=TRUE, overwrite_layer= TRUE)


## End prepare spacial data ###################################################
```

C.2.3 Prepare MQ table

```r
##%#############################################%##
#                                                  #
#   Diploma Thesis                                 #
#   TopKriging prediction with                     #
#   diversion consideration                        #
#                                                  #
#   MQ table for rtop                              #
#   Creator:                                       #
#   nikolaus.weber@tuwien.ac.at                    #
#   Editor:                                        #
#   nikolaus.weber@tuwien.ac.at                    #
#   Last edit:                                     #
#   21.11.2020                                     #
#                                                  #
##%#############################################%##

## Libs ##########################################################################

    if (!require("data.table")) install.packages("data.table", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("dplyr")) install.packages("dplyr", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("readxl")) install.packages("readxl", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("hydroTSM")) install.packages("hydroTSM", dependencies = TRUE, repos="https://cloud.r-project.org/")

    library(data.table)
    library(dplyr)
    library(readxl)
    library(hydroTSM) # used for days per year


## MQ table for rtop ##############################################################

    ## load MQ_rnet_gauges (list of gauges with runoff data)
    MQ_rnet_gauges <- as.data.table(read.csv2(file = "./raw_data/Jahresabfluesse/20200921_Jahresabfluesse_alle_Pegel_IWAG_updated.csv", header = T)) ## mean annual q [m3/s]
    MQ_rnet_gauges <- MQ_rnet_gauges[MQ_m3_s >= 0, ]
    MQ_rnet_gauges <- MQ_rnet_gauges[,.(ID = pegel_nr, YEAR = jahr, obsDAYS = anzahl_werte, MQ = MQ_m3_s)]
    MQ_rnet_gauges <- MQ_rnet_gauges[YEAR >= 2009 & YEAR <= 2017,] # 5631 obs.
    MQ_rnet_gauges$YEAR <- as.numeric(MQ_rnet_gauges$YEAR)

    ## dismiss gauges
    gauges_dismiss <- setDT(read_excel("./data/Diversion_data/Diversion_data_table.xlsx", sheet = "DIV_obs_gauges_dismiss"))
    MQ_rnet_gauges <- MQ_rnet_gauges[!ID %in% gauges_dismiss$HZBR_NR,] # 5559 obs.

    ## check for gauges with not enough observations days per year
    table(MQ_rnet_gauges$obsDAYS)
    MQ_rnet_gauges[,.N, by=.(YEAR,obsDAYS)] %>% dplyr::arrange(YEAR, -obsDAYS)
    ## exclude gauges with not enough observations days per year
    MQ_rnet_gauges <- MQ_rnet_gauges[obsDAYS >= 356]

    ## add AREAS for disturbed gauges (A_oro & A_eff)
    DIV_obs <- setDT(read_excel("./data/Diversion_data/Diversion_data_table.xlsx", sheet = "DIV_obs"))
    # check for broad consistency
    DIV_obs[, A_gauges := rnet_gauges$Area[match(DIV_obs$HZBR_NR, rnet_gauges$ID)]]
    DIV_obs[ abs((A_oro - A_gauges)/A_gauges*100) > 2,] # 3 obs. checked
    # merge
    DIV_obs <- DIV_obs[, list(HZBR_NR, A_oro, A_eff)]
    MQ_rnet_gauges <- merge(MQ_rnet_gauges,DIV_obs,
                            by.x = "ID", by.y = "HZBR_NR",
                            all.x = T)

    # which gauge stations are not in observations
    outliers <- DIV_obs[!HZBR_NR %in% observations$ID_GAUGE]
    outliers[HZBR_NR %in% gauges_Q_AT] # 3 obs. checked
    outliers[!HZBR_NR %in% gauges_Q_AT] # 28 obs. checked(AT gauges are without Q,
```

```
    BY improvable if finer predictionLocations in BY)

    ## add AREAS for undisturbed gauges
    MQ_rnet_gauges[is.na(A_oro), A_oro :=
    rnet_gauges$Area[match(MQ_rnet_gauges[is.na(A_oro),]$ID, rnet_gauges$ID)]]
    MQ_rnet_gauges[is.na(A_eff), A_eff := A_oro]

    # check for not included gauges
    MQ_rnet_gauges[is.na(A_oro) | A_oro == 0,.(ID), by ="ID"] # 13 obs.checked

    # remove gauges with unknown area or zero area
    MQ_rnet_gauges <- MQ_rnet_gauges[!is.na(A_oro) & A_oro != 0,] # 5487 obs.

    rm(gauges_dismiss,DIV_obs,outliers)

    # write number of observations per year
    MQ_rnet_gauges[,.N, by=.(YEAR)] %>% dplyr::arrange(N)
    write.csv2(MQ_rnet_gauges[,.N,
    by=.(YEAR)],"./data/Diversion_data/Gauges_per_year.csv", row.names = F)


## calc the specific runoff ##################################################

    # get days per year
    YEARdays <- data.table(YEAR = 2009:2017,
                           YEARdays = unlist(lapply(2009:2017, FUN = hydroTSM::diy,
                           out.type = "nmbr")))

    # merge with table
    MQ_rnet_gauges[, YEARdays := YEARdays$YEARdays[match(MQ_rnet_gauges$YEAR,
    YEARdays$YEAR)]] # q = Q/A

    #specific runoff
    MQ_rnet_gauges[, `:=`(q_nat = MQ/A_oro, # q = Q/A
                          q_eff = MQ/A_eff)] # [m3/s/km2]

    # change unit to [mm]
    MQ_rnet_gauges[, `:=`(q_nat_mm = signif(q_nat*(3.6*24*YEARdays),digits = 3),
                          q_eff_mm = signif(q_eff*(3.6*24*YEARdays),digits = 3))]
                          #[mm/a]
    # (calc m3/s/km2 -> mm/a) [m3/s/km2 = 3.6*24*365.242 mm/a]
    # more exact with: hydroTSM::diy(YEAR, out.type = "nmbr")

    # View and check
    MQ_rnet_gauges[q_nat_mm > 2500, list(ID, A_oro, mean(q_nat_mm), mean(q_eff_mm))
    , by="ID"]


## available gauges different by year ########################################

    for (i in 2009:2017) {
      print(paste("Year", i))
      print(MQ_rnet_gauges[YEAR == i, .N])
    }
    rm(i)


## statistics ################################################################

    summary(MQ_rnet_gauges[, .(YEAR, q_nat_mm, q_eff_mm)])
    hist(MQ_rnet_gauges[, q_nat_mm], n = 30, ylim = c(0,20))
    hist(MQ_rnet_gauges[, q_eff_mm], n = 30, ylim = c(0,20))
    quantile(MQ_rnet_gauges$q_nat_mm, probs = c(0.01, 0.99))
    quantile(MQ_rnet_gauges$q_eff_mm, probs = c(0.01, 0.99))


## output ####################################################################

    write.csv2(MQ_rnet_gauges, "./data/prepareSPACIAL/MQ_table_gauges.csv",
    row.names = F)


## save image ################################################################
```

```
    save.image("./data/prepareSPACIAL/01_saveVAR_MQ_table.RData")


## End View & plot Data ######################################################
```

C.3 Child script: Processing & Post-Processing

C.3.1 Interpolation & Post-Processing

```r
##%##############################################%##
#                                                 #
#   Diploma Thesis                                #
#   TopKriging prediction with                    #
#   diversion consideration                       #
#                                                 #
#   rtop Interpolation                            #
#   Creator:                                      #
#   nikolaus.weber@tuwien.ac.at                   #
#   Editor:                                       #
#   nikolaus.weber@tuwien.ac.at                   #
#   Last edit:                                    #
#   01.12.2020                                    #
#                                                 #
##%##############################################%##


## Libs ######################################################################

if (!require("data.table")) install.packages("data.table", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("rtop")) install.packages("rtop", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("sp")) install.packages("sp", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("dplyr")) install.packages("dplyr", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("rgdal")) install.packages("rgdal", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("sf")) install.packages("sf", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("raster")) install.packages("raster", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("tmap")) install.packages("tmap", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("leaflet")) install.packages("leaflet", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("hydroGOF")) install.packages("hydroGOF", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("ggplot2")) install.packages("ggplot2", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("ggsci")) install.packages("ggsci", dependencies = TRUE, repos="https://cloud.r-project.org/")

library(data.table) # for fast & easy table handling
library(rtop) # for TopKriging
library(sp) # dependency of rtop
library(dplyr) # for easy data processing
library(rgdal) # for spatial projection/transformation operations
library(sf) # for easy handing of spatial objects
library(raster) # for raster objects
library(tmap) # for plotting thematic maps
library(leaflet) # for interactive thematic maps
library(hydroGOF) # for Statistics like NSE, ...
library(ggplot2) # for Plots
library(ggsci) # Color Scales for ColorBlind


## rtop interpolation function ###############################################

rtop_interp <- function(MQyear = 2009, Div = TRUE,
                        OL_Limit_Q = FALSE, OL_Limit = c(50, 2900))
{
# starting time
print(Sys.time())


## rtop interpolation ########################################################

# plotting parameter (Legend range [mm])
mm_at      = c(0,OL_Limit[1],100,seq(500,1500,200),2000,OL_Limit[2],8000) # for specific runoff [mm]
mm_col     = rev(bpy.colors(length(mm_at)))
```

```r
mm_at_var  = c(-10000,0,10000,50000,100000,500000) # for variance of specific runoff [mm]
mm_col_var = rev(bpy.colors(length(mm_at)))
tmap_mode("plot")

# rtop parameter
params = list(
gDist = TRUE,          # Use Ghosh-distance to reduce computation time
cloud = FALSE,         # logical; if TRUE use the variogram cloud, if FALSE use binned variogram
rresol = 25,           # Minimum number of discretization points in each element (area or line) (default = 25)
singularSolve = TRUE, # logical; set TRUE if kriging matrices are singular (when two or more areas being (almost) identical )
nclus = 1   # option to use parallel processing (number of workers for parallel processing) (library(parallel) & detectCores() )
)

# Create a column with the specific runoff:
MQ_tbl = MQ_rnet_gauges[YEAR == MQyear,] # subset MQ table
if (Div == T) {
  observations$obs      = MQ_tbl$q_eff_mm   [match(observations$ID_GAUGE, MQ_tbl$ID)] # add specific runoff [mm]
} else if (Div == F) {
  observations$obs      = MQ_tbl$q_nat_mm   [match(observations$ID_GAUGE, MQ_tbl$ID)] # add specific runoff [mm]
} else stop("STOP should be Div = T/F")

# Build an rTopObject
rtopObj = createRtopObject(observations[!is.na(observations$obs),], # subset to existing gauge data
                           predictionLocations,
                           formulaString = obs~1,
                           params = params)

# Fit a variogram (function also creates a sample Variogram
rtopObj = rtopFitVariogram(rtopObj)

# produce some diagnostic plots for the sample variogram and the fitted variogram model
#rtopObj = checkVario(rtopObj, cloud = T, identify = T, acor = 0.000001)
source("./analysis/2.1_plot_rtop_checkVario_2.0.R") # functions to plot rtop::checkVario
png(paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_01_plot_diagPlots_1.png"), width = 400, height = 300)
TK_checkVario.1(rtopObj, acor = 0.000001) # [1] plot dispersion variance
dev.off()
png(paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_01_plot_diagPlots_2.png"), width = 400, height = 250)
TK_checkVario.2(rtopObj, dcor = 0.001, cloud =  T, identify = F) # [2] plot Variogramm Cloud
dev.off()
png(paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_01_plot_diagPlots_3.png"), width = 400, height = 400)
TK_checkVario.3(rtopObj, acor = 0.000001) # [3] plot Variogramm Gamma
dev.off()
png(paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_01_plot_diagPlots_4.png"), width = 500, height = 420)
TK_checkVario.4(rtopObj, acor = 0.000001) # [4] plot VariogrammFit -> !!! try to adjust margins and remove plot question !!!!
dev.off()

print(summary(rtopObj$observations$obs))

# Cross-validation
rtopObj = rtopKrige(rtopObj,cv=TRUE)

# save Cross-Validation predictions
predCV = st_as_sf(rtopObj$predictions)
rtopObj$predictionsCV = rtopObj$predictions
## add scenario
predCV$MQyear <- MQyear
predCV$Div <- Div
```

```r
predCV$OL_Limit_Q <- OL_Limit_Q

# calc Cross-Validation Statistics
predCV_Stat = tibble(MQyear = MQyear,
                     Div = Div,
                     name =c("NSE", "mNSE"),
                     value = c(round(hydroGOF::NSE(sim = predCV$var1.pred, obs = predCV$obs),3), # NSE
                               round(hydroGOF::mNSE(sim = predCV$var1.pred, obs = predCV$obs, j=1),3)), # mNSE
                     Explanation = c("Nash-Sutcliffe efficiency",
                                     "Modified Nash-Sutcliffe efficiency (j=1)"))
print(predCV_Stat)

# save Outliers Limit
fwrite(predCV_Stat,
       paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_02_table_pred_CV_Stat.csv"),
       sep = ";", dec = ",")

# TopKriging (Predict at prediction locations)
rtopObj = rtopKrige(rtopObj)

# save TopKriging predictions
predTK = st_as_sf(rtopObj$predictions)
## add scenario
rtopObj$observations$MQyear <- MQyear
rtopObj$observations$Div <- Div
rtopObj$observations$OL_Limit_Q <- OL_Limit_Q
predTK$MQyear <- MQyear
predTK$Div <- Div
predTK$OL_Limit_Q <- OL_Limit_Q

# save Observations & predictions_CV & predictions_TK
fwrite(setDT(rtopObj$observations@data),
       paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_01_table_obs.csv"),
       sep = ";", dec = ",")
fwrite(setDT(st_drop_geometry(predCV)),
       paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_02_table_pred_CV.csv"),
       sep = ";", dec = ",")
fwrite(setDT(st_drop_geometry(predTK)),
       paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_02_table_pred_TK.csv"),
       sep = ";", dec = ",")

# plot Observations & predictions_CV & predictions_TK#
## Observations
tmap_save(
  tm_shape(arrange(st_as_sf(observations[!is.na(observations$obs),]),-EZGE_AREA)) +
    tm_polygons("obs", id = "obs", palette = mm_col, breaks = mm_at) +
    tm_layout(main.title =  paste0("Observations | MQyear=",MQyear," | Div=", Div),
              main.title.size = 1, legend.position = c("left" , "top"), legend.title.size = 0.9),
  paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_01_plot_obs.png"))
## CV predictions
tmap_save(
  tm_shape(arrange(predCV, -EZGE_AREA)) +
    tm_polygons("var1.pred", id = "var1.pred", palette = mm_col, breaks = mm_at) +
    tm_layout(main.title =  paste0("Predictions CrossValidation | MQyear=",MQyear," | Div=", Div),
              main.title.size = 1, legend.position = c("left" , "top"), legend.title.size = 0.9) +
    tm_credits(paste0("   NSE=",
                     round(hydroGOF::NSE(sim = predCV$var1.pred, obs = predCV$obs),2), # NSE
                     "\nmNSE=",
                     round(hydroGOF::mNSE(sim = predCV$var1.pred, obs = predCV$obs, j=1),2)), # mNSE)
               position = c("right", "BOTTOM"), size = 1.1),
  paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_02_plot_pred_CV.png"))
## CV Variance
tmap_save(
  tm_shape(arrange(predCV, -EZGE_AREA)) +
    tm_polygons("var1.var", id = "var1.var", palette = mm_col_var, breaks = mm_at_var) +
```

```
    tm_layout(main.title =  paste0("Predictions CV variance | MQyear=",MQyear," | Div=", Div),
              main.title.size = 1, legend.position = c("left" , "top"), legend.title.size = 0.9),
  paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_02_plot_pred_CV_var.png"), units = "cm", width = 14)
## CV residuals
tmap_save(
  tm_shape(arrange(predCV, -EZGE_AREA)) +
    tm_polygons("residual", id = "residual", palette = "div", breaks = c(-8000,-1000,-100,100,1000,8000)) +
    tm_layout(main.title =  paste0("Predictions CV residuals | MQyear=",MQyear," | Div=", Div),
              main.title.size = 1, legend.position = c("left" , "top"), legend.title.size = 0.9),
  paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_02_plot_pred_CV_res.png"), units = "cm", width = 14)
## TK predictions
tmap_save(
  tm_shape(arrange(predTK, -EZGE_AREA)) +
    tm_polygons("var1.pred", id = "var1.pred", palette = mm_col, breaks = mm_at) +
    tm_layout(main.title =  paste0("Predictions TopKriging | MQyear=",MQyear," | Div=", Div),
              main.title.size = 1, legend.position = c("left" , "top"), legend.title.size = 0.9),
  paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_02_plot_pred_TK.png"))
## TK Variance
tmap_save(
  tm_shape(arrange(predTK, -EZGE_AREA)) +
    tm_polygons("var1.var", id = "var1.var", palette = mm_col_var, breaks = mm_at_var) +
    tm_layout(main.title =  paste0("Predictions TK variance | MQyear=",MQyear," | Div=", Div),
              main.title.size = 1, legend.position = c("left" , "top"), legend.title.size = 0.9),
  paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_02_plot_pred_TK_var.png"), units = "cm", width = 14)


## Outliers ##################################################################

# add specific runoff [mm]
predTK_OR = predTK %>% mutate( q_TK_OR    = round(var1.pred,0),
                               qvar_TK_OR = round(var1.var ,0))

# count Outliers
Outliers = tibble(Outliers_Limit = OL_Limit,
                  N = c(data.table(predTK_OR)[q_TK_OR < OL_Limit[1] , .N],
                        data.table(predTK_OR)[q_TK_OR > OL_Limit[2] , .N]),
                  MinMax_q_pred_mm = c(min(predTK_OR$q_TK_OR, na.rm = T),
                                       max(predTK_OR$q_TK_OR, na.rm = T)),
                  MQyear = MQyear, Div = Div)
show(Outliers)

# save Outliers Limit & Statistic
fwrite(Outliers,
       paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_03_table_pred_TK_OR_outliers_Stat.csv"),
       sep = ";", dec = ",")

# save Outliers
fwrite(predTK_OR %>% filter(q_TK_OR < OL_Limit[1] | q_TK_OR > OL_Limit[2]) %>%
         st_drop_geometry () %>% arrange(q_TK_OR) %>% setDT(),
       paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_03_table_pred_TK_OR_outliers.csv"),
       sep = ";", dec = ",")

# plot only Outliers
predTK_OR = predTK_OR %>%
  mutate(Outliers = ifelse(q_TK_OR <= OL_Limit[1], paste0("<", OL_Limit[1]," mm"),
                           ifelse(q_TK_OR >= OL_Limit[2], paste0(">", OL_Limit[2]," mm"), NA)))
tmap_save(
```

```r
    tm_shape(arrange(predTK_OR, -EZGE_AREA)) +
      tm_polygons("Outliers", id = "Outliers", palette = c("firebrick4", "dodgerblue4"), lwd = 0.1, colorNA = NULL) +
      tm_layout(main.title =  paste0("Predictions TopKriging outliers only | MQyear=",MQyear," | Div=", Div),
                main.title.size = 1, legend.position = c("left" , "top"),
                legend.title.size = 0.9),
    paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_03_plot_pred_TK_OR_OL_only.png"), units = "cm", width = 14)

  # restrict Outliers
  if (OL_Limit_Q == T ) {
    predTK_OR = predTK_OR %>%
      mutate(q_TK_OR = replace(q_TK_OR, which(q_TK_OR < OL_Limit[1]), OL_Limit[1]),
             q_TK_OR = replace(q_TK_OR, which(q_TK_OR > OL_Limit[2]), OL_Limit[2]))

    # plot map after Outlier_restriction
    tmap_save(
      tm_shape(arrange(predTK_OR, -EZGE_AREA)) +
        tm_polygons("q_TK_OR", id = "q_TK_OR", palette = mm_col, breaks = mm_at) +
        tm_layout(main.title =  paste0("Predictions TopKriging outliers restricted | MQyear=",MQyear," | Div=", Div),
                  main.title.size = 1, legend.position = c("left" , "top"),
                  legend.title.size = 0.9),
      paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_03_plot_pred_TK_OR.png"))

    # save pred_TK_OR
    fwrite(predTK_OR %>% st_drop_geometry () %>% setDT(),
           paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_03_table_pred_TK_OR.csv"),
           sep = ";", dec = ",")
  }

  # write shape Files
  #st_write(obj = arrange(predTK_OR, -EZGE_AREA),
  #         dsn = paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_03_geo_pred_TK_OR.shp"),
  #         layer = "predTK_OR",
  #         driver = "ESRI Shapefile")


  ## Transform into raster #####################################################

  # settings
  raster_res = 2000 # raster resolution [m]

  # Create a raster from scratch using raster
  predTK_raster = raster(xmn = bbox(STOBIMO_EZG)[1,1]-10000,  # set minimum x coordinate
                         xmx = bbox(STOBIMO_EZG)[1,2]+20000,  # set maximum x coordinate
                         ymn = bbox(STOBIMO_EZG)[2,1]-10000,  # set minimum y coordinate
                         ymx = bbox(STOBIMO_EZG)[2,2]+20000,  # set maximum y coordinate
                         res = c(raster_res,raster_res),      # resolution in c(x,y) direction
                         crs = proj4string(STOBIMO_EZG))      # set Coordinate Reference System

  # Assign values from predTK_OR top Layer to raster
  predTK_raster = raster::rasterize(x = arrange(predTK_OR, EZGE_AREA),
                                    y = predTK_raster,
                                    field = "q_TK_OR",
                                    fun = 'first')
  summary(predTK_raster)

  # plot raster
  tmap_save(
    tm_shape(predTK_raster) +
      tm_raster("layer", palette = mm_col, breaks = mm_at, title = "q_TK_OR") +
      tm_layout(main.title =  paste0("Predictions TopKriging Raster | MQyear=",MQyear," | Div=", Div),
                main.title.size = 0.9, legend.position = c("left" , "top"),
                legend.title.size = 0.9),
```

```r
  paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_04_plot_pred_TK_raster.png"))

# extract the raster values to STOBIMO_EZG
predTK_MORE = raster::extract(y=STOBIMO_EZG,
                                  x=predTK_raster,
                                  fun=mean,
                                  na.rm=T,
                                  small=T,
                                  weights=T,
                                  cellnumbers=T,
                                  df = T)
summary(predTK_MORE)


## transform to STOBIMO_EZG ##################################################

## add results to STOBIMO_EZG spatial object
predTK_MORE = cbind(ID_MORE = STOBIMO_EZG$ID_MORE, q_TK_OR = predTK_MORE)
STOBIMO_EZG$q_mm_sim <- round(predTK_MORE$q_TK_OR.layer[match(STOBIMO_EZG$ID_MORE,
predTK_MORE$ID_MORE)],0)
#STOBIMO_EZG$MQ_m3_s <-
signif(STOBIMO_EZG$q_mm*STOBIMO_EZG$AREAKM2_korr/(3.6*24*YEARdays[YEAR == MQyear,
YEARdays]),3)
## add scenario
STOBIMO_EZG$MQyear <- MQyear
STOBIMO_EZG$Div <- Div
STOBIMO_EZG$OL_Limit_Q <- OL_Limit_Q

# plot STOBIMO_EZG
tmap_save(
  tm_shape(STOBIMO_EZG) +
    tm_polygons("q_mm_sim", id = "q_mm_sim", palette = mm_col, breaks = mm_at) +
    tm_layout(main.title =  paste0("Predictions TopKriging STOBIMO_EZG |
    MQyear=",MQyear," | Div=", Div),
              main.title.size = 0.9, legend.position = c("left" , "top"),
              legend.title.size = 0.9),
  paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_05_plot_pred_TK_STOBIMO_EZG.png"))

# save STOBIMO_EZG
fwrite(STOBIMO_EZG@data %>% setDT(),
       paste0("data/",saveDir,"/", MQyear,"_Div_",
       Div,"_05_table_pred_TK_STOBIMO_EZG.csv"),
       sep = ";", dec = ",")


## write process steps values ################################################

# compare process steps
if (OL_Limit_Q == T) {
  process_Steps <- rbind(data.table(Step = "obs",          q_mm =
  rtopObj$observations$obs),
                         data.table(Step = "pred_CV",      q_mm = predCV$var1.pred),
                         data.table(Step = "pred_TK",      q_mm = predTK$var1.pred),
                         data.table(Step = "pred_TK_OR",   q_mm = predTK_OR$q_TK_OR),
                         data.table(Step = "STOBIMO",      q_mm =
                         STOBIMO_EZG$q_mm_sim))
} else
  if (OL_Limit_Q == F) {
    process_Steps <- rbind(data.table(Step = "obs",          q_mm =
    rtopObj$observations$obs),
                           data.table(Step = "pred_CV",      q_mm = predCV$var1.pred),
                           data.table(Step = "pred_TK",      q_mm = predTK$var1.pred),
                           data.table(Step = "STOBIMO",      q_mm =
                           STOBIMO_EZG$q_mm_sim))
  }
## add scenario
process_Steps$MQyear <- MQyear
process_Steps$Div <- Div
process_Steps$OL_Limit_Q <- OL_Limit_Q

fwrite(process_Steps,
       paste0("data/",saveDir,"/", MQyear,"_Div_",
       Div,"_06_table_process_Steps_comparison.csv"),
```

```
        sep = ";", dec = ",")


## ending time
print(Sys.time())
}


## End rtop interpolation ####################################################
```

C.3.2 TK Diagnostic plots

```
##%#############################################%##
#                                                #
#   Diploma Thesis                               #
#   TopKriging prediction with                   #
#   diversion consideration                      #
#                                                #
#   diagnostic plots                             #
#   Creator:                                     #
#   nikolaus.weber@tuwien.ac.at                  #
#   Editor:                                      #
#   nikolaus.weber@tuwien.ac.at                  #
#   Last edit:                                   #
#   27.10.2020                                   #
#                                                #
##%#############################################%##

## Libs ###############################################################

if (!require("rtop")) install.packages("rtop", dependencies = TRUE,
repos="https://cloud.r-project.org/")

library(rtop) # for topKriging


## Libs ###############################################################

 # all of the following codes are original from the rtop-package (rtop::checkVario)
 # and are modified to standardized plotting of the diagnostic plots

## functions for plotting ####################################################

# Source: rtop:::checkVario.rtop
TK_checkVario.1 <- function (object, acor = 1, log = "xy", cloud = FALSE,
                             gDist = TRUE, acomp = NULL, curveSmooth = FALSE, params
                             = list(),
                             ...)
{
  params = getRtopParams(object$params, newPar = params, ...)
  dots = list(...)
  variogramModel = object$variogramModel
  sampleVariogram = object$variogram
  if (is.null(sampleVariogram))
    sampleVariogram = object$variogramCloud
  observations = object$observations
  formulaString = object$formulaString
  amul = object$params$amul
  varFit = object$varFit
  abins = TK_adfunc(NULL, observations, amul)
  observations$acl = findInterval(observations$area, abins)
  observations$n = 1
  obsvar = aggregate(observations@data[, as.character(formulaString[[2]])],
                     by = list(acl = observations$acl), FUN = var)
  obsvar$area = aggregate(observations$area, by = list(acl = observations$acl),
                          FUN = mean)[, 2] * acor
  obsvar$n = aggregate(observations$n, by = list(acl = observations$acl),
                       FUN = sum)[, 2]
  obsvar$n = obsvar$n/max(obsvar$n) * 20
  obsvar = obsvar[!is.na(obsvar$x), ]
  plot(obsvar$area, obsvar$x, xlab = "area", ylab = "variance",
       cex = sqrt(obsvar$n), pch = 16, log = log) # , ylim = c(1e+01,5e+07)
}

# Source: rtop:::checkVario.rtop
TK_checkVario.2 <- function (object, dcor = 1, log = "xy", cloud = FALSE,
                             gDist = TRUE, acomp = NULL, curveSmooth = FALSE, params
                             = list(),
                             ...)
{
  params = getRtopParams(object$params, newPar = params, ...)
  dots = list(...)
  sampleVariogram = object$variogram
  if (is.null(sampleVariogram))
    sampleVariogram = object$variogramCloud
```

```
  observations = object$observations
  formulaString = object$formulaString


  if (cloud | is(sampleVariogram, "rtopVariogramCloud")) {
    print("Creating cloud variogram; this might take some time")
    if (!is(sampleVariogram, "rtopVariogramCloud")) {
      if (!("variogramCloud" %in% names(object)))
        object$variogramCloud = rtopVariogram(observations,
                                              formulaString, params, cloud = TRUE)
      clvar = object$variogramCloud
    }
    else clvar = sampleVariogram
    if (gDist) {
      if (!("gdistObs" %in% names(object))) {
        if (!("dObs" %in% names(object)))
          object$dObs = rtopDisc(observations, params = params)
        dObs = object$dObs
        object$gDistObs = gDist(dObs, dObs, params = params)
      }
      gdists = object$gDistObs
      gDiag = diag(gdists)
      clvar$gdist = 0
      for (ic in 1:dim(clvar)[1]) {
        ia = clvar$acl1[ic]
        ib = clvar$acl2[ic]
        clvar$dist[ic] = gdists[ia, ib] - 0.5 * (gDiag[ia] +
                                                   gDiag[ib])
      }
    }
    clvar$np = clvar$ord
    if (!"identify" %in% names(dots) | !dev.interactive())
      dots$identify = FALSE
    cdots = which(names(dots) %in% names(formals(TK_rtopVariogramModel)))
    if (length(cdots) > 0)
      dots = dots[-cdots]
    clvar$dist = clvar$dist*dcor
    #par("ylog")
    print(plot(clvar, xlab = "distance", unlist(dots))) #, ylim = c(0.0,5e07) , log = "y"
    #print(clvar)
  }
}

# Source: rtop:::checkVario.rtop
TK_checkVario.3 <- function (object, acor = 1, log = "xy", cloud = FALSE,
                             gDist = TRUE, acomp = NULL, curveSmooth = FALSE, params = list(),
                             ...)
{
  sampleVariogram = object$variogram
  if (is.null(sampleVariogram))
    sampleVariogram = object$variogramCloud
  varFit = object$varFit

  if (!is.null(varFit) & is(sampleVariogram, "rtopVariogram")) {
    gammar = varFit[, c("np", "gamma", "gammar")]
    gammar$nnp = sqrt(gammar$np)/max(sqrt(gammar$np)) * 20
    gmax = max(gammar[, c("gamma", "gammar")])
    gmin = quantile(c(gammar$gammar, gammar$gamma), 0.05)
    nnp = 0
    plot(gammar ~ gamma, gammar,
         xlim = c(ifelse(length(grep("x", log)) > 0, gmin, 0), gmax), #xlim = c(1e+01,1e+08),
         ylim = c(ifelse(length(grep("x", log)) > 0, gmin, 0), gmax), #ylim = c(1e+01,1e+08),
         cex = sqrt(nnp), xlab = "gamma",
         ylab = "gamma regularized", log = log)
    abline(0, 1)
  }
  else if (!is.null(varFit) & is(sampleVariogram, "rtopVariogramCloud")) {
    gammar = varFit[, c("np", "gamma", "gammar")]
    gammar = gammar[order(gammar$gamma), ]
```

```
      ng = dim(gammar)[1]
      groups = ifelse(ng > 200, 20, ng/10)
      npg = ng/groups
      gammar$group = c(1:ng)%/%npg
      ngammar = aggregate(list(gamma = gammar$gamma, gammar = gammar$gammar),
                          by = list(gammar$group), FUN = mean)
      ngammar = cbind(ngammar, aggregate(list(gammav = gammar$gamma,
                                              gammarv = gammar$gammar), by =
                                              list(gammar$group),
                                         FUN = var))
      xmax = max(c(ngammar$gamma, ngammar$gammar))
      xmin = quantile(c(ngammar$gammar, ngammar$gamma), 0.05)
      nnp = 0
      plot(gammar ~ gamma, ngammar,
           xlab = "regularized gamma", ylab = "gamma",
           xlim = c(ifelse(length(grep("x", log)) > 0, xmin, 0), xmax), #xlim =
           c(1e+01,1e+08),
           ylim = c(ifelse(length(grep("x", log)) > 0, xmin, 0), xmax), #ylim =
           c(1e+01,1e+08),
           cex = sqrt(nnp),
           pch = 16, log = log)
      errorBar(ngammar$gammar, ngammar$gamma, upper = sqrt(ngammar$gammav))
      abline(0, 1)
    }
  }

  # Source: rtop::checkVario.rtopVariogramModel
  TK_rtopVariogramModel <- function (object, sampleVariogram = NULL, observations =
  NULL,
                                     areas = NULL, dists = NULL, acomp = NULL, params
                                     = list(),
                                     compVars = list(), acor = 1, log = "xy", legx =
                                     NULL,
                                     legy = NULL, plotNugg = TRUE, curveSmooth =
                                     FALSE, ...)
  {
    variogramModel = object
    params = getRtopParams(params, ...)

    if (is.null(areas))
      areas = params$amul
    if (is.null(dists))
      dists = params$dmul
    if (length(areas) == 1)
      areas = TK_adfunc(sampleVariogram, observations, areas)
    if (length(dists) == 1)
      dists = TK_dfunc(sampleVariogram, observations, dists)
    Srl = list()
    icomb = 0
    polylist = list()
    aavg = areas[1:(length(areas) - 1)]
    dists = c(0, dists)
    adists = dists[1:length(dists)]
    for (iarea in 1:(length(areas) - 1)) {
      area = mean(c(areas[iarea], areas[iarea + 1]))
      aavg[iarea] = area
      for (idist in 1:(length(dists))) {
        icomb = icomb + 1
        ddist = ifelse(idist == 1, 0, mean(c(dists[idist],
                                             dists[idist - 1])))
        if (iarea == 1)
          adists[idist] = ddist
        cs = sqrt(area)/2
        x1 = ddist - cs
        x2 = ddist + cs
        y1 = -cs
        y2 = cs
        boun = cbind(x = c(x1, x2, x2, x1, x1), y = c(y1,
                                                      y1, y2, y2, y1))
        polyBoun = Polygon(boun)
        Srl[[icomb]] = Polygons(list(polyBoun), ID = as.character(icomb))
      }
    }
```

```
    polys = SpatialPolygons(Srl)
    vmats = list()
    iplot = 0
    na = length(areas)
    if (is.null(acomp) | length(acomp) == 1) {
      if (is.null(acomp))
        acomp = 5
      if (!is.null(sampleVariogram) & is(sampleVariogram, "rtopVariogram")) {
        samp = aggregate(sampleVariogram$np, by = list(acl1 = sampleVariogram$acl1,
                                                       acl2 = sampleVariogram$acl2),
                                                       sum)
        if (acomp > dim(samp)[1])
          acomp = dim(samp)[1]
        acomp = samp[order(samp$x, decreasing = TRUE)[1:acomp],
                     1:2]
      }
      else {
        aacomp = expand.grid(a1 = c(1:(na - 1)), a2 = c(1:(na -
                                                              1)))
        aacomp = aacomp[aacomp$a1 >= aacomp$a2, ]
        if (acomp > dim(aacomp)[1])
          acomp = dim(aacomp)[1]
        acomp = aacomp[sample(dim(aacomp)[1], acomp), ]
      }
    }
    else {
      acomp = acomp[acomp$acl1 < length(areas) & acomp$acl2 <
                      length(areas), ]
    }
    vmats = matrix(0, ncol = length(dists), nrow = dim(acomp)[1])
    for (iplot in 1:dim(acomp)[1]) {
      i1 = acomp[iplot, 2]
      i2 = acomp[iplot, 1]
      ld = length(adists)
      poly1 = polys[unique(c((i1 - 1) * ld + 1, ((i2 - 1) *
                                                  ld + 1):(i2 * ld)))]
      lobject = createRtopObject(SpatialPolygonsDataFrame(poly1,
                                                          data = data.frame(obs =
                                                          c(1:length(poly1))),
                                                          match.ID = FALSE),
                                 params = params, formulaString = obs ~ 1)
      lobject$variogramModel = variogramModel
      nadists = adists
      if (i1 != i2)
        nadists = c(0, nadists)
      overlapObs = TK_findVarioOverlap(data.frame(a1 = poly1[1]@polygons[[1]]@area,
                                                  a2 = poly1[2]@polygons[[1]]@area,
                                                  dist = nadists))
      lobject$overlapObs = t(matrix(rep(overlapObs, ld + (i1 !=
                                                            i2)), ncol = ld + (i1 !=
                                                            i2)))
      vmat = varMat(lobject, cv = TRUE)$varMatObs
      if (i1 == i2) {
        vmats[iplot, 2:ld] = vmat[1, 2:ld]
      }
      else {
        vmats[iplot, ] = vmat[1, 2:(ld + 1)]
      }
    }
    if (inherits(sampleVariogram, "rtopVariogramCloud")) {
      xmin = min(sampleVariogram$dist)/1.3
    }
    else {
      xmin = min(sampleVariogram$dist[sampleVariogram$np >
                                        2]/1.3)
    }
    xmax = max(adists)
    pdists = 10^seq(log10(xmin), log10(xmax), length.out = 100)
    pvar = apply(as.matrix(pdists), 1, TK_varioEx, variogramModel = variogramModel) +
      ifelse(plotNugg, TK_nuggEx(1, variogramModel) * acor, 0)
    ymin = max(min(vmats[vmats > 0]), min(sampleVariogram$gamma))
    ymax = max(pvar)
    if (acor != 1) {
```

```
    Rver = R.Version()
    if (as.numeric(Rver$major) * 100 + as.numeric(Rver$minor) >=
        214) {
      xTicks = axTicks(1, c(xmin, xmax, 3), usr = c(log10(xmin),
                                                    log10(xmax)), log = TRUE,
                                                    nintLog = Inf)
    }
    else xTicks = axTicks(1, c(xmin, xmax, 3), usr = c(log10(xmin),
                                                       log10(xmax)), log = TRUE)
    xlabs = xTicks * sqrt(acor)
  }
  else {
    xTicks = NULL
    xlabs = TRUE
  }
  plot(pdists, pvar, ylim = c(ymin, ymax), xlim = c(xmin, xmax), #c(5e+02,5e+05)
       log = log, type = "l", col = "black", lwd = 2,
       ylab = "gamma", xlab = "distance", xaxt = "n")
  axis(1, at = xTicks, labels = xlabs)
  legende = list(text = "point", col = c("black"),
                 lty = c(1), pch = 16)
  bcols = bpy.colors(8) # Original: c("red", "blue", "green", "orange", "brown",
  "violet", "yellow", "salmon")
  cols1 = bcols[1:length(areas)]
  cols2 = bcols[1:dim(acomp)[1]]
  for (iplot in 1:dim(acomp)[1]) {
    i1 = acomp[iplot, 2]
    i2 = acomp[iplot, 1]
    ld = length(adists)
    if (i1 == i2) {
      lt = 1
      lcol = cols1[i1]
    }
    else {
      lt = 2
      lcol = cols2[iplot]
    }
    xx = adists
    yy = vmats[iplot, 1:ld]
    if (curveSmooth) {
      if (is.numeric(curveSmooth))
        df = curveSmooth
      else df = length(adists) - 3
      xx = sort(c(xx, seq(min(xx), max(xx), length.out = 1000)))
      yy = predict(smooth.spline(adists, yy, df = df),
                   xx)$y
    }
    lines(xx, yy, lty = lt, lwd = 2, col = lcol)
    legende$text = c(legende$text, paste(aavg[i1] * acor,
                                         "vs", aavg[i2] * acor))
    legende$col = c(legende$col, lcol)
    legende$lty = c(legende$lty, lt)
    if (!is.null(sampleVariogram) & is(sampleVariogram, "rtopVariogram")) {
      ppts = sampleVariogram[sampleVariogram$acl2 == i1 &
                             sampleVariogram$acl1 == i2, ]
      lpch = 16 + lt
      np = 0
      points(gamma ~ dist, ppts, col = lcol, pch = lpch,
             cex = sqrt(sqrt(np/max(sampleVariogram$np) *
                             60)))
      legende$pch = c(legende$pch, lpch)
    }
  }
  if (length(compVars) > 0) {
    for (ic in 1:length(compVars)) {
      cvar = compVars[ic]
      xx = adists
      if (curveSmooth)
        xx = sort(c(xx, seq(min(xx), max(xx), length.out = 1000)))
      clines = variogramLine(cvar[[1]], dist_vector = xx)
      lines(clines, lty = 3, lwd = 2, col = cols2[ic])
      legende$text = c(legende$text, names(cvar))
      legende$col = c(legende$col, cols2[ic])
```

```
      legende$lty = c(legende$lty, 3)
      legende$pch = c(legende$pch, 16)
    }
  }
  if (is.null(legx))
    legx = ifelse(length(grep("x", log)) > 0, (max(adists)/log(xmax/xmin,5)-1)*0.8, max(adists) * 0.50)
  if (is.null(legy))
    legy = ifelse(length(grep("y", log)) > 0, sqrt(ymin * ymax/1.5), ymax * 0.35)
  warn = options("warn")
  options(warn = -1)
  legend(legx, legy, legende$text, col = legende$col, lty = legende$lty, # xmax/5, 5e+04
         lwd = rep(2, length(legende$pch)), pch = legende$pch,
         merge = TRUE)
  checkVarioRes = list(vmats = rbind(vmats, pvar), acomp = acomp)
  options(warn = warn$warn)
  invisible(checkVarioRes)
}


## functions dependencies ####################################################

# Source: rtop:::checkVario.rtop
TK_checkVario.4 <- function (object, acor = 1, log = "xy", cloud = FALSE,
                             gDist = TRUE, acomp = NULL, curveSmooth = FALSE, params = list(),
                             ...)
{
  params = getRtopParams(object$params, newPar = params, ...)
  dots = list(...)
  variogramModel = object$variogramModel
  sampleVariogram = object$variogram
  if (is.null(sampleVariogram))
    sampleVariogram = object$variogramCloud
  observations = object$observations


  if (is.null(variogramModel)) {
    if (is.null(sampleVariogram))
      sampleVariogram = rtopVariogram(observations)
    TK_rtopVariogramModel(sampleVariogram, observations, params = params,
                          log = log, curveSmooth = curveSmooth, acomp = acomp,
                          ...)
  }
  else {
    if (is.null(sampleVariogram)) {
      object$checkVario = TK_rtopVariogramModel(object$variogramModel,
                                                observations = object$observations,
                                                params = params,
                                                acor = acor, log = log, curveSmooth = curveSmooth,
                                                acomp = acomp, ...)
    }
    else {
      object$checkVario = TK_rtopVariogramModel(object$variogramModel,
                                                sampleVariogram = sampleVariogram,
                                                observations = object$observations,
                                                params = params, acor = acor, log = log, curveSmooth = curveSmooth,
                                                acomp = acomp, ...)
    }
  }
}

# Source: rtop:::dfunc
TK_dfunc <- function (sampleVariogram, observations, dmul)
{
  if (is.null(sampleVariogram)) {
    dmax = sqrt(bbArea(bbox(observations)))/2
    dmin = min(dist(coordinates(observations)))
  }
  else if (is(sampleVariogram, "rtopVariogramCloud")) {
```

```
    dmax = max(sampleVariogram$dist)
    dmin = min(sampleVariogram$dist)
  }
  else {
    dmax = max(sampleVariogram$dist)
    dmin = min(sampleVariogram$dist[sampleVariogram$np >
                                          2])
  }
  if (dmin < dmax/1e+08)
    dmin = dmax/1e+08
  Rver = R.Version()
  if (as.numeric(Rver$major) * 100 + as.numeric(Rver$minor) >=
      214) {
    dists = axTicks(1, c(dmin/5, dmax * 10, dmul), usr = c(log10(dmin/5) -
                                                  1, log10(dmax) + 1),
                                                  log = TRUE, nintLog =
                                                  Inf)
  }
  else {
    dists = axTicks(1, c(dmin/5, dmax * 10, dmul), usr = c(log10(dmin/5) -
                                                  1, log10(dmax) + 2),
                                                  log = TRUE)
  }
  dists[(min(which(dists > dmin)) - 1):(max(which(dists < dmax)) +
                                            1)]
}

# Source: rtop:::adfunc
TK_adfunc <- function (sampleVariogram, observations, amul)
{
  if (is.null(sampleVariogram)) {
    if ("area" %in% names(observations)) {
      area = observations$area
    }
    else area = unlist(lapply(observations@polygons, FUN = function(poly) poly@area))
  }
  else area = c(sampleVariogram$a1, sampleVariogram$a2)
  amax = max(area)
  amin = min(area)
  Rver = R.Version()
  if (as.numeric(Rver$major) * 100 + as.numeric(Rver$minor) >=
      214) {
    areas = axTicks(1, c(amin/5, amax * 10, amul), usr = c(log10(amin/5) -
                                                  1, log10(amax) + 1),
                                                  log = TRUE, nintLog =
                                                  Inf)
  }
  else {
    areas = axTicks(1, c(amin/5, amax * 10, amul), usr = c(log10(amin/5) -
                                                  1, log10(amax) + 2),
                                                  log = TRUE)
  }
  areas = areas[(min(which(areas > amin)) - 1):(max(which(areas <
                                                  amax)) + 1)]
  areas
}

# Source: rtop:::findVarioOverlap
TK_findVarioOverlap <- function (vario)
{
  overlap = function(a1, a2, dist) {
    ad = sqrt(a1)/2
    ad[2] = sqrt(a2)/2
    if (ad[1] + ad[2] > dist) {
      Srl = list()
      for (i in 1:2) {
        pt1 = c(0, ifelse(i == 1, 0, dist))
        x1 = pt1[1] - ad[i]
        x2 = pt1[1] + ad[i]
        y1 = pt1[2] - ad[i]
        y2 = pt1[2] + ad[i]
        boun = data.frame(x = c(x1, x2, x2, x1, x1),
                          y = c(y1, y1, y2, y2, y1))
```

```
        Srl[[i]] = Polygon(SpatialPoints(boun))
      }
      cArea = TK_commonArea(Srl[[1]], Srl[[2]])
    }
    else cArea = 0
    cArea[[1]] * a1
  }
  mapply(FUN = overlap, vario$a1, vario$a2, vario$dist)
}

# Source: rtop:::commonArea
TK_commonArea <- function (objecti, objectj)
{
  bi = bbox(objecti)
  bj = bbox(objectj)
  iarea = TK_bbArea(bi)
  jarea = TK_bbArea(bj)
  sdim = sqrt((iarea + jarea)/2)
  bl = list()
  for (i in 1:2) bl[[i]] = max(bi[[i]], bj[[i]])
  for (i in 3:4) bl[[i]] = min(bi[[i]], bj[[i]])
  if (bl[[3]] >= bl[[1]] & bl[[4]] >= bl[[2]]) {
    larea = TK_bbArea(bl)
  }
  else {
    larea = 0
  }
  ilarea = larea/iarea
  jlarea = larea/jarea
  return(list(ilarea, jlarea))
}

# Source: rtop:::bbArea
TK_bbArea <- function (bb)
{
  xd = bb[[3]] - bb[[1]]
  yd = bb[[4]] - bb[[2]]
  abs(xd) * abs(yd)
}

# Source: rtop:::varioEx
TK_varioEx <- function (skor, variogramModel)
{
  model = variogramModel$model
  params = variogramModel$params
  res = 0
  imod = TK_imodel(model)
  vres = .Fortran("varioex", res, skor, length(params),
                  params, imod)
  return(vres[[1]])
}

# Source: rtop:::imodel
TK_imodel <- function (model)
{
  as.integer(switch(model, Exp = 1, Ex1 = 2, Gau = 3, Ga1 = 4,
                    Gho = 5, Sph = 6, Sp1 = 7, Fra = 8))
}

# Source: rtop:::nuggEx
TK_nuggEx <- function (ared, variogramModel)
{
  model = variogramModel$model
  params = variogramModel$params
  res = 0
  return(params[3] * ared)
}

## End ######################################################################
```

C.4 Child script: Validation, Comparison & Data Export

C.4.1 Calc splitting factor

```
##%############################################%##
#                                                #
#   Diploma Thesis                               #
#   TopKriging prediction with                   #
#   diversion consideration                      #
#                                                #
#   calc splitting factor                        #
#   Creator:                                     #
#   nikolaus.weber@tuwien.ac.at                  #
#   Editor:                                      #
#   nikolaus.weber@tuwien.ac.at                  #
#   Last edit:                                   #
#   02.12.2020 by Nikolaus Weber                 #
#                                                #
##%############################################%##


## Libs ####################################################################

    if (!require("data.table")) install.packages("data.table", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("dplyr")) install.packages("data.table", dependencies = TRUE, repos="https://cloud.r-project.org/")
    if (!require("readxl")) install.packages("readxl", dependencies = TRUE, repos="https://cloud.r-project.org/")

    library(data.table)
    library(dplyr)
    library(readxl)


## Set input parameter #####################################################

## Parameter
saveDir <-  "2020-12-01_STOBIMO_all_V46" # e.g. "2020-11-18_STOBIMO_2009_V43"


## Load data ################################################################

    # STOBIMO (MoRE) AU with runoff data
    STOBIMO <- setDT(readr::read_csv2(paste0("data/", saveDir, "/2009_Div_TRUE_05_table_pred_TK_STOBIMO_EZG.csv"), na = "NA") %>%
                       dplyr::select(ID_MORE, TO_ID_MORE, AREAKM2_korr))
      # or
      #STOBIMO <- tibble(STOBIMO_EZG@data) %>% dplyr::select(ID_MORE, TO_ID_MORE, TO_ID_2_MORE, q_mm) %>%
      #           dplyr::mutate(MQ_m3_s = signif(q_mm*AREAKM2_korr/(3.6*24*YEARdays[YEAR == MQyear, YEARdays]),3))

    # load updated flowtree:
    FlowTree <- setDT(readr::read_csv2("data/Diversion_data/MoRE_flow_tree_upd_2.csv", na = "NA"))

    # add colum to FlowTree
    FlowTree[bifurcation == "Verzweigung 2. Ordnung", FT_split := 2
    ][bifurcation == "Verzweigung 1. Ordnung", FT_split := 1
    ][bifurcation == "keine Verzweigung", FT_split := 0]

    # Diversion Areas for Splitting
    DIV_Area <- tibble(read_excel("./data/Diversion_data/Diversion_data_table.xlsx", sheet = "DIV_MORE", na = "NA"))


## prepare data #############################################################

    # Hyd_short (hydraulic short circuit UEB_ID_MORE is equal to upstream AU)
    hyd_short <- DIV_Area %>% dplyr::filter(hyd_short == "T") # 5 obs.
    DIV_Area <- DIV_Area %>% dplyr::filter(hyd_short == "F") # 69 obs.

    # subtract and add Area & Q for Hydraulic short circuit
    STOBIMO[, `:=`(AREAKM2_wHS = AREAKM2_korr)
    ][ # substract AREA_DIV & Q_DIV from giving Hyd_short AU
      ID_MORE %in% hyd_short$ID_MORE, `:=`(AREAKM2_wHS = AREAKM2_korr -
```

```
        hyd_short$A_Div[match(ID_MORE, hyd_short$ID_MORE)])
    ][ # add AREA_DIV & Q_DIV from receiving Hyd_short AU
      ID_MORE %in% hyd_short$TO_ID_2_MORE, `:=`(AREAKM2_wHS = AREAKM2_korr +
      hyd_short$A_Div[match(ID_MORE, hyd_short$TO_ID_2_MORE)])
    ]
    rm(hyd_short)

    # add AREA_DIV for all others AU
    STOBIMO[, AREA_DIV := DIV_Area$A_Div[match(ID_MORE, DIV_Area$ID_MORE)]
    ][ # replace NAs with 0
      is.na(AREA_DIV), AREA_DIV := 0]


## input for CalcFlowTree ####################################################

    # input
    input_vars <- STOBIMO[,.(from_id = ID_MORE, varAREA = AREAKM2_wHS, splitAREA =
    AREA_DIV)]
    #input_vars <- STOBIMO[,.(from_id = ID_MORE, varAREA = 1, splitAREA = 0)]

    ## find and write upstream AUs
    find_upstream_main <- function(id){
      ol <-paste(FlowTree[to_ID==id & FT_split %in% c(0,1),from_ID], collapse =";")
      return(ifelse(length(ol)==0L,NA,ol))
    }
    find_upstream_split <- function(id){
      ol <-paste(FlowTree[to_ID==id & FT_split == 2,from_ID], collapse =";")
      return(ifelse(length(ol)==0L,NA,ol))
    }
    input_vars$upstream_main <- sapply(input_vars$from_id, FUN= find_upstream_main)
    input_vars$upstream_split <- sapply(input_vars$from_id, FUN= find_upstream_split)


## Calculation of CalcFlowTree AREA ##########################################

    # Source of code: Steffen Kittlaus (skittlaus@iwag.tuwien.ac.at)

    ## Calculation of total AREA
    input_vars$resultAREA <- NULL
    input_vars$resultAREA <- numeric()
    for(loop in unique(FlowTree$calc_loop)){
      print(paste("Start Loop Nr.", loop))
      au_to_calculate <- FlowTree[calc_loop==loop,from_ID]
      #print(paste("AUs to calculate:",paste(au_to_calculate, collapse = ",")))
      for (i in au_to_calculate) {
        #print(paste("AU to calculate:",i))
        ## If there is no upstream AU, save value of variable as resultAREA:
        if(nchar(input_vars[from_id == i,upstream_main])==0L &
        nchar(input_vars[from_id == i,upstream_split])==0L){
                 input_vars[from_id == i]$resultAREA <- input_vars[from_id == i,
                 varAREA]
          #print("Headwater - no upstream!")
          }
        else{
          AU_upstream_main  <- unlist(strsplit(input_vars[from_id ==
          i,upstream_main], split=";"))
          AU_upstream_split <- unlist(strsplit(input_vars[from_id ==
          i,upstream_split],split=";"))
          if (length(AU_upstream_main)>0 & length(AU_upstream_split)>0) {
            input_vars[from_id ==i]$resultAREA <- sum(input_vars[from_id %in%
            AU_upstream_main,  .(A_split=resultAREA-splitAREA)],na.rm = T) +
                                                sum(input_vars[from_id %in%
                                                AU_upstream_split,
                                                .(A_split=splitAREA)],
                                                na.rm = T) +
                                                    input_vars[from_id ==i, varAREA]
          } else if (length(AU_upstream_main)>0) {
            input_vars[from_id ==i]$resultAREA <- sum(input_vars[from_id %in%
            AU_upstream_main,  .(A_split=resultAREA-splitAREA)],na.rm = T) +
                                                    input_vars[from_id ==i, varAREA]
          } else {
            input_vars[from_id ==i]$resultAREA <- sum(input_vars[from_id %in%
            AU_upstream_split, .(A_split=splitAREA)],           na.rm = T) +
```

```
                                                      input_vars[from_id ==i, varAREA]
          }
        }
        #print(paste("AU",i, "Calculated"))
      }
      #print(paste("End Loop Nr.", loop))
    }


## Calculation of SplittingFactor ############################################

    # SplittingFactor = RM_FCT_Q_SPLIT == SF_Q.Split)
    input_vars[, SF_Q.Split := splitAREA/resultAREA]

    # NAs
    #write.csv2(input_vars[is.na(SF_Q.Split), ], "Fehlende Gebiete im
    Abflussbaum.csv")
    #FlowTree[to_ID %in% input_vars[is.na(SF_Q.Split), ]$from_id]

    # statistics
    summary(input_vars$SF_Q.Split)
    summary(input_vars[SF_Q.Split > 0, SF_Q.Split]) # without zeros


## output ####################################################################

    # save SplittingFactor & upstream AUs
    readr::write_csv2(input_vars[,.(from_id, upstream_main, upstream_split,
    SF_Q.Split)],
                      "data/Diversion_data/STOBIMO_SF_Q.Split.csv")
    # save total AREA
    readr::write_csv2(input_vars[,.(from_id, splitAREA, resultAREA)],
                      "data/Diversion_data/STOBIMO_totalAREA.csv")


## End calc SplittingFactor ##################################################
```

C.4.2 Validation

```r
##%##############################################%##
#                                                 #
#   Diploma Thesis                                #
#   TopKriging prediction with                    #
#   diversion consideration                       #
#                                                 #
#   Validation                                    #
#   Creator:                                      #
#   nikolaus.weber@tuwien.ac.at                   #
#   Editor:                                       #
#   nikolaus.weber@tuwien.ac.at                   #
#   Last edit:                                    #
#   02.12.2020                                    #
#                                                 #
##%##############################################%##


## Libs ###################################################################

if (!require("data.table")) install.packages("data.table", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("hydroGOF")) install.packages("hydroGOF", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("dplyr")) install.packages("dplyr", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("readr")) install.packages("dplyr", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("readxl")) install.packages("readxl", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("hydroGOF")) install.packages("hydroGOF", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("ggplot2")) install.packages("ggplot2", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("ggsci")) install.packages("ggsci", dependencies = TRUE, repos="https://cloud.r-project.org/")

library(data.table) # for fast & easy table handling
library(hydroGOF)   # for Statistics like NSE, ...
library(dplyr)      # for easy data processing
library(readr)      # for easy data reading
library(readxl)     # for reading MS Excel files
library(ggplot2)    # for Plots
library(ggsci)      # Color Scales for ColorBlind


## Set input parameter #######################################################

## Parameter
saveDir <-  "2020-12-01_STOBIMO_all_V46" # e.g. "2020-11-18_STOBIMO_2009_V43"


## Load data #################################################################

## Load image
#load("./data/prepareSPACIAL/03_validation.RData")

# load observation runoff
DIV_MQ_obs <- setDT(read_excel("./data/Diversion_data/Diversion_data_table.xlsx", sheet = "DIV_MORE_q_long", na = "NA")) # Div table
MQ_rnet_gauges <- setDT(read.csv2("./data/prepareSPACIAL/MQ_table_gauges.csv")) # MQ table

# load diversions MORE
gauge_DIV <- setDT(read_excel("./data/Diversion_data/Diversion_data_table.xlsx", sheet = "DIV_MORE", na = "NA"))

# load diversions gauges
gauge_Obs <- setDT(read_excel("./data/Diversion_data/Diversion_data_table.xlsx", sheet = "DIV_obs", na = "NA"))

# load previous used data for MORE (STOBIMO)
STOBIMO_prev <- setDT(read_excel("raw_data/MoRE-Model/Export_Eingangsdaten_Abfluss.xlsx", na = "NA"))
```

```r

# load q sums for MORE
MQ_Div_obs_sum_list <- setDT(read_excel("./data/Diversion_data/Diversion_data_table.xlsx", sheet = "DIV_MORE_q_sum", na = "NA"))


## loop calculation ##########################################################

files <- list.files(path=paste0("data/",saveDir,"/"), pattern="table_pred_TK_STOBIMO_EZG.csv", full.names=T, recursive=FALSE, ignore.case=TRUE)
# file <- files[2]
for(file in files) {
  print(file)


## Post-processing ###########################################################

STOBIMO = fread(file, sep = ";", dec = ",")
STOBIMO = STOBIMO[,.(ID_MORE, TO_ID_MORE, TO_ID_2_MORE, Anteil_Ueberleitung,
               HZB_PEGEL1, HZB_PEGEL2, HZB_BEM, AREAKM2_korr, q_mm_sim,
               MQyear, Div)]

MQyear = STOBIMO$MQyear[1]
Div    = STOBIMO$Div[1]
STOBIMO[, YEARdays := YEARdays[YEAR %in% MQyear]$YEARdays]

# calculate runoff
STOBIMO[,MQ_sim := signif(q_mm_sim*AREAKM2_korr/(3.6*24*YEARdays),3)]

# reduce Diversion Area for PumpStorage and modify Hyd_short
if (Div == T) {
  # Diversion Areas for Splitting
  DIV_Area <- setDT(read_excel("./data/Diversion_data/Diversion_data_table.xlsx", sheet = "DIV_MORE", na = "NA"))

  # add Pumpstorage information
  STOBIMO[, PumpStorage := DIV_Area$PumpStorage[match(ID_MORE, DIV_Area$ID_MORE)]][is.na(PumpStorage), PumpStorage := "F"]
  # add "A_Div out of MQ" information
  STOBIMO[, A_Div_wQ    := DIV_Area$A_Div_wQ    [match(ID_MORE, DIV_Area$ID_MORE)]][is.na(A_Div_wQ),    A_Div_wQ    := "F"]
  # add "hydraulic shortage" information
  STOBIMO[, hyd_short   := DIV_Area$hyd_short   [match(ID_MORE, DIV_Area$ID_MORE)]][is.na(hyd_short),   hyd_short   := "F"]
  # add Type information
  STOBIMO[, Type        := DIV_Area$Typ         [match(ID_MORE, DIV_Area$ID_MORE)]][is.na(hyd_short),   Type        := NA]


  # add column of Div Information
  STOBIMO[A_Div_wQ == "T",    Div_Info := "MQ_to_AREA"]
  STOBIMO[is.na(Div_Info),    Div_Info := "DivAREA"]

  # Hyd_short (hydraulic short circuit UEB_ID_MORE is equal to upstream AU)
  HYD_short <- DIV_Area %>% dplyr::filter(hyd_short == "T") # 5 obs.
  DIV_Area  <- DIV_Area %>% dplyr::filter(hyd_short == "F") # 69 obs.

  # subtract and add Q for Hydraulic short circuit
  STOBIMO[, `:=`(MQ_sim_cor = MQ_sim)
  ][ # substract Q_DIV from giving Hyd_short AU
    ID_MORE %in% HYD_short$ID_MORE,
    `:=`(MQ_sim_cor     = MQ_sim - HYD_short$A_Div[match(ID_MORE, HYD_short$ID_MORE)] / AREAKM2_korr * MQ_sim,
         MQ_Div_sim_HS =          HYD_short$A_Div[match(ID_MORE, HYD_short$ID_MORE)] / AREAKM2_korr * MQ_sim)
  ][ # add Q_DIV from receiving Hyd_short AU
    ID_MORE %in% HYD_short$TO_ID_2_MORE,
    `:=`(MQ_sim_cor     = MQ_sim + HYD_short$A_Div[match(ID_MORE, HYD_short$TO_ID_2_MORE)] / AREAKM2_korr * MQ_sim)
  ]
  # add Diversion area
```

```
  STOBIMO[, A_Div := gauge_DIV$A_Div[match(ID_MORE, gauge_DIV$ID_MORE)]]

} else if (Div == F) {
  STOBIMO[, `:=`(MQ_sim_cor = MQ_sim)]
}

# save STOBIMO_EZG
fwrite(STOBIMO,
       paste0("data/",saveDir,"/", MQyear,"_Div_",
       Div,"_06_table_pred_TK_STOBIMO_EZG_cor.csv"),
       sep = ";", dec = ",")


## Validation #################################################################

# compare with previous used data
STOBIMO <- merge.data.table(STOBIMO,
                            STOBIMO_prev[,.(ID_MORE = FlächenId,
                                            YEAR    = Jahr,
                                            MQ_prev = Wert)][YEAR == MQyear,],
                            by = "ID_MORE", all.x = T)
STOBIMO[, q_mm_prev := round(MQ_prev / AREAKM2_korr *(3.6*24*YEARdays),digits = 0)]

# load updated flowtree:
FlowTree <- as.data.table(read_csv2("data/Diversion_data/MoRE_flow_tree_upd_2.csv",
                                    col_types = cols(step        = col_integer(),
                                                     from_ID     = col_integer(),
                                                     to_ID       = col_integer(),
                                                     bifurcation = col_character(),
                                                     calc_loop   = col_integer())))

FlowTree[bifurcation == "Verzweigung 2. Ordnung", FT_split := 2
][bifurcation == "Verzweigung 1. Ordnung",        FT_split := 1
][bifurcation == "keine Verzweigung",             FT_split := 0]

# load SplittingFactor & upstream AUs
SF_upAU <- setDT(readr::read_csv2("data/Diversion_data/STOBIMO_SF_Q.Split.csv"))


# add SplittingFactor and Upstream AUs to STOBIMO
if (Div == T) {
  STOBIMO[, Anteil_Ueberleitung := SF_upAU$SF_Q.Split    [match(ID_MORE,
  SF_upAU$from_id)]]
  STOBIMO[, upstream_main       := SF_upAU$upstream_main [match(ID_MORE,
  SF_upAU$from_id)]]
  STOBIMO[, upstream_split      := SF_upAU$upstream_split[match(ID_MORE,
  SF_upAU$from_id)]]
} else if (Div == F) {
  STOBIMO[, Anteil_Ueberleitung := 0L]
  STOBIMO[, upstream_main       := SF_upAU$upstream_main [match(ID_MORE,
  SF_upAU$from_id)]]
  STOBIMO[, upstream_split      := as.character("")]
}

# remove NAs
STOBIMO[is.na(upstream_main),  upstream_main  := as.character("")]
STOBIMO[is.na(upstream_split), upstream_split := as.character("")]

# function input
input_vars <- STOBIMO[,.(from_id = ID_MORE, var = MQ_sim_cor, split =
Anteil_Ueberleitung, upstream_main, upstream_split)]

## Calculation of total runoff
input_vars$result <-NULL
input_vars$result <-numeric()
for(loop in unique(FlowTree$calc_loop)){
  print(paste("Start Loop Nr.", loop,"/ 69"))
  au_to_calculate <- FlowTree[calc_loop==loop,from_ID]
  #print(paste("AUs to calculate:",paste(au_to_calculate, collapse = ",")))
  for (i in au_to_calculate) {
    #print(paste("AU to calculate:",i))
    ## If there is no upstream AU, save value of variable as result:
    if(nchar(input_vars[from_id == i,upstream_main])==0L & nchar(input_vars[from_id
```

```
    == i,upstream_split])==0L){
      input_vars[from_id == i]$result <- input_vars[from_id == i, var]
      #print("Headwater - no upstream!")
      }
    else{
      AU_upstream_main  <- unlist(strsplit(input_vars[from_id ==
      i,upstream_main],split=";"))
      AU_upstream_split <- unlist(strsplit(input_vars[from_id ==
      i,upstream_split],split=";"))
      if (length(AU_upstream_main)>0 & length(AU_upstream_split)>0) {
        input_vars[from_id ==i]$result <- sum(input_vars[from_id %in%
        AU_upstream_main,  .(q_split=result*(1-split))],na.rm = T)+
                                          sum(input_vars[from_id %in%
                                          AU_upstream_split,
                                          .(q_split=result*split)],    na.rm = T)+
                                          input_vars[from_id ==i, var]
      } else if (length(AU_upstream_main)>0) {
        input_vars[from_id ==i]$result <- sum(input_vars[from_id %in%
        AU_upstream_main,  .(q_split=result*(1-split))],na.rm = T)+
                                          input_vars[from_id ==i, var]
      } else {
        input_vars[from_id ==i]$result <- sum(input_vars[from_id %in%
        AU_upstream_split, .(q_split=result*split)],    na.rm = T)+
                                          input_vars[from_id ==i, var]
      }
    }
    #print(paste("AU",i, "Calculated"))
  }
  #print(paste("End Loop Nr.", loop))
}

#summary(input_vars$result)

# add to table
STOBIMO$MQ_tot_sim <- signif(input_vars$result[match(STOBIMO$ID_MORE,
input_vars$from_id)],3)

# calc natural runoff & diversion runoff
if (Div == T) {
  STOBIMO[, `:=`(MQ_eff_sim = MQ_tot_sim*(1-Anteil_Ueberleitung),
                 MQ_Div_sim = MQ_tot_sim*Anteil_Ueberleitung)]
} else if (Div == F) {
  STOBIMO[, `:=`(MQ_eff_sim = MQ_tot_sim,
                 MQ_Div_sim = 0)]
}

# add observation runoff
MQ_Div_obs <-  DIV_MQ_obs[YEAR == MQyear]
MQ_tbl  <- MQ_rnet_gauges[YEAR == MQyear,]
if (Div == T) {
  STOBIMO[, `:=`(MQ_eff_obs    = MQ_tbl$MQ          [match(HZB_PEGEL1, MQ_tbl$ID)],
                 Gauge_A_oro = MQ_tbl$A_oro     [match(HZB_PEGEL1, MQ_tbl$ID)],
                 MQ_Div_obs    = MQ_Div_obs$MQ_m3_s [match(ID_MORE,
                 MQ_Div_obs$ID_MORE)])]
} else if (Div == F) {
  STOBIMO[, `:=`(MQ_eff_obs    = MQ_tbl$MQ          [match(HZB_PEGEL1, MQ_tbl$ID)],
                 Gauge_A_oro = MQ_tbl$A_oro     [match(HZB_PEGEL1, MQ_tbl$ID)],
                 MQ_Div_obs    = 0)]
}

# workaround # removed gauges due to  non-compliance with the validation method
STOBIMO[ID_MORE == 12225, MQ_eff_obs := NA] #correction after GIS analysis (gauge
station: 205229   Ebensee (Unterlangbath))
STOBIMO[ID_MORE == 40065, MQ_eff_obs := NA] #correction after GIS analysis (gauge
station: 2319     Ova da Cluozza - Zernez)
STOBIMO[ID_MORE == 70055, MQ_eff_obs := NA] #correction after GIS analysis (gauge
station: 18226009     Miesbach / Schlierach)

# add information if gauge is affected
STOBIMO[HZB_PEGEL1 %in% gauge_Obs$HZBR_NR, Div_Bias := TRUE][is.na(Div_Bias),
Div_Bias := FALSE]

# subset diversions with only sum of observed runoff
```

```
if (Div == T) {
  STOBIMO[,`:=`(Div_sum = MQ_Div_obs_sum_list$group [match(ID_MORE, MQ_Div_obs_sum_list$ID_MORE)])]
  MQ_Div_obs_sum <- STOBIMO[!is.na(Div_sum),
                            .(MQ_Div_sum = sum(MQ_Div_sim)),
                            by = .(Div_sum)]
  STOBIMO[!is.na(Div_sum),
          `:=`(MQ_Div_sim = MQ_Div_obs_sum$MQ_Div_sum [match(Div_sum, MQ_Div_obs_sum$Div_sum)])]
  STOBIMO[ID_MORE %in% MQ_Div_obs_sum_list[first_of_group == "F", ID_MORE],
          `:=`(MQ_Div_sim = NA, MQ_Div_obs = NA)]
}

# gauge correction
STOBIMO[ID_MORE == 40040, # because gauge station Inn - S-Chanf 2462 is a Total runoff Station
        `:=`(MQ_eff_sim = MQ_eff_sim + MQ_Div_sim)]
STOBIMO[ID_MORE == 10980, # because sum of gauge station 208199 & 208157
        `:=`(MQ_eff_obs = MQ_tbl[ID %in% c(208199, 208157), sum(MQ)])]


# add simulated MQ_Div from Hydraulic shortage (HS)
if (Div == T) {
STOBIMO[!is.na(MQ_Div_sim_HS) , MQ_Div_sim := MQ_Div_sim_HS]
}

# add country information
STOBIMO[ID_MORE > 80000                 , Country := "Others"] # 6 obs.
STOBIMO[ID_MORE %between% c(70000,79999), Country := "DE"] # 106 obs.
STOBIMO[ID_MORE %between% c(60000,69999), Country := "Others"] # 4 obs.
STOBIMO[ID_MORE %between% c(50000,59999), Country := "Others"] # 1 obs.
STOBIMO[ID_MORE %between% c(40000,49999), Country := "CH"] # 24 obs.
STOBIMO[ID_MORE %between% c(10005,39999), Country := "AT"] # 754 obs.


# save table STOBIMO
fwrite(STOBIMO,
       paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_07_table_STOBIMO_MQs.csv"),
       sep = ";", dec = ",")

# save table efficiency comparison
if (Div == T) {
  fwrite(data.table(year = MQyear, Div = Div,
                    runoff = c("MQ_eff","MQ_eff","MQ_Div","MQ_Div"),
                    Stat = c("NSE","mNSE","NSE","mNSE"),
                    Value  = c(round(hydroGOF::NSE (sim = STOBIMO$MQ_eff_sim, obs = STOBIMO$MQ_eff_obs),4),
                               round(hydroGOF::mNSE(sim = STOBIMO$MQ_eff_sim, obs = STOBIMO$MQ_eff_obs),4),
                               round(hydroGOF::NSE (sim = STOBIMO$MQ_Div_sim, obs = STOBIMO$MQ_Div_obs),4),
                               round(hydroGOF::mNSE(sim = STOBIMO$MQ_Div_sim, obs = STOBIMO$MQ_Div_obs),4)),
                    Commend = c("Effective runoff comparison","Effective runoff comparison",
                                "Diversion runoff comparison","Diversion runoff comparison")),
         paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_07_table_MQ_Stat_comp.csv"),
         sep = ";", dec = ",")
} else if (Div == F) {
  fwrite(data.table(year = MQyear, Div = Div,
                    runoff = c("MQ_eff","MQ_eff"),
                    Stat = c("NSE","mNSE"),
                    Value  = c(round(hydroGOF::NSE (sim = STOBIMO$MQ_eff_sim, obs = STOBIMO$MQ_eff_obs),4),
                               round(hydroGOF::mNSE(sim = STOBIMO$MQ_eff_sim, obs = STOBIMO$MQ_eff_obs),4)),
                    Commend = c("Effective runoff comparison","Effective runoff comparison")),
         paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_07_table_MQ_Stat_comp.csv"),
         sep = ";", dec = ",")
```

```
}


## plot comparison with previous used data ###################################

ggplot(STOBIMO, aes(x = q_mm_prev, y = q_mm_sim)) +
  geom_point(alpha = 0.1) +
  geom_abline(intercept = c(0,0), slope = 1, size=0.3) +
  labs(title = paste0("STOBIMO watershed specific runoff\n sim to previous (MQyear=", MQyear,", Div=", Div,")")) +
  xlab("previous") + ylab("sim") + coord_fixed(ratio = 1) +
  scale_x_log10(limits = c(20, 3000)) + scale_y_log10(limits = c(20, 3000)) +
  scale_shape_manual(values = c(20)) + scale_color_npg() +
  theme( plot.title = element_text(size=9), aspect.ratio = 1,
         legend.text = element_text(size=9), legend.title = element_blank(),
         legend.position =  "none", legend.background = element_blank()) +
  annotate("text", x = 1000, y = 30,
           label = paste0("    NSE = ",round(hydroGOF::NSE(sim = STOBIMO$MQ_sim, obs = STOBIMO$MQ_prev),2),
                          "\n mNSE = ", round(hydroGOF::mNSE(sim = STOBIMO$MQ_sim, obs = STOBIMO$MQ_prev),2))) +
  ggsave(paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_07_plot_MQ_prev.png"), width = 4.1, height = 4.4)
}


## End Validation ############################################################
```

C.4.3 Result comparison

```
##%###############################################%##
#                                                   #
#   Diploma Thesis                                  #
#   TopKriging prediction with                      #
#   diversion consideration                         #
#                                                   #
#   Compare Validation results                      #
#   Creator:                                        #
#   nikolaus.weber@tuwien.ac.at                     #
#   Editor:                                         #
#   nikolaus.weber@tuwien.ac.at                     #
#   Last edit:                                      #
#   11.12.2020                                      #
#                                                   #
##%###############################################%##


## Libs ###################################################################

if (!require("data.table")) install.packages("data.table", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("hydroGOF")) install.packages("hydroGOF", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("dplyr")) install.packages("dplyr", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("readr")) install.packages("dplyr", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("readxl")) install.packages("readxl", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("hydroGOF")) install.packages("hydroGOF", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("ggplot2")) install.packages("ggplot2", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("ggpubr")) install.packages("ggpubr", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("ggsci")) install.packages("ggsci", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("gridExtra")) install.packages("gridExtra", dependencies = TRUE, repos="https://cloud.r-project.org/")

library(data.table) # for fast & easy table handling
library(hydroGOF)   # for Statistics like NSE, ...
library(dplyr)      # for easy data processing
library(readr)      # for easy data reading
library(readxl)     # for reading MS Excel files
library(ggplot2)    # for Plots
library(ggpubr)    # for Plots
library(ggsci)      # Color Scales for ColorBlind
library(gridExtra)  # to plot tables as images


## Load data ##############################################################

saveDir  <- "2020-12-01_STOBIMO_all_V46" # e.g. "2020-11-18_STOBIMO_2009_V43"

# load data
MQ_rnet_gauges <- fread("./data/prepareSPACIAL/MQ_table_gauges.csv", sep = ";", dec = ",")
# calc total runoff
totalQ <- MQ_rnet_gauges[,.(totalQ_10e9 = signif(sum(MQ*3600*24*YEARdays)/10^9,3)), by = .(YEAR)]
# load diversions gauges
gauge_Obs <- setDT(read_excel("./data/Diversion_data/Diversion_data_table.xlsx", sheet = "DIV_obs", na = "NA"))


## per year number of diversion affected gauges ##############################

data_all <- data.table(ID_GAUGE=integer(), MQyear=integer(), Div=logical())
files <- list.files(path=paste0("data/",saveDir,"/"), pattern="Div_TRUE_01_table_obs.csv", full.names=T, recursive=FALSE, ignore.case=TRUE)
for(file in files) {
  #file <- files [1]
```

```r
  data = fread(file, sep = ";", dec = ",")
  #ncol(data)
  #colnames(data)
  # bind data
  data_all = rbind(data_all,data[, .(ID_GAUGE, MQyear,Div)])
}
data_N <- data.table(MQyear = unique(data_all$MQyear),
           Total = data_all[, .N, by = .(MQyear)][,.(N)],
           DivBias = data_all[ID_GAUGE %in% gauge_Obs$HZBR_NR, .N, by =
           .(MQyear)][,.(N)])
data_N[, perc := round(DivBias.N/Total.N*100,0)]
data_N


## plot total runoff ##########################################################

# plot total runoff
ggplot(data = totalQ) +
  geom_col(aes(x =  YEAR, y = totalQ_10e9), position = "dodge") +
  scale_x_continuous(limits = c(2008.5,2017.5), breaks = 2009:2017) +
  scale_y_continuous(limits = c(0,1000)) +
  theme(strip.background =  element_rect(fill = "white", colour = "black"),
  legend.position =  "bottom",
        legend.text = element_text(size=9), legend.title = element_text(size=9),
        panel.grid.major.x = element_blank(), panel.grid.minor.x = element_blank()) +
  labs(title = paste("Annual total runoff per year")) + xlab(NULL) + ylab("annual
  total runoff in km³") +
  ggsave(paste0("data/",saveDir,"/All_09_plot_year_comparison_Qtot.png"), width =
  5.0, height = 3.2)


## compare Outliers ###########################################################

compare_OL <- F
if (compare_OL == T) {
  data_all <- data.table(Outliers_Limit=integer(), N=integer(),
  MinMax_q_DIFF_mm=numeric(), file =character())
  files <- list.files(path=paste0("data/",saveDir,"/"),
  pattern="_table_pred_TK_DIFF_outliers_Stat.csv", full.names=T, recursive=FALSE,
  ignore.case=TRUE)
  for(file in files) {
    print(file)
    data = fread(file, sep = ";", dec = ",")
    data[, file := file]
    # bind data
    data_all = rbind(data_all,data)
  }

  # plot table
  png(paste0("data/",saveDir,"/All_09_plot_year_comparison_OL.png"), height=810,
  width=320)
  p<-tableGrob(data_all[, .(Outliers_Limit, N, MinMax_q_DIFF_mm)])
  grid.arrange(p)
  dev.off()

  # plot statistic
  png(paste0("data/",saveDir,"/All_09_plot_year_comparison_OL_Stat.png"), height=80,
  width=250)
  p<-tableGrob(data_all[, .(OL_min = min(MinMax_q_DIFF_mm), OL_max =
  max(MinMax_q_DIFF_mm),
                            N_min  = min(N),                 N_max  = max(N))])
  grid.arrange(p)
  dev.off()

  # write table
  fwrite(data_all, paste0("data/",saveDir,"/All_09_table_year_comparison_OL.csv"),
  sep = ";", dec = ",")
  }


## per year compare process steps ############################################

data_all <- data.table(Step=character(), q_mm=numeric(),
```

```r
                             MQyear=integer(), Div=logical(),
                             OL_Limit_Q=logical())
files <- list.files(path=paste0("data/",saveDir,"/"), pattern="table_process_Steps_comparison.csv", full.names=T, recursive=FALSE, ignore.case=TRUE)
#file <- files[1]
for(file in files) {
  data = fread(file, sep = ";", dec = ",")

  n_Kategorie <-  paste0(levels(factor(data$Step)),"\nn = ",table(data$Step))
  MQyear <- data$MQyear[1]
  Div <- data$Div[1]

  ggplot(data, aes(x = Step, y = q_mm, fill = Step, color = Step)) +
    geom_violin(aes(fill = Step), color = "transparent", alpha = 0.70, width = 1.03) +
    geom_boxplot(outlier.alpha = 0.0, coef = 0,
                 color = "black", width = 0.15, size = 0.5) +
    theme(legend.position = "none") +
    scale_fill_npg() + scale_y_log10(breaks = breaks, minor_breaks = minor_breaks, limits = c(13, 7500)) +
    labs(title = paste0("Process steps comparison (MQyear=", MQyear,", Div=", Div,")")) +
    xlab(NULL) + ylab("specific runoff in [mm]") +
    scale_x_discrete(labels = n_Kategorie) +
    theme( plot.title = element_text(size=9), legend.position =  "none",
           axis.title.y = element_text(color = "grey20", size = 9)) +
    ggsave(paste0("data/",saveDir,"/", MQyear,"_Div_", Div,"_06_plot_process_OL_comparison_log_2.png"), width = 4.5, height = 4.0)

  # bind data
  data_all = rbind(data_all,data)

}

# combined process steps by Div for year 2009
data_all <- data_all[MQyear==2009]

n_Kategorie <-  paste0(levels(factor(data_all[Div==T]$Step)),"\nn = ",table(data_all[Div==T]$Step))
MQyear <- data_all$MQyear[1]
Div <- data_all$Div[1]

ggplot(data_all[], aes(x = Step, y = q_mm, fill = Step, color = Step)) +
  facet_grid(.~Div) +
  geom_violin(aes(fill = Step), color = "transparent", alpha = 0.70, width = 1.03) +
  geom_boxplot(outlier.alpha = 0.0, coef = 0,
               color = "black", width = 0.15, size = 0.5) +

  theme(legend.position = "none") +
  scale_fill_npg() + scale_y_log10(breaks = breaks, minor_breaks = minor_breaks, limits = c(13, 7500)) +
  labs(title = paste0("Process steps comparison by diversion consideration (MQyear=", MQyear,")")) +
  xlab(NULL) + ylab("specific runoff in [mm]") +
  scale_x_discrete(labels = n_Kategorie) +
  theme( plot.title = element_text(size=9), legend.position =  "none",
         axis.title.y = element_text(color = "grey20", size = 9)) +
  ggsave(paste0("data/",saveDir,"/", MQyear,"_Div_TRUE_FALSE_06_plot_process_OL_comparison_log_2_2009.png"), width = 8.5, height = 4.0)


## per year compare MQ_div and q_Div #########################################

data_all <- data.table(ID_MORE=integer(), HZB_PEGEL1=integer(), MQyear=integer(), Div=logical(),
                             MQ_Div_sim=numeric(), MQ_Div_obs=numeric(),
                             q_Div_obs=numeric(), q_Div_sim=numeric(),
                             Div_Info=character(), A_Div=numeric(), Type=character(),
                             Country=character())
files <- list.files(path=paste0("data/",saveDir,"/"), pattern="TRUE_07_table_STOBIMO_MQs.csv", full.names=T, recursive=FALSE, ignore.case=TRUE)
```

```r
for(file in files) {
  #file <- files [1]
  data = fread(file, sep = ";", dec = ",")
  #ncol(data)
  data[,`:=`(q_Div_obs = MQ_Div_obs/A_Div, q_Div_sim = MQ_Div_sim/A_Div)]
  #colnames(data)
  #plot MQ by category
  ggplot(data[!is.na(MQ_Div_obs)], aes(x = MQ_Div_obs, y = MQ_Div_sim, color =
  Div_Info)) +
    geom_point(size=0.8) + #facet_grid(.~Type) +
    geom_abline(intercept = c(0,0), slope = 1, size=0.3) +
    labs(title = paste0("Diversion runoff by Diversion Area transformation
    \n(MQyear=", data$MQyear[1],", Div=", data$Div[1],")")) +
    xlab("obs") + ylab("sim") + coord_fixed(ratio = 1) +
    scale_x_log10(limits = c(0.1, 100), minor_breaks=NULL) + scale_y_log10(limits =
    c(0.1, 100), minor_breaks=NULL) +
    scale_color_manual(values = c("red","grey18")) +
    scale_shape_manual(values = c(1,1)) + #scale_color_npg() +
    theme( plot.title = element_text(size=9), aspect.ratio = 1,
           legend.text = element_text(size=9), legend.title = element_blank(),
           legend.position =  c(.21,.89), legend.background = element_blank()) +
    annotate("text", x = 30, y = 0.17,
             label = paste0("    NSE = ", round(hydroGOF::NSE(sim = data$MQ_Div_sim,
             obs = data$MQ_Div_obs),2),
                            "\n mNSE = ", round(hydroGOF::mNSE(sim =
                            data$MQ_Div_sim, obs = data$MQ_Div_obs),2))) +
    ggsave(paste0("data/",saveDir,"/", data$MQyear[1],"_Div_",
    data$Div[1],"_07_plot_MQDiv_Stat.png"), width = 4.1, height = 4.4)

  #plot q by category
  ggplot(data[!is.na(MQ_Div_obs)], aes(x = q_Div_obs, y = q_Div_sim, color =
  Div_Info)) +
    geom_point(size=0.8) + #facet_grid(.~Type) +
    geom_abline(intercept = c(0,0), slope = 1, size=0.3) +
    labs(title = paste0("Diversion specific runoff by Diversion Area transformation
    \n(MQyear=", data$MQyear[1],", Div=", data$Div[1],")")) +
    xlab("obs") + ylab("sim") + coord_fixed(ratio = 1) +
    scale_x_continuous(limits = c(0.00, 1.00)) + scale_y_continuous(limits = c(0.00,
    1.00)) +
    scale_color_manual(values = c("red","grey18")) +
    scale_shape_manual(values = c(1,1)) + #scale_color_npg() +
    theme( plot.title = element_text(size=9), aspect.ratio = 1,
           legend.text = element_text(size=9), legend.title = element_blank(),
           legend.position =  c(.21,.89), legend.background = element_blank()) +
    ggsave(paste0("data/",saveDir,"/", data$MQyear[1],"_Div_",
    data$Div[1],"_07_plot_qDiv_Stat.png"), width = 4.1, height = 4.4)

  # bind data
  data_all = rbind(data_all,data[!is.na(MQ_Div_obs),
                               .(ID_MORE, HZB_PEGEL1, MQyear, Div, MQ_Div_sim,
                               MQ_Div_obs,
                                 q_Div_obs, q_Div_sim, Div_Info, A_Div, Type,
                                 Country)])
}


## per year compare MQ_eff runoff prediction ################################

data_all <- data.table(HZB_PEGEL1=integer(), MQyear=integer(), Div=logical(),
                       MQ_eff_sim=numeric(), MQ_eff_obs=numeric(), Div_Bias=logical())
files <- list.files(path=paste0("data/",saveDir,"/"),
pattern="_07_table_STOBIMO_MQs.csv", full.names=T, recursive=FALSE, ignore.case=TRUE)
for(file in files) {
  #file <- files [1]
  data = fread(file, sep = ";", dec = ",")
  #ncol(data)

  #colnames(data)
  # MQ Effective Comparison (natural runoff) all gauges
  ggplot(data[!is.na(MQ_eff_obs)], aes(x = MQ_eff_obs, y = MQ_eff_sim, color =
  Div_Bias)) +
    geom_point(size = 0.8, alpha = 1.0) +
    geom_abline(intercept = c(0,0), slope = 1, size=0.3) +
```

```
    scale_x_log10(limits = c(0.1, 10000)) + scale_y_log10(limits = c(0.1, 10000)) +
    scale_color_manual(values = c("red","grey18")) +
    scale_shape_manual(values = c(1,1)) +
    labs(title = paste0("Effective runoff by Div Affection(MQyear=",
    data$MQyear[1],", Div=", data$Div[1],")")) +
    xlab("obs") + ylab("sim") + coord_fixed(ratio = 1) +
    theme( plot.title = element_text(size=9), aspect.ratio = 1,
           legend.text = element_text(size=9), legend.title = element_text(size=9),
           legend.position =  c(.13,.87), legend.background = element_blank()) +
    annotate("text", x = 1000, y = 0.2,
             label = paste0("    NSE = ", round(hydroGOF::NSE(sim = data$MQ_eff_sim,
             obs = data$MQ_eff_obs),2),
                            "\n mNSE = ", round(hydroGOF::mNSE(sim =
                            data$MQ_eff_sim, obs = data$MQ_eff_obs),2))) +
    ggsave(paste0("data/",saveDir,"/",  data$MQyear[1],"_Div_",
    data$Div[1],"_07_plot_MQ_eff_Stat.png"), width = 4.4, height = 5.0)

  # MQ Effective Comparison (natural runoff) only DivBias
  data_DB <- data[!is.na(MQ_eff_obs) & Div_Bias == TRUE]
  ggplot(data_DB, aes(x = MQ_eff_obs, y = MQ_eff_sim)) +
    geom_point(size = 0.8, alpha = 1.0) +
    geom_abline(intercept = c(0,0), slope = 1, size=0.3) +
    scale_x_log10(limits = c(0.1, 10000)) + scale_y_log10(limits = c(0.1, 10000)) +
    #scale_color_manual(values = c("red","grey18")) +
    #scale_shape_manual(values = c(1,1)) +
    labs(title = paste0("Effective runoff by Div Affection(MQyear=",
    data_DB$MQyear[1],", Div=", data_DB$Div[1],")")) +
    xlab("obs") + ylab("sim") + coord_fixed(ratio = 1) +
    theme( plot.title = element_text(size=9), aspect.ratio = 1,
           legend.text = element_text(size=9), legend.title = element_text(size=9),
           legend.position =  c(.13,.87), legend.background = element_blank()) +
    annotate("text", x = 1000, y = 0.2,
             label = paste0("    NSE = ", round(hydroGOF::NSE(sim =
             data_DB$MQ_eff_sim, obs = data_DB$MQ_eff_obs),2),
                            "\n mNSE = ", round(hydroGOF::mNSE(sim =
                            data_DB$MQ_eff_sim, obs = data_DB$MQ_eff_obs),2))) +
    ggsave(paste0("data/",saveDir,"/",  data_DB$MQyear[1],"_Div_",
    data_DB$Div[1],"_07_plot_MQ_eff_Stat_onlyDivBias.png"), width = 4.4, height = 5.0)

  # identify the outliers
  # MQ
  #plot(data_DB$MQ_eff_obs,data_DB$MQ_eff_sim,  log="xy")
  #identify(data_DB$MQ_eff_obs,data_DB$MQ_eff_sim,  log="xy")


  # MQ Effective Comparison (natural runoff) by country
  ggplot(data[!is.na(MQ_eff_obs)], aes(x = MQ_eff_obs, y = MQ_eff_sim, color =
  Div_Bias)) +
    geom_point(size=0.8, alpha = 1.0) +  facet_grid(.~Country) +
    geom_abline(intercept = c(0,0), slope = 1, size=0.3) +
    scale_x_log10(limits = c(0.1, 10000)) + scale_y_log10(limits = c(0.1, 10000)) +
    labs(title = paste0("Effective runoff by country (MQyear=", data$MQyear[1],",
    Div=", data$Div[1],")")) +
    xlab("obs") + ylab("sim") + coord_fixed(ratio = 1) +
    scale_color_manual(values = c("red","grey18")) +
    scale_shape_manual(values = c(1,1)) +
    theme( plot.title = element_text(size=9), aspect.ratio = 1,
           legend.text = element_text(size=9), legend.title = element_text(size=9),
           legend.position =  c(.81,.79), legend.background = element_blank()) +
    ggsave(paste0("data/",saveDir,"/",  data$MQyear[1],"_Div_",
    data$Div[1],"_07_plot_MQ_eff_Stat_2.png"), width = 8.1, height = 3.1)


  # bind data
  data_all = rbind(data_all,data[!is.na(MQ_eff_obs), .(HZB_PEGEL1, MQyear, Div,
  MQ_eff_sim, MQ_eff_obs, Div_Bias)])
}

## all years compare CV specific runoff prediction ###########################

data_all <- data.table(MQyear=integer(), Div=logical(), ID_GAUGE=integer(),
                       obs=numeric(), CV_pred=numeric())
files <- list.files(path=paste0("data/",saveDir,"/"), pattern="_02_table_pred_CV",
```

```
full.names=T, recursive=FALSE, ignore.case=TRUE)
files <- files[!files %like% "_Stat"]
for(file in files) {
  data = fread(file, sep = ";", dec = ",")
  # bind data
  data_all = rbind(data_all,data[,.(MQyear, Div, ID_GAUGE, obs, CV_pred = var1.pred)])
}

# plot CV comparison all gauges
ggplot(data = melt(data_all[,.(NSE  = round(hydroGOF::NSE (sim = CV_pred, obs =
obs),2),
                               mNSE = round(hydroGOF::mNSE(sim = CV_pred, obs =
                               obs),2)),
                            by = .(MQyear,Div)],
                   id.vars = c("MQyear", "Div"), variable.name = "Stat")) +
  geom_col(aes(x =  MQyear, y = value, fill = Div), position = "dodge") +
  scale_x_continuous(limits = c(2008.5,2017.5), breaks = 2009:2017) + #
  scale_y_continuous(limits = c(0,1)) +
  facet_grid(.~Stat) + coord_cartesian(ylim = c(0.0,1.0)) +
  theme(strip.background =  element_rect(fill = "white", colour = "black"),
  legend.position =  "bottom",
        legend.text = element_text(size=9), legend.title = element_text(size=9),
        panel.grid.major.x = element_blank(), panel.grid.minor.x = element_blank()) +
  labs(title = paste("Cross-Validation comparison - Model efficiency coefficient"))
  + xlab(NULL) + ylab(NULL) +
  ggsave(paste0("data/",saveDir,"/All_09_plot_year_comparison_CV.png"), width = 8.5,
  height = 4.7)

# write table
fwrite(dcast(data_all[,.(NSE  = round(hydroGOF::NSE (sim = CV_pred, obs = obs),2),
                         mNSE = round(hydroGOF::mNSE(sim = CV_pred, obs = obs),2),
                         N    = .N),
                      by = .(MQyear,Div)],
             MQyear ~ Div,
             value.var = c("NSE", "mNSE", "N")),
       paste0("data/",saveDir,"/All_09_table_year_comparison_CV.csv"), sep = ";",
       dec = ",")

# plot CV comparison onlyDivBias gauges
ggplot(data = melt(data_all[ID_GAUGE %in% gauge_Obs$HZBR_NR,
                            .(NSE  = round(hydroGOF::NSE (sim = CV_pred, obs =
                            obs),2),
                              mNSE = round(hydroGOF::mNSE(sim = CV_pred, obs =
                              obs),2)),
                            by = .(MQyear,Div)],
                   id.vars = c("MQyear", "Div"), variable.name = "Stat")) +
  geom_col(aes(x =  MQyear, y = value, fill = Div), position = "dodge") +
  scale_x_continuous(limits = c(2008.5,2017.5), breaks = 2009:2017) + #
  scale_y_continuous(limits = c(0,1)) +
  facet_grid(.~Stat) + coord_cartesian(ylim = c(0.0,1.0)) +
  theme(strip.background =  element_rect(fill = "white", colour = "black"),
  legend.position =  "bottom",
        legend.text = element_text(size=9), legend.title = element_text(size=9),
        panel.grid.major.x = element_blank(), panel.grid.minor.x = element_blank()) +
  labs(title = paste("Cross-Validation comparison - Model efficiency coefficient -
  only diversion affected gauges")) + xlab(NULL) + ylab(NULL) +
  ggsave(paste0("data/",saveDir,"/All_09_plot_year_comparison_CV_onlyDivBias.png"),
  width = 8.5, height = 4.7)

# write table onlyDivBias
fwrite(dcast(data_all[ID_GAUGE %in% gauge_Obs$HZBR_NR,
                      .(NSE  = round(hydroGOF::NSE (sim = CV_pred, obs = obs),2),
                        mNSE = round(hydroGOF::mNSE(sim = CV_pred, obs = obs),2),
                        N    = .N),
                      by = .(MQyear,Div)],
             MQyear ~ Div,
             value.var = c("NSE", "mNSE","N")),
       paste0("data/",saveDir,"/All_09_table_year_comparison_CV_onlyDivBias.csv"),
       sep = ";", dec = ",")

## all years compare MQ_eff runoff prediction ################################

data_all <- data.table(HZB_PEGEL1=integer(), MQyear=integer(), Div=logical(),
```

```
                    MQ_eff_sim=numeric(), MQ_eff_obs=numeric(), Div_Bias=logical())
files <- list.files(path=paste0("data/",saveDir,"/"),
pattern="_07_table_STOBIMO_MQs.csv", full.names=T, recursive=FALSE, ignore.case=TRUE)
for(file in files) {
  #file <- files [1]
  data = fread(file, sep = ";", dec = ",")
  #ncol(data)
  #colnames(data)
  # bind data
  data_all = rbind(data_all,data[!is.na(MQ_eff_obs), .(HZB_PEGEL1, MQyear, Div,
  MQ_eff_sim, MQ_eff_obs, Div_Bias)])
}

# plot MQeff prediction comparison all gauges
ggplot(data = melt(data_all[,.(NSE  = round(hydroGOF::NSE (sim = MQ_eff_sim, obs =
MQ_eff_obs),2),
                              mNSE = round(hydroGOF::mNSE(sim = MQ_eff_sim, obs =
                              MQ_eff_obs),2)),
                           by = .(MQyear,Div)],
                  id.vars = c("MQyear", "Div"), variable.name = "Stat")) +
  geom_col(aes(x =  MQyear, y = value, fill = Div), position = "dodge") +
  scale_x_continuous(limits = c(2008.5,2017.5), breaks = 2009:2017) + #
  scale_y_continuous(limits = c(0,1)) +
  facet_grid(.~Stat) + coord_cartesian(ylim = c(0.0,1.0)) +
  theme(strip.background =  element_rect(fill = "white", colour = "black"),
  legend.position =  "bottom",
        legend.text = element_text(size=9), legend.title = element_text(size=9),
        panel.grid.major.x = element_blank(), panel.grid.minor.x = element_blank()) +
  labs(title = paste("MQeff prediction comparison - Model efficiency coefficient"))
  + xlab(NULL) + ylab(NULL) +
  ggsave(paste0("data/",saveDir,"/All_09_plot_year_comparison_MQeff.png"), width =
  8.5, height = 4.7)

# write table All
fwrite(dcast(data_all[,.(NSE  = round(hydroGOF::NSE (sim = MQ_eff_sim, obs =
MQ_eff_obs),2),
                         mNSE = round(hydroGOF::mNSE(sim = MQ_eff_sim, obs =
                         MQ_eff_obs),2),
                         N    = .N),
                      by = .(MQyear,Div)],
             MQyear ~ Div,
             value.var = c("NSE", "mNSE","N")),
       paste0("data/",saveDir,"/All_09_table_year_comparison_MQeff.csv"), sep = ";",
       dec = ",")

# plot MQeff prediction comparison Div_Bias == TRUE
ggplot(data = melt(data_all[Div_Bias == TRUE,
                            .(NSE  = round(hydroGOF::NSE (sim = MQ_eff_sim, obs =
                            MQ_eff_obs),2),
                              mNSE = round(hydroGOF::mNSE(sim = MQ_eff_sim, obs =
                              MQ_eff_obs),2)),
                            by = .(MQyear,Div)],
                  id.vars = c("MQyear", "Div"), variable.name = "Stat")) +
  geom_col(aes(x =  MQyear, y = value, fill = Div), position = "dodge") +
  scale_x_continuous(limits = c(2008.5,2017.5), breaks = 2009:2017) + #
  scale_y_continuous(limits = c(0,1)) +
  facet_grid(.~Stat) + coord_cartesian(ylim = c(0.0,1.0)) +
  theme(strip.background =  element_rect(fill = "white", colour = "black"),
  legend.position =  "bottom",
        legend.text = element_text(size=9), legend.title = element_text(size=9),
        panel.grid.major.x = element_blank(), panel.grid.minor.x = element_blank()) +
  labs(title = paste("MQeff prediction comparison - Model efficiency coefficient -
  only diversion affected gauges")) + xlab(NULL) + ylab(NULL) +

  ggsave(paste0("data/",saveDir,"/All_09_plot_year_comparison_MQeff_onlyDivBias.png"),
   width = 8.5, height = 4.7)


# write table Div_Bias == TRUE
fwrite(dcast(data_all[Div_Bias == TRUE,
                      .(NSE  = round(hydroGOF::NSE (sim = MQ_eff_sim, obs =
                      MQ_eff_obs),2),
                        mNSE = round(hydroGOF::mNSE(sim = MQ_eff_sim, obs =
```

```
                      MQ_eff_obs),2),
                      N     = .N),
                    by = .(MQyear,Div)],
             MQyear ~ Div,
             value.var = c("NSE", "mNSE","N")),

       paste0("data/",saveDir,"/All_09_table_year_comparison_MQeff_onlyDivBias.csv"),
       sep = ";", dec = ",")

# MQ Effective Comparison (natural runoff) per year ALL gauges
## Div TRUE
ggplot(data_all[!is.na(MQ_eff_obs) & Div == TRUE], aes(x = MQ_eff_obs, y =
MQ_eff_sim, color = Div_Bias)) +
  geom_point(size = 0.8, alpha = 1.0) + facet_wrap("MQyear", ncol = 3) +
  geom_abline(intercept = c(0,0), slope = 1, size=0.3) +
  scale_x_log10(limits = c(0.1, 10000), minor_breaks = NULL) + scale_y_log10(limits
  = c(0.1, 10000), minor_breaks = NULL) +
  scale_color_manual(values = c("red","black")) +
  #scale_shape_manual(values = c(1,1)) +
  labs(title = paste0("Effective runoff by MQyear by Div Affection (Div=",
  data_all[Div == TRUE]$Div[1],")")) +
  xlab("obs") + ylab("sim") + coord_fixed(ratio = 1) +
  theme( plot.title = element_text(size=9), aspect.ratio = 1,
         legend.text = element_text(size=9), legend.title = element_text(size=9),
         legend.position =  c(.05,.94), legend.background = element_blank()) +

  ggsave(paste0("data/",saveDir,"/All_09_plot_year_comparison_MQ_eff_Div_TRUE_all.png"
  ), width = 8.1, height = 8.4)
## Div FALSE
ggplot(data_all[!is.na(MQ_eff_obs) & Div == FALSE], aes(x = MQ_eff_obs, y =
MQ_eff_sim, color = Div_Bias)) +
  geom_point(size = 0.8, alpha = 1.0) + facet_wrap("MQyear", ncol = 3) +
  geom_abline(intercept = c(0,0), slope = 1, size=0.3) +
  scale_x_log10(limits = c(0.1, 10000), minor_breaks = NULL) + scale_y_log10(limits
  = c(0.1, 10000), minor_breaks = NULL) +
  scale_color_manual(values = c("red","black")) +
  #scale_shape_manual(values = c(1,1)) +
  labs(title = paste0("Effective runoff by MQyear by Div Affection (Div=",
  data_all[Div == FALSE]$Div[1],")")) +
  xlab("obs") + ylab("sim") + coord_fixed(ratio = 1) +
  theme( plot.title = element_text(size=9), aspect.ratio = 1,
         legend.text = element_text(size=9), legend.title = element_text(size=9),
         legend.position =  c(.05,.94), legend.background = element_blank()) +

  ggsave(paste0("data/",saveDir,"/All_09_plot_year_comparison_MQ_eff_Div_FALSE_all.png
  "), width = 8.1, height = 8.4)

# MQ Effective Comparison (natural runoff) per year only Div Affected gauges
## Div TRUE
ggplot(data_all[!is.na(MQ_eff_obs) & Div_Bias == TRUE & Div == TRUE], aes(x =
MQ_eff_obs, y = MQ_eff_sim)) +
  geom_point(size = 0.8, alpha = 1.0) + facet_wrap("MQyear", ncol = 3) +
  geom_abline(intercept = c(0,0), slope = 1, size=0.3) +
  scale_x_log10(limits = c(0.1, 10000), minor_breaks = NULL) + scale_y_log10(limits
  = c(0.1, 10000), minor_breaks = NULL) +
  #scale_color_manual(values = c("red","grey18")) +
  #scale_shape_manual(values = c(1,1)) +
  labs(title = paste0("Effective runoff by MQyear - only Div Affected gauges (Div=",
  data_all[Div == TRUE]$Div[1],")")) +
  xlab("obs") + ylab("sim") + coord_fixed(ratio = 1) +
  theme( plot.title = element_text(size=9), aspect.ratio = 1,
         legend.text = element_text(size=9), legend.title = element_text(size=9),
         legend.position =  c(.13,.87), legend.background = element_blank()) +

  ggsave(paste0("data/",saveDir,"/All_09_plot_year_comparison_MQ_eff_Div_TRUE_onlyDivB
  ias.png"), width = 8.1, height = 8.4)
## Div FALSE
ggplot(data_all[!is.na(MQ_eff_obs) & Div_Bias == TRUE & Div == FALSE], aes(x =
MQ_eff_obs, y = MQ_eff_sim)) +
  geom_point(size = 0.8, alpha = 1.0) + facet_wrap("MQyear", ncol = 3) +
  geom_abline(intercept = c(0,0), slope = 1, size=0.3) +
  scale_x_log10(limits = c(0.1, 10000), minor_breaks = NULL) + scale_y_log10(limits
  = c(0.1, 10000), minor_breaks = NULL) +
```

```
    #scale_color_manual(values = c("red","grey18")) +
    #scale_shape_manual(values = c(1,1)) +
    labs(title = paste0("Effective runoff by MQyear - only Div Affected gauges (Div=",
    data_all[Div == FALSE]$Div[1],")")) +
    xlab("obs") + ylab("sim") + coord_fixed(ratio = 1) +
    theme( plot.title = element_text(size=9), aspect.ratio = 1,
           legend.text = element_text(size=9), legend.title = element_text(size=9),
           legend.position =  c(.13,.87), legend.background = element_blank()) +

    ggsave(paste0("data/",saveDir,"/All_09_plot_year_comparison_MQ_eff_Div_FALSE_onlyDiv
    Bias.png"), width = 8.1, height = 8.4)


  ## all years compare MQ_Div and q_Div ########################################

  data_all <- data.table(ID_MORE=integer(), HZB_PEGEL1=integer(), MQyear=integer(),
  Div=logical(),
                         MQ_Div_sim=numeric(), MQ_Div_obs=numeric(),
                         q_Div_obs=numeric(), q_Div_sim=numeric(),
                         Div_Info=character(), A_Div=numeric(), Type=character(),
                         Country=character())
  files <- list.files(path=paste0("data/",saveDir,"/"),
  pattern="TRUE_07_table_STOBIMO_MQs.csv", full.names=T, recursive=FALSE,
  ignore.case=TRUE)
  for(file in files) {
    #file <- files [1]
    data = fread(file, sep = ";", dec = ",")
    #ncol(data)
    data[,`:=`(q_Div_obs = MQ_Div_obs/A_Div, q_Div_sim = MQ_Div_sim/A_Div)]
    #colnames(data)
    # bind data
    data_all = rbind(data_all,data[!is.na(MQ_Div_obs),
                                   .(ID_MORE, HZB_PEGEL1, MQyear, Div, MQ_Div_sim,
                                   MQ_Div_obs,
                                     q_Div_obs, q_Div_sim, Div_Info, A_Div, Type,
                                     Country)])
  }

  # plot MQ_Div prediction comparison all diversions
  ggplot(data = melt(data_all[,.(NSE  = round(hydroGOF::NSE (sim = MQ_Div_sim, obs =
  MQ_Div_obs),2),
                                 mNSE = round(hydroGOF::mNSE(sim = MQ_Div_sim, obs =
                                 MQ_Div_obs),2)),
                              by = .(MQyear)],
                     id.vars = c("MQyear"), variable.name = "Stat")) +
    geom_col(aes(x =  MQyear, y = value, fill = "blue"), position = "dodge") +
    facet_grid(.~Stat) +
    scale_x_continuous(limits = c(2008.5,2017.5), breaks = 2009:2017) +
    coord_cartesian(ylim = c(0.0,1.0)) +
    theme(strip.background =  element_rect(fill = "white", colour = "black"),
    legend.position =  "none",
          legend.text = element_text(size=9), legend.title = element_text(size=9),
          panel.grid.major.x = element_blank(), panel.grid.minor.x = element_blank()) +
    labs(title = "Diversion MQ runoff -  model prediction efficiency coefficient") +
    xlab(NULL) + ylab(NULL) +
    ggsave(paste0("data/",saveDir,"/All_09_plot_year_comparison_MQ_Div.png"), width =
    8.5, height = 4.7)

  # write table prediction comparison all diversions by Div Year
  fwrite(data_all[,.(NSE  = round(hydroGOF::NSE (sim = MQ_Div_sim, obs = MQ_Div_obs),2),
                     mNSE = round(hydroGOF::mNSE(sim = MQ_Div_sim, obs =
                     MQ_Div_obs),2),
                   N    = .N),
                by = .(MQyear)],
         paste0("data/",saveDir,"/All_09_table_year_comparison_MQ_Div.csv"), sep =
         ";", dec = ",")

  # write table prediction comparison all diversions by Div type
  fwrite(data_all[,.(NSE  = round(hydroGOF::NSE (sim = MQ_Div_sim, obs = MQ_Div_obs),2),
                   mNSE = round(hydroGOF::mNSE(sim = MQ_Div_sim, obs = MQ_Div_obs),2),
                   N    = .N),
                by = .(Type)],
         paste0("data/",saveDir,"/All_09_table_year_comparison_MQ_Div_type.csv"), sep
```

```
       = ";", dec = ",")

# write table prediction comparison all diversions by source of Div AREA
fwrite(data_all[,.(NSE  = round(hydroGOF::NSE (sim = MQ_Div_sim, obs = MQ_Div_obs),2),
                   mNSE = round(hydroGOF::mNSE(sim = MQ_Div_sim, obs = MQ_Div_obs),2),
                   N    = .N),
                by = .(Div_Info)],
       paste0("data/",saveDir,"/All_09_table_year_comparison_MQ_Div_MQarea.csv"),
       sep = ";", dec = ",")

# compare MQ_Div by category
ggplot(data_all, aes(x = MQ_Div_obs, y = MQ_Div_sim, color = Div_Info)) +
  geom_point(size=0.8) +  facet_wrap(vars(Type), ncol = 2) +
  geom_abline(intercept = c(0,0), slope = 1, size=0.3) +
  labs(title = "Diversion MQ runoff by diversion category for years 2009-17") +
  xlab("obs") + ylab("sim") + coord_fixed(ratio = 1) +
  scale_x_log10(limits = c(0.1, 100), minor_breaks=NULL) +
  scale_y_log10(limits = c(0.1, 100), minor_breaks=NULL) +
  scale_color_manual(values = c("red","grey18")) +
  scale_shape_manual(values = c(1,1)) + #scale_color_npg() +
  theme( plot.title = element_text(size=9), aspect.ratio = 1,
         legend.text = element_text(size=9), legend.title = element_blank(),
         legend.position =  c(.11,.95), legend.background = element_blank()) +
  ggsave(paste0("data/",saveDir,"/All_09_plot_type_comparison_MQ_Div_Stat_2.png"),
  width = 6.4, height = 6.1)
# compare q_Div by category
ggplot(data_all, aes(x = q_Div_obs, y = q_Div_sim, color = Div_Info)) +
  geom_point(size=0.8) +  facet_wrap(vars(Type), ncol = 2) +
  geom_abline(intercept = c(0,0), slope = 1, size=0.3) +
  labs(title = "Diversion MQ specific runoff by diversion category for years
  2009-17") +
  xlab("obs") + ylab("sim") + coord_fixed(ratio = 1) +
  scale_x_continuous(limits = c(0, 1.00), minor_breaks=NULL) +
  scale_y_continuous(limits = c(0, 1.00), minor_breaks=NULL) +
  scale_color_manual(values = c("red","grey18")) +
  scale_shape_manual(values = c(1,1)) + #scale_color_npg() +
  theme( plot.title = element_text(size=9), aspect.ratio = 1,
         legend.text = element_text(size=9), legend.title = element_blank(),
         legend.position =  c(.11,.95), legend.background = element_blank()) +
  ggsave(paste0("data/",saveDir,"/All_09_plot_type_comparison_q_Div_Stat_2.png"),
  width = 6.4, height = 6.1)

# identify the outliers
# MQ
plot(data_all$MQ_Div_obs,data_all$MQ_Div_sim,  log="xy")
identify(data_all$MQ_Div_obs,data_all$MQ_Div_sim,  log="xy")
# q
plot(data_all$q_Div_obs,data_all$q_Div_sim)
identify(data_all$q_Div_obs,data_all$q_Div_sim)

## compare total runoff with CV  #############################################

data_all <- data.table(MQyear=integer(), Div=logical(), ID_GAUGE=integer(),
                       obs=numeric(), CV_pred=numeric())
files <- list.files(path=paste0("data/",saveDir,"/"), pattern="_02_table_pred_CV",
full.names=T, recursive=FALSE, ignore.case=TRUE)
files <- files[!files %like% "_Stat"]
for(file in files) {
  data = fread(file, sep = ";", dec = ",")
  # bind data
  data_all = rbind(data_all,data[,.(MQyear, Div, ID_GAUGE, obs, CV_pred = var1.pred)])
}
## prepare
Qtot_CV <- MQ_rnet_gauges[,.(totalQ_10e9 = signif(sum(MQ*3600*24*YEARdays)/10^9,3)),
by = .(YEAR)]
data_all2 <- dcast(data_all[,.(NSE  = hydroGOF::NSE (sim = CV_pred, obs = obs),
                               mNSE = hydroGOF::mNSE(sim = CV_pred, obs = obs)),
                            by = .(MQyear,Div)],
                   MQyear~Div, value.var = c("NSE","mNSE"))
data_all2[, `:=`(NSE_diff = (NSE_TRUE - NSE_FALSE), mNSE_diff = (mNSE_TRUE -
mNSE_FALSE))]
Qtot_CV[, `:=`(NSE  = data_all2$NSE_diff [match(YEAR, data_all2$MQyear)],
               mNSE = data_all2$mNSE_diff[match(YEAR, data_all2$MQyear)])]
```

```
Qtot_CV_long <- melt(Qtot_CV,
                     id.vars = c("YEAR","totalQ_10e9"),
                     variable.name = "Stat",
                     value.name = c("value"))

## plot
ggplot(data = Qtot_CV_long, aes(x = value, y =  totalQ_10e9, color = Stat, shape =
Stat)) +
  geom_point(size = 2) +  #facet_grid(.~Stat) +
  geom_smooth(method = "lm", se = F, linetype = 2, size = 0.5) +
  scale_y_continuous(limits = c(650,950)) +
  theme(strip.background =  element_rect(fill = "white", colour = "black"),
  legend.position =  "bottom",
        legend.text = element_text(size=9), legend.title = element_text(size=9)) +
  labs(title = "Validation of assumptions - CV") +
  xlab("Diff between Div T/F CV Validation") + ylab("total annual runoff per year in
  km³") +
  ggsave(paste0("data/",saveDir,"/All_09_plot_year_comparison_Qtot_CV.png"), width =
  4.0, height = 3.5)

# Spearman's rank correlation test
cor.test(Qtot_CV_long[Stat == "NSE"]$totalQ_10e9,Qtot_CV_long[Stat == "NSE"]$value,
         method = "spearman")
cor.test(Qtot_CV_long[Stat == "mNSE"]$totalQ_10e9,Qtot_CV_long[Stat == "mNSE"]$value,
         method = "spearman")


## compare total runoff with MQ_eff ##########################################

data_all <- data.table(HZB_PEGEL1=integer(), MQyear=integer(), Div=logical(),
                       MQ_eff_sim=numeric(), MQ_eff_obs=numeric(), Div_Bias=logical())
files <- list.files(path=paste0("data/",saveDir,"/"),
pattern="_07_table_STOBIMO_MQs.csv", full.names=T, recursive=FALSE, ignore.case=TRUE)
for(file in files) {
  #file <- files [1]
  data = fread(file, sep = ";", dec = ",")
  #ncol(data)
  #colnames(data)
  # bind data
  data_all = rbind(data_all,data[!is.na(MQ_eff_obs), .(HZB_PEGEL1, MQyear, Div,
  MQ_eff_sim, MQ_eff_obs, Div_Bias)])
}
## prepare
Qtot_MQ_eff <- MQ_rnet_gauges[,.(totalQ_10e9 =
signif(sum(MQ*3600*24*YEARdays)/10^9,3)), by = .(YEAR)]
data_all2 <- dcast(data_all[,.(NSE  = hydroGOF::NSE (sim = MQ_eff_sim, obs =
MQ_eff_obs),
                              mNSE = hydroGOF::mNSE(sim = MQ_eff_sim, obs =
                              MQ_eff_obs)),
                            by = .(MQyear,Div)],
                   MQyear~Div, value.var = c("NSE","mNSE"))

data_all2[, `:=`(NSE_diff = (NSE_TRUE - NSE_FALSE), mNSE_diff = (mNSE_TRUE -
mNSE_FALSE))]
Qtot_MQ_eff[, `:=`(NSE  = data_all2$NSE_diff [match(Qtot_MQ_eff$YEAR,
data_all2$MQyear)],
                   mNSE = data_all2$mNSE_diff[match(Qtot_MQ_eff$YEAR,
                   data_all2$MQyear)])]
Qtot_MQ_eff_long <- melt(Qtot_MQ_eff,
                         id.vars = c("YEAR","totalQ_10e9"),
                         variable.name = "Stat",
                         value.name = c("Value"))

## plot
ggplot(data = Qtot_MQ_eff_long, aes(x = Value, y =  totalQ_10e9, color = Stat, shape
= Stat)) +
  geom_point(size = 2) +  #facet_grid(.~Stat) +
  geom_smooth(method = "lm", se = F, linetype = 2, size = 0.5) +
  scale_y_continuous(limits = c(650,950)) +
  theme(strip.background =  element_rect(fill = "white", colour = "black"),
  legend.position =  "bottom",
        legend.text = element_text(size=9), legend.title = element_text(size=9)) +
  labs(title = "Validation of assumptions - MQ_eff") +
```

```
  xlab("Diff between Div T/F MQ_eff Validation") + ylab("total annual runoff per year in km³") +
  ggsave(paste0("data/",saveDir,"/All_09_plot_year_comparison_Qtot_MQ_eff.png"), width = 4.0, height = 3.5)

# Spearman's rank correlation test
cor.test(Qtot_MQ_eff_long[Stat == "NSE"]$totalQ_10e9,Qtot_MQ_eff_long[Stat == "NSE"]$Value,
         method = "spearman")
cor.test(Qtot_MQ_eff_long[Stat == "mNSE"]$totalQ_10e9,Qtot_MQ_eff_long[Stat == "mNSE"]$Value,
         method = "spearman")

## MQ_eff runoff prediction comparison of each gauge MQ_eff #################

data_all <- data.table(ID_MORE=integer(), HZB_PEGEL1=integer(), MQyear=integer(), Div=logical(),
                       MQ_eff_sim=numeric(), MQ_eff_obs=numeric(), Div_Bias=logical())
files <- list.files(path=paste0("data/",saveDir,"/"), pattern="_07_table_STOBIMO_MQs.csv", full.names=T, recursive=FALSE, ignore.case=TRUE)
for(file in files) {
  #file <- files [1]
  data = fread(file, sep = ";", dec = ",")
  #ncol(data)
  #colnames(data)
  # bind data
  data_all = rbind(data_all,data[!is.na(MQ_eff_obs), .(ID_MORE, HZB_PEGEL1, MQyear, Div, MQ_eff_sim, MQ_eff_obs, Div_Bias)])
}

data_diff <- dcast(data_all[Div == TRUE,
                            .(HZB_PEGEL1, MQeff_Diff_p = round((MQ_eff_obs-MQ_eff_sim)/MQ_eff_obs*100,0)),
                            by = .(ID_MORE, MQyear)],
                   HZB_PEGEL1+ID_MORE~MQyear, value.var = c("MQeff_Diff_p"))

## write table with correlation for each gauge
fwrite(data_diff,
       paste0("data/",saveDir,"/All_11_table_cor_MQ_eff_per_gauge.csv"), sep = ";", dec = ",")

## End result comparison ###################################################
```

C.4.4 Data Export to MoRE

```
##%###############################################%##
#                                                   #
#   Diploma Thesis                                  #
#   TopKriging prediction with                      #
#   diversion consideration                         #
#                                                   #
#   Export to MoRE Model                            #
#   Creator:                                        #
#   nikolaus.weber@tuwien.ac.at                     #
#   Last edit:                                      #
#   02.12.2020 by Nikolaus Weber                    #
#                                                   #
##%###############################################%##


## Libs ####################################################################

if (!require("dplyr")) install.packages("dplyr", dependencies = TRUE, repos="https://cloud.r-project.org/")
if (!require("data.table")) install.packages("data.table", dependencies = TRUE, repos="https://cloud.r-project.org/")

library(dplyr)
library(data.table)


## create Input for MoRE Model ##############################################

## set Input Parameter
saveDir <- "2020-12-01_STOBIMO_all_V46" # e.g. "2020-11-18_STOBIMO_2009_V43"

## load diversions MORE table
gauge_DIV <- setDT(read_excel("./data/Diversion_data/Diversion_data_table.xlsx", sheet = "DIV_MORE", na = "NA"))
AUs_noSplit <- gauge_DIV[Div2to_ID == "T" | hyd_short == "T", ID_MORE] # exclude AUs with hydraulic short circuit & TO_ID == TO_2_ID


## Create BI_Q_net for MoRE-Input ############################################

data_all <- data.table(FlächenId = integer(), Jahr = numeric(), Variable = character(), Wert = numeric(),
                       Variante = integer(), Name_Eingangsdatensatz = character(), Datum = character())
files <- list.files(path=paste0("data/",saveDir,"/"), pattern="TRUE_07_table_STOBIMO_MQs.csv", full.names=T, recursive=FALSE, ignore.case=TRUE)
#file <- files[1]
for(file in files) {
  data_MQ = fread(file, sep = ";", dec = ",")
  data_MQ = data_MQ[,.(FlächenId = ID_MORE, Jahr = MQyear, Variable = "BI_Q_net", Wert = round(MQ_sim_cor,4), Variante = 2,
                  Name_Eingangsdatensatz = "TK_NW", Datum = as.character(format(Sys.Date(), "%d.%m.%Y")))]
  # bind data
  data_all = rbind(data_all,data_MQ)
}


## Create RM_FCT_Q_SPLIT for MoRE-Input ######################################

year_min <- min(data_all$Jahr)
year_max <- max(data_all$Jahr)

# load SplittingFactor & upstream AUs
SF_upAU <- fread("data/Diversion_data/STOBIMO_SF_Q.Split.csv", sep = ";", dec = ",")
#year <- 2009
for(year in year_min:year_max) {
  data_SF = SF_upAU[,.(FlächenId = from_id, Jahr = year, Variable = "RM_FCT_Q_SPLIT", Wert = round(SF_Q.Split,4), Variante = 2,
                     Name_Eingangsdatensatz = "TK_NW", Datum = as.character(format(Sys.Date(), "%d.%m.%Y")))]
  data_SF[FlächenId %in% AUs_noSplit, Wert := 0] # remove Splitting Factor for AUs
```

```
    with hydraulic short circuit & TO_ID == TO_2_ID
    data_SF[FlächenId %in% "10070", Wert := 0] # remove Splitting Factor for AU 10070
    because in MoRE it is different than in my model
    # bind data
    data_all = rbind(data_all,data_SF)
  }
  str(data_all)


  ## Export for MORE model input ##################################################

  fwrite(data_all, file=paste0("data/",saveDir,"/All_10_Export2MORE_zeitbezogene
  AU-Variablen.csv"),
         sep = ";", dec = ",")


  ## End Export to MoRE Model #####################################################

```